

## Part 1: Populations and Probabilities

Given an initial genome population:

$$G = \{n_g \geq x \geq 1 | g_1, g_2, g_3, \dots, g_x, \dots, g_{n_g-1}, g_{n_g}\}$$

Where  $g_x$  is the  $x^{\text{th}}$  genome in the population G, and  $n_g$  is the number of genomes present in the population.

We define:

$$M = \{n_m \geq x \geq 1 | m_1, m_2, \dots, m_x, \dots, m_{n_m-1}, m_{n_m}\}$$

Where  $m_x$  is the  $x^{\text{th}}$  genome in the population of mutated offspring M of parents G, and  $n_m$  is the number of offspring produced by G and included in M.

We further define  $n_e$  as the number of individuals expected to survive from M and produce further offspring, and  $P_s(m_x)$  as the chance of survival of individual  $m_x$  undergoing the selective pressure that will reduce the population M from  $n_m$  to  $n_e$ .

The sum total of all the probabilities for M will thus equal  $n_e$ , or:

$$\sum P_s(M) = n_e$$

And the average chance of survival, (or the chance of survival for an individual with an average genome,) will thus be:

$$P_s(m_{\text{avg}}) = \frac{\sum P_s(M)}{n_m} = \frac{n_e}{n_m}$$

$$\boxed{P_s(m_{\text{avg}}) = \frac{n_e}{n_m}}$$

## Part 2: Definition of Fitness

The fitness of an individual  $F(m_x)$  is then defined such that:

$$\frac{F(m_a)}{F(m_b)} = \frac{P_s(m_a)}{P_s(m_b)}$$

In this case, an individual's fitness will have a direct relation to its chance of survival relative to the other individuals in its population.

If, then, we define:

$$F(m_a) \equiv \alpha_m P_s(m_a)$$

we obtain the following:

$$\frac{F(m_a)}{F(m_b)} = \frac{\alpha_m P_s(m_a)}{\alpha_m P_s(m_b)}$$

which is equivalent to our original definition, for  $\alpha_m$ , a positive scalar.

Here  $\alpha_m$  is arbitrary in relation to the individual's probability, but not when individuals are compared against each other; the scalar  $\alpha_m$  must be consistent throughout the population, but need not be consistent across generations.

Since:

$$P_s(m_{avg}) = \frac{n_e}{n_m}$$

then:

$$\alpha_m P_s(m_{avg}) = \alpha_m \frac{n_e}{n_m}$$

$$F(m_{avg}) = \alpha_m \frac{n_e}{n_m}$$

$$\alpha_m = F(m_{avg}) \frac{n_m}{n_e}$$

And since:

$$F(m_{avg}) = \frac{\sum F(M)}{n_m}$$

then:

$$\alpha_m = \frac{\sum F(M) \frac{n_m}{n_e}}{n_m}$$

$$\alpha_m = \frac{\sum F(M)}{n_e}$$

So the scalar  $\alpha_m$  is dependent on the total fitness and the expected number of survivors alone. Moreover, we can now determine an individual's chance of success using the following:

$$F(m_x) = \alpha_m P_s(m_x)$$

$$P_s(m_x) = \frac{F(m_x)}{\alpha_m}$$

$$P_s(m_x) = \frac{F(m_x)}{\sum F(M)} n_e$$

An individual's survival, then, is not so much dependent on the total number of individuals to choose from as the total fitness of those individuals. (Of course, more individuals means more total fitness, so the values are still related.) The number of expected survivors, then, acts as an indicator of selective pressure – the lower  $n_e$  is set to, the higher the selective pressure must be to bring about this change.

An individual's chance of success can also be determined as follows:

$$F(m_x) = \alpha_m P_s(m_x)$$

$$F(m_x) = F(m_{avg}) \frac{n_m}{n_e} P_s(m_x)$$

$$P_s(m_x) = \frac{F(m_x) \frac{n_e}{n_m}}{F(m_{avg})}$$

Which is an individual's fitness compared to the average fitness, times the average survival rate. This is, perhaps, the more intuitive of the two formulas.

*Part 3: Limits to  $\alpha_m$  and  $n_e$*

The limiting factor to  $n_e$ , and, consequently  $\alpha_m$  is the probability of survival of an individual with the maximum fitness value in the population ( $P_s(m_{max})$ ).

Since no probability of survival can be above 1.0 (100%), we have:

$$F(m_{max}) = \alpha_m P_s(m_{max})$$

$$\frac{F(m_{max})}{\alpha_m} = P_s(m_{max}) \leq 1$$

$$\frac{F(m_{max})}{\alpha_m} \leq 1$$

$$\boxed{\alpha_m \geq F(m_{max})}$$

And hence,

$$F(m_{max}) \leq \alpha_m = \frac{\sum F(M)}{n_e}$$

$$F(m_{max}) \leq \frac{\sum F(M)}{n_e}$$

$$\boxed{n_e \leq \frac{\sum F(M)}{F(m_{max})}}$$

or,

$$\boxed{n_e \leq n_m \frac{F(m_{avg})}{F(m_{max})}}$$

A fitness scalar must therefore be at least as large as the maximum fitness value that is present in the population. Likewise, the number of expected survivors cannot exceed the total fitness divided by the maximum fitness value included in its population.

#### Part 4: Evolutionary Advancement

For evolutionary advancement to occur, the average fitness of the population must increase from generation to generation:

$$\boxed{F(m'_{avg}) > F(g_{avg})}$$

**Evolutionary Advancement**

Where  $m'_{avg}$  is an average surviving child, and  $g_{avg}$  is an average parent.

This can also be defined as follows:

$$\boxed{\frac{\sum F(M')}{n_{m'}} > \frac{\sum F(G)}{n_g}}$$

**Evolutionary Advancement**

Where  $n_{m'}$  is the number of children that survive to reproduce and M' is the actual child population.

There are four main ways that this increase in the average fitness can be brought about. These include:

1. Positive mutations can occur.
2. Better fit parents can produce multiple children and these children can out-survive their neighbors.
3. The size of the population can change.
4. Environmental conditions can change, altering or shifting the fitness values of individuals relative to one another.

The first method occurs whenever a positive mutation is gained. For example, given a parent population:

$$F(G) = \{22, 20, 18, 17\}, F(g_{avg}) = 19.25$$

If  $g_1$  mutates such that  $F(m'_1) = 23$ , we have:

$$F(M') = \{23, 20, 18, 17\}, F(m'_{avg}) = 19.50$$

An average increase in the fitness of the population.

The second method occurs when a parent has two neutral children:

$$F(G) = \{22, 20, 18, 17\}, F(g_{avg}) = 19.25$$

$$F(M) = \{22, 22, 20, 18, 17\}, n_e = 4$$

$$F(M') = \{22, 22, 20, 18\}, F(m'_{avg}) = 21.25$$

An average increase in the fitness of the population.

The third method occurs when the population size changes:

$$F(G) = \{22, 20, 18, 17\}, F(g_{avg}) = 19.25$$

$$F(M) = \{22, 20, 18, 17\}, n_e = 2$$

$$F(M') = \{22, 20\}, F(m'_{avg}) = 21.00$$

An average increase in the fitness of the population.

A combination of these three methods is also possible:

$$F(G) = \{22, 20, 18, 17\}, F(g_{avg}) = 19.25$$

$$F(M) = \{23, 22, 20, 18, 17\}, n_e = 2$$

$$F(M') = \{23, 22\}, F(m'_{avg}) = 22.50$$

An average increase in the fitness of the population.

The fourth method involves changes in the conditions that underly the individual fitness values. In general, it would work in the following way:

$$F(G) = \{22, 20, 18, 17\}, F(g_{avg}) = 19.25$$

$$F(M) = \{22, 20, 18, 17\}, F(m_{avg}) = 19.25$$

- Environmental Change -

$$F(M') = \{29, 14, 16, 21\}, F(m'_{avg}) = 20.00$$

An average increase in the fitness of the population.

This change comes about wholly according to the environmental changes. (The genome remained the same as it was before, only the fitness levels changed due to the increase in the population's ability to survive under the new conditions.)

For novel advancement to occur, (advancement that does not come about only by dominance of latent genetic information,) the first of the four methods *must* occur. In point of fact, the first of the methods must be the overall driving force of evolution if it is to produce new and more complex structures. While the second and third methods do increase the average fitness, they do so merely by homogenizing the population towards the higher genomes already present in the population. They can be used to aid in the preservation of the more fit genomes,

but they cannot drive the population to higher and more advanced genomes.

The fourth method can potentially increase the average, but this increase could only be due to either pure chance (which is unlikely, or at least no more likely than the first method could be), or by genetic features that had already been developed in previous environments (in which case the real advancement would have come earlier from the first method). For actual advancement from lower to higher forms, the actual genome must be changed – this fact is unavoidable.

## Part 5: Natural Selection

The total final expected fitness of a population of genomes after selection will depend on the fitness values of the genomes that make up the population. In general, the expected fitness contribution per individual will be that individual's fitness times its probability of survival. The total final expected fitness of a population of genomes after selection will thus be:

$$\sum F_e(M') = \sum (F(m_x)P_s(m_x))$$

$$\sum F_e(M') = \frac{\sum (F(m_x)^2)}{\alpha_m}$$

$$\sum F_e(M') = \frac{\sum (F(m_x)^2)}{\sum F(M)} n_e$$

$$F_e(m'_{avg}) = \frac{\sum (F(m_x)^2)}{\sum F(M)}$$

$$\boxed{F_e(m'_{avg}) = \frac{\sum (F(M)^2)}{\sum F(M)}}$$

Given that the standard variance of the population M is:

$$\sigma_m^2 = \frac{1}{n_m} \left( \sum_{i=1}^{n_m} (F(m_i) - F(m_{avg}))^2 \right)$$

$$\sigma_m^2 = \frac{1}{n_m} \left( \sum_{i=1}^{n_m} (F(m_i)^2 + F(m_{avg})(F(m_{avg}) - 2F(m_i))) \right)$$

$$\sigma_m^2 = \frac{1}{n_m} \left( \sum (F(M)^2) + F(m_{avg}) \sum_{i=1}^{n_m} (F(m_{avg}) - 2F(m_i)) \right)$$

$$\sigma_m^2 = \frac{1}{n_m} \left( \sum (F(M)^2) + F(m_{avg})(F(m_{avg})n_m - 2 \sum_{i=1}^{n_m} F(m_i)) \right)$$

$$\sigma_m^2 = \frac{1}{n_m} \left( \sum (F(M)^2) + F(m_{avg})(\sum F(M) - 2 \sum F(M)) \right)$$

$$\sigma_m^2 = \frac{1}{n_m} \left( \sum (F(M)^2) - F(m_{avg}) \sum F(M) \right)$$



$$\sigma_m^2 = \frac{1}{n_m} (\sum (F(M)^2) - n_m F(m_{avg})^2)$$

$$\sigma_m^2 = \frac{\sum (F(M)^2)}{n_m} - F(m_{avg})^2$$

Which gives us:

$$\sigma_m^2 + F(m_{avg})^2 = \frac{\sum (F(M)^2)}{n_m}$$

$$\sum (F(M)^2) = n_m (\sigma_m^2 + F(m_{avg})^2)$$

Substituting in our original equation, we get:

$$F_e(m'_{avg}) = \frac{n_m (\sigma_m^2 + F(m_{avg})^2)}{\sum F(M)}$$

$$F_e(m'_{avg}) = \frac{n_m (\sigma_m^2 + F(m_{avg})^2)}{n_m F(m_{avg})}$$

$$F_e(m'_{avg}) = F(m_{avg}) + \frac{\sigma_m^2}{F(m_{avg})}$$

Thus, the expected average fitness after selection depends on the average fitness before selection and the standard variance from that average (the standard variance before selection).

Selection according to fitness, then, (i.e. our model of Natural Selection,) will raise the average expected fitness by  $\sigma_m^2 / F(m_{avg})$ , the standard variance over the average fitness of the group that the selection is acting upon. Since the standard variance is always positive, we must have:

$$F_e(m'_{avg}) \geq F(m_{avg})$$

However, selection can be of no advantage where there is no variation in the population. If  $\sigma_m^2 = 0$  there will be no expected change in the average fitness of the population after selection takes place.

## Part 6: Maximum Variance

To determine the maximum theoretical standard variance from an average value specified for a value-bounded set, the squares must be maximized. If the range of values is from  $p_a$  to  $p_b$ , and the average value is  $m$ , the maximum variance can be calculated fairly simply.

The largest variances possible will be found where the largest possible gaps are. These are from  $p_a$  to  $m = m - p_a$ , and from  $m$  to  $p_b = p_b - m$ . The values attained from each of these gaps must contribute to an average value of  $m$ . To do this, we require:

$$m = \frac{q_a p_a + q_b p_b}{q_a + q_b}$$

Where  $q_a$  is the quantity of  $p_a$  and  $q_b$  is the quantity of  $p_b$ .

The standard variance will then be:

$$\begin{aligned}\sigma_{max}^2 &= \frac{q_a(m - p_a)^2 + q_b(p_b - m)^2}{q_a + q_b} \\ \sigma_{max}^2 &= \frac{q_a(m^2 + p_a^2 - 2m p_a) + q_b(p_b^2 + m^2 - 2m p_b)}{q_a + q_b} \\ \sigma_{max}^2 &= \frac{(q_a + q_b)m^2}{(q_a + q_b)} + \frac{q_a p_a^2 + q_b p_b^2}{(q_a + q_b)} - 2m \frac{(q_a p_a + q_b p_b)}{q_a + q_b}\end{aligned}$$

Simplifying the first term in this equation and substituting the previous equation into it, we obtain:

$$\sigma_{max}^2 = m^2 + \frac{q_a p_a^2 + q_b p_b^2}{(q_a + q_b)} - 2m^2$$

$$\boxed{\sigma_{max}^2 = \frac{q_a p_a^2 + q_b p_b^2}{(q_a + q_b)} - m^2}$$

Again, from the first equation we can directly derive:

$$q_a p_a = m(q_a + q_b) - q_b p_b$$

And its counterpart:

$$q_b p_b = m(q_a + q_b) - q_a p_a$$

Substituting these into our working equation, we have:

$$\sigma_{max}^2 = \frac{p_a(m(q_a+q_b)-q_b p_b) + p_b(m(q_a+q_b)-q_a p_a)}{(q_a+q_b)} - m^2$$

$$\sigma_{max}^2 = \frac{m(q_a+q_b)(p_a+p_b)}{(q_a+q_b)} - \frac{(q_a+q_b)(p_a p_b)}{(q_a+q_b)} - m^2$$

$$\sigma_{max}^2 = m(p_a+p_b) - p_a p_b - m^2$$

$$\sigma_{max}^2 = m(p_b-m) - p_a(p_b-m)$$

$$\boxed{\sigma_{max}^2 = (m-p_a)(p_b-m)}$$

Which is simply the product of the two maximum gaps.

The maximum standard variance for the population M, where  $F(m_{avg})$  is the average fitness of the population and the maximum fitness is  $F(m)_{max}$  will therefore be:

$$\boxed{\sigma_{max}^2 = F(m_{avg})(F(m)_{max} - F(m_{avg}))}$$

Therefore, the maximum possible average expected value after natural selection must always be:

$$F_e(m'_{avg})_{max} = F(m_{avg}) + \frac{\sigma_{max}^2}{F(m_{avg})}$$

$$F_e(m'_{avg})_{max} = F(m_{avg}) + \frac{F(m_{avg})(F(m)_{max} - F(m_{avg}))}{F(m_{avg})}$$

$$F_e(m'_{avg})_{max} = F(m_{avg}) + F(m)_{max} - F(m_{avg})$$

$$\boxed{F_e(m'_{avg})_{max} = F(m)_{max}}$$

Which should be expected, and therefore verifies the above equations.

## Part 7: Expected Fitness Change After Mutation

Given a initial genome sequence in which the following conditions are met:

- a) The genome sequence can be split into  $w$  elements, each having  $k$  possible permutations.
- b) The mutation rate  $r$  per element is consistent across the genome.
- c) A single mutation can transform any given element into any of the other possible element permutations.

If  $F(h_x)$  is the fitness contribution of an individual element, and if  $F_e(h)_{rnd}$  is the expected fitness for a random element, (the average combined fitness of all of the permutations possible for an element,) we have:

$$F_e(g)_{rnd} = w \cdot F_e(h)_{rnd}$$

That is, the expected fitness for a random genome is equal to the expected fitness for a random element times the number of elements in the genome.

Now, for a given element,  $F(h_x)$ , the expected fitness for a random element can be expressed as follows:

$$F_e(h)_{rnd} = \frac{1}{k} F(h_x) + \sum_{i=1, i \neq x}^k \frac{1}{k} F(h_i)$$

$$F_e(h)_{rnd} = \frac{1}{k} (F(h_x) + \sum_{i=1, i \neq x}^k F(h_i))$$

Therefore,

$$\sum_{i=1, i \neq x}^k F(h_i) = k \cdot F_e(h)_{rnd} - F(h_x)$$

Now on mutation, for a true mutation to occur, the element must change from what it was before into one of the other possible element permutations. Thus, the expected fitness after mutation ( $F_e(t_x)_{mut}$ ) must be:

$$F_e(t_x)_{mut} = \sum_{i=1, i \neq x}^k \frac{F(h_i)}{(k-1)}$$

$$F_e(t_x)_{mut} = \frac{1}{(k-1)} \sum_{i=1, i \neq x}^k F(h_i)$$

And, from what we discovered above, this means:

$$F_e(t_x)_{mut} = \frac{k \cdot F_e(h)_{rnd} - F(h_x)}{(k-1)}$$

The fitness change per element can then be calculated fairly easily:

$$F_e(t_x)_{mut} = F(h_x) + \Delta F_e(t_x)_{mut}$$

$$\Delta F_e(t_x)_{mut} = F_e(t_x)_{mut} - F(h_x)$$

$$\Delta F_e(t_x)_{mut} = \frac{k \cdot F_e(h)_{rnd} - F(h_x)}{(k-1)} - F(h_x)$$

$$\Delta F_e(t_x)_{mut} = \frac{k \cdot F_e(h)_{rnd} - F(h_x) - (k-1) \cdot F(h_x)}{(k-1)}$$

$$\Delta F_e(t_x)_{mut} = \frac{k \cdot F_e(h)_{rnd} - k \cdot F(h_x)}{(k-1)}$$

$$\Delta F_e(t_x)_{mut} = \frac{k}{(k-1)} (F_e(h)_{rnd} - F(h_x))$$

Given that  $r$  is the mutation rate per element, the expected fitness change per element for the next generation will then be:

$$\Delta F_e(t_x) = r \cdot \Delta F_e(t_x)_{mut}$$

$$\Delta F_e(t_x) = \frac{r k}{(k-1)} (F_e(h)_{rnd} - F(h_x))$$

Summing over the entire genome, we have:

$$\Delta F_e(m_x) = \sum_{i=1}^w \left( \frac{r k}{(k-1)} (F_e(h)_{rnd} - F(h_x)) \right)$$

$$\Delta F_e(m_x) = \frac{r k}{(k-1)} \sum_{i=1}^w (F_e(h)_{rnd} - F(h_x))$$

$$\Delta F_e(m_x) = \frac{r k}{(k-1)} \left( \sum_{i=1}^w F_e(h)_{rnd} - \sum_{i=1}^w F(h_x) \right)$$

$$\Delta F_e(m_x) = \frac{r k}{(k-1)} (w \cdot F_e(h)_{rnd} - F(g_x))$$

$$\Delta F_e(m_x) = \frac{r k}{(k-1)} (F_e(g)_{rnd} - F(g_x))$$

To get the average expected change over the entire population, we simply take the sum of the changes in the population, and divide by the population size.

Since our calculations assume a fixed number of children per parent, we can calculate the expected average according to the parents' size as follows:

$$\Delta F_e(m_{avg}) = \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{r k}{(k-1)} (F_e(g)_{rnd} - F(g_x))$$

$$\Delta F_e(m_{avg}) = \frac{r k}{n_g (k-1)} \sum_{i=1}^{n_g} (F_e(g)_{rnd} - F(g_x))$$

$$\Delta F_e(m_{avg}) = \frac{n_g \cdot r \cdot k}{n_g (k-1)} F_e(g)_{rnd} - \frac{r k}{n_g (k-1)} \sum_{i=1}^{n_g} F(g_x)$$

$$\Delta F_e(m_{avg}) = \frac{r k}{(k-1)} F_e(g)_{rnd} - \frac{n_g \cdot r \cdot k}{n_g (k-1)} F(g_{avg})$$

$$\Delta F_e(m_{avg}) = \frac{r k}{(k-1)} (F_e(g)_{rnd} - F(g_{avg}))$$

The function that results is linear, with a maximum value of:

$$\frac{r k}{(k-1)} F_e(g)_{rnd} \quad \text{for} \quad F(g_{avg}) = 0$$

And a minimum value of:

$$\frac{r k}{(k-1)} (F_e(g)_{rnd} - F(g)_{max}) \quad \text{for} \quad F(g_{avg}) = F(g)_{max}$$

A value of zero is attained when  $F(g_{avg}) = F_e(g)_{rnd}$ .

As long as  $F(g_{avg})$  is less than the expected random value, the fitness can be expected to increase. After a fitness value above the expected random value is obtained, natural selection must be utilized in order to overcome the negative trend.

