Statistical Mechanics Theory of Genetic Algorithms

Part I — Overview

Jonathan Shapiro

Department of Computer Science, University of Manchester

September 6, 1999

Preview of Coming Attractions

The statistics mechanics theory of genetic algorithms:

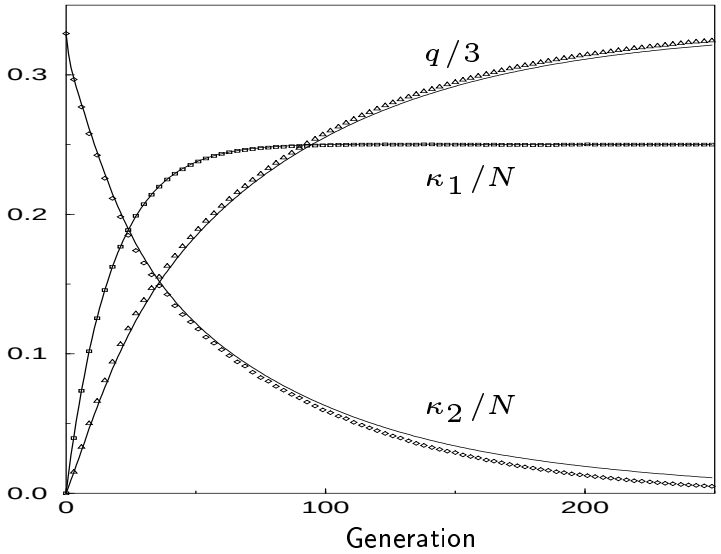
- Describes the *average dynamics* of genetic algorithms acting on specific problems.
- Requires approximations and assumptions, mostly controllable.
- Includes finite population effects.
- Dynamical equations can be difficult to solve, but the computation is independent of the population size and of the string size.
- Leads to accurate predictions

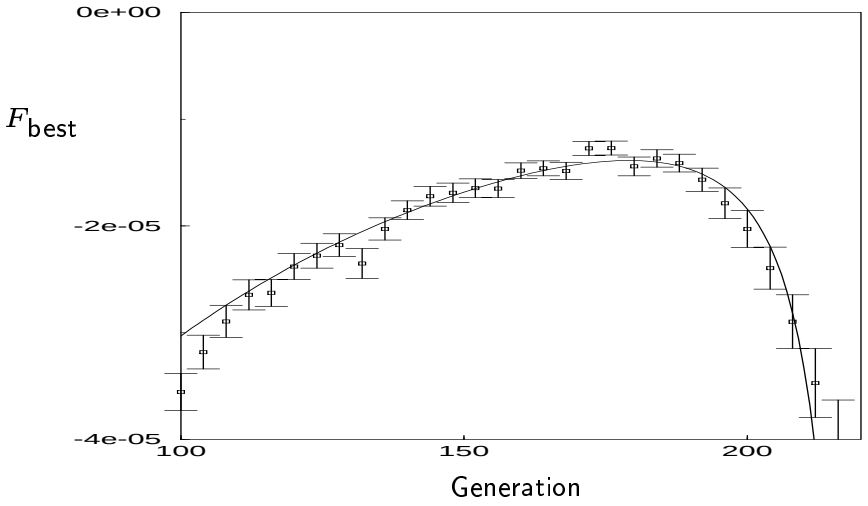
Example: Subset Sum — A weakly NP hard problem:

Reference: (Rattray '96)

- There are L files with sizes f_i for the ith file.
- The back-up disk has size D and is not big enough to hold all of the files.
- Find the subset of the files which comes closest to filling the disk.

κ_1 is the mean amount stored on the disk; κ_2 is the variance, q is a measure of the similarity of the solutions in the population.





The optimum phenotype was $R_{\text{opt}} = N/4$. The other parameters were $N = 80, L = 150, p_m = 0, \beta_s = 0.03$ and uniform crossover was used with $a = 0.5$.

What system will we study

- Dynamics of a finite population of strings
- Undergoing mutation + selection + crossover
- Many components in each string
- Complex ("realistic") selection models — problems with some characteristics of realistic problems.

Mostly interested in applied (directional) selection.
Natural selection also of interest

A simple genetic algorithm

Representation: fixed length string of L bits. $x_i \leftarrow \{0, 1\}$. E.g.

$$\mathbf{x} = 1001110$$

Quantity to be optimized: *Objective-function* or *trait*; a function of the strings, $E(\mathbf{x})$.

Population: of N strings. Fixed size.

Selection weight: N parents are chosen from the current population; probability of selection string α is

$$p^\alpha = \frac{W(\mathbf{x}^\alpha)}{\sum_\gamma W(\mathbf{x}^\gamma)}.$$

Recombination/Crossover: This operation produces offspring from a pair of parents by *randomly* combining parts of each parent.

Single-point crossover:

$$A_1 = 1111111|1111111 \quad \rightarrow \quad A_1' = 0000000|1111111$$
$$A_2 = 0000000|0000000 \quad \rightarrow \quad A_2' = 1111111|0000000$$

Multi-point crossover:

$$A_1 = 111|111111|11111 \quad \rightarrow \quad A_1' = 000|111111|00000$$
$$A_2 = 000|000000|00000 \quad \rightarrow \quad A_2' = 111|000000|11111$$

Uniform crossover: Each site is independently chosen em at random from parent 1 or parent 2.

Mutation: This operation simply changes characters in the offspring string at random. Each character is changed with a small probability.

Goals of the Theory

To answer questions such as:

- How do GAs work (insight)?
- On what problems do GAs excel?
- How can genetic algorithm search be optimized?

By solving the dynamics of evolutionary search on specific problems:

To do this, we must

- Derive a *deterministic* set of equations describing the evolution
- Solve it for the specific problem.

We have to find a set of equations of motion which are deterministic and simple enough to solve.

Aside — Schema theory does not achieve this

- Does not produce equations of motion which can be iterated.

  expectation value = actual values in particular population

Usual approach is to (implicitly) assume infinite population so that both sides of equation involve expectation values.

Basic Idea

To get simple equations: Describe the population in terms of a few *macroscopic* quantities — statistics of the population.

To get deterministic equations: Describe the population in terms of average behaviour of the macroscopics.

To close the loop: Use a statistical inference method (maximum entropy).

Levels of Description

Microscopic description: Describe the system in terms of the actual details of the constituents.

Macroscopic description: Describe the system in terms of a few statistical, and large scale properties of the system.

Example — electrons in a circuit

Microscopic description: The quantum-mechanical wave function of the 10^{24} conduction elections interacting with themselves, the other elections, the nuclei, etc.

Macroscopic description: $V = IR$.

Points about microscopic description:

- It is potentially rigorous, but . . .
- It is extremely difficult to solve,
- Were a solution to exist, it would be extremely complicated; difficult to obtain useful information.

Points about the macroscopic description:

- It is approximate,
- It is easy to solve and directly related to the useful information,
- There is missing information. A method of inference may be required.

Statistical Mechanics = macroscopic description + MaxEnt inference

Elements of Statistical Mechanics

Image a system which can be in states: $s_1, s_2, \ldots, s_{N_\Omega} \in \Omega$.

(Huge number of states) E.g.

- Molecules in a glass of H_2O,
- genes in a population of genotypes.

Dynamics — stochastic, at least in initial conditions.

Remove the stochasticity by considering an *ensemble* of systems.
Goal is to predict behaviour on average over the ensemble.

Microscopic knowledge: Probability of finding a member of the ensemble in state $s_i, P(s_i)$. N_Ω numbers.

Our actual knowledge: Some average (macroscopic) quantity Q.

$$Q = \sum_{s_i \in \Omega} \tilde{Q}(s_i) P(s_i).$$

Problem: Suppose you want to know about some other quantity, G,

$$G = \sum_{s_i \in \Omega} \tilde{G}(s_i) P(s_i)$$

if we don't know the P's, how can we do it.

Maximum Entropy Inference: Choose the assignment of the $\{P(s_i)\}$ which is consistent with Q and otherwise has the most "uncertainty" (or assumes the least information).

Uncertainty of a distribution: Given by the entropy

$$H = -\sum_{s_i \in \Omega} P(s_i) \log P(s_i),$$

determines the amount of information required to remove the uncertainty.

Maximum Entropy Distribution: Optimize

$$H = -\sum_{s_i \in \Omega} P(s_i) \log P(s_i) + \lambda \sum_{s_i \in \Omega} \tilde{Q}(s_i) P(s_i)$$

$$P(s_i) = \frac{e^{-\lambda \tilde{Q}(s_i)}}{\sum_{s'_i \in \Omega} e^{-\lambda \tilde{Q}(s'_i)}}$$

The Lagrange multiplier, λ is chosen so that

$$\frac{\sum_{s_i \in \Omega} e^{-\lambda \tilde{Q}(s_i)} \tilde{Q}(s_i)}{\sum_{s'_i \in \Omega} e^{-\lambda \tilde{Q}(s'_i)}} = Q.$$

Statistical Mechanics: describe system by

- Considering averages over ensembles
- Using maximum entropy distribution to infer other quantities.

Maximum Entropy Inference

Maximum Entropy Distribution:

- Occupies almost all of probability distribution space, as $N_\Omega \to \infty$,
- Is an assumption (of microscopic chaos).

Works well if

- Q is "self-averaging; averages of Q are similar to typical ones,
- Q explains most of the non-random aspects of the system (otherwise further macroscopic quantities will be required).

Application to Evolutionary Dynamics

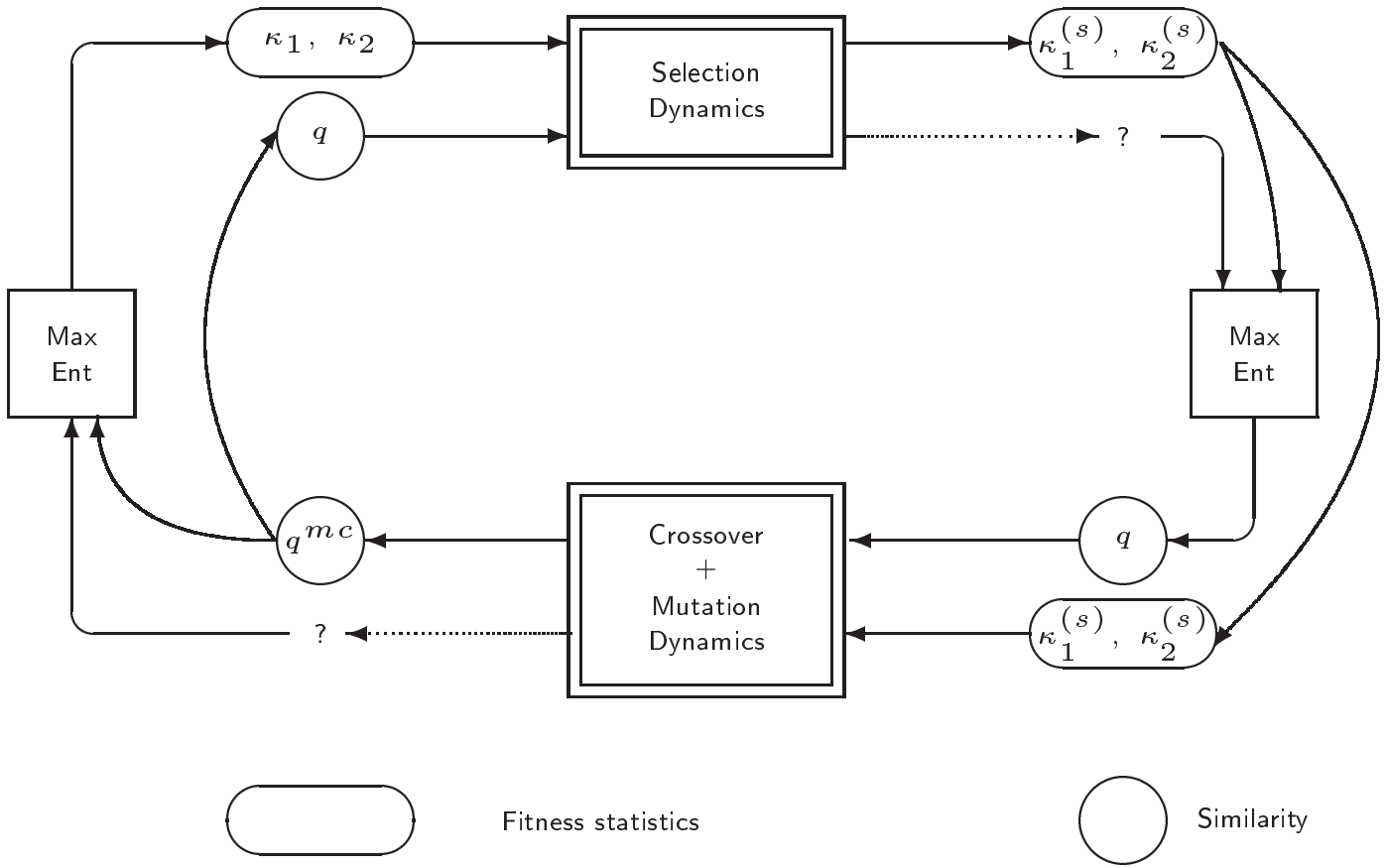
Microscopic approach: Markov chain and quadratic dynamical systems methods.

Microscopic approach:

- Predict average statistics of E, e.g. mean standard deviation, best, etc. (Called $\kappa_1, \kappa_2, \ldots$).
- Predict average configurational statistics, e.g. a measure of how similar the strings in the population are q.
- **Problem:** It is not possible to relate configurational statistics to objective function statistics.

Maximum Entropy inference is used to close this loop.

Closing the loop

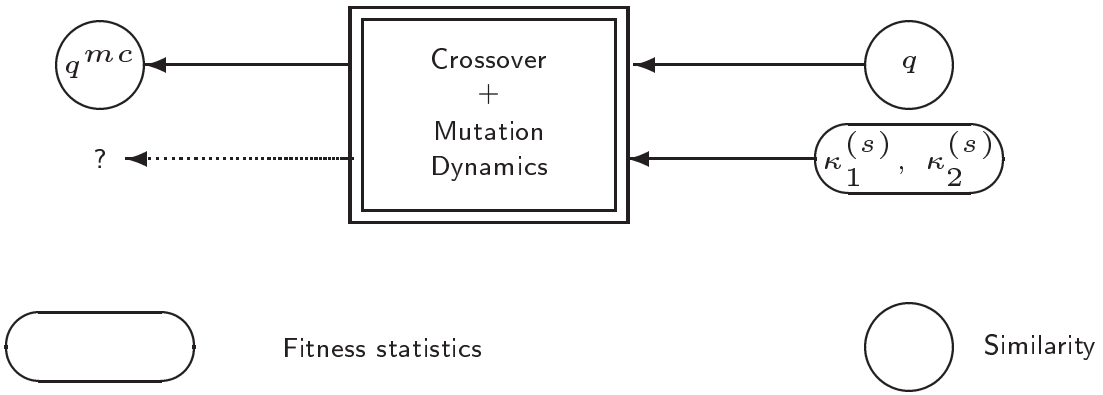


Closing the loop

Fitness statistics

Similarity

Closing the loop



Closing the loop