Statistical Mechanics Theory of Genetic Algorithms

Part II — Illustration of the details

Jonathan Shapiro

Department of Computer Science, University of Manchester

September 6, 1999

The problem

Variables: L binary variables, $x_i = \{0, 1\}$

Problem instance: Defined by L "weights" $J_i \in \mathcal{R}$. Determine the importance of each site.

Objective Function: Weighted sum of variables, $\boxed{E = \sum_{i=1}^{L} J_i x_i}$.

The J's are fixed (for a given problem); optimize over the x's.

Optimize using a GA consisting of a population size N of binary strings of length L.

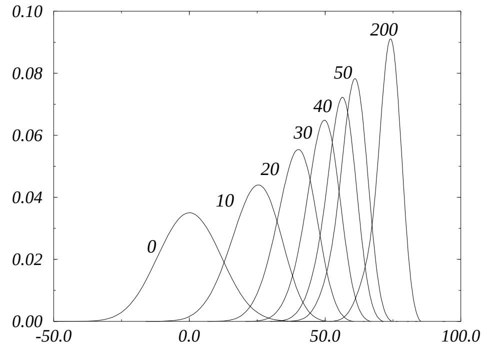
Notes:

- A trivial optimization problem. The solution is $x_i = iiif\, J_i > 0$.
- If we let the objective function be a *function* of weighted sum of variables, hard problems can be considered (see below, Rattray 1996, Rogers and Prügel-Bennett, this workshop).

$$E = \mathcal{E}\left(\sum_{i=1}^{L} J_i x_i\right).$$

Choice of macroscopics — first cut

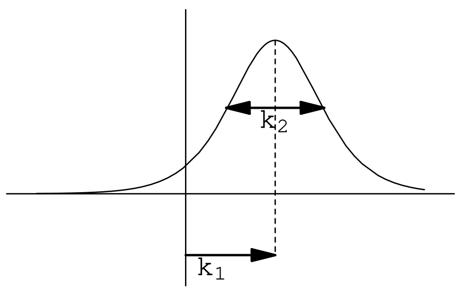
Consider the distribution of the objective function in the population averaged over runs of the genetic algorithm. $\boxed{\rho_t(E)}$.



Try to evolve *statistics* of this distribution (e.g. mean, variance, best, etc).

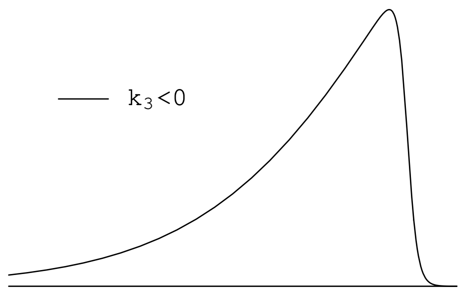
Cumulants

$$\begin{aligned}
\kappa_1 &= \text{mean} \\
\kappa_2 &= \text{variance} \\
\kappa_3 &= \text{related to ``skewness''} \\
\kappa_4 &= \text{related to ``kurtosis'' or ``excess''} \\
&\vdots \\
\kappa_n &= \text{measure of } n\text{th moment relative to a Gaussian}
\end{aligned}$$

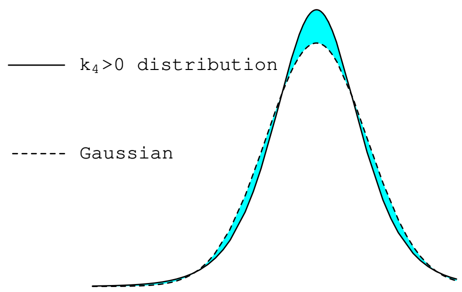


First two cumulants:

A distribution with negative "skewness":



A distribution with positive kurtosis:



Cumulants average well. They contain all information about the distribution.

Choice of Macroscopics — second cut

We will also require a measure of the similarity of the strings in the populations. Particularly important for crossover.

Correlation:

$$q = \frac{1}{LN^2} \sum_i \sum_\mu \sum_\nu J_i^2 (2x_i^\mu - 1)(2x_i^\nu - 1),$$

$$q = \begin{cases} \frac{1}{L} \sum_i J_i^2 & \sim 1; \quad \text{population completely correlated} \\ \frac{1}{LN} \sum_i J_i^2 & \sim \frac{1}{N}; \quad \text{population completely uncorrelated} \end{cases}$$

Selection Dynamics

Let $W(E^\mu)$ — Selection "weight" for a string with objective function E^μ.

The selection weight represents the selection operator:

$$W(E) = \begin{cases} E; & \text{“Fitness proportional selection”} \\ \exp(sE); & \text{Boltzmann selection} \\ \Theta(E - E_0); & \text{Culling selection} \\ \int_{-\infty}^{E} \rho(E')dE'; & \text{Ranking selection} \\ \text{etc.} & \end{cases}$$

We must average over the selection process to find the average statistics of the distribution after selection $\rho_t^{(s)}$.

How to think about selection

There are two stages of averaging:

Average over the ways of selecting from set $\{E^\mu\}$:

- Selection is stochastic; we are predicting properties of the averages over this stochasticity.

Average over all ways of selection N from $\rho_t(E)$ to produce $\{E^\mu\}$:

- This is not real stochasticity. It is an approximation made necessary because the actual levels $\{E^\mu\}$ *are unknown*. All that is known is the distribution $\rho_t(E)$ (or its statistics) with which they are consistent.
- The simplest assumption is made — that they are drawn independently from $\rho_t(E)$. A maximum entropy assumption.

Example — change in mean due to selection

$$\kappa_1 = \sum_\mu E^\mu$$

$$\Downarrow$$

$$\sum_\mu E^\mu \boxed{n^\mu} \longleftarrow$$ Average over number of times μ selected n^μ.

$$\Downarrow$$

$$\sum_\mu \boxed{E^\mu} \frac{W^\mu}{\sum_\nu W^\nu} \longleftarrow$$ Average over ways of producing $\{E^\mu\}$

$$\Downarrow$$

$$\kappa_1^{(s)} = \boxed{\boxed{\left\langle \frac{\sum_\mu E^\mu W^\mu}{\sum_\mu W^\mu} \right\rangle}} \longleftarrow \langle\cdot\rangle = \prod_\mu^N \int dE^\mu \rho_t(E^\mu)\cdot$$

A complex multi-dimensional integral.

Expanding around the infinite population solution

In an infinite population: the situation is simple,

$$\kappa_1^{(s)} = \left\langle \frac{\sum_\mu E^\mu W^\mu}{\sum_\mu W^\mu} \right\rangle = \frac{\overline{EW}}{\overline{W}}$$

where $\bar{\cdot}$ denotes averaging over the distribution $\rho_t(E)$.

An expansion for weak selection, large N: (A non-rigorous method):

Write the expression as

$$\kappa_1^{(s)} = \left\langle \frac{\sum_\mu E^\mu W^\mu}{N\overline{W} + \sum_\mu \left(W^\mu - \overline{W}\right)} \right\rangle$$

and expand in small $\sum_\mu \left(W^\mu - \overline{W}\right)$

$$\kappa_1^{(s)} = \left\langle \frac{\sum_\mu E^\mu W^\mu}{N\overline{W}} \left[1 - \frac{\sum_\mu \left(W^\mu - \overline{W}\right)}{N\overline{W}} + \frac{\sum_\mu \left(W^\mu - \overline{W}\right)^2}{\left(N\overline{W}\right)^2} - \dots \right] \right\rangle .$$

Example: let $W = 1 + sE$ with selection strength s. Then

$$\kappa_1^{(s)} = \kappa_1 + s\left(1 - \frac{1}{N}\right)\kappa_2$$

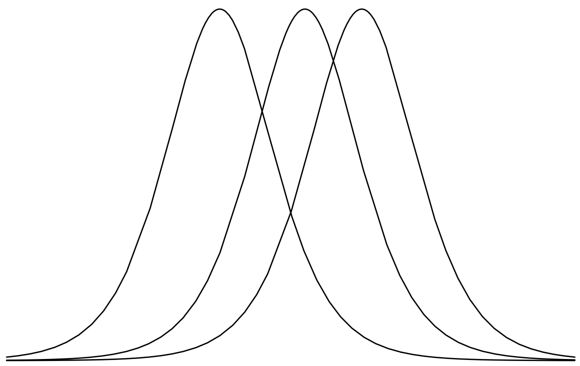
Carrying on . . . ,

$$
\begin{aligned}
\kappa_1^{(s)} &= \kappa_1 + s\left(1 - \frac{1}{N}\right)\kappa_2, \\
\kappa_2^{(s)} &= \left(1 - \frac{1}{N}\right)\kappa_2 + s\left(1 - \frac{3}{N} + \frac{2}{N^2}\right)\kappa_3, \\
\kappa_3^{(s)} &= \left(1 - \frac{3}{N} + \frac{2}{N^2}\right)\kappa_3 + s\left(1 - \frac{7}{N} + \frac{12}{N^2} - \frac{6}{N^3}\right)\kappa_4 \\
&\quad - s\frac{6}{N}\left(1 - \frac{1}{N}\right)\kappa_2^2, \\
\kappa_4^{(s)} &= \left(1 - \frac{7}{N} + \frac{12}{N^2} - \frac{6}{N^3}\right)\kappa_4 - \frac{6}{N}\left(1 - \frac{1}{N}\right)\kappa_2^2 \\
&\quad + s\left(1 - \frac{15}{N} + \frac{50}{N^2} - \frac{60}{N^3}\right)\kappa_5 - \frac{60s}{N}\left(1 - \frac{3}{N} + \frac{6}{N^2}\right)\kappa_2\kappa_3
\end{aligned}
$$

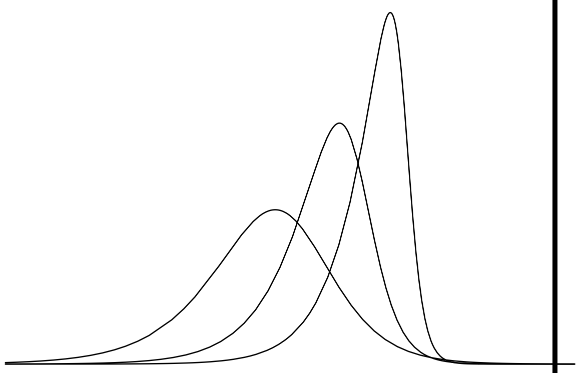
Punch lines about selection and cumulants

- Finite population introduces higher cumulants
- The stronger the selection pressure, the more important these effects are.

Infinite Population:



Finite Population:



Evolving the correlation due to selection

q relates to $\{\mathbf{x}^\mu\}$ not $\{E^\mu\}$. How to proceed?

1. **Use maximum entropy inference:** In principle, this would be

$$P_{\text{ME}}(\{\mathbf{x}\}|\kappa_1,\ldots) = Z^{-1} e^{-\lambda_1 \tilde{\kappa_1}(\{\mathbf{x}\}) - \lambda_2 \tilde{\kappa_2}(\{\mathbf{x}\}) - \ldots}$$

where Z is the normalization. Lagrange multipliers λ are chosen to give the cumulants we know. This would be used to compute $q^{(s)}$.

For calculational reasons, we use

$$\boxed{P_{\text{ME}}(\mathbf{x}|\kappa_1, q) = Z^{-1} e^{-\lambda_1 \tilde{\kappa_1}(\mathbf{x}) - \lambda_2 \tilde{q}(\mathbf{x}) - \ldots}}$$

and use to calculate $\kappa_2^{(s)}$.

This underestimates correlations due to "founder effects" — selection can only increase fitness due to duplication.

2. **Putting in founder effects:** divide the correlation after selection into two parts:

$$q^{(s)} \quad \to \quad \underbrace{\tfrac{1}{L}\textstyle\sum_i J_i^2 \displaystyle\sum_\mu n^\mu n^\mu}_{\substack{\text{founder effect}\\ \text{due to duplication}}} \quad + \quad \underbrace{\sum_\mu \sum_\nu q^{\mu\nu} n^\mu n^\nu}_{\substack{\text{natural increase}\\ \text{treated via MaxEnt}}}$$

Founder effects: Can be treated as before by expanding around the infinite population,

$$\left\langle \sum_\mu \left(\frac{W^\mu}{\sum_\nu W^\nu} \right)^2 \right\rangle \sim \frac{\overline{W^2}}{N\overline{W}^2} + \cdots$$

$$\sim \frac{1}{N}\left(1 + \kappa_2 s^2 + \ldots\right)$$

Natural increase: Would occur in an infinite population, due to the decrease in state space for higher fitness states. Calculated via MaxEnt.

Punch lines about selection and correlations

Correlation increases due to two effects

Infinite population increase: due to smaller space occupied by higher fitness states. Computed from MaxEnt given the first two cumulants.

Duplication effects: An essential finite population effect. Selection can only increase fitness through duplication

Crossover Dynamics

Ingredients:

1. Introduce some new crossover variables,

$$C_i^{\mu\nu} = \begin{cases} 1; & \text{if site } i \text{ in } \mu \text{ is replaced by site } i \text{ in } \nu \\ 0; & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mathbf{x}^\mu &\rightarrow \mathbf{x}^\mu + (\mathbf{x}^\nu - \mathbf{x}^\mu) \cdot \mathbf{C}^{\mu\nu} \\ E^\mu &\rightarrow E^\mu + \Delta E^{\mu\nu} \end{aligned}$$

Statistics of C's determined by the crossover model. For simplicity, consider uniform crossover with probability a of crossing at each site.

2. Apply crossover on the *average* mating pool. E.g.

$$\kappa_1^{(c)} = \sum_\mu n^\mu E^\mu + \sum_\mu \sum_\nu n^\mu n^\nu \Delta E^{\mu\nu}$$

Average mating pool: $n^\mu \rightarrow \frac{W^\mu}{\sum_{\mu'} W^{\mu'}}$

Average over $C^{\mu\nu}$: $\Delta E^{\mu\nu} \rightarrow$ $\boxed{\text{work it out}}$

$$\kappa_1^{(c)} = \kappa_1^{(s)}$$

3. Carrying on . . .

$$
\begin{aligned}
q^{(c)} &= q \\
\kappa_2^{(c)} &= \kappa_2 - 2a(1-a)\left[\kappa_2 - \frac{1}{4}\left(\sum_i J_i^2 - Lq\right)\right] \\
\kappa_3^{(c)} &= \kappa_3 - 3a(1-a)\left(\kappa_3 - \kappa_3^{\mathrm{LE}}\right) \\
\kappa_4^{(c)} &= \kappa_4 - 2a(1-a)(2-a(1-a))\left(\kappa_4 - \kappa_4^{\mathrm{LE}}\right)
\end{aligned}
$$

The Linkage Equilibrium (LE) cumulants: cumulants of a distribution where every site is drawn independently a binomial:

$$P_{\mathrm{LE}}(x_i^\mu = 1) = \frac{1}{N}\sum_\nu x_i^\nu \equiv \tau_i.$$

But how does the frequency of 1*'s in the current population depend upon the site?* I.e. what is $\tau_i(J_i)$?

We answer that question by use of our maximum entropy distribution:

$$
\begin{aligned}
P_{\mathsf{ME}}(\{x_i^\mu\}) &= \prod_i P_{\mathsf{ME}}(\tau_i) \\
P_{\mathsf{ME}}(\tau_i) &= Z_i^{-1} e^{-\lambda_1 J_i \tau_i - \lambda_2 J_i^2 \tau_i^2}
\end{aligned}
$$

Punch lines about crossover

- Crossover keeps κ_1 and q unchanged,
- Increase in diversity (variance) is related to how similar the strings in the population are (q)
- Higher cumulants are returned towards the LE fixed point which depends upon frequency of 1's at a site as a function of J_i. Computed via MaxEnt.

Mutation Dynamics

Similar to crossover. Introduce mutation variables to the average mating pool, average over them. Mutation rate is γ.

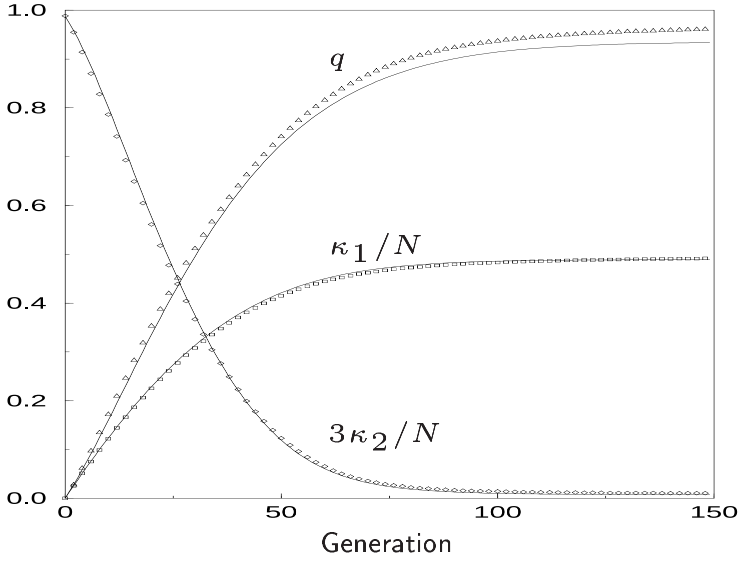
$$
\begin{aligned}
q^{(m)} &= (1-2\gamma)^2 q \\
\kappa_1^{(m)} &= \kappa_1 + 2\gamma(\kappa_1^{\text{random}} - \kappa_1) \\
\kappa_2^{(m)} &= \kappa_2 + 4\gamma(1-\gamma)(\kappa_2^{\text{random}} - \kappa_2)
\end{aligned}
$$

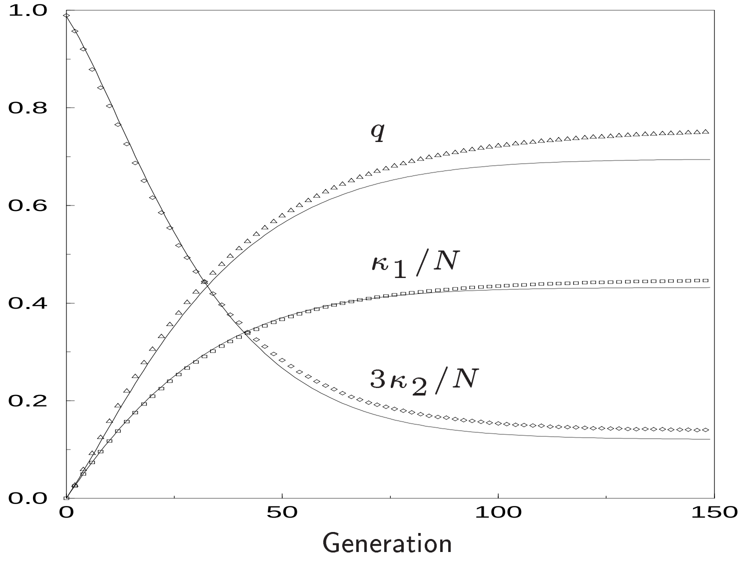
Superscript "random" denotes cumulants of a random population.

The higher cumulants, like those of crossover, relate to the frequency of 1's as a function of J_i, $\tau(J_i)$, which is computed via maximum entropy.

Putting it together

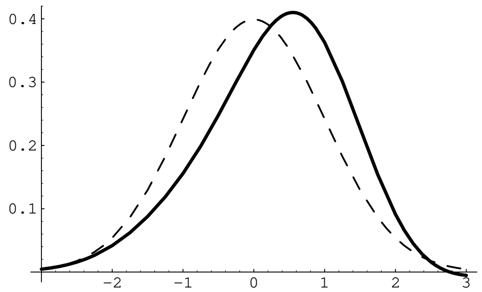
Iterate the process using 4 (or two) cumulants and q.



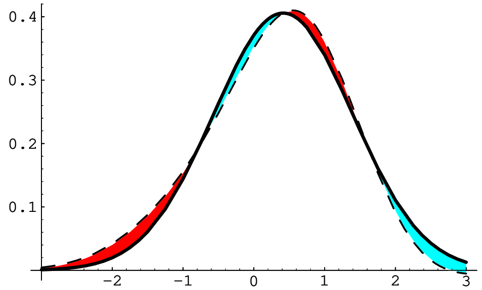


Selection and Crossover Effects

Selection increases the mean, reduces the variance, and introduces skewness



Crossover restores higher cumulants such as skewness to random values, mediated by the degree of correlation in the population.



Sources of error

Truncation: Only small number of the cumulants are used. If selection is strong or crossover is weak it may be important to use higher cumulants (Prügel-Bennett 1997).

Averaging non-linear quantities: Selection introduces non-linearities, e.g. κ_2^2. The ensemble average of κ_2^2 is not the same as square of the ensemble average of κ_2. In this case it is necessary to study the fluctuations (Prügel-Bennett 1997).

Maximum Entropy assumption: is not always valid (Rattray 1996). It may be necessary to use different macroscopics, but it is not clear what those are.