

Stochastic optimization algorithms

Lecture 6, 20180914

Evolutionary algorithms: Properties

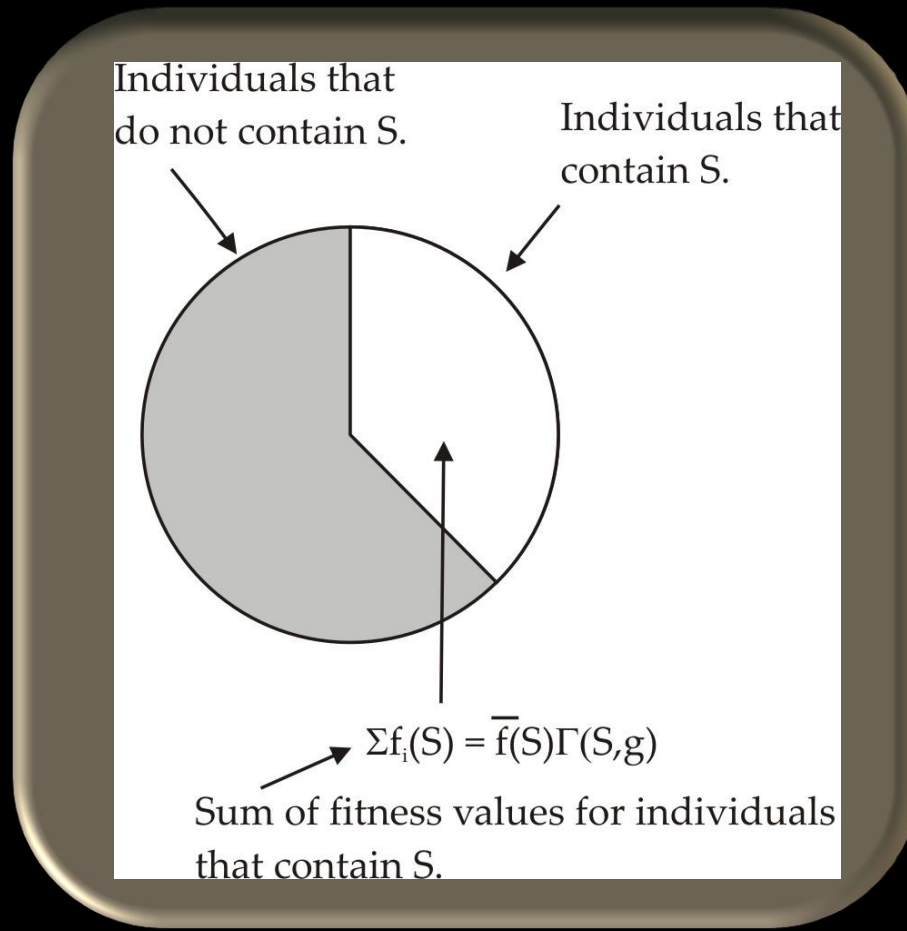
Today's learning goals

- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs.
 - Derive expressions for the expected running time for a simple GA.
 - Derive the optimal mutation rate for a simple GA.
 - Explain the concept of premature convergence
 - List methods for avoiding premature convergence

The schema theorem

- Schema = pattern consisting of 1,0,x, where x is a wild card (represents both 0 and 1).
- Example: 100xx1 represents 100001, 100011, 100101, and 100111.
- Different schemata have different values.
- GAs treat schemata in such a way as to increase (in the population) the number of schemata associated with high fitness.

The schema theorem



The schema theorem

$$E(\Gamma(S, g + 1)) \geq \frac{\bar{F}_S}{\bar{F}} \Gamma(S, g) \left(1 - p_c \frac{d(S)}{m - 1} \right) (1 - p_{\text{mut}})^{o(S)}$$

- Derivation pp. 174-176 (Appendix B2.1)

The schema theorem

- Building blocks: Schemata with
 - Low defining length
 - Low order
 - High fitness (\bar{F}_S)
- Building block hypothesis: GAs manipulate building blocks in an efficient way.
- No proof, but shown in empirical tests.
- The schema theorem does not, in fact, say very much that is specific (or useful) about GAs.

Today's learning goals

- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs.
 - Derive expressions for the expected running time for a simple GA.
 - Derive the optimal mutation rate for a simple GA.
 - Explain the concept of premature convergence
 - List methods for avoiding premature convergence



Infinite-population models

- The analytical treatment of GAs becomes simpler (at least in some cases) if one lets the population size (N) tend to infinity.
- Note: The chromosome length (m) remains finite!
- Enumeration of the 2^m possible strings: $j = 1, 2, 3, \dots, 2^m$.
- When N is infinite one obtains probability distributions (instead of frequencies) for each string (j): $p = p(j)$.

Infinite-population models

- Selection, crossover, and mutation operators combined:

$$\mathcal{G}(p) = \mathcal{G}_m(p) \circ \mathcal{G}_c(p) \circ \mathcal{G}_s(p)$$

- If one considers selection only (in proportion to fitness):

$$\mathcal{G}_s(p) = \frac{F(j)p(j)}{\sum_{j \in \Omega} F(j)p(j)} = \frac{F(j)p(j)}{\bar{F}}$$

Ω = set of all possible chromosomes

Infinite-population models

- Additional simplification: Consider **functions of unitation**, i.e. function in which the fitness f only *depends on the number of ones in the chromosomes*.
- Example: the Onemax function $f(j) = j$, where (NOTE!) j = the *number of ones* in the (binary) chromosome.
- From now on, $p_q(j)$ denotes the probability distribution for chromosomes with j ones (thus $m - j$ zeros) in generation q , where j thus ranges from 0 to m (*not* 2^m as before).

Infinite-population models

- Initial distribution, assuming random initialization:

$$p_1(j) = 2^{-m} \binom{m}{j}$$

Infinite-population models

- One can now compute the average fitness in the first generation as

$$\bar{F}_1 = \sum_{j=0}^m j p_1(j) = 2^{-m} \sum_{j=0}^m j \binom{m}{j} = \frac{m}{2}$$

- See Eq. (B17).

Infinite-population models

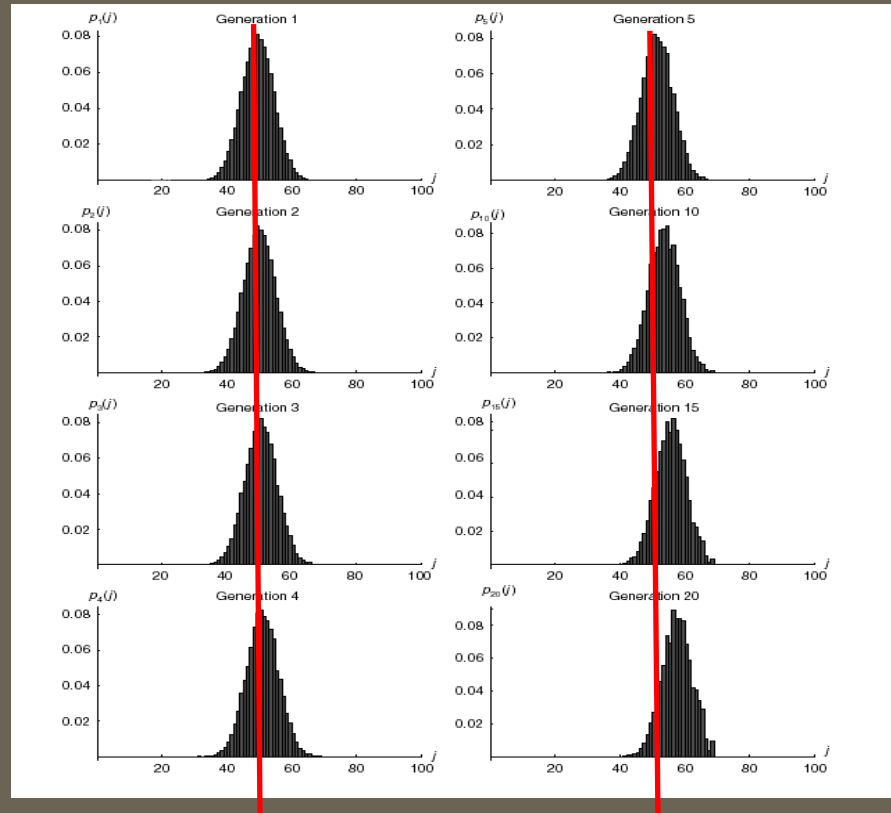
- The probability distribution of the second generation then becomes

$$p_2(j) = \frac{jp_1(j)}{\sum_{j=0}^m jp_1(j)} = \frac{jp_1(j)}{\bar{F}_1} = 2^{1-m} \frac{j}{m} \binom{m}{j}$$

Probability of getting a chromosome with j ones, assuming Onemax fitness

- In principle, one can proceed analytically to compute the probability distribution in generations 3, 4, 5, ... (but the equations soon become very messy).

Infinite population models



Infinite population models

- Thus, one can get an exact description of the evolution of the probability distribution $p_q(j)$.
- So far only selection. Crossover is difficult to treat analytically. One can treat mutation, though, in a simplified way:
- Consider a GA, applied to the Onemax function where, with probability p_μ , *exactly* one gene mutates.



Infinite population models

- In that case, one finds (Appendix B2.3.5):

$$p_2(j) = 2^{1-m} \left(\frac{j}{m} + p_\mu \frac{m-2j}{m^2} \right) \binom{m}{j}$$

- Here, selection has a positive effect for $j > m/2$, whereas mutation has a negative (immediate) effect.
- At some point the effects balance each other out .

Today's learning goals

- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications 
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs. 
 - Derive expressions for the expected running time for a simple GA.
 - Derive the optimal mutation rate for a simple GA.
 - Explain the concept of premature convergence
 - List methods for avoiding premature convergence

Expected running time

- Consider a very simple GA with a single individual, which is modified by mutation only, such that the new individual is kept if it is better than the old one.
- Let $p_{\text{mut}} = k/m$, where $k \ll m$.
- Apply this GA to the Onemax problem.
- The expected running time (number of evaluations L) is then (pp. 181-182):

$$E(L) \approx e^k \frac{m}{k} \ln \frac{m}{2}$$

Today's learning goals

- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications ✓
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs. ✓
 - Derive expressions for the expected running time for a simple GA. ✓
 - Derive the optimal mutation rate for a simple GA.
 - Explain the concept of premature convergence
 - List methods for avoiding premature convergence

Optimal mutation rate

- Consider the same simple GA as in the runtime computation, and the same (Onemax) problem.
- Using the equation for the probability of an improvement:

$$P(l, p_{\text{mut}}) = (1 - p_{\text{mut}})^{m-l} (1 - (1 - p_{\text{mut}})^l),$$

... one obtains (see pp. 182-183) $p_{\text{mut}}^* = \frac{1}{m}$.

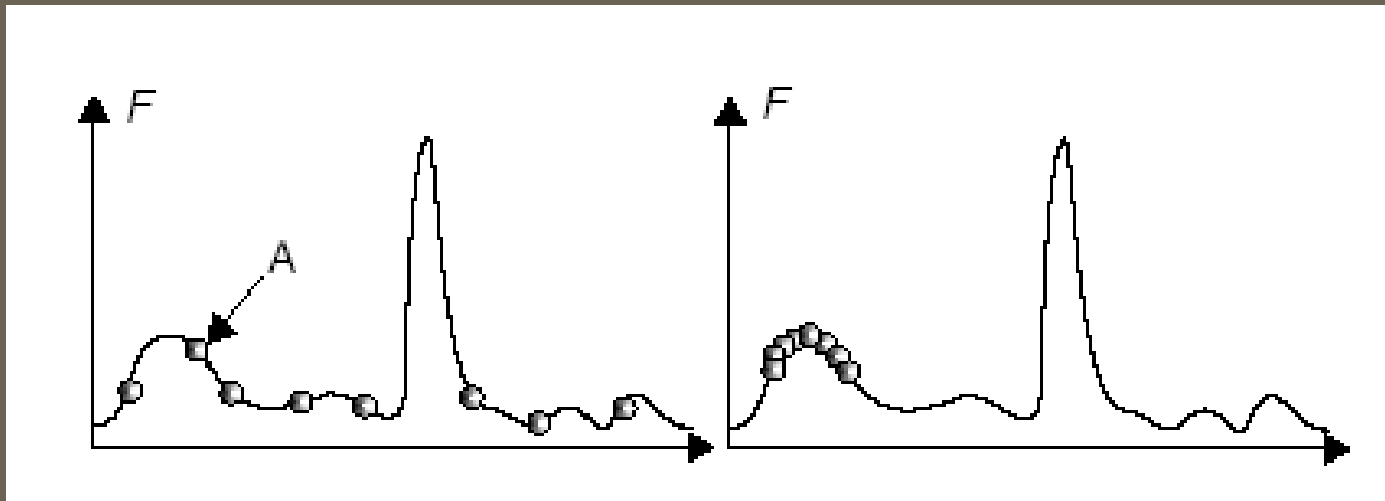
- This mutation rate (one or a few times $1/m$) typically works well for most fitness functions (with binary chromosomes).

Today's learning goals

- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications ✓
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs. ✓
 - Derive expressions for the expected running time for a simple GA. ✓
 - Derive the optimal mutation rate for a simple GA. ✓
 - Explain the concept of premature convergence
 - List methods for avoiding premature convergence

Premature convergence

- Since GAs are very efficient in their search for an optimum, they may get stuck at a local optimum, a phenomenon known as **premature convergence**:



Today's learning goals

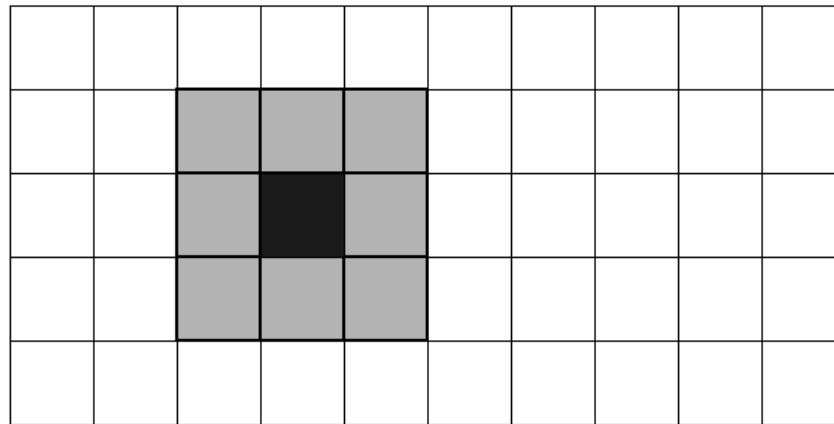
- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications ✓
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs. ✓
 - Derive expressions for the expected running time for a simple GA. ✓
 - Derive the optimal mutation rate for a simple GA. ✓
 - Explain the concept of premature convergence ✓
 - List methods for avoiding premature convergence ✓

Premature convergence

- Premature convergence can be avoided in many different ways:
 - With fitness ranking (if roulette-wheel selection is used, not needed if tournament selection is used!)
 - Reducing the crossover probability,
 - Using varying mutation rates (see pp. 69-71; Fig. 3.15),
 - Introducing mating restrictions (e.g. diffusion models).

Diffusion models

- Place the individuals on an imaginary grid.
- For any selected individual, allow mating only with one of its neighbors:



Premature convergence

- However, an alternative approach is simply to restart the GA, with a different random number sequence.
- It is often a good idea to make a few short trial runs to find good parameter settings; see e.g. Tables 3.1-3.3.

Today's learning goals

- After this lecture you should be able to
 - Derive and explain the schema theorem, and its implications ✓
 - Derive expressions for the result of selection and mutation in infinite-population models of GAs. ✓
 - Derive expressions for the expected running time for a simple GA. ✓
 - Derive the optimal mutation rate for a simple GA. ✓
 - Explain the concept of premature convergence ✓
 - List methods for avoiding premature convergence ✓