on Evolutionary Computation. IEEE World Congress on Computational Intelligence (ICEC94), 508–512.

Whitley, D. (1989). 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* (pp. 116 – 121), San Mateo, CA. Morgan Kaufmann Publishers.

Genetic Algorithm Laboratory, University of Urbana-Champaign, Urbana, Illinois 61801.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm. *Evolutionary Computation*, 1(1).

Nagaraja, H. (1982). Selection differentials. *Encyclopedia of Statistical Science*, 334–336.

Shapiro, J., Prügel-Bennett, A., & Rattray, M. (1994). A statistical mechanical formulation of the dynamics of genetic algorithms. In Terence C. Fogarty, editor, *Evolutionary Computing AISB Workshop.* Springer , LNCS 865.

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

Thierens, D., & Goldberg, D. (1994b). Elitist recombination: an integrated selection recombination ga. In *Proceedings of the First IEEE Conference*

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

De la Maza, M., & Tidor, B. (1993). 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* (pp. 124–131), San Mateo, CA. Morgan Kaufmann Publishers.

Goldberg, D.E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. In G. Rawlins, editor, *Foundations of Genetic Algorithms* (pp. 69–93), San Mateo. Morgan Kaufmann.

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

Koza, J.R. (1992). *Genetic programming: on the programming of computers by means of natural selection.* The MIT Press, Cambridge, Massachusetts.

Miller, B.L. & Goldberg, D.E. (1995). Genetic algorithms, tournament selection, and the effects of noise. Technical Report 95006, The Illinois *tional Conference on Genetic Algorithms (ICGA95)*, San Francisco, CA. Morgan Kaufmann Publishers.

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

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

Blickle, T., & Thiele, L. (1995). A mathematical analysis of tournament selection. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA95)* (pp. 9–16), San Francisco. Morgan Kaufmann Publishers.

Blickle, T. (1996). *Theory of Evolutionary Algorithms and Application to System-Synthesis.* PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland, 1996. ETH diss no. 11894, VDF-Verlag Zurich, ISBN 3 7281 2433 8.

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

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

References

Arnold, B.C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A First Course in Order Statistics.* Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Bäck, Th. (1994a). *Evolutionary Algorithms in Theory and Practice.* PhD thesis, Fachbereich Informatik, Universität Dortmund.

Bäck, Th. (1994b). 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)*, 57–62.

Bäck, Th. (1995). Generalized convergence models for tournament- and (μ,λ)-selection. In L. Eshelman, editor, *Proceedings of the Sixth Interna-*

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

- Some interesting features of selection schemes could be proven, e.g., the concatenation of several tournament selections (Theorem 3.3) and the equivalence of binary tournament and linear ranking (Theorem 6.1).

- Important characteristics of selection schemes, like selection intensity, selection variance, and loss of diversity were derived.

- 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 "second order" properties of selection schemes to be compared. This gives a well grounded basis for deciding which selection scheme is applicable, assuming that for a particular problem the impact of these properties on the optimization process is known.

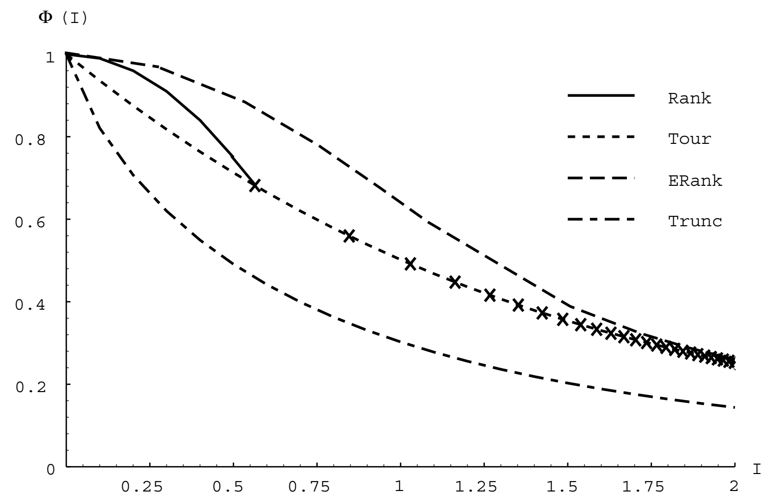

Figure 16: The dependence of the selection variance Φ on the selection intensity I for tournament selection (Tour), truncation selection (Trunc), ranking selection (Rank), and exponential ranking selection (ERank). Note that for tournament selection only the points marked as crosses on the graph correspond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

the selection, i.e., the fitness distributions after selection were compared with the fitness distributions before selection. Although this idea is not new, the consequent realization of this idea led to a powerful framework that allowed the following results to be obtained:

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

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

An interpretation of the results is 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 decide the “appropriate” selection method for a particular optimization problem. Under the assumption that a higher variance is advantageous to the optimization process, exponential ranking selection reveals itself to be the best selection scheme.

7 Conclusions

In this paper some common selection schemes were analyzed. The analysis was based on fitness distributions and examined only the local behavior of 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 Fig. 14 one has a useful tool at hand to make the right decision for a particular problem.

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

6.5 Comparison of the Selection Variance

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

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

can be seen, in almost all cases this difference is positive, i.e., truncation selection has a higher loss of diversity as tournament selection *for arbitrary fitness distributions*. Furthermore, the computed loss of diversity seems to be an accurate approximation and upper bound for this difference.

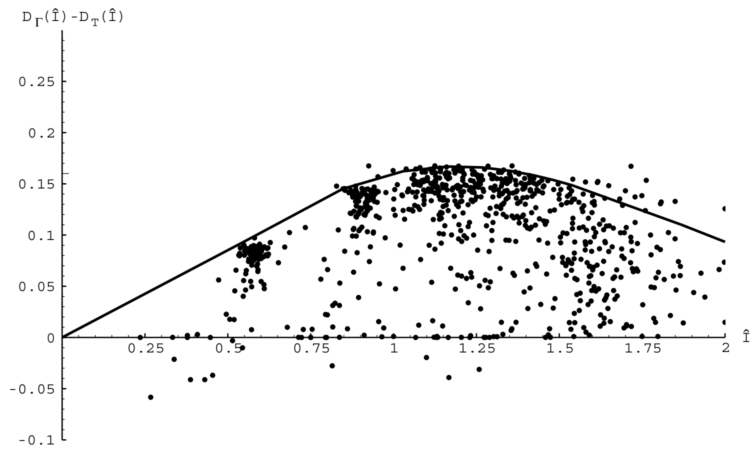

Figure 15: The measured difference of the loss of diversity of truncations selection and tournament selection as a function of the realized selection intensity. The solid line gives the difference according to the standardized selection intensity (as displayed in Fig. 14), the points give the differences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

Supposing that a lower loss of diversity is desirable as it reduces the risk of premature convergence, one expects that truncation selection should be schemes, respectively. This means that more "genetic material" is lost using truncation selection.

The graph in Fig. 14 is based on the standardized selection intensity as given by Definition 2.10, i.e., a normal distribution of the initial fitness values has been assumed. Furthermore, continuous distributions have been used to derive the above results. In order to estimate their generality and the generality of the continuous calculations, the following experiment has been set up: 800 different fitness distribution have been randomly created and used as input to tournament selection and truncation selection. Eight different parameter settings for the selection methods have been used (corresponding to the first eight entries in Table 3). The *realized selection intensity* $\hat{I}$ and the loss of diversity have been measured for each fitness distribution. The realized selection intensity has been computed according to the general definition of selection intensity (Def. 2.9), i.e.,

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

The difference between the loss of diversity of both methods for a given selection intensity is shown in Fig. 15. The solid line is the theoretical expected difference according to the standardized selection intensity (as in Fig. 14). The dots are the measured differences of the loss of diversity. As

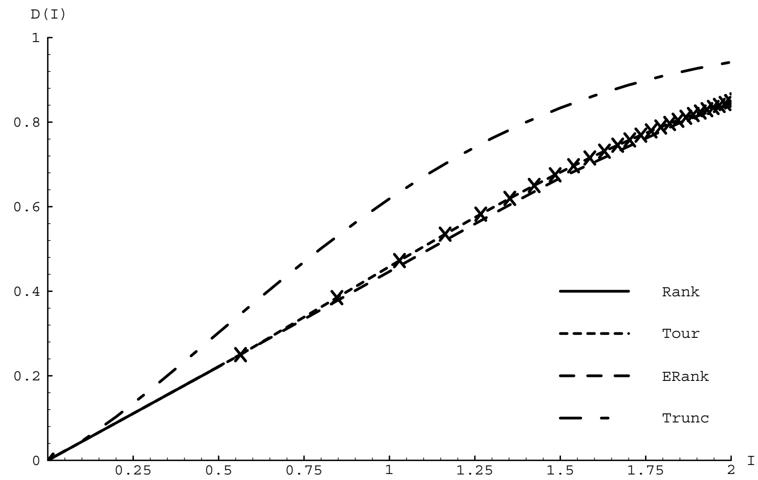

Figure 14: The dependence of the loss of diversity D on the selection intensity I for tournament selection (Tour), truncation selection (Trunc), linear ranking selection (Rank) and exponential ranking selection (ERank). Note that for tournament selection only the points marked as crosses on the graph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

lection intensity. Tournament and exponential ranking behave almost identical. Ranking selection only covers the lower quarter of the I-axis, as the maximum selection intensity of ranking selection is $I_R(0) = \frac{1}{\sqrt{\pi}} \approx 0.56$. In this region it behaves also very similar to exponential ranking. To achieve the same selection intensity more bad individuals are replaced using truncation selection than using tournament selection or one of the ranking selection

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

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

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 and to be able to compare the loss of diversity. This measure is chosen to be the selection intensity. The loss of diversity of the selection methods is viewed as a function of the selection intensity. To calculate the corresponding graph one first computes the value of the parameter of a selection method (i.e., t for tournament selection, T for truncation selection, η^- for linear ranking selection, and κ for exponential ranking selection) that is necessary to achieve a certain selection intensity. With this value the loss of diversity is then obtained using the corresponding equations, i.e., (22), (30), (40), (49). Figure 14 shows the result of this comparison: the loss of diversity for the different selection schemes in dependence of the sefact that $p(k_c) = 1$ so the convergence time for the special case of $p_0 = 0.5$ is given by $k_c = \frac{\pi}{2}\frac{\sqrt{n}}{I}$. It is straightforward to give the convergence time for any other selection method, by substituting I with the corresponding terms derived in the preceding sections.

For tournament selection one calculates

$$k_{T,c}(t) \approx \frac{\pi}{2}\sqrt{\frac{l}{2(\ln t - \ln\sqrt{4.14\ln t})}} \tag{56}$$

for truncation selection

$$k_{\Gamma,c}(T) = T\frac{\pi\sqrt{\pi l}}{\sqrt{2}}e^{\frac{f_c^2}{2}} \tag{57}$$

for linear ranking selection

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

and for exponential ranking selection

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

6.4 Comparison of Loss of Diversity

Table 4 summarizes the loss of diversity for the selection methods. It is difficult to compare these relations directly as they depend on different parameters that are characteristic of the specific selection method, e.g., the

I	0.34	0.56	0.84	1.03	1.16
Ω_T:t	-	2	3	4	5
Ω_R:η^-	0.4	0	-	-	-
Ω_Γ:T	0.8	0.66	0.47	0.36	0.30
Ω_E:κ	0.29	0.12	0.032	$9.8 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$

I	1.35	1.54	1.87	2.16
Ω_T:t	7	10	20	40
Ω_Γ:T	0.22	0.15	0.08	0.04
Ω_E:κ	$4.7 \cdot 10^{-4}$	$2.5 \cdot 10^{-5}$	10^{-9}	$2.4 \cdot 10^{-18}$

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

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

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

1993) a prediction is made for a genetic algorithm optimizing the ONEMAX (or bit-counting) function. Here the fitness is given by the number of 1's in the binary string of length l. Uniform crossover is used and assumed to be a random process 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 the authors calculate

$$p(k) = \frac{1}{2}\left(1 + \sin(\frac{I}{\sqrt{l}}k + \arcsin(2p_0 - 1))\right) \tag{55}$$

where p_0 denotes the fraction of 1's in the initial random population and $p(k)$ the fraction of 1's in generation k. Convergence is characterized by the

Theorem 6.1 *The expected fitness distributions of linear ranking selection with* $\eta^- = \frac{1}{N}$ *and tournament selection with* $t = 2$ *are identical, i.e.*

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

Proof:

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

□

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

6.3 Comparison of the Selection Intensity

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

The importance of the selection intensity is based on the fact that the behavior of a simple Genetic Algorithm can be predicted if the fitness distribution is normally distributed. In (Mühlenbein & Schlierkamp-Voosen,

6.2 The Complement Selection Schemes: Tournament and Linear Ranking

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

Algorithm 6: (Universal Selection Method)

Input: The population $\vec{J} = (J_1, \ldots, J_N)$,

the fitness values of the population $\vec{\phi} = (\phi_1, \ldots, \phi_N)$.

Output: The population after selection $\vec{J'}$

universal_selection($\vec{J}, \vec{\phi}$):

$\vec{s} \leftarrow$ fitness_distribution($\vec{\phi}$)

$\vec{r} \leftarrow$ reproduction_rate($\vec{s}$)

$\vec{J'} \leftarrow$ SUS($\vec{J}, \vec{r}$)

return $\vec{J'}$

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 are any performance advantages in using SUS, it makes the run of a selection method more "predictable".

To be able to apply SUS one has to know the expected number of offspring of each individual. Baker has applied this sampling method only to linear ranking selection as here the expected number of offspring is known by construction (see Section 4). As the offspring values have been derived for the selection methods discussed in the previous sections (see Table 4.1), it is possible to use stochastic universal sampling for all these selections schemes. This leads to an unified view of selection schemes and allows the construction of an "universal selection method" in the following way: First the fitness distribution of the population is computed. Next the expected reproduction rates are calculated using the equations derived in the preceding sections (Table 4.1). In the last step SUS is used to obtain the new population after selection (Algorithm 6). The time complexity of the universal selection method is $\mathcal{O}(N \ln N)$ as the fitness distribution has to be computed. Hence, if "tournament selection" is performed with this algorithm the lower mean variation is paid with a higher computational complexity.

Algorithm 5: (Stochastic Universal Sampling)

Input: The population $\vec{J}$ and the reproduction rate for each

fitness value $\vec{R} = (R_1, \ldots, R_n)$

Output: The population after selection $\vec{J}'$

SUS($\vec{J}$,$\vec{R}$):

$sum \leftarrow 0$

$j \leftarrow 1$

$ptr \leftarrow$ random[0,1)

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

$sum \leftarrow sum + R_k$ where R_k 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 $\{\vec{J}'\}$

in the outcome is observed (see also Section 2 and Theorem 2.1). This is also the case for tournament selection although there is no explicitly used roulette wheel sampling.

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

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

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

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

derived using the exact offspring equations (17), (27), (37), and (45) with some additional algebraic manipulations.

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

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

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 mean variation variance with an relative error of less than 2.8% for $\kappa \in [10^{-20}, 0.8]$:

$$\Phi_E(\kappa) \approx 0.1 + \kappa^{\frac{1}{4}}(e^{\frac{-\kappa}{2}} + 0.2428\kappa^{-\frac{15}{64}}) \quad (52)$$

The selection variance is shown in Fig. 12.

6 Comparison of Selection Schemes

In this section the selection methods are compared according to their properties derived in the preceding sections. First the reproduction rates of selection methods are compared and an unified view of selection schemes is derived. Section 6.3 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.

6.1 Reproduction Rate and Universal Selection

The reproduction rate gives the number of expected offspring of an individual with a certain fitness value after selection. In the preceding sections only the reproduction rate for the continuous case has been considered. Table 4.1 gives the equations for the discrete (exact) case. They have been

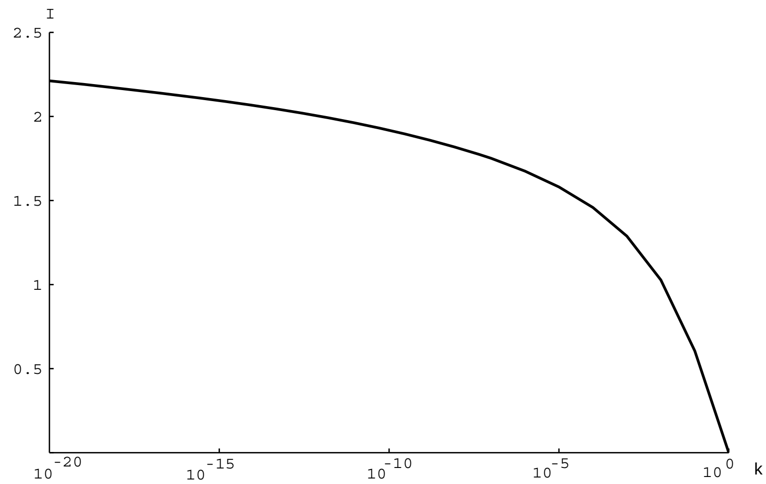

Figure 13: The selection intensity $I_E(\kappa)$ of exponential ranking selection. Note the logarithmic scale of the κ-axis.

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

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

Similar, an approximation for the selection variance of exponential ranking selection can be found. The following formula approximates the selection

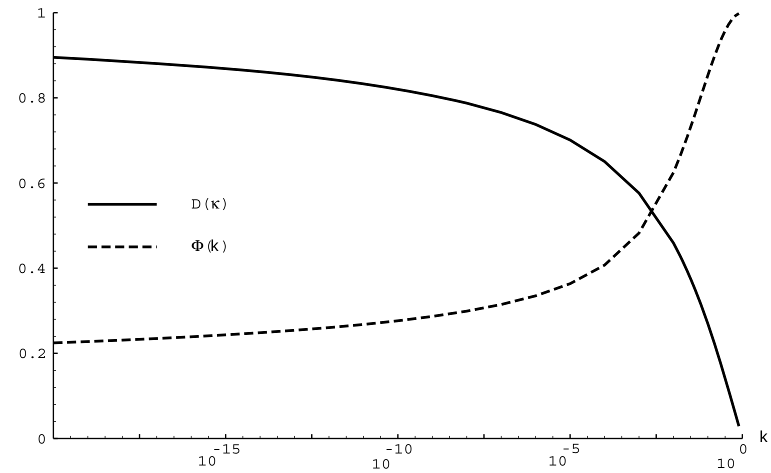

Figure 12: The loss of diversity $D_E(\kappa)$ (solid line) and the selection variance $\Phi_E(\kappa)$ (broken line) of exponential ranking selection. Note the logarithmic scale of the κ-axis.

5.3 Selection Intensity and Selection Variance

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

rate $\bar{R}(f_n) = \frac{\ln\kappa}{\kappa-1}$. Hence a natural explanation of the variable κ is obtained, as $\frac{\bar{R}(f_0)}{\bar{R}(f_n)} = \kappa$: it describes the ratio of the reproduction rate of the worst and the best individual. Note that $c < 1$ and hence $c^N \ll 1$ for large N, i.e., the interesting region of values for κ is in the range from $[10^{-20}, 1]$.

5.2 Loss of Diversity

Theorem 5.3 *The loss of diversity $D_E(\kappa)$ of exponential ranking selection is*

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

Proof: ¿From the demand $R(f_z) = 1$ one calculates:

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

Using Theorem 2.2 leads to:

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

□

The loss of diversity is shown in Fig. 12.

Proof: As the continuous form of (44) is given by $\bar{p}(x) = \frac{c^{N-x}}{\int_0^N c^{N-x}}$ and $\int c^x = \frac{1}{\ln c} c^x$ one calculates:

$$\begin{aligned} \bar{S}^*(f) &= N \frac{c^N \ln c}{c^N - 1} \int_0^{\bar{S}(f)} c^{-x} \, dx \\ &= -N \frac{c^N}{c^N - 1} [c^{-x}]_0^{\bar{S}(f)} \\ &= -N \frac{c^N}{c^N - 1} \left(1 + c^{-S(f)}\right) \end{aligned}$$

As $\bar{s}^*(f) = \frac{d\bar{S}^*(f)}{df}$, (46) follows. □

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

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

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

5.1 Reproduction Rate

Corollary 5.1 *The reproduction rate of exponential ranking selection is*

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

This is directly obtained by substituting (47) in (10).

Equation (48) shows that the worst individual has the lowest reproduction rate $\bar{R}(f_0) = \frac{\kappa \ln \kappa}{\kappa - 1}$ and the best individual has the highest reproduction

Example 5.1 *Using the discrete fitness distribution from Example 2.1 the fitness distribution shown in Figure 11 is obtained after applying exponential ranking selection with* $c = 0.986$ *(*$N = 250$*).*

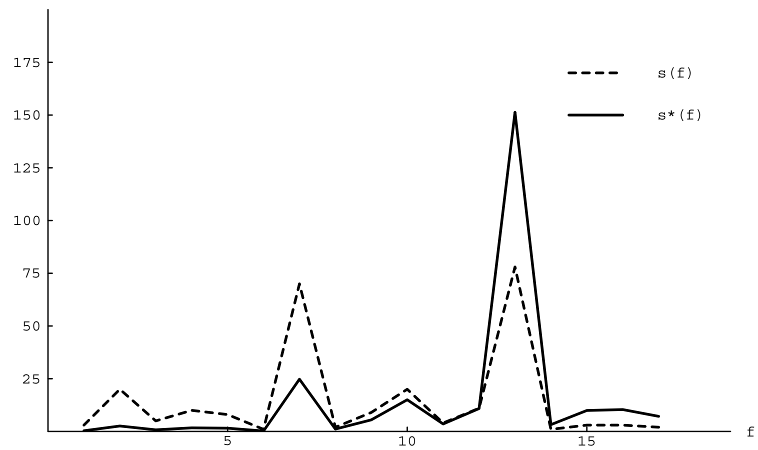

Figure 11: The initial fitness distribution (dashed) and the resulting expected fitness distribution after applying exponential ranking selection with $c = 0.986$ ($N = 250$).

Theorem 5.2 *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 parameter* c *on the distribution* $\bar{s}$ *is*

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

Proof: First the expected number of individuals with fitness f_i or worse is calculated, 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 $S^*(f_i)$ least fit individuals:

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

and with the substitution $k = N - j$

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

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

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

□

Algorithm 4: (Exponential Ranking Selection)

Input: The population $J = (J_1, \ldots, J_N)$,

the fitness values of the population $\vec{\phi} = (\phi_1, \ldots, \phi_N)$,

the ranking base $c \in]0, 1]$

Output: The population after selection $\vec{J'} = (J'_1, \ldots, J'_N)$

exponential_ranking(c,$\vec{J}$, $\vec{\phi}$):

$\vec{J}^* \leftarrow$ sorted population $\vec{J}$ according to fitness

with least fit individual at the first position

$sum_0 \leftarrow 0$

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

$sum_i \leftarrow sum_{i-1} + p_i$ (Equation 44)

od

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

$r \leftarrow$ random[0,1[

$J'_i \leftarrow J^*_l$ such that $sum_{l-1} \leq r < sum_l$

od

return $\vec{J'}$

tial ranking selection with parameter c on the distribution s is

$$\Omega_E^*(s, c, N)(f_i) = s^*(f_i) = N \frac{c^N}{c^N - 1} c^{-S(f_i)} \left(c^{s(f_i)} - 1 \right) \qquad (45)$$

5 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 to 1 the lower is the "exponentiality" of the selection method. The meaning and the influence of this parameter will be discussed in detail. Again the rank N is assigned to the best individual and the rank 1 to the worst individual. Hence the probabilities of the individuals are given by

$$p_i = \frac{c^{N-i}}{\sum_{j=1}^{N} c^{N-j}} \; ; \qquad i \in \{1, ..., N\} \tag{43}$$

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

$$p_i = \frac{c-1}{c^N - 1} c^{N-i} \; ; \qquad i \in \{1, ..., N\} \tag{44}$$

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 5.1 *The expected fitness distribution after performing exponen-*

4.4 Selection Variance

Theorem 4.5 *The selection variance of ranking is*

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

Proof: Substituting (38) into the definition equation (16) leads to

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

Using the relations $\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}$ and $\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy\, dx = \pi$ leads to

$$\begin{aligned}
\Phi_R(\eta^-) &= \eta^- + (1-\eta^-) - I_R(\eta^-)^2 \\
&= 1 - I_R(\eta^-)^2
\end{aligned}$$

□

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

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

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

Using the relations $\int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} = 0$ and $\int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} \left(\int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \right)^2 \; dx = \sqrt{2}\pi$ (41) follows. □

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

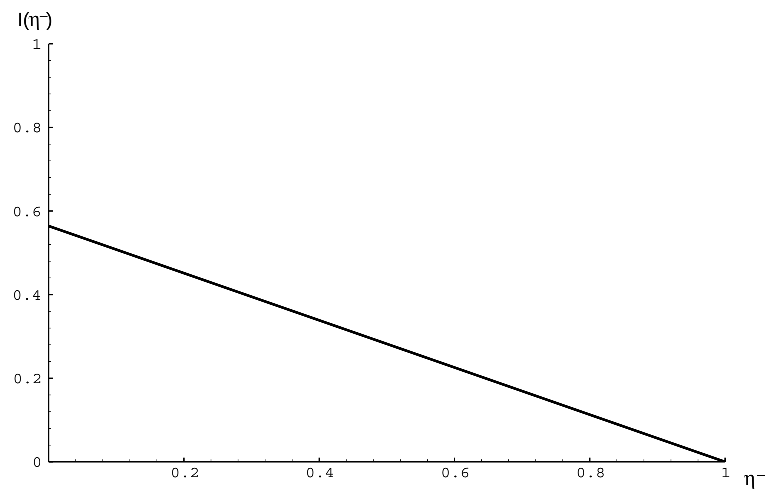

Figure 10: The selection intensity $I_R(\eta^-)$ of ranking selection.

□

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. The loss of diversity is depicted in Fig. 9.

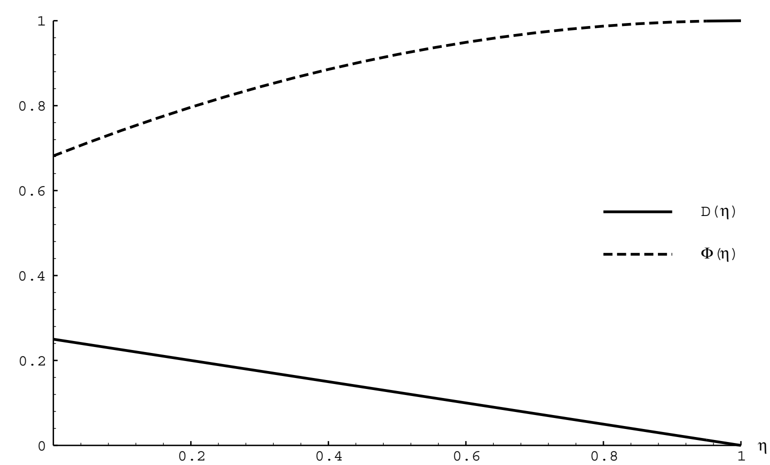

Figure 9: The loss of diversity $D_R(\eta)$ (solid line) and the selection variance $\Phi_R(\eta)$ (broken line) of linear ranking selection.

4.3 Selection Intensity

Theorem 4.4 *The selection intensity of ranking selection is*

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

4.1 Reproduction Rate

Corollary 4.1 *The reproduction rate of ranking selection is*

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

This is directly obtained by substituting (38) in (10).

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

4.2 Loss of Diversity

Theorem 4.3 *The loss of diversity $D_R(\eta^-)$ of ranking selection is*

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

Proof: Using Theorem 2.2 and realizing that $S(f_z) = \frac{N}{2}$ it follows that:

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

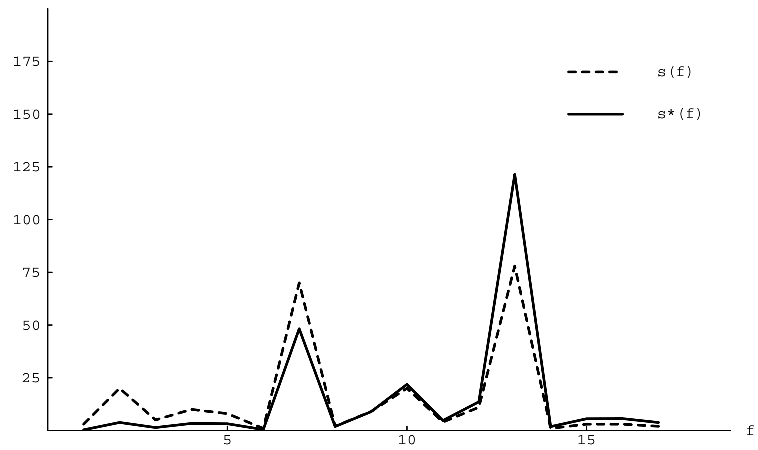

Figure 8: The initial fitness distribution (dashed) and the resulting expected fitness distribution after applying linear ranking selection with $\eta^- = 0.1$.

tion $\overline{\Omega}_R$ *with* η^- *on the distribution* $\bar{s}$ *is*

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

Proof: As the continuous form of (35) is given by $\bar{p}(x) = \frac{1}{N}(\eta^- + \frac{\eta^+ - \eta^-}{N}x)$ one calculates $\bar{S}(f)$ using $\eta^+ = 2 - \eta^-$:

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

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

$S^*(f_i)$ least fit individuals:

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

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

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

□

Example 4.1 *Using the discrete fitness distribution from Example 2.1 the fitness distribution shown in Figure 8 is obtained after applying linear ranking selection with* $\eta^- = 0.1$.

Theorem 4.2 *Let* $\bar{s}(f)$ *be the continuous fitness distribution of the population. Then the expected fitness distribution after performing ranking selec-*

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

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

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 (35) with $\eta^+ = c$ (Bäck, 1994b).

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

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

Proof: First the expected number of individuals with fitness f_i or worse is calculated, 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

Algorithm 3: (Linear Ranking Selection)

Input: The population $\vec{J} = (J_1, \ldots, J_N)$,

the fitness values of the population $\vec{\phi} = (\phi_1, \ldots, \phi_N)$,

the reproduction rate of the worst individual $\eta^- \in [0, 1]$

Output: The population after selection $\vec{J'} = (J'_1, \ldots, J'_N)$

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

$\vec{J}^* \leftarrow$ sorted population $\vec{J}$ according fitness $\vec{\phi}$

with least fit individual at the first position

$sum_0 \leftarrow 0$

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

$sum_i \leftarrow sum_{i-1} + p_i$ (Equation 35)

od

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

$r \leftarrow$ random[0,1[

$J'_i \leftarrow J^*_l$ such that $sum_{l-1} \leq r < sum_l$

od

return $\vec{J'}$

4 Linear Ranking Selection

Linear ranking selection was first suggested by Baker to eliminate the serious disadvantages of fitness proportional selection (Grefenstette & Baker, 1989; Whitley, 1989). For linear 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. The selection probability is linearly assigned to the individuals according to their rank:

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

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

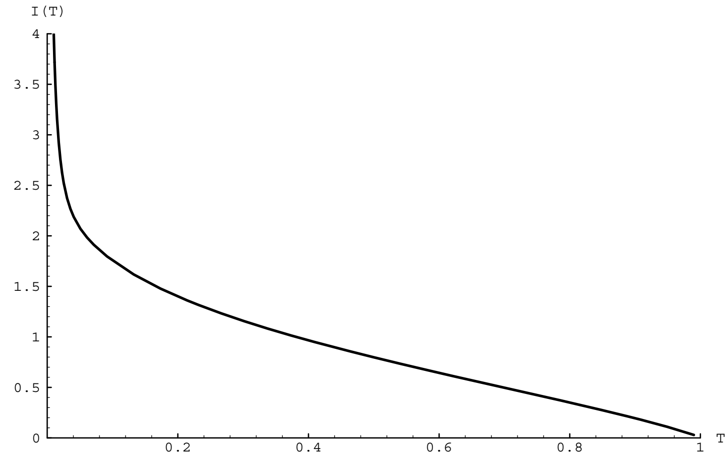

Figure 7: The Selection intensity $I_\Gamma(T)$ of truncation selection.

Proof: The substitution of (28) in the definition equation (16) gives

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

A partial integration with $u(f) = f$ and $\frac{\partial v}{\partial f} = f e^{-\frac{f^2}{2}}$ yields

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

Substituting (31) and (32) simplifies this this equation to (34). □

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

replaced by f_c. Here f_c is determined by

$$\bar{S}(f_c) = (1-T)N = 1 - T \tag{33}$$

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

Hence, I_Γ can be computed as

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

Here f_c is determined by (33). Solving (33) for T yields

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

□

A lower bound for the selection intensity reported by (Nagaraja, 1982) is $I_\Gamma(T) \le \sqrt{\frac{1-T}{T}}$.

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

3.9 Selection Variance

Theorem 3.8 *The selection variance of truncation selection is*

$$\Phi_\Gamma(T) = 1 - I_\Gamma(T)(I_\Gamma(T) - f_c) \tag{34}$$

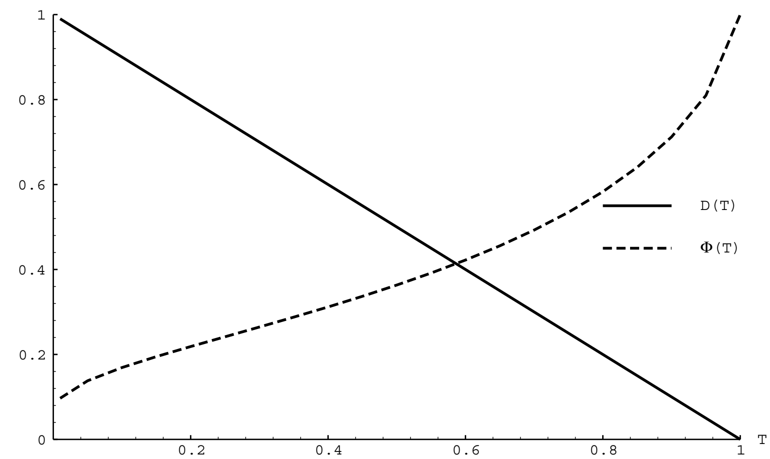

Figure 6: The loss of diversity $D_\Gamma(t)$ (solid line) and the selection variance $\Phi_\Gamma(t)$ (broken line) of truncation selection.

Theorem 3.7 *The selection intensity of truncation selection is*

$$I_\Gamma(T) = \frac{1}{T}\frac{1}{\sqrt{2\pi}}e^{-\frac{f_c^2}{2}} \tag{31}$$

where f_c is determined by

$$T = \int_{f_c}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{f^2}{2}}df. \tag{32}$$

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

3.6 Reproduction Rate

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

Corollary 3.2 *The reproduction rate of truncation selection is*

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

This is directly obtained by substituting (28) in (10).

3.7 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

$$D_\Gamma(T) = 1 - T \tag{30}$$

The loss of diversity is depicted in Fig. 6.

3.8 Selection Intensity

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

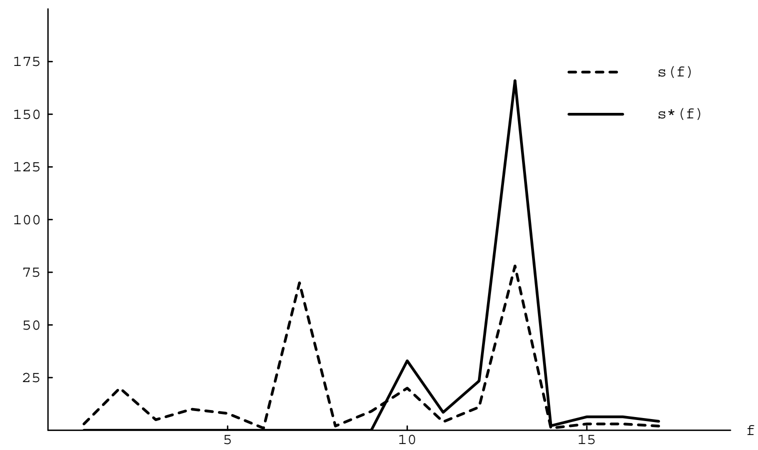

Figure 5: The initial fitness distribution (dashed) and the resulting expected fitness distribution after applying truncation selection with truncation threshold $T = 0.47$.

with threshold T is

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

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. $\Box$

properties are derived.

Theorem 3.5 *The expected fitness distribution after performing truncation selection with threshold T on the distribution s is*

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

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

Example 3.2 *Using the discrete fitness distribution from Example 2.1 the fitness distribution shown in Figure 5 is obtained after applying truncation selection with threshold $T = 0.47$.*

Theorem 3.6 *Let $\bar{s}(f)$ be the continuous distribution of the population. Then the expected fitness distribution after performing truncation selection*

Algorithm 2: (Truncation Selection)

Input: The population $\vec{J} = (J_1, \ldots, J_N)$,

the fitness values of the population $\vec{\phi} = (\phi_1, \ldots, \phi_N)$,

the truncation threshold $T \in [0, 1]$

Output: The population after selection $\vec{J'} = (J'_1, \ldots, J'_N)$

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

$\vec{J}^* \leftarrow$ sorted population $\vec{J}$ according fitness $\vec{\phi}$

with least fit individual at the first position

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

$r \leftarrow$ uniformrandom$\{$ $[(1-T)N], \ldots, N\}$

$J'_i \leftarrow J^*_r$

od

return $\vec{J'}$

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

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

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

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

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

The outline of the algorithm is given by Algorithm 2.

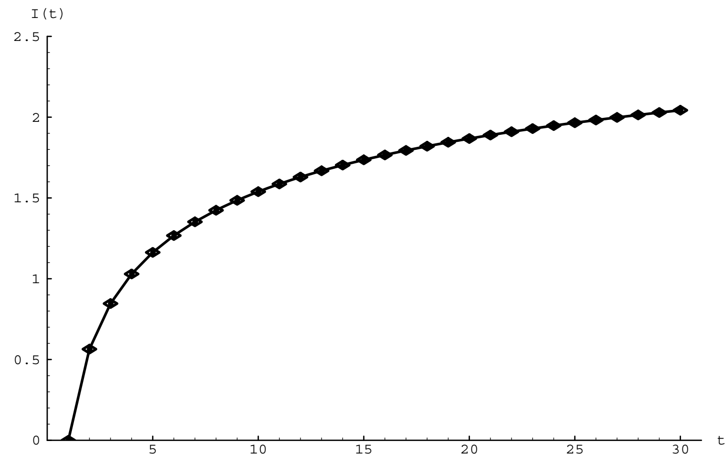

Figure 4: Dependence of the selection intensity $I_T(t)$ on the tournament size t.

3.5 Selection Variance

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

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

For a binary tournament one calculates

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

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

garaja, 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 & Goldberg 1994a) in a completely different manner. However their formulation can not be extended to other tournament sizes.

For larger tournament sizes (23) can be accurately evaluated by numerical integration. The result is shown in Fig. 4 for a tournament size from 1 to 30. The following equation approximates (23) 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% (Blickle & Thiele, 1995)

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

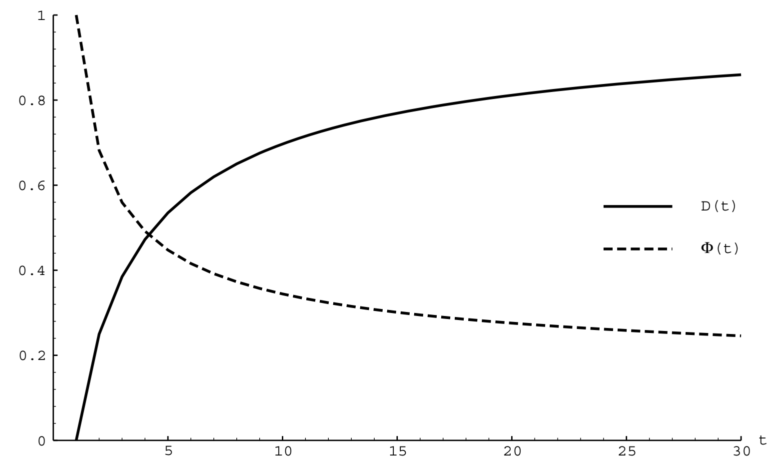

Figure 3: The loss of diversity $D_T(t)$ (solid line) and the selection variance $\Phi_T(t)$ (broken line) of tournament selection.

3.4 Selection Intensity

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

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

These integral equations can be solved analytically for the cases $t = 1, \ldots, 5$ ((Blickle & Thiele, 1995; Bäck, 1995; Arnold, Balakrishnan & Na-

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

3.2 Reproduction Rate

Corollary 3.1 *The reproduction rate of tournament selection is*

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

This is directly obtained by substituting (18) in (10).

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

3.3 Loss of Diversity

Theorem 3.4 *The loss of diversity D_T of tournament selection is*

$$D_T(t) = t^{-\frac{1}{t-1}} - t^{-\frac{t}{t-1}} \tag{22}$$

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

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

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

3.1 Concatenation of Tournament Selection

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

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

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

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

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

Equation (17) shows the strong influence of the tournament size t on the behavior of the selection scheme. Obviously for $t = 1$ (in average) the unchanged initial distribution is obtained as $\Omega_T^*(s,1)(f_i) = N\left(\frac{S(f_i)}{N} - \frac{S(f_{i-1})}{N}\right) = S(f_i) - S(f_{i-1}) = s(f_i)$.

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

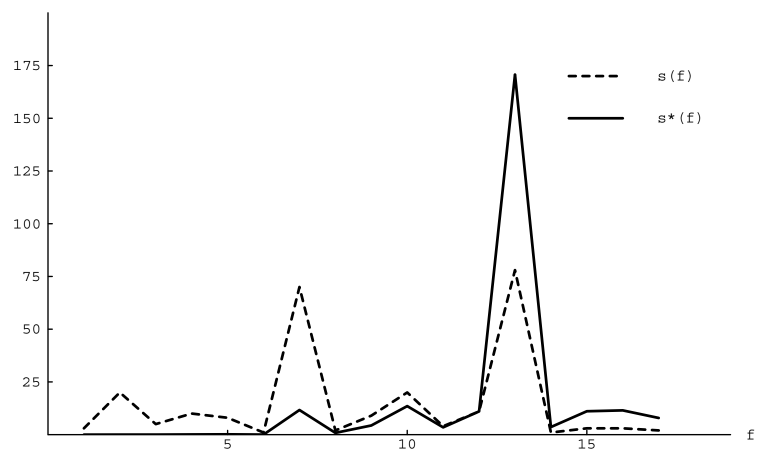

Figure 2: The initial fitness distribution (dashed) and the resulting expected fitness distribution after applying tournament selection with a tournament size of 3.

distribution. The results presented here are published in (Blickle & Thiele, 1995) where also the proofs of the cited theorems can be found.

Algorithm 1: (Tournament Selection)

Input: The population $\vec{J} = (J_1, \ldots, J_N)$,

the fitness values of the population $\vec{\phi} = (\phi_1, \ldots, \phi_N)$.

the tournament size $t \in \{1, 2, ..., N\}$

Output: The population after selection $\vec{J'} = (J'_1, \ldots, J'_N)$

tournament(t,$\vec{J}$, $\vec{\phi}$):

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

$J'_i \leftarrow$ best fit individual out of t uniform random picked individuals from $\vec{J}$

od

return $\vec{J'}$

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

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

3 Tournament Selection

In tournament selection a group of t individuals is randomly chosen from the population. They may be drawn from the population with or without replacement. This group takes part in a "tournament", i.e., a winning individual is determined depending on its fitness value. The best individual having the highest fitness value is usually chosen deterministically although occasionally stochastic selection may be used. In both cases only the winner is inserted into the next population and the process is repeated N times to obtain a new population. Often, tournaments are held between two individuals (binary tournament). However, this can be generalized to an arbitrary group of size t called the *tournament size*.

The following description assumes that the individuals are drawn with replacement and the winning individual is deterministically selected (Algorithm 1). 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 section, the entire fitness distribution after selection can be predicted. The prediction is made for the discrete (exact) fitness distribution as well as for a continuous fitness

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 one means that the variance is not changed by selection. A selection variance less than one corresponds to a decrease in variance. The lowest possible value of Φ_Ω is zero, which means that the variance of the fitness values of the population after selection is itself zero. Again the term the "selection variance" is used as equivalent for "standardized selection variance".

In the following sections several selection schemes are examined in detail. They all have in common, that they are scale and translation invariant such that all properties defined above can be derived. However, fitness proportional selection can not be examined with this methodology as it is not translation invariant (De la Maza & Tidor, 1993). Furthermore, some properties discussed here including the loss of diversity are difficult to investigate for fitness proportional selection. The crucial point is the explicit occurrence of the fitness value in the expected fitness distribution after selection. Hence an analysis is only possible if some further assumptions on the initial fitness distribution are made.

2.8 Selection Variance

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

Definition 2.11 (Selection variance) *The* selection variance Φ *is the normalized expected variance of the fitness distribution of the population after applying the selection method* Ω *to the fitness distribution* $\bar{s}(f)$*, i.e.*

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

For comparison the standardized selection variance is used.

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

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

that is equivalent to

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

leads to the following definition.

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

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

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

Recently more and more researchers are using this term to characterize selection schemes (Thierens & Goldberg 1994a; Thierens & Goldberg 1994b; Bäck, 1995; Miller & Goldberg, 1995).

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

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

$$I = \frac{\bar{M}^* - \bar{M}}{\bar{\sigma}} \tag{12}$$

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 $\mathcal{G}(0, 1)$ as initial fitness distribution

□

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

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

2.7 Selection Intensity

The term "selection intensity" or "selection pressure" is often used in different contexts and for different properties of a selection method. Goldberg and Deb (Goldberg & Deb, 1991) and Bäck (Bäck, 1994b) use the "takeover time" to define the selection pressure. Whitley calls the parameter c (see Section 4) of his ranking selection method selection pressure.

In this paper the term "selection intensity" is used in the same way as in population genetics (Bulmer, 1980). (Mühlenbein & Schlierkamp-Voosen, 1993) have adopted the definition and applied it to Genetic Algorithms.

lost that was contained in the bad individuals. The number of individuals that are replaced corresponds to the strength of the "loss of diversity". This leads to the following definition.

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

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

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

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

Proof: As $\bar{R}(f)$ increases strictly monotonously in f, f_z is uniquely defined. 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

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

the variance of the fitness distribution $\bar{s}^*(f)$ after selection:

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

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

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

2.5 Reproduction Rate

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

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

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

2.6 Loss of Diversity

During every selection phase bad individuals are lost and replaced by copies of better individuals. Thereby a certain amount of "genetic material" is the result of the continuous case can be derived out of the discrete case we restrict ourselves to the continuous case. The main reason for that is that discrete calculations are complicated and messy and the continuous ones are quite handy. Nevertheless the results for the continuous case are relevant for the exact discrete case. As an example a result obtained with continuous arithmetics is experimentally verified with discrete distributions in Section 6.4.

2.3 Average Fitness

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

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

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

2.4 Fitness Variance

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

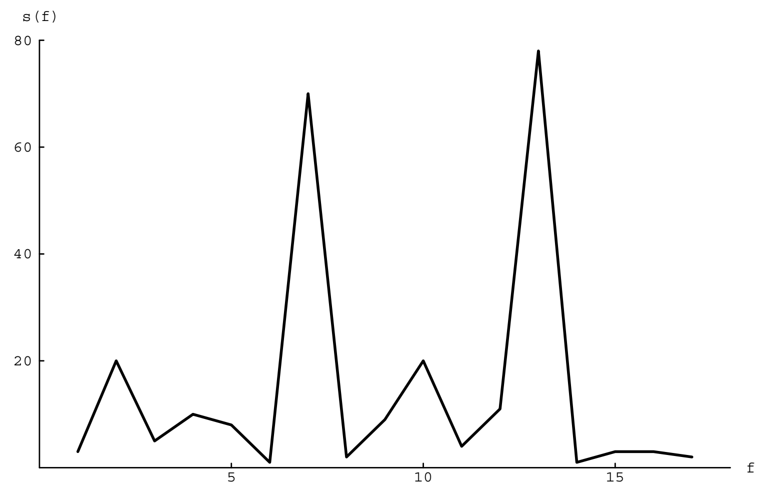

Figure 1: The fitness distribution $s(f)$ for the 6-multiplexer problem ($N = 250$).

denotes the continuous cumulative fitness distribution.

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

$$\mathcal{G}(\mu, \sigma)(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

is chosen.

Now the aspects of the fitness distribution that will be compared are introduced. The definitions given all refer to continuous distributed fitness values. Although “discrete calculations” are more general and accurate and

Definition 2.4 (Cumulative fitness distribution) *Let n be the number of unique fitness values and $f_1 < ... < f_{n-1} < f_n$ $(n \leq N)$ 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.*

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

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

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

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

All functions in the continuous case will be denoted with a bar, e.g. $\bar{s}(f)$ instead of $s(f)$. Sums are replaced by integrals, for example

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

Theorem 2.1 *The variance in obtaining the fitness distribution s' is*

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

Proof: $s^*(f_i)$ denotes the expected number of individuals with fitness value f_i after selection. It is obtained by doing N experiments of the form "select an individual from the population using a certain selection mechanism". Hence the selection probability of an individual with fitness value f_i is given by $p_i = \frac{s^*(f_i)}{N}$. For each fitness value there exists a Bernoulli trial, i.e. an individual with fitness f_i is selected. As the variance of a Bernoulli experiment with N trials is given by $\sigma^2 = Np(1-p)$, (2) is obtained using $p = p_i$. □

The index s in σ_s stands for "sampling" as it is the mean variance due to the sampling of the finite population. The variance of (2) is obtained by performing the selection method in N independent experiments. It is possible to reduce the variance almost completely by using more sophisticated sampling algorithms to select the individuals. Baker's "stochastic universal sampling" algorithm (SUS) (Baker, 1987), which is an optimal sampling algorithm, is introduced in Section 6 when the different selection schemes are compared.

Definition 2.1 (Fitness distribution) *The function* $s : \mathbb{R} \mapsto Z_0^+$ *assigns to each fitness value* $f \in \mathbb{R}$ *the number of individuals in a population* $\vec{J} \in \mathbf{J}^N$ *carrying this fitness value.* s *is called the* fitness distribution *of a population* $\vec{J}$.

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

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

$$s' = \Omega(s, par\_list) \tag{1}$$

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

As the selection methods are probabilistic the next definition is introduced.

Definition 2.3 $\Omega^*(s, par\_list)$ *denotes the* expected fitness distribution *after applying the selection method* Ω *to the fitness distribution* s*. The notation* $s^* = \Omega^*(s, par\_list)$ *will be used as abbreviation.*

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

- new properties of exponential ranking selection (selection intensity, selection variance and loss of diversity).

- The convergence of an Evolutionary Algorithm optimizing the ONE-MAX function can be predicted for several selection schemes.

- The unified description enables a new kind of comparison of selection schemes. In Section 6 for the first time several properties can be compared simultaneously.

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

2.2 Fitness Distribution

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

There exists only a finite number of different fitness values $f_1, \ldots, f_n$, $(n \leq N)$ 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.

of a Genetic Algorithm that also makes use of fitness distributions.

The consequent use of fitness distributions has many advantages:

- Fitness distributions are a powerful framework giving an unified view of the selection schemes and allowing several up to now independently and isolated obtained aspects of these selection schemes (such as selection intensity or takeover time) to be derived with one single methodology.

- Several new properties of selection schemes can be derived for the first time, e.g.

  - the expected fitness distribution of tournament selection (Theorem 3.1), truncation selection (Theorem 3.5), linear ranking selection (Theorem 4.1), and exponential ranking selection (Theorem 5.1).

  - a new important characteristic of selection schemes, called *loss of diversity* (Definition 2.8)

  - an important new property of tournament selection (the concatenation of tournament selection - Theorem 3.3).

  - new approximations for properties of tournament selection (selection intensity and the selection variance) vidual. Bäck has analyzed the most prominent selection schemes used in Evolutionary Algorithms with respect to their takeover time (Bäck, 1994b; Bäck, 1994a).

In (Mühlenbein & 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 describes the change in the average fitness of the population due to selection. They derive the selection intensity for proportional selection and truncation selection. Here this work will be extended to tournament selection, linear and exponential ranking selection.

An analysis based on the behavior of the best individual or on the average population fitness only describes one aspect of a selection method. Here 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 characterization of the population by its fitness distribution has also been used by other researchers, but in a more informal way. In (Mühlenbein & Schlierkamp-Voosen, 1993) the fitness distribution is used to calculate some properties of truncation selection. In (Shapiro, Prügel-Bennett & Rattray 1994) a statistical mechanics approach is taken to describe the dynamics linear and exponential ranking selection will be considered. A large part of the analysis is also applicable to steady-state selection schemes as the selection probabilities are the same as in the generational case.

The methodology will be described in the next section. In Section 3 the most important results from (Blickle & Thiele, 1995) concerning tournament selection are collected. In the subsequent sections an analysis of truncation selection, linear and exponential ranking is given. Finally the selection schemes are compared in Section 6.

2 Description of Selection Schemes

2.1 Related Work

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

Goldberg (Goldberg & Deb, 1991) introduced the term *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 crossover or mutation is used. Hence, this analysis is restricted to the behavior of the best indi-

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

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

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

In this paper a description of selection schemes is used based on the fitness distribution of the population before and after selection as introduced in (Blickle & Thiele, 1995). This framework is used to analyze common selection schemes. In particular, tournament selection, truncation selection,

1 Introduction

Evolutionary Algorithms (EAs) 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* $\vec{J} = \{J_1, J_2, ..., J_N\} \in \mathbf{J}^N$ is modified according to the natural evolutionary process: after initialization, selection $\omega : \mathbf{J}^N \mapsto \mathbf{J}^N$ and variation $\Xi : \mathbf{J}^N \mapsto \mathbf{J}^N$ are executed in a loop until some termination criterion is reached. Each run of the loop is called a generation and $P(\tau)$ denotes the population at generation τ.

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

With this a mathematical analysis of tournament selection, truncation selection, ranking selection and exponential ranking selection is carried out that allows an exact prediction of the fitness values after selection. The correspondence of binary tournament selection and ranking selection in the expected fitness distribution is proven. Furthermore several properties of selection schemes are derived (selection intensity, selection variance, loss of diversity) and the three selection schemes are compared using these properties.

A Comparison of Selection Schemes used in Evolutionary Algorithms

Tobias Blickle[*] and Lothar Thiele[†]

Computer Engineering and Communication Networks Lab (TIK)

Swiss Federal Institute of Technology (ETH)

Gloriastrasse 35, 8092 Zurich

Switzerland

December 27, 1996

Abstract

Evolutionary Algorithms are a common probabilistic optimization method based on the model of natural evolution. One important operator in these algorithms is the selection scheme for which in this paper a new description model based on fitness distributions is introduced.

[*] blickle@tik.ee.ethz.ch

[†] thiele@tik.ee.ethz.ch