# A Comparison of Selection Schemes used in Genetic Algorithms

Tobias Blickle and Lothar Thiele
Computer Engineering and Communication Networks Lab (TIK)
Swiss Federal Institute of Technology (ETH)
Gloriastrasse 35, 8092 Zurich
Switzerland
{blickle,thiele}@tik.ee.ethz.ch

## Abstract

Genetic 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 a new description model is introduced in this paper. With this a mathematical analysis of tournament selection, truncation selection, linear and exponential ranking selection and proportional selection is carried out that allows an exact prediction of the fitness values after selection. The further analysis derives the selection intensity, selection variance, and the loss of diversity for all selection schemes. For completion a pseudo-code formulation of each method is included. The selection schemes are compared and evaluated according to their properties leading to an unified view of these different selection schemes. Furthermore the correspondence of binary tournament selection and ranking selection in the expected fitness distribution is proven.

# Foreword

This paper is the revised and extended version of the TIK-Report No. 11 from April, 1995. The main additions to the first edition are the analysis of exponential ranking selection and proportional selection. Proportional selection is only included for completeness - we believe that it is a very unsuited selection method and we will show this (like it has be done by other researchers, too) based on a mathematical analysis in chapter 7. Furthermore for each selection scheme a pseudo-code notation is given and a short remark on time complexity is included.

The main correction concerns the approximation formula for the selection variance of tournament selection. The approximation given in the first edition was completely wrong. In this report the approximation formula is derived by a genetic algorithm, or better speaking by the genetic programming optimization method. The used method is described in appendix A and also applied to derive an analytic approximation for the selection intensity and selection variance of exponential ranking selection.

We hope that this report summarizes the most important facts for these five selection schemes and gives all researches a well founded basis to chose the appropriate selection scheme for their purpose.

Tobias Blickle                                               Zürich, Dec., 1995

1

# Contents

# Chapter 1

# Introduction

*Genetic Algorithms (GA)* are probabilistic search algorithms characterized by the fact that a number $N$ of potential solutions (called *individuals $J_i \in \mathbf{J}$*, where $\mathbf{J}$ represents the space of all possible individuals) of the optimization problem simultaneously sample the search space. This *population $P = \{J_1, J_2, ..., J_N\}$* is modified according to the natural evolutionary process: after initialization, selection $\omega : \mathbf{J}^N \mapsto \mathbf{J}^N$ and recombination $\Xi : \mathbf{J}^N \mapsto \mathbf{J}^N$ are executed in a loop until some termination criterion is reached. Each run of the loop is called a generation and $P(\tau)$ denotes the population at generation $\tau$.

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 $f : \mathbf{J} \mapsto \mathbf{R}$. Recombination changes the genetic material in the population either by crossover or by mutation in order to exploit new points in the search space.

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. As this balance is critical for the behavior of the GA it is of great interest to know the properties of the selection and recombination operators to understand their influence on the convergence speed.

Some work has been done to classify the different selection schemes such as *proportionate selection, ranking selection, tournament selection.* Goldberg [Goldberg and Deb, 1991] introduced the term of *takeover time*. The takeover time is the number of generations that is needed for a single best individual to fill up the whole generation if no recombination is used. Recently Bäck [Bäck, 1994] has analyzed the most prominent selection schemes used in Evolutionary Algorithms with respect to their takeover time. In [Mühlenbein and Schlierkamp-Voosen, 1993] the *selection intensity* in the so called *Breeder Genetic Algorithm (BGA)* is used to measure the progress in the population. The selection intensity is derived for proportional selection and truncation selection. De la Maza and Tidor [de la Maza and Tidor, 1993] analyzed several selection methods according

4

to their scale and translation invariance.

An analysis based on the behavior of the best individual (as done by Goldberg and Bäck) or on the average population fitness (as done by Mühlenbein) only describes one aspect of a selection method. In this paper a selection scheme is described by its interaction on the distribution of fitness values. Out of this description several properties can be derived, e.g. the behavior of the best or average individual. The description is introduced in the next chapter. In chapter 3 an analysis of the tournament selection is carried out and the properties of the tournament selection are derived. The subsequent chapters deal with truncation selection, ranking selection, and exponential ranking selection. Chapter 7 is devoted to proportional selection that represents some kind of exception to the other selection schemes analyzed in this paper. Finally all selection schemes are compared.

# Chapter 2

# Description of Selection Schemes

In this chapter we introduce a description of selection schemes that will be used in the subsequent chapters to analyze and compare several selection schemes, namely tournament selection, truncation selection, and linear and exponential ranking selection and fitness proportional selection. The description is based on the fitness distribution of the population before and after selection as introduced in [Blickle and Thiele, 1995]. It is assumed that selection and recombination are done sequentially: first a selection phase creates an intermediate population $P'(\tau)$ and then recombination is performed with a certain probability $p_c$ on the individuals of this intermediate population to get the population for the next generation (Fig. 2.1). Recombination includes crossover and mutation or any other operator that changes the "genetic material". This kind of description differs from the common paradigms where selection is made to obtain the individuals for recombination [Goldberg, 1989; Koza, 1992]. But it is mathematically equivalent and allows to analyze the selection method separately.

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

**Definition 2.0.1 (Fitness distribution)** *The function $s : \mathbf{R} \mapsto Z_0^+$ assigns to each fitness value $f \in \mathbf{R}$ the number of individuals in a population $P \in \mathbf{J}^N$ carrying this fitness value. $s$ is called the* fitness distribution *of a population $P$.*

The characterization of the population by its fitness distribution has also been used by other researches, but in a more informal way. In [Mühlenbein and Schlierkamp-Voosen, 1993] the fitness distribution is used to calculate some properties of truncation selection. In [Shapiro *et al.*, 1994] a statistical mechanics approach is taken to describe the dynamics of a Genetic Algorithm that makes use of fitness distributions, too.
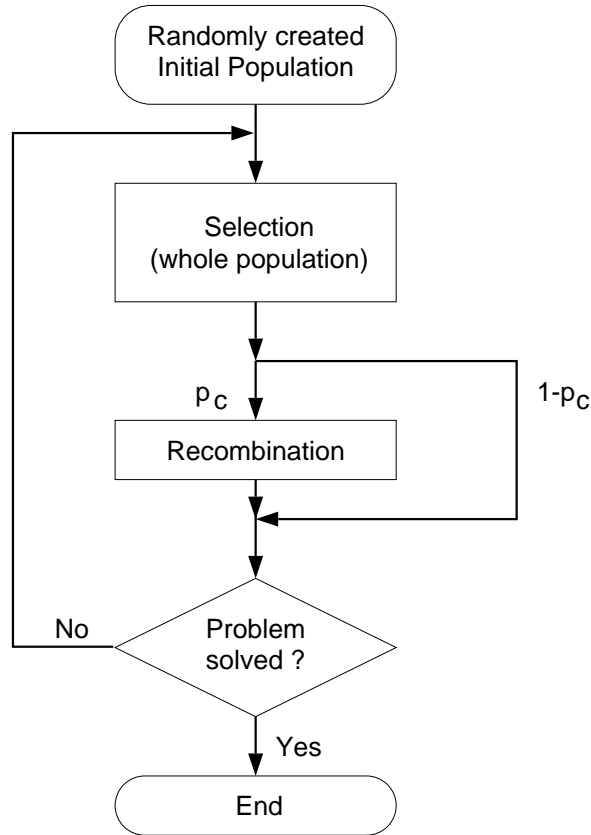
Figure 2.1: Flowchart of the Genetic Algorithm.

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

**Definition 2.0.2 (Selection method)** *A selection method $\Omega$ is a function that transforms a fitness distribution s into an new fitness distribution s':*

$$s' = \Omega(s, par\_list) \tag{2.1}$$

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

As the selection methods are probabilistic we will often make use of the expected fitness distribution.

**Definition 2.0.3 (Expected fitness distribution)** *$\Omega^*$ denotes the expected fitness distribution after applying the selection method $\Omega$ to the fitness distribution s, i.e.*

$$\Omega^*(s, par\_list) = E(\Omega(s, par\_list)) \tag{2.2}$$

*The notation $s^* = \Omega^*(s, par\_list)$ will be used as abbreviation.*

7

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

**Theorem 2.0.1** *The variance in obtaining the fitness distribution $s'$ is*

$$\sigma_s^2 = s^* \left( 1 - \frac{s^*}{N} \right) \tag{2.3}$$

**Proof:** $s^*(f_i)$ denotes the expected number of individuals with fitness value $f_i$ after selection. It is obtained by doing $N$ experiments "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 $p_i = \frac{s^*(f_i)}{N}$. To each fitness value there exists a Bernoulli trial "an individual with fitness $f_i$ is selected". As the variance of a Bernoulli trial with $N$ trials is given by $\sigma^2 = Np(1-p)$, (2.3) is obtained using $p_i$. $\qquad\square$

The index $s$ in $\sigma_s$ stands for "sampling" as it is the mean variance due to the sampling of the finite population.

The variance of (2.3) 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. We will introduce Baker's "stochastic universal sampling" algorithm (SUS) [Baker, 1987], which is an optimal sampling algorithm when we compare the different selection schemes in chapter 8.

**Definition 2.0.4 (Cumulative fitness distribution)** *Let $n$ be the number of unique fitness values and $f_1 < ... < f_{n-1} < f_n$ $(n \le N)$ the ordering of the fitness values with $f_1$ denoting the worst fitness occurring in the population and $f_n$ denoting the best fitness in the population.*

*$S(f_i)$ denoting the number of individuals with fitness value $f_i$ or worse and is called* cumulative fitness distribution, *i.e.*

$$S(f_i) = \begin{cases} 0 & : \quad i < 1 \\ \sum_{j=1}^{j=i} s(f_j) & : \quad 1 \le i \le n \\ N & : \quad i > n \end{cases} \tag{2.4}$$

**Example 2.0.1** *As an example of a discrete fitness distribution we use the initial fitness distribution of the "wall-following-robot" from Koza [Koza, 1992]. This distribution is typical of problems solved by genetic programming (many bad and only very few good individuals exist). Figure 2.2 shows the distribution $s(f)$ (left) and the cumulative distribution $S(f)$ (right).*

We will now describe the distribution $s(f)$ as a continuous distribution $\bar{s}(f)$ allowing the following properties to be easily derived. To do so, we assume
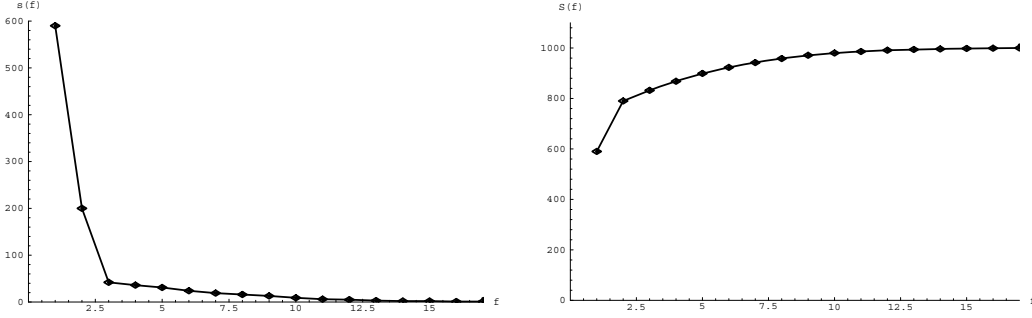
Figure 2.2: The fitness distribution $s(f)$ and the cumulative fitness distribution $S(f)$ for the "wall-following-robot" problem.

continuous distributed fitness values. The range of the function $\bar{s}(f)$ is $f_0 < f \leq f_n$, using the same notation as in the discrete case.

We denote all functions in the continuous case with a bar, e.g. we write $\bar{s}(f)$ instead of $s(f)$. Similar sums are replaced by integrals, for example

$$\bar{S}(f) = \int_{f_0}^{f} \bar{s}(x)\,dx \tag{2.5}$$

denotes the continuous cumulative fitness distribution.

**Example 2.0.2** *As an example for a continuous fitness distribution we chose the Gaussian distribution $G(\mu, \sigma)$ with*

$$G(\mu, \sigma)(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.6}$$

*The distribution $\bar{s}_G(f) = NG(\mu, \sigma)(f)$ with $\sigma = 30, \mu = 100, N = 1000$ and $f_0 = -\infty, f_n = +\infty$ is shown in the interesting region $f \in [0, 200]$ in Figure 2.3 (left). The right graph in this figure shows the cumulative fitness distribution $\bar{S}_G(f)$.*

We will now introduce the aspects of the fitness distribution we want to compare. The definitions given will all refer to continuous distributed fitness values.

## 2.1 Average Fitness

**Definition 2.1.1 (Average fitness)** *$\bar{M}$ denotes the average fitness of the population before selection and $\bar{M}^*$ denotes the expected average fitness after selection:*

$$\bar{M} = \frac{1}{N} \int_{f_0}^{f_n} \bar{s}(f)\,f\,df \tag{2.7}$$

$$\bar{M}^* = \frac{1}{N} \int_{f_0}^{f_n} \bar{s}^*(f)\,f\,df \tag{2.8}$$

9
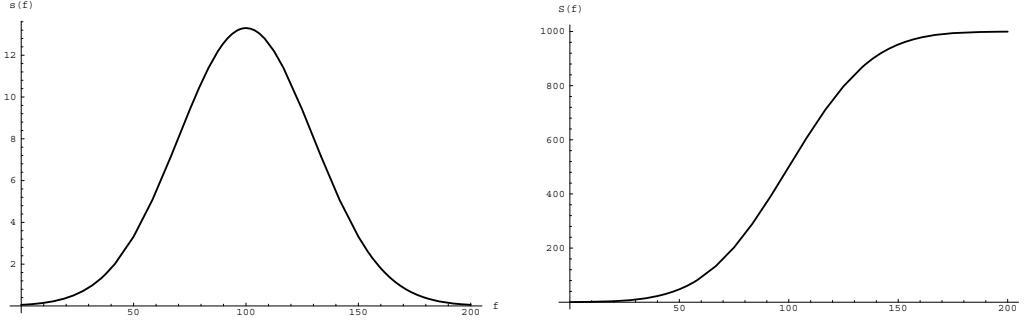
Figure 2.3: The fitness distribution $\bar{s}_G(f)$ (left) and the cumulative fitness distribution $\bar{S}_G(f)$ (right).

## 2.2    Fitness Variance

**Definition 2.2.1 (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:*

$$\bar{\sigma}^2 = \frac{1}{N} \int_{f_0}^{f_n} \bar{s}(f)\,(f - \bar{M})^2\,df = \frac{1}{N} \int_{f_0}^{f_n} f^2 \bar{s}(f)\,df - \bar{M}^2 \tag{2.9}$$

$$(\bar{\sigma}^*)^2 = \frac{1}{N} \int_{f_0}^{f_n} \bar{s}^*(f)\,(f - \bar{M}^*)^2\,df = \frac{1}{N} \int_{f_0}^{f_n} f^2 \bar{s}^*(f)\,df - \bar{M}^{*2} \tag{2.10}$$

Note the difference of this variance to the variance in obtaining a certain fitness distribution characterized by theorem 2.0.1

## 2.3    Reproduction Rate

**Definition 2.3.1 (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*

$$\bar{R}(f) = \left\{ \begin{array}{ccc} \frac{\bar{s}^*(f)}{\bar{s}(f)} & : & \bar{s}(f) > 0 \\ 0 & : & \bar{s}(f) = 0 \end{array} \right. \tag{2.11}$$

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 $\bar{R}(f) < 1$.

10

## 2.4 Loss of Diversity

During every selection phase bad individuals will be lost and be 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 2.4.1 (Loss of diversity)** *The* loss of diversity $p_d$ *is the proportion of individuals of a population that is not selected during the selection phase.*

**Theorem 2.4.1** *If the reproduction rate $\bar{R}(f)$ increases monotonously in $f$, the loss of diversity of a selection method is*

$$p_d = \frac{1}{N} \left( \bar{S}(f_z) - \bar{S}^*(f_z) \right) \tag{2.12}$$

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

**Proof:** For all fitness values $f \in (f_0, f_z]$ the reproduction rate is less than one. 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

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

$\square$

The loss of diversity should be as low as possible because a high loss of diversity increases the risk of premature convergence.

In his dissertation [Baker, 1989], Baker has introduced a similar measure called "reproduction rate RR". RR gives the percentage of individuals that is selected to reproduce, hence $RR = 100(1 - p_d)$.

## 2.5 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 [Goldberg and Deb, 1991] and Bäck [Bäck, 1994] use the "takeover time" to define the selection pressure. Whitley calls the parameter $c$ (see chapter 5) of his ranking selection method selection pressure.

We use the term "selection intensity" in the same way it is used in population genetic [Bulmer, 1980]. Mühlenbein has adopted the definition and applied it to genetic algorithms [Mühlenbein and Schlierkamp-Voosen, 1993]. Recently more and more researches are using this term to characterize selection schemes [Thierens and Goldberg, 1994a; Thierens and Goldberg, 1994b; Bäck, 1995; Blickle and Thiele, 1995].

The change of the average fitness of the population due to selection is a reasonable measure for selection intensity. In population genetic the term selection intensity was introduced to obtain a normalized and dimension-less measure. The idea is to measure the progress due to selection by the so called "selection differential", i.e. the difference between the population average fitness after and before selection. Dividing this selection differential by the mean variance of the population fitness leads to the desired dimension-less measure that is called the selection intensity.

**Definition 2.5.1 (Selection intensity)** *The* selection intensity *of a selection method $\Omega$ for the fitness distribution $\bar{s}(f)$ is the standardized quantity*

$$I = \frac{\bar{M}^* - \bar{M}}{\bar{\sigma}} \tag{2.13}$$

By this, 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 $G(0,1)$ as initial fitness distribution leads to the following definition.

**Definition 2.5.2 (Standardized selection intensity)** *The* standardized selection intensity $I_\Omega$ *is the expected average fitness value of the population after applying the selection method $\Omega$ to the normalized Gaussian distribution $G(0,1)(f) = \frac{1}{\sqrt{2\pi}}e^{-\frac{f^2}{2}}$:*

$$I_\Omega = \int_{-\infty}^{\infty} f \, \overline{\Omega}^*(G(0,1))(f) \, df \tag{2.14}$$

The "effective" average fitness value of a Gaussian distribution with mean $\mu$ and variance $\sigma^2$ can easily be derived as $\bar{M}^* = \sigma I + \mu$. 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 except proportional selection. 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 2.5.1. In the remainder of this paper we use the term "selection intensity" as equivalent for "standardized selection intensity" as our intention is the comparison of selection schemes.

## 2.6   Selection Variance

In addition to the selection intensity we introduce the term of "selection variance". The definition is analogous to the definition of the selection intensity, but here we are interested in the the new variance of the fitness distribution after selection.

**Definition 2.6.1 (Selection variance)** *The* selection variance *is the normalized expected variance of the fitness distribution of the population after applying the selection method $\Omega$ to the fitness distribution $\bar{s}(f)$, i.e.*

$$V = \frac{(\bar{\sigma}^*)^2}{\bar{\sigma}^2} \qquad (2.15)$$

For comparison the standardized selection variance is of interest.

**Definition 2.6.2 (Standardized selection variance)** *The* standardized selection variance $V_\Omega$ *is the normalized expected variance of the fitness distribution of the population after applying the selection method $\Omega$ to the normalized Gaussian distribution $G(0,1)$.*

$$V_\Omega = \int_{-\infty}^{\infty} (f - I_\Omega)^2 \, \overline{\Omega}^*(G(0,1))(f) \, df \qquad (2.16)$$

*that is equivalent to*

$$V_\Omega = \int_{-\infty}^{\infty} f^2 \, \overline{\Omega}^*(G(0,1))(f) \, df - I_\Omega^2 \qquad (2.17)$$

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 reports a decrease in variance. The lowest possible value of $V_\Omega$ is zero, which means that the variance of the fitness values of population after selection is itself zero. Again we will use the term the "selection variance" as equivalent for "standardized selection variance".

13

# Chapter 3

# Tournament Selection

Tournament selection works as follows: Choose some number $t$ of individuals randomly from the population and copy the best individual from this group into the intermediate population, and repeat $N$ times. Often tournaments are held only between two individuals (binary tournament) but a generalization is possible to an arbitrary group size $t$ called *tournament size*.

The pseudo code of tournament selection is given by algorithm 1.

---

Algorithm 1: (Tournament Selection)

**Input:** The population $P(\tau)$ the tournament size $t \in \{1, 2, ..., N\}$
**Output:** The population after selection $P(\tau)'$

tournament($t, J_1, ..., J_N$):
    **for** $i \leftarrow 1$ **to** $N$ **do**
        $J_i' \leftarrow$ best fit individual out of $t$ randomly picked
        individuals from $\{J_1, ..., J_N\}$;
    **od**
    **return** $\{J_1', ..., J_N'\}$

---

The outline of the algorithm shows that tournament selection can be implemented very efficiently as no sorting of the population is required. Implemented in the way above it has the time complexity $\mathcal{O}(N)$.

Using the notation introduced in the previous chapter, the entire fitness distribution after selection can be predicted. The prediction will be made for the discrete (exact) fitness distribution as well as for a continuous fitness distribution. These results were first published in [Blickle and Thiele, 1995]. The calculations assume that tournament selection is done with replacement.

**Theorem 3.0.1** *The expected fitness distribution after performing tournament*

*selection with tournament size t on the distribution s is*

$$\Omega_T^*(s,t)(f_i) = s^*(f_i) = N\left(\left(\frac{S(f_i)}{N}\right)^t - \left(\frac{S(f_{i-1})}{N}\right)^t\right) \tag{3.1}$$

**Proof:** We first calculate the expected number of individuals with fitness $f_i$ or worse, i.e. $S^*(f_i)$. An individual with fitness $f_i$ or worse can only win the tournament if all other individuals in the tournament have a fitness of $f_i$ or worse. This means we have to calculate the probability that all $t$ individuals have a fitness of $f_i$ or worse. As the probability to choose an individual with fitness $f_i$ or worse is given by $\frac{S(f_i)}{N}$ we get

$$S^*(f_i) = N\left(\frac{S(f_i)}{N}\right)^t \tag{3.2}$$

Using this equation and the relation $s^*(f_i) = S^*(f_i) - S^*(f_{i-1})$ (see Definition 2.0.4) we obtain (3.1). $\qquad\square$

Equation (3.1) shows the strong influence of the tournament size $t$ on the behavior of the selection scheme. Obviously for $t = 1$ we obtain (in average) the unchanged initial distribution as $\Omega_T^*(s,1)(f_i) = N\left(\frac{S(f_i)}{N} - \frac{S(f_{i-1})}{N}\right) = S(f_i) - S(f_{i-1}) = s(f_i)$.

In [Bäck, 1994] the probability for the individual number $i$ to be selected by tournament selection is given by $p_i = N^{-t}((N - i + 1)^t - (N - i)^t)$, under the assumption that the individuals are ordered according to their fitness value $f(J_1) \leq f(J_2) \leq ... \leq f(J_N)$. Note that Bäck uses an "reversed" fitness function where the best individual has the lowest index. For comparison with our results we transform the task into an maximization task using $j = N - i + 1$:

$$p_j = N^{-t}(j^t - (j-1)^t) \qquad 1 \leq j \leq N \tag{3.3}$$

This formula is as a special case of (3.1) with *all individuals having a different fitness value*. Then $s(f_i) = 1$ for all $i \in [1, N]$ and $S(f_i) = i$ and $p_i = \frac{s^*(f_i)}{N}$ yields the same equation as given by Bäck. Note that (3.3) is not valid if some individuals have the same fitness value.

**Example 3.0.1** *Using the discrete fitness distribution from Example 2.0.1 (Figure 2.2) we obtain the fitness distribution shown in Figure 3.1 after applying tournament selection with a tournament size $t = 10$. In addition to the expected distribution there are also the two graphs shown for $s^*(f) - \sigma_s(f)$ and $s^*(f) + \sigma_s(f)$. Hence a distribution obtained from one tournament run will lie in the given interval (the confidence interval) with a probability of 68%.*

*The high agreement between the theoretical derived results and a simulation is verified in Figure 3.2. Here the distributions according to (3.1) and the average of 20 simulation are shown.*
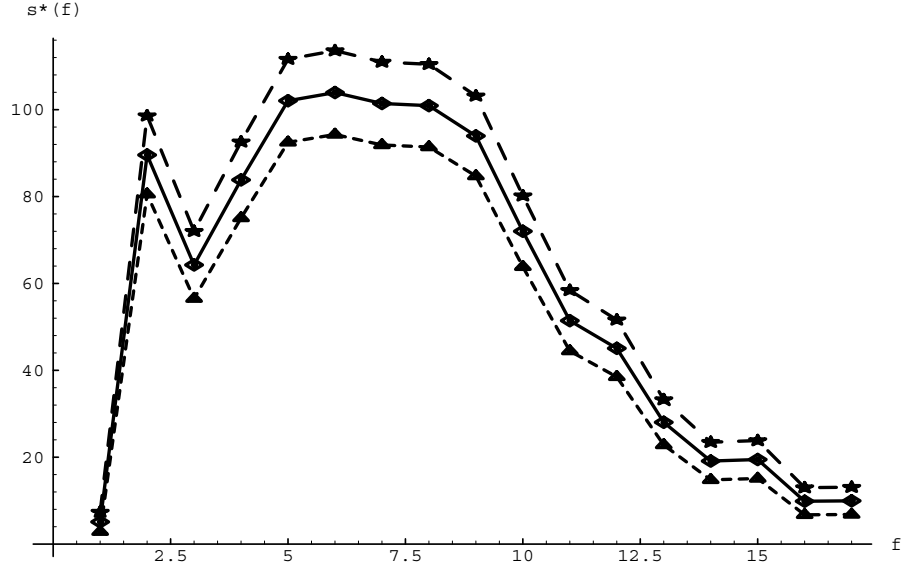
Figure 3.1: The resulting expected fitness distribution and the confidence interval of 68% after applying tournament selection with a tournament size of 10.

In example 3.0.1 we can see a very high variance in the distribution that arises from fact that the individuals are selected in $N$ independent trials. In chapter 8.1 we will meet the so called "stochastic universal sampling" method that minimizes this mean variance.

**Theorem 3.0.2** *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*

$$\overline{\Omega}_T^*(\bar{s}, t))(f) = \bar{s}^*(f) = t\bar{s}(f)\left(\frac{\bar{S}(f)}{N}\right)^{t-1} \tag{3.4}$$

**Proof:** Analogous to the proof of the discrete case the probability of an individual with fitness $f$ or worse to win the tournament is given by

$$\bar{S}^*(f) = N\left(\frac{\bar{S}(f)}{N}\right)^t \tag{3.5}$$

As $\bar{s}^*(f) = \frac{d\bar{S}^*(f)}{df}$, we obtain (3.4). $\qquad\qquad\square$

**Example 3.0.2** *Figure 3.3 shows the resulting fitness distributions after applying tournament selection on the Gaussian distribution from Example 2.0.2.*
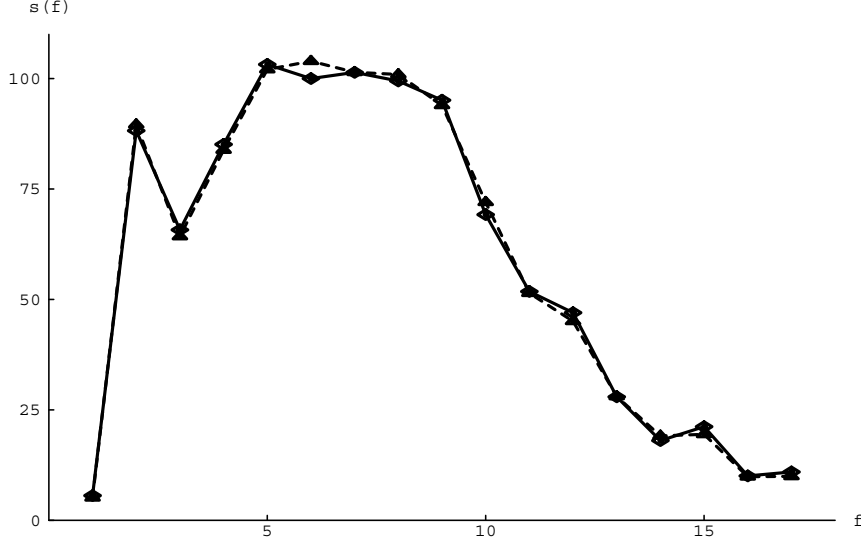
16

Figure 3.2: Comparison between theoretical derived distribution (—) and simulation (- - -) for tournament selection (tournament size $t = 10$).

## 3.1 Concatenation of Tournament Selection

An interesting property of the tournament selection is the concatenation of several selection phases. Assume an arbitrary population with the fitness distribution $\bar{s}$. We apply first tournament selection with tournament size $t_1$ 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 the tournament size $t_1 t_2$ is applied to the initial distribution $\bar{s}$.

**Theorem 3.1.1** *Let $\bar{s}$ be a continuous fitness distribution and $t_1, t_2 \geq 1$ two tournament sizes. Then the following equation holds*

$$\overline{\Omega}_T^*(\overline{\Omega}_T^*(\bar{s}, t_1), t_2)(f) = \overline{\Omega}_T^*(\bar{s}, t_1 t_2)(f) \tag{3.6}$$

**Proof:**

$$\overline{\Omega}_T^*(\overline{\Omega}_T^*(\bar{s}, t_1), t_2)(f) = t_2 \overline{\Omega}_T^*(\bar{s}, t_1)(f) \left( \frac{1}{N} \int_{f_0}^{f} \overline{\Omega}_T^*(\bar{s}, t_1)(x)\, dx \right)^{t_2 - 1}$$

$$= t_2 t_1 \bar{s}(f) \left( \frac{1}{N} \int_{f_0}^{f} \bar{s}(x)\, dx \right)^{t_1 - 1} \left( \frac{1}{N} \int_{f_0}^{f} t_1 \bar{s}(x) \left( \frac{1}{N} \int_{f_0}^{x} \bar{s}(y)\, dy \right)^{t_1 - 1} dx \right)^{t_2 - 1}$$

As

$$\int_{f_0}^{f} t_1 \bar{s}(x) \left( \frac{1}{N} \int_{f_0}^{x} \bar{s}(y)\, dy \right)^{t_1 - 1} dx = N \left( \frac{1}{N} \int_{f_0}^{f} \bar{s}(x)\, dx \right)^{t_1}$$

17
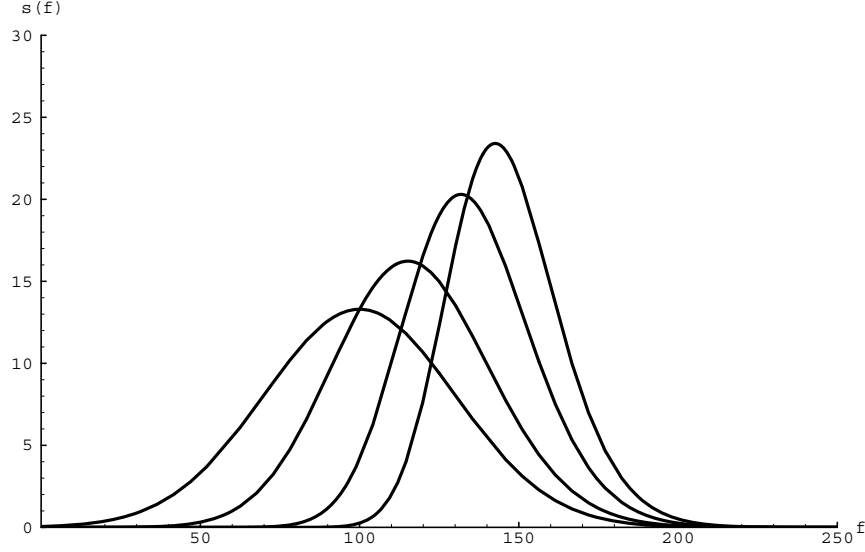
Figure 3.3: Gaussian fitness distribution approximately leads again to Gaussian distributions after tournament selection (from left to right: initial distribution, $t = 2$, $t = 5$, $t = 10$).

we can write

$$\overline{\Omega}_T^*(\overline{\Omega}_T^*(\bar{s}, t_1), t_2)(f) = t_2 t_1 \bar{s}(f) \left( \frac{1}{N} \int_{f_0}^f \bar{s}(x)\, dx \right)^{t_1 - 1} \left( \left( \frac{1}{N} \int_{f_0}^f \bar{s}(x)\, dx \right)^{t_1} \right)^{t_2 - 1}$$

$$= t_2 t_1 \bar{s}(f) \left( \frac{1}{N} \int_{f_0}^f \bar{s}(x)\, dx \right)^{t_1 - 1} \left( \frac{1}{N} \int_{f_0}^f \bar{s}(x)\, dx \right)^{t_1(t_2 - 1)}$$

$$= t_2 t_1 \bar{s}(f) \left( \frac{1}{N} \int_{f_0}^f \bar{s}(x)\, dx \right)^{t_1 t_2 - 1} = \overline{\Omega}_T^*(\bar{s}, t_1\, t_2)(f)$$

$\square$

In [Goldberg and Deb, 1991] the proportion $P_\tau$ of best-fit individuals after $\tau$ selections with tournament size $t$ (without recombination) is given to

$$P_\tau = 1 - (1 - P_0)^{t^\tau} \tag{3.7}$$

This can be obtained as a special case from Theorem 3.1.1, if only the best-fit individuals are considered.

**Corollary 3.1.1** *Let $\bar{s}(f)$ be a fitness distribution representable as*

$$\bar{s}(f) = \beta g(f) \left( \frac{\int_{f_0}^f g(x)\, dx}{N} \right)^{\beta - 1} \tag{3.8}$$

18

with $\beta \geq 1$ and $\int_{f_0}^{f_n} g(x)\,dx = N$. Then the expected distribution after tournament with tournament size $t$ is

$$\bar{s}^*(f) = \beta\,t\,g(f)\left(\frac{\int_{f_0}^{f} g(x)\,dx}{N}\right)^{\beta t - 1} \tag{3.9}$$

**Proof:** If we assume that $\bar{s}(f)$ is the result of applying tournament selection with tournament size $\beta$ on the distribution $g(f)$, (3.9) is directly obtained using Theorem 3.1.1. $\qquad\qquad\square$

## 3.2 Reproduction Rate

**Corollary 3.2.1** *The reproduction rate of tournament selection is*

$$\bar{R}_T(f) = \frac{\bar{s}^*(f)}{\bar{s}(f)} = t\left(\frac{\bar{S}(f)}{N}\right)^{t-1} \tag{3.10}$$

This is directly obtained by substituting (3.4) in (2.11).

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 3.3.1** *The loss of diversity $p_{d,T}$ of tournament selection is*

$$p_{d,T}(t) = t^{-\frac{1}{t-1}} - t^{-\frac{t}{t-1}} \tag{3.11}$$

**Proof:** $\bar{S}(f_z)$ can be determined using (3.10) (refer to Theorem 2.4.1 for the definition of $f_z$):

$$\bar{S}(f_z) = N\,t^{-\frac{1}{t-1}} \tag{3.12}$$

Using Definition 2.4.1 and (3.12) we obtain:

$$
\begin{aligned}
p_{d,T}(t) &= \frac{1}{N}\left(\bar{S}(f_z) - \bar{S}^*(f_z)\right)\\
&= \frac{\bar{S}(f_z)}{N} - \left(\frac{\bar{S}(f_z)}{N}\right)^{t}\\
&= t^{-\frac{1}{t-1}} - t^{-\frac{t}{t-1}}
\end{aligned}
$$

$$\square$$

It turns out that the number of individuals lost increases with the tournament size (see Fig. 3.4). About the half of the population is lost at tournament size $t = 5$.
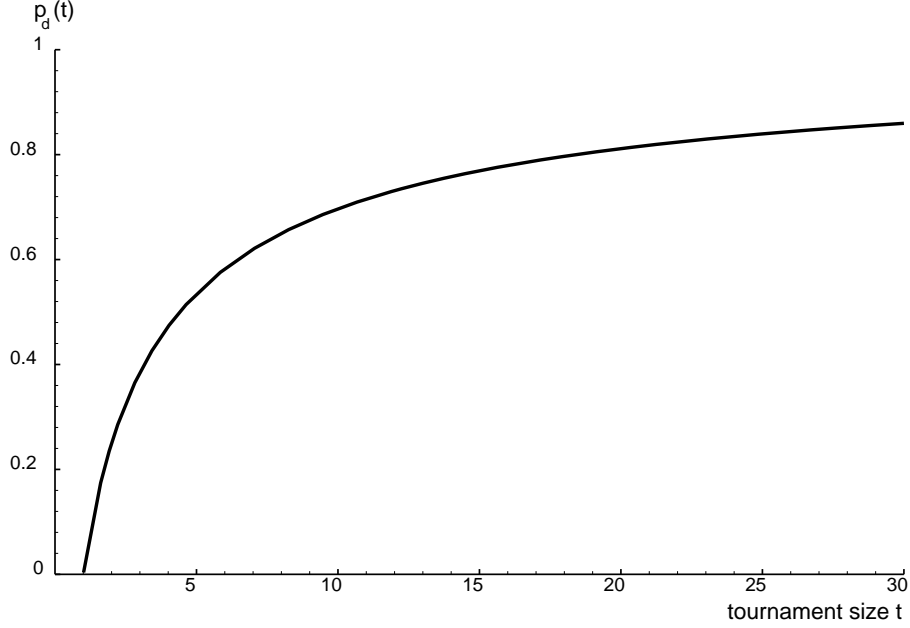
Figure 3.4: The loss of diversity $p_{d,T}(t)$ for tournament selection.

## 3.4 Selection Intensity

To calculate the selection intensity we calculate the average fitness of the population after applying tournament selection on the normalized Gaussian distribution $G(0,1)$. Using Definition 2.1.1 we obtain

$$I_T(t) = \int_{-\infty}^{\infty} t\, x\, \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left( \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right)^{t-1} dx \qquad (3.13)$$

These integral equations can be solved analytically for the cases $t = 1, \ldots, 5$ ([Blickle and Thiele, 1995; Bäck, 1995; Arnold *et al.*, 1992]):

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

For a tournament size of two Thierens and Goldberg derive the same average fitness value [Thierens and Goldberg, 1994a] in a completely different manner. But their formulation can not be extended to other tournament sizes.

For larger tournament sizes (3.13) can be accurately evaluated by numerical integration. The result is shown on the left side of Figure 3.5 for a tournament size from 1 to 30. But an explicit expression of (3.13) may not exist. By means of the steepest descent method (see, e.g. [Henrici, 1977]) an approximation for large tournament sizes can be given. But even for small tournament sizes this approximation gives acceptable results.

The calculations lead to the following recursion equation:

$$I_T(t)^k \approx \sqrt{c_k(\ln(t) - \ln(I_T(t)^{k-1}))} \tag{3.14}$$

with $I_T(t)^0 = 1$ and $k$ the recursion depth. The calculation of the constants $c_k$ is difficult. Taking a rough approximation with $k = 2$ the following equation is obtained that approximates (3.13) with an relative error of less than 2.4% for $t \in [2, 5]$, for tournament sizes $t > 5$ the relative error is less than 1%:

$$I_T(t) \approx \sqrt{2(\ln(t) - \ln(\sqrt{4.14 \ \ln(t)}))} \tag{3.15}$$
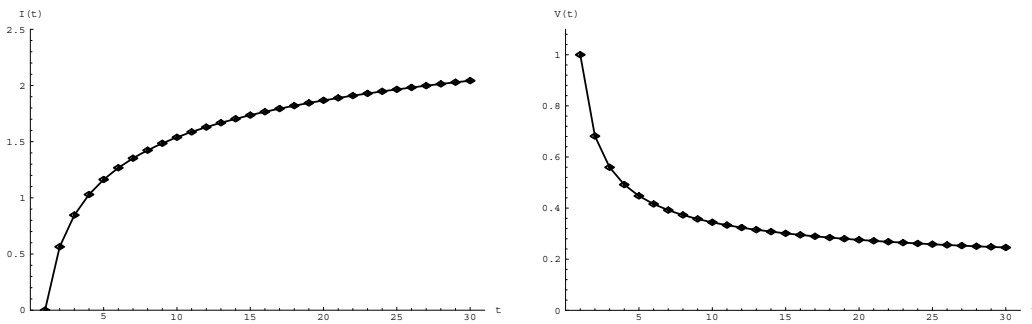


Figure 3.5: Dependence of the selection intensity (left) and selection variance (right) on the tournament size $t$.

## 3.5   Selection Variance

To determine the selection variance we need to solve the equation

$$V_T(t) = \int_{-\infty}^{\infty} t \ (x - I_T(t))^2 \ \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left( \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right)^{t-1} dx \tag{3.16}$$

For a binary tournament we have

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

21

Here again (3.16) can be solved by numerical integration. The dependence of the selection variance on the tournament size is shown on the right of Figure 3.5.

To obtain a useful analytic approximation for the selection variance, we perform a symbolic regression using the genetic programming optimization method. Details about the way the data was computed can be found in appendix A. The following formula approximates the selection variance with an relative error of less than 1.6% for $t \in \{1, \ldots, 30\}$:

$$V_T(t) \approx \sqrt{\frac{2.05 + t}{3.14t^{\frac{3}{2}}}} \qquad ; t \in \{1, \ldots, 30\} \qquad (3.17)$$

# Chapter 4

# Truncation Selection

In Truncation selection with threshold $T$ only the fraction $T$ best individuals can be selected and they all have the same selection probability. This selection method is often used by breeders and in population genetic [Bulmer, 1980; Crow and Kimura, 1970]. Mühlenbein has introduced this selection scheme to the domain of genetic algorithms [Mühlenbein and Schlierkamp-Voosen, 1993]. This method is equivalent to $(\mu, \lambda)$-selection used in evolution strategies with $T = \frac{\mu}{\lambda}$ [Bäck, 1995].

The outline of the algorithm is given by algorithm 2.

---

Algorithm 2: (Truncation Selection)

**Input:** The population $P(\tau)$, the truncation threshold $T \in [0, 1]$
**Output:** The population after selection $P(\tau)'$

truncation($T, J_1, ..., J_N$):
    $\bar{J} \leftarrow$ sorted population $J$ according fitness
            with worst individual at the first position
    **for** $i \leftarrow 1$ **to** $N$ **do**
        $r \leftarrow$ random$\{ [(1 - T)N], \ldots, N \}$
        $J'_i \leftarrow \bar{J}_r$
    **od**
    **return** $\{J'_1, ..., J'_N\}$

---

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

Although this method has been investigated several times we will describe this selection method using the methods derived here, as additional properties can be observed.

**Theorem 4.0.1** *The expected fitness distribution after performing truncation se-*

*lection with threshold $T$ on the distribution $s$ is*

$$\Omega_\Gamma^*(s, T)(f_i) = s^*(f_i) = \begin{cases} 0 & : \quad S(f_i) \leq (1 - T)N \\ \frac{S(f_i) - (1-T)N}{T} & : \quad S(f_{i-1}) \leq (1 - T)N < S(f_i) \\ \frac{s(f_i)}{T} & : \quad else \end{cases} \quad (4.1)$$

**Proof:** The first case in (4.1) gives zero offspring to individuals with a fitness value below the truncation threshold. The second case reflects the fact that threshold may lie within $s_i$. Then only the fraction above the threshold ($S_i - (1-T)N$) may be selected. These fraction is in average copied $\frac{1}{T}$ times. The last case in (4.1) gives all individuals above the threshold the multiplication factor $\frac{1}{T}$ that is necessary to keep the population size constant. $\qquad\square$

**Theorem 4.0.2** *Let $\bar{s}(f)$ be the continuous distribution of the population. Then the expected fitness distribution after performing truncation selection with threshold $T$ is*

$$\overline{\Omega}_\Gamma^*(\bar{s}, T)(f) = \begin{cases} \frac{\bar{s}(f)}{T} & : \quad \bar{S}(f) > (1 - T)N \\ 0 & : \quad else \end{cases} \quad (4.2)$$

**Proof:** As $\bar{S}(f)$ gives the cumulative fitness distribution, it follows from the construction of truncation selection that all individuals with $\bar{S}(f) < (1 - T)N$ are truncated. As the population size is kept constant during selection, all other individuals must be copied in average $\frac{1}{T}$ times. $\qquad\square$

## 4.1 Reproduction Rate

**Corollary 4.1.1** *The reproduction rate of truncation selection is*

$$\bar{R}_\Gamma(f) = \begin{cases} \frac{1}{T} & : \quad \bar{S}(f) > (1 - T)N \\ 0 & : \quad else \end{cases} \quad (4.3)$$

## 4.2 Loss of Diversity

By construction of the selection method only the fraction $T$ of the population will be selected, i.e. the loss of diversity is

$$p_{d,\Gamma}(T) = 1 - T \quad (4.4)$$

## 4.3 Selection Intensity

The results presented in this subsection have been already derived in a different way in [Crow and Kimura, 1970].

**Theorem 4.3.1** *The selection intensity of truncation selection is*

$$I_\Gamma(T) = \frac{1}{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{f_c^2}{2}} \tag{4.5}$$

*where $f_c$ is determined by $T = \int_{f_c}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{f^2}{2}} df$.*

**Proof:** The selection intensity is defined as the average fitness of the population after selection assuming an initial normalized Gaussian distribution $G(0,1)$, hence $I = \int_{-\infty}^{\infty} \bar{\Omega}(G(0,1))(f) \, f \, df$. As 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

$$\bar{S}(f_c) = (1 - T)N = 1 - T \tag{4.6}$$

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

So we can compute

$$\begin{aligned} I_\Gamma(T) &= \int_{f_c}^{\infty} \frac{1}{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{f^2}{2}} f \, df \\ &= \frac{1}{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{f_c^2}{2}} \end{aligned}$$

Here $f_c$ is determined by (4.6). Solving (4.6) for $T$ yields

$$\begin{aligned} T &= 1 - \int_{-\infty}^{f_c} \frac{1}{\sqrt{2\pi}} e^{-\frac{f^2}{2}} df \\ &= \int_{f_c}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{f^2}{2}} df \end{aligned}$$

$\square$

A lower bound for the selection intensity reported by [Mühlenbein and Voigt, 1995] is $I_\Gamma(T) \leq \sqrt{\frac{1-T}{T}}$.

Figure 4.1 shows on the left the selection intensity in dependence of parameter $T$.

## 4.4 Selection Variance

**Theorem 4.4.1** *The selection variance of truncation selection is*

$$V_\Gamma(T) = 1 - I_\Gamma(T)(I_\Gamma(T) - f_c) \tag{4.7}$$

25
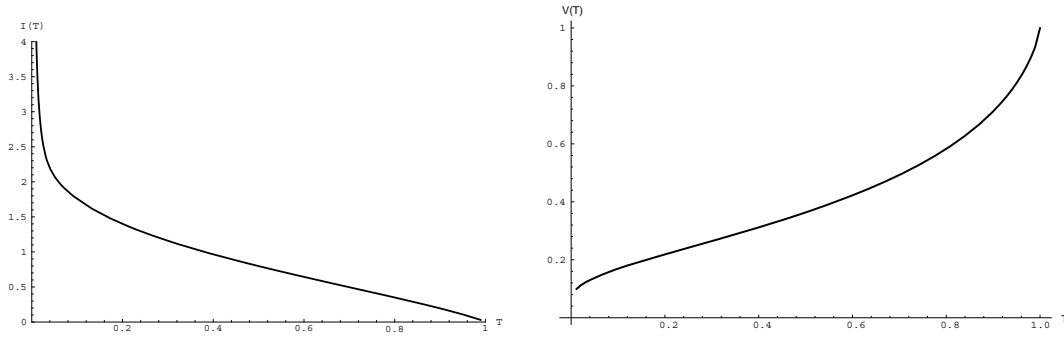
Figure 4.1: Selection intensity (left) and selection variance (right) of truncation selection.

**Sketch of proof:** The substitution of (4.2) in the definition equation (2.17) gives

$$V_\Gamma(T) = \int_{f_c}^{\infty} f^2 \frac{1}{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{f^2}{2}} \, df - I_\Gamma(T))^2$$

After some calculations this equation can be simplified to (4.7). □

The selection variance is plotted on the right of Figure 4.1. (4.7) has also been derived in [Bulmer, 1980].

# Chapter 5

# Linear Ranking Selection

Ranking selection was first suggested by Baker to eliminate the serious disadvantages of proportionate selection [Grefenstette and Baker, 1989; Whitley, 1989]. For ranking selection the individuals are sorted according their fitness values and the rank $N$ is assigned to the best individual and the rank 1 to the worst individual. The selection probability is linearly assigned to the individuals according to their rank:

$$p_i = \frac{1}{N} \left( \eta^- + (\eta^+ - \eta^-) \frac{i-1}{N-1} \right) ; \qquad i \in \{1, \ldots, N\} \tag{5.1}$$

Here $\frac{\eta^-}{N}$ is the probability of the worst individual to be selected and $\frac{\eta^+}{N}$ the probability of the best individual to be selected. As 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, i.e. a different selection probability, even if they have the same fitness value.

Koza [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 (5.1) is possible by $\eta^- = \frac{2}{r_m+1}$ and $\eta^+ = \frac{2r_m}{r_m+1}$.

Whitley [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

$$j = \lfloor \frac{N}{2(c-1)} \left( c - \sqrt{c^2 - 4(c-1)\chi} \right) \rfloor \tag{5.2}$$

where $c$ is a parameter called "selection bias". Bäck has shown that for $1 < c \leq 2$ this method is almost identical to the probabilities in (5.1) with $\eta^+ = c$ [Bäck, 1994].

```
Algorithm 3: (Linear Ranking Selection)

Input: The population $P(\tau)$ and the reproduction rate of the worst
        individual $\eta^- \in [0,1]$
Output: The population after selection $P(\tau)'$

linear_ranking($\eta^-$,$J_1$, ..., $J_N$):
        $\bar{J} \leftarrow$ sorted population $J$ according fitness
                     with worst individual at the first position
        $s_0 \leftarrow 0$
        for $i \leftarrow 1$ to $N$ do
                     $s_i \leftarrow s_{i-1} + p_i$ (Equation 5.1)
        od
        for $i \leftarrow 1$ to $N$ do
                     $r \leftarrow$ random[0,$s_N$[
                     $J'_i \leftarrow \bar{J}_l$ such that $s_{l-1} \le r < s_l$
        od
        return $\{J'_1, ..., J'_N\}$
```

The pseudo-code 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, i.e. $\mathcal{O}(N \log N)$.

**Theorem 5.0.2** *The expected fitness distribution after performing ranking selection with $\eta^-$ on the distribution $s$ is*

$$\Omega_R^*(s, \eta^-)(f_i) = s^*(f_i) = s(f_i)\frac{N\eta^- - 1}{N - 1} + \frac{1 - \eta^-}{N - 1}\left(S(f_i)^2 - S(f_{i-1})^2\right) \quad (5.3)$$

**Proof:** We first calculate the expected number of individuals with fitness $f_i$ or worse, i.e. $S^*(f_i)$. As the individuals are sorted according to their fitness value this number is given by the sum of the probabilities of the $S^*(f_i)$ less fit individuals:

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

$$= \eta^- S(f_i) + \frac{\eta^+ - \eta^-}{N - 1}\sum_{j=1}^{S(f_i)} j - 1$$

$$= \eta^- S(f_i) + \frac{\eta^+ - \eta^-}{N - 1}\frac{1}{2}S(f_i)\left(S(f_i) - 1\right)$$

As $\eta^+ = 2 - \eta^-$ and $s^*(f_i) = S^*(f_i) - S^*(f_{i-1})$ we obtain

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

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

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

$\square$

**Example 5.0.1** *As an example we use again the fitness distribution of the "wall-following-robot" from Example 2.0.1. The resulting distribution after ranking selection with $\eta^- = 0.1$ is shown in Figure 5.1. Here again the confidence interval is shown. A comparison between theoretical analysis and the average of 20 simulations is shown in Figure 5.2. Again a very high agreement with the theoretical results is observed.*
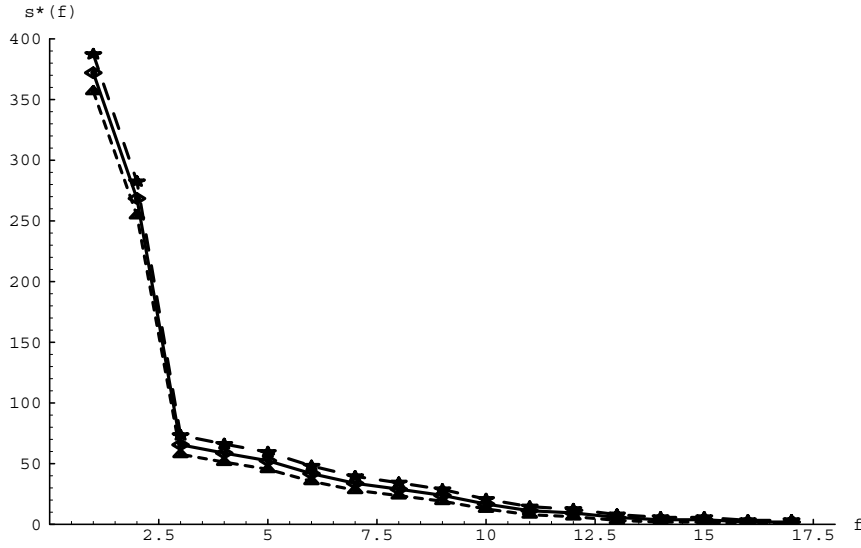


Figure 5.1: The resulting expected fitness distribution and the confidence interval of 68% after applying ranking selection with $\eta^- = 0.1$.

**Theorem 5.0.3** *Let $\bar{s}(f)$ be the continuous fitness distribution of the population. Then the expected fitness distribution after performing ranking selection $\overline{\Omega}_R$ with $\eta^-$ on the distribution $\bar{s}$ is*

$$\overline{\Omega}_R^*(\bar{s}, \eta^-)(f) = \bar{s}^*(f) = \eta^- \bar{s}(f) + 2\frac{1 - \eta^-}{N} \bar{S}(f)\bar{s}(f) \tag{5.4}$$

**Proof:**  As the continuous form of (5.1) is given by $\bar{p}(x) = \frac{1}{N}(\eta^- + \frac{\eta^+ - \eta^-}{N}x)$ we calculate $\bar{S}(f)$ using $\eta^+ = 2 - \eta^-$:

$$\bar{S}^*(f) = N \int_0^{\bar{S}(f)} \bar{p}(x)\, dx$$

29
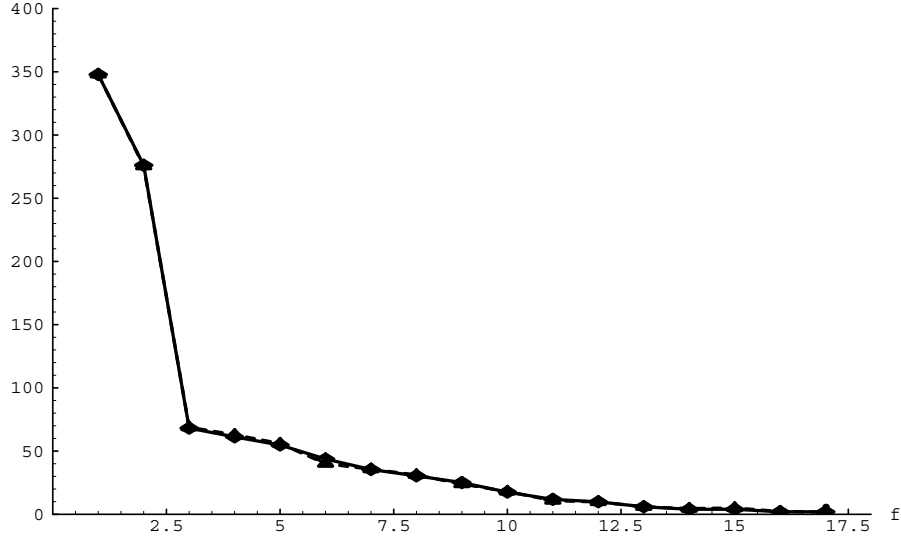
Figure 5.2: Comparison between theoretical derived distribution (—) and the average of 20 simulations (- - -) for ranking selection with $\eta^- = \frac{1}{N}$.

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

As $\bar{s}^*(f) = \frac{d\bar{S}^*(f)}{df}$, (5.4) follows.  $\square$

**Example 5.0.2** *Figure 5.3 shows the the initial continuous fitness distribution $\bar{s}_G$ and the resulting distributions after performing ranking selection.*

## 5.1   Reproduction Rate

**Corollary 5.1.1** *The reproduction rate of ranking selection is*

$$
\bar{R}_R(f) = \eta^- + 2\frac{1-\eta^-}{N} \bar{S}(f) \tag{5.5}
$$

This equation shows that the worst fit individuals have the lowest reproduction rate $\bar{R}(f_0) = \eta^-$ and the best fit individuals have the highest reproduction rate $\bar{R}(f_n) = 2 - \eta^- = \eta^+$. This can be derived from the construction of the method as $\frac{\eta^-}{N}$ is the selection probability of the worst fit individual and $\frac{\eta^+}{N}$ the one of the best fit individual.
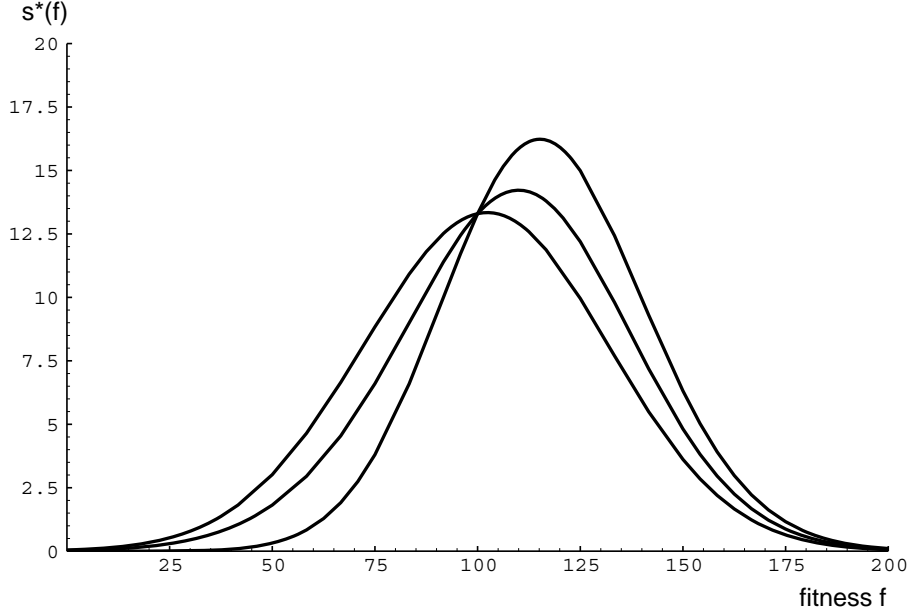
Figure 5.3: Gaussian fitness distribution $\bar{s}_G(f)$ and the resulting distributions after performing ranking selection with $\eta^- = 0.5$ and $\eta^- = 0$ (from left to right).

## 5.2  Loss of Diversity

**Theorem 5.2.1** *The loss of diversity $p_{d,R}(\eta^-)$ of ranking selection is*

$$p_{d,R}(\eta^-) = (1 - \eta^-)\frac{1}{4} \tag{5.6}$$

**Proof:**  Using Theorem 2.4.1 and realizing that $S(f_z) = \frac{N}{2}$ we calculate:

$$
\begin{aligned}
p_{d,R}(\eta^-) &= \frac{1}{N}\left(\bar{S}(f_z) - \bar{S}^*(f_z)\right) \\
&= \frac{1}{N}\left(\bar{S}(f_z) - \eta^-\bar{S}(f_z) - \frac{1-\eta^-}{N}\bar{S}(f_z)^2\right) \\
&= \frac{1}{N}\left(\frac{N}{2} - \eta^-\frac{N}{2} - \frac{1-\eta^-}{N}\frac{N^2}{4}\right) \\
&= \frac{1}{4}(1 - \eta^-)
\end{aligned}
$$

$\square$

Baker has derived this result using his term of "reproduction rate" [Baker, 1989].

Note that the loss of diversity is again independent of the initial distribution.

31

## 5.3  Selection Intensity

**Theorem 5.3.1** *The selection intensity of ranking selection is*

$$I_R(\eta^-) = (1 - \eta^-)\frac{1}{\sqrt{\pi}} \tag{5.7}$$

**Proof:**  Using the definition of the selection intensity (Definition 2.5.2) and using the Gaussian function for the initial fitness distribution we obtain

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

As the first summand is 0 and $\int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \, dx = \sqrt{\pi}$ we obtain (5.7). $\square$

The selection intensity of ranking selection is shown in Figure 5.4 (left) in dependence of the parameter $\eta^-$.
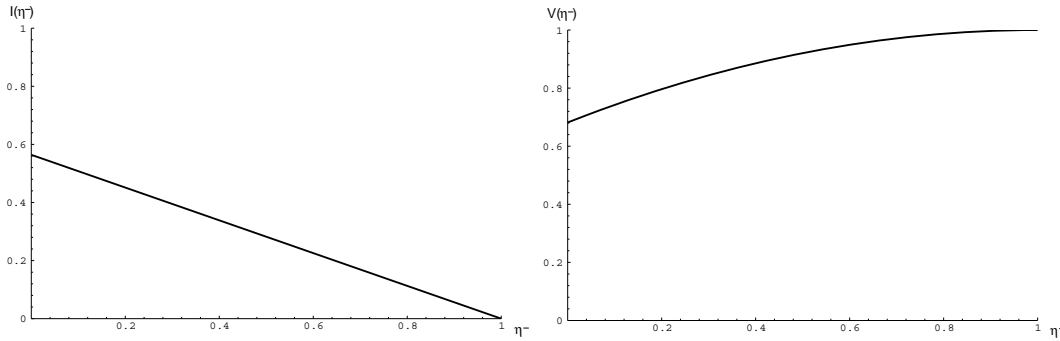


Figure 5.4:  Selection intensity (left) and selection variance (right) of ranking selection.

## 5.4  Selection Variance

**Theorem 5.4.1** *The selection variance of ranking is*

$$V_R(\eta^-) = 1 - \frac{(1 - \eta^-)^2}{\pi} = 1 - I_R(\eta^-)^2 \tag{5.8}$$

**Proof:**  Substituting (5.4) into the definition equation (2.17) leads to

$$V_R(\eta^-) = \int_{-\infty}^{\infty} f^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{f^2}{2}} \left( \eta^- + 2(1 - \eta-) \int_{-\infty}^{f} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right) df - I_R(\eta^-)^2$$

$$V_R(\eta^-) = \frac{\eta^-}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f^2 e^{-\frac{f^2}{2}} \, df$$

$$+ \frac{1 - \eta^-}{\pi} \int_{-\infty}^{\infty} f^2 e^{-\frac{f^2}{2}} \int_{-\infty}^{f} e^{-\frac{y^2}{2}} \, dy \, df$$

$$- I_R(\eta^-)^2$$

Using the relations B.7 and B.8 we obtain

$$V_R(\eta^-) = \eta^- + (1 - \eta^-) - I_R(\eta^-)^2$$
$$= 1 - I_R(\eta^-)^2$$

$\square$

The selection variance of ranking selection is plotted on the right of Figure 5.4.

# Chapter 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. We will discuss the meaning and the influence of this parameter in detail in the following. 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

$$p_i = \frac{c^{N-i}}{\sum_{j=1}^{N} c^{N-j}} \; ; \qquad i \in \{1, ..., N\} \tag{6.1}$$

The sum $\sum_{j=1}^{N} c^{N-j}$ normalizes the probabilities to ensure that $\sum_{i=1}^{N} p_i = 1$. As $\sum_{j=1}^{N} c^{N-j} = \frac{c^N - 1}{c - 1}$ we can rewrite the above equation:

$$p_i = \frac{c - 1}{c^N - 1} c^{N-i} \; ; \qquad i \in \{1, ..., N\} \tag{6.2}$$

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

**Theorem 6.0.2** *The expected fitness distribution after performing exponential ranking selection with c on the distribution s is*

$$\Omega_E^*(s, c, N)(f_i) = s^*(f_i) = N \frac{c^N}{c^N - 1} c^{-S(f_i)} \left( c^{s(f_i)} - 1 \right) \tag{6.3}$$

Algorithm 4: (Exponential Ranking Selection)

**Input:** The population $P(\tau)$ and the ranking base $c \in ]0,1]$
**Output:** The population after selection $P(\tau)'$

exponential_ranking($c, J_1, ..., J_N$):
      $\bar{J} \leftarrow$ sorted population $J$ according to fitness
             with worst individual at the first position
      $s_0 \leftarrow 0$
      **for** $i \leftarrow 1$ **to** $N$ **do**
             $s_i \leftarrow s_{i-1} + p_i$ (Equation 6.2)
      **od**
      **for** $i \leftarrow 1$ **to** $N$ **do**
             $r \leftarrow$ random$[0, s_N[$
             $J_i' \leftarrow \bar{J}_l$ such that $s_{l-1} \leq r < s_l$
      **od**
      **return** $\{J_1', ..., J_N'\}$

**Proof:** We first calculate the expected number of individuals with fitness $f_i$ or worse, i.e. $S^*(f_i)$. As the individuals are sorted according to their fitness value this number is given by the sum of the probabilities of the $S^*(f_i)$ less 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 $k = N - j$

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

$$=\; N \frac{c-1}{c^N - 1} \left( \sum_{k=0}^{N-1} c^k - \sum_{k=0}^{N-S(f_i)-1} c^k \right)$$

$$=\; N \frac{c-1}{c^N - 1} \left( \frac{c^N - 1}{c - 1} - \frac{c^{N-S(f_i)}}{c - 1} \right)$$

$$=\; N \left( 1 - \frac{c^N}{c^N - 1} c^{-S(f_i)} \right)$$

As $s^*(f_i) = S^*(f_i) - S^*(f_{i-1})$ we obtain

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

$$= N\frac{c-1}{c^N-1}c^{-S(f_i)}\left(c^{s(f_i)}-1\right)$$

$\square$

**Example 6.0.1** *As an example we use again the fitness distribution of the "wall-following-robot" from Example 2.0.1. The resulting distribution after exponential ranking selection with $c = 0.99$ and $N = 1000$ is shown in Figure 6.1 as a comparison to the average of 20 simulations. Again a very high agreement with the theoretical results is observed.*
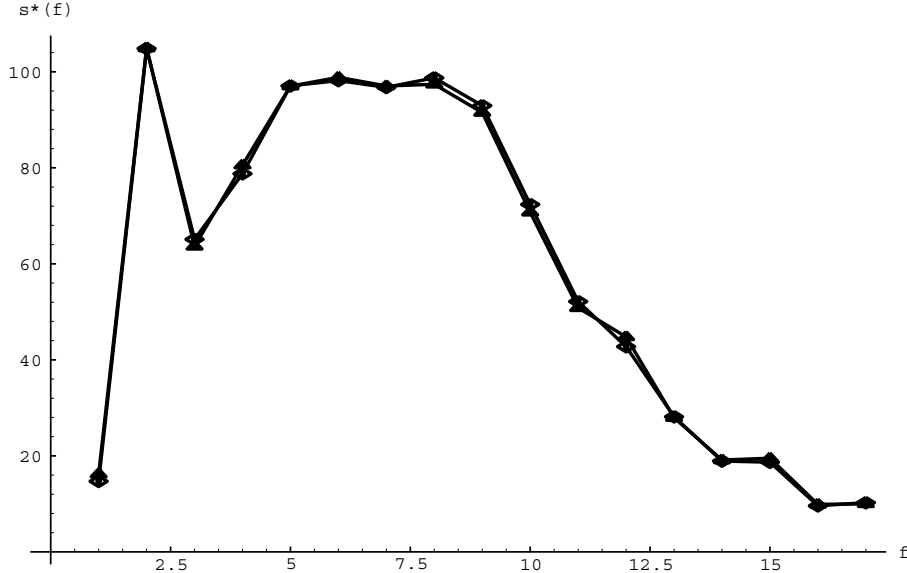


Figure 6.1: Comparison between theoretical derived distribution (—) and the average of 20 simulations (- - -) for ranking selection with $c = 0.99$.

**Theorem 6.0.3** *Let $\bar{s}(f)$ be the continuous fitness distribution of the population. Then the expected fitness distribution after performing exponential ranking selection $\overline{\Omega}_E$ with $c$ on the distribution $\bar{s}$ is*

$$\overline{\Omega}_E^*(\bar{s},c)(f) = \bar{s}^*(f) = N\frac{c^N}{c^N-1}\ln c\,\bar{s}(f)\,c^{-\bar{S}(f)} \tag{6.4}$$

**Proof:** As the continuous form of (6.2) is given by $\bar{p}(x) = \frac{c^{N-x}}{\int_0^N c^{N-x}}$ and $\int c^x = \frac{1}{\ln c}c^x$ we calculate:

$$\bar{S}^*(f) = N\frac{c^N\ln c}{c^N-1}\int_0^{\bar{S}(f)} c^{-x}\,dx$$

36

$$= -N \frac{c^N}{c^N - 1} [c^{-x}]_0^{\bar{S}(f)}$$

$$= N \frac{c^N}{c^N - 1} \left( 1 - c^{-S(f)} \right)$$

As $\bar{s}^*(f) = \frac{d\bar{S}^*(f)}{df}$, (6.4) follows. $\qquad\qquad \square$

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

$$\overline{\Omega}_E^*(s, \alpha)(f) = \bar{s}^*(f) = \frac{\alpha \ln \alpha}{\alpha - 1} \bar{s}(f) \alpha^{-\frac{\bar{S}(f)}{N}} \tag{6.5}$$

The meaning of $\alpha$ will become apparent in the next section.

## 6.1  Reproduction Rate

**Corollary 6.1.1** *The reproduction rate of exponential ranking selection is*

$$\bar{R}_E(f) = \frac{\alpha \ln \alpha}{\alpha - 1} \alpha^{-\frac{\bar{S}(f)}{N}} \tag{6.6}$$

This equation shows that the worst fit individuals have the lowest reproduction rate $\bar{R}(f_0) = \frac{\alpha \ln \alpha}{\alpha - 1}$ and the best fit individuals have the highest reproduction rate $\bar{R}(f_n) = \frac{\ln \alpha}{\alpha - 1}$. Hence we obtain a natural explanation of the variable $\alpha$, as $\frac{\bar{R}(f_0)}{\bar{R}(f_n)} = \alpha$: 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 $N$, i.e. the interesting region of values for $\alpha$ is in the range from $10^{-20}, \ldots, 1$.

## 6.2  Loss of Diversity

**Theorem 6.2.1** *The loss of diversity $p_{d,E}(\alpha)$ of exponential ranking selection is*

$$p_{d,E}(\alpha) = \frac{1 - \ln \frac{\alpha - 1}{\alpha \ln \alpha}}{\ln \alpha} - \frac{\alpha}{\alpha - 1} \tag{6.7}$$

**Proof:** First we calculate from the demand $R(f_z) = 1$ :

$$\frac{\bar{S}(f_z)}{N} = -\frac{\ln \frac{\alpha - 1}{\alpha \ln \alpha}}{\ln \alpha} \tag{6.8}$$

Using Theorem 2.4.1 we obtain:

$$p_{d,E}(\alpha) = \frac{1}{N} \left( \bar{S}(f_z) - \bar{S}^*(f_z) \right)$$

37

$$
= \quad -\frac{\ln\frac{\alpha-1}{\alpha\ln\alpha}}{\ln\alpha} - \frac{\alpha}{\alpha-1}\left(1 - \alpha^{\frac{\ln\frac{\alpha-1}{\alpha\ln\alpha}}{\ln\alpha}}\right)
$$

$$
= \quad -\frac{\ln\frac{\alpha-1}{\alpha\ln\alpha}}{\ln\alpha} - \frac{\alpha}{\alpha-1}\left(1 - \frac{\alpha-1}{\alpha\ln\alpha}\right)
$$

$$
= \quad \frac{1 - \ln\frac{\alpha-1}{\alpha\ln\alpha}}{\ln\alpha} - \frac{\alpha}{\alpha-1}
$$

$\square$

The loss of diversity is shown in figure 6.2.
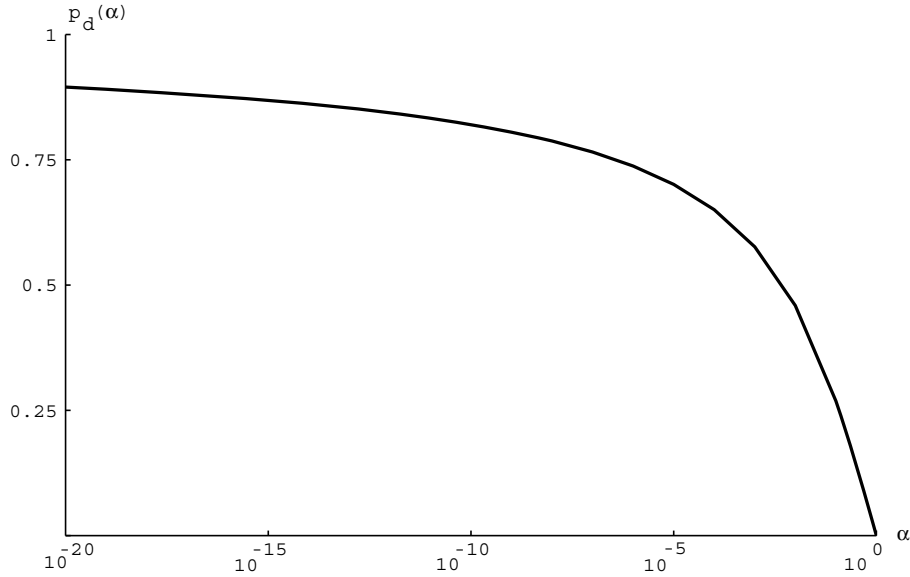


Figure 6.2: The loss of diversity $p_{d,E}(\alpha)$ for exponential ranking selection. Note the logarithmic scale of the $\alpha$-axis.

## 6.3   Selection Intensity and Selection Variance

The selection intensity and the selection variance are very difficult to calculate for exponential ranking. If we recall the definition of the selection intensity (definition 2.5.2) we see that the integral of the Gaussian function occurs as exponent in an indefinite integral. Hence we restrict ourselves here to numerical calculation of the selection intensity as well as of the selection variance. The selection intensity and the selection variance of exponential ranking selection is shown in Figure 6.3 in dependence of the parameter $\alpha$.

An approximation formula can be derived using the genetic programming optimization method for symbolic regression (see Appendix A). The selection
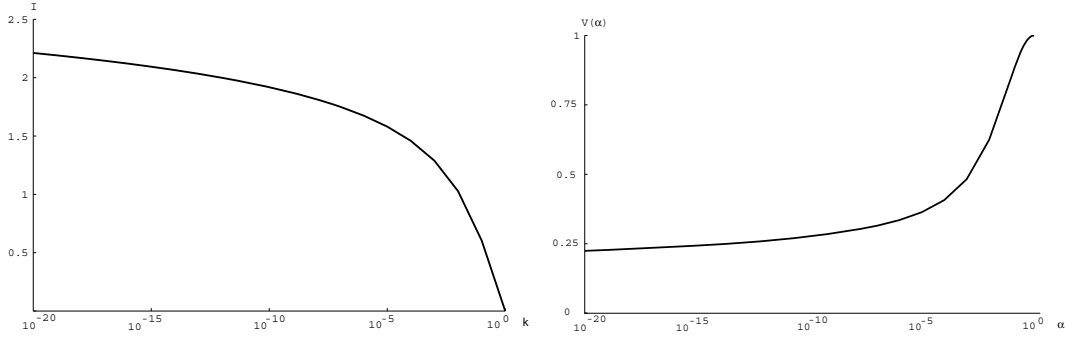
Figure 6.3: Selection intensity (left) and selection variance (right) of exponential ranking selection. Note the logarithmic scale of the $\alpha$-axis.

intensity of exponential ranking selection can be approximated with a relative error of less than 6% for $\alpha \in [10^{-20}, 0.8]$ by

$$I_E(\alpha) \approx 0.588 \frac{\ln \ln \frac{\pi}{\alpha}}{3.69^{\alpha}} \tag{6.9}$$

Similar, 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 5% for $\alpha \in [10^{-20}, 0.8]$:

$$V_E(\alpha) \approx \ln \left( 1.2 + \frac{2.8414}{2.225\alpha - \ln \alpha} \right) \tag{6.10}$$

# Chapter 7

# Proportional Selection

Proportional selection is the original selection method proposed for genetic algorithms by Holland [Holland, 1975]. We include the analysis of the selection method mostly because of its fame.

---

Algorithm 5: (Proportional Selection)

**Input:** The population $P(\tau)$
**Output:** The population after selection $P(\tau)'$

proportional$(J_1, ..., J_N)$:
    $s_0 \leftarrow 0$
    **for** $i \leftarrow 1$ **to** $N$ **do**
        $s_i \leftarrow s_{i-1} + \frac{f_i}{M}$
    **od**
    **for** $i \leftarrow 1$ **to** $N$ **do**
        $r \leftarrow$ random$[0, s_N[$
        $J_i' \leftarrow J_l$ such that $s_{l-1} \leq r < s_l$
    **od**
    **return** $\{J_1', ..., J_N'\}$

---

The probability of an individual to be selected is simply proportionate to its fitness value, i.e.

$$p_i = \frac{f_i}{NM} \tag{7.1}$$

Algorithm 5 displays the method using a pseudo code formulation. The time complexity of the algorithm is $\mathcal{O}(N)$.

Obviously, this mechanism will only work if all fitness values are greater than zero. Furthermore the selection probabilities strongly depend on the scaling of the fitness function. As an example, assume a population of 10 individuals with the best individual having a fitness value of 11 and the worst a fitness value of

1. The selection probability for the best individual is hence $p_b \approx 16.6\%$ and for the worst $p_w \approx 1.5\%$. If we now translate the fitness function by 100, i.e. we just add a the constant value 100 to every fitness value, we calculate $p_b' \approx 10.4\%$ and $p_w' \approx 9.5\%$. The selection probabilities of the best and the worst individual are now almost identical. This undesirable property arises from the fact that proportional selection is not translation invariant (see e.g. [de la Maza and Tidor, 1993]). Because of this several *scaling methods* have been proposed to keep proportional selection working, e.g. linear static scaling, linear dynamic scaling, exponential scaling, logarithmic scaling [Grefenstette and Baker, 1989]; sigma truncation [Brill *et al.*, 1992]. Another method to improve proportional selection is the "over selection" of a certain percentage of the best individuals, i.e. to force that 80 % of all individuals are taken from the best 20 % of the population. This method was used in [Koza, 1992]. In [Mühlenbein and Schlierkamp-Voosen, 1993] it is already stated that "these modifications are necessary, not tricks to speed up the algorithm". The following analysis will confirm this statement.

**Theorem 7.0.1** *The expected fitness distribution after performing proportional selection on the distribution s is*

$$\Omega_P^*(s)(f_i) = s^*(f) = s(f)\frac{f}{M} \tag{7.2}$$

# 7.1 Reproduction Rate

**Corollary 7.1.1** *The reproduction rate of proportional selection is*

$$\bar{R}_P(f) = \frac{f}{\bar{M}} \tag{7.3}$$

The reproduction rate is proportionate to the fitness value of an individual. If all fitness values are close together (as it was in the example at the beginning of this chapter) all individuals have almost the same reproduction rate $R \approx 1$. Hence no selection takes place anymore.

# 7.2 Selection Intensity

As proportional selection is not translation invariant our original definition of standardized selection intensity cannot be applied. We will cite here the results obtained by Mühlenbein and Schlierkamp-Voosen [Mühlenbein and Schlierkamp-Voosen, 1993].

**Theorem 7.2.1** *[Mühlenbein and Schlierkamp-Voosen, 1993] The standardized selection intensity of proportional selection is*

$$I_P = \frac{\bar{\sigma}}{\bar{M}} \tag{7.4}$$

*where $\bar{\sigma}$ is the mean variance of the fitness values of the population before selection.*

**Proof:** See [Mühlenbein and Schlierkamp-Voosen, 1993].  □

The other properties we are interested in like the selection variance an the loss of diversity are difficult to investigate for proportional selection. The crucial point is the explicit occurrence of the fitness value in the expected fitness distribution after selection (7.2). Hence an analysis is only possible if we make some further assumptions on the initial fitness distribution. This is why other work on proportional selection assume some special functions to be optimized (e.g. [Goldberg and Deb, 1991]).

Another weak point is that the selection intensity even in the early stage of the optimization (when the variance is high) is too low. Measurements on a broad range of problems showed sometimes a negative selection intensity. This means that in some cases (due to sampling) there is a *decrease* in average population fitness. Seldom a very high selection intensity occurred ($I \approx 1.8$) if a *super-individual* was created. But the measured average selection intensity was in range of 0.1 to 0.3.

All the undesired properties together led us to the conclusion that proportional selection is a very unsuited selection scheme. Informally one can say that the only advantage of proportional selection is that it is so difficult to prove the disadvantages.

# Chapter 8

# Comparison of Selection Schemes

In the subsequent sections the selection methods are compared according to their properties derived in the preceding chapters. First we will compare the reproduction rates of selection methods and derive an unified view of selection schemes. Section 8.2 is devoted to the comparison of the selection intensity and gives a convergence prediction for 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.

We will take into account proportional selection only in the first two subsections when the reproduction rate and the selection intensity are analyzed. In other comparisons it is neglected as it withdraws itself an analysis of the properties we are interested in.

## 8.1    Reproduction Rate and Universal Selection

The reproduction rate simply gives the number of expected offspring of an individual with a certain fitness value after selection. But in the preceding chapters only the reproduction rate for the continuous case have been considered. Table 8.1 gives the equations for the discrete (exact) case. They have been derived using the exact offspring equations (3.1), (4.1), (5.3), (6.3) and (7.2) and doing some simple algebraic manipulations.

The examples in the preceding chapter showed a large mean variation of the fitness distributions after selection. In the following, we will see that this mean variation can be almost completely eliminated by using the reproduction rate and the so called "stochastic universal sampling". As can be seen from table 8.1 we can calculate the expected distribution in advance without carrying out a "real" selection method. This calculation also enables us to use stochastic universal sampling (SUS) [Baker, 1987] for all selection schemes discussed herein.

The SUS algorithm can be stated to be an optimal sampling algorithm. It has zero bias, i.e. no deviation between the expected reproduction rate and the

| Selection Method | Reproduction Rate |
|---|---|
| Tournament | $R_T(f_i) = \frac{N}{s(f_i)}\left(\left(\frac{S(f_i)}{N}\right)^t - \left(\frac{S(f_{i-1})}{N}\right)^t\right)$ |
| Truncation | $R_\Gamma(f_i) = \begin{cases} 0 & : & S(f_i) \le (1-T)N \\ \frac{S(f_i)-(1-T)N}{s(f_i)T} & : & S(f_{i-1}) \le (1-T)N < S(f_i) \\ \frac{1}{T} & : & else \end{cases}$ |
| Linear Ranking | $R_R(f_i) = \frac{N\eta^- - 1}{N-1} + \frac{1-\eta^-}{N-1}\left(2S(f_i) - s(f_i)\right)$ |
| Exponential Ranking | $R_E(f_i) = \frac{N}{s(f_i)}\frac{\alpha \ln \alpha}{\alpha-1}\alpha^{-\frac{S(f_i)}{N}}\left(\alpha^{-\frac{s(f_i)}{N}} - 1\right)$ |
| Proportional | $R_P(f_i) = \frac{f_i}{M}$ |

Table 8.1: Comparison of the reproduction rate of the selection methods for discrete distributions.

algorithmic sampling frequency. Furthermore, SUS has a minimal spread, i.e. the range of the possible values for $s'(f_i)$ is

$$s'(f_i) \in \{\lfloor s^*(f_i)\rfloor, \lceil s^*(f_i)\rceil\} \tag{8.1}$$

The outline of the SUS algorithm is given by algorithm 6. The standard sampling mechanism uses one spin of a roulette wheel (divided into segments for each individual with an 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. As these trials are independent of each other a relatively high variance in the outcome is observed (see also chapter 2 and theorem 2.0.1). This is also the case for tournament selection although there is no explicitly used roulette wheel sampling. In contrary for SUS only a single spin of the wheel is necessary as the roulette has $N$ markers for the "winning individuals" and hence all individuals are chosen at once.

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, i.e. the mean variation is minimal. Even though it is not clear whether there 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 has applied this sampling method only to linear ranking selection as here the expected number of offspring is known by construction (see chapter 5). As we have derived this offspring values for the selection methods discussed in the previous chapters it is possible to use stochastic universal sampling for all these selections schemes. Hence, we may obtain a unified view of selection schemes, if we neglect the way the reproduction rates were derived and construct an "universal selection method" in the following way: First we compute

the fitness distribution of the population. Next the expected reproduction rates are calculated using the equations derived in the proceeding chapters and summarized in table 8.1. In the last step SUS is used to obtain the new population after selection. This algorithm is given in algorithm 7 and the SUS algorithm is outlined by algorithm 6.

---

Algorithm 6: (Stochastic Universal Sampling)

**Input:** The population $P(\tau)$ and the reproduction rate for each
fitness value $R_i \in [0, N]$
**Output:** The population after selection $P(\tau)'$

SUS($R_1, \ldots, R_n, J_1, \ldots, J_N$):
    $sum \leftarrow 0$
    $j \leftarrow 1$
    $ptr \leftarrow \text{random}[0,1)$
    **for** $i \leftarrow 1$ **to** $N$ **do**
        $sum \leftarrow sum + R_i$ where $R_i$ is the reproduction rate
            of individual $J_i$
        **while** ($sum > ptr$) **do**
            $J'_j \leftarrow J_i$
            $j \leftarrow j + 1$
            $ptr \leftarrow ptr + 1$
        **od**
    **od**
    **return** $\{J'_1, ..., J'_N\}$

---

Algorithm 7: (Universal Selection Method)

**Input:** The population $P(\tau)$
**Output:** The population after selection $P(\tau)'$

universal_selection($J_1, \ldots, J_N$):
    $s \leftarrow \text{fitness\_distribution}(J_1, \ldots, J_N)$
    $r \leftarrow \text{reproduction\_rate}(s)$
    $J' \leftarrow \text{SUS}(r, J)$
    **return** $J'$

---

The time complexity of the universal selection method is $\mathcal{O}(N \ln N)$ as the fitness distribution has to be computed. Hence, if we perform "tournament selection" with this algorithm we pay the lower mean variation with a higher com-

putational complexity.

## 8.2 Comparison of the Selection Intensity

| Selection Method | Selection Intensity |
|---|---|
| Tournament | $I_T(t) \approx \sqrt{2(\ln t - \ln(\sqrt{4.14 \ln t}))}$ |
| Truncation | $I_\Gamma(T) = \frac{1}{T}\frac{1}{\sqrt{2\pi}}e^{-\frac{f_c^2}{2}}$ |
| Linear Ranking | $I_R(\eta^-) = (1 - \eta^-)\frac{1}{\sqrt{\pi}}$ |
| Exponential Ranking | $I_E(\alpha) \approx 0.588\frac{\ln\ln\frac{\pi}{\alpha}}{3.69\alpha}$ |
| Fitness Proportionate | $I_P = \frac{\bar{\sigma}}{M}$ |

Table 8.2: Comparison of the selection intensity of the selection methods.

As the selection intensity is a very important property of the selection method, we give in table 8.3 some settings for the three selection methods that yield the same selection intensity.

| 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 |
| $\Omega_E$:$\alpha$ | 0.29 | 0.12 | 0.032 | $9.8 \cdot 10^{-3}$ | $3.5 \cdot 10^{-3}$ |
| $\Omega_E$:$c(N = 1000)$ | 0.999 | 0.998 | 0.997 | 0.995 | 0.994 |
| 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 | |
| $\Omega_E$:$\alpha$ | $4.7 \cdot 10^{-4}$ | $2.5 \cdot 10^{-5}$ | $10^{-9}$ | $2.4 \cdot 10^{-18}$ | |
| $\Omega_E$:$c(N = 1000)$ | 0.992 | 0.989 | 0.979 | 0.960 | |

Table 8.3: Parameter settings for truncation selection $\Omega_\Gamma$, tournament selection $\Omega_T$, linear ranking selection $\Omega_R$, and exponential ranking selection $\Omega_E$ to achieve the same selection intensity $I$.

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 $n$. Uniform crossing-over is used and assumed to be random process which 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 sufficiently large population Mühlenbein calculates

$$p(\tau) = \frac{1}{2}\left(1 + \sin(\frac{I}{\sqrt{n}}\tau + \arcsin(2p_0 - 1))\right) \tag{8.2}$$

where $p_0$ denotes the fraction of 1's in the initial random population and $p(\tau)$ the fraction of 1's in generation $\tau$. Convergence is characterized by the fact that $p(\tau_c) = 1$ so the convergence time for the special case of $p_0 = 0.5$ is given by $\tau_c = \frac{\pi}{2}\frac{\sqrt{n}}{I}$. Mühlenbein derived this formula for truncation selection, where only the selection intensity is used. Thereby 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.

For tournament selection we have

$$\tau_{T,c}(t) \approx \frac{\pi}{2}\sqrt{\frac{n}{2(\ln t - \ln\sqrt{4.14\ln t})}} \tag{8.3}$$

for truncation selection

$$\tau_{\Gamma,c}(T) = T\frac{\pi\sqrt{\pi n}}{\sqrt{2}}e^{\frac{f_c^2}{2}} \tag{8.4}$$

for linear ranking selection

$$\tau_{\Gamma,c}(\eta^-) = \frac{\pi\sqrt{\pi n}}{2(1 - \eta^-)} \tag{8.5}$$

and for exponential ranking selection

$$\tau_{E,c}(\alpha) \approx 2.671\frac{\sqrt{n}3.69^\alpha}{\ln\ln\frac{\pi}{\alpha}} \tag{8.6}$$

## 8.3  Comparison of Loss of Diversity

Table 8.4 summarizes the loss of diversity for the selection methods. It is difficult to compare these relations directly as they depend on different parameters that are characteristic for the specific selection method, e.g., the tournament size $t$ for tournament selection, the threshold $T$ for truncation selection, etc. Hence, one has to look for an independent measure to eliminate these parameters

| Selection Method | Loss of Diversity |
|---|---|
| Tournament | $p_{d,T}(t) = t^{-\frac{1}{t-1}} - t^{-\frac{t}{t-1}}$ |
| Truncation | $p_{d,\Gamma}(T) = 1 - T$ |
| Linear Ranking | $p_{d,R}(\eta^-) = (1 - \eta^-)\frac{1}{4}$ |
| Exponential Ranking | $p_{d,E}(\alpha) = \frac{1 - \ln\frac{\alpha-1}{\alpha\ln\alpha}}{\ln\alpha} - \frac{\alpha}{\alpha-1}$ |

Table 8.4: Comparison of the loss of diversity of the selection methods

and to be able to compare the loss of diversity. We chose this measure 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, $\eta^-$ for linear ranking selection, and $\alpha$ 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. (3.11), (4.4), (5.6), (6.7). Figure 8.1 shows the result of this comparison: the loss of diversity for the different selection schemes in dependence of the selection intensity. 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.

If we suppose that a lower loss of diversity is desirable as it reduces the risk of premature convergence, we expect that truncation selection should be outperformed by the other selection methods. But in general it depends on the problem and on the representation of the problem to be solved whether a low loss of diversity is "advantageous". But with figure 8.1 one has a useful tool at hand to make the right decision for a particular problem.

Another interesting fact can be observed if we look again at table 8.4: The loss of diversity is independent of the initial fitness distribution. Nowhere in the derivation of these equations a certain fitness distribution was assumed and nowhere the fitness distribution $\bar{s}(f)$ occurs in the equations. In contrary, 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.

## 8.4   Comparison of the Selection Variance

We use again the same mechanism to compare the selection variance we used in the preceding section, i.e., the selection variance is viewed as a function of the
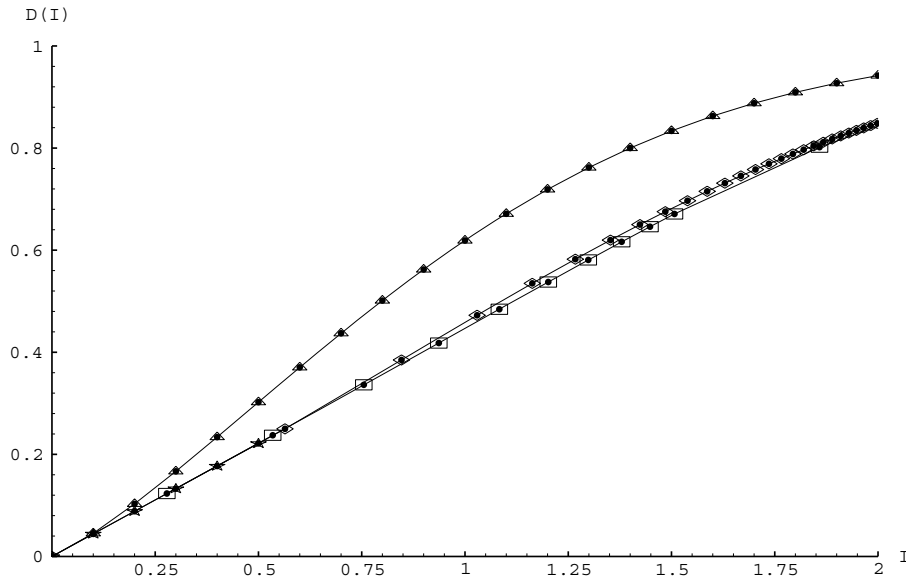
Figure 8.1: The dependence of the loss of diversity $p_d$ on the selection intensity $I$ for tournament selection ($\diamond$—$\diamond$), truncation selection ($\triangle$ − − $\triangle$), linear ranking selection ($\star$ - - $\star$), and exponential ranking selection ($\square$ − - $\square$). Note that for tournament selection only the dotted points on the graph correspond to valid (integer) tournament sizes.

selection intensity.

Figure 8.2 shows the dependence of the selection variance on the selection intensity. It can be seen clearly 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 may be difficult as it depends on the optimization task and the kind of problem to be solved whether a high selection variance is advantageous or not. But again this graph may help to decide for the "appropriate" selection method for a particular optimization problem. If we accept the assumption that a higher variance is advantageous to the optimization process, exponential ranking selection selection reveals itself to be the best selection scheme. In [Mühlenbein and Voigt, 1995] it is stated that "if two selection selection methods have the same selection intensity, the method giving the higher standard deviation of the selected parents is to be preferred". From this point of view exponential ranking selection should be the "best" selection method.

| Selection Method | Selection Variance |
|---|---|
| Tournament | $V_T(t) \approx \sqrt{\frac{2.05+t}{3.14t^{\frac{3}{2}}}}$ |
| Truncation | $V_\Gamma(T) = 1 - I_\Gamma(T)(I_\Gamma(T) - f_c)$ |
| Linear Ranking | $V_R(\eta^-) = 1 - I_R^2(\eta^-)$ |
| Exponential Ranking | $V_E(\alpha) \approx \ln\left(1.2 + \frac{2.8414}{2.225\alpha - \ln\alpha}\right)$ |

Table 8.5: Comparison of the selection variance of the selection methods.

## 8.5 The Complement Selection Schemes: Tournament and Linear Ranking

If we compare the several properties of tournament selection and linear ranking selection we observe that binary tournament behaves similar to a linear ranking selection with a very small $\eta^-$. And indeed it is possible to prove that binary tournament and linear ranking with $\eta^- = \frac{1}{N}$ have identical average behavior.

**Theorem 8.5.1** *The expected fitness distributions of linear ranking selection with $\eta^- = \frac{1}{N}$ and tournament selection with $t = 2$ are identical, i.e.*

$$\Omega_R^*(s, \frac{1}{N}) = \Omega_T^*(s, 2) \tag{8.7}$$

**Proof:**

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

$\square$

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

By this we see the complementary character of the two selection schemes. For lower selection intensities ($I \leq \frac{1}{\sqrt{\pi}}$) linear ranking selection is the appropriate selection mechanism as for selection intensities ($I \geq \frac{1}{\sqrt{\pi}}$) tournament selection is better suited. At the border the two section schemes are identical.

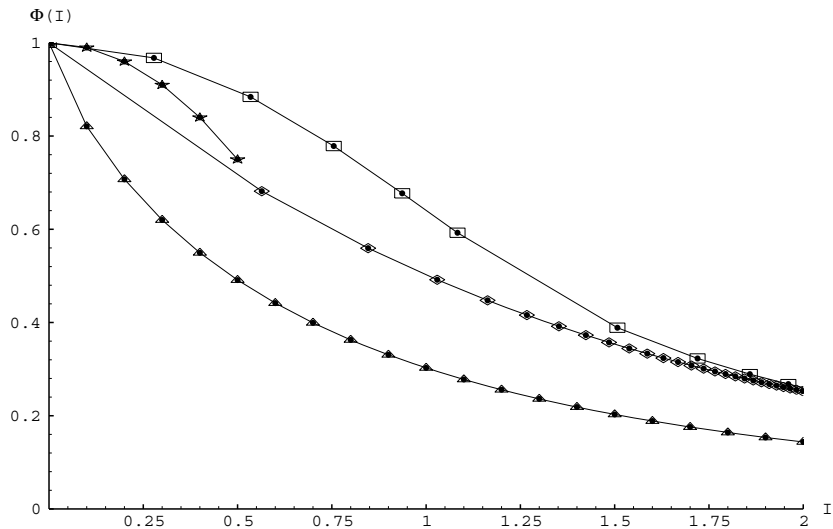Figure 8.2: The dependence of the selection variance $V$ on the selection intensity $I$ for tournament selection ($\diamond$—$\diamond$), truncation selection ($\triangle$ – – $\triangle$), ranking selection ($\star$ - - $\star$), and exponential ranking selection ($\square$ – - $\square$). Note that for tournament selection only the dotted points on the graph correspond to valid (integer) tournament sizes.

# Chapter 9

# Conclusion

In this paper a unified and systematic approach to analyze selection methods was developed and applied to the selection schemes tournament selection, truncation selection, linear and exponential ranking selection, and proportional selection. This approach is based on the description of the population using fitness distributions. Although this idea is not new, the consequent realization of this idea led to a powerful framework that gave an unified view of the selection schemes and allowed several up to now independently and isolated obtained aspects of these selection schemes to be derived with one single methodology. Besides some interesting features of selection schemes could be proven, e.g. the concatenation of several tournament selections (theorem 3.1.1) and the equivalence of binary tournament and linear ranking (theorem 8.5.1).

Furthermore the derivation of the major characteristics of a selection scheme, i.e. the selection intensity, the selection variance and the loss of diversity, could easily be achieved with this approach. The selection intensity was used to obtain a convergence prediction of the simple genetic algorithm with uniform crossover optimizing the ONEMAX function. The comparison of the loss of diversity and the selection variance based on the selection intensity allowed for the first time to compare "second order" properties of selection schemes. This comparison gives a well grounded basis to decide which selection scheme should be used, if the impact of these properties on the optimization process is known for the particular problem.

The one exception in this paper is proportional selection, that withdraws itself from a detailed mathematical analysis. But based on some basic analysis and some empirical observations we regard proportional selection to be a very unsuited selection scheme.

The presented analysis can easily be extended to other selection schemes and other properties of selection schemes.

# Appendix A

# Deriving Approximation Formulas Using Genetic Programming

In this chapter we describe the way the approximation formulas for the selection variance of tournament selection (3.17), the selection intensity of exponential ranking selection (6.9), and the selection variance of exponential ranking (6.10) were obtained.

In general we use the same approach as Koza in his first book on genetic programming [Koza, 1992]. Genetic Programming (GP) is an optimization method based on natural evolution similar to genetic algorithms. The major difference is that GP uses *trees* to represent the individuals where GA uses bit-strings. The tree structure can represent functional dependencies or complete computer programs. Hence we can use this optimization method to obtain an analytic approximation of a data set. Given are a certain number of data points $(x_i, y_i)$ and we want to find an analytic expression that approximates the functional dependence $y = u(x)$.

The fitness function is to minimize the maximum relative error over all data points $(x_i, y_i)$. If an arithmetic exception occurs during the evaluation of an individual (such as division by zero) the individual is punished by a very high error score (100.000).

The parameter for the optimization are:

- population size 10.000

- maximum tree size 15

- maximum number of generations 30

- tournament selection with tournament size 5

- reducing redundancy using marking crossover [Blickle and Thiele, 1994]

- use of one step hill-climbing to adjust the $RFPC$ numbers

The last two items need further explanation: the marking crossover introduced in [Blickle and Thiele, 1994] works as follows. During the evaluation of the fitness function all edges in the tree of the individual are marked. The edges that remain unmarked after calculating the fitness value are said to be *redundant*, because they were never used for fitness calculation. The crossover operator now only selects the edges for crossover that are marked, because only changes at these edges may lead to individuals with a different fitness score. With this approach an increase in performance of almost 50% for the 6-multiplexer problem was achieved [Blickle and Thiele, 1994].

"One step hill-climbing" works in the following way: after evaluation the fitness of an individual, successively all random constants in the trees are change by a little amount $\pm\delta$. If this change leads to a better individual it is accepted, otherwise rejected. In our experiments, the setting is $\delta = 0.1$.

The very large population size was chosen because only small trees were allowed.

No further tuning of the parameters was made, as well as no comparison of performance with other possible optimization methods (e.g. simulated annealing) as this is beyond the scope of this paper. The intention was only to find one good approximation for each data set. The problem was programmed on a SPARC Station 20 using the YAGPLIC library [Blickle, 1995]. A run over 30 generations took about 15 minutes CPU time. The given solution were found after 15 - 23 generations.

## A.1   Approximating the Selection Variance of Tournament Selection

The operators and terminal provided to the optimization method for this problem were

$$
\begin{aligned}
F &= \{Plus, Subtract, Times, Divide, Log, Sqrt\} \\
T &= \{t, \pi, RFPC\}
\end{aligned}
$$

were $RFPC$ is a random floating point number in the range from [-10,10] once determined at creation time of the population. These sets were chosen with some knowledge in mind about the possible dependency.

The following approximation was found with maximum relative error of 1.66%:
$V_T(t) \approx$ Sqrt[Divide[Plus[Sqrt[Plus[Log[Pi],Pi]],t],Times[Times[t,Pi],Sqrt[t]]]]. After simplifying this expression and some local fine tuning of the constants (3.17)

is obtained that approximates the selection variance of tournament selection with an relative error of less than 1.6% for $t \in \{1, \ldots, 30\}$:

$$V_T(t) \approx \sqrt{\frac{2.05 + t}{3.14t^{\frac{3}{2}}}} \qquad (3.17)$$

Table A.1 displays the numerical calculated values for the selection variance, the approximation by (3.17) and the relative error of the approximation for the tournament sizes $t = 1, \ldots, 30$.

## A.2 Approximating the Selection Intensity of Exponential Ranking Selection

The operators and terminal provided to the optimization method for this problem were

$$\begin{aligned} F &= \{Plus, Subtract, Times, Divide, Log, Sqrt, Exp\} \\ T &= \{\alpha, RFPC\} \end{aligned}$$

were $RFPC$ is a random floating point number in the range from [-10,10] once determined at creation time of the population.

The GP found the following approximation with an relative error of 6.3 %:
$I_E(\alpha)$=Divide[Log[Log[Divide[$\pi, \alpha$]]], Times[Sqrt[Power[Plus[8.040000, 5.468000], $\alpha$ ]], Exp[Times[3.508000, 0.150000]]]].

After some local fine tuning of the real constants and some simplifications (6.9) is obtained, that approximates the selection intensity of exponential ranking selection with and relative error of less than 5.8%.

$$I_E(\alpha) \approx 0.588 \frac{\ln \ln \frac{\pi}{\alpha}}{3.69^\alpha} \quad (6.9)$$

Table A.2 displays again the numerical calculated values for the selection intensity, the approximation by (6.9) and the relative error of the approximation.

## A.3 Approximating the Selection Variance of Exponential Ranking Selection

The operators and terminal provided to the optimization method for this problem were

$$\begin{aligned} F &= \{Plus, Subtract, Times, Divide, Log, Sqrt, Exp\} \\ T &= \{\alpha, RFPC\} \end{aligned}$$

were $RFPC$ is a random floating point number in the range from [-10,10] once determined at creation time of the population.

One solution with an accuracy of 5.4% found by GP was

$V_E(\alpha) \approx$ Log[Subtract[Divide[2.840000,Subtract[Times[Exp[0.796000],$\alpha$],Log[$\alpha$]]],-1.196000]].

Further manual tuning of the constants led to approximation formula 6.10:

$$V_E(\alpha) \approx \ln\left(1.2 + \frac{2.8414}{2.225\alpha - \ln\alpha}\right) \quad (6.10)$$

Table A.3 displays again the numerical calculated values for the selection variance, the approximation by (6.10) and the relative error of the approximation.

| Tournament size $t$ | $V_T(t)$ | Approximation (3.17) | rel. Error in % |
|---|---|---|---|
| 1 | 1 | 0.985314748118875 | 1.468525188112524 |
| 2 | 0.6816901138160949 | 0.6751186081382552 | 0.964001904186694 |
| 3 | 0.5594672037973512 | 0.5561984979283774 | 0.5842533479688358 |
| 4 | 0.4917152368747342 | 0.4906341319420119 | 0.2198640293503141 |
| 5 | 0.4475340690206629 | 0.4480145502588547 | 0.1073619354261157 |
| 6 | 0.4159271089832759 | 0.4175510364657733 | 0.3904355949452292 |
| 7 | 0.3919177761267493 | 0.39438993519557 8 | 0.6307851338769677 |
| 8 | 0.3728971432867331 | 0.3760023889838275 | 0.832735179927804 |
| 9 | 0.357353326357783 | 0.3609311128657064 | 1.00119020701134 |
| 10 | 0.3443438232607686 | 0.3482720218045281 | 1.140777989441309 |
| 11 | 0.3332474427030835 | 0.3374316422588417 | 1.255583395274936 |
| 12 | 0.3236363870477149 | 0.3280026037472064 | 1.349111803935622 |
| 13 | 0.3152053842122778 | 0.3196949671601049 | 1.424335741931216 |
| 14 | 0.3077301024704087 | 0.3122960960698358 | 1.483765664383183 |
| 15 | 0.3010415703137873 | 0.3056460664995608 | 1.52952171388625 |
| 16 | 0.2950098090102839 | 0.299621986989894 | 1.563398178210858 |
| 17 | 0.2895330036877659 | 0.2941276564766719 | 1.586918496469898 |
| 18 | 0.2845301297414324 | 0.2890865414042389 | 1.601381079377116 |
| 19 | 0.2799358049283933 | 0.2844368844693661 | 1.607897047011976 |
| 20 | 0.2756966156185853 | 0.2801282213768255 | 1.60742116775606 |
| 21 | 0.2717684436810235 | 0.2761188509706837 | 1.600777202362149 |
| 22 | 0.2681144875238161 | 0.2723739652629051 | 1.588678694101006 |
| 23 | 0.2647037741277227 | 0.2688642452925616 | 1.571746069185771 |
| 24 | 0.2615098815029825 | 0.2655647916870945 | 1.55057627681493 |
| 25 | 0.2585107005876581 | 0.2624542995840844 | 1.525507063135685 |
| 26 | 0.2556866644747772 | 0.2595144145665026 | 1.49704721581323 |
| 27 | 0.2530210522851858 | 0.2567292244751288 | 1.465558757444203 |
| 28 | 0.2504992994478195 | 0.2540848544627421 | 1.431363290367017 |
| 29 | 0.248108653859635 2 | 0.2515691413740026 | 1.394746801667437 |
| 30 | 0.24583789644110 1 | 0.249171369706327 | 1.355963955713685 |

Table A.1: Approximation of the selection variance of tournament selection.

| $\alpha$ | $I_E(\alpha)$ | Approximation (6.9) | rel. error in % |
|---|---|---|---|
| $1.10^{-20}$ | 2.21187 | 2.26634 | 2.46276 |
| $1.10^{-19}$ | 2.19127 | 2.23693 | 2.08369 |
| $1.10^{-18}$ | 2.16938 | 2.20597 | 1.6866 |
| $1.10^{-17}$ | 2.14604 | 2.17329 | 1.26989 |
| $1.10^{-16}$ | 2.12104 | 2.13869 | 0.83183 |
| $1.10^{-15}$ | 2.09416 | 2.10192 | 0.370482 |
| $1.10^{-14}$ | 2.0651 | 2.06269 | 0.116247 |
| $1.10^{-13}$ | 2.03349 | 2.02067 | 0.630581 |
| $1.10^{-12}$ | 1.99889 | 1.97541 | 1.17477 |
| $1.10^{-11}$ | 1.96069 | 1.92637 | 1.75086 |
| $1.10^{-10}$ | 1.91813 | 1.87286 | 2.36022 |
| $1.10^{-9}$ | 1.87015 | 1.81399 | 3.00251 |
| $1.10^{-8}$ | 1.81525 | 1.74857 | 3.67343 |
| $1.10^{-7}$ | 1.7513 | 1.67495 | 4.35951 |
| $1.10^{-6}$ | 1.67494 | 1.59077 | 5.02548 |
| 0.00001 | 1.58068 | 1.49248 | 5.58 |
| 0.0001 | 1.4585 | 1.37426 | 5.77587 |
| 0.001 | 1.28826 | 1.22496 | 4.91391 |
| 0.01 | 1.02756 | 1.01518 | 1.20517 |
| 0.0158489 | 0.958452 | 0.959374 | 0.0961498 |
| 0.0251189 | 0.88211 | 0.895999 | 1.57453 |
| 0.0398107 | 0.797944 | 0.823028 | 3.14361 |
| 0.0630957 | 0.705529 | 0.738058 | 4.61058 |
| 0.1 | 0.604719 | 0.638636 | 5.60873 |
| 0.125893 | 0.55122 | 0.58292 | 5.75089 |
| 0.158489 | 0.495745 | 0.523127 | 5.52332 |
| 0.199526 | 0.438398 | 0.459482 | 4.8093 |
| 0.251189 | 0.379315 | 0.392562 | 3.49231 |
| 0.316228 | 0.318668 | 0.323416 | 1.49006 |
| 0.398107 | 0.256659 | 0.253675 | 1.16253 |
| 0.501187 | 0.193519 | 0.185613 | 4.08562 |
| 0.630957 | 0.129509 | 0.122102 | 5.71875 |
| 0.794328 | 0.0649044 | 0.0663754 | 2.26635 |

Table A.2: Approximation of the selection intensity of exponential ranking selection.

| $\alpha$ | $V_E(\alpha)$ | Approximation (6.10) | rel. error in % |
|---|---|---|---|
| $1.10^{-20}$ | 0.224504 | 0.232462 | 3.54445 |
| $1.10^{-19}$ | 0.227642 | 0.235033 | 3.24672 |
| $1.10^{-18}$ | 0.231048 | 0.237881 | 2.95725 |
| $1.10^{-17}$ | 0.234767 | 0.241055 | 2.67849 |
| $1.10^{-16}$ | 0.238849 | 0.244614 | 2.41344 |
| $1.10^{-15}$ | 0.243361 | 0.248632 | 2.16573 |
| $1.10^{-14}$ | 0.248386 | 0.253204 | 1.93978 |
| $1.10^{-13}$ | 0.254032 | 0.258454 | 1.74096 |
| $1.10^{-12}$ | 0.260441 | 0.264544 | 1.57569 |
| $1.10^{-11}$ | 0.267807 | 0.271695 | 1.45147 |
| $1.10^{-10}$ | 0.276403 | 0.280208 | 1.37665 |
| $1.10^{-9}$ | 0.286619 | 0.290515 | 1.35936 |
| $1.10^{-8}$ | 0.299052 | 0.303252 | 1.40448 |
| $1.10^{-7}$ | 0.314661 | 0.319393 | 1.50382 |
| $1.10^{-6}$ | 0.335109 | 0.340517 | 1.61388 |
| 0.00001 | 0.363607 | 0.36936 | 1.58242 |
| 0.0001 | 0.407156 | 0.411119 | 0.973227 |
| 0.001 | 0.482419 | 0.47699 | 1.12538 |
| 0.01 | 0.624515 | 0.595566 | 4.63544 |
| 0.0158489 | 0.664523 | 0.631164 | 5.01997 |
| 0.0251189 | 0.70839 | 0.672819 | 5.02134 |
| 0.0398107 | 0.755421 | 0.721681 | 4.46638 |
| 0.0630957 | 0.804351 | 0.778705 | 3.18843 |
| 0.1 | 0.853263 | 0.843853 | 1.1029 |
| 0.125893 | 0.876937 | 0.878753 | 0.207166 |
| 0.158489 | 0.899606 | 0.914171 | 1.61901 |
| 0.199526 | 0.920882 | 0.948648 | 3.01517 |
| 0.251189 | 0.940366 | 0.979962 | 4.21072 |
| 0.316228 | 0.957663 | 1.00499 | 4.94227 |
| 0.398107 | 0.972403 | 1.0198 | 4.87463 |
| 0.501187 | 0.98425 | 1.02009 | 3.6411 |
| 0.630957 | 0.992927 | 1.00213 | 0.927139 |
| 0.794328 | 0.998221 | 0.964101 | 3.41803 |

Table A.3: Approximation of the selection variance of exponential ranking selection.

59

# Appendix B

# Used Integrals

$$\int xe^{-\frac{x^2}{2}} = -e^{-\frac{x^2}{2}} \tag{B.1}$$

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \, dx = \sqrt{2\pi} \tag{B.2}$$

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \, dx = \pi \tag{B.3}$$

$$\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} = 0 \tag{B.4}$$

$$\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \, dx = \sqrt{\pi} \tag{B.5}$$

$$\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} \left( \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \right)^2 \, dx = \sqrt{2}\pi \tag{B.6}$$

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \, dx = \sqrt{2\pi} \tag{B.7}$$

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \, dx = \pi \tag{B.8}$$

$$\int_{-\infty}^{\infty} t \, e^{-\frac{x^2}{2}} \left( \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \right)^{t-1} \, dx = (2\pi)^{\frac{t}{2}} \tag{B.9}$$

# Appendix C

# Glossary

| | |
|---|---|
| $\alpha$ | Parameter of Exponential Ranking ($\alpha = c^N$) |
| $c$ | Basis for Exponential Ranking |
| $\eta^-$ | Selection Probability of worst fit Individual in Ranking Selection |
| $f$ | Fitness Value |
| $f(J)$ | Fitness Value of Individual $J$ |
| $G(\mu, \sigma)$ | Gaussian Distribution with Mean $\mu$ and Variance $\sigma^2$ |
| $I$ | Selection Intensity |
| $J$ | Individual |
| $\mathbf{J}$ | Space of all Possible Individuals |
| $M$ | Average Population Fitness |
| $N$ | Population Size |
| $\Omega$ | Selection Method |
| $\Omega_E$ | Exponential Ranking Selection |
| $\Omega_T$ | Tournament Selection |
| $\Omega_\Gamma$ | Truncation Selection |
| $\Omega_P$ | Proportional Selection |
| $\Omega_R$ | Ranking Selection |
| $p_c$ | Crossover Probability |
| $p_d$ | Loss of Diversity |
| $P$ | Population |
| $R$ | Reproduction Rate |
| $\mathbf{R}$ | Set of Real Numbers |

$s$    (Discrete) Fitness Distribution

$\bar{s}$    (Continuous) Fitness Distribution

$S$    Cumulative (Discrete) Fitness Distribution

$\bar{S}$    Cumulative (Continuous) Fitness Distribution

$\bar{\sigma}$    Mean Variance of the Population Fitness

$t$    Tournament Size

$T$    Truncation Threshold

$\tau$    Generation

$\tau_c$    Convergence Time (in Generations) for the ONEMAX Example

$V$    Selection Variance

$\mathbf{Z}$    Set of Integers

# Bibliography

[Arnold *et al.*, 1992] B.C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1992.

[Bäck, 1994] Thomas Bäck. Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence (ICEC94)*, pages 57–62, 1994.

[Bäck, 1995] Thomas Bäck. Generalized convergence models for tournament- and $(\mu,\lambda)$-selection. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA95)*, San Francisco, CA, 1995. Morgan Kaufmann Publishers.

[Baker, 1987] J. E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms*, pages 14–21, Cambridge, MA, 1987. Lawrence Erlbaum Associates.

[Baker, 1989] J. E. Baker. *An Analysis of the Effects of Selection in Genetic Algorithms*. PhD thesis, Graduate School of Vanderbilt University, Nashville, Tennessee, 1989.

[Blickle and Thiele, 1994] Tobias Blickle and Lothar Thiele. Genetic programming and redundancy. In J. Hopf, editor, *Genetic Algorithms within the Framework of Evolutionary Computation (Workshop at KI-94, Saarbrücken)*, pages 33–38. Max-Planck-Institut für Informatik (MPI-I-94-241), 1994.

[Blickle and Thiele, 1995] Tobias Blickle and Lothar Thiele. A mathematical analysis of tournament selection. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA95)*, pages 9–16, San Francisco, CA, 1995. Morgan Kaufmann Publishers.

[Blickle, 1995] Tobias Blickle. *YAGPLIC - User Manual*. Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich, 1995.

[Brill *et al.*, 1992] F. Z. Brill, D. E. Brown, and W. N. Martin. Fast genetic selection of features for neural network classifiers. *IEEE Transactions on Neural Networks*, 2(3):324–328, March 1992.

[Bulmer, 1980] M.G. Bulmer. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, 1980.

[Crow and Kimura, 1970] J.F. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. Harper and Row, New York, 1970.

[de la Maza and Tidor, 1993] Michael de la Maza and Bruce Tidor. An analysis of selection procedures with particular attention paid to proportional and bolzmann selection. In Stefanie Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 124–131, San Mateo, CA, 1993. Morgan Kaufmann Publishers.

[Goldberg and Deb, 1991] David E. Goldberg and Kalyanmoy Deb. A comparative analysis of selection schemes used in genetic algorithms. In G. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 69–93, San Mateo, 1991. Morgan Kaufmann.

[Goldberg, 1989] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1989.

[Grefenstette and Baker, 1989] John J. Grefenstette and James E. Baker. How genetic algorithms work: A critical look at implicit parallelism. In J. David Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 20 – 27, San Mateo, CA, 1989. Morgan Kaufmann Publishers.

[Henrici, 1977] P. Henrici. *Applied and Computational Complex Analysis*, volume 2. A Wiley-Interscience Series of Texts, Monographs, and Tracts, 1977.

[Holland, 1975] John H. Holland. *Adaption in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI, 1975.

[Koza, 1992] John R. Koza. *Genetic programming: on the programming of computers by means of natural selection*. The MIT Press, Cambridge, Massachusetts, 1992.

[Mühlenbein and Schlierkamp-Voosen, 1993]
Heinz Mühlenbein and Dirk Schlierkamp-Voosen. Predictive models for the breeder genetic algorithm. *Evolutionary Computation*, 1(1), 1993.

[Mühlenbein and Voigt, 1995] Heinz Mühlenbein and Hans-Michael Voigt. Gene pool recombination in genetic algorithms. In I. H. Osman and J. P. Kelly, editors, *Proceedings of the Metaheuristics Inter. Conf.*, Norwell, 1995. Kluwer Academic Publishers.

[Shapiro *et al.*, 1994] Jonathan Shapiro, Adam Prügel-Bennett, and Magnus Rattray. A statistical mechanical formulation of the dynamics of genetic algorithms. In Terence C. Fogarty, editor, *Evolutionary Computing AISB Workshop*. Springer , LNCS 865, 1994.

[Thierens and Goldberg, 1994a] D. Thierens and D. Goldberg. Convergence models of genetic algorithm selection schemes. In Yuval Davidor, Hans-Paul Schwefel, and Reinhard Männer, editors, *Parallel Problem Solving from Nature - PPSN III*, pages 119 – 129, Berlin, 1994. Lecture Notes in Computer Science 866 Springer-Verlag.

[Thierens and Goldberg, 1994b] Dirk Thierens and David Goldberg. Elitist recombination: an integrated selection recombination ga. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence (ICEC94)*, pages 508–512, 1994.

[Whitley, 1989] Darrell Whitley. The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In J. David Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 116 – 121, San Mateo, CA, 1989. Morgan Kaufmann Publishers.