Calculating the Expected Loss of Diversity of Selection Schemes

Tatsuya Motoki motoki@ie.niigata-u.ac.jp
Department of Information Engineering, Niigata University, Ikarashi 2-8050, Niigata 950-2181, Japan

Abstract

This paper concerns a measure of selective pressure, called "loss of diversity," that denotes the proportion of unselected individuals during the selection phase. We probabilistically calculate the expected value and variance of loss of diversity in tournament selection, truncation selection, linear ranking selection, and exponential ranking selection. From numerical results, we observe that in tournament selection, many more individuals are expected to be lost than with Blickle and Thiele's static estimate. We also observe that tournament and exponential ranking schemes potentially bring about nearly equivalent selection behaviors but have different types of control parameters.

Keywords

Genetic algorithms, selection method, selective pressure, loss of diversity, tournament selection, truncation selection, ranking selection.

1 Introduction

Evolutionary Algorithms (EAs) are probabilistic search algorithms that maintain a population of search points and repeatedly evolve the population by selection, recombination, and mutation operations. Due to their ability to search complex spaces, EAs have been used in a variety of problems. Starting from the current candidate solutions in the population, EAs generate new candidate solutions through recombination and mutation, and then exploit this exploration through selection of better solutions.

Selection operation plays an important role of focusing the search effort on promising regions in the search space and controlling the speed of convergence; therefore, some researchers have analyzed common selection schemes such as tournament selection, fitness-proportionate selection, ranking selection, and truncation selection. For example, Goldberg and Deb (1991) introduced a measure of selective pressure called *takeover time* and compared selection schemes with respect to that measure; the takeover time is defined to be the number of generations needed for the single best individual to fill up the whole population when only the selection operation is applied repeatedly. Blickle and Thiele (1995, 1997) considered fitness distributions to give a unified view of selection schemes and introduced a natural measure of selective pressure called *loss of diversity*; then, for some selection schemes, they calculated expected fitness distribution after selection, loss of diversity, selection intensity, and selection variance. Schell and Wegenkittl (2001) viewed the selection process as a two-step procedure consisting of assignment of selection probabilities and sampling according to this probability distribution; then, they analyzed stochastic properties of sampling methods such as roulette-wheel sampling and stochastic universal sampling.

Table 1: Comparison of Blickle and Thiele's estimates of loss of diversity with experimental values.

Tournament Size	Estimate by Simulation (Average)								Blickle & Thiele's Estimate
	Population Size								
	2	3	4	5	7	10	20	50	
1	0.250	0.300	0.321	0.328	0.336	0.347	0.363	0.355	nan
2	0.314	0.370	0.379	0.390	0.411	0.407	0.431	0.426	0.250
3	0.390	0.438	0.460	0.472	0.476	0.493	0.495	0.503	0.385
4	0.445	0.503	0.517	0.533	0.558	0.550	0.551	0.557	0.472
5	0.470	0.549	0.571	0.586	0.595	0.597	0.611	0.605	0.535
6	0.482	0.587	0.611	0.615	0.623	0.625	0.641	0.638	0.582
7	0.493	0.611	0.636	0.644	0.653	0.661	0.656	0.666	0.620
8	0.496	0.626	0.661	0.675	0.680	0.688	0.692	0.693	0.650
9	0.498	0.644	0.684	0.691	0.698	0.706	0.713	0.716	0.675
10	0.499	0.649	0.699	0.707	0.719	0.727	0.728	0.733	0.697
15	0.500	0.664	0.736	0.773	0.781	0.787	0.789	0.796	0.769
20	0.500	0.666	0.747	0.789	0.819	0.820	0.829	0.828	0.811
25	0.500	0.667	0.749	0.796	0.843	0.845	0.854	0.854	0.840
30	0.500	0.667	0.750	0.797	0.849	0.865	0.867	0.873	0.860

Among these theoretical subjects of selection, this paper concerns loss of diversity. As explained by Whitley (1989) and Banzhaf et al. (1998, Section 8.4.3), in order to maintain exploration ability in evolutionary search, we need to maintain genetic diversity in the population. To avoid great computational effort in measuring genetic diversity, some researchers pay attention to diversity of fitness in the population instead of genetic diversity, under the assumption that fitness differences reflect differences in genotypes of individuals. Even if we have a good measure of genetic diversity in the population, the amount of loss of genetic diversity during the selection phase would generally depend not only upon selection schemes, but also upon arrangements of genes before selection. For comparing selection schemes, we want a measure that reflects the amount of loss of genetic diversity during the selection phase and does not depend on arrangements of genes before selection. Such a measure is given by Blickle and Thiele (1995, 1997). They defined loss of diversity to be the proportion of individuals of a population that are not selected during the selection phase. For some selection schemes, they also give static estimates of loss of fitness diversity based on the difference of the expected fitness distribution after selection and the original fitness distribution before selection; of course, these estimates do not coincide with experimental results on the loss of diversity defined above. Their estimation is only an approximation to the loss of diversity because they calculate loss of fitness diversity, and because they implicitly assume that selection is done deterministically based on the expected distribution of individuals after selection. So, we now probabilistically calculate expected value and variance of the loss of diversity.

In the following sections, we use the phrase "loss of diversity" to mean the above measure of selective pressure defined by Blickle and Thiele. Section 2 describes Blickle and Thiele's estimates of the loss of fitness diversity and compare their estimates with experimental results on the loss of diversity. In Sections 3 to 6, we calculate the expected loss of diversity in tournament selection , truncation selection, linear ranking selection, and exponential ranking selection, respectively. In Section 7, we calculate variance of the loss of diversity. We summarize our work in Section 8.

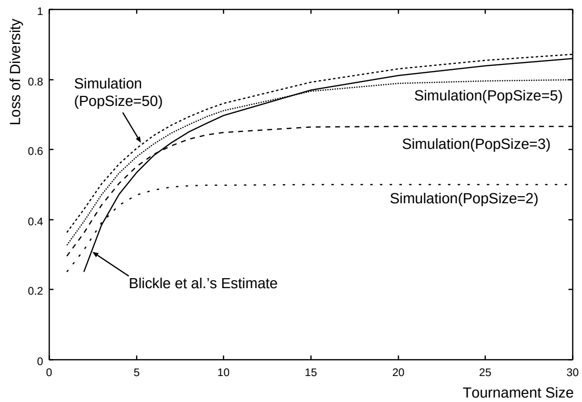

Figure 1: Blickle and Thiele's estimates vs. experimental results under fixed population size.

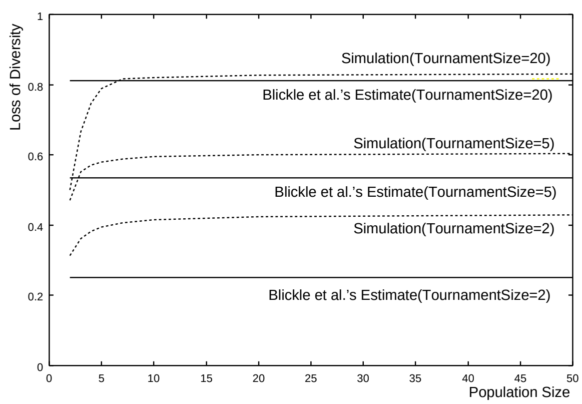

Figure 2: Blickle and Thiele's estimates vs. experimental results under fixed tournament size.

2 Loss of Diversity

Blickle and Thiele (1995, 1997) gave the definition of loss of diversity and its estimates for some selection schemes as follows.

DEFINITION 2.1 (Blickle and Thiele): *The loss of diversity is the proportion of individuals of a population that are not selected during the selection phase.*

THEOREM 2.2 (Blickle and Thiele, 1997):
(a) *For tournament selection with tournament size t, loss of (fitness) diversity is*

$$D_T(t) \approx t^{-1/(t-1)} - t^{-t/(t-1)}.$$

(b) *For truncation selection with threshold T, loss of (fitness) diversity is*

$$D_\Gamma(T) \approx 1 - T.$$

(c) *For linear ranking selection with lowest reproduction rate η^-, loss of (fitness) diversity is*

$$D_R(\eta^-) \approx \frac{1}{4}(1 - \eta^-).$$

(d) *For exponential ranking selection with ranking base c, loss of (fitness) diversity is*

$$D_E(c) \approx \frac{1 - \ln(\kappa - 1) + \ln(\kappa \ln \kappa)}{\ln \kappa} - \frac{\kappa}{\kappa - 1},$$

where $\kappa = c^N$ and N denotes the population size.

To show how this theoretical result on tournament selection agrees with experimental results on the loss of diversity, we repeatedly perform the tournament selection operation and measure the average loss of diversity for various combinations of tournament size t and population size N. Numerical results of this simulation and estimates by Theorem 2.2(a) are given in Table 1 and Figures 1 and 2, which show the disagreement of Theorem 2.2(a) with experimental results on the loss of diversity. Looking to Blickle and Thiele's estimation, we observe that they start the calculation as follows:

$$D_T(t) = \frac{1}{N} \int_{f_0}^{f_z} (\bar{s}(x) - \bar{s}^*(x)) dx,$$

where $\bar{s}(x)$ denotes the density of individuals with fitness value x before selection, $\bar{s}^*(x)$ denotes the expected density of individuals with fitness value x after selection, f_0 denotes the least (worst) fitness value, and f_z denotes the fitness value such that $\bar{s}(f_z) = \bar{s}^*(f_z)$. This estimation is only an approximation, because it corresponds to loss of fitness diversity, because it ignores the possibility that individuals with fitness value above f_z may be lost during a selection phase, and so on.

To obtain an estimate of the loss of diversity that agrees with practice, we should not utilize the expected fitness distribution after selection but should probabilistically calculate the expected number of individuals that are not selected during the selection phase. In the following sections, we will perform such probabilistic calculation where we assume individuals are selected one by one, that is, roulette-wheel sampling is adopted. When an individual is the kth worst one in the population, we say the individual has *rank k*. For simplicity, we assume that every individual has a unique rank number between 1 (worst one) and the population size (best one).

3 Expected Loss of Diversity of Tournament Selection

In usual tournament selection, a prespecified number (called *tournament size*) of individuals are randomly drawn from the population with replacement; the best of them is selected into the mating pool. This process is repeated a number of times equal to the population size.

LEMMA 3.1: *For any nonnegative integer a, we have*

$$\sum_{k=1}^{N} k^a = \begin{cases} \dfrac{N^{a+1}}{a+1} & \textit{if } a=0 \\ \dfrac{N^{a+1}}{a+1} + \dfrac{N^a}{2} & \textit{if } a=1 \\ \dfrac{N^{a+1}}{a+1} + \dfrac{N^a}{2} + \dfrac{aN^{a-1}}{12} + \mathcal{O}(N^{a-2}) & \textit{if } a \geq 2 \end{cases}$$

PROOF: The proof proceeds by induction on a.
Basis, $a \leq 1$. It obviously holds.
Induction step, $a \geq 2$. First, we obtain

$$(N+1)^{a+1} - 1 = \sum_{k=1}^{N} \{(k+1)^{a+1} - k^{a+1}\}$$

$$
\begin{aligned}
&= \sum_{k=1}^{N} \left\{ \sum_{i=0}^{a+1} \binom{a+1}{i} k^{a+1-i} - k^{a+1} \right\} \\
&= \sum_{k=1}^{N} \left\{ \binom{a+1}{1} k^{a} + \sum_{i=2}^{a+1} \binom{a+1}{i} k^{a+1-i} \right\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\sum_{k=1}^{N} k^a &= \frac{1}{\binom{a+1}{1}} \left\{ (N+1)^{a+1} - 1 - \sum_{k=1}^{N}\sum_{i=2}^{a+1} \binom{a+1}{i} k^{a+1-i} \right\} \\
&= \frac{1}{a+1} \left\{ \sum_{i=0}^{a+1} \binom{a+1}{i} N^i - 1 - \sum_{i=2}^{a+1} \binom{a+1}{i} \sum_{k=1}^{N} k^{a+1-i} \right\} \\
&= \frac{1}{a+1} \left\{ N^{a+1} + \binom{a+1}{a} N^a + \binom{a+1}{a-1} N^{a-1} + \mathcal{O}(N^{a-2}) \right. \\
&\qquad - \binom{a+1}{2} \sum_{k=1}^{N} k^{a-1} - \binom{a+1}{3} \sum_{k=1}^{N} k^{a-2} \\
&\qquad \left. - \sum_{i=4}^{a+1} \binom{a+1}{i} \sum_{k=1}^{N} k^{a+1-i} \right\}.
\end{aligned}
$$

Using the induction hypothesis, we proceed with the calculation as follows:

$$
\begin{aligned}
\sum_{k=1}^{N} k^a &= \frac{1}{a+1} \left\{ N^{a+1} + \binom{a+1}{a} N^a + \binom{a+1}{a-1} N^{a-1} + \mathcal{O}(N^{a-2}) \right. \\
&\qquad - \binom{a+1}{2} \left(\frac{N^a}{a} + \frac{N^{a-1}}{2} + \mathcal{O}(N^{a-2}) \right) \\
&\qquad - \binom{a+1}{3} \left(\frac{N^{a-1}}{a-1} + \mathcal{O}(N^{a-2}) \right) \\
&\qquad \left. - \sum_{i=4}^{a+1} \binom{a+1}{i} \mathcal{O}(N^{a+2-i}) \right\} \\
&= \frac{N^{a+1}}{a+1} + \frac{1}{a+1} \left\{ \binom{a+1}{a} - \binom{a+1}{2} \frac{1}{a} \right\} N^a \\
&\qquad + \frac{1}{a+1} \left\{ \binom{a+1}{a-1} - \binom{a+1}{2} \frac{1}{2} - \binom{a+1}{3} \frac{1}{a-1} \right\} N^{a-1} + \mathcal{O}(N^{a-2}) \\
&= \frac{N^{a+1}}{a+1} + \frac{N^a}{2} + \frac{aN^{a-1}}{12} + \mathcal{O}(N^{a-2}). \qquad \square
\end{aligned}
$$

THEOREM 3.2: *Let $D_T(t,N)$ denote the expected loss of diversity in tournament selection with tournament size t and population size N. Then*

(a) $D_T(t,N) = \dfrac{1}{N} \displaystyle\sum_{k=1}^{N} \left(1 - \frac{k^t - (k-1)^t}{N^t} \right)^N.$

(b) $\displaystyle\lim_{N\to\infty} D_T(t,N) = \sum_{a=2}^{\infty} \frac{(-t)^a}{a!\,(at-a+1)}$. *Specifically,* $\displaystyle\lim_{N\to\infty} D_T(1,N) = e^{-1} = 0.3678\cdots$

and $\lim_{N\to\infty} D_T(2,N) = \frac{1}{2}(1-e^{-2}) = 0.4323\cdots$.

PROOF: (a) We now call a (multi)set of t randomly chosen individuals a *tournament group* and use $\Pr(E)$ to denote the probability of an event E. For any $k \in \{1,2,3,\ldots,N\}$, we can calculate the probability p_k of the event that occurs when the individual of rank k is selected in a single selection process:

$$\begin{aligned} p_k &= \Pr(k \text{ is the highest rank among a (randomly generated) tournament group}) \\ &= \Pr(\text{a tournament group consists of individuals of rank} \le k) \\ &\quad - \Pr(\text{a tournament group consists of individuals of rank} \le k-1) \\ &= \frac{k^t-(k-1)^t}{N^t}. \end{aligned} \tag{1}$$

Thus, the probability of the event that occurs when the individual of rank k is never selected in N selection processes is

$$\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N;$$

the expected number of unselected individuals in the selection phase is

$$\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N,$$

and so

$$D_T(t,N) = \frac{1}{N}\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N.$$

(b) We can proceed with calculation of (a) as follows:

$$\begin{aligned} D_T(t,N) &= \frac{1}{N}\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N \\ &= \frac{1}{N}\sum_{k=1}^{N}\left(1-\binom{t}{1}\frac{k^{t-1}}{N^t}+\binom{t}{2}\frac{k^{t-2}}{N^t}+\frac{\mathcal{O}(k^{t-3})}{N^t}\right)^N \\ &= \frac{1}{N}\sum_{k=1}^{N}\sum_{a=0}^{N}\sum_{b=0}^{N-a}\sum_{c=0}^{N-a-b}\binom{N}{a,b,c,N-a-b-c}\left(-t\frac{k^{t-1}}{N^t}\right)^a\left(\binom{t}{2}\frac{k^{t-2}}{N^t}\right)^b\left(\frac{\mathcal{O}(k^{t-3})}{N^t}\right)^c \\ &= \sum_{a=0}^{N}\sum_{b=0}^{N-a}\left\{\binom{N}{a,b,N-a-b}\frac{(-t)^a\binom{t}{2}^b}{N^{at+bt+1}}\sum_{k=1}^{N}k^{at+bt-a-2b}\right. \\ &\quad \left.+\sum_{c=1}^{N-a-b}\binom{N}{a,b,c,N-a-b-c}\frac{(-t)^a\binom{t}{2}^b}{N^{at+bt+ct+1}}\sum_{k=1}^{N}\mathcal{O}(k^{at+bt+ct-a-2b-3c})\right\}. \end{aligned}$$

By applying Lemma 3.1, we continue the calculation:

$$D_T(t,N) = \sum_{a=0}^{N}\sum_{b=0}^{N-a}\left\{\binom{N}{a,b,N-a-b}\frac{(-t)^a\binom{t}{2}^b}{N^{at+bt+1}}\left(\frac{N^{at+bt-a-2b+1}}{at+bt-a-2b+1}\right.\right.$$

$$
\begin{aligned}
&\quad + \frac{N^{at+bt-a-2b}}{2} + \frac{at+bt-a-2b}{12} N^{at+bt-a-2b-1} + \mathcal{O}(N^{at+bt-a-2b-2}) \Bigg) \\
&\quad + \sum_{c=1}^{N-a-b} \binom{N}{a,b,c,N-a-b-c} \frac{(-t)^a \binom{t}{2}^b}{N^{at+bt+ct+1}} \mathcal{O}(N^{at+bt+ct-a-2b-3c+1}) \Bigg\} \\
= &\sum_{a=0}^{N} \sum_{b=0}^{N-a} \Bigg\{ \frac{N(N-1)(N-2)\cdots(N-a-b+1)}{a!\,b!} (-t)^a \binom{t}{2}^b \Bigg(\frac{N^{-a-2b}}{at+bt-a-2b+1} \\
&\quad + \frac{N^{-a-2b-1}}{2} + \frac{at+bt-a-2b}{12} N^{-a-2b-2} + \mathcal{O}(N^{-a-2b-3}) \Bigg) \\
&\quad + \sum_{c=1}^{N-a-b} \binom{N}{a,b,c,N-a-b-c} (-t)^a \binom{t}{2}^b \mathcal{O}(N^{-a-2b-3c}) \Bigg\} \\
= &\sum_{a=0}^{N} \frac{(-t)^a}{a!\,(at-a+1)} + \frac{1}{N} \Bigg\{ \sum_{a=0}^{N} \frac{(-t)^a}{a!\,(at-a+1)} (-1-2-\cdots-(a-1)) \\
&\quad + \sum_{a=0}^{N} \frac{(-t)^a}{a!} \frac{1}{2} + \sum_{a=0}^{N} \frac{(-t)^a \binom{t}{2}}{a!(at+t-a-1)} \Bigg\} + \mathcal{O}(N^{-2}).
\end{aligned}
$$

So, we obtain the equation

$$\lim_{N\to\infty} D_T(t,N) = \sum_{a=0}^{\infty} \frac{(-t)^a}{a!\,(at-a+1)}.$$

Specifically, when $t=1$,

$$\lim_{N\to\infty} D_T(1,N) = \sum_{a=0}^{\infty} \frac{(-1)^a}{a!} = e^{-1};$$

when $t=2$,

$$\lim_{N\to\infty} D_T(2,N) = \sum_{a=0}^{\infty} \frac{(-2)^a}{a!(a+1)} = \frac{1}{(-2)} \left\{ \sum_{i=0}^{\infty} \frac{(-2)^i}{i!} - 1 \right\} = \frac{1}{2}(1-e^{-2}). \qquad \square$$

Utilizing Theorem 3.2(a), we can numerically calculate $D_T(t,N)$ for every pair of tournament size t and population size N; utilizing Theorem 3.2(b), we can also numerically calculate $\lim_{N\to\infty} D_T(t,N)$ for every t. Numerical results of the function $D_T(t,N)$ are summarized in Table 2 and Figures 3 and 4, which agree with results by simulation in Table 1. Given a population size N, we can adjust $D_T(t,N)$ within a wide interval $(3.7, 1-\frac{1}{N})$ through the tournament size parameter t. The expected loss of diversity $D_T(t,N)$ monotonically increases in t and N. However, in the subdomain $N \geq 10$, the function $D_T(t,N)$ only increases slowly in N; in fact, for any $N \geq 10$, $|\lim_{x\to\infty} D_T(t,x) - D_T(t,N)| \leq 0.02$. When $t=1$, tournament selection acts as random sampling; even in such a case, about 35% of the population is expected to be lost for any population size $N \geq 10$. When $t=3$, about 50% of the population is expected to be lost for any $N \geq 10$. Many more individuals will be lost than with Blickle and Thiele's static estimate (Theorem 2.2(a)).

4 Expected Loss of Diversity of Truncation Selection

Mühlenbein and Schlierkamp-Voosen's (1993) *truncation selection* has a control parameter $T \in (0,1]$ called *threshold*. Let N denote the population size; then, in this selection

Table 2: Expected loss of diversity $D_T(t, N)$ of tournament selection.

Tournament Size t	Population Size N								
	2	3	4	5	7	10	20	50	inf
1	0.250	0.296	0.316	0.328	0.340	0.349	0.358	0.364	0.368
2	0.312	0.362	0.383	0.394	0.406	0.415	0.424	0.429	0.432
3	0.391	0.442	0.461	0.471	0.482	0.489	0.497	0.502	0.504
4	0.441	0.504	0.522	0.532	0.542	0.548	0.555	0.559	0.561
5	0.470	0.551	0.570	0.579	0.589	0.595	0.601	0.604	0.606
6	0.485	0.586	0.608	0.617	0.625	0.631	0.637	0.640	0.642
7	0.492	0.612	0.638	0.647	0.655	0.661	0.667	0.669	0.671
8	0.496	0.629	0.663	0.672	0.680	0.686	0.691	0.694	0.695
9	0.498	0.641	0.683	0.693	0.701	0.706	0.712	0.714	0.716
10	0.499	0.650	0.698	0.711	0.718	0.724	0.730	0.732	0.733
15	0.500	0.664	0.737	0.767	0.780	0.784	0.791	0.793	0.794
20	0.500	0.666	0.747	0.789	0.817	0.820	0.827	0.830	0.830
25	0.500	0.667	0.749	0.796	0.837	0.846	0.851	0.854	0.855
30	0.500	0.667	0.750	0.799	0.848	0.864	0.869	0.872	0.873

scheme, an individual is randomly selected from the best $\lceil TN \rceil$ individuals; this type of selection is repeated N times.

THEOREM 4.1: *Let $D_\Gamma(T, N)$ denote the expected loss of diversity in truncation selection with threshold $T \in (0, 1]$ and population size N. Then*

(a) $D_\Gamma(T, N) = \dfrac{1}{N}\left\{\lfloor (1-T)N \rfloor + \lceil TN \rceil \left(1 - \dfrac{1}{\lceil TN \rceil}\right)^N\right\}$.

(b) $\lim\limits_{N\to\infty} D_\Gamma(T, N) = (1-T) + Te^{-1/T}$.

PROOF: (a) In truncation selection with threshold T and population size N, individuals with rank $\leq (1-T)N$ are necessarily discarded and remaining individuals are assigned the equal selection probability. So we can calculate the probability p_k of the event that occurs when the individual of rank k is selected in a single selection process:

$$p_k = \begin{cases} 0 & \text{if } 1 \leq k \leq \lfloor (1-T)N \rfloor \\ \dfrac{1}{N - \lfloor (1-T)N \rfloor} & \text{if } \lfloor (1-T)N \rfloor + 1 \leq k \leq N. \end{cases} \tag{2}$$

Thus, the probability of the event that occurs when the individual of rank k is never selected in N selection processes is

$$\begin{cases} 1 & \text{if } 1 \leq k \leq \lfloor (1-T)N \rfloor \\ \left(1 - \dfrac{1}{N - \lfloor (1-T)N \rfloor}\right)^N & \text{if } \lfloor (1-T)N \rfloor + 1 \leq k \leq N; \end{cases}$$

the expected number of unselected individuals in the selection phase is

$$\begin{aligned} &\lfloor (1-T)N \rfloor + (N - \lfloor (1-T)N \rfloor)\left(1 - \frac{1}{N - \lfloor (1-T)N \rfloor}\right)^N \\ &= \lfloor (1-T)N \rfloor + \lceil TN \rceil \left(1 - \frac{1}{\lceil TN \rceil}\right)^N, \end{aligned}$$

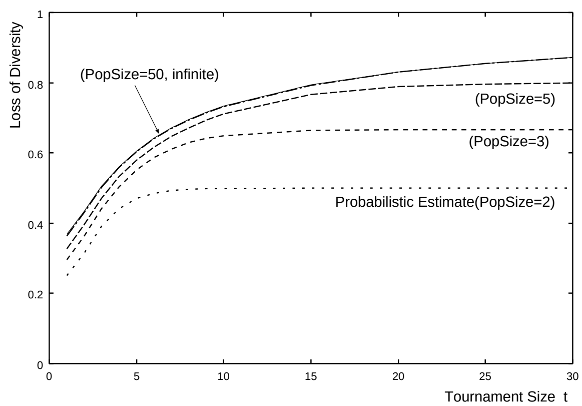

Figure 3: Expected loss of diversity $D_T(t, N)$ of tournament selection under fixed population size N.

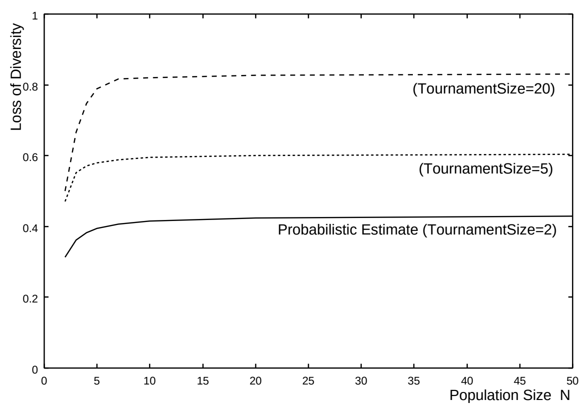

Figure 4: Expected loss of diversity $D_T(t, N)$ of tournament selection under fixed tournament size t.

and so

$$D_\Gamma(T, N) = \frac{1}{N}\left\{\lfloor (1-T)N \rfloor + \lceil TN \rceil \left(1 - \frac{1}{\lceil TN \rceil}\right)^N\right\}.$$

(b) Utilizing (a), we carry out the following calculation.

$$\begin{aligned}
\lim_{N\to\infty} D_\Gamma(T, N) &= \lim_{N\to\infty}\left\{(1-T) + T\left(1 - \frac{1}{TN}\right)^N\right\} \\
&= (1-T) + T\left\{\lim_{N\to\infty}\left(1 - \frac{1}{TN}\right)^{TN}\right\}^{1/T} \\
&= (1-T) + T\left(e^{-1}\right)^{1/T} \\
&= (1-T) + Te^{-1/T}.
\end{aligned}$$

□

Numerical results of the function $D_\Gamma(T,N)$ are summarized in Table 3 and Figures 5 and 6. Given a population size N, we can adjust $D_\Gamma(T,N)$ within a wide interval $(3.7, 1-\frac{1}{N})$ through the threshold parameter T. The expected loss of diversity $D_\Gamma(T,N)$ monotonically decreases in T (Figure 5). Due to the ceiling and the floor functions, $D_\Gamma(T,N)$ is not monotonic in N (Figure 6), but for the subdomain $N \geq 20$, it only increases slowly in N; in fact, for any $N \geq 20$, $|\lim_{x\to\infty} D_\Gamma(T,x) - D_\Gamma(T,N)| \leq 0.05$. When $T=1$, truncation selection acts as random sampling; in such a case, Theorem 4.1(b) indicates that for any sufficiently large N, the expected loss of diversity is nearly equal to e^{-1}, which agrees with $D_T(1,N)$ in Theorem 3.2(b).

5 Expected Loss of Diversity of Linear Ranking Selection

In Baker's (1985) *linear ranking selection*, each individual is assigned the selection probability

$$p_k \quad = \quad \frac{1}{N}\left(\eta^- + (2-2\eta^-)\frac{k-1}{N-1}\right), \tag{3}$$

where k is the rank of the individual, η^- is a prespecified value satisfying $0 \leq \eta^- \leq 1$ (called the *lowest reproduction rate*), and N denotes the population size; then, roulette-wheel sampling with selection probabilities $\{p_k\}$ is repeated N times.

THEOREM 5.1: *Let $D_R(\eta^-,N)$ denote the expected loss of diversity in linear ranking selection with the lowest reproduction rate $\eta^- \in [0,1]$ and population size N. Then*

(a) $$D_R(\eta^-,N) \quad = \quad \frac{1}{N}\sum_{k=1}^{N}\left(1-\frac{1}{N}\left(\eta^- + (2-2\eta^-)\frac{k-1}{N-1}\right)\right)^N$$

$$\approx \quad \begin{cases} \left(1-\dfrac{1}{N}\right)^N & \textit{if } \eta^-=1 \\ \dfrac{N-1}{(2-2\eta^-)(N+1)}\left\{\left(1-\dfrac{\eta^-}{N}+\dfrac{1-\eta^-}{N(N-1)}\right)^{N+1}\right. & \\ \qquad\qquad \left. -\left(1-\dfrac{2-\eta^-}{N}-\dfrac{1-\eta^-}{N(N-1)}\right)^{N+1}\right\} & \textit{if } 0\leq\eta^-<1. \end{cases}$$

(b) $$\lim_{N\to\infty} D_R(\eta^-,N) = \begin{cases} e^{-1} & \textit{if } \eta^-=1 \\ \frac{1}{(2-2\eta^-)}\left(e^{-\eta^-} - e^{-(2-\eta^-)}\right) & \textit{if } 0\leq\eta^-<1. \end{cases}$$

PROOF: (a) In linear ranking selection with the lowest reproduction rate η^- and population size N, the probability p_k of the event that occurs when the individual of rank k is selected in a single selection process is given by Equation (3). Therefore, the probability of the event that occurs when the individual of rank k is never selected in N selection processes is

$$(1-p_k)^N;$$

the expected number of unselected individuals in the selection phase is

$$\sum_{k=1}^{N}(1-p_k)^N,$$

Table 3: Expected loss of diversity $D_\Gamma(T, N)$ of truncation selection.

Thresh-old T	Population Size N								
	2	3	4	5	7	10	20	50	inf
0.001	0.500	0.667	0.750	0.800	0.857	0.900	0.950	0.980	0.999
0.010	0.500	0.667	0.750	0.800	0.857	0.900	0.950	0.980	0.990
0.100	0.500	0.667	0.750	0.800	0.857	0.900	0.900	0.900	0.900
0.200	0.500	0.667	0.750	0.800	0.717	0.800	0.801	0.801	0.801
0.300	0.500	0.667	0.531	0.613	0.597	0.705	0.708	0.710	0.711
0.400	0.500	0.417	0.531	0.613	0.597	0.623	0.628	0.631	0.633
0.500	0.500	0.417	0.531	0.479	0.505	0.554	0.561	0.565	0.568
0.600	0.250	0.417	0.398	0.479	0.436	0.497	0.505	0.510	0.513
0.700	0.250	0.296	0.398	0.390	0.436	0.450	0.459	0.464	0.468
0.800	0.250	0.296	0.316	0.390	0.382	0.410	0.420	0.426	0.429
0.900	0.250	0.296	0.316	0.328	0.340	0.377	0.387	0.393	0.396
1.000	0.250	0.296	0.316	0.328	0.340	0.349	0.358	0.364	0.368

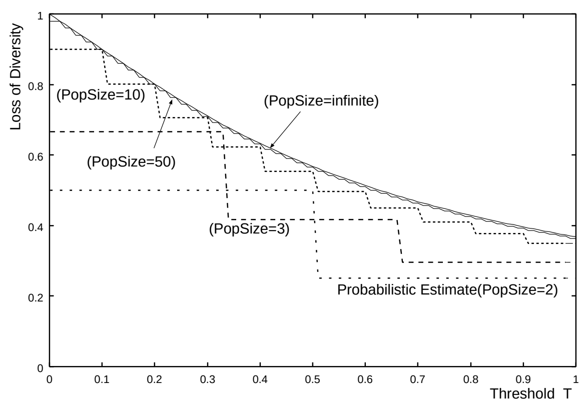

Figure 5: Expected loss of diversity $D_\Gamma(T, N)$ of truncation selection under fixed population size N.

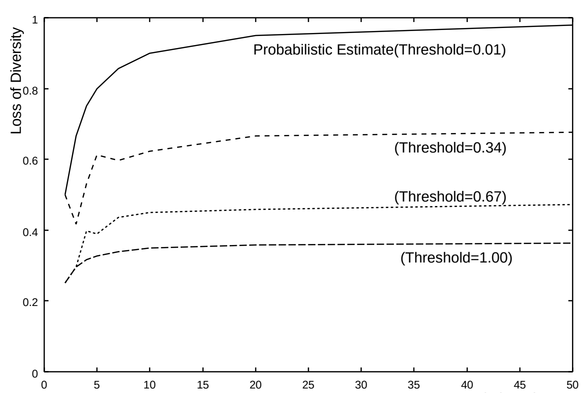

Figure 6: Expected loss of diversity $D_\Gamma(T, N)$ of truncation selection under fixed threshold T.

and so

$$D_R(\eta^-, N) = \frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{1}{N}\left(\eta^- + (2-2\eta^-)\frac{k-1}{N-1}\right)\right)^N.$$

Replacing the summation by integral, we yield an approximation.

$$\begin{aligned} D_R(\eta^-, N) &\approx \frac{1}{N}\int_{1/2}^{N+1/2}\left(1 - \frac{1}{N}\left(\eta^- + (2-2\eta^-)\frac{x-1}{N-1}\right)\right)^N dx \\ &= \begin{cases} \left(1-\frac{1}{N}\right)^N & \text{if } \eta^- = 1 \\ \frac{N-1}{(2-2\eta^-)(N+1)}\left\{\left(1-\frac{\eta^-}{N}+\frac{1-\eta^-}{N(N-1)}\right)^{N+1} \right. \\ \qquad \left. -\left(1-\frac{2-\eta^-}{N}-\frac{1-\eta^-}{N(N-1)}\right)^{N+1}\right\} & \text{if } 0 \le \eta^- < 1 \end{cases} \end{aligned}$$

(b) When $N \to \infty$, the error in the above approximation approaches to zero. So, in the case where $\eta^- = 1$, we can proceed with the calculation as follows:

$$\lim_{N\to\infty} D_R(\eta^-, N) = \lim_{N\to\infty}\left(1-\frac{1}{N}\right)^N = e^{-1}.$$

In the case where $0 \le \eta^- < 1$, we can proceed as follows:

$$\begin{aligned} &\lim_{N\to\infty} D_R(\eta^-, N) \\ &= \lim_{N\to\infty}\frac{N-1}{(2-2\eta^-)(N+1)}\left\{\left(1-\frac{\eta^-}{N}+\frac{1-\eta^-}{N(N-1)}\right)\left(\left(1-\frac{\eta^-}{N}+\frac{1-\eta^-}{N(N-1)}\right)^{N/\eta^-}\right)^{\eta^-}\right. \\ &\qquad \left. -\left(1-\frac{2-\eta^-}{N}-\frac{1-\eta^-}{N(N-1)}\right)\left(\left(1-\frac{2-\eta^-}{N}-\frac{1-\eta^-}{N(N-1)}\right)^{N/(2-\eta^-)}\right)^{2-\eta^-}\right\} \\ &= \frac{1}{2-2\eta^-}\left(e^{-\eta^-} - e^{-(2-\eta^-)}\right). \end{aligned}$$

□

Numerical results of the function $D_R(\eta^-, N)$ are summarized in Table 4 and Figures 7 and 8. For population sizes $N \ge 10$, the function $D_R(\eta^-, N)$ only varies within a narrow interval $[0.349, 0.432]$. The expected loss of diversity $D_R(\eta^-, N)$ monotonically decreases in η^- (Figure 7). The function $D_R(\eta^-, N)$ is not necessarily monotonic in N (Figure 8), but in the subdomain $N \ge 10$, it only increases slowly; in fact, for any $N \ge 10$, $|\lim_{x\to\infty} D_R(\eta^-, x) - D_R(\eta^-, N)| \le 0.02$. When $\eta^- = 1$, linear ranking selection acts as random sampling; in such a case, Theorem 5.1(b) indicates that for any sufficiently large N, $D_R(1, N)$ is nearly equal to e^{-1}, which agrees with $D_T(1, N)$ in Theorem 3.2(b).

6 Expected Loss of Diversity of Exponential Ranking Selection

In *exponential ranking selection* (Michalewicz, 1992), each individual is assigned the selection probability

Table 4: Expected loss of diversity $D_R(\eta^-, N)$ of linear ranking selection.

Lowest Reprod. Rate η^-	Population Size N								
	2	3	4	5	7	10	20	50	inf
0.000	0.500	0.444	0.436	0.433	0.431	0.431	0.431	0.432	0.432
0.100	0.452	0.416	0.413	0.412	0.413	0.415	0.417	0.418	0.420
0.200	0.410	0.391	0.392	0.394	0.398	0.401	0.404	0.407	0.408
0.300	0.372	0.369	0.374	0.379	0.384	0.388	0.393	0.397	0.399
0.400	0.340	0.350	0.359	0.365	0.372	0.378	0.384	0.388	0.390
0.500	0.312	0.333	0.346	0.353	0.362	0.369	0.376	0.380	0.383
0.600	0.290	0.320	0.335	0.344	0.354	0.361	0.370	0.375	0.378
0.700	0.273	0.310	0.327	0.337	0.348	0.356	0.365	0.370	0.373
0.800	0.260	0.302	0.321	0.332	0.343	0.352	0.361	0.367	0.370
0.900	0.253	0.298	0.318	0.329	0.341	0.349	0.359	0.365	0.368
1.000	0.250	0.296	0.316	0.328	0.340	0.349	0.358	0.364	0.368

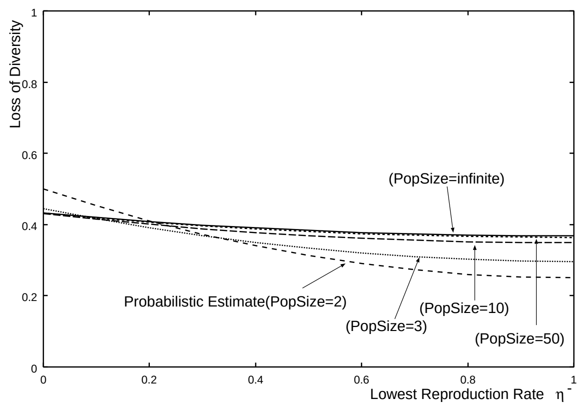

Figure 7: Expected loss of diversity $D_R(\eta^-, N)$ of linear ranking selection under fixed population size N.

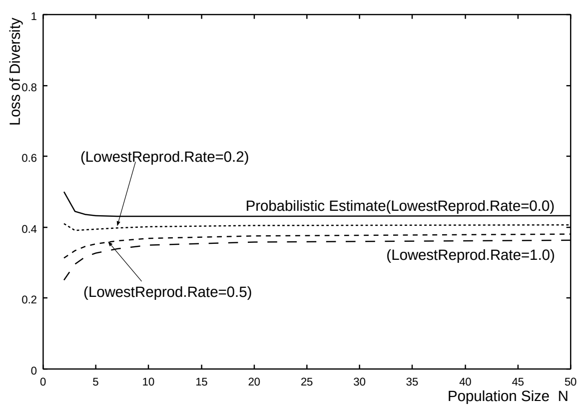

Figure 8: Expected loss of diversity $D_R(\eta^-, N)$ of linear ranking selection under fixed lowest reproduction rate η^-.

$$p_k \;=\; \frac{c^{N-k}}{\sum_{i=1}^{N} c^{N-i}} = \begin{cases} \dfrac{1}{N} & \text{if } c=1 \\ \dfrac{1-c}{1-c^N}c^{N-k} & \text{if } 0<c<1, \end{cases} \tag{4}$$

where k is the rank of the individual, c is a prespecified value satisfying $0 < c \leq 1$ (called *ranking base*), and N denotes the population size; then, roulette-wheel sampling with selection probabilities $\{p_k\}$ is repeated N times.

THEOREM 6.1: *Let $D_E(c,N)$ denote the expected loss of diversity in exponential ranking selection with ranking base $c \in (0,1]$ and population size N. Then*

(a) $D_E(c,N) = \begin{cases} \left(1-\dfrac{1}{N}\right)^N & \textit{if } c=1 \\ \dfrac{1}{N}\displaystyle\sum_{k=1}^{N}\left(1-\dfrac{1-c}{1-c^N}c^{N-k}\right)^N & \textit{if } 0<c<1. \end{cases}$

(b) $\lim\limits_{N\to\infty} D_E(c,N) = \begin{cases} e^{-1} & \textit{if } c=1 \\ 1 & \textit{if } 0<c<1. \end{cases}$

PROOF: (a) We can proceed with the proof in a similar manner as with the first half of the proof of Theorem 5.1 (a).

(b) When $c=1$, we can immediately derive the conclusion:

$$\lim_{N\to\infty} D_E(c,N) = \lim_{N\to\infty}\left(1-\frac{1}{N}\right)^N = e^{-1}.$$

When $0<c<1$, we observe that for any fixed positive integer k, we have

$$\begin{aligned} \lim_{N\to\infty}\left(1-\frac{1-c}{1-c^N}c^{N-k}\right)^N &= \lim_{N\to\infty}\left[\left(1-\frac{1}{\frac{1-c^N}{1-c}c^{-(N-k)}}\right)^{\frac{1-c^N}{1-c}c^{-(N-k)}}\right]^{\frac{1-c}{1-c^N}c^{N-k}N} \\ &= \lim_{N\to\infty}\exp\left(-\frac{1-c}{1-c^N}c^{N-k}N\right) \\ &= \lim_{N\to\infty}\exp\left(-\frac{1-c}{1-c^N}c^{N-\frac{\log N}{\log(1/c)}-k}\right) \\ &= 1. \end{aligned}$$

So, we next examine which values of $k \in [0, N-1]$ satisfy the approximation $\left(1-\frac{1-c}{1-c^N}c^k\right)^N \approx 1$. For any small value $\epsilon > 0$, we can repeatedly transform inequalies:

$$\begin{aligned} \left(1-\frac{1-c}{1-c^N}c^k\right)^N \geq 1-\epsilon \quad &\Leftrightarrow \quad \left(\frac{1}{c}\right)^k \geq \frac{1-c}{1-c^N}\frac{1}{1-(1-\epsilon)^{1/N}} \\ &\Leftrightarrow \quad k \geq \frac{1}{\log(1/c)}\left\{\log\frac{1-c}{1-c^N} - \log(1-(1-\epsilon)^{1/N})\right\}. \end{aligned}$$

Now, note that for any real number m, the function $(1+x)^m$ can be represented by an infinite Maclaurin's series, $(1+x)^m = 1 + \frac{m}{1!}x + \frac{m(m-1)}{2!}x^2 + \frac{m(m-1)(m-2)}{3!}x^3 + \cdots$ when $-1 < x < 1$; by utilizing this mathematical fact, we can give a simple lower bound for $\log(1-(1-\epsilon)^{1/N})$:

$$\begin{aligned}
&\log(1-(1-\epsilon)^{1/N}) \\
&= \log\left(1 - \left\{1 - \frac{1}{N}\epsilon + \frac{\frac{1}{N}\left(\frac{1}{N}-1\right)}{2!}\epsilon^2 - \frac{\frac{1}{N}\left(\frac{1}{N}-1\right)\left(\frac{1}{N}-2\right)}{3!}\epsilon^3 + \cdots\right\}\right) \\
&= \log\left(\frac{1}{N}\epsilon + \frac{\frac{1}{N}\left(1-\frac{1}{N}\right)}{2!}\epsilon^2 + \frac{\frac{1}{N}\left(1-\frac{1}{N}\right)\left(2-\frac{1}{N}\right)}{3!}\epsilon^3 + \cdots\right) \\
&\geq \log\frac{\epsilon}{N}.
\end{aligned}$$

Thus, we can deduce the claim that for any small value $\epsilon > 0$ and any nonnegative integer $k \geq \frac{1}{\log(1/c)}\left\{\log\frac{1-c}{1-c^N} - \log\frac{\epsilon}{N}\right\}$, the inequality $\left(1 - \frac{1-c}{1-c^N}c^k\right)^N \geq 1-\epsilon$ holds, and hence,

$$\begin{aligned}
&\frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{1-c}{1-c^N}c^{N-k}\right)^N \\
&= \frac{1}{N}\sum_{k=0}^{N-1}\left(1 - \frac{1-c}{1-c^N}c^{k}\right)^N \\
&\geq \frac{1}{N}(1-\epsilon)\left[N - 1 - \frac{1}{\log(1/c)}\left\{\log\frac{1-c}{1-c^N} - \log\frac{\epsilon}{N}\right\}\right] \\
&= 1 - \epsilon - \frac{1-\epsilon}{N} - \frac{1-\epsilon}{N\log(1/c)}\left\{\log N - \log\frac{1-c^N}{1-c} + \log\frac{1}{\epsilon}\right\}.
\end{aligned}$$

Letting $N \to \infty$, we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{1-c}{1-c^N}c^{N-k}\right)^N \geq 1-\epsilon.$$

Since we can assign an arbitrarily small positive value to ϵ, we conclude that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{1-c}{1-c^N}c^{N-k}\right)^N \geq 1.$$

From this inequality and the trivial fact

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{1-c}{1-c^N}c^{N-k}\right)^N \leq \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}(1-0)^N = 1,$$

we obtain the final conclusion:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{1-c}{1-c^N}c^{N-k}\right)^N = 1. \qquad \square$$

Numerical results of the function $D_E(c, N)$ are summarized in Table 5 and Figures 9 and 10. When $c = 1$, exponential ranking selection acts as random sampling; in such a case, Theorem 6.1(b) indicates that for any sufficiently large N, $D_E(c, N)$ is nearly equal to e^{-1}, which agrees with $D_T(1, N)$ in Theorem 3.2(b). When $0 < c < 1$, the expected loss of diversity $D_E(c, N)$ monotonically decreases in c (Figure 9) and rapidly increases in N (Figure 10). Function $D_E(c, N)$ is apt to take a value near 1; even if we greatly enlarge the population, the number of surviving individuals will only increase a little. If we want to keep $D_E(c, N)$ in a certain moderate level for various population sizes, we must change the value of ranking base parameter c depending on population size; in this respect, the exponential ranking scheme has a different type of control parameter than tournament, truncation, and linear ranking schemes, although Bäck (1994, 1996) and Julstrom (1999) observed that tournament selection with tournament size t and exponential ranking selection with ranking base c behave nearly equivalently if population size N is sufficiently large and $c = \left(1 - \frac{1}{N}\right)^t$.

7 Variance of Loss of Diversity

We now calculate variance of loss of diversity. Given a random variable X, we use $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ to denote the expected value and variance of X, respectively.

THEOREM 7.1: *Consider an arbitrary selection procedure in which individuals are selected one by one according to some selection probability distribution. Let p_k be the selection probability assigned to the individual of rank k, and let N be the population size. Then*

(a) *the variance of loss of diversity is*

Var(*LossOfDiversity*)

$$= \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}(1-p_k-p_m)^N - \frac{1}{N^2}\left\{\sum_{k=1}^{N}(1-p_k)^N\right\}^2 + \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N.$$

(b) $\mathrm{Var}(\mathit{LossOfDiversity}) \le \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N - \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^{2N}$

(c) $\lim_{N\to\infty} \mathrm{Var}(\mathit{LossOfDiversity}) = 0$

PROOF: (a) By using random variables,

$L =$ the number of individuals not selected during the selection phase

and

$$L_k = \begin{cases} 1 & \text{if the individual of rank } k \text{ is not selected during the selection phase} \\ 0 & \text{otherwise,} \end{cases}$$

we calculate the variance of loss of diversity as follows:

$$\begin{aligned}
\mathrm{Var}(\mathit{LossOfDiversity}) &= \mathrm{Var}\left(\frac{L}{N}\right) \\
&= \mathrm{E}\left(\left(\frac{L}{N}\right)^2\right) - \left\{\mathrm{E}\left(\frac{L}{N}\right)\right\}^2 \\
&= \frac{1}{N^2}\mathrm{E}\left((L_1+L_2+\cdots+L_N)^2\right) - \frac{1}{N^2}\left\{\mathrm{E}(L_1+L_2+\cdots+L_N)\right\}^2 \\
&= \frac{1}{N^2}\sum_{k=1}^{N}\mathrm{E}(L_k^2) + \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}\mathrm{E}(L_kL_m) - \frac{1}{N^2}\left\{\sum_{k=1}^{N}\mathrm{E}(L_k)\right\}^2.
\end{aligned}$$

Table 5: Expected loss of diversity $D_E(c, N)$ of exponential ranking selection.

Ranking Base c	Population Size N								
	2	3	5	7	10	50	200	1000	inf
0.100	0.417	0.576	0.715	0.778	0.829	0.952	0.985	0.996	1.000
0.200	0.361	0.501	0.646	0.719	0.782	0.937	0.980	0.995	1.000
0.300	0.322	0.441	0.582	0.664	0.736	0.921	0.975	0.994	1.000
0.400	0.296	0.393	0.521	0.605	0.686	0.903	0.968	0.992	1.000
0.500	0.278	0.358	0.464	0.543	0.628	0.880	0.960	0.990	1.000
0.600	0.266	0.332	0.414	0.480	0.561	0.850	0.949	0.987	1.000
0.700	0.258	0.314	0.374	0.422	0.486	0.805	0.932	0.982	1.000
0.800	0.253	0.304	0.347	0.376	0.416	0.731	0.902	0.973	1.000
0.900	0.251	0.298	0.332	0.349	0.366	0.577	0.828	0.950	1.000
0.950	0.250	0.297	0.329	0.342	0.353	0.446	0.717	0.912	1.000
0.980	0.250	0.296	0.328	0.340	0.349	0.379	0.523	0.823	1.000
0.995	0.250	0.296	0.328	0.340	0.349	0.365	0.382	0.568	1.000
1.000	0.250	0.296	0.328	0.340	0.349	0.364	0.367	0.368	0.368

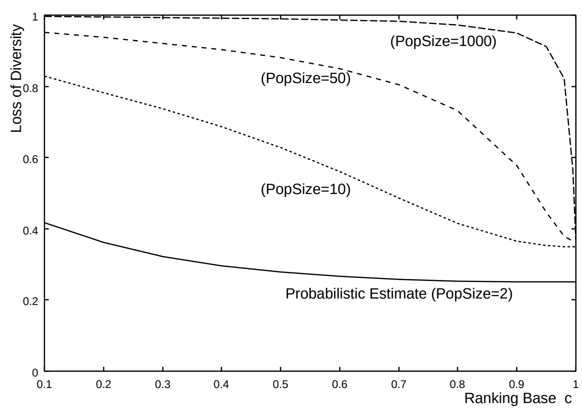

Figure 9: Expected loss of diversity $D_E(c, N)$ of exponential ranking selection under fixed population size N.

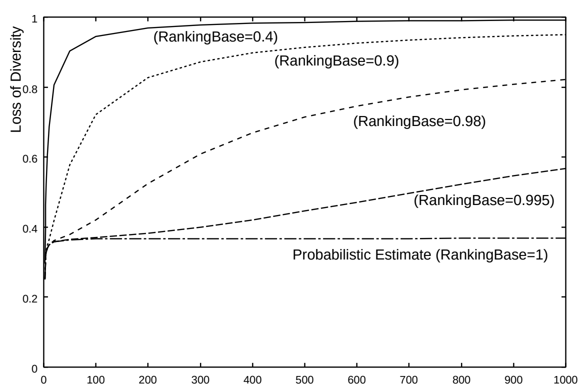

Figure 10: Expected loss of diversity $D_E(c, N)$ of exponential ranking selection under fixed ranking base c.

Now, we have $L_k^2 = L_k$; from the choice of p_k, we also obtain

$$\begin{aligned} \mathrm{E}(L_k) &= \Pr(\text{the individual of rank } k \text{ is not selected during the selection phase}) \\ &= (1-p_k)^N, \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}(L_k L_m) &= \Pr(\text{the individuals of rank } k \text{ and } m \\ &\qquad\quad \text{are both unselected during the selection phase}) \\ &= (1-p_k-p_m)^N. \end{aligned}$$

From these equations, the required equation follows.

(b) If $k \neq m$, then from the choice of p_k's, it must hold that $1 - p_k - p_m \geq 0$. So we can continue calculation of (a) as follows:

$$\begin{aligned} &\mathrm{Var}(\text{LossOfDiversity}) \\ &\leq \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}(1-p_k-p_m+p_kp_m)^N - \frac{1}{N^2}\left\{\sum_{k=1}^{N}(1-p_k)^N\right\}^2 + \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N \\ &= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{N}(1-p_k-p_m+p_kp_m)^N - \frac{1}{N^2}\sum_{k=1}^{N}(1-2p_k+p_k^2)^N \\ &\qquad -\frac{1}{N^2}\left\{\sum_{k=1}^{N}(1-p_k)^N\right\}^2 + \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N \\ &= \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N\sum_{m=1}^{N}(1-p_m)^N - \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^{2N} \\ &\qquad -\frac{1}{N^2}\left\{\sum_{k=1}^{N}(1-p_k)^N\right\}^2 + \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N \\ &= \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N - \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^{2N}. \end{aligned}$$

(c) From the inequality in (b), we have

$$\begin{aligned} 0 &\leq \lim_{N\to\infty}\mathrm{Var}(\text{LossOfDiversity}) \\ &\leq \lim_{N\to\infty}\frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N \leq \lim_{N\to\infty}\frac{1}{N^2}\sum_{k=1}^{N}1 = \lim_{N\to\infty}\frac{1}{N} = 0. \end{aligned}$$

Thus, $\lim_{N\to\infty}\mathrm{Var}(\text{LossOfDiversity}) = 0$. □

COROLLARY 7.2: (a) *For tournament selection with tournament size t and population size N, the variance of loss of diversity is*

$$
\begin{aligned}
&\mathrm{Var}(\textit{LossOfDiversity})\\
&= \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}\left(1-\frac{k^t-(k-1)^t}{N^t}-\frac{m^t-(m-1)^t}{N^t}\right)^N\\
&\qquad -\frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N\right\}^2+\frac{1}{N^2}\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N\\
&\le \frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^N-\sum_{k=1}^{N}\left(1-\frac{k^t-(k-1)^t}{N^t}\right)^{2N}\right\}.
\end{aligned}
$$

(b) *For truncation selection with threshold T and population size N, the variance of loss of diversity is*

$$
\begin{aligned}
&\mathrm{Var}(\textit{LossOfDiversity})\\
&= \frac{\lceil TN\rceil}{N^2}\left\{(\lceil TN\rceil-1)\left(1-\frac{2}{\lceil TN\rceil}\right)^N-\lceil TN\rceil\left(1-\frac{1}{\lceil TN\rceil}\right)^{2N}+\left(1-\frac{1}{\lceil TN\rceil}\right)^N\right\}\\
&\le \frac{\lceil TN\rceil}{N^2}\left\{\left(1-\frac{1}{\lceil TN\rceil}\right)^N-\left(1-\frac{1}{\lceil TN\rceil}\right)^{2N}\right\}.
\end{aligned}
$$

(c) *For linear ranking selection with the lowest reproduction rate η^- and population size N, the variance of loss of diversity is*

$$
\begin{aligned}
\mathrm{Var}(\textit{LossOfDiversity}) &= \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}\left(1-\frac{1}{N}\left(2\eta^-+(2-2\eta^-)\frac{k+m-2}{N-1}\right)\right)^N\\
&\qquad -\frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1}{N}\left(\eta^-+(2-2\eta^-)\frac{k-1}{N-1}\right)\right)^N\right\}^2\\
&\qquad +\frac{1}{N^2}\sum_{k=1}^{N}\left(1-\frac{1}{N}\left(\eta^-+(2-2\eta^-)\frac{k-1}{N-1}\right)\right)^N\\
&\le \frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1}{N}\left(\eta^-+(2-2\eta^-)\frac{k-1}{N-1}\right)\right)^N\right.\\
&\qquad \left.-\sum_{k=1}^{N}\left(1-\frac{1}{N}\left(\eta^-+(2-2\eta^-)\frac{k-1}{N-1}\right)\right)^{2N}\right\}.
\end{aligned}
$$

(d) *For exponential ranking selection with ranking base c and population size N, if $0 < c < 1$ then the variance of loss of diversity is*

$$
\begin{aligned}
\mathrm{Var}(\textit{LossOfDiversity}) &= \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}\left(1-\frac{1-c}{1-c^N}c^{N-k}-\frac{1-c}{1-c^N}c^{N-m}\right)^N\\
&\qquad -\frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1-c}{1-c^N}c^{N-k}\right)^N\right\}^2+\frac{1}{N^2}\sum_{k=1}^{N}\left(1-\frac{1-c}{1-c^N}c^{N-k}\right)^N\\
&\le \frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1-c}{1-c^N}c^{N-k}\right)^N-\sum_{k=1}^{N}\left(1-\frac{1-c}{1-c^N}c^{N-k}\right)^{2N}\right\};
\end{aligned}
$$

if $c=1$ then the variance of loss of diversity is

$$\begin{aligned}
&\mathrm{Var}(\mathit{LossOfDiversity}) \\
&= \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}\left(1-\frac{2}{N}\right)^N - \frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1}{N}\right)^N\right\}^2 + \frac{1}{N^2}\sum_{k=1}^{N}\left(1-\frac{1}{N}\right)^N \\
&\leq \frac{1}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1}{N}\right)^N - \sum_{k=1}^{N}\left(1-\frac{1}{N}\right)^{2N}\right\}.
\end{aligned}$$

PROOF: Assertions (a), (c), and (d) immediately follow from Theorem 7.1(a) and (b), and the selection probabilities given in Equations (1), (3), and (4). As for assertion (b), from Equation (2) we have

$$p_k = \begin{cases} 0 & \text{if } 1 \leq k \leq \lfloor (1-T)N \rfloor \\ \dfrac{1}{\lceil TN \rceil} & \text{if } \lfloor (1-T)N \rfloor + 1 \leq k \leq N; \end{cases}$$

so we can proceed with calculation of Theorem 7.1(a) as follows:

$$\begin{aligned}
&\mathrm{Var}(\mathrm{LossOfDiversity}) \\
&= \frac{2}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{k-1}(1-p_k-p_m)^N - \frac{1}{N^2}\left\{\sum_{k=1}^{N}(1-p_k)^N\right\}^2 + \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N \\
&= \frac{2}{N^2}\left\{\frac{1}{2}\lfloor (1-T)N \rfloor(\lfloor (1-T)N \rfloor - 1) + \lceil TN \rceil \lfloor (1-T)N \rfloor \left(1-\frac{1}{\lceil TN \rceil}\right)^N\right. \\
&\quad \left. + \frac{1}{2}\lceil TN \rceil(\lceil TN \rceil - 1)\left(1-\frac{2}{\lceil TN \rceil}\right)^N\right\} \\
&\quad - \frac{1}{N^2}\left\{\lfloor (1-T)N \rfloor + \lceil TN \rceil\left(1-\frac{1}{\lceil TN \rceil}\right)^N\right\}^2 \\
&\quad + \frac{1}{N^2}\left\{\lfloor (1-T)N \rfloor + \lceil TN \rceil\left(1-\frac{1}{\lceil TN \rceil}\right)^N\right\} \\
&= \frac{\lceil TN \rceil}{N^2}\left\{(\lceil TN \rceil - 1)\left(1-\frac{2}{\lceil TN \rceil}\right)^N\right. \\
&\quad \left. - \lceil TN \rceil\left(1-\frac{1}{\lceil TN \rceil}\right)^{2N} + \left(1-\frac{1}{\lceil TN \rceil}\right)^N\right\}.
\end{aligned}$$

We can also proceed with calculation of Theorem 7.1(b) as follows:

$$\begin{aligned}
\mathrm{Var}(\mathrm{LossOfDiversity}) &\leq \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^N - \frac{1}{N^2}\sum_{k=1}^{N}(1-p_k)^{2N} \\
&= \frac{1}{N^2}\left\{\lfloor (1-T)N \rfloor + \lceil TN \rceil\left(1-\frac{1}{\lceil TN \rceil}\right)^N\right\} \\
&\quad - \frac{1}{N^2}\left\{\lfloor (1-T)N \rfloor + \lceil TN \rceil\left(1-\frac{1}{\lceil TN \rceil}\right)^{2N}\right\} \\
&= \frac{\lceil TN \rceil}{N^2}\left\{\sum_{k=1}^{N}\left(1-\frac{1}{\lceil TN \rceil}\right)^N - \sum_{k=1}^{N}\left(1-\frac{1}{\lceil TN \rceil}\right)^{2N}\right\}. \qquad \square
\end{aligned}$$

Table 6: The variance of loss of diversity in tournament selection.

Tournament Size t	Population Size N								
	2	3	4	5	7	10	20	50	200
1	0.250	0.189	0.161	0.143	0.120	0.100	0.070	0.044	0.022
2	0.242	0.190	0.162	0.144	0.121	0.100	0.071	0.045	0.022
3	0.207	0.183	0.155	0.138	0.116	0.096	0.068	0.043	0.021
4	0.161	0.174	0.148	0.130	0.109	0.091	0.064	0.041	0.020
5	0.119	0.160	0.142	0.124	0.104	0.086	0.061	0.038	0.019
6	0.086	0.143	0.136	0.119	0.098	0.082	0.058	0.037	0.018
7	0.062	0.124	0.129	0.115	0.094	0.078	0.055	0.035	0.017
8	0.044	0.105	0.121	0.111	0.090	0.075	0.053	0.033	0.017
9	0.031	0.088	0.112	0.108	0.087	0.072	0.051	0.032	0.016
10	0.022	0.073	0.102	0.103	0.084	0.069	0.049	0.031	0.015
15	0.004	0.027	0.056	0.074	0.077	0.059	0.042	0.027	0.013
20	0.001	0.010	0.028	0.046	0.065	0.055	0.037	0.024	0.012
25	0.000	0.004	0.014	0.027	0.050	0.053	0.034	0.021	0.011
30	0.000	0.001	0.007	0.016	0.036	0.049	0.031	0.020	0.010

Table 7: The variance of loss of diversity in truncation selection.

Threshold T	Population Size N								
	2	3	4	5	7	10	20	50	200
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.200	0.000	0.000	0.000	0.000	0.018	0.004	0.006	0.004	0.002
0.300	0.000	0.000	0.083	0.048	0.055	0.022	0.019	0.013	0.007
0.400	0.000	0.144	0.083	0.048	0.055	0.043	0.033	0.021	0.011
0.500	0.000	0.144	0.083	0.103	0.084	0.061	0.044	0.028	0.014
0.600	0.250	0.144	0.140	0.103	0.102	0.074	0.053	0.033	0.017
0.700	0.250	0.189	0.140	0.130	0.102	0.084	0.059	0.037	0.019
0.800	0.250	0.189	0.161	0.130	0.113	0.091	0.064	0.040	0.020
0.900	0.250	0.189	0.161	0.143	0.120	0.096	0.068	0.043	0.021
1.000	0.250	0.189	0.161	0.143	0.120	0.100	0.070	0.044	0.022

Table 8: The variance of loss of diversity in linear ranking selection.

Lowest Reprod. Rate η^-	Population Size N								
	2	3	4	5	7	10	20	50	200
0.000	0.000	0.157	0.146	0.134	0.116	0.098	0.070	0.044	0.022
0.100	0.147	0.174	0.155	0.140	0.120	0.100	0.071	0.045	0.022
0.200	0.192	0.184	0.161	0.144	0.122	0.102	0.072	0.045	0.023
0.300	0.218	0.189	0.163	0.146	0.122	0.102	0.072	0.045	0.023
0.400	0.233	0.192	0.164	0.146	0.122	0.102	0.072	0.045	0.023
0.500	0.242	0.192	0.164	0.146	0.122	0.102	0.071	0.045	0.022
0.600	0.247	0.192	0.163	0.145	0.121	0.101	0.071	0.045	0.022
0.700	0.249	0.191	0.163	0.144	0.121	0.100	0.071	0.045	0.022
0.800	0.250	0.190	0.162	0.143	0.120	0.100	0.070	0.044	0.022
0.900	0.250	0.189	0.161	0.143	0.120	0.100	0.070	0.044	0.022
1.000	0.250	0.189	0.161	0.143	0.120	0.100	0.070	0.044	0.022

Table 9: The variance of loss of diversity in exponential ranking selection.

Ranking Base c	Population Size N								
	2	3	5	7	10	20	50	200	1000
0.100	0.186	0.152	0.105	0.079	0.057	0.026	0.011	0.003	0.001
0.200	0.224	0.183	0.123	0.090	0.063	0.032	0.013	0.003	0.001
0.300	0.239	0.194	0.134	0.100	0.071	0.037	0.015	0.004	0.001
0.400	0.246	0.197	0.144	0.110	0.080	0.042	0.017	0.004	0.001
0.500	0.248	0.196	0.149	0.119	0.089	0.047	0.020	0.005	0.001
0.600	0.250	0.194	0.150	0.125	0.098	0.054	0.023	0.006	0.001
0.700	0.250	0.192	0.149	0.126	0.104	0.062	0.027	0.007	0.001
0.800	0.250	0.190	0.146	0.124	0.105	0.071	0.033	0.009	0.002
0.900	0.250	0.189	0.144	0.121	0.102	0.074	0.043	0.012	0.003
0.950	0.250	0.189	0.143	0.120	0.100	0.071	0.047	0.017	0.004
0.980	0.250	0.189	0.143	0.120	0.100	0.070	0.045	0.023	0.006
0.995	0.250	0.189	0.143	0.120	0.100	0.070	0.044	0.022	0.010
1.000	0.250	0.189	0.143	0.120	0.100	0.070	0.044	0.022	0.010

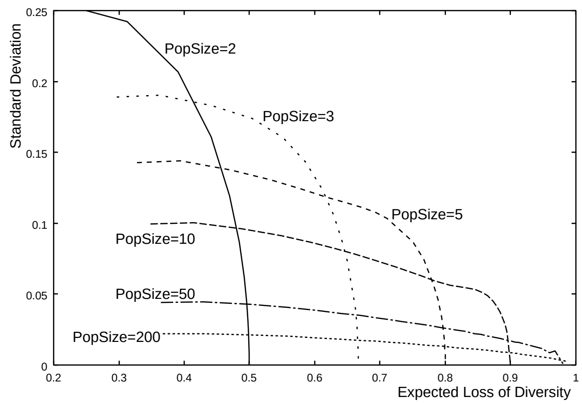

Figure 11: Possible pairs of expected value and variance of loss of diversity in tournament selection.

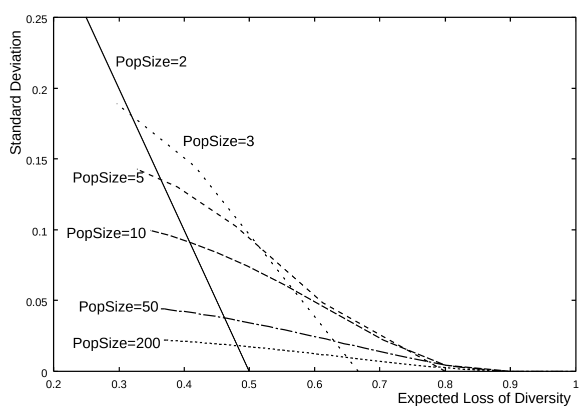

Figure 12: Possible pairs of expected value and variance of loss of diversity in truncation selection.

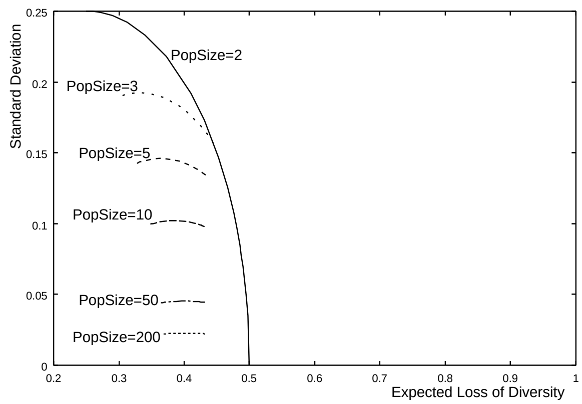

Figure 13: Possible pairs of expected value and variance of loss of diversity in linear ranking selection.

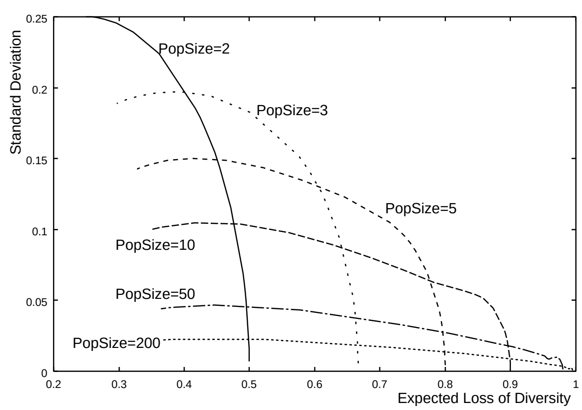

Figure 14: Possible pairs of expected value and variance of loss of diversity in exponential ranking selection.

Tables 6 to 9 give numerical results of standard deviation $\sqrt{\mathrm{Var(LossOfDiversity)}}$ in tournament selection, truncation selection, linear ranking selection, and exponential ranking selection, respectively. Figures 11 to 14 show what pair of values is possible as a pair of expected value and standard deviation of loss of diversity with fixed population size. In these figures, we observe respective characteristics of four selection schemes. For example, in Figure 12, we observe that the deviation clearly falls as the expected value increases, which agrees with the fact that in truncation selection most of dismissals are performed deterministically. In Figure 13, we observe that both expected value and deviation only vary within a narrow range; therefore, in linear ranking selection, the speed of evolution will only slightly vary through the lowest reproduction parameter η^-. We also observe that Figure 14 gives almost the same locus as Figure 11; so exponential ranking and tournament schemes are seen to bring about nearly equivalent selection behaviors via some transformation between control parameters, as Bäck (1996) and Julstrom (1999) observed.

8 Summary

Although the loss of diversity is considered to be a fundamental measure of selection pressure, study about this subject was only performed by Blickle and Thiele (1995, 1997); their estimates do not agree with numerical results by simulation. So we probabilistically calculated the expected value and variance of the loss of diversity in tournament selection, truncation selection, linear ranking selection, and exponential ranking selection. Our theoretical results agree with numerical results by simulation.

From our results, we understood that even in random sampling, $e^{-1}\times 100 \approx 36.8\%$ of the population is apt to be lost during the selection phase. From numerical results, we showed that in tournament selection, many more individuals are expected to be lost than with Blickle and Thiele's static estimate. We observed that tournament and exponential ranking schemes potentially bring about nearly equivalent selection behaviors but have different types of control parameters; we also observed that in linear ranking selection, the evolution speed will only vary slightly through the lowest reproduction parameter η^-.

Acknowledgments

The author would like to thank anonymous reviewers and Dr. Robert E. Smith for their helpful comments.

References

Bäck, T. (1994). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, pages 57–62, IEEE Press, Piscataway, New Jersey.

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York.

Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. In Grefenstette, J. J., editor, *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*, pages 101–111, Lawrence Erlbaum, Hillsdale, New Jersey.

Banzhaf, W. et al. (1998). *Genetic Programming An Introduction*. Morgan Kaufmann, San Francisco, California.

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

Blickle, T. and Thiele, L. (1997). A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394.

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

Julstrom, B. A. (1999). It's all the same to me: Revisiting rank-based probabilities and tournaments. In *Proceedings of the Congress on Evolutionary Computation*, pages 1501–1505, IEEE Press, Piscataway, New Jersey.

Michalewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin, Germany.

Mühlenbein, H. and Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm I. Continuous parameter optimization. *Evolutionary Computation*, 1(1):25–49.

Schell, T. and Wegenkittl, S. (2001). Looking beyond selection probabilities: Adaptation of the χ^2 measure for the performance analysis of selection methods in GAs. *Evolutionary Computation*, 9(2):243–256.

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