Efficient Sampling with Small Populations: a Genetic Algorithm Satisfying Detailed Balance

Chris Watkins
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey, TW20 0EX, UK
C.J.Watkins@rhul.ac.uk

Yvonne Buttkewitz
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey, TW20 0EX, UK
Yvonne.Buttkewitz@rhul.ac.uk

ABSTRACT

We present a population genetic algorithm which satisfies detailed balance, and which has a stationary distribution that factorises into an explicit form for arbitrary fitness functions. For a population size of 1, it is the Metropolis algorithm with a 'breeder' proposal distribution; it extends to larger populations in a natural way, and the stationary (that is, the mutation-selection equilibrium) distribution is exactly known in a simple form for any population size. We term this algorithm exchangeable breeding tuple product sampling (EBT).

EBT is closely related to some non-parametric Bayesian Markov-chain Monte Carlo sampling algorithms. EBT can also be viewed as a generalisation of the Moran process.

Categories and Subject Descriptors

F.2 [**Analysis of algorithms and problem complexity**]: General; G.3 [**Probability and statistics**]: Probabilistic algorithms (including Monte Carlo)

Keywords

Theory; Genetic Algorithms; Evolutionary Computation; EDA; MCMC; Exchangeability

1. INTRODUCTION

Evolution can be viewed as a learning algorithm, in that populations evolve improved solutions to the problem of how to survive and reproduce. These improvements are the result of empirical experience in the sense that each organism is an experiment that is evaluated according to its lifetime reproductive success. Can evolution be connected in a natural way to existing learning algorithms, or is it a learning process of a different type? We propose some probability models of evolution that are analogous to Bayesian inference. For these models, standard MCMC sampling algorithms can have exactly the same form as evolutionary computation.

GECCO '15 July 11-15, 2015, Madrid, Spain

ACM ISBN 978-1-4503-3488-4/15/07.

DOI: http://dx.doi.org/10.1145/2739482.2764708

2. EXCHANGEABLE BREEDING MODELS

We model breeding as conditional sampling from an exchangeable probability distribution $p_B(\cdot)$ over genomes. Exchangeable distributions are surveyed in [3, 6].

In our algorithms, a new genome is always sampled conditionally given an existing population of genomes. Given a population G comprising N genomes $(g_1, \ldots, g_N)$, we may: breed children $c_1, \ldots, c_M$ sequentially by $c_{i+1} \sim p_B(\cdot|G, c_1, \ldots, c_i)$ and $p_B()$ is the probability of breeding the population with no selection. The effect of exchangeability is to allow the interchange of ancestors and descendants, which allows construction of a reversible Markov chain of populations. The exchangeable breeding (EB) models we suggest are some of the most popular probability models in machine learning.

3. EXCHANGEABLE BREEDING WITH TOURNAMENTS (EBT)

Any genetic algorithm, operating under constant conditions, generates a Markov chain of populations, since each population depends only upon the previous population. Provided there is mutation, this Markov chain is irreducible, and therefore has a unique stationary distribution[4]. The stationary distribution is a basic property of a genetic algorithm and we desire to know what it is. Unfortunately, calculating the stationary distribution even of simple genetic algorithms is hard [8].

If a Markov chain satisfies the *detailed balance* conditions, calculating the stationary distribution may be easy. Consider a Markov chain over populations; let the probability that if the current population is G, then the next population is G' be $p(G \to G')$, and let the stationary distribution be $p_\pi(G)$. The detailed balance conditions are that for all populations G, G':

$$p_\pi(G)p(G \to G') = p_\pi(G')p(G' \to G) \qquad (1)$$

Most importantly, if we know the transition probabilities $p(\cdot \to \cdot)$, and if we can find or guess a non-negative function $\pi(G)$ such that the detailed balance equations (1) hold for π in place of p_π, then it follows that the stationary distribution $p_\pi(G) = \frac{\pi(G)}{Z}$, for some normalising constant Z.

The following genetic algorithm satisfies detailed balance. Let there be a fitness function f, such that for all genomes g, $f(g) > 0$. At the start of each generation, let the current population be N genomes $g_1, \ldots, g_N$, with fitnesses $f_1, \ldots, f_N$. In each generation, one new genome is bred, and one genome is removed (the genome that is bred may also be the one that is then removed).

Algorithm 1:

1. Breed g_{N+1} by sampling from p_B conditional on the current population. That is: $g_{N+1} \sim p_B(\cdot \mid g_1, \ldots, g_N)$
2. Calculate $f_{N+1} = f(g_{N+1})$. Add g_{N+1} into the population.
3. Select one genome to remove; genome i is removed from the population with probability

$$Pr(\text{remove } g_i) = \frac{\frac{1}{f_i}}{\frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}} \tag{2}$$

The genomes $\{g_1, \ldots, g_{N+1}\} \setminus \{g_i\}$ become the next population of N genomes.

THEOREM 1. *If p_B is exchangeable, Algorithm 1 defines a Markov chain of (overlapping) populations that satisfies detailed balance, and the stationary distribution is proportional to π, where*

$$\pi(g_1, \ldots, g_N) = p_B(g_1, \ldots, g_N) f(g_1) \cdots f(g_N) \tag{3}$$

PROOF. Let $G = (g_1, \ldots, g_N)$, $G^+ = (g_1, \ldots, g_{N+1})$, $G' = (g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_{N+1})$ and $R_f = \frac{1}{f_1} + \cdots + \frac{1}{f_{N+1}}$. Observe that

$$p(G \to G') = \frac{p_B(g_{N+1} \mid G)}{f_i R_f}, \text{ and } p(G' \to G) = \frac{p_B(g_i \mid G')}{f_{N+1} R_f}.$$

$$\text{So } \pi(G)p(G \to G') = p_B(G) f_1 \cdots f_N \; p_B(g_{N+1} \mid G) \frac{\frac{1}{f_i}}{R_f}$$

$$= p_B(G^+) \frac{f_1 \cdots f_{N+1}}{f_i f_{N+1}} \frac{1}{R_f}$$

which is symmetric in f_i, f_{N+1}, so
$\pi(G)p(G \to G') = \pi(G')p(G' \to G)$. $\square$

Algorithm 1 can be extended to produce $K > 1$ children in each generation, and to have tournament selection similar to, but also different from types of tournament selection commonly used in genetic algorithms, as in [1].

3.1 Relations to existing evolutionary models and nature

EBT with a population of $N = 1$ is exactly the Metropolis algorithm with proposal distribution $p_B(\cdot \mid g_1)$, and the stationary distribution is $\pi(g) \propto p_B(g)f(g)$. EBT therefore seems a natural generalisation of Metropolis-Hastings to a population algorithm. Note that EBT is quite different from the 'Differential Evolution' algorithm of [7]: in our notation, differential evolution has stationary distribution proportional to $f(g_1) \cdots f(g_N)$, whereas the stationary distribution of EBT is proportional to $p_B(g_1, \ldots, g_N) f(g_1) \cdots f(g_N)$.

Our algorithm 1 is an overlapping-generations Moran process. The more usual Wright-Fisher (WF) process is a non-overlapping generations model, in which each new generation consists entirely of children none of the previous parents: the genetic algorithms of [2] were of this type.

A significant difference between EBT and WF is that for small population sizes, the level of fitness in WF sampling is biased downwards as compared to EB. To obtain insight into the biased sampling of WF for small populations, consider a population of size 1. EBT with a population of 1 is the Metropolis-Hastings algorithm with proposal distribution $p_B(\cdot \mid g_1)$, and the stationary distribution is $\pi(g) \propto p_B(g)f(g)$. An analogous WF GA would have a different sampling procedure; in each generation, it would generate 2 independent children $g_2, g_3 \sim p_B(\cdot \mid g_1)$, and it would then select the next generation g_1' to be g_2 with probability $\frac{f_2}{f_2+f_3}$ and g_3 otherwise. If $g_1 \sim \pi(\cdot)$, where π is stationary distribution of EBT, then the expected fitness of the next GA sample $E(f(g_1') \mid g_1) \leq E(f(g_1))$, for any f. We therefore conjecture that the mean log fitness in the stationary distribution of the GA is lower than for EBT for all fitness functions; this is hard to prove because the stationary distribution of the GA is itself hard to characterise. For sufficiently large population, and sufficiently weak selection, this sampling bias becomes small, and algorithms that replace their populations at each generation will have similar behaviour to EBT. For many natural species, biological evolution itself is such a population-replacement algorithm. However, there appears to be no virtue in a literal simulation of nature here: unbiased sampling enables small populations to have similar behaviour to large populations.

3.1.1 Non-parametric Bayesian MCMC inference

The factorisation for the stationary distribution of EBT:

$$\pi(G) \propto p_B(G) \prod_{i=1}^{N} f(g_i) \tag{4}$$

is reminiscent to the equation for a Bayes posterior distribution, with p_B as the prior, f as the likelihoods, and G as the latent parameters. In Gibbs-within-Metropolis sampling schemes, we may specify a different fitness function for each of the N elements.In non-parametric clustering using a Dirichlet Process prior, as in [5], genomes correspond to cluster-parameters, and $f_i(g_i)$ is the likelihood of cluster-parameter g_i according to data case i. Neal's algorithm 5 in [5] can in be regarded as an evolutionary algorithm, similar to EBT. Many non-parametric distributions used in Bayesian MCMC Gibbs-sampling schemes are possible exchangeable distributions that can be used with EBT.

4. REFERENCES

[1] D. Goldberg. *Genetic algorithms in search, optimization, and machine learning.* Addison-wesley, 1989.

[2] J. Holland. *Adaptation in natural and artificial systems.* University of Michigan press, 1975.

[3] O. Kallenberg. *Probabilistic symmetries and invariance principles.* Springer Science & Business Media, 2006.

[4] J. G. Kemeny and J. L. Snell. *Finite Markov chains: with a new appendix" Generalization of a fundamental matrix".* Springer, 1976.

[5] R. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, pages 249–265, 2000.

[6] P. Orbanz and Y. W. Teh. Bayesian nonparametric models. In *Encyclopedia of Machine Learning*, pages 81–89. Springer, 2010.

[7] C. J. Ter Braak. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249, 2006.

[8] M. Vose. *The simple genetic algorithm: foundations and theory.* The MIT Press, 1999.