Modeling Alternate Selection Schemes For Genetic Algorithms

Michael D. Vose

Computer Science Dept., 107 Ayres Hall
The University of Tennessee
Knoxville, TN 37996-1301

vose@cs.utk.edu

ABSTRACT

Beginning with a representational framework of which genetic algorithms are a special case, the ranking and tournament selection schemes are defined and formalized as mathematical functions. The main result is that ranking and tournament selection are diffeomorphisms of the representation space. Explicit algorithms are also developed for computing their inverses.

1 Introduction

Although there has been recent progress in the mathematical analysis of Genetic Algorithms (GAs), the gap between practice and theory is large. When applied as optimization tools, GAs are used with various operators which have not yet been formalized or considered outside an empirical context.

Besides the standard practice of creating adhoc recombination operators, a variety of selection schemes are used. Practitioners often prefer some substitute for classical proportional selection[1]. This paper concerns the mathematical formalization of two popular alternatives: ranking selection and tournament selection.

By way of comparison with Hollands "schema theorem" [2], this paper is not concerned with estimating the change of schemata from one generation to the next. The population is taken as the fundamental object of interest and analysis is focused on the behavior of selection with respect to it.

This paper begins with a representational framework for a general type of heuristic search of which genetic algorithms are a special case. Next, the ranking and tournament selection schemes are defined and formalized as functions. Some of their basic mathematical properties are then considered.

[1] In the proportional selection scheme, population members are chosen in proportion to their fitness.

Apart from their formalization as functions, the main contribution of this paper is proving that the ranking and tournament selection schemes are diffeomorphisms of the representation space. Explicit algorithms are also developed for computing their inverses.

1.1 Notation

The set of integers is denoted by Z, and the symbol $\Re$ denotes the set of real numbers. Angle brackets $<\cdots>$ denote a tuple which is to be regarded as a column vector. The vector with all components 1 is denoted by $\mathbf{1}$. Indexing of vectors begins with 0. For any collection C of vectors, αC denotes the collection whose members are those of C multiplied by α.

Composition of functions f and g is $f \circ g(x) = f(g(x))$. Modulus (or absolute value) is denoted by $|\cdot|$. Square brackets $[\cdots]$ are, besides their standard use as specifying a closed interval of real numbers, used to denote an *indicator function*: if *expr* is an expression which may be true or false, then

$$[expr] = \begin{cases} 1 & \text{if } expr \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

The delta function is $\delta_{i,j} = [i = j]$.

2 Representation

From an abstract perspective, a genetic algorithm can be thought of as an initial collection of elements P_0 chosen from some *search space* Ω of finite cardinality n together with some nondeterministic *transition rule* τ which from P_i will produce another collection P_{i+1}. In general, τ will be iterated to produce a sequence of collections

$$P_0 \xrightarrow{\tau} P_1 \xrightarrow{\tau} P_2 \xrightarrow{\tau} \ldots$$

The beginning collection P_0 is referred to as the *initial population*, the first population (or *generation*) is P_1, the second generation is P_2, and so on. Populations are multisets.

Obtaining a good representation for populations is a first step towards characterizing populations geometrically. Define the *simplex* to be the set

$$\Lambda = \{<x_0, ..., x_{n-1}> : \mathbf{1}^T x = 1,\ x_j \geq 0\}$$

An element p of Λ corresponds to a population according to the following rule for defining its components

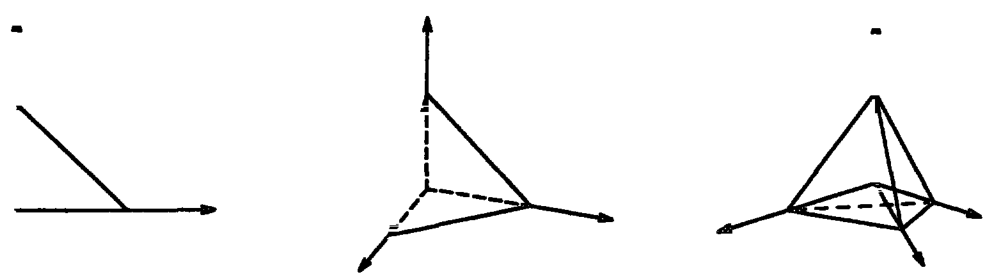
$$p_j = \text{the proportion in the population of the } j\text{th element of } \Omega$$

Since Ω may be enumerated, it can without loss of generality be thought of as $\{0, 1, \ldots, n-1\}$. For example, if $n = 6$ then population $\{1, 0, 3, 1, 1, 3, 2, 2, 4, 0\}$ would be represented by the vector $p = <.2, .3, .2, .2, .1, .0>$ given that

coordinate	corresponding element of Ω	percentage of P_0
p_0	0	2/10
p_1	1	3/10
p_2	2	2/10
p_3	3	2/10
p_4	4	1/10
p_5	5	0/10

The cardinality of each generation $P_0, P_1, \ldots$ is a parameter r called the *population size.* Hence the proportional representation given by p unambiguously determines a population once r is known. The vector p is referred to as a *population vector.* The distinction between population and population vector will often be blurred, because the population size is fixed and they are equivalent in that context. In particular, τ may be thought of as mapping the current population vector to the next.

To get a feel for the geometry of the representation space, Λ is shown in the following sequence of diagrams for $n = 2, 3, 4$. The figures represent Λ (a line segment, triangle, and solid tetrahedron). The arrows show the coordinate axes of the ambient space (the projection of the coordinate axes are being viewed in the second figure, which is three dimensional, and in the last figure where the ambient space is four dimensional).



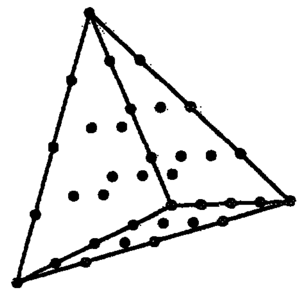
In general, Λ is a tetrahedron of dimension $n-1$ contained in an ambient space of dimension n. Note that each vertex of Λ corresponds to a unit basis vector of the ambient space; Λ is their convex hull. For example, the vertices of the solid tetrahedron (the right most figure) are at the basis vectors $e_0 =< 1, 0, 0, 0 >$, $e_1 =< 0, 1, 0, 0 >$, $e_2 =< 0, 0, 1, 0 >$, and $e_3 =< 0, 0, 0, 1 >$. They correspond (respectively) to the following populations: r copies of 0, r copies of 1, r copies of 2, and r copies of 3.

It should be realized that not every point of Λ corresponds to a finite population. In fact, only those rational points with common denominator r correspond to populations of size r.

They are

$$\frac{1}{r}X_n^r \;=\; \frac{1}{r}\{<x_0,\ldots,x_{n-1}>: \; x_j \in Z,\; x_j \geq 0,\; 1^T x = r\}$$

For example, the points corresponding to $\frac{1}{4}X_4^4$ ($n = 4$ and $r = 4$) are the dots in the following figure



As $r \to \infty$, these rational points become dense in Λ. Since a rational point may represent arbitrarily large populations, a point p of Λ carries little information concerning population size. A natural view is therefore that Λ corresponds to populations of indeterminate size. This is but one of several useful interpretations. Another is that Λ corresponds to sampling distributions over Ω: since the components of p are nonnegative and sum to 1, p may be viewed as indicating that i is sampled with probability p_i.

To complete the picture of the genetic algorithm would require that the details of the stochastic transition function τ be filled in. However, the remaining details very much depend upon which GA variant is being used. Moreover, most of the remaining details are tangential to the focus of this paper, which concerns selection, and so they will be left unspecified.

3 Selection

The symbol s will be used for three equivalent (though different) things. This overloading of s does not take long to get used to because context makes meaning clear. The benefits are clean and elegant presentation and the ability to use a common symbol for ideas whose differences are often conveniently blurred.

First, $s \in \Lambda$ can be regarded as a *selection distribution* describing the probability s_i with which i is selected (with replacement) from the current population for participation in forming the next generation. A selected element is an intermediate step towards producing the next population, not typically a member of it. In total, $2r$ such selections are typically made, the aggregate of which is sometimes referred to as the *gene pool*.

Second, $s : \Lambda \to \Omega$ can be regarded as a *selection function* which is nondeterministic. The result $s(p)$ of applying s to p is i with probability given by the ith component s_i of the selection distribution. Of course, for there to be a nontrivial dependence on p, the selection distribution must be some function $\mathcal{F}$ of p. The function $\mathcal{F} : \Lambda \to \Lambda$ is referred to as the *selection scheme*.

Third, $s \in \Lambda$ can be regarded as a population vector.

In analogy with survival of the fittest, an integral part of $\mathcal{F}$ is a *fitness function* $f : \Omega \to \Re$ which is used (in a variety of ways) to determine a selection scheme. The fitness function is assumed to be injective.[2] The value $f(i)$ is called the *fitness* of i. Through the identification $f_i = f(i)$, the fitness function may be regarded as a vector.

3.1 Ranking selection

Ranking selection refers to the selection function corresponding to the selection scheme

$$\mathcal{F}(x)_i = \int_{\sum[f_j<f_i]x_j}^{\sum[f_j\leq f_i]x_j} \varrho(y)\, dy$$

where ϱ is any continuous increasing probability density over $[0,1]$.

For example, let $n = 4$ and $p = <.2, .3, .1, .4>$. If the fitness function is $f(x) = \ln(1+x)$, then $f(0) < f(1) < f(2) < f(3)$. Assuming $\varrho(y) = 2y$ gives

i	$\sum[f_j < f_i]x_j$	$\mathcal{F}(x)_i$	$s_i = \mathcal{F}(p)_i$
0	0	$\int_0^{x_0} \varrho(y)\, dy$	0.04
1	x_0	$\int_{x_0}^{x_0+x_1} \varrho(y)\, dy$	0.21
2	$x_0 + x_1$	$\int_{x_0+x_1}^{x_0+x_1+x_2} \varrho(y)\, dy$	0.11
3	$x_0 + x_1 + x_2$	$\int_{x_0+x_1+x_2}^{1} \varrho(y)\, dy$	0.64

Hence 0 is chosen by the selection function for the gene pool with probability 0.04, while 2 is chosen with probability 0.11. This example illustrates that the selection distribution depends on how the elements of Ω are ranked by f; any increasing fitness function ($i < j \Longrightarrow f(i) < f(j)$) would yield these same results. Also note that no reference to population size was made (this example is independent of r).

In practice, the ranking selection function is not implemented by computing a selection distribution over the search space so as to choose i with probability s_i. Instead, the following procedure is followed.

[2]In practice f is not injective, but an arbitrarily small perturbation would make it so. Because the assumption simplifies analysis, it is made throughout this paper.

1. Sort the elements of the population by increasing fitness, obtaining $q_0 \leq \ldots \leq q_{r-1}$.
2. Choose element q_i (with replacement) with probability t_i where t is some fixed probability vector of dimension r (i.e., $t_0 + \cdots + t_{r-1} = 1$) such that $i < j \Longrightarrow t_i < t_j$.

This procedural method yields identical results, as will now be shown. Let the elements of the population P be indexed so that $f(q_0) \leq \ldots \leq f(q_{r-1})$. The selection distribution is

$$s_i = \int_{\sum[f_j < f_i]p_j}^{\sum[f_j \leq f_i]p_j} \varrho(y)\, dy$$

If $i \notin P$ then $p_i = 0$, therefore

$$\sum_j [f_j \leq f_i]\, p_j = \sum_j [f_j < f_i]\, p_j$$

Hence $s_i = 0$ since the upper and lower limits of the integral coincide. If $i \in P$, the lower limit is

$$\begin{aligned} \frac{1}{r} \sum_j [f_j < f_i] \sum_{k \in P} \delta_{j,k} &= \frac{1}{r} \sum_{k \in P} \sum_j [f_j < f_i]\, \delta_{j,k} \\ &= \frac{1}{r} \sum_{k \in P} [f_k < f_i] \\ &= \{\text{the smallest } j \text{ such that } i = q_j\}/r \end{aligned}$$

Similarly, the upper limit is $\{\text{the largest } j \text{ such that } i = q_{j-1}\}/r$. It follows that

$$s_i = \sum [i = q_j] \int_{j/r}^{(j+1)/r} \varrho(y)\, dy$$

Now define

$$t_j = \int_{j/r}^{(j+1)/r} \varrho(y)\, dy$$

and consider choosing q_j from P with probability t_j. The probability of choosing i is

$$\sum [i = q_j]\, t_j$$

which is identical to s_i. This establishes the following.

Theorem 3.1 *Sorting the population by fitness followed by choosing members (based on rank) according to the distribution t implements the ranking selection function.*

Previous to this work, only the procedural definition of ranking selection was available. The preceding theorem establishes the correctness of the functional formulation provided by the ranking selection scheme. This alternate representation has interesting properties, some of which will be explored in section 4.

3.2 Tournament selection

Tournament selection refers to the selection function corresponding to the selection scheme

$$\mathcal{F}(x)_i \;=\; k! \sum_{v \in X_n^k} \mathcal{F}'(v/k)_i \prod_{j<n} \frac{x_j^{v_j}}{v_j!}$$

where $k > 1$ is an integer parameter and $\mathcal{F}'$ is any ranking selection scheme. Observe that, as in the case of ranking selection, there is no dependence on population size. In practice, the tournament selection function is implemented as follows.

1. Uniformly select (with replacement) k members from the population.
2. Sort those elements by increasing fitness, obtaining $q_0 \leq q_1 \leq \cdots \leq q_{k-1}$.
3. Choose element q_i (with replacement) with probability t_i where t is some fixed probability vector of dimension k (i.e., $t_0 + \cdots + t_{k-1} = 1$) such that $i < j \Longrightarrow t_i < t_j$.

As for the case of ranking selection, only the procedural definition of tournament selection was available prior to this work. The correctness of the functional formulation provided by the tournament selection scheme will now be shown. The following identity, referred to as the *multinomial theorem* will be used.

$$(\mathbf{1}^T x)^k \;=\; k! \sum_{v \in X_n^k} \prod_{j<n} \frac{x_j^{v_j}}{v_j!}$$

In particular,

$$\mathbf{1}^T \mathcal{F}(x) \;=\; k! \sum_{v \in X_n^k} \int_0^1 \varrho(y)\, dy \prod_{j<n} \frac{x_j^{v_j}}{v_j!} \;=\; k! \sum_{v \in X_n^k} \prod_{j<n} \frac{x_j^{v_j}}{v_j!} \;=\; 1$$

The proof of the following lemma (the result will be applied later in chapter 4) will also be useful.

Lemma 3.2 *Let the next population result from r independent identically distributed random choices, where the probability of choosing i (with replacement) is w_i. Then the expected next population vector is w.*

Sketch of proof: The first step is to determine for each possible population vector the probability that it represents the next generation. Feasible populations are $\frac{1}{r} X_n^r$. To obtain a general population vector v/r, it must happen that v_0 choices out of r are 0, which has probability

$$\binom{r}{v_0} w_0^{v_0}$$

and v_1 choices out of the remaining $r - v_0$ must be 1, which has probability

$$\binom{r - v_0}{v_1} w_1^{v_1}$$

and so on until finally v_{n-1} choices out of the remaining $r - v_0 - \cdots - v_{n-2}$ must be $n-1$, which has probability

$$\binom{r - v_0 - \cdots - v_{n-2}}{v_{n-1}} w_{n-1}^{v_{n-1}}$$

The product of these probabilities reduces (after expanding the binomial coefficients) to

$$r! \prod_{j<n} \frac{w_j^{v_j}}{v_j!}$$

It follows that the expectation is given by

$$r! \sum_{v \in X_n^r} \frac{v}{r} \prod_{j<n} \frac{w_j^{v_j}}{v_j!}$$

Applying the operator $\sum e_i x_i \frac{\partial}{\partial x_i}$ to both sides of the multinomial theorem yields

$$k\, x\, (\mathbf{1}^T x)^{k-1} \;=\; k! \sum_i e_i\, x_i \sum_{v \in X_n^k} \frac{v_i x_i^{v_i - 1}}{v_i!} \prod_{j \neq i} \frac{x_j^{v_j}}{v_j!} \;=\; k! \sum_{v \in X_n^k} v \prod_j \frac{x_j^{v_j}}{v_j!}$$

Using this formula to simplify the expectation completes the proof. □

Note that the statement of lemma 3.2 is independent of r. It therefore holds independent of population size.

Observe that choosing among equally likely population members from (a population with population vector) p results in picking i with probability p_i. As was seen in the proof of lemma 3.2,

$$k! \prod_j \frac{p_j^{kq_j}}{(kq_j)!} \quad = \quad \Pr\{\text{ population } q \text{ results from uniformly choosing } k \text{ members of } p\ \}$$

Applying tournament selection to the result q, the probability of obtaining q and selecting i is

$$k!\, \mathcal{F}'(q)_i \prod_j \frac{p_j^{kq_j}}{(kq_j)!}$$

Summing over all possible $q = v/k$ yields

$$k! \sum_{v \in X_n^k} \mathcal{F}'(v/k)_i \prod_{j<n} \frac{x_j^{v_j}}{v_j!}$$

This establishes the following.

Theorem 3.3 *Applying the ranking selection function to the result of k uniform choices (with replacement) from the population implements the tournament selection function with parameter k.*

This section concludes with an alternate representation for the tournament selection scheme. Let ψ be a permutation of $<0, 1, \ldots, n-1>$ such that $i < j \implies f(\psi_i) < f(\psi_j)$. Observe that $\mathcal{F}'(v/k)_{\psi_i}$ depends only on $\{v_{\psi_j} : j \le i\}$. Thus it makes sense to define $\mathcal{F}'_\psi(<y_0, \ldots, y_i>)_i$ as $\mathcal{F}'(z)_{\psi_i}$ where z satisfies $0 \le j \le i \implies z_{\psi_j} = y_j$. Therefore

$$\begin{aligned}\mathcal{F}(x)_{\psi_i} &= k! \sum_{u=0}^{k} \sum_{w \in X_{i+1}^u} \mathcal{F}'_\psi(w/k)_i \sum_{v \in X_n^k} \prod_{j \le i} [v_{\psi_j} = w_j] \prod_{j<n} \frac{x_j^{v_j}}{v_j!} \\ &= k! \sum_{u=0}^{k} \sum_{w \in X_{i+1}^u} \mathcal{F}'_\psi(w/k)_i \prod_{j \le i} \frac{x_{\psi_j}^{w_j}}{w_j!} \sum_{v \in X_{n-i-1}^{k-u}} \prod_{j>i} \frac{x_{\psi_j}^{v_j}}{v_j!}\end{aligned}$$

Applying the multinomial theorem in the last sum and rearranging yields

$$\sum_{u=0}^{k} \binom{k}{u} (1 - \sum_{j \le i} x_{\psi_j})^{k-u} \, u! \sum_{w \in X_{i+1}^u} \mathcal{F}'_\psi(w/k)_i \prod_{j \le i} \frac{x_{\psi_j}^{w_j}}{w_j!}$$

Next suppose $w_{\psi_i} = v$ and $\sum_{j<i} w_{\psi_j} = u - v$. Then $\mathcal{F}'_\psi(w/k)_i = \sum_{j<v} t_{j+u-v}$ where the t_j are as defined in section 3.1 for the choice $\mathcal{F} = \mathcal{F}'$ and $r = k$. It follows that the sum above is

$$\begin{aligned}&\sum_{u=0}^{k} \binom{k}{u} (1 - \sum_{j \le i} x_{\psi_j})^{k-u} \sum_{v=0}^{u} u! \sum_{w \in X_i^{u-v}} \sum_{j<v} t_{j+u-v} \frac{x_{\psi_i}^v}{v!} \prod_{j<i} \frac{x_{\psi_j}^{w_j}}{w_j!} \\ = &\sum_{u=0}^{k} \binom{k}{u} (1 - \sum_{j \le i} x_{\psi_j})^{k-u} \sum_{v=0}^{u} \binom{u}{v} x_{\psi_i}^v \sum_{j<v} t_{j+u-v} \, (u-v)! \sum_{w \in X_i^{u-v}} \prod_{j<i} \frac{x_{\psi_j}^{w_j}}{w_j!}\end{aligned}$$

Applying the multinomial theorem in the last sum and rearranging establishes the following.

Theorem 3.4 *The tournament selection scheme satisfies*

$$\mathcal{F}(x)_{\psi_i} = \sum_{u=0}^{k} \binom{k}{u} (1 - \sum_{j \le i} x_{\psi_j})^{k-u} \sum_{v=0}^{u} \binom{u}{v} x_{\psi_i}^v \, (\sum_{j<i} x_{\psi_j})^{u-v} \sum_{j<v} t_{j+u-v}$$

where the t_j are as defined in section 3.1 for the choice $\mathcal{F} = \mathcal{F}'$ and $r = k$.

4 Basic Properties

Now that ranking and tournament selection have been formalized in terms of mathematical functions, some of their basic properties will be considered. First note that if p represents the current population, then by lemma 3.2 the expected gene pool is $\mathcal{F}(p)$ for any selection scheme and any population size.

It follows that the expected gene pool is a continuously differentiable function of the current population (the tournament selection scheme is polynomial, and the ranking selection scheme has continuous partials by the continuity of ϱ).

Finally, it will be shown that the ranking and tournament selection schemes are diffeomorphisms of Λ by indicating how to compute their inverses. Consider first the problem of recovering $x \in \Lambda$ from $\mathcal{F}(x)$.

For ranking selection, the i th component of $\mathcal{F}(x)$ is

$$\int_{\eta_i}^{\eta_i+x_i} \varrho(y)\, dy$$

where $\eta_{\psi_i} = \sum_{j<i} x_{\psi_j}$. Since ϱ is positive, this is an increasing function of x_i for any η_i. Hence the zero of

$$h_i(z) \;=\; \int_{\eta_i}^{\eta_i+z} \varrho(y)\, dy \;-\; \mathcal{F}(x)_i$$

is at $z = x_i$ and can be found numerically provided η_i is known. Because ϱ is increasing, h_i has positive second derivative and is concave up with root in the interval $[0, 1-\eta_i]$. Using Newton's method, an acceptable initial guess is $z = 1 - \eta_i$, and an improved guess z' can be obtained by

$$z' \;=\; z - \frac{h_i(z)}{\varrho(\eta_i + z)}$$

For this method, convergence is assured, and the components of x may be solved in the order $x_{\psi_0}, x_{\psi_1}, \ldots$ since then the η_i are known:

$$\begin{aligned} \eta_{\psi_0} &= 0 \\ \eta_{\psi_{i+1}} &= \eta_{\psi_i} + x_{\psi_i} \end{aligned}$$

The case of tournament selection is similar. To simplify notation, let x_{ψ_i} be abbreviated by z, and let $\sum_{j<v} t_{j+u-v}$ be abbreviated by $w(u,v)$. Using the representation provided by theorem 3.4, the partial of $\mathcal{F}(x)_{\psi_i}$ with respect to z is

$$\sum_{u=0}^{k}\sum_{v=0}^{u} \frac{k!\,\eta_{\psi_i}^{u-v} w(u,v)}{(k-u)!\,(u-v)!\,v!} \left\{(1-z-\eta_{\psi_i})^{k-u} v z^{v-1} - (k-u)(1-z-\eta_{\psi_i})^{k-u-1} z^{v}\right\}$$

$$= \sum_{u=0}^{k}\sum_{v=1}^{u} \frac{k!\,\eta_{\psi_i}^{u-v} w(u,v)(1-z-\eta_{\psi_i})^{k-u} z^{v-1}}{(k-u)!\,(u-v)!\,(v-1)!} - \sum_{u=0}^{k-1}\sum_{v=0}^{u} \frac{k!\,\eta_{\psi_i}^{u-v} w(u,v)(1-z-\eta_{\psi_i})^{k-u-1} z^{v}}{(k-u-1)!\,(u-v)!\,v!}$$

Making the change of variables $u = u' + 1$, $v = v'$ in the first sum (of the right hand side above), and $u = u'$, $v = v' - 1$ in the second sum, and then recombining them yields

$$\sum_{u'=0}^{k-1}\sum_{v'=1}^{u'+1} \frac{k!\,\eta_{\psi_i}^{u'-v'+1}(1-z-\eta_{\psi_i})^{k-u'-1} z^{v'-1}}{(k-u'-1)!\,(u'-v'+1)!\,(v'-1)!} \left\{w(u'+1, v') - w(u', v'-1)\right\}$$

$$= \sum_{u=0}^{k-1} \sum_{v=1}^{u+1} \frac{k!\, \eta_{\psi_i}^{u-v+1} (1-z-\eta_{\psi_i})^{k-u-1} z^{v-1}}{(k-u-1)!\,(u-v+1)!\,(v-1)!} t_u$$

$$= \sum_{u=0}^{k-1} \binom{k}{u} t_u\,(k-u)\,(1-z-\eta_{\psi_i})^{k-u-1} \sum_{v=1}^{u+1} \binom{u}{v-1} z^{v-1}\, \eta_{\psi_i}^{u-v+1}$$

$$= \sum_{u=0}^{k-1} \binom{k}{u} t_u\,(k-u)\,(1-z-\eta_{\psi_i})^{k-u-1} (z+\eta_{\psi_i})^u$$

The second derivative is computed similarly,

$$\frac{\partial^2}{\partial z^2} \mathcal{F}(x)_{\psi_i}$$

$$= \sum_{u=0}^{k-1} \binom{k}{u} t_u\,(k-u) \Big\{ (1-z-\eta_{\psi_i})^{k-u-1} u (z+\eta_{\psi_i})^{u-1}$$

$$- (k-u-1)(1-z-\eta_{\psi_i})^{k-u-2} (z+\eta_{\psi_i})^u \Big\}$$

$$= \sum_{u=1}^{k-1} \frac{k!\,(1-z-\eta_{\psi_i})^{k-u-1}\,(z+\eta_{\psi_i})^{u-1}}{(u-1)!\,(k-u-1)!} t_u - \sum_{u=0}^{k-2} \frac{k!\,(1-z-\eta_{\psi_i})^{k-u-2}\,(z+\eta_{\psi_i})^{u}}{u!\,(k-u-2)!} t_u$$

Making the change of variables $u = u' + 1$ in the first sum (of the right hand side above), replacing u' with u and then recombining the sums yields

$$\sum_{u=0}^{k-2} \frac{k!\,(1-z-\eta_{\psi_i})^{k-u-2}\,(z+\eta_{\psi_i})^{u}}{u!\,(k-u-2)!} \{t_{u+1} - t_u\}$$

Note that the first derivative is positive when $0 < z + \eta_{\psi_i} < 1$, and since ϱ is increasing (i.e., $t_{u+1} - t_u > 0$) the second derivative is also positive. As in the case of ranking selection, the zero of the function

$$h_i(z) \;=\; \sum_{u=0}^{k} \binom{k}{u} (1-\eta_i-z)^{k-u} \sum_{v=0}^{u} \binom{u}{v} z^v\, \eta_i^{u-v} \sum_{j<v} t_{j+u-v} \;-\; \mathcal{F}(x)_i$$

can be found numerically provided η_i is known. Like before, h_i is increasing and concave up with root $z = x_i$ in the interval $[0, 1-\eta_i]$. Using Newton's method, an acceptable initial guess is $z = 1 - \eta_i$, and an improved guess z' can be obtained by

$$z' = z - h_i(z) \,/\, \sum_{u=0}^{k-1} \binom{k}{u} t_u\,(k-u)\,(1-z-\eta_i)^{k-u-1} (z+\eta_i)^u$$

For this method, convergence is assured, and the components of x may be solved in the order $x_{\psi_0}, x_{\psi_1}, \ldots$ since then the η_i are known as before.

Theorem 4.1 $\mathcal{F}^{-1} : \Lambda \to \Lambda$

Sketch of proof: In the case of ranking selection, any $v \in \Lambda$ satisfies $v_i \geq 0$ and $\mathbf{1}^T v = 1$. Since ϱ is a probability density on $[0,1]$, there exists corresponding ξ_i such that $0 = \xi_0 \leq \xi_1 \leq \cdots \leq \xi_n = 1$ and

$$\int_{\xi_i}^{\xi_{i+1}} \varrho(y)\, dy = v_i$$

It follows that the equations which are solved numerically to compute $\mathcal{F}^{-1}(v)$ have non-negative solutions which sum to 1 (the root x_{ψ_i} of h_{ψ_i} is $\xi_{i+1} - \xi_i$).

For tournament selection, that the equations (with variable z)

$$v_i = \sum_{u=0}^{k} \binom{k}{u} (1 - \eta_i - z)^{k-u} \sum_{v=0}^{u} \binom{u}{v} z^v \eta_i^{u-v} \sum_{j<v} t_{j+u-v}$$

have solutions x_i follows from the fact that the right hand side varies from 0 to $1 - \sum_{j<i} v_j$ for $z \in [0, 1-\eta_i]$. To see this, induct on i. First note that the case $z = 0$ is trivial. Next regard v_i as being defined by the right hand side for the choice of $z = x_i$ and set $x_j = 0$ for $j > h$. To establish the claim at $i = h$, let $z = x_h = 1 - \eta_h$ so that what needs to be proved is $\mathbf{1}^T v = 1$ (since $v_j = 0$ for $j > h$ by choice of x_j). It follows from the multinomial theorem (applied to the first representation for tournament selection) that

$$\mathbf{1}^T v = \mathbf{1}^T \mathcal{F}(x) = 1$$

□

The continuous differentiability of $\mathcal{F}^{-1}$ in the interior of Λ follows from the inverse function theorem provided that the differential of $\mathcal{F}$ is nonsingular. For tournament selection, theorem 3.4 shows $\mathcal{F}(x)_{\psi_i}$ to depend only on $\{x_{\psi_j} : j \leq i\}$, which is also true of ranking selection (as noted in the proof of theorem 3.4). Hence the differential is triangular in a suitable basis with general diagonal element given by the partial of $\mathcal{F}(x)_{\psi_i}$ with respect to x_{ψ_i}. By the discussion preceding theorem 4.1, these partials are all positive, hence the differential has positive determinant.

According to theorems 3.1 and 3.3, the faces of Λ (of any dimension) are invariant under $\mathcal{F}$. Moreover, since a face corresponds to element(s) of Ω being excluded from the population, a face is literally the "Λ" corresponding to an "Ω" for some smaller value of n (i.e., a face is the representation space corresponding to a subset of the search space). When considering the differentiability of $\mathcal{F}$ in a face of Λ, it therefore makes most sense to consider $\mathcal{F}$ as having that face as its domain. By what has already been shown, $\mathcal{F}$ is a diffeomorphism in the interior of that face.

Applying the previous paragraph to all faces of Λ (of all dimensions) completes both the definition (of what precisely – in this paper – is meant by) and proof of the result that $\mathcal{F}$ is a diffeomorphism.

5 Conclusion

This paper began with a representational framework for a general type of heuristic search of which genetic algorithms are a special case. Using this framework, the ranking and tournament selection schemes were defined and formalized as mathematical functions.

The main result is that ranking and tournament selection are diffeomorphisms of the representation space. Explicit algorithms were also developed for computing their inverses.

In the case of a binary GA with classical crossover and mutation, it is known that the function $\mathcal{G} : \Lambda \to \Lambda$ which maps the current population p to the expected next generation $\mathcal{G}(p)$ has the form $\mathcal{G} = \mathcal{M} \circ \mathcal{F}$ where $\mathcal{M}$ is invertible [2]. Hence this paper establishes the invertibility of $\mathcal{G}$ when either ranking or tournament selection is used.

6 Acknowledgements

This research was supported by the National Science Foundation: IRI-9224917.

7 Reference

[1] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley.

[2] Vose, M. D. (in press). *The Simple Genetic Algorithm: foundations and theory.* Cambridge, MA: MIT press.