

Exact Schema Theorems for GP with 1-point Crossover

Case for representation as a Bernoulli Trial [1]

- Basic objective of schema theorems is to provide a description for how the population schema membership varies over time.
- As we are already aware, the search operators are capable of both creating and removing instances of a particular schema from the population.
- Generally,
 - Crossover may be described in terms of one of two processes,
 - Reproduction – individuals already possessing schema H are copied.
 - Crossover – children may or may not possess schema H of the parents, however, they also have the potential to introduce schema not currently sampled in the population or alter the population schema sampling.
 - Mutation
 - May result in schema removed by crossover being reintroduced;
 - Might remove (or not) a schema created by crossover;
 - Introduce a completely new schema.
- Schema theorems provide a basis for iteratively expressing properties of a population, say proportion of individuals possessing a given schema, in terms of the current generation.
 - Properties usually expressed as an expected value, $E[\cdot]$, of the proportion/ count, $m(\cdot)$, of schema H , at the next generation.
 - $E[m(H, t + 1)]$
- Question:
 - Can we investigate the properties of the underlying distribution governing the expectation without having to provide a definitive analytical expression for the probability of schema H survival?
- Consider:
 - Irrespective of how the survival of schema H is actually described there is only one outcome when creating new members for the next population,
 - Schema H either survives (is propagated) or it does not!

- Assumption,
 - Selection/ crossover/ mutation are independent processes;
- Implication,
 - Chance of children matching schema H is the same.
- Such an assumption defines a binary event, $\alpha(H, t)$, denoting the probability that a child samples H , or the *total transmission probability*.
- Such an event is modeled by a Bernoulli trial¹, and therefore conforms to a binomial probability density function.
- Implication – a binomial distribution immediately provides for the following statements,
 - The number of instances of schema H at the next generation, $m(H, t + 1)$, is modeled by,

$$P(m(H, t + 1) = k) = \binom{M}{k} \alpha(H, t)^k (1 - \alpha(H, t))^{M-k}$$

- Where M is the population size.
- The mean and variance of such a schema now also naturally follow from the properties of the binomial distribution,

$$E[m(H, t + 1)] = M\alpha(H, t)$$

$$\text{Var}[m(H, t + 1)] = M\alpha(H, t) (1 - \alpha(H, t))$$

- It also follows that the signal-to-noise ratio for schema H can be defined as the ratio of mean to standard deviation ($\text{SD} = \sqrt{\text{variance}}$)

$$\left(\frac{S}{N} \right)^{\text{def}} = \sqrt{M} \sqrt{\frac{\alpha(H, t)}{1 - \alpha(H, t)}}$$

- Has the potential to enable studies on the convergence properties of GAs [2] or the likelihood of schema extinction or growth [1].
- There we will concentrate on the problem of estimating $\alpha(H, t)$ exactly.

¹ You may have come across this when modeling a coin toss

Exact Schema Theory for GA [4]

- Approach taken by Poli is to develop a microscopic (exact) schema theorem utilizing the concept of “total schema transmission probability,” $\alpha(H, t)$.
- This is not the only approach. A tautology of schema models, figure 1², summarizes the cross-section of approaches in terms of,
 - Degree of approximation (exactness) in the model;
 - Granularity of the predicted quantity (left hand side of schema equation);
 - Granularity of the elements utilized to make the prediction (right hand side of schema equation).
- The approach of Poli will provide an Exact schema theory for,
 - *Microscopic* – program specific properties, and then;
 - *Macroscopic* – properties of schema.
- For clarity in the following we concentrate on the specific case of fixed length binary GAs without mutation.
- The first major result for exact schema theorems was derived by Stephens and Waelbroek [4].
 - Summarized briefly as follows in order to provide the basis for the more general results of Poli [3, 5].
- For the special case of fixed length GA and 1-point crossover, recognize that it is necessary to express the action of crossover as either,
 - P (schema H reproduced without crossover);
 - P (schema H survives crossover or is created by crossover).

Case of reproduction without crossover

- From the last lecture,

$$P(H, t) \equiv \frac{m(H, t)f(H, t)}{M\bar{f}(t)}$$

- Describes the rate of reproduction without crossover (or mutation)
- Thus,
 - P (schema H reproduced without crossover) = $(1 - p_{crossover})P(H, t)$

² From GECCO 2002 Tutorial Program, Foundations of Genetic Programming, Langdon W.B., Poli R. July 10.

Case of schema surviving or being introduced by crossover

- Principle observation that Stephens and Waelbroek make is that for one point crossover, schema H only survive if the,
 - contribution from LHS of crossover point *and*;
 - contribution from RHS of crossover point, *match* schema H .
- The probability of this happening is estimated by considering all possible crossover points.
- Thus, let
 - $L(H, i)$ denote the schema from instance H with “don’t care” symbols replacing genes at index $i + 1$ to N ;
 - $R(H, i)$ denote the schema from instance H with “don’t care” symbols replacing genes at index 1 to i ;
- E.g. for $H = **110$

Then

i	$L(H, i)$	$R(H, i)$
1	*****	**110
2	*****	**11*
3	**1**	**1**
4	**11*	*****
5	**110	*****

- $P(K, t)$ is the selection probability of a schema K at generation t .
- Stephens and Waelbroek demonstrate that the total transmission probability for a fixed length binary GA is [4],

$$\alpha(H, t) = (1 - p_{crossover}) + \frac{p_{crossover}}{N} \sum_{i=0}^{N-1} P(L(H, i), t) P(R(H, i), t)$$

Example

- For $H = 01^*$ what is the total transmission probability?

$$\alpha(01^*, t) = (1 - p_{crossover}) P(01^*, t) + \frac{p_{crossover}}{3} \left\{ \begin{aligned} &P(***, t) P(01^*, t) + P(0**, t) P(0**, t) \\ &+ P(01^*, t) P(***, t) \end{aligned} \right\}$$

- but $P (**, t) = 1$, so
- $\alpha(01^*, t) = (1 - \frac{P_{crossover}}{3}) P (01^*, t) + \frac{P_{crossover}}{3} P (0^{**}, t) P (0^{**}, t)$

Exact Schema Theory for GP with fixed size-and-shape population and 1-point Crossover

- Approach the problem in two steps,
 - Can we develop a GP schema theory from the special case where all individuals are of the *same* size and shape?
 - If so can we then generalize this result to programs with *different* size and shape?
- Poli [3, 5] uses the previous result of Stephens and Weadbroeck [4] to describe the total transmission probability along the same general lines, i.e.
 - $\alpha(H, t) = P$ (schema H reproduced)
 $+ P$ (schema H survives crossover or created by crossover)

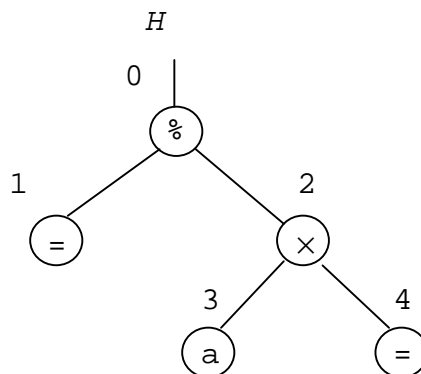
Case 1 – P (schema H reproduced)

- Takes the exactly the same form as developed above

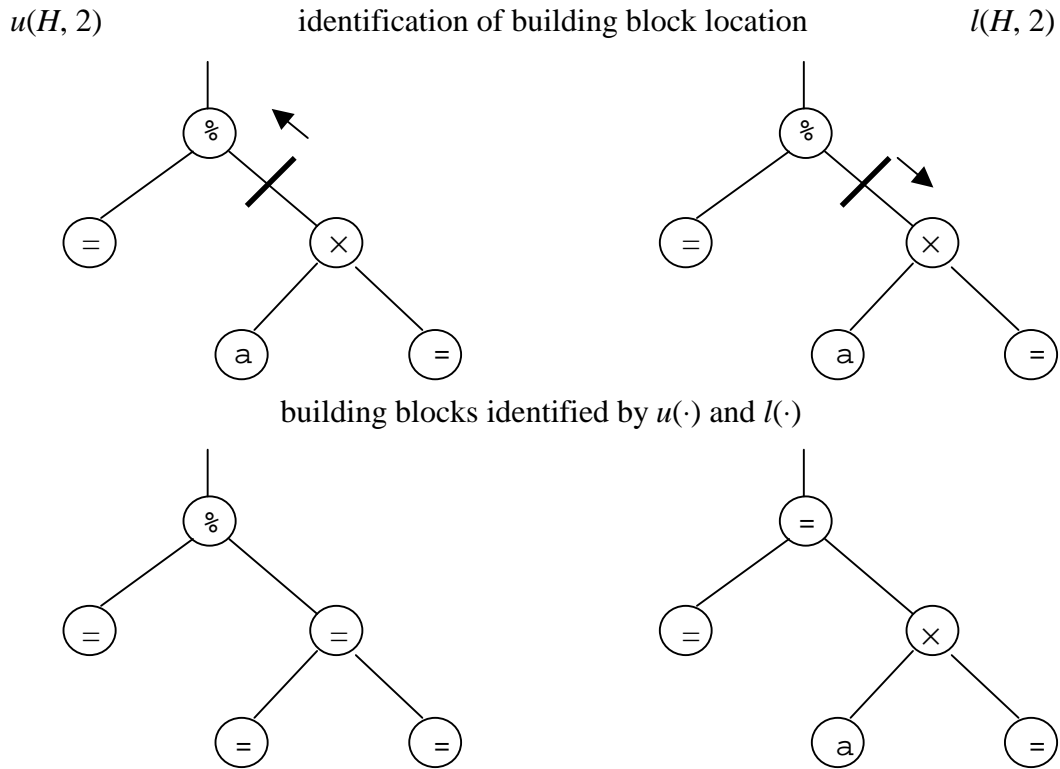
Case 2 – P (schema H survives crossover or created by crossover)

- There are $N(H)$ crossover points, e.g.,

For individual ($\% = (\times a =)$), corresponding index points are $\{0, 1, 2, 3, 4\}$, or



- Schema H is ‘duplicated’ iff
 - First parent, h_1 , has correct lower region with respect to $l(\cdot)$ crossover point;
 - Second parent, h_2 , has correct upper region with respect to $u(\cdot)$ crossover point;
 - See following figure, or Figure 3 [5].



- Let i denote crossover index,
- For schema H to survive or be created, then
 - $l(H, i), u(H, i) \in H$
- The total probability of survival / creation for schema H under any selection of the crossover point i is given by,

$$\frac{P_{crossover}}{N(H)} \sum_{i=0}^{N(H)-1} P(l(H, i))P(u(H, i))$$

- Or,

$$\alpha(H, t) = (1 - p_{crossover}) P(H, t) + \frac{P_{crossover}}{N(H)} \sum_{i=0}^{N(H)-1} P(l(H, i))P(u(H, i))$$

- Comment,
 - Fine as far as it goes, but in order to generalize this result to the case of varying size and shape individuals, the schema needs to be a little more flexible.

Hyperschema [3]

- In order to generalize this result to the case of populations with individuals with different sizes and shapes, but still using 1-point crossover, the (rooted tree) schema definition requires extension to include arbitrary size trees.
- To do so, introduce an additional “don’t care” symbol, ‘#’, this time able to represent any subtree.
- Note,
 - ‘#’ supports the description of programs with arbitrary shape and size.
 - The resulting ‘hyperschema’ are still rooted Tree-schema
 - Not possible to place a # between a terminal and the root node.
- *Definition 1* – GP Hyperschema [3]
 - “A GP hyperschema is a rooted tree composed of internal nodes from the set $F \cup \{=, \#\}$ and leaves from $T \cup \{=, \#\}$, where F and T are the function set and the terminal set used in a GP run. The operator = is a “don’t care” symbol which stands for exactly one node, while the operator # stands for any valid subtree.”
- Example
 - (+ = (% x #)) is the hyperschema denoting all programs with
 - a) Root node of addition;
 - b) 1st argument of root is any valid terminal;
 - c) 2nd argument of root is the % operator;
 - d) 1st argument of % is the terminal x;
 - e) 2nd argument of % is any legal subtree.

Microscopic exact GP schema theorem [3]

- A microscopic schema theorem considers the details of individual schema from generation to generation.
 - Provide such a schema by defining the concept of lower order schema in the same way as above.
 - The ensuing schema theory still requires fixed point crossover, but does not constrain all trees to the same size and shape.
- To this end let,

- $L(H, i)$ denote the hyperschema defined by replacing all nodes between crossover point i and root with '=' and any subtrees connected to a '=' with '#';
 - $U(H, i)$ denote the hyperschema defined by replacing the *subtree* following crossover point i with '#'.
 - And the following condition holds,
 - IF the crossover point, i , resides within the common region of both parents, but is not within the region of schema H ,
 - THEN $L(H, i), U(H, i) = \{\emptyset\}$
- Example,

$$\begin{array}{rcccccc}
 i & & 0 & & 1 & & 2 & & 3 & & 4 \\
 H = & & (& \times & = & (& + & \times & = &) &) \\
 L(H, 1) & \rightarrow & (= & = & \#) \\
 U(H, 1) & \rightarrow & (\times & \# & (+ & \times & =)) \\
 L(H, 3) & \rightarrow & (= & \# & (= & \times & \#)) \\
 U(H, 3) & \rightarrow & (\times & = & (+ & \# & =))
 \end{array}$$

- Where this also has a pictorial interpretation, Figure 3 [3], or Figure 4 [5].

Methodology:

- Given a working concept for how to define lower order schema, $L(\cdot)$ and $H(\cdot)$, we are now in a position to formulate $\alpha(H, t)$.
- The total transmission probability needs to encompass,
 - All possible ways in which crossover may be applied per parent;
 - Probability that both parents upper and lower are compatible with schema H .
- *Derivation* [3]:

a) Define a measurement function, $\delta(h \in H) = \begin{cases} 1 & \text{if children} \in H \\ 0 & \text{otherwise} \end{cases}$

- a. $\delta(h_1 \in L(H, i)) \times \delta(h_2 \in L(H, i))$ therefore returns 1 if both parents contain the required lower order schema necessary to build schema H and 0 otherwise.

- b) Let $P(h_1, h_2, i, t)$ denote the conditional probability that crossover point i is selected, given parents h_1, h_2 and the probability that both parents h_1, h_2 are selected at generation t , or

$$= P(i | h_1, h_2) P(h_1, t) P(h_2, t)$$

- c) The proportion of times that the offspring of h_1 and $h_2 \in H$ may now be modeled as the expectation of the sum of: parents and crossover points from (b), modulated by the schema membership from (a) over all combinations of parents and common crossover points³ $i \in C(h_1, h_2)$, or

$$\sum_{h_1} \sum_{h_2} \sum_{i \in C(h_1, h_2)} \delta(h_1 \in L(H, i)) \delta(h_2 \in U(H, i)) \times P(i | h_1, h_2) P(h_1, t) P(h_2, t)$$

- d) For the 1-point crossover to function,

$P(i | h_1, h_2)$ is always in the common region shared between each parent, $i \in C(h_1, h_2)$, or

$$P(i | h_1, h_2) = \frac{1}{NC(h_1, h_2)}$$

i.e. all crossover points within the shared region are equally possible.

- e) Substituting (d) into (c) provides theorem 1.

Theorem 1 – Microscopic Exact GP Schema Theorem [3]

- “The total transmission probability for a fixed-size-and-shape GP schema H under 1-point crossover and no mutation is

$$\alpha(H, t) = (1 - p_{crossover}) P(H, t) + p_{crossover} \sum_{h_1} \sum_{h_2} \frac{P(h_1, t) P(h_2, t)}{NC(h_1, h_2)} \times \sum_{i \in C(h_1, h_2)} \delta(h_1 \in L(H, i)) \delta(h_2 \in U(H, i))$$

where the first two summations are over all the individuals in the population;

- $NC(h_1, h_2)$ is the number of nodes in the tree fragment representing the common region between program h_1 and program h_2 ;
- $C(h_1, h_2)$ is the set of indices of the crossover points in the common region;
- $\delta(x)$ is a function which returns 1 if x is true, 0 otherwise;
- $U(H, i)$ and $L(H, i)$ are defined as above.”

³ We are still using 1-point crossover, so crossover points are common, hence limited to the common region of parents h_1 and h_2 .

- Observation

- Under the restriction that only individuals with the same size and shape as H are included in the first two summations of *Theorem 1*, then [3]
- $h_1, h_2 \in G(H) \rightarrow NC(h_1, h_2) = N(H)$, and;
 $\rightarrow C(h_1, h_2) = \{1, 2, \dots, N(H) - 1\}$

$$\alpha(H, t) \geq (1 - p_{crossover})P(H, t) +$$

$$\frac{p_{crossover}}{N(H) - 1} \sum_{i=1}^{N(H)-1} \sum_{h_1 \in G(H)} P(h_1, t) \delta(h_1 \in L(H, i)) \sum_{h_2 \in G(H)} P(h_2, t) \delta(h_2 \in U(H, i))$$

- Where this has an equivalent form, providing theorem 2.

Theorem 2 – GP Schema Theorem with Schema Creation Correction [3]

- “A lower bound for the total transmission probability for a fixed-size-and-shape GP schema H under one-point crossover and no mutation is

$$\alpha(H, t) \geq (1 - p_{crossover})P(H, t) + \frac{p_{crossover}}{N(H) - 1} \sum_{i=1}^{N(H)-1} P(L(H, i) \cap G(H), t) P(U(H, i) \cap G(H), t)$$

- the equality applying when all the programs in the population sample $G(H)$.”
- Consider,
 - In the previous lecture we identified an expression for the GP schema theorem in terms of a lower bound, which only incorporated disruptive effects.
 - Given such a lower bound and the above expression incorporating schema creation under the same constraints to schema shape and size, can we identify the factors contributing to schema creation?

- Method

- Dividing the lower bound by ‘ M ’ provides a lower bound for total transmission probability.
- For equivalence, also let $p_{mutate} = 0$ and $p_{diff}(t) = 1$;
- The lower bound with disruption alone now has the form,

$$\alpha(H, t) \geq P(H, t) \left\{ 1 - p_{crossover} \left[1 - P(G(H), t) + \frac{L(H)}{N(H) - 1} (P(G(H), t) - P(H, t)) \right] \right\}$$

- It follows that the difference between the above lower bound with disruption alone and theorem 2 is now,

$$\Delta\alpha(H,t) = \frac{P_{crossover}}{N(H)-1} \sum_{i \in B(H)} (P(L(H,i),t)P(U(H,i),t) - P(H,t)^2)$$

- Where $B(H)$ denotes the index set for which the crossover point lies between defining nodes of schema H .
- Observation,
 - $L(H, i) \subset H$ and $U(H, i) \subset H$; as a consequence, both contain more than schema H , thus
 - $P(L(H, i), t) \geq P(H, t)$ and $P(U(H, i), t) \geq P(H, t)$
 - and $\Delta\alpha(H, t) \geq 0$
 - The above expression for $\Delta\alpha(H, t)$ therefore represents the creative contributions from crossover.

Macroscopic Schema Theorem [5]

- The above microscopic formulation, although exact is a function of properties of individuals.
- A macroscopic formulation would express properties of schema and possibly provide better insight to the propagation of schema, as opposed to individuals.
- Given that a working microscopic schema theorem already exists, it should be possible to transform this directly into an equivalent macroscopic theorem.
- Consider [5],
 - The set of order zero schema $\{G_1, G_2, \dots\}$ define a disjoint set of programs, the union of which cover the search space, or

$$\sum_j \delta(h_1 \in G_j) = 1$$

$$\text{and } \delta(h_1 \in L(H, i)) \equiv \sum_j \delta(h_1 \in L(H, i) \cap G_j) \equiv \sum_j \delta(h_1 \in L(H, i))\delta(h_1 \in G_j)$$

(likewise for the case of $U(H, i)$)

- If this is substituted into the expression for the microscopic schema theorem and the summations reorganized taking note of the conditions under which $h_i \in G_{j,k}$ and then the following theorem for macroscopic schema theory follows [5].

Theorem 3 – Macroscopic Exact GP Schema Theorem [5]

- “The total transmission probability for a fixed-size-and-shape GP schema H under one-point crossover and no mutation is,

$$\alpha(H, t) = (1 - p_{crossover})P(H, t) + p_{crossover} \sum_j \sum_k \frac{1}{NC(G_j, G_k)} \sum_{i \in (G_j, G_k)} P(L(H, i) \cap G_j, t) P(U(H, i) \cap G_k, t)$$

- where the sets $L(H, i) \cap G_j$ and $U(H, i) \cap G_k$ are either fixed-size-and-shape schemata or the empty set $\{\emptyset\}$.”

Summary

- Concept of “total transmission probability” or $\alpha(H, t)$ used to denote the survival of schema in terms of a binary event, it either survives or it does not.
 - Immediately enables us to make predictions regarding general statistical properties of schema.
- Work of Stephens and Waelbroeck [4] provided concepts of,
 - Reproduction as survival without crossover;
 - Model for creative and disruptive effects of crossover in terms of schema H being ‘reproduced’ from left and right hand strings with respect to all string index positions.
- The concept of a hyperschema – *Definition 1* – provides the basis for introducing arbitrary subtrees into rooted tree schema.
- *Microscopic exact schema – Theorem 1* – derived by Poli for an exact model of survival for an individual program at generation t .
 - Special case of same size-and-shape individuals – *Theorem 2* – Poli is able to explicitly identify the components responsible for schema creation with respect to the previous lower bound.
- *Microscopic exact schema – Theorem 1* – derived by Poli for an exact model of survival for schema at generation t .
- Such models will provide the basis for the formal analysis of code bloat in GP.

References⁴

1. Poli R., Langdon W.B., O'Reilly U.-M. (1998) "Analysis of Schema Variance and Short Term Extinction Likelihoods," 3rd Annual Conference on Genetic Programming, Koza J.R. et al., (eds), Morgan Kaufmann, pp 284-292.
2. Poli R. (2000) "Recursive Conditional Schema Theorem, Convergence and Population Sizing in Genetic Algorithms," Foundations of Genetic Algorithms 6, Spears W., Martin W. (eds) pp 143-163.
3. Poli R., (2000) "Hyperschema Theory for GP with One-Point Crossover, Building Blocks, and some New Results in GA Theory," EuroGP'2000, Proceedings of the European Conference on Genetic Programming. Poli R., et al. (eds), Lecture Notes in Computer Science, Volume 1802, pp 163-180.
4. Stephens C.R., Waelbroeck H., (1999) "Schema Evolution and Building Blocks," Evolutionary Computation, 7(2), pp 109-124.
5. Poli R. (2000) "Exact Schema Theorem and Effective Fitness for GP with One-Point Crossover," Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2000, pp 469-476.

⁴ All are available from library reserve loans.