Modelling the Dynamics of a Steady State Genetic Algorithm

Alex Rogers and Adam Prügel-Bennett
ISIS Research Group
Department of Electronics and Computer Science
University of Southampton
Highfield, Southampton, England
ar97r@ecs.soton.ac.uk
apb@ecs.soton.ac.uk

Abstract

A comparison is made between the dynamics of steady state and generational genetic algorithms using the statistical mechanics approach developed by Prügel-Bennett, Shapiro and Rattray. It is shown that the loss of variance of the population under steady state selection — genetic drift — occurs at twice the rate of generational selection. By considering a simple ones counting problem with selection and mutation, it is shown that, with weak selection, the steady state genetic algorithm can reproduce the dynamics of the generational genetic algorithm at half the computational cost in terms of function evaluations.

1 Introduction

Since the popularisation of the genetic algorithm (GA) by Holland [1], there have been two strategies for reproducing the population members. In one, the entire population is replaced simultaneously — the generational GA. In the steady state GA, populations overlap. One or two population members are reproduced in each time frame. The study of these two schemes dates back to De Jong's introduction of the term 'generation gap' in 1975 [2,3] to describe the size of the population overlap. Comparisons of the two types appear in the literature but are difficult as implementation differences often cause a change in selection pressures which masks the effect of the selection scheme [4,5]. A full understanding of the differences between the two selection schemes does not exist.

The formalisation developed by Prügel-Bennett, Shapiro and Rattray [6,7,8] for describing the dynamics of a generational GA provides the framework for a comparison of the two types of selection. In this formalisation, rather than follow the microscopic details of the evolving system, a statistical mechanics approach is taken and several macroscopic variables which describe the system are modelled. The macroscopic variables used are the cumulants of the fitness distribution. The first two cumulants of a distribution are familiar as the mean and variance. Higher cumulants describe the deviation of the distribution from a Gaussian — the third and fourth describing the skewness and kurtosis of the distribution.

In this paper, we develop models of simple generational and steady state GA which allow their selection schemes to be compared. For a simple fitness function based on ones counting of a binary string with mutation, we can solve the dynamics analytically and directly compare the dynamics of the two types. We find that the loss of variance of the population due to selection — genetic drift — in the steady state GA is twice as fast as that of the generational GA — a result observed by De Jong in terms of allele loss in an earlier paper [3]. By rescaling the parameters of the steady state GA, we can reproduce the dynamics of the generational GA but with half the computational cost in terms of function evaluations. This seems to give some indication of a possible performance improvement through using steady state selection. Whether this effect persists with the introduction of the crossover operator — the next stage of the work — is still open to investigation.

We start in section 2 by introducing the model genetic algorithm. In section 3 we calculate the dynamics of this model under steady state selection and compare these results to the generational case. Section 4 introduces a more realistic model with a ones counting fitness function and mutation. In the conclusion, section 5, we discuss the scaling effects found and its implications for the performance of the steady state GA.

Although the analysis of the dynamics involves some mathematics, the findings are all readily apparent in the simulations. Solving the dynamics analytically gives further insight into what is happening in the simulations.

2 The Model Steady State Genetic Algorithm

Initially a simple model of a steady state GA is considered where only selection is applied. We consider a population where each member has an assigned fitness, F. Later we shall consider a problem with a real fitness function but for the initial discussion the extra detail is not important. We only assume that the fitness function gives rise to some form of unimodal fitness distribution across the population.

We select a fit member, μ, from the population of P members, using a Boltzmann weighting,

$$W_\mu = \frac{\mathrm{e}^{\beta \mathrm{F}_\mu}}{Z}, \tag{1}$$

where β is a parameter allowing control of the selection strength and Z is the normalisation factor over the population,

$$Z = \sum_{\mu=1}^{P} \mathrm{e}^{\beta \mathrm{F}_\mu}. \tag{2}$$

The fit member is replicated and replaces an unfit population member, ν, drawn with the inverse weighting,

$$W_\nu = \frac{e^{-\beta F_\nu}}{Z'}, \tag{3}$$

where again Z' is the normalisation factor over the population.

The use of this selection/deletion strategy is important as it allows the selection strength to be clearly parameterised. Other selection/deletion schemes such as ranking selection and least fit deletion have the selection strength implicit in the implementation. Using Boltzmann weighting in this form, will allow us to compare these results directly with those from a generational GA using Boltzmann roulette wheel selection.

The results described in this paper can be found directly through simulation of this simple GA in its steady state and generational forms. The mathematical analysis does however give significant insights and explains the features seen in the simulation experiments.

3 Calculating the Dynamics

The formalisation developed by Prügel-Bennett, Shapiro and Rattray consists of modelling the cumulants of the fitness distribution as it evolves at each generation. The cumulants describe the fitness distribution of the population. For a finite population of members with fitness F, the first four cumulants are,

$$\begin{aligned}
\kappa_1 &= \langle F \rangle \\
\kappa_2 &= \langle F^2 \rangle - \langle F \rangle^2 \\
\kappa_3 &= \left\langle (F - \langle F \rangle)^3 \right\rangle \\
&= \langle F^3 \rangle - 3\langle F \rangle \langle F^2 \rangle + 2\langle F \rangle^3 \\
\kappa_4 &= \left\langle (F - \langle F \rangle)^4 \right\rangle - 3\,(\kappa_2)^2 \\
&= \langle F^4 \rangle - 4\langle F^3 \rangle \langle F \rangle - 3\langle F^2 \rangle^2 + 12\langle F^2 \rangle \langle F \rangle^2 - 6\langle F \rangle^4,
\end{aligned} \tag{4}$$

where $\langle \ldots \rangle$ represents the expected values averaged over the population.

In this paper we model the cumulants by calculating the changes in the moments of the population fitness distribution — $\langle F^n \rangle$ where n is the exponent of fitness — after each selection/deletion process. Using the expressions in equation 4, we can then calculate the changes in the cumulants.

3.1 Selection

To calculate the change in expected population fitness, we must find the expected fitness of the selected and deleted population members. For a finite population, the fitness of the selected member F_μ, is given by the summing the population fitness over the Boltzmann weighting,

$$\langle F_\mu \rangle_S = \sum_{\mu=1}^{P} W_\mu F_\mu, \tag{5}$$

where $\langle \ldots \rangle_S$ represents the expected value after selection.

This represents the selected value in a finite population. We wish however to model the dynamics of the average case, the result seen when many individual runs of the GA are averaged together. To do this we consider the selection to take place from a continuous distribution averaged across an infinite number of these finite populations. This we call an ensemble. We can rewrite the discrete summations above as integrals over this ensemble,

$$\langle F_\mu \rangle_S = \frac{\int_{-\infty}^{\infty} \rho(F)\, \mathrm{e}^{\beta \mathrm{F}} \mathrm{F} \mathrm{dF}}{\int_{-\infty}^{\infty} \rho(F')\, \mathrm{e}^{\beta \mathrm{F}'} \mathrm{dF}'}, \tag{6}$$

where $\rho(F)$ is the fitness distribution of the ensemble and the denominator is the normalisation factor over this distribution.

Strong selection implies that we are picking members in the extremes of the distribution. In these ranges, there are likely to be differences between the real finite population and the continuous integral. We can correct for this effect but it requires a numerical integration. If we consider the case of weak selection, we can use these expressions and integrate analytically.

To perform the integral we must construct the distribution of the ensemble. This must be a continuous function with the same cumulants as the finite population. A convenient form is an expansion around a Gaussian using the Gram-Charlier expansion,

$$\rho(F) = \frac{1}{\sqrt{2\pi K_2}} \exp\left(\frac{-(F-K_1)^2}{2K_2}\right)\left[1 + \sum_{n=3}^{N} \frac{K_n}{n! K_2^{n/2}} H_n\left(\frac{F-K_1}{\sqrt{K_2}}\right)\right] \tag{7}$$

where $H_n(x)$ are Hermite polynomials,

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{\mathrm{d^n}}{\mathrm{dx^n}}\left(\mathrm{e}^{-\frac{x^2}{2}}\right) \tag{8}$$

and n is the number of cumulants used.

When n is two, the expansion has exactly the form of a Gaussian. Adding in further terms distorts the Gaussian to give the desired distribution. In most cases considered here, the first four cumulants are sufficient. In this case,

$$\begin{aligned} H_3(x) &= x^3 - 3x \\ H_4(x) &= x^4 - 6x^2 + 3. \end{aligned} \tag{9}$$

For the weak selection expansion we can perform these integrals analytically. To first order, for the first two exponents, the expected fitnesses in terms of the ensemble cumulants are,

$$\begin{aligned} \langle F_\mu \rangle_S &= K_1 + K_2\beta \\ \langle F_\mu^2 \rangle_S &= K_2 + K_1^2 + (2K_1K_2 + K_3)\beta, \end{aligned} \tag{10}$$

where the terms K_n is the nth cumulant of the ensemble distribution.

3.2 Finite Population Effects

In equation 4 we used the expression κ_n for the cumulants of a finite population, whilst in equation 10 we used K_n as the cumulants of the ensemble distribution. This distinction is intentional as the cumulants of a finite population will always be slightly different from the cumulants of the continuous distribution from which they are drawn. This is known as the finite sample effect. For large populations, these effects are small. For small populations and high cumulants they can become significant. Using the identities below, we can switch between the actual finite population cumulants and the ensemble cumulants,

$$\begin{aligned}
\kappa_1 &= K_1 \\
\kappa_2 &= P_2 K_2 \\
\kappa_3 &= P_3 K_3 \\
\kappa_4 &= P_4 K_4 - 6P_2 \left(K_2\right)^2 / P
\end{aligned} \tag{11}$$

where $P_2 = \left(1 - \frac{1}{P}\right)$, $P_3 = P_2\left(1 - \frac{2}{P}\right)$ and $P_4 = P_2\left(1 - \frac{6}{P} + \frac{6}{P^2}\right)$.

3.3 Calculating the Cumulants

Having calculated the expected fitness of the selected fit and unfit members, we can update the moments of the population fitness distribution after the selection/deletion process,

$$\langle F^n \rangle_S = \langle F^n \rangle + \frac{\langle F_\mu^n \rangle_S}{P} - \frac{\langle F_\nu^n \rangle_S}{P}. \tag{12}$$

Using the original expressions for the cumulants of a finite population given in equation 4 and the expressions for the selected population members in equation 10, we can find the change in cumulants with each selection/deletion process in terms of the initial cumulants. The first two cumulants to first order are,

$$\begin{aligned}
\langle \kappa_1 \rangle_S &= \kappa_1 + \frac{2K_2}{P}\beta \\
\langle \kappa_2 \rangle_S &= \kappa_2 - \frac{2K_2}{P^2} + \frac{2K_3}{P}\beta.
\end{aligned} \tag{13}$$

Using these expressions and those for the higher order cumulants (up to four) we can model the evolution of a distribution through this steady state selection. Figure 1 shows the model predictions plotted against simulation data averaged over one thousand runs. Initially the population is a Gaussian distribution and through selection converges to a fixed value.

3.4 Random Deletion

Another form of deletion scheme commonly used in comparisons of steady state GA and generational GA is that of deleting a member at random [5]. In this case the expected fitness value of the deleted member, ν is simply the mean value of the population fitness,

$$\langle F_\nu^n \rangle_S = \langle F^n \rangle. \tag{14}$$

The terms for the first two cumulants to first order are,

$$\begin{aligned}
\langle \kappa_1 \rangle_S &= \kappa_1 + \frac{K_2}{P}\beta \\
\langle \kappa_2 \rangle_S &= \kappa_2 - \frac{2K_2}{P^2} + P_2\frac{K_3}{P}\beta,
\end{aligned} \tag{15}$$

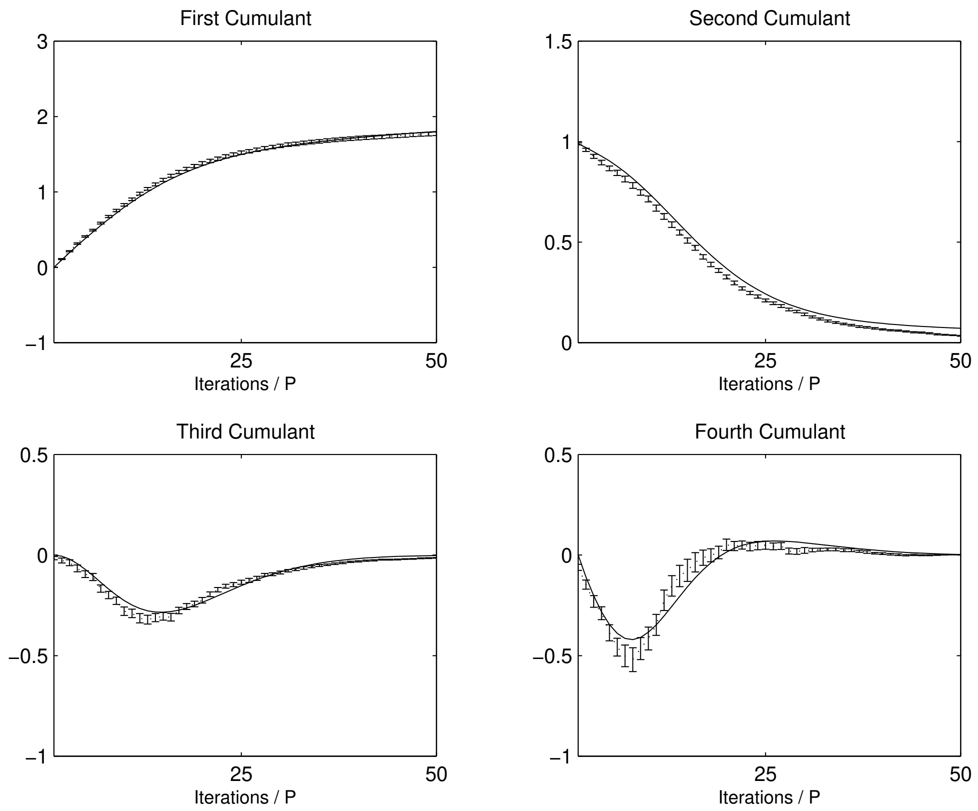

Figure 1. Theory predictions plotted against experimental data averaged over 1000 runs for a simple steady state GA using weak selection. The population size is 100 and the selection strength is 0.05.

where P_2 is the finite sample correction.

The factor of two in the terms containing the selection strength β is a function of the different selection pressures. Boltzmann deletion exhibits twice the selection pressure of the random deletion scheme.

The rate at which variance is decreasing with each selection/deletion process — the κ_2 term independent of selection strength — is the same in the two cases.

3.5 Comparison with Generational GA Results

The analytical expressions describing the changes in cumulants after each selection/deletion process for the steady state GA can be compared to those derived by Prügel-Bennett *et al.* for the changes after one generation of a generational GA,

$$\begin{aligned}\langle\kappa_1\rangle_S &= \kappa_1 + K_2\beta \\ \langle\kappa_2\rangle_S &= \kappa_2 - \frac{K_2}{P} + K_3\beta. \qquad (16)\end{aligned}$$

The expansion for all three selection schemes — Boltzmann deletion, random deletion and generational — are very similar. The two steady state expressions have an extra factor of P as they describe the change after each selection/deletion process. In this weak selection case, the cumulants do not change significantly with each selection and we can consider a quasi-static case and compare the two expressions directly despite this factor of P.

The factor of two in the term independent of selection strength, describes the rate at which the variance of the population decreases under the selection scheme — the genetic drift. When $\beta = 0$ and we are selecting and deleting at random, these expansions become exact,

$$\begin{aligned} \langle \kappa_2 \rangle_S &= \kappa_2 - \frac{K_2}{P} && \text{generational selection} \\ \langle \kappa_2 \rangle_S &= \kappa_2 - \frac{2K_2}{P^2} && \text{steady state selection.} \end{aligned} \tag{17}$$

By expansion it can be shown that the rate at which the population variance decreases under steady state selection is twice that of generational selection — to an accuracy of $1/P^2$ over one generation. This same effect was commented on by De Jong in terms of allele loss with changing generation gap [3]. Figure 2 shows the decreasing variance, κ_2, due to this selection difference.

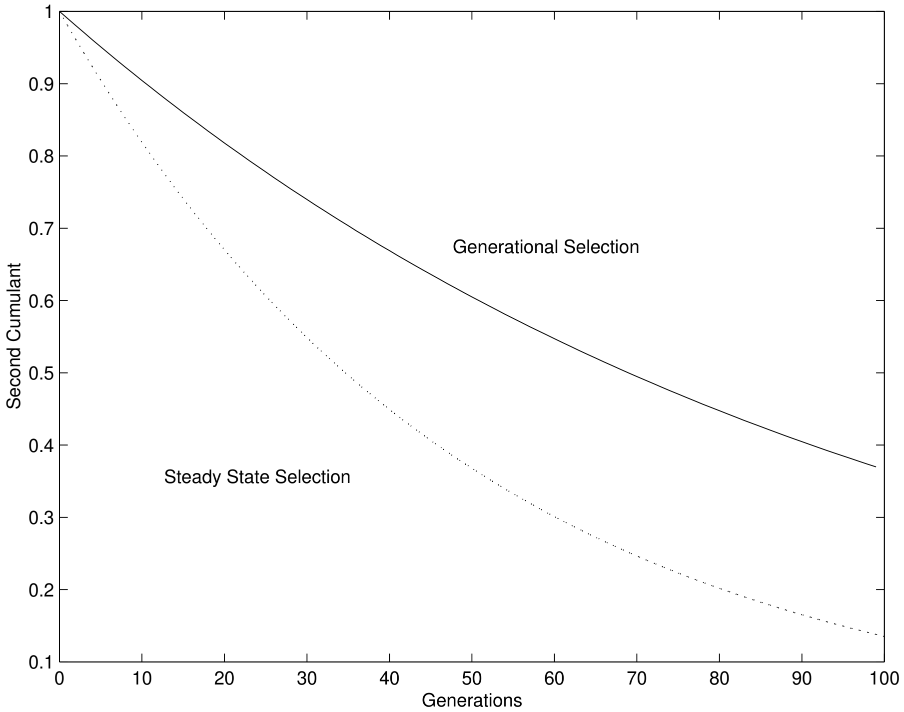

Figure 2. Comparison of generational and steady state selection when β=0 - random selection. Steady state data is plotted at P selection/deletion processes to one generation of the generational GA. Population size is 100.

3.6 Discussion

We have already noted that if we double the selection strength of the steady state GA with random deletion, it matches the Boltzmann deletion terms to first order. Taking this further by rescaling the time scale of the steady state GA — $P/2$ selection/deletion processes equating to one generation of the generational GA — the dynamics of all three types will match. This is shown by simulation results in Figure 3.

This result suggests that in this simple case, we can reproduce the dynamics of the generational GA with a steady state GA but at half the computational cost in terms of function evaluations.

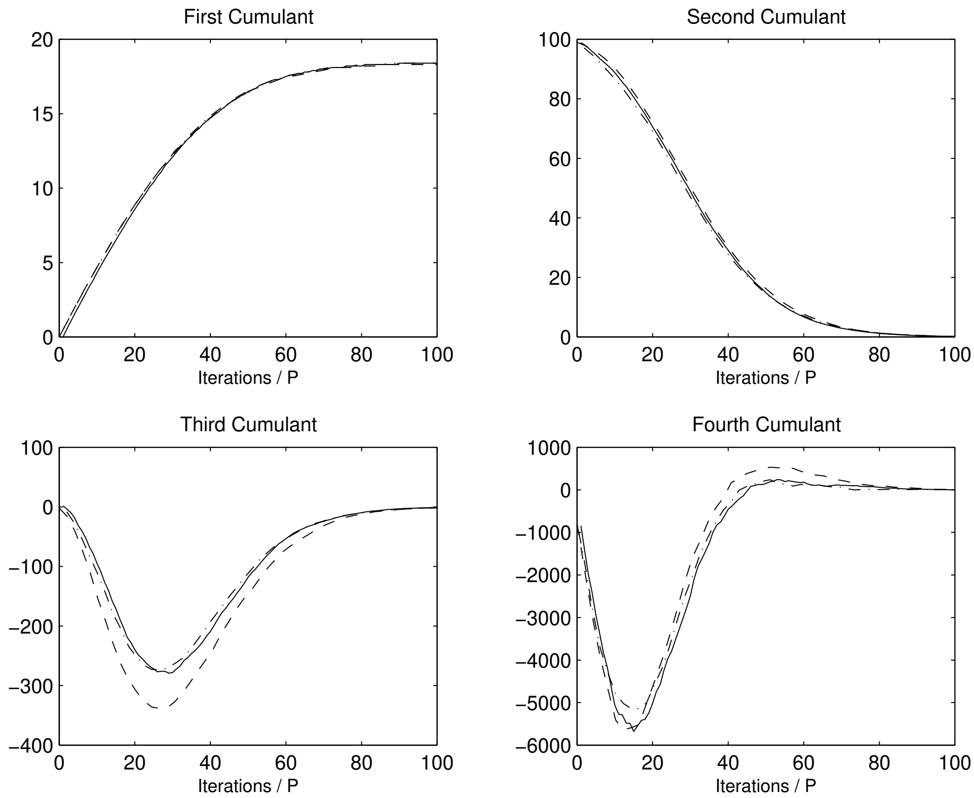

Figure 3. Comparison of steady state and generational GA (solid line - generational, dot-dash line - steady state with Boltzmann deletion and dash line - steady state with random deletion) when parameters are rescaled. Population size is 100 and for the generational GA $\beta = 0.05$.

4 Ones Counting and Mutation

In the analysis so far, we have not considered the fitness function other than to assume that it gives a smooth unimodal distribution. We can extend this analysis to a more realistic problem by considering a binary string and ones counting.

For population member, μ, consisting of a string of length L, the fitness is,

$$F^{\mu} = \sum_{i=1}^{L} S_i^{\mu}, \tag{18}$$

where S_i^{μ} are the bits of the binary string and may take values,

$$S_i^{\mu} = \{-1, 1\}. \tag{19}$$

We pick a fit member using Boltzmann weighting, replicate it with a fixed point mutation rate and replace an unfit member with the mutated replica.

4.1 Calculating the Effects of Mutation

During mutation,

$$S_i^{\mu} \rightarrow -S_i^{\mu} \quad \text{with probability } \gamma. \tag{20}$$

The fitness of population member, μ, after mutation is given by,

$$\langle F_{\mu} \rangle_M = \Gamma F_{\mu}, \tag{21}$$

where,

$$\Gamma = (1 - 2\gamma), \tag{22}$$

and $\langle \ldots \rangle_M$ represents the expected fitness after mutation.

Extending this to the other exponents is an exercise in combinatorics,

$$\begin{aligned}
\left\langle F_{\mu}^2 \right\rangle_M &= \Gamma^2 F_{\mu}^2 + L\left(1 - \Gamma^2\right) \\
\left\langle F_{\mu}^3 \right\rangle_M &= \Gamma^3 F_{\mu}^3 + \Gamma F_{\mu}\left(1 - \Gamma^2\right)(3L - 2) \\
\left\langle F_{\mu}^4 \right\rangle_M &= \Gamma^4 \left(F_{\mu}^4 + 2F_{\mu}^2 (4 - 3L) + 3L(L - 2)\right) \\
&\quad + 2\Gamma^2 \left(F_{\mu}^2 - L\right)(3L - 4) + L(3L - 2).
\end{aligned} \tag{23}$$

Mutation acts to drive the population back to its initial random distribution centred around zero. Selection acts against this pressure, trying to correlate the population to a member with high fitness. An equilibrium is reached between these two forces.

As before we can calculate the change in cumulants by the change in the expected moments of the population fitness distribution. However we must apply the mutation operator to the selected population member which is replicated,

$$\langle F^n \rangle_{S,M} = \langle F^n \rangle + \frac{\langle F_{\mu}^n \rangle_{S,M}}{P} - \frac{\langle F_{\nu}^n \rangle_S}{P}, \tag{24}$$

where $\langle \ldots \rangle_{S,M}$ represents the expected value after selection and mutation.

Figure 4 shows the simulation results for this problem averaged over ten thousand runs against the theory predictions. In this case an analytical expression for all four cumulants becomes unwieldy and we perform the integral in equation 6 numerically.

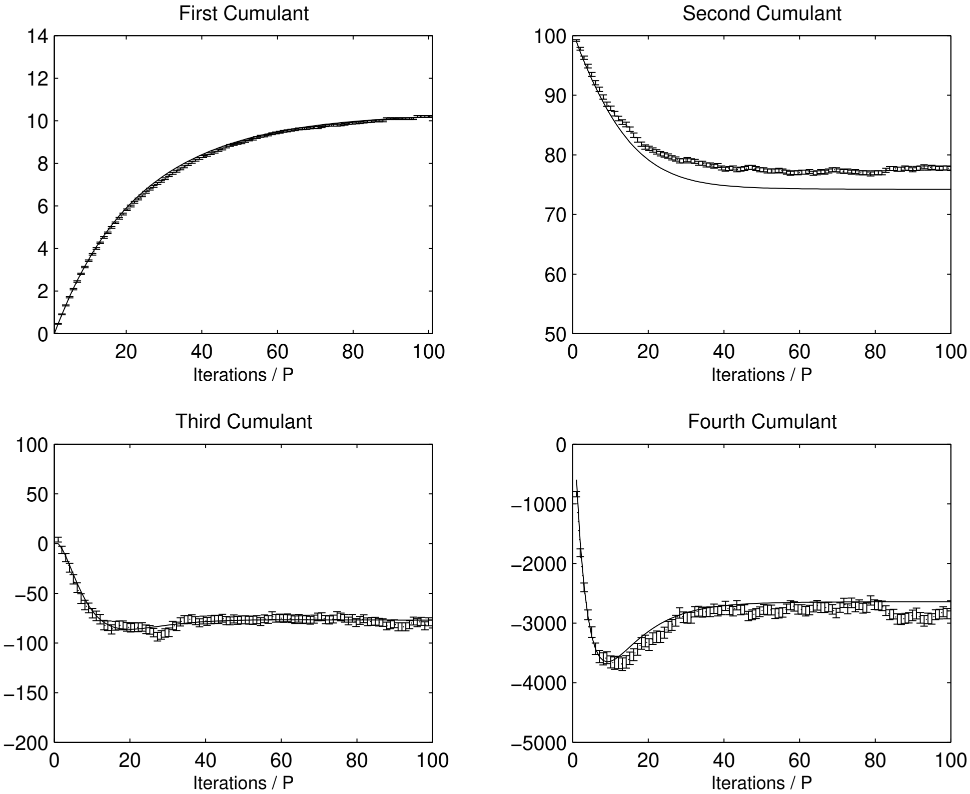

Figure 4. Comparison of simulations averaged of 10,000 runs using the weak selection expansion for steady state GA with random deletion on the ones counting problem. The population size is 100, the string length is 100, $\beta = 0.05$ and $\gamma = 0.02$.

4.2 Discussion

Prügel-Bennet *et al.* calculated the effect of mutation on the whole population and for the first two cumulants derived the results,

$$\begin{aligned} \langle \kappa_1 \rangle_M &= \Gamma \kappa_1 \\ \langle \kappa_2 \rangle_M &= \Gamma^2 \kappa_2 + L\left(1 - \Gamma^2\right). \end{aligned} \tag{25}$$

By considering again the quasi-static case where the population cumulants do not change significantly over $P/2$ selection/deletion processes, we can compare the steady state dynamics to the generational case. Expanding analytically the steady state analysis for the first two cumulants and assuming we make $P/2$ selection/deletion processes, we obtain the same expressions as the generational GA but with a doubling of the mutation rate.

Thus we can rescale both the mutation and selection parameters to make the steady state GA match the dynamics of the generational GA but again with half the computational cost. Figure 5 shows the three schemes plotted together.

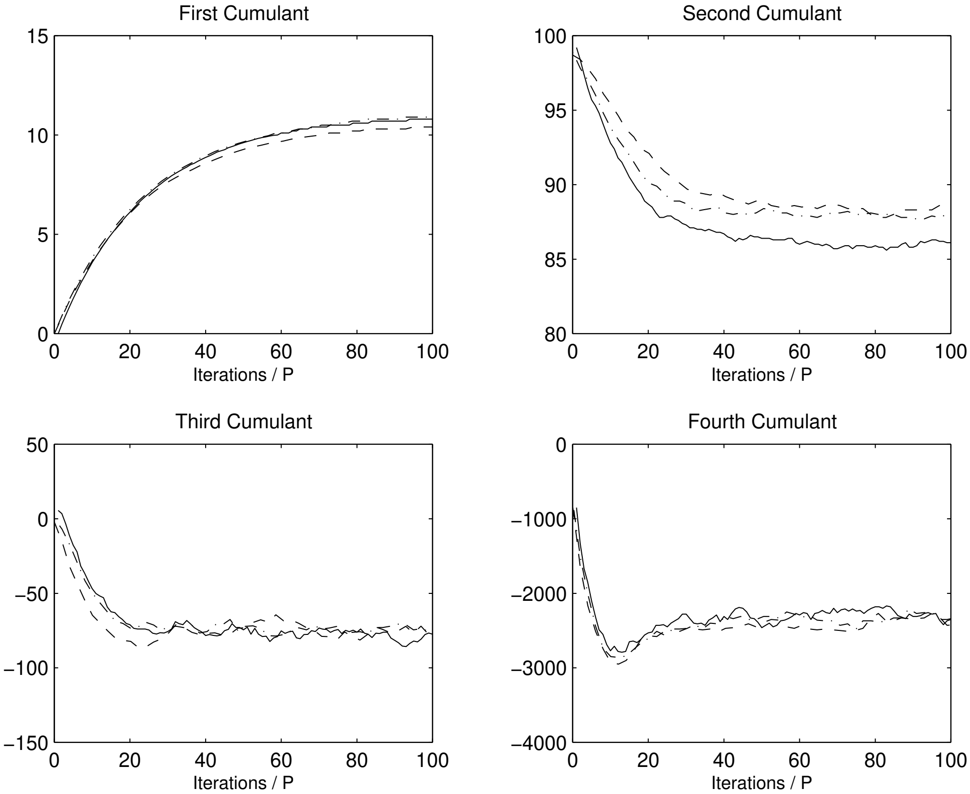

Figure 5. Comparison of steady state and generational GA (solid line - generational, dot-dash line - steady state with Boltzmann deletion and dash line - steady state with random deletion) for a ones counting problem with mutation when parameters are rescaled. The population size is 100, the string length is 100. For the generational GA $\beta = 0.05$ and $\gamma = 0.01$. For steady state with Boltzmann deletion $\beta = 0.05$ and $\gamma = 0.02$. For steady state with random deletion $\beta = 0.1$ and $\gamma = 0.02$.

5 Conclusions

The formalisation developed by Prügel-Bennet, Shapiro and Rattray provides a very good basis on which comparisons can be made. The careful choice of selection method — Boltzmann roulette wheel — allows comparison directly from simulation without changes in selection pressure. The mathematical analysis gives a further insight and explains the simulation observations.

From this analysis it is shown that the rate of decrease of population variance — genetic drift — under steady state selection is twice that compared to the generational GA. Whilst only selection and mutation operators are considered, it appears that we can rescale the parameters of the steady state GA to match the dynamics of the generational GA. However, the steady state GA is evolving at twice the rate - $P/2$ selection/deletion processes compare with one generation of the generational GA. This gives an indication as to a possible performance advantage of the steady state GA. It evolves with the same dynamics but with half the computational cost. Alternatively, for the same computational cost, double the population size can be used.

The work presented here is part of an ongoing study of steady state and generational GA. The next stage, is to include the crossover operator. It is suspected that at this point the rescaling will break down as the effect of crossover is dependent on the correlation of the population. However it appears that through this formalisation and by looking at a harder problem than ones counting, some useful insights into the differences between generational and steady state GA will be achieved.

References

[1] Holland, J. H. (1975) *Adaption in Natural and Artificial Systems.*The University of Michigan Press, Ann Arbor, MI, 1975.

[2] DeJong K. (1975) *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems,* Doctoral Thesis, Department of Computer and Communication Sciences, University of Michigan, Ann Arbor.

[3] DeJong K. Sarma J. (1992) *Generation Gaps Revisited.* Foundations of Genetic Algorithms 2. Morgan Kaufmann, 1992.

[4] Vavak F. Fogarty T. (1996) *Comparison of Steady State and Generational Genetic Algorithms for use in Nonstationary Environments.* Proceedings of the 1996 IEEE Conference on Evolutionary Computation, IEEE Press.

[5] Syswerda G. (1991) *A Study of Reproduction in Generational and Steady-State Genetic Algorithms.* Foundations of Genetic Algorithms. Morgan Kaufmann, 1991.

[6] Prügel-Bennett, A., Shapiro, J.L. and Rattray, M. (1994) *A Statistical Mechanics Formulation of the Dynamics of Genetic Algorithms.* Lecture notes in Computer Science **865**, pages 17–27, Berlin, 1994. Springer-Verlag.

[7] Prügel-Bennett, A. and Shapiro, J.L. (1997) *The dynamics of a Genetic Algorithm for simple random Ising systems.* Physica D 104 (1997) 75-114

[8] Prügel-Bennett, A. (1997) *Modelling Evolving Populations.* Journal of Theoretical Biology, 1997.