Distribution Replacement: How survival of the worst can out perform survival of the fittest

Howard Tripp
Surrey Space Centre,
University of Surrey, Guildford,
Surrey, GU2 7XH, UK.
+44 (0)1483 683419
h.tripp@surrey.ac.uk

Phil Palmer
Surrey Space Centre,
University of Surrey, Guildford,
Surrey, GU2 7XH, UK.
+44 (0)1483 683419
p.palmer@surrey.ac.uk

ABSTRACT

A new family of "Distribution Replacement" operators for use in steady state genetic algorithms is presented. Distribution replacement enforces the members of the population to conform to an arbitrary statistical distribution, defined by its Cumulative Distribution Frequency, relative to the current best individual. As new superior individuals are discovered, the distribution "stretches" to accommodate the increased diversity, the exact opposite of convergence. Decoupling the maintenance of an optimal set of parents from the production of superior children allows the search to be freed from the traditional overhead of evolving a population of maximal fitness and, more significantly, avoids premature convergence. The population distribution has a significant effect on performance for a given problem, and in turn, the type of problem affects the performance of different distributions. Keeping mainly good individuals naturally does well on simple problems (as do distributions that exclude "median" individuals). With deceptive problems however, distributions which keep mainly bad individuals are shown to be superior to other replacement operators and also outperform classical generational genetic algorithms. In all cases, the uniform distribution proves suboptimal. This paper explains the details of distribution replacement, simulation experiments and discussions on the extension of this idea to a dynamic distribution.

Categories and Subject Descriptors

I.2.2 [**Artificial Intelligence**]: Automatic Programming – *automatic analysis of algorithms.*

General Terms

Algorithms, Theory.

Keywords

Genetic algorithms, steady state genetic algorithm, distribution replacement, replacement, CDF, survival of the worst, survival of the extremes, keep second best, random keep best, normal distribution, beta distribution, DeJong functions, Saw tooth.

GECCO'07, July 7–11, 2007, London, England, United Kingdom.

1. INTRODUCTION

Evolutionary computing (specifically genetic algorithms (GAs) in this paper) has been studied for many years as a powerful mechanism for navigating the search space of optimisation problems. GAs maintain a population of candidate solutions. Each solution is assigned a fitness value according to some objective function and a new generation of the population is then created using selection, crossover and mutation operators to splice, mutate and/or copy individuals from the old population; this process is repeated until some termination condition is triggered.

As Wiegand [9] points out, there are essentially two different kinds of selection in GAs, *parent selection* for breeding new candidate solutions and *survival selection* to choose a subset of the population to be carried forward to the next generation. The importance of this distinction is almost never recognised and it is standard practice to perform both functions using a single *Selection* operator. Attempting to do two jobs with one operator results in the enviable trade-off that that neither job is performed optimally, and usually understood in terms of selection pressure (convergence to the best solution) being inversely proportional to population diversity (divergence to the best parents) [6, 7]. By recognising this as the root cause of the trade-off however, the dependency can be broken as it is *parent selection* (which is referred to from here on as **Selection**) that controls the search for the optimum solution; and it is *survival selection* (referred to as **Replacement**) that controls population diversity. These functions (and hence their effects) can therefore be chosen independently.

Legg proposes his Fitness Uniform Deletion Strategy (FUDS) which is "…based on the insight that we are not primarily interested in a population converging to maximal fitness, but only in a single individual of maximal fitness." [2, 5]. This insight is powerful indeed, as the removal of the requirement on population convergence breaks the convergence/divergence (or selection pressure/diversity) trade-off. The population remains diverse at all times, whilst simultaneously searching for the single individual of maximal fitness (I_{MAX}). Along with I_{MAX}, other individuals should now be carried forward to the next generation not based on their current fitness, but on their potential as parents to produce children with a higher fitness than I_{MAX}.

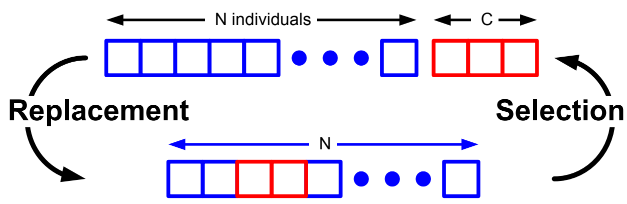


Figure 1: The Replacement-Selection Loop

The Replacement operator is already a necessary part of a Steady State GA (in each generation children are added to the existing population, before it is then pruned), see Figure 1. In fact it is the explicit use of a Replacement operator (such as replace worst, replace parent, DeJong Crowding etc.) that defines the difference between steady state and classical GAs. The reason that the steady state GA is not the "standard" GA is likely down to a lack of analysis on the Replacement operator. Without the insight of the separation of responsibility between Selection and Replacement, the implementation of Replacement can seem an unnecessary overhead. However, this paper demonstrates that on deceptive problems, the addition of a suitable Replacement operator is a simple method to improve performance over a classical generational genetic algorithm.

In essence, the separation of responsibilities is such that *Replacement* seeks to maintain the optimal set of *parents*; *Selection* seeks to match parents into breeding pairs optimally; and *mutation/crossover* seek to recombine those breeding pairs into new *children* optimally. This paper uses standard approaches for selection and mutation/crossover as a baseline for the introduction of novel replacement strategies and their analysis.

Legg's FUDS [2, 5] can best be described as a replacement operator that enforce the fitness distribution of individuals in the population to conform to a uniform distribution. To the author's knowledge this is the only example of a distribution replacement strategy currently in the literature and the aim of this paper is to extend this work to other distribution shapes, and examine the effects on performance (see section 2). Section 3 describes the experimental setup and parameter ranges used for investigation. Section 4 presents the results of the simulations and sections 5 and 6 discuss general conclusions and future potential.

2. REPLACEMENT

2.1 Standard Replacement Strategies

Consider Figure 1 where a steady state GA has a population of size N. In a single generation the selection, mutation and crossover operators splice and mutate individuals from the population to generate a set of children of size C. These children are added back into the population resulting in an expanded population of size N+C. It is now the Replacement operator's task to choose C individuals from expanded population to delete, and once again return the population to size N. Standard approaches to this are as follows:

- **Random** – Remove C individuals randomly.
- **Truncation** – Remove the C individuals with the lowest fitness scores.
- **Parent** – Remove C of the parents chosen by the previous selection operator.

It can be seen that none of these operators explicitly control population diversity, or take any account of the child solution generated by the selection operator. The following operators were therefore developed to try to maintain population diversity by considering the similarity of the individuals in the population rather than fitness. The similarity is normally determined by the Hamming distance between the two individuals (i.e. the minimum number of bit flips required to change one individual into the other in the binary representation) [10]:

- **DeJong Crowding** – For each child, randomly select a subset of the old population (with the size of the subset defined as the crowding factor). Now remove the individual from the subset that is the most similar to the child. [1]
- **Contribution to Diversity / Remove Worst (CDRW)** – For each child, consider the subset of the old population with lower fitnesses. Now remove the member of this subset with the lowest contribution to diversity (defined as minimum average Hamming distance from all other members of the population). Or, in the case that the child has the lowest contribution to diversity, simply remove the individual with the lowest fitness [6].

A major disadvantage of these two approaches is the additional computation involved in calculating the similarity between two individuals. This is a relatively small overhead for DeJong Crowding (scales linearly with crowding factor), but the effect is much larger for CDRW which scales with a square of population size. In CDRW every child must compare itself with every individual in the population, and further, every individual in the population must have its contribution to diversity updated every time any individual is added or removed.

2.2 Distribution Replacement

The replacement operator's objective is to evolve the population towards the optimal set of parents. For any set of selection, mutation and crossover operators, on a specific problem, there will be a particular distribution of parents for which their performance is most effective. Distribution Replacement simply maintains that particular distribution of parents in order for the other operators to work optimally. Any type of statistical distribution can be maintained such as Exponential, Binomial, Gaussian, Beta, or a numerical distribution that cannot be easily mathematically represented.

As every individual in the population has a fitness, the fitness distribution of the whole population can be mapped onto the desired statistical distribution. Rather than considering individual population members this abstraction allows for a reduction in computational overhead. Simple metrics are usually used to define particular distributions for example:

- The difference between the maximum fitness in the population and the mean of the population is a measure of what is normally meant by selection pressure (the smaller the difference the higher the pressure).

- The standard deviation of the population fitnesses is a simple measure of population diversity[1].

Of course the actual fitness function will have an impact on these measures, as will any scaling function, but in general, the metrics are intuitive and practical.

2.2.1 Implementing Distribution Replacement

Any statistical distribution (continuous or discreet) can be defined by its Cumulative Distribution Function, CDF (or it's probability density function, pdf). This CDF is normalized (and truncated if it is defined over an infinite range) to give CDF_N that it is valid only in the range [0,1], so that it can be mapped onto the population. The algorithm for removing an individual from the population is then (with the effect shown in Figure 2):

1. Add the child individual and re-sort the population according to fitness (ascending).
2. Identify the individual in the population with maximum fitness (final individual), I_{MAX}.
3. The CDF_N is normalised by I_{MAX}'s fitness, and the population size, to give the theoretical CDF that is required for this generation, CDF_T. CDF_T can be used to translate the index, n, of a particular individual (with fitness f_n) into the theoretical fitness, $f_{Tn,}$ that it should have:

$$f_{Tn} = CDF_T(n) = CDF_N(n / popsize) / I_{MAX}$$

4. The ordered list of individuals can now be considered as a set of discreet points that should fall upon the CDF_T curve. The individual causing the maximum misalignment[2] to the correlation between the discreet points and the CDF_T curve needs to be removed.
5. Rather than introduce the overhead of calculating the misalignment for every individual, it is a sufficient approximation to use a standard binary decomposition process for efficiency:
   - 5.1. Consider the fitness of a pivot, p, located at the median of a range (initially the whole population).
   - 5.2. If $f_p \geq f_{Tp}$, there are too many individuals below the pivot with an actual fitness higher than their theoretical CDF_T values.
   - 5.3. The pivot becomes the new upper bound of the range (similarly, the lower bound if the fitness of the pivot is too low), and repeat from 5.1 until the range contains only a single individual.
6. Remove the single individual in the range.

From this algorithm, several facts about distribution replacement can be discerned:

- The individual with maximum fitness, I_{MAX} is never removed (compared with Roulette Wheel and many other selection schemes in classical GAs that make no such guarantee).
- The shape of the distribution is normalised by I_{MAX}, so as its fitness increases through the generations so the (non-normalised) mean and deviation of the actual population will also increase.
- A new CDF_T is only calculated when a new individual supersedes I_{MAX} which occurs much less often than once per generation.
- As this is a binary decomposition process, doubling the size of the population adds only one extra pivot comparison operation. Therefore the number of times an individual in the population is compared with CDF_T scales with logarithm of the population size, O(log(N)). I.e. Significantly less than CDRW, $O(N^2)$, although more than Random and Truncation O(constant).

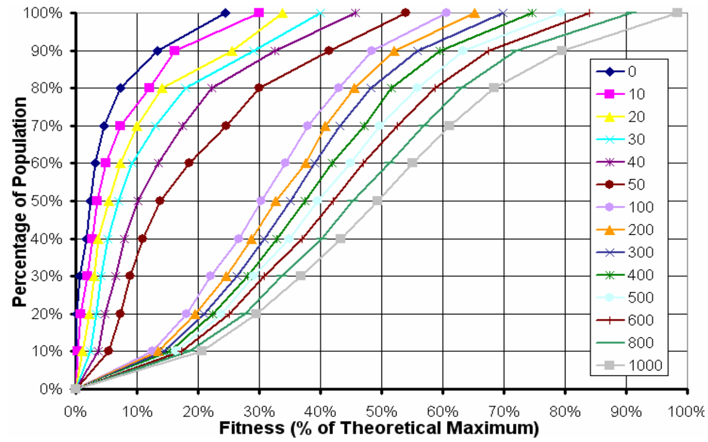


Figure 2: The CDF of a population (at various generations) which converges to a Normal distribution and stretches as better individuals are discovered

2.3 Other Replacement Strategies

In order to benchmark distribution replacement, two further simple replacement strategies are proposed and used in this paper:

- **Random Keep Best** (RandKB) – As Random (above), but never remove I_{MAX}.
- **Remove Second Best** (2^{nd}Best) – Always remove the individual with the second-highest fitness score. This results in a large group with low fitnesses and a single maximal "explorer".

3. EXPERIMENT & SIMULATION SETUP

The DeJong fitness function test set [1] is a set of five different mathematical functions that give highly different search spaces. Evaluating the various replacement/selection strategies and parameters for each of the functions gives a good indication about the type of problems to which each is suited and allows for direct comparisons with other work:

- **Function 1 "Sphere"** (DF1) - Smooth, unimodal, symmetric, a measure of general efficiency.

---

[1] As the mean and standard deviation of a population varies in each generation, both these values must be normalised by the maximum fitness in that generation.

[2] The individual at index, n, with the lowest value of $(2f_n^2 - f_{n-1}^2 - f_{n+1}^2) / (2f_{Tn}^2 - f_{Tn-1}^2 - f_{Tn+1}^2)$.

- **Function 2 "Rosenbrock"** (DF2) - Has a narrow, sharp ridge running around a parabola, tests ability to discover good directions.
- **Function 3 "Step"** (DF3) – Real numbers rounded to integers - representative of the problem of flat surfaces (with no direction information).
- **Function 4 "Quartic"** (DF4) – Simple, unimodal function padded with gaussian noise, tests ability on noisy data.
- **Function 5 "Shekel's Foxholes"** (DF5) - Multiple (24) local optima which is difficult and deceptive.

Figure 4 shows the two-dimensional representations of the five functions, however functions 1-4 were actually implemented as 10-dimensional functions to make the problems more challenging.

Three different Selection operators were used: Standard Roulette wheel, Rank (i.e. Roulette wheel on ranking rather than fitness values) and Random selection. Eight different basic Replacement operators were used (as discussed above in section 2): Random, Truncation, Parent, DeJong Crowding (Factors 1,2,3…9,10 & 20), CDRW, Random Keep Best (RandKB) and Remove Second Best (2^{nd}Best) (see sections 2.1 and 2.3). Three distribution replacement operators were used:

- **FUDS** (uniform distribution) proposed by Legg.
- **Beta** distribution. Controlled by two parameters α and β. Parameter sets were chosen to skew the distribution form uniform to extreme positive skew, negative skew and "U-shaped" (see Figure 3).
- **Normal** (Gaussian) distribution. Controlled by two parameters μ (mean) and σ (variance). A large range of μ/σ pairs was used, including "spikes", "gentle inclines" and "flat" distributions. The [-∞, +∞] range had to be normalised and truncated, as described in 2.2.1, so it is worth noting that the actual mean and variance of the population no longer directly corresponded to μ and σ of the distribution respectively.

In some situations the Normal and Beta distributions tended towards each other and simpler replacement strategies as can be seen in Figure 3.

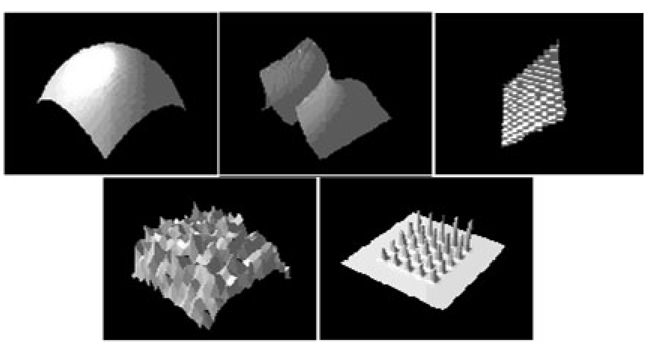
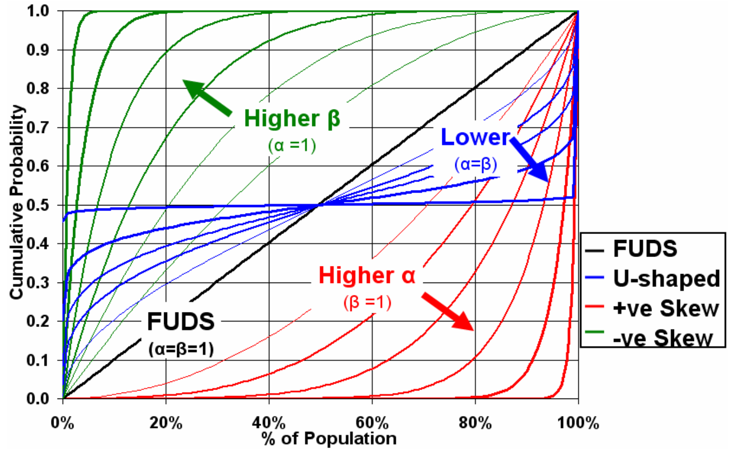


Figure 4: 2D Representations of the 5 DeJong Fitness Functions [www.denizyuret.com]]

As this was a steady state GA, there was no benefit in a child individual being a direct clone of a parent. Each child was



Normal parameters	Beta parameters	Other distribution
$\mu \rightarrow 1$ (I_{max})	$\alpha << \beta$	2^{nd}Best (-ve skew)
$\mu \rightarrow 0$	$\alpha >> \beta$	Truncation (+ve skew)
$\sigma >> 1$	$\alpha = \beta = 1$	FUDS

Figure 3: The CDF of the Beta distribution as α and β parameters are varied along with a table of the asymptotic tendencies of Normal and Beta distribution replacement

therefore a (single point) crossover of two parents, or a mutated[3] version of a single parent, with crossovers and mutations being mutually exclusive on any individual child. The overall fraction of crossovers to mutations was varied so that crossovers occurred 0%, 5%, 20%, 40%, 60%, 80% 95% and 100% of the time.

A population size of 30 was used with 3 (possible) replacements per generation for 1000 generation (10000 for DF5). Each individual simulation run was also the averaged over a minimum of 50 separate trials.

4. RESULTS

4.1 Effect of Distribution shape

For any particular configuration of fitness function, selection operator and mutation/crossover fraction, the plots of performance against Normal's μ and σ or Beta's α and β generate smooth continuous curves and surfaces. With the Normal distribution as σ increases (tending towards FUDS) the surface flattens. At low σ, μ has a significant effect (for example see Figure 5). Crucially, the point giving the best performance on the surface (or a subset of all maximal points) can always be found on the minimal σ line. Equally the worst point can also always be found on the minimal σ line. Hence high σ, or FUDS always gives a mid-range performance, or at best, a *non-uniquely* optimal performance.

This flattening at high σ and extreme ranges of performance at low as σ was a fundamental characteristic of the Normal distribution across all fitness functions, selection operators and crossover/mutation fractions. With optimal performance at low σ and either low or high (never median) μ, it is evident that the classic normal bell-shape, (or even skewed-bell-shape) was not

[3] A random number (<½ total number of bits) of bit flips. Each flip was performed on a random bit, so multiple flips per bit were possible, i.e. average of 11.4% bits changed per mutation.

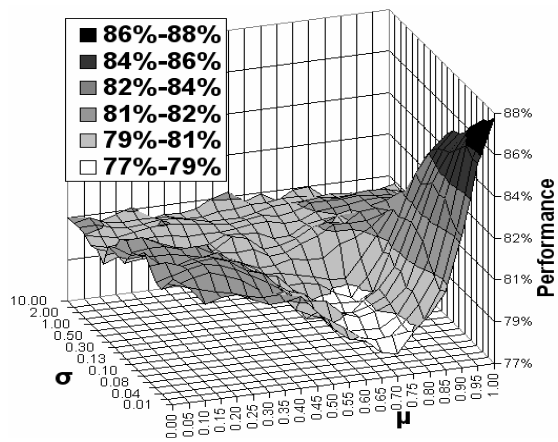


Figure 5: Surface plot of Performance (% of theoretical max achieved after 250 generations) vs Normal μ and σ parameters on DF3 with Roulette Wheel selection and a Crossover fraction of 5%.

the most effective distribution for Replacement. In other words, keeping mainly average individuals was never as effective as keeping mainly good or mainly bad individuals. As can be seen from Figure 3, such distributions are more easily modelled with the Beta distribution. Hence this same effect can be seen in Figure 6 and Figure 7, where moving away from the symmetrical FUDS ($\alpha = \beta = 1$ at the bottom in all cases) towards more extreme skews improves performance. From here on only the Beta distribution is discussed.

Considering now Figure 6.2.3, with its optimal performance at large α (when $\beta = 1$). This is exactly the well known effect that the Truncation replacement operator already exploits. In other words, keeping only the best individuals with minimal diversity. High +ve skew mirrors the effect of Truncation replacement very accurately, doing consistently well on the unimodal DFs 1, 2, 3 and 4[4] and consistently badly on the deceptive multimodal DF5 (Figure 6.1.3), across all selection operators and crossover fractions. The explanations for the Beta –ve skew or U-shaped's performance is not so familiar however.

The Beta –ve skew tends towards keeping the single best and all the worst individuals each generation, and further as β increases (when $\alpha = 1$) it asymptotically approaches the 2[nd]Best replacement strategy. For all selection schemes, performance improves with more extreme –ve skew on the deceptive and multi-modal DF5. Truncation replacement (i.e. extreme +ve skew) is inherently unsuited to this problem due to the massive loss of diversity. –ve skew (or 2[nd]Best) on the other hand manages to maintain more diversity and hence avoids the premature convergence problem (see section 4.3).

The Beta U-shaped distribution (Figure 6.1.1 & 6.2.1) whereby only the extremely good and bad individuals are preserved, generally mirrors the performance of +ve skew (Truncation replacement). The U-shaped distribution can take advantage of both the +ve and –ve skew effects, although it is able to do neither optimally and was therefore always the median result out of the three distribution skews. As $\alpha = \beta$, the distribution was symmetrical and so simply had the effect of halving the useful population which naturally impacted performance. It stands to reason that there is a spectrum of possible non-symmetrical U-shaped distributions with +ve and –ve skews being at opposite ends. Although this spectrum was not considered in this work its implication is discussed later (see section 5).

[4] Hence the reason these functions are averaged together

4.1.1 Other Replacement strategies

Random and Parent replacement both perform badly in every situation. DeJong Crowding's performance, although better than Random and Parent was also poor, but improved with higher crowding factors. Random-Keep-Best performed surprisingly well, although never the best, it should certainly always be used in place of Random replacement as it is a trivial extension to implement and results in a significant performance improvement. CDRW was impressive and proved very effective indeed, out-performing, or in the top few strategies, in almost every situation. The down side of course is the significant run time overhead to calculate and maintain the real time diversity measures. When its performance is evaluated against actual CPU overheads or time, it is likely to drop down the rankings significantly. Although the author acknowledges that this particular CDRW implementation was not coded for optimal CPU efficiency, simulation runs involving CDRW nevertheless took 2 or 3 orders of magnitude longer to run than the other replacement strategies.

4.2 Effect of Selection operator

The principle of the selection operator is to identify the best possible parents. As this measure is relative to the population as a whole, controlling the diversity profile using a replacement operator clearly has a significant effect.

Again, there is a clear distinction between the performance on the unimodal and multimodal functions. As can be seen from Figure 7, for the multimodal DJ5 a -ve skewed distribution is the most effective replacement strategy irrespective of the selection operator. Maximising diversity is fundamental to the DF5 problem which is achieved by maximising the -ve skew and hence

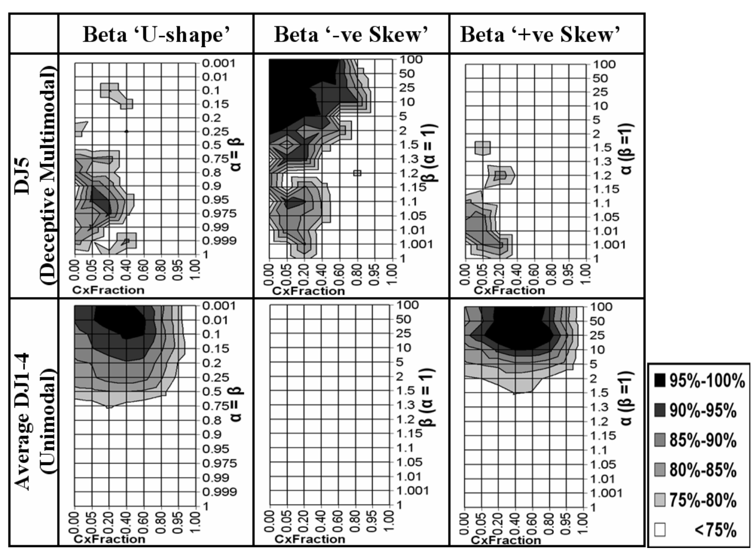


Figure 6: Contour plot of relative performance of a (α, β, Crossover-fraction)-tuple expressed as a ranking (i.e. 100% is best, 0% the worst). Top row is DF5, Bottom row the average of DFs 1-4. Averaged over all 3 selection functions. [Individual graphs referenced by row.column notation (e.g. Figure 6.2.3 for "Av DJ1-4, Beta +ve Skew")]

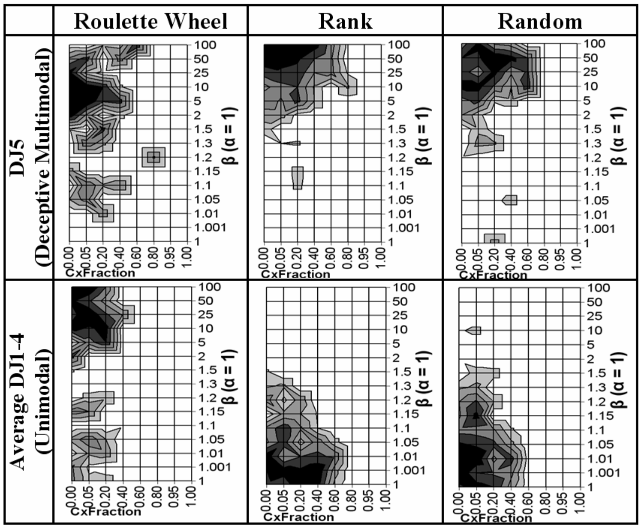


Figure 7: Relative performance ranking as Figure 6 (same scale). However ranking is only for the Beta '-ve Skew' not across the whole parameter range and is now broken down by selection function.

relegating effects of the selection operator to the point where it only has a secondary effect on multimodal problems.

The selection operator's influence is more evident in the unimodal problems. Roulette Wheel takes the relative fitness of individuals in the population into account when making selections, unlike Rank and Random which were only affected by the distribution of the population in an indirect manner. Figure 7.2.2 and 7.2.3 show that on the unimodal problems increasing -ve skew is inversely proportional to performance and these operators performed far more effectively on the +ve skewed distributions. These operators are most effective when the mean fitness of the population is as close to I_{MAX} as possible (for unimodal problems).

Consider Figure 7.2.1, as the –ve skew increases (tending towards 2^{nd}Best) the relative fitness of I_{MAX} becomes significantly higher than the rest the population. Hence with Roulette Wheel selection, the probability of the best few individuals being selected is proportional to the –ve skew. It is therefore almost always the best few individuals that are mutated, or crossed-over. In effect, all the "effort" or "selection pressure" of the GA is being concentrated on these top individuals. Unfortunately this concentration results in a reduction of the effective population size, and in extreme cases tends towards the classic (1+1) GA (which has a population size of 1 and simply replaces this individual when a new better individual is found through random search) [3]. However, it never actually reaches pure random search as it is still based upon mutations/crossover of the best individual and so suffers from the premature convergence problem. Hence, although Roulette wheel is relatively more effective on –ve skewed distributions compared to the other shapes. Both Rank and Random selection outperform Roulette wheel for the -ve skewed case on DJ5 (see Figure 9).

4.3 Diversity

Figure 8 shows the population diversity through the generations for various replacement operators. The initially high diversity of the randomly initialised population falls off as selection and replacement operate over the generations, imposing structure on the population. High mutation rates allow for more "random jumping" than high crossover rates and so the drop-off in the high mutation rate case is much slower. Across all selection operators and mutation/crossover fractions the performance of the replacement operators was highly consistent and so the following generalisations can be made:

- Parent replacement usually has the highest diversity (although associated poor performance), as removing the best individuals every time generates a lot of churn.
- Truncation replacement always has the lowest diversity due to population convergence (possibly suboptimally). CDRW is the next lowest after Truncation, as it is in fact a modified version of Truncation.
- DeJong crowding does extremely well under high mutation rates. It has mid range performance at high crossover rates, as in this case the children are more likely to be similar to the best individuals causing the removal of worse individuals and hence tending towards Truncation.
- The Random, Distribution based and 2^{nd}Best replacement strategies all have mid range performance. Distribution and 2^{nd}Best do much better than Random and RandKB at maintaining diversity at high crossover rates, and all have similar performance at low crossover rates.
- 2^{nd}Best usually has lower diversity than the distribution replacements, but is always much better (around twice as good) as Truncation. Even though 2^{nd}Best is effectively the mirror image to Truncation, it maintains higher diversity because there are vastly more combinations of solutions with equally bad fitness scores than there are with equal good ones. In other words, although the deviation of fitness values is similar in the two cases, the diversity of the population (as measured by Hamming Distance) is not.

These graphs graphically demonstrate the success of the distribution replacements at maintaining diversity. The fact that in many cases they are comparable with random replacement is impressive indeed.

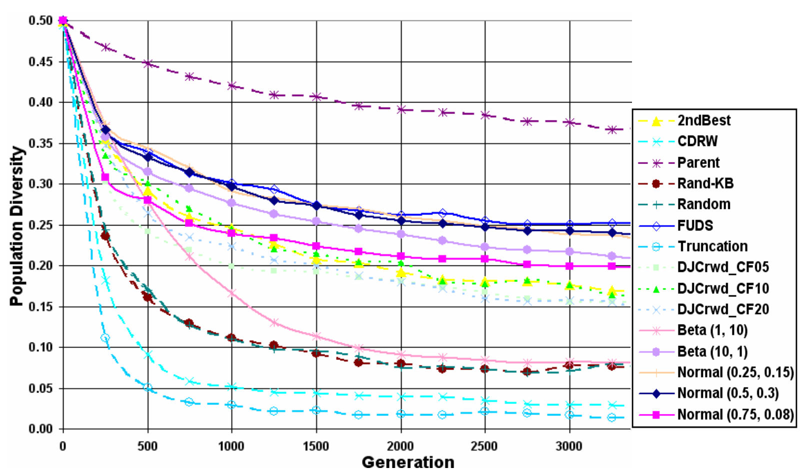


Figure 8: Population Diversity (average Hamming distance from a theoretical all 0's individual) vs Generation using Roulette Wheel selection with a crossover fraction of 95% averaged over all 5 DeJong Functions

4.4 Comparison with Classical GAs

Up to now only steady state GA's have been considered. This section compares the performance against the more common "Classical GA" that does not use a specific Replacement operator.

Figure 9 (left hand side) shows a comparison between the three selection operators used on their own, in the classical scenario, and used with a -ve skewed distribution replacement, in a steady state scenario, on DJ5. The population size was fixed at 30, however, a large range of mutation, crossover and skew parameters was simulated (each simulation is the average of 100 separate runs). For each of the different cases, the best performing combination of these parameters is presented (which was different in each case). CDRW (the best performing non-distribution Replacement operator) is also included (using its best set of parameters and best Selection operator). For the classical GA, an elitism of 1 was introduced so as to guarantee the best individual was always carried forward into subsequent generations (without this elitism the classical GA results are significantly worse).

The nature of the deceptive and multimodal DJ5 problem means that improved performance is achieved by not concentrating on the best individuals. In other words, pure random search is highly effective and is represented by the (1+1) GA. Accordingly, this is most effectively mimicked by Random followed by Rank, and finally Roulette wheel selection. For each selection operator, it can be seen that the introduction of the explicit -ve distribution replacement improves its performance (against number of fitness function evaluations) over its classical counterpart and other distribution operators. Moreover, -ve distribution replacement accounts for 3 out of 4 of the top results.

Figure 9 (right hand side) show the same comparison using the Sawtooth principle [4]. With a Sawtooth GA (either steady-state or classical) the population size is monotonically decreased each generation. When it reaches a minimum population size, it is increased to the maximum size and the monatomic decreases starts again (hence the graph of population size vs generation looks like a serrated sawtooth). When the population size is expanded to maximum, it is filled with new randomly drawn individuals. Sawtooth therefore combines the benefits of both the variable population size and multi-restart principles.

A population size of 55 decreasing to 5 (i.e. average 30) was used, however the rate of the decrease was varied. Again the very best performing set of parameters (now including rate of decrease) is presented. Sawtooth GA's clearly outperform their standard counterparts. The performance improvement found by using -ve skewed distributions is again evident and graphically demonstrated by the fact that all three distribution replacements are now able to outperform the random search of (1+1) GA. With Sawtooth, CDRW is also able to escape from its premature convergence and match the performance of –ve skew. However, as previously discussed, CDRW entails a significant and highly detrimental computation overhead that is hidden on these plots.

5. DISCUSSION

The discoveries and methodologies outlined in this paper open up some exciting possibilities for further extension. It is clear that a simple Gaussian distribution, centred near the middle of population fitness (relative to I_{MAX}), always yields sub optimal results no matter what the variance. Similarly, FUDS (the uniform distribution) also generated sub optimal results. As FUDS provided the original foundation for this work, this paper has certainly extended the understanding of this family of Distribution Replacement strategies.

The -ve skewed distributions proved to be highly successful (across a large range of crossover fractions) on the most taxing multimodal DF5. Not only did the "survival of the worst" strategy outperform other distribution shapes but it was more effective than standard classical GAs. Survival of the worst was also successful with Roulette Wheel on unimodal problems or when the fitness of the children was less dependent on the fitness of the parents (high mutation rates) which is a major characteristic of non-linear and dynamic problems.

The superior performance of the negatively skewed distributions (survival of the worst strategy) in these situations leads the author to hypothesis that this strategy of maintaining a large diversity in

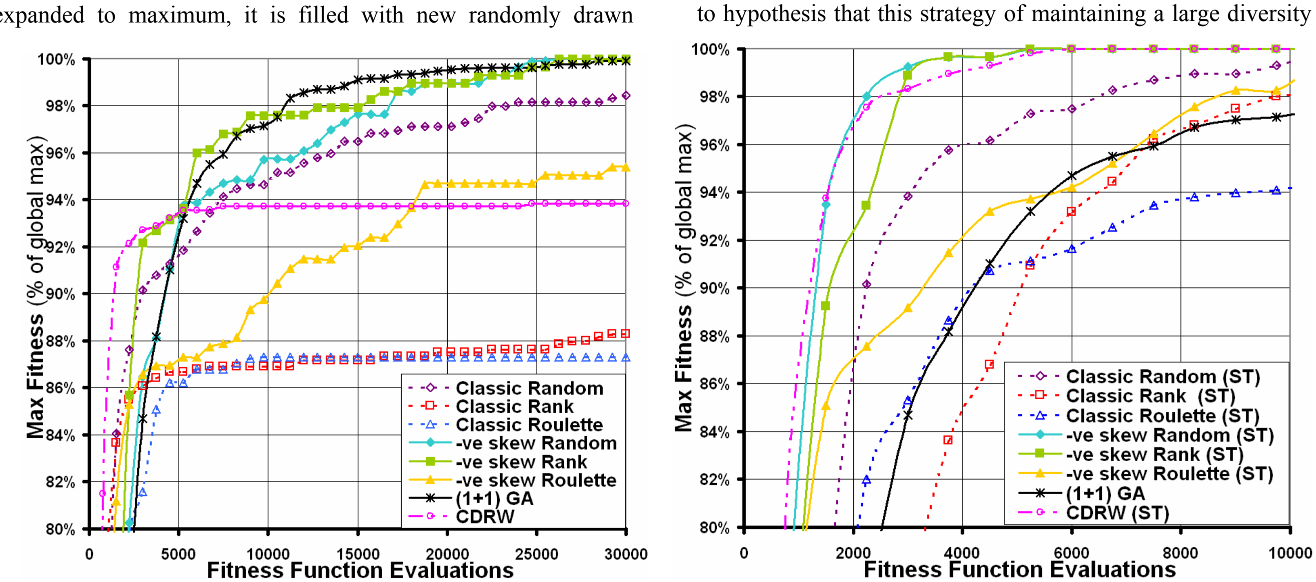


Figure 9: Comparison between performance of negatively skewed distributions and the classical GA (using the same selection function but without a replacement operator) on DJ5. Right graph is the same comparison but using the "Saw tooth" approach.

reserve will really start to out class survival of the fittest in highly complex, deceptive, non-linear and/or dynamic problems. Applications into co-evolution (where agents inherently change each other's learning environments) and "simultaneous multiple learners" multiagent systems [8] seem to be the fields which might benefit most directly from this approach. Further work with more dynamic and difficult benchmarks would be beneficial along with an ongoing analysis of different distribution shapes (e.g. non-symmetrical U-shapes).

CDRW certainly proved to be the best performer out of the non-distribution replacement strategies. However, its superior performance is coloured by the high overhead of continuously maintaining the contribution to diversity metrics that it relies upon. By contrast, distribution replacement has a significantly lower overhead and is intuitive and straightforward to implement. Of course further work is needed to empirically gauge this difference in performance against CPU overhead rather than generational, or fitness function evaluation metrics.

Clearly different distribution shapes are suited to different types of problem, and in the real world, with scant knowledge about the actual problem, the choice of distribution shapes maybe not be obvious. With a better understanding of the effect of distribution shape the possibility of a dynamic distribution could be realised (most likely on based upon a non-symmetrical U-shape). As the search progresses it effectively samples the fitness landscape and this knowledge about the characteristics of the problem can be used to continually modify the distribution to a more effective shape.

Another major area for research is into tailored selection strategies. Roulette wheel, Rank and Random selection were chosen in this paper due to their widespread usage, however they are inherently designed for a generational GA architecture, rather than the steady-state GA with explicit distribution based replacement. In this paper there was no attempt to design an optimal selection operator. Some simple reworking of the Rank and Roulette Wheel operators to truly take advantage of the explicitly controlled (and guaranteed) distribution of parents provided by Distribution Replacement is likely to produce some sizeable performance improvements.

6. CONCLUSION

This paper has taken the single example of a prior distribution based replacement strategy (FUDS) and extended it to a new family (or class) of replacement strategies. This family provides a mechanism to explicitly control the diversity of population and a simple method for its implementation. Specifically, it has shown that the only prior example of a distribution replacement strategy (FUDS) is always sub-optimal (or at best, non-uniquely optimal), on the standard benchmarking problems represented by the DeJong fitness function test set.

Analysis on the effects of distribution shape revealed that the "survival of the extremes" was a serious contender to the standard "survival of the fittest" approach in the simplest problems. The novel keep-the-worst strategy proved most effective on the taxing and deceptive problems (even outperforming some advanced classical GAs), and in situations where the fitness of children differed from the fitness of parents more radically, indicating its likely suitably on highly dynamic and challenging problems.

In the real world, the structure of such problems is generally not known a-priori and is therefore not possible to choose the optimal distribution for the problem. However, as the GA runs, knowledge about the problem space is improved which can in turn be used to optimise a dynamic population distribution for which this work provides a foundation.

7. REFERENCES

[1]. DeJong, K.A. Analysis of the Behavior of a Class of Genetic Adaptive Systems *Computer and Communication Sciences*, University of Michigan, Ann Arbor, 1975.

[2]. Hutter, M. and Idsia, M.L. Fitness uniform selection to preserve genetic diversity. *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*.

[3]. Jansen, T. and Wiegand, R.P., Exploring the Explorative Advantage of the Cooperative Coevolutionary (1+ 1) EA. in *Genetic and Evolutionary Computation*, (2003).

[4]. Koumousis, V.K. and Katsaras, C.P. A Saw-Tooth Genetic Algorithm Combining the Effects of Variable Population Size and Reinitialization to Enhance Performance. *Evolutionary Computation, IEEE Transactions on*, *10* (1). 19-28.

[5]. Legg, S. and Hutter, M. Fitness uniform deletion: a simple way to preserve diversity. *Proceedings of the 2005 conference on Genetic and evolutionary computation*. 1271-1278.

[6]. Lozano, M., Herrera, F. and Cano, J.R. Replacement strategies to preserve useful diversity in steady-state genetic algorithms. *Press, March*.

[7]. Ortiz-Boyer, D., Hervas-Martinez, C. and Garcia-Pedrajas, N. CIXL2: A Crossover Operator for Evolutionary Algorithms Based on Population Features. *Journal of Artificial Intelligence Research*, *24*. 1-48.

[8]. Panait, L. and Luke, S. Cooperative Multi-Agent Learning: The State of the Art. *Autonomous Agents and Multi-Agent Systems*, *11* (3). 387-434.

[9]. Wiegand, R.P. An Analysis of Cooperative Coevolutionary Algorithms, George Mason University, 2003.

[10]. Wikipedia (2007). Hamming Distance: http://en.wikipedia.org/wiki/Hamming_distance