Analysis of Schema Variance and Short Term Extinction Likelihoods

Riccardo Poli
School of Computer Science
The University of Birmingham
Edgbaston
Birmingham, B15 2TT, UK
R.Poli@cs.bham.ac.uk
Phone: +44-121-414-3739

W.B. Langdon
School of Computer Science
The University of Birmingham
Edgbaston
Birmingham, B15 2TT, UK
W.B.Langdon@cs.bham.ac.uk
Phone: +44-121-414-4791

Una-May O'Reilly
Artificial Intelligence Lab
MIT
545 Tech Sq
Cambridge, MA, 02143
unamay@ai.mit.edu
Phone: +1-617-258-5978

ABSTRACT

This paper first analyses the impact of variance on schema transmission. Working from an exact derivation of the expected variance in schema transmission, it derives and analyses the signal-to-noise ratio for schemata. The paper then presents short term schema transmission probability results that focus on newly created schemata in the population. The analysis reveals the relative dependencies between schema transmission, population size, schema measured fitness, schema fragility and schema creation.

1 Introduction

While Holland's Schema Theorem (see [Holland, 1992] and [Goldberg, 1989]) in its original presentation and interpretation presents many technical difficulties, the GA and GP communities have gained a great deal of analytic knowledge by challenging it and coming to understand both its inadequacies and its content (see for example [Grefenstette, 1993] and [Altenberg, 1995]). As recently as December of 1997 schema understanding, in the context of crossover analysis and the building block hypothesis, was a topic of discussion on the genetic programming list [GP-List, 1997]. It seems that (rightfully) each researcher has a unique perspective on the usefulness of schema analysis, the validity of the building block hypothesis and the relationship between the two.

It is our position that schema analysis is useful for explaining what happens to subsolutions of different fitness and structure in the course of one generation. While a schema theorem makes a prediction (as an expectation) one generation time step in advance and only focuses on one schema at a time, schema processing accounts yield valuable insights.

Schema analysis can also be combined with experimentation to yield a substantiated and further improved account of GP [Poli and Langdon, 1997a].

Schema analysis does not lead to one "correct" GP schema theorem. Different GP schema definitions and formulations of schema theorems should not be seen as competitive but should be intellectually organised so as to distinguish the significant properties and issues of focus that each author tried to address [Poli and Langdon, 1998]. To date, GP schema research has offered different schema definitions [Koza, 1992, O'Reilly and Oppacher, 1995, Whigham, 1995, Poli and Langdon, 1997b, Rosca, 1997] and different Schema Theorems. Depending on whether the definition allows only one or more than one schema instance to appear in a program, the resulting schema theorem expresses a lower bound on either the expected number of programs sampling a schema or the expected number of instances of the schema in the population from one generation to the next [Poli and Langdon, 1998].

In this paper our results concerning schemata fall into two areas and are independent of a particular schema definition. The areas are: 1) results which derive the variance in the expected number of individuals sampling a schema and which examine the signal-to-noise ratio of a schema, and 2) probability results on the short term likelihood of schema instances being transmitted or becoming extinct.

In the first area, our focus is on analysing the second moment of a schema sample distribution as well as on the first moment. We start by formulating a Schema Variance Theorem expressing an estimate of the *variance* in the number of individuals sampling a schema that are propagated from one generation to the next. We extend this formulation, in contrast to other schema theorems, to account, not only for the propagation of a schema through the propagation of the individuals that sample it in the current population, but also for the additional schema samples that appear in the next generation because of spontaneous creation due to crossover. The derivation of the variance permits us to formulate the signal-

to-noise ratio of a schema's generation-based propagation and spontaneous creation in terms of a Schema Signal-to-Noise Ratio Theorem. We analyse it relative to population size as transmission probability is varied or decomposed.

The analysis of the short term schema transmission probability constitutes the second area of results. Our analysis focuses on the schemata which are sampled only once in the population. The rationale for this is as follows. The important schemata for the success of a run are those of above average fitness. New schemata come into existence when crossover produces a new individual. Often these are sampled by only one individual (for example towards the end of a run, when finding new above average individuals becomes very rare). Thus, it is important to analyse the extinction likelihood of a schema with only one sample. We derive the likelihood of schema extinction in one or two generations. The analysis suggests a reconsideration of fitness scaling techniques and the value of a steady state selection algorithm.

We now proceed in the following manner. In Section 2 we present the results related to variance analysis and signal-to-noise ratio. Section 3 states our derivations in standard schema theorem terms. In Section 4 we present the schema extinction likelihood results. In Section 5 we summarise our results and briefly comment on future work.

2 Variance and Signal-to-Noise Ratio Analysis

In this section derivations have the following sequence: (1) An individual h is considered. Under selection, the expected number of copies of h and the variance in the number of copies of h in the next generation is derived. (2) The individuals that sample the schema H are considered. Under selection the average number of individuals matching H and the variance in the number of individuals matching H in the next generation is derived. (3) Crossover is integrated into the expected value and variance derivations by including the probability that individuals sampling H still sample H after crossover. (4) An exact expression of the expected number of individuals matching H is derived by considering the probability that new samples of H are created. The total schema transmission probability is defined and the variance in the number of individuals matching H is derived. (5) The signal-to-noise ratio of the number of individuals sampling a schema is derived in terms of the total transmission probability.

2.1 Propagation of Individuals under Selection

Following the notation used in our previous work [Poli and Langdon, 1997b, Poli and Langdon, 1997a, Poli and Langdon, 1998], let us denote with $m(h, t + \frac{1}{2})$ the actual number of copies of individual h in the mating pool produced after selection at generation t. If fitness proportionate selection is used then

$$E[m(h, t + \frac{1}{2})] = \frac{f(h)}{\bar{f}(t)}$$

where $f(h)$ is the fitness of h and $\bar{f}(t)$ is the mean fitness of the individuals in the population. Each selection step can be modelled as a Bernoulli trial (either an individual is selected or it is not). So, $m(h, t + \frac{1}{2})$ will follow a binomial distribution with $p = p(h, t)$ and $q = 1 - p(h, t)$, where $p(h, t) = \frac{f(h)}{\bar{f}(t)M}$ is the selection probability of individual h and M is the population size. So,

$$Var[m(h, t + \frac{1}{2})] = Mp(h, t)\,(1 - p(h, t))$$

This is a well known result (see for example [Goldberg et al., 1991, page 14]).

2.2 Propagation of Schemata under Selection

Now let us concentrate on a subset of individuals at generation t: those matching a schema H. The number of individuals sampling H in the mating pool is

$$m(H, t + \frac{1}{2}) = \sum_{h \in H} m(h, t + \frac{1}{2})$$

where the summation is performed over all the individuals in the population matching the schema H at generation t. As a result

$$E[m(H, t + \frac{1}{2})] = \sum_{h \in H} E[m(h, t + \frac{1}{2})] = Mp(H, t),$$

where $p(H, t)$ is the schema selection probability, defined as

$$p(H, t) = \sum_{h \in H} p(h, t).$$

Each application of selection can be seen as a Bernoulli trial where each individual inserted in the mating pool either samples H or it does not. So, $m(H, t + \frac{1}{2})$ is a binomial stochastic variable with $p = p(H, t)$, whereby

$$Var[m(H, t + \frac{1}{2})] = Mp(H, t)\,(1 - p(H, t))$$

2.3 Survival of Schemata under Crossover

Crossover will cause some of the individuals matching H in the mating pool to produce offspring not sampling H. If we call $p_s(H, t)$ the probability that individuals in H will survive crossover (in the sense that their offspring will still sample H) and we assume that the only possible outcomes of a crossover

operation are schema disruption or schema survival, we can compute the expected number of individuals in H in the next generation

$$
\begin{aligned}
E[m(H, t+1)] &= p_s(H, t)E[m(H, t+\frac{1}{2})] \\
&= Mp_s(H, t)p(H, t).
\end{aligned} \tag{1}
$$

Note that, in general, $p_s(H, t)$ depends not only on the characteristics of the particular schema considered but also on the contents of the rest of the population (hence the dependence on t). In the following, schemata with relatively large survival probability will be qualified as *robust*, while schemata with small survival probability will be called *fragile*. Although we will consider only the effects of crossover in the rest of the paper, it is worth noting that most of the results presented below remain valid if we consider $p_s(H, t)$ to represent the probability that individuals in H will survive mutation or any combination of mutation and crossover.

An alternative way of looking at the selection/crossover process is to consider it a Bernoulli trial (as, again, each time an individual is produced, either the individual samples H or it does not). So, we can compute the variance of the stochastic variable $m(H, t+1)$

$$
Var[m(H, t+1)] = Mp_s(H, t)p(H, t)(1-p_s(H, t)p(H, t)). \tag{2}
$$

2.4 Schema Creation

If we term $p_c(H, t)$ the probability that offspring which sample H are *created* by parents not sampling H, then the expected number of individuals sampling a schema H at generation $t + 1$ can be expressed *exactly* as

$$
\begin{aligned}
E[m(H, t+1)] &= p_s(H, t)E[m(H, t+\frac{1}{2})] + \\
&\quad p_c(H, t)(M - E[m(H, t+\frac{1}{2})]) \\
&= M\Big[p_s(H, t)p(H, t) + \\
&\quad p_c(H, t)\Big(1 - p(H, t)\Big)\Big]
\end{aligned}
$$

Again, $m(H, t+1)$ can be seen as a binomial stochastic variable. Therefore, if we define the total schema transmission probability

$$
\alpha = p_s(H, t)p(H, t) + p_c(H, t)(1 - p(H, t)),
$$

we can write

$$
E[m(H, t+1)] = M\alpha \tag{3}
$$

and

$$
Var[m(H, t+1)] = M\alpha(1 - \alpha) \tag{4}
$$

which is a parabola with a maximum of $0.25M$ at $\alpha = 0.5$. Therefore the standard deviation $StdDev[m(H, t+1)]$ of any schema is at most $0.5\sqrt{M}$. For example, if $M = 10,000$,

$StdDev[m(H, t + 1)] \leq 50$ which is 0.5% of the population size. The total schema transmission probability α corresponds to the expected proportion of population sampling H at generation $t + 1$. Thus, Equation 3 can be seen as a reformulation of the "gain and losses" equation in [Whitley, 1992].

2.5 Signal-to-Noise Ratio

We define signal-to-noise ratio as the division of the mean of a distribution by its standard deviation: $\frac{\mu}{\sigma}$.[1] It expresses a second order property of the distribution as one value. The distribution of relevance to schema processing is the samples of a schema in the next generation. When the signal-to-noise ratio is infinite, one expects that the propagation of the analysed schema will occur exactly as predicted. When the signal-to-noise ratio is less than one, one expects that the actual number of instances of a schema in the next generation is essentially random with respect to its current status.

From Equations 3 and 4 we can define the signal-to-noise ratio for schema H as

$$
\left(\frac{S}{N}\right) \overset{def}{=} \frac{E[m(H, t+1)]}{\sqrt{Var[m(H, t+1)]}},
$$

obtaining:

$$
\left(\frac{S}{N}\right) = \sqrt{M}\sqrt{\frac{\alpha}{1 - \alpha}} \tag{5}
$$

which is a monotonic increasing function of α.

For $\alpha \ll 1$ this equation becomes

$$
\left(\frac{S}{N}\right) \approx \sqrt{M\alpha}
$$

while

$$
\lim_{\alpha \to 1}\left(\frac{S}{N}\right) = \infty
$$

Figure 1 shows a plot of $\left(\frac{S}{N}\right)/\sqrt{M}$.

An interesting special case of Equation 5 is when $p_c(H, t) = 0$, thus obtaining:

$$
\left(\frac{S}{N}\right) = \sqrt{M}\sqrt{\frac{p_s(H, t)p(H, t)}{1 - p_s(H, t)p(H, t)}} \tag{6}
$$

3 Schema Theorems

Equations 1, 2 and 6 can be formulated in the manner of a general schema theorem and the corresponding schema variance and schema signal-to-noise ratio theorems. The probability $p(H, t)$ can be expressed as:

$$
p(H, t) = \frac{m(H, t)}{M}\frac{f(H, t)}{\bar{f}(t)}
$$

where $f(H, t)$ is the measured schema fitness.

---

[1] This differs from the definition of signal-to-noise ratio used by other authors, e.g. in [Kargupta, 1995]
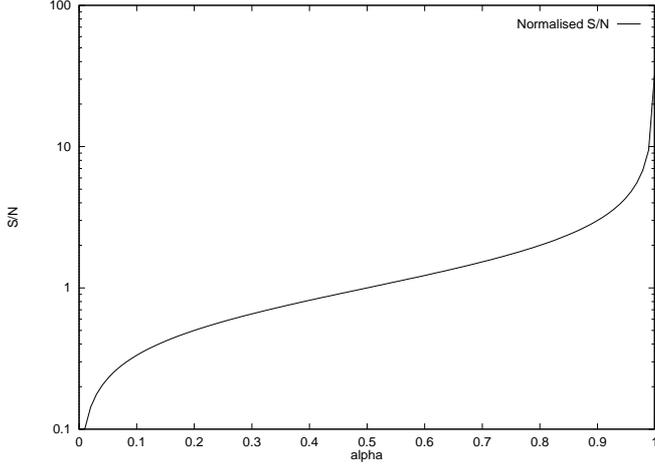
Figure 1: Normalised signal-to-noise ratio vs. total schema transmission probability.

However, Equations 1, 2 and 6 should be viewed as approximations because they do not account for a crossover event that generates an offspring sampling H from parents which did not sample H. In contrast to the established style of schema theorems, they are left as approximations rather than being derived into upper/lower bounds. On the contrary, Equations 3, 4 and 5 are entirely general. In the following we derive more general theorems formulating upper and lower bounds for schema mean, variance and signal-to-noise ratio by considering the best and worse case scenarios for the values of $p_s(H,t)$.

Let $p_{s_{\min}}(H,t)$ be the minimum of $p_s(H,t)$. This is normally known in each schema theory thanks to probabilistic calculations on the likelihood of schema disruption (see for example [Poli and Langdon, 1998]). For instance, for a simple GA with one-point crossover, with probability p_{XO}, and no mutation applied to binary strings of length N, $p_{s_{\min}}(H,t) = 1 - p_{XO}\frac{\mathcal{L}(H)}{N-1}(1 - p(H,t))$ where $\mathcal{L}(H)$ is the defining length of the schema H.

Thus a lower bound for α is

$$\alpha_{\min} = p_{s_{\min}}(H,t)p(H,t) + p_c(H,t)\Big(1 - p(H,t)\Big).$$

We can obtain an upper limit for α by considering selection without crossover and, by implication, assuming $p_s(H,t) = 1$. Therefore,

$$\alpha_{\max} = p(H,t) + p_c(H,t)(1 - p(H,t)).$$

These limits are functions of the value of $p_c(H,t)$. Using them it is possible to obtain lower and upper bounds for $E[m(H,t+1)]$, $Var[m(H,t+1)]$ and $\left(\frac{S}{N}\right)$.

For the expected value we obtain:

$$E_{\min} \leq E[m(H,t+1)] \leq E_{\max} \tag{7}$$

where

$$E_{\min} = M\alpha_{\min} \quad \text{and} \quad E_{\max} = M\alpha_{\max}.$$

For the variance we obtain:

$$V_{\min} \leq Var[m(H,t+1)] \leq V_{\max} \tag{8}$$

where

$$V_{\min} = \begin{cases} V_1 & \text{if } E_{\max} < 0.5M \\ V_2 & \text{if } E_{\min} > 0.5M \\ \min(V_1, V_2) & \text{otherwise} \end{cases}$$

$$V_{\max} = \begin{cases} V_2 & \text{if } E_{\max} < 0.5M \\ V_1 & \text{if } E_{\min} > 0.5M \\ \frac{M}{4} & \text{otherwise} \end{cases}$$

and

$$V_1 \stackrel{def}{=} E_{\min}(1 - \frac{1}{M}E_{\min}), \qquad V_2 \stackrel{def}{=} E_{\max}(1 - \frac{1}{M}E_{\max}).$$

The different cases in these equations are due to the non-monotonicity of $Var[m(H,t+1)]$ both with and without crossover.

Finally for the signal-to-noise ratio we obtain:

$$\left(\frac{S}{N}\right)_{\min} \leq \left(\frac{S}{N}\right) \leq \left(\frac{S}{N}\right)_{\max} \tag{9}$$

where

$$\left(\frac{S}{N}\right)_{\min} = \sqrt{M}\sqrt{\frac{\alpha_{\min}}{1 - \alpha_{\min}}}$$

$$\left(\frac{S}{N}\right)_{\max} = \sqrt{M}\sqrt{\frac{\alpha_{\max}}{1 - \alpha_{\max}}}$$

Equations 7, 8 and 9 are general schema mean, schema variance and schema signal-to-noise ratio theorems which include also the effects of schema creation.

3.1 Analysis of the Schema Signal-to-Noise Ratio Theorem

Figure 2 shows plots of $\left(\frac{S}{N}\right)/\sqrt{M}$ as a function of $p_c(H,t)$ and $p(H,t)$ for different values of $p_s(H,t)$. The uppermost plot can be interpreted as the maximum normalised S/N ratio $\left(\frac{S}{N}\right)_{\max}/\sqrt{M}$. This becomes very large when either $p_c(H,t)$ or $p(H,t)$ are large. In this situation sampling noise effects on the processing of robust schemata are limited even with small populations. On the contrary, if $p_c(H,t)$ and $p(H,t)$ are both small, the maximum normalised S/N ratio becomes quite small. This means that even the schemata less affected by crossover can suffer from severe sampling noise if the population is not large enough. The other plots in the figure represent the behaviour of progressively less robust schemata. They can be interpreted as representing the value of $\left(\frac{S}{N}\right)_{\min}/\sqrt{M}$ for different values of $p_{s_{\min}}(H,t)$. So, the value of the normalised S/N ratio will always stay between the "blankets" represented by the uppermost plot and one of these plots.
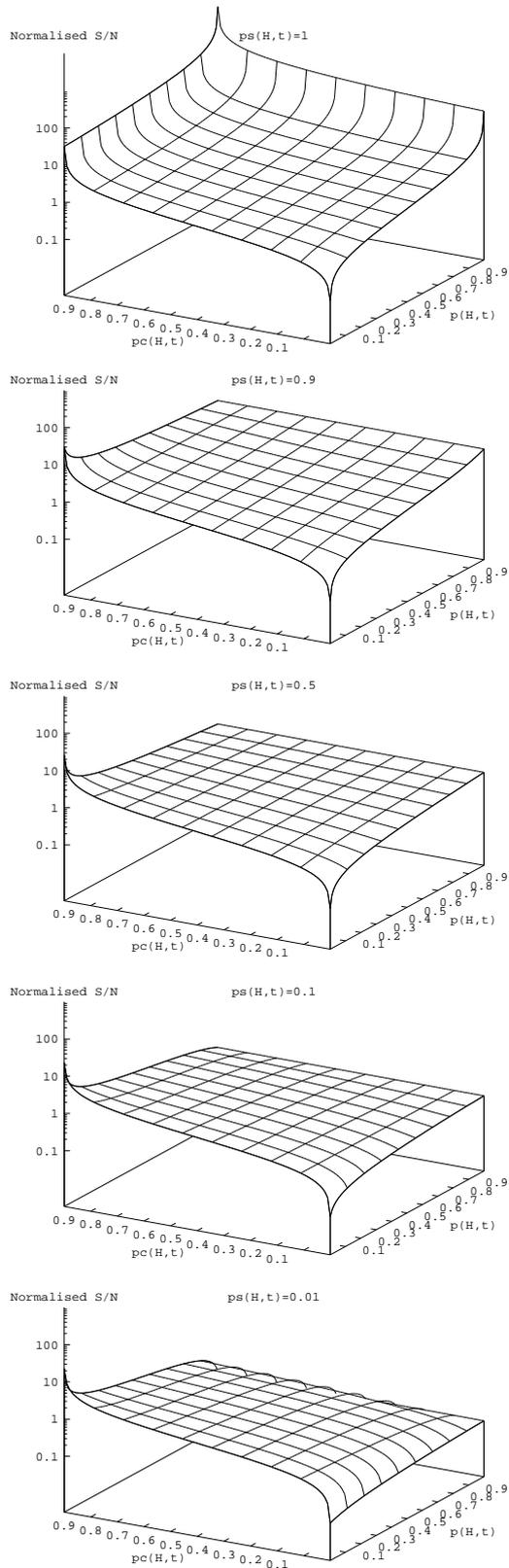
Figure 2: Normalised signal-to-noise ratio for different schema survival probabilities.

The bottommost plot shows a paradoxical effect which occurs with fragile schemata: the signal-to-noise ratio may get worse as the schema spreads in the population.

This effect is clearly shown in Figure 3 (at the end of the paper) which contains plots of $\left(\frac{S}{N}\right)/\sqrt{M}$ as a function of $p(H,t)$ for different values of $p_s(H,t)$ and $p_c(H,t)$ (these represent cross sections of the surfaces in Figure 2).[2]

These plots show another interesting phenomenon: when the schema selection probability is near zero (for example because the schema has only one instance in the population or because it is severely below average fitness), the signal-to-noise ratio depends very little on the robustness of the schema (all the plots have a common starting point for $p(H,t) = 0$). However, if also $p_c(H,t)$ is very small then, as soon as $p(H,t)$ starts growing, robust schemata will have significantly better S/N ratios than fragile ones.

4 Schema Extinction Likelihoods

For any reasonable population size, only a relatively tiny proportion of all possible schemata are present in the initial population. This implies that, until the population converges, crossover continues to introduce new schemata into the population. Often, a schema that is better than ones of the previous generation is identified when it appears in a single, novel individual of above average fitness. Thus, it is important to investigate the effects of noise on the ability of a newly created schema to survive and spread. A way of assessing this is to compute the probability of extinction for a new schema at the next generation.[3]

4.1 Schema Extinction in One Generation

As mentioned in Section 2.4, $m(H, t + 1)$ can be seen as a binomial stochastic variable. Therefore,

$$\Pr\{m(H, t+1) = k\} = \binom{M}{k}\alpha^k(1-\alpha)^{M-k}. \quad (10)$$

So, the probability of any schema becoming extinct at the next generation is

$$\Pr\{m(H, t+1) = 0\} = (1-\alpha)^M$$

as $\binom{M}{0} = 1$. So, the probability of extinction in one generation is a monotonic decreasing function of the total transmission probability of the schema. Since $(1-\alpha)^M = e^{M\log(1-\alpha)}$

---

[2] The right-bottom plot in the figure does not show the signal-to-noise ratio for $p_c(H,t) = 1$ and $p_s(H,t) = 1$ as all the points are at infinity.

[3] The term "extinction" is used here to indicate the situation in which a schema sampled by at least one individual in one generation is sampled by zero individuals at some point in future generations. This is different from a more biological interpretation of the term which would mean that all the instances of a schema are not propagated due to the disruptive effects of crossover. In fact, with our terminology a schema becomes extinct if all the existing instances of a schema are disrupted *and* no new instance is created by crossover between parents not sampling the schema.

we can write

$$\Pr\{m(H, t+1) = 0\} = e^{M \log(1-\alpha)}$$

So, in general, the probability of extinction goes quickly to zero as α grows, and for practical purposes there is no chance of losing a schema if $\alpha > 1 - e^{-\frac{4}{M}} \approx 4/M$ for any reasonable value of M.

For newly created schemata typically $m(H, t) = 1$ and

$$p(H, t) = \frac{f(H, t)}{M \bar{f}(t)} = \frac{f(h)}{M \bar{f}(t)} = p(h, t),$$

where h is the only individual in H at generation t. Therefore,

$$\alpha = p_s(H, t)p(h, t) + p_c(H, t)(1 - p(h, t)).$$

Let us now consider the worst case scenario, i.e. when schema creation is absent in the current generation ($p_c(H, t) = 0$).[4] In this case

$$
\begin{aligned}
\Pr\{m(H, t+1) = 0\} &= (1 - p_s(H, t)p(h, t))^M \\
&= \left(1 - \frac{p_s(H, t)f(h)}{M \bar{f}(t)}\right)^M
\end{aligned}
$$

If H is an average fitness schema (i.e. h is an average fitness individual such that $f(h) = \bar{f}$), $p(h, t) = 1/M$ and the previous equation simplifies further:

$$\Pr\{m(H, t+1) = 0\} = (1 - p_s(H, t)/M)^M \approx e^{-p_s(H, t)}$$

the final approximation being valid for any reasonable value of M. Newly created average fitness schemata have a probability of extinction (in one generation) of at least 37% ($1/e$) for any reasonable population size unless schema creation helps. If a schema is fragile this probability can actually become much bigger, thus emphasising even more the rescuing role of schema creation. Obviously better than average schemata will tend to suffer less from this effect. However, as emphasised in Figure 4, unless they are significantly better than average (e.g. $\alpha > 2/M$), they still have a remarkably high extinction probability.

The picture obtained from these observations is relatively gloomy. However, reality is even worse. Newly created schemata can become extinct in other ways: a) it is possible for a new schema to survive for one generation without getting additional individuals sampling it and then becomes extinct after a further generation, b) it is also possible that a new schema is sampled by two individuals in the next generation, but becomes extinct after a further generation, and so on. Obviously the more individuals a schema contains the smaller the probability of extinction. Therefore, we can obtain a good underestimate of the real extinction probability in two generations by considering case a) only.

---

[4]Obviously at some earlier generation $p_c(H, t)$ must have been different from 0.
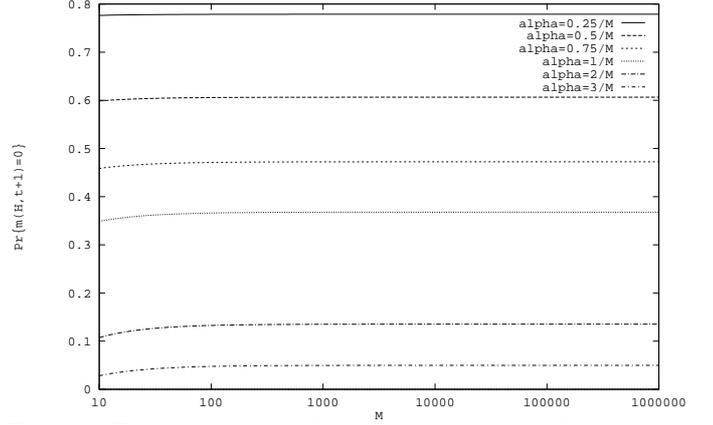

Figure 4: Extinction probability in one generation for newly created schemata as a function of the population size M, for different schema transmission probabilities α.

4.2 Schema Extinction in Two Generations

The probability of any schema being sampled by exactly one individual in the next generation is

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

as $\binom{M}{1} = M$. As $(1-\alpha)^{M-1} = \frac{1}{1-\alpha}e^{M \log(1-\alpha)}$, we obtain

$$\Pr\{m(H, t+1) = 1\} = \frac{M\alpha}{1-\alpha}e^{M \log(1-\alpha)}.$$

From this equation it is easy to infer that, in the absence of schema creation, newly created average fitness schemata have a probability of being sampled by only one individual also in the next generation (thus barely surviving) of approximately $p_s(H, t)e^{-p_s(H, t)}$ (which is at most 37%) for any reasonable population size.

The probability of extinction in two generations for any schema is:

$$
\begin{aligned}
\Pr\{m(H, t+2) = 0\} &\approx \Pr\{m(H, t+1) = 0\} \\
&\quad + \Pr\{m(H, t+2) = 0 | m(H, t+1) = 1\} \\
&\qquad \cdot \Pr\{m(H, t+1) = 1\} \\
&= (1-\alpha)^M + M\alpha(1-\alpha)^{M-1}(1-\alpha(t+1))^M
\end{aligned}
$$

where $\alpha(t+1)$ is the value of α one generation ahead. If we assume that the features of the schema H and the mean population fitness do not change significantly in one generation (i.e. $\alpha(t+1) \approx \alpha$) then the probability of extinction in two generations can be approximated as:

$$
\begin{aligned}
\Pr\{m(H, t+2) = 0\} \\
\approx (1-\alpha)^M(1 + M\alpha(1-\alpha)^{M-1}) \\
= e^{M \log(1-\alpha)}\left(1 + \frac{M\alpha}{1-\alpha}e^{M \log(1-\alpha)}\right).
\end{aligned}
$$

In the absence of schema creation, $\alpha = p_s(H,t)f(h)/(M\bar{f}(t)) \ll 1$ for any reasonable population size. Thus, $\log(1 - \alpha) \approx -\alpha$. So, newly created schemata have a probability of becoming extinct in two generations of approximately

$$\Pr\{m(H,t+2) = 0\}$$

$$\approx e^{-\frac{p_s(H,t)f(h)}{\bar{f}(t)}} \left(1 + \frac{\frac{p_s(H,t)f(h)}{\bar{f}(t)}}{1 - \frac{p_s(H,t)f(h)}{M\bar{f}(t)}} e^{-\frac{p_s(H,t)f(h)}{\bar{f}(t)}}\right)$$

$$\approx e^{-\frac{p_s(H,t)f(h)}{\bar{f}(t)}} \left(1 + \frac{p_s(H,t)f(h)}{\bar{f}(t)} e^{-\frac{p_s(H,t)f(h)}{\bar{f}(t)}}\right).$$

Using Taylor series expansion about $\frac{p_s(H,t)f(h)}{\bar{f}(t)} = 1$ we obtain

$$\Pr\{m(H,t+2) = 0\} \approx 1.0064 - 0.5032 \times \frac{p_s(H,t)f(h)}{\bar{f}(t)}.$$

So, even very robust average fitness schemata will have *at least a 50% probability of extinction in two generations*. This suggests that in order for slightly above average schemata to survive for long enough to spread reliably in the population, creation must guarantee a reasonable supply of schemata for more than one generation. These results corroborate recent theoretical results on the importance of schema creation [Stephens and Waelbroeck, 1997].

4.3 Role of Fitness Scaling, Elitism and Steady-State GAs

The last equation highlights an interesting role of fitness scaling techniques. If we adopt a technique that amplifies the differences in fitness of the individuals in the population without altering $\bar{f}(t)$, then newly created above average fitness schemata will have a higher probability of survival. This observation is actually independent of the selection method used. The higher the selective pressure imposed, the better the chances of survival of new good schemata.

It is also interesting to consider the effect of elitism (where the best individual in each generation is always copied unchanged in the next). If a new best of generation individual is discovered, the newly created schemata it contains will be shielded from sampling noise and schema disruption until a better individual is discovered, thus allowing an easier propagation of highly fit schemata.

Extending this idea, it is also arguable that some types of steady state GAs can do much better than generational GAs in terms of chances of survival of newly created schemata. For example, if one uses a steady state GA where the worst individual in the population is replaced after each crossover operation, newly created above average schemata will survive for at least as long as they remain above average, i.e., as observed in [Syswerda, 1991], this form of steady state GA grants elitism status to all good members of the population.

5 Conclusions and Future Work

In this paper we have obtained, for the first time, a schema variance theorem which expresses the variance in the number of individuals sampling a schema at the next generation in terms of the information available on such a schema and the current population. Using this and a reformulation of the standard schema theorem, we obtained a schema signal-to-noise ratio theorem, which expresses how confident one can be of the schema theorem predictions of the expected growth or decay in the number of individuals sampling a schema. When the signal-to-noise ratio is less than one the actual number of instances of a schema in the next generation is essentially random. When the signal-to-noise ratio is infinite the propagation of schemata occurs exactly as predicted.

For each theorem we provide three versions: the first ones are approximation based on the assumption (used in past schema theorems) that no schema creation occurs (Equations 1, 2 and 6), the second ones are exact formulations which include schema creation effects (Equations 3, 4 and 5), the final ones give upper and lower bounds obtained using the minimum and maximum schema survival probabilities (Equations 7, 8 and 9).

The analysis of the second and third versions of the schema signal-to-noise ratio theorem has revealed two interesting phenomena. Firstly, the signal-to-noise ratio of very fragile schemata may get worse as they spread in the population. Secondly, when the schema selection probability is near zero (e.g. for newly created schemata with only one instance in the population) the signal-to-noise ratio depends very little on the robustness of the schema.

We have also studied the short-term extinction probability of schemata in general, with particular attention paid to newly created ones. This analysis is general and includes schema creation. In the special case where no creation is present, for any reasonable population size, a newly created schema of average fitness which is only sampled once has an extinction probability of at least 37% in one generation. This drastic extinction probability is decreased if a schema is better than average in fitness, but it becomes small only if the schema is several times better than average. The extension of the analysis to two generations shows that the extinction probability, in fact, increases. Even a very robust average fitness schema has at least a 50% probability of becoming extinct in the next two generations.

These and the previous results show the importance of considering schema creation in the analysis of schema propagation, as also indicated by other researchers in the field.

These findings also suggest a reconsideration of fitness scaling techniques, elitism and steady-state GAs as such techniques improve the chances of survival of new fit schemata. Future work, theoretical and experimental, will need to be devoted to understanding the effects of these techniques on schema processing.

More theoretical work is necessary to specialise the theory presented to different GP and GA schema theories, to deepen

our analysis, and to estimate the schema creation probability. Experiments will also be necessary to corroborate our results and to provide measures of schema creation and schema survival probabilities.

Acknowledgements

References

[Altenberg, 1995] Altenberg, L. (1995). The Schema Theorem and Price's Theorem. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms 3*, pages 23–49, Estes Park, Colorado, USA. Morgan Kaufmann.

[Goldberg, 1989] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Massachusetts.

[Goldberg et al., 1991] Goldberg, D. E., Deb, K., and Clark, J. H. (1991). Genetic algorithms, noise, and the sizing of populations. IlliGAL Report No 91010, Department of General Engineering, University of Illinois at Urbana-Champaign.

[GP-List, 1997] GP-List (1997). Genetic programming email list: genetic-programming@lists.stanford.edu.

[Grefenstette, 1993] Grefenstette, J. J. (1993). Deception considered harmful. In Whitley, L. D., editor, *Foundations of Genetic Algorithms 2*, San Mateo, CA. Morgan Kaufman.

[Holland, 1992] Holland, J. (1992). *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, Massachusetts, second edition.

[Kargupta, 1995] Kargupta, H. (1995). Signal-to-noise, crosstalk, and long range problem difficulty in genetic algorithms. In Eshelman, L. J., editor, *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 193–200, Pittsburgh, PA, USA. Morgan Kaufmann Publishers.

[Koza, 1992] Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.

[O'Reilly and Oppacher, 1995] O'Reilly, U.-M. and Oppacher, F. (1995). The troubling aspects of a building block hypothesis for genetic programming. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms 3*,

pages 73–88, Estes Park, Colorado, USA. Morgan Kaufmann.

[Poli and Langdon, 1997a] Poli, R. and Langdon, W. B. (1997a). An experimental analysis of schema creation, propagation and disruption in genetic programming. In Back, T., editor, *Genetic Algorithms: Proceedings of the Seventh International Conference*, pages 18–25, Michigan State University, East Lansing, MI, USA. Morgan Kaufmann.

[Poli and Langdon, 1997b] Poli, R. and Langdon, W. B. (1997b). A new schema theory for genetic programming with one-point crossover and point mutation. In Koza, J. R., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M., Iba, H., and Riolo, R. L., editors, *Genetic Programming 1997: Proceedings of the Second Annual Conference*, pages 278–285, Stanford University, CA, USA. Morgan Kaufmann.

[Poli and Langdon, 1998] Poli, R. and Langdon, W. B. (1998). A review of theoretical and experimental results on schemata in genetic programming. In Banzhaf, W., Poli, R., Schoenauer, M., and Fogarty, T. C., editors, *Proceedings of the First European Workshop on Genetic Programming*, Paris. Springer-Verlag.

[Rosca, 1997] Rosca, J. P. (1997). Analysis of complexity drift in genetic programming. In Koza, J. R., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M., Iba, H., and Riolo, R. L., editors, *Genetic Programming 1997: Proceedings of the Second Annual Conference*, pages 286–294, Stanford University, CA, USA. Morgan Kaufmann.

[Stephens and Waelbroeck, 1997] Stephens, C. R. and Waelbroeck, H. (1997). Effective degrees of freedom in genetic algorithms and the block hypothesis. In Bäck, T., editor, *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)*, East Lansing. Morgan Kaufmann.

[Syswerda, 1991] Syswerda, G. (1991). A study of reproduction in generational and steady state genetic algorithms. In Rawlins, G. J. E., editor, *Foundations of genetic algorithms*, pages 94–101, Indiana University. Morgan Kaufmann.

[Whigham, 1995] Whigham, P. A. (1995). A schema theorem for context-free grammars. In *1995 IEEE Conference on Evolutionary Computation*, volume 1, pages 178–181, Perth, Australia. IEEE Press.

[Whitley, 1992] Whitley, D. (1992). An executable model of a simple genetic algorithm. In Whitley, D., editor, *Foundations of Genetic Algorithms Workshop (FOGA-92)*, Vail, Colorado.
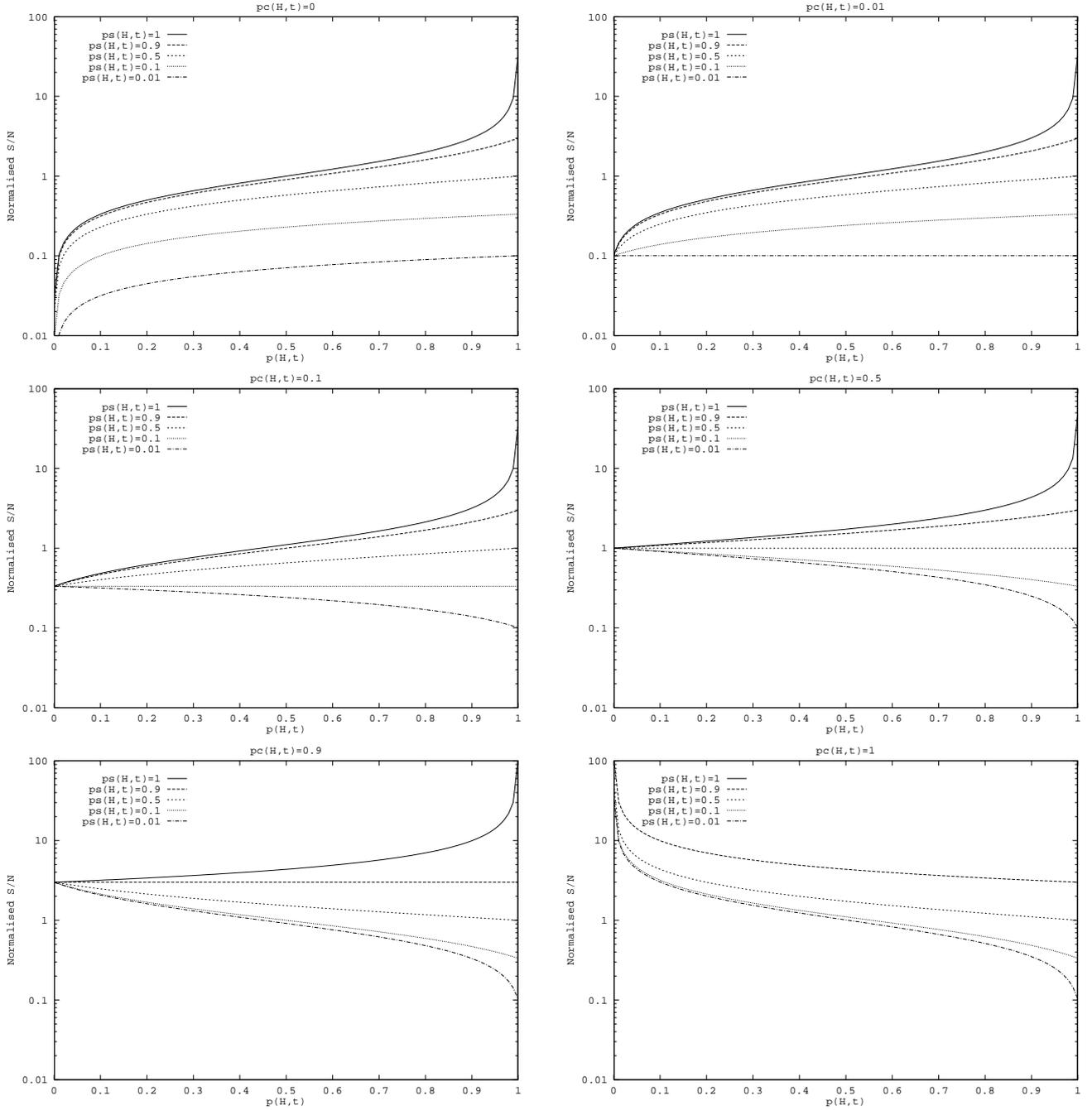
Figure 3: Normalised signal-to-noise ratio for different schema survival and schema creation probabilities.