

Infinite Population GA Tutorial

Copyright © Michael D. Vose
The University of Tennessee, Knoxville
September 2004
All rights reserved.

Introduction

Judging from comments, questions, talks, and papers over the past few years, there is some degree of confusion regarding what relevance—if any—infinite population GAs have to finite population GAs. This tutorial should either clear up some confusion (if you may have been confused), or else review what you already knew (if you had it all straight to begin with).

What *is* an infinite population GA? Is it merely a thought experiment, or can one actually compute with an infinite number of population members?

To begin with, I would like you to adopt a certain mind set, which is basically to let go of superstitions about the infinite. Consider, as an analogy of sorts, the Riemann integral. In a sense that can be made precise, a Riemann integral of a continuous function over a closed interval is the infinite sum of infinitesimally small things... so, yes, it evidently *is* the case that computing with an infinite number of things is possible! One might suspect this analogy on the grounds that the infinite is balanced by the infinitesimal.

In that case, however, consider the infinite sum

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$$

Does a new objection arise? Even if you question this new analogy—because the infinite sum is being *defined* as a limit—your mind set will do. Whereas the infinite population GA is *not* defined as a limit, it *is* shown to equal a limit of finite population GAs as the population size goes to infinity.

A Finite Population GA

You probably know what a finite GA is, and have C code to prove it! But your Turing machine is not necessarily the most perspicuous description possible. And if we would like our opinions concerning GAs to be certainties, then there are genuine advantages to expressing our GA using nice functions (differentiable, perhaps?) over nice domains (compact subsets of Euclidean space, perhaps?).

With an eye toward making it easy to take the limit as the population size goes to infinity, it might be particularly convenient if certain limits involving populations were trivial. An example of a really trivial limit is

$$\lim_{n \rightarrow \infty} P_n$$

where $P_1 = P_2 = \dots$. Fortunately, there is a simple way to represent populations which has some of these desirable properties.

Example 1

To begin with, let r denote the population size, and denote the search space by Ω . For simplicity, let Ω be the set of length ℓ binary strings. For instance, if $\ell = 3$ and $r = 7$ then a population could be

$$P = \{010, 000, 110, 010, 011, 000, 010\}$$

Let this population be represented by the following “population vector” in \mathfrak{R}^n , where $n = |\Omega|$ (by convention, vectors delimited by angle brackets are interpreted as column vectors),

$$p = \left\langle \frac{2}{7}, \frac{0}{7}, \frac{3}{7}, \frac{0}{7}, \frac{0}{7}, \frac{0}{7}, \frac{1}{7}, \frac{0}{7} \right\rangle$$

In general, p has n components—one for each element of Ω —and p_i is the proportion of i contained in the population. When $\Omega = \{0, \dots, n-1\}$ this makes sense, and without loss of generality that is the case. Note that points in Λ with rational coordinates represent finite populations.

What pleasant properties are associated with this representation? First, note that the set of population vectors, including their limits, is the compact set

$$\Lambda = \{ \langle x_0, \dots, x_{n-1} \rangle : x_i \geq 0, \sum x_i = 1 \}$$

Second, the set Λ has an alternate interpretation. If d is a probability distribution over Ω , it can be regarded as a vector such that d_i is the probability of i . Note that $d \in \Lambda$. Hence elements of Λ can be interpreted as populations or as probability distributions over Ω .

Third, note that the following population (of size $r = 14$) is also represented by the same population vector p as displayed above.

$$P' = \{010, 010, 000, 000, 110, 110, 010, 010, 011, 011, 000, 000, 010, 010\}$$

This is because P' contains the same proportion of elements as does P . More generally, let P_n be the population of size $r_n = 7n$ containing $r_n p_i$ copies of i . Then each population P_n is represented by the population vector p (the proportion of i contained in P_n is $r_n p_i / r_n = p_i$). Moreover, if by “the limit as $r_n \rightarrow \infty$ of P_n ” we mean that same limit but of the population vectors representing the P_n , then the limit is really trivial!

Having settled on a convenient representation for populations, the next step is to describe how the next generation is produced from the current population.

Before proceeding, it should be remarked that this tutorial concerns the *simple GA* (described in the book “The Simple Genetic Algorithm: Foundations and Theory” by Michael D. Vose). It is a generational, John Holland style genetic algorithm, conforming to the following scheme:

1. Let P be some initial population containing r binary strings of length ℓ .
2. Choose (with replacement) parents u and v from the current population P (by any selection scheme).
 - Cross u and v (by any collection of crossover masks and probabilities) to produce children u' and v' .
 - Mutate u' and v' (by any collection of mutation masks and probabilities) to produce u'' and v'' .
 - Keep, with uniform probability, one of u'' and v'' for the next generation.
3. If the next generation contains fewer than r members, repeat step 2.
4. Replace P by the new generation just formed and go to step 2.

First notice that at the end of step 2, some child is kept for the next generation with some probability which depends on the current population p . Assemble these probabilities together into a vector $\mathcal{G}(p)$ so that

$$\mathcal{G}(p)_i = \text{the probability that } i \text{ is kept for the next generation}$$

Next, note that the next generation depends only on the current population (we assume the crossover and mutation probabilities are fixed, as is the fitness function which influences selection). Hence the GA is a *Markov chain*. Let p be the population vector representing the current population, and let $\tau(p)$ denote the population vector describing the next generation. The probabilities that $\tau(p) = q$ (for various p and q) are assembled into a *transition matrix* Q .

Theorem 1

$$Q_{q,p} = r! \prod \frac{(\mathcal{G}(p)_j)^{r q_j}}{(r q_j)!}$$

The stochastic function τ is the Simple Genetic Algorithm, and the evolution from an initial population p is the sequence

$$p \mapsto \tau(p) \mapsto \tau(\tau(p)) \mapsto \dots$$

Infinite Population GA

The infinite population GA corresponding to τ is defined to be \mathcal{G} , and the evolution from an initial population p is the sequence

$$p \mapsto \mathcal{G}(p) \mapsto \mathcal{G}(\mathcal{G}(p)) \mapsto \dots$$

There are a number of questions which naturally arise, but the following demand immediate attention.

- If \mathcal{G} acts on infinite populations, then what *is* an infinite population, and how is it p ?
- By definition, $\mathcal{G}(p)$ is a probability vector such that

$$\mathcal{G}(p)_i = \text{the probability that } i \text{ is kept for the next generation}$$

How is that a population?

The first question is answered as follows. Let P_1, P_2, \dots be any sequence of populations with respective sizes r_1, r_2, \dots such that $\lim_{i \rightarrow \infty} r_i = \infty$. Let the population vector representing P_i be p_i . Since p_1, p_2, \dots is an infinite sequence of points in the compact set Λ , there exists a convergent subsequence tending toward some point $p \in \Lambda$. Such a point p is *defined* to represent an infinite population. It follows from example 1 that every element of Λ represents an infinite population. As has been previously observed, points in Λ with rational coordinates represent finite populations as well.

The second question has already been answered. As previously observed, elements of Λ can be interpreted as populations or as probability distributions over Ω . Thus the probability distribution $\mathcal{G}(p)$ may be regarded as representing an infinite population.

Making Sense of the definition

To legitimize calling \mathcal{G} the infinite population GA, one might show that in some sense $\tau \rightarrow \mathcal{G}$ as $r \rightarrow \infty$. The following results prepare the way.

Let \mathcal{E} denote expectation, so that $\mathcal{E}(\tau(p))$ is the expected next generation given current population p

$$\mathcal{E}(\tau(p)) = \sum_q Q_{q,p} q$$

Let \mathcal{V} denote the expected squared deviation from the mean, so that

$$\mathcal{V}(\tau(p)) = \mathcal{E}(\|\tau(p) - \mathcal{E}(\tau(p))\|^2) = \sum_q Q_{q,p} \|q - \mathcal{E}(\tau(p))\|^2$$

Theorem 2

$$\mathcal{E}(\tau(p)) = \mathcal{G}(p)$$

Theorem 3

$$\mathcal{V}(\tau(p)) = \frac{1 - \mathcal{G}(p)^T \mathcal{G}(p)}{r}$$

According to Theorem 2, the expected next generation is given by the infinite population GA. This is as one might anticipate; given infinite populations, the various random choices involved in obtaining the next generation ought to average out to yield the expectation. But whether or not that aligns with your intuition, consider that Theorems 2 and 3 imply

$$\lim_{r \rightarrow \infty} \mathcal{E}(\|\tau(p) - \mathcal{G}(p)\|^2) = 0$$

Therefore, as population size goes to infinity, τ converges to \mathcal{G} in mean square.

Theorem 4 (*Chebyshev's Inequality*) For all $\varepsilon > 0$,

$$\text{Prob} (|Y| \geq \varepsilon) \leq \frac{E(Y^2)}{\varepsilon^2}$$

Choosing $Y = \|\tau(p) - \mathcal{E}(\tau(p))\|$ in Chebyshev's Inequality yields

Theorem 5 For all $\varepsilon > 0$ and all $p \in \Lambda$,

$$\text{Prob} (\|\tau(p) - \mathcal{G}(p)\| \geq \varepsilon) < \frac{1}{r\varepsilon^2}$$

Connecting the Finite and Infinite

It is not uncommon to encounter opinions to the effect that the infinite population model \mathcal{G} can give very misleading information about finite population GAs, or gives an unreliable estimate for their dynamics.

In view of the results presented so far, caution is called for when interpreting statements of that kind. To the extent such statements are true, they are simply referring to a misapplication of the infinite population model. ***Any tool can be misused!***

Whereas there *is* a provably correct connection between the evolutionary sequences

$$p \mapsto \tau(p) \mapsto \tau(\tau(p)) \mapsto \dots \tag{1}$$

$$p \mapsto \mathcal{G}(p) \mapsto \mathcal{G}(\mathcal{G}(p)) \mapsto \dots \tag{2}$$

the relationship is not what one might naively suppose. As emphasized by Vose (1999) in "What are Genetic Algorithms? A Mathematical Perspective" (*IMA Volume 111, Evolutionary Algorithms*),

"Because taking the limit, as population size goes to infinity, yields a deterministic system whose dynamics coincide with that induced by the function which produces the expected next generation, the infinite population model can be used to determine the expected behavior of a genetic algorithm.

The previous conclusion is slippery, because the statement is so vague. Following are a list of specific interpretations, each of which is a *false* assertion.

- The path followed by the infinite population model is the expected path followed by a GA.
- If the infinite population model converges to a population q , then q is representative of the GA's steady state distribution.
- If the infinite population model indicates that certain elements are likely to emerge in early generations, then it probable for these elements to likewise emerge during genetic search.

This list of errors is not exhaustive, but these are not uncommon misconceptions. Equally illusory is the ignis fatuus that because such notions are false, the infinite population model is not of fundamental importance to the theory of genetic search."

The "infinite population model" refers to \mathcal{G} , and evidence of it's relevance to finite population genetic search is (so far) provided by Theorems 1, 2, 3, 5. In particular, \mathcal{G} determines τ . The "infinite population model" neither refers to the evolutionary sequence (2), nor does it propose that sequence as an approximation for (1), unless one is abusing the term! The paper "Modeling Finite Populations" (Adam Prügler-Bennett (2003) in FOGA 7) provides a nice example demonstrating how (1) and (2) may be quite different.

What then is the relevance of the evolutionary sequence (2) to finite population behavior? To a large extent, that is the subject of the book "The Simple Genetic Algorithm: Foundations and Theory" (one should be cautioned however that the book needs to be read completely through to properly understand the answer).

Transients

Theorem 6 *Given $k > 0$, $\varepsilon > 0$ and $\gamma < 1$, there exists N such that with probability at least γ and for all $0 \leq t \leq k$*

$$r > N \implies \|\tau^t(x) - \mathcal{G}^t(x)\| < \varepsilon$$

The simplest connection between (1) and (2) has to do with transients. According to Theorem 6, as $r \rightarrow \infty$, and with probability converging to 1, the initial transient of a population trajectory under τ converges to following the path determined by iterates of \mathcal{G} , and that transient occupies a time span diverging to infinity. Consequently, large populations will predominately appear near some fixed point ω of \mathcal{G} if the evolutionary path $p, \mathcal{G}(p), \mathcal{G}^2(p), \dots$ approaches a fixed point relatively quickly.

Note that Theorem 6 does not indicate how large N is, and that might not align with what one would either expect or desire. The universe is indifferent to our whims; it is as it is, independent of whether we happen to find that pleasant or convenient for any particular purpose.

Punctuated Equilibria

The description above appears in contrast to the fact that with positive mutation τ is an ergodic Markov chain; every state must be visited infinitely often. This is reconciled in *punctuated equilibria*: random events may eventually move the system to a population p' contained within the basin of attraction (with respect to the underlying dynamical system corresponding to \mathcal{G}) of a different fixed point ω' . Since the behavior of GAs are time independent, the anticipated behavior of large populations approximates the trajectory $p', \mathcal{G}(p'), \mathcal{G}^2(p'), \dots$ — as if p' were the initial population — to reach a new temporary stasis in the vicinity of ω' .

This cycle of a period of relative stability followed by a sudden change to a new dynamic equilibrium is the picture provided as $r \rightarrow \infty$ when (2) converges to a limit point which depends non trivially on p . Although these conclusions hold for sufficiently large r , how large the population size must be before these qualitative aspects are typically exhibited is an open question.

Steady State

Let \mathfrak{S} be the set of fixed points of \mathcal{G} (i.e., the set of populations p for which $\mathcal{G}(p) = p$). If (2) converges for all p , \mathcal{G} is said to be *focused*. If τ is an ergodic Markov chain, then \mathcal{G} is said to be *ergodic*. In that case let $\pi(A)$ be the proportion of time that (1) spends in A , averaged over infinitely many generations,

$$\pi(A) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{0 \leq t < T} [\tau^t(p) \in A]$$

where $[expression]$ denotes 1 if *expression* is true, and 0 otherwise.

Theorem 7 *Suppose \mathcal{G} is focused and ergodic. For every open set U containing \mathfrak{S} ,*

$$\lim_{r \rightarrow \infty} \pi(U) = 1$$

Theorem 7 indicates that large populations predominately spend time near fixed points of \mathcal{G} . If the differential $d\mathcal{G}_x$ of \mathcal{G} at x is continuous in x and has no eigenvalues of unit modulus when x is a fixed point, then \mathcal{G} is called *hyperbolic*. It is known that if proportional selection is used, then the set of fitness functions for which \mathcal{G} is hyperbolic is dense and open. Moreover, in that case \mathcal{G} has only finitely many fixed points in Λ .

Theorem 7 suggests that GAs are adept at locating regions in the vicinity of fixed points of \mathcal{G} ; it is natural to wonder about transition probabilities from one such region to another.¹ If the transition

¹Theorem 7 speaks in a proven manner to large populations; empirical evidence indicates smaller population GAs have a general propensity for fixed points of \mathcal{G} as well (the reader is cautioned that the sense in which this is true might differ from a naive interpretation).

probabilities from temporary stasis in the vicinity of one fixed point to temporary stasis near another can be determined, then τ could be approximated by a Markov chain over the fixed points of \mathcal{G} .

Subject to technical conditions (that \mathcal{G} is hyperbolic, ergodic, and has a complete Lyapunov function), the goal of constructing a meta level Markov chain, as described in the previous paragraph, has been achieved in the large population case.

Let $\rho = x_0, \dots, x_k$ be a sequence of points from Λ , referred to as a *path* of length k from x_0 to x_k . The *cost* of ρ is

$$|\rho| = \alpha_{x_0, x_1} + \dots + \alpha_{x_{k-1}, x_k}$$

where

$$\alpha_{u,v} = \sum_j v_j \ln \frac{v_j}{\mathcal{G}(u)_j}$$

and it is assumed that mutation is positive so that \mathcal{G} maps Λ into its interior. Let the stable fixed points of \mathcal{G} in Λ be $\{\omega_0, \dots, \omega_w\}$ and define

$$\rho_{\omega_i, \omega_j} = \inf \{ |\rho| : \rho \text{ is a path from } \omega_i \text{ to } \omega_j \}$$

Let C_r be a Markov chain defined over $\{0, \dots, w\}$ with $i \rightarrow j$ transition probability (for $i \neq j$)

$$\exp\{-r \rho_{\omega_i, \omega_j} + o(r)\}$$

Up to the uncertainty in the $o(r)$ terms, the desired Markov chain is C_r . As the book “The Simple Genetic Algorithm: Foundations and Theory” demonstrates, C_r cannot possibly be appropriate for small r , because unstable fixed points, complex fixed points, and stable fixed points of \mathcal{G} which are outside Λ make no contribution to C_r . However, as $r \rightarrow \infty$ the steady state distribution of the Markov chain τ converges to that of C_r , which is formally the sense in which C_r approximates τ .

The *fixed point graph* is the weighted, directed, complete graph \mathcal{I} on the states of C_r with edge $i \rightarrow j$ having weight $\rho_{\omega_i, \omega_j}$. A *tributary* is a tree containing every vertex of \mathcal{I} such that all edges point towards the root. Let T_k be a tributary rooted at k ; the cost of T_k is the sum of its edge weights, and the fixed point corresponding to T_k is ω_k .

Theorem 8 *As $r \rightarrow \infty$, the measure π converges to point mass at the fixed point corresponding to the minimum cost tributary of \mathcal{I} (provided it exists).*

Quotients

Managing complexity involves quotients (or some generalization thereof) if by “managing complexity” one intends to reduce complexity while simultaneously maintaining important aspects of fidelity. The following diagram is an abstraction of the general scheme being considered. In that illustration, $x \in X$ represents state and $h : X \rightarrow X$ transforms state. Complexity is managed by Ξ , which maps state into a simpler form, and by \tilde{h} which has reduced complexity by virtue of transforming simplified state.

$$\begin{array}{ccc} x & \xrightarrow{h} & h(x) \\ \Xi \downarrow & & \downarrow \Xi \\ \Xi x & \xrightarrow{\tilde{h}} & \Xi h(x) \end{array}$$

Maintaining important aspects of fidelity is interpreted to mean the diagram commutes; both paths from x to $\Xi h(x)$ yield identical results. Thus Ξ can be regarded as defining what aspects of fidelity *are maintained*—if leeway exists in choosing it—or what aspects of fidelity are *capable of preservation*—if there is virtually no leeway. The reduced complexity model \tilde{h} is the *quotient* of h corresponding to the *coarse graining* Ξ , and h is said to be compatible with Ξ if a quotient exists.

Whereas modeling h in an approximate fashion (by relaxing commutativity of the diagram) is interesting, it is also interesting to inquire whether one can do better than approximation, and if so, then

how? Moreover, knowledge of what it is that can be exact may identify a useful starting point for what it is that later will be approximated or perturbed from.

Let \equiv be an arbitrary equivalence relation over Ω , and let $\{0^*, \dots, k^*\}$ be equivalence class representatives. The linear operator Ξ with matrix

$$\Xi_{i^*,j} = [i^* \equiv j]$$

lifts \equiv to populations by

$$p \equiv q \iff \Xi p = \Xi q$$

In a stochastic setting, τ is said to be *compatible with \equiv* iff

$$p \equiv p' \implies \forall q. \text{Prob} \{\tau(p) \equiv q\} = \text{Prob} \{\tau(p') \equiv q\}$$

In that case, $\tilde{\tau}$ defined by $\tilde{\tau}(\Xi x) = \Xi \tau(x)$ is referred to as the *quotient* of τ .

Theorem 9 *A quotient $\tilde{\tau}$ exists if and only if a corresponding quotient $\tilde{\mathcal{G}}$ of \mathcal{G} exists. Moreover, the transition matrix for $\tilde{\tau}$ is obtained from*

$$Q_{q,p} = r! \prod \frac{(\mathcal{G}(p)_j)^{rq_j}}{(rq_j)!}$$

by replacing \mathcal{G} by $\tilde{\mathcal{G}}$, p by Ξp , and q by Ξq .

Theorem 9 reduces the stochastic case to a more tractable deterministic setting, and commutativity of the diagram on the previous page is therefore of particular interest when $h = \mathcal{G}$. As this *is* a tutorial, it bears emphasis that theorem 9 is neutral as to whether quotients do or do not exist; it merely provides a means by which that question can be answered in a provably correct way.²

Conclusion

This tutorial has introduced the “infinite population model” and explained some of its connections to finite population GAs. The most direct connection is that \mathcal{G} defines τ as indicated in theorem 1.

Perhaps the second most direct connection is that whether or not quotients for τ exist (with respect to Ξ as defined above) is answered by the corresponding question for \mathcal{G} .

Statistical connections between τ and \mathcal{G} are given in theorems 2 and 3, and a consequent approximation result is provided by theorem 5.

Qualitative relationships between τ as a Markov chain and the infinite population model are provided by theorems 6, 7, and 8.

The reader is referred to the book “The Simple Genetic Algorithm: Foundations and Theory” (by Michael D. Vose) for details and further results.

²While on the subject of quotients, it can be fairly said that there is perhaps even greater confusion with respect to that topic than there is concerning the infinite population model. There are a growing number of opinions (and one would expect papers soon to follow) which misrepresent and misinterpret the meaning and significance of quotients.