Sex as Gibbs Sampling: Modelling Evolution with a Tractable Markov Chain

Chris Watkins
Yvonne Buttkewitz

Department of Computer Science
Royal Holloway University of London

{C.J.Watkins, Yvonne.Buttkewitz}@rhul.ac.uk

January 2013

Plan of Talk

1. Take evolutionary inspiration seriously

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*
   - *prior* is the equilibrium distribution under breeding with no selection (ie drift). Standard population genetics.

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*
   - *prior* is the equilibrium distribution under breeding with no selection (ie drift). Standard population genetics.
   - Reversible Markov chain: energy function, detailed balance

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*
   - *prior* is the equilibrium distribution under breeding with no selection (ie drift). Standard population genetics.
   - Reversible Markov chain: energy function, detailed balance
   - Can compute un-normalised equilibrium probability of any given population.

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*
   - *prior* is the equilibrium distribution under breeding with no selection (ie drift). Standard population genetics.
   - Reversible Markov chain: energy function, detailed balance
   - Can compute un-normalised equilibrium probability of any given population.
4. Standard MCMC methods give many algorithms that converge to same equilibrium distribution: superevolution?

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*
   - *prior* is the equilibrium distribution under breeding with no selection (ie drift). Standard population genetics.
   - Reversible Markov chain: energy function, detailed balance
   - Can compute un-normalised equilibrium probability of any given population.
4. Standard MCMC methods give many algorithms that converge to same equilibrium distribution: superevolution?
5. A moment of embarrassment: model looks too simple.

Plan of Talk

1. Take evolutionary inspiration seriously
2. Propose a simple evolutionary algorithm
3. Observe that in this algorithm,
   breeding $\equiv$ Gibbs sampling on a probability model.
   - Explicit representation of mutation-selection equilibrium distribution, factorised as *prior* $\times$ *fitness*
   - *prior* is the equilibrium distribution under breeding with no selection (ie drift). Standard population genetics.
   - Reversible Markov chain: energy function, detailed balance
   - Can compute un-normalised equilibrium probability of any given population.
4. Standard MCMC methods give many algorithms that converge to same equilibrium distribution: superevolution?
5. A moment of embarrassment: model looks too simple.
6. Suggest tentatively how *prior* might explain some aspects of effectiveness of evolution of sexual populations.

Evolution

- A Powerful Learning Algorithm
  - spontaneously self-optimised
  - remarkably fast, considering level of complexity produced
- Sexual Evolution
  - All organisms that you can actually see are from sexual lineages
  - Evolution of sexual populations astonishingly effective.
  - We will consider a model of sexual evolution.
- We know how evolution works
  - Genetic recombination, mutation, and selection are well understood.
  - Co-regulation of gene expression and development: outlines known.
- Evolution breaks every software engineering rule
  - flat syntactic representation, with random re-assortments
  - no protection of code inside modules: any transcription factor can connect to any gene
  - more than 13% of cloned code
  - only full-system testing, but plenty of it.

Modelling Evolution

Find a model that

- is simple enough to analyse and be widely applicable
- models what is *essential* to effectiveness of evolution,
- leaves out *accidental* aspects of genetic mechanisms

(Of course we can only know which aspects of genetic mechanisms are essential, and which accidental, after looking at the success of a model...)

Evolution as a Markov Chain

Assuming constant environment, evolution is naturally viewed as a Markov chain.

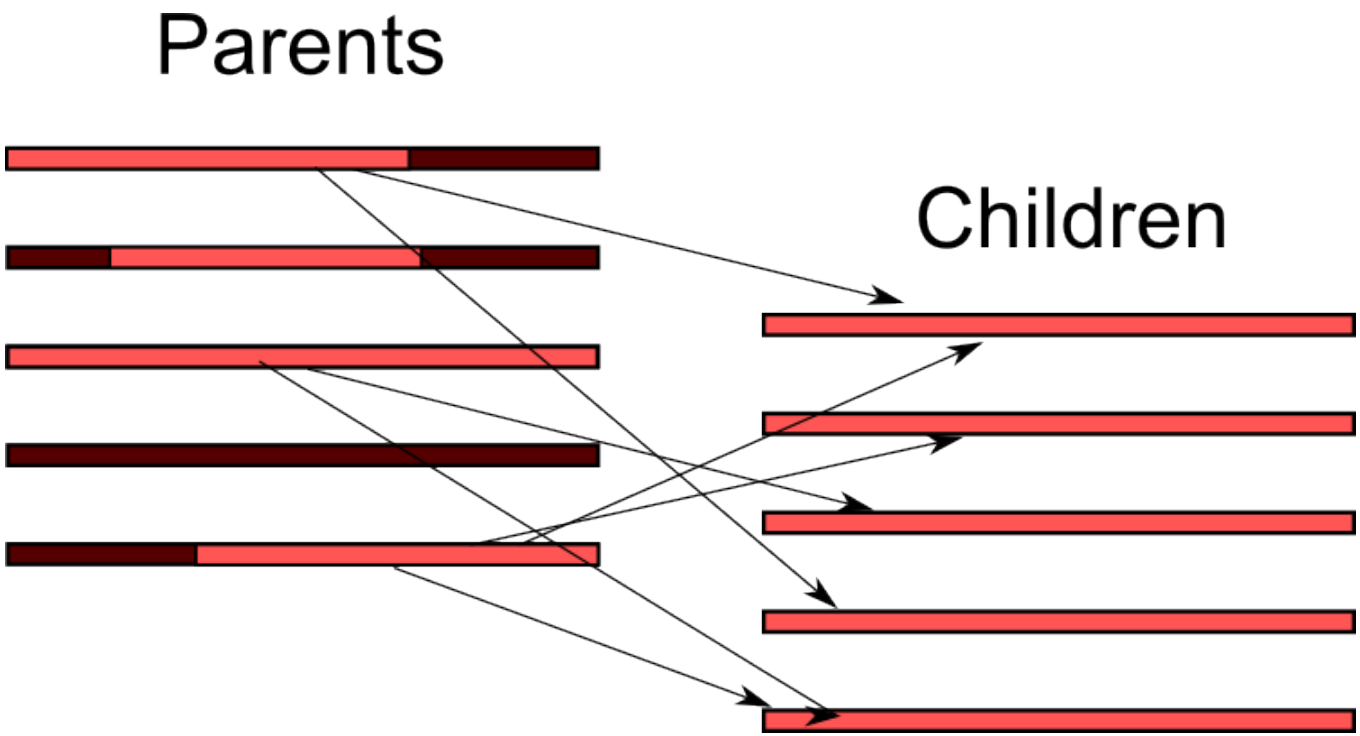
- *states* of Markov chain are *populations*
- *transitions* between populations are episodes of breeding, mutation, and selection

Mutation ensures Markov chain is connected: therefore there is unique equilibrium distribution over populations.

Seek to characterise this *mutation-selection equilibrium*

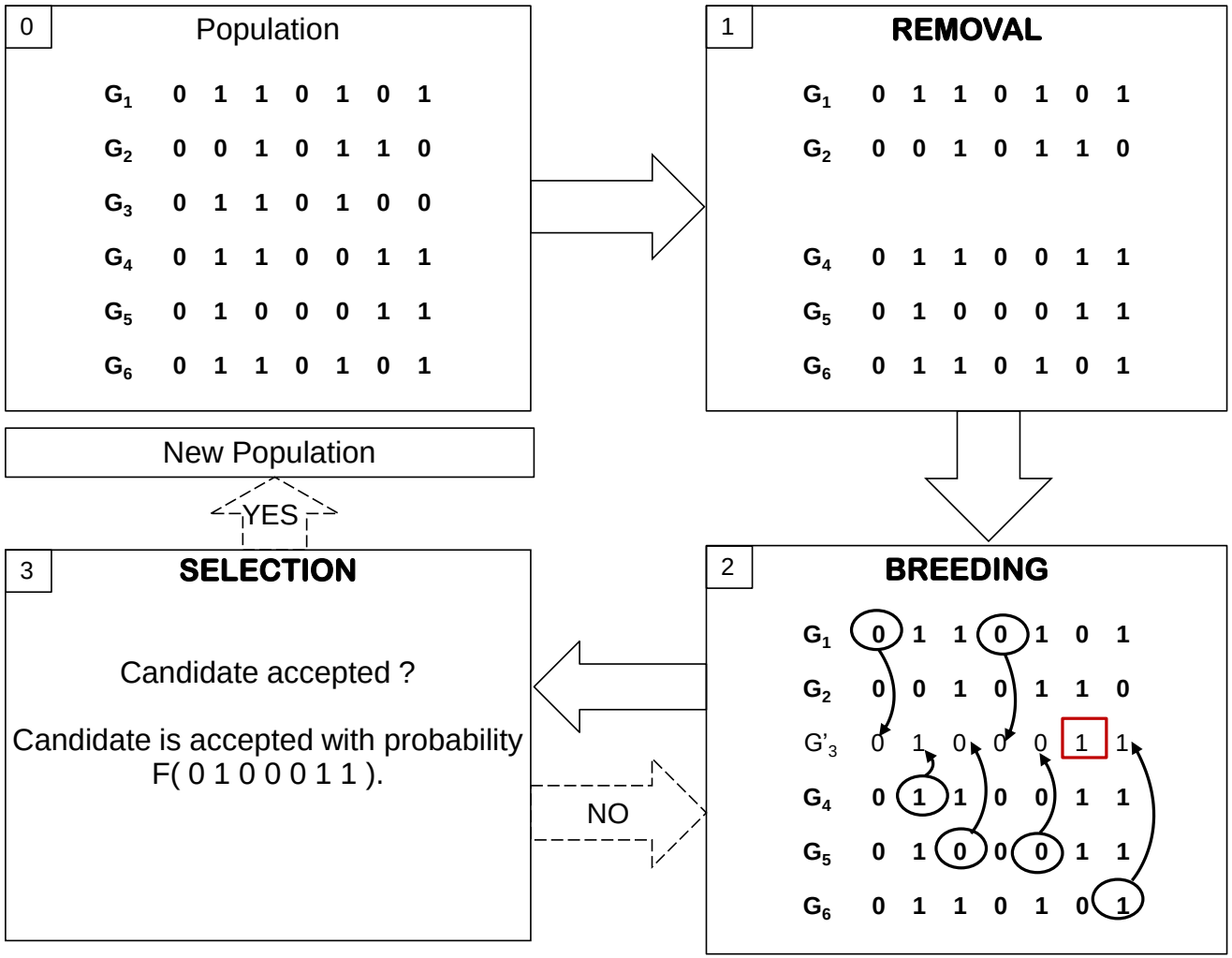
(EC may be used for optimisation: we concentrate on the equilibrium distribution of the Markov chain.)

Irreversibility of Breeding, in GAs and in Nature



Every child sequence is a direct copy of some parent sequence, but not vice-versa.
Hence $p(parents|children) \neq p(children|parents)$
Markov chain not reversible; no detailed balance.
Looks complicated.

Gibbs Breeding Algorithm, or 'Die-First Moran Process'



Key Simplifications

- Linkage equilibrium/N-way crossover
  - In nature (and most GAs) each child is generated from two parents; here, each child generated from whole population.
  - Same assumption as UMDA, RPP, PBIL and its variants, simple EDA, also made by Mackay, Baum.
  - No compelling reason to suppose that breeding from 2 parents with restricted crossover is any better?
- Mutation as Gibbs sampling with existing alleles
  - Beta-binomial distribution if alleles are 0/1
  - Dirichlet process is simplest approach: other mutation distributions possible, e.g. Dirichlet diffusion process
- $0 < F(\mathbf{g}) \leq 1$, breeding by proposal and rejection

Defensible model of evolution with nice properties.
Actual evolution, or other evolutionary algorithms, will not have these properties exactly.

A Tiny Genetic Algorithm: only one gene

Values	Action	Proposed value for θ_k is
θ_1		
θ_2		$\tilde{\theta}_k \sim (1-u)\mathrm{Uniform}(\theta_{:\setminus k})+uH$
$\vdots$		
θ_{k-1}		where $u > 0$ is the
	resample	*mutation rate* and
θ_{k+1}		the p.d. H is the *mutation*
$\vdots$		*distribution*
θ_N		

Models of this type proposed by Ewens, and Kingman ('House of Cards' model) in 1970s

A Tiny Gibbs Breeding Algorithm: Finite Sample from Dirichlet Process

Values	Action
θ_1	
θ_2	
$\vdots$	
θ_{k-1}	
	resample
θ_{k+1}	
$\vdots$	
θ_N	

Proposed value for θ_k is

$$\tilde{\theta}_k \sim \frac{\text{Uniform}(\theta_{:\setminus k})}{N-1+\alpha} + \frac{\alpha}{N-1+\alpha} H$$

where $\alpha > 0$ is the *concentration parameter* and
the p.d. H is the *base distribution*

Breeding is Gibbs sampling from predictive distribution of Dirichlet Process – identical to 'tiny GA' if we match up the parameters.

Mutation Rate and Concentration Parameter

α is a function of mutation rate u and population size N:

$$u = \frac{\alpha}{N - 1 + \alpha}, \quad \text{that is,} \quad \alpha = (N - 1)\frac{u}{1 - u}$$

Role of concentration parameter

$\theta_1, \ldots, \theta_N$ will contain repeated values.

Since $\theta_1, \ldots, \theta_N$ are exchangeable, for any $i \neq j$, we can suppose they are the first and second vales of the generative Polya Urn process, so:

$$p(\theta_i \neq \theta_j) = \frac{\alpha}{\alpha + 1}$$

This does not depend on N.

Regimes with small $\alpha < 1$ are of interest because repeated values are probable.

Tiny GBA with selection (by rejection sampling)

Instead of accepting $\tilde{\theta}_k$ immediately, resample θ_k by:

repeat generate proposal $\tilde{\theta}_k$
until $\tilde{\theta}_k$ is accepted with probability $f(\tilde{\theta}_k)$

where $0 < f(\theta) \leq 1$ is a *fitness function*.

Equilibrium joint probability factorises as:

$$p(\theta_1, \ldots, \theta_N) = \frac{1}{Z(N, H, \alpha, f)} DP(\theta_1, \ldots, \theta_N | H, \alpha) \prod_{i=1}^{N} f(\theta_i)$$

Acceptance may also be done by Metropolis-Hastings rule.

Key Property: Gibbs Breeding

Consider populations $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_N)$ and $\mathbf{G}' = (\mathbf{g}_1, \ldots, \mathbf{g}'_\mathbf{k}, \ldots, \mathbf{g}_N)$.
Breed using distribution over populations q in which genomes are exchangeable.

$$p(\mathbf{G} \to \mathbf{G}') = C\ q(\mathbf{g}'_k | \mathbf{G}_{:\setminus k}) f(\mathbf{g}'_k)$$
$$p(\mathbf{G}' \to \mathbf{G}) = C\ q(\mathbf{g}_k | \mathbf{G}_{:\setminus k}) f(\mathbf{g}_k)$$

Then $\pi(\mathbf{G}) = \frac{1}{Z} q(\mathbf{G}) \prod_{i=1}^{N} f(\mathbf{g}_i)$, and

$$\pi(\mathbf{G}) p(\mathbf{G} \to \mathbf{G}') = \frac{C}{Z} q(\mathbf{G}_{:\setminus k}) q(\mathbf{g}_k | \mathbf{G}_{:\setminus k}) q(\mathbf{g}'_k | \mathbf{G}_{:\setminus k}) \left(\prod_{i=1}^{N} f(\mathbf{g}_i) \right) f(\mathbf{g}'_k)$$
$$= \pi(\mathbf{G}') p(\mathbf{G}' \to \mathbf{G})$$

giving detailed balance.

Marginals of a Dirichlet Process

What is the effect of the fitness function on the marginal distribution of θ?
For large α, nearly all elements will be distinct:

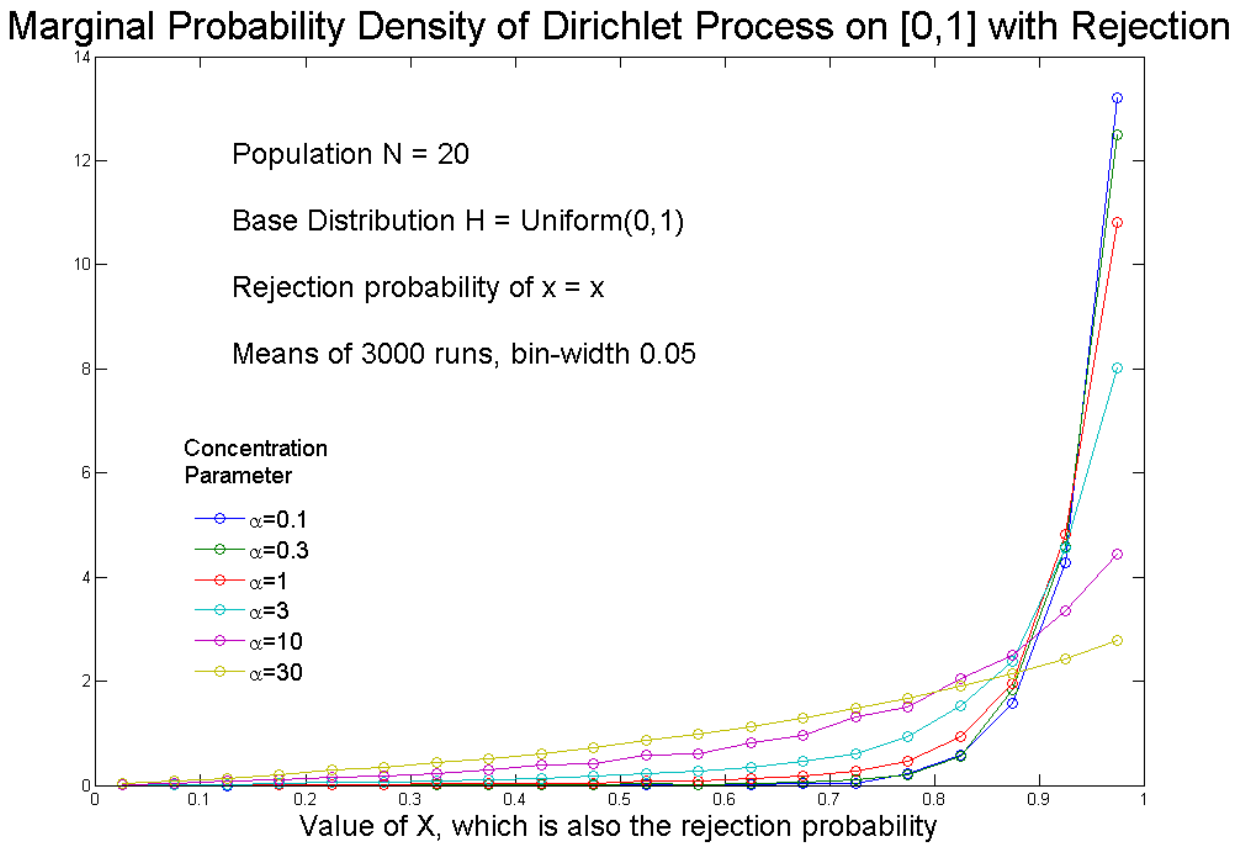
$$p(\theta) \propto H(\theta) f(\theta)$$

As $\alpha \to 0$, nearly all elements will be identical:

$$p(\theta) \propto H(\theta) f(\theta)^N$$

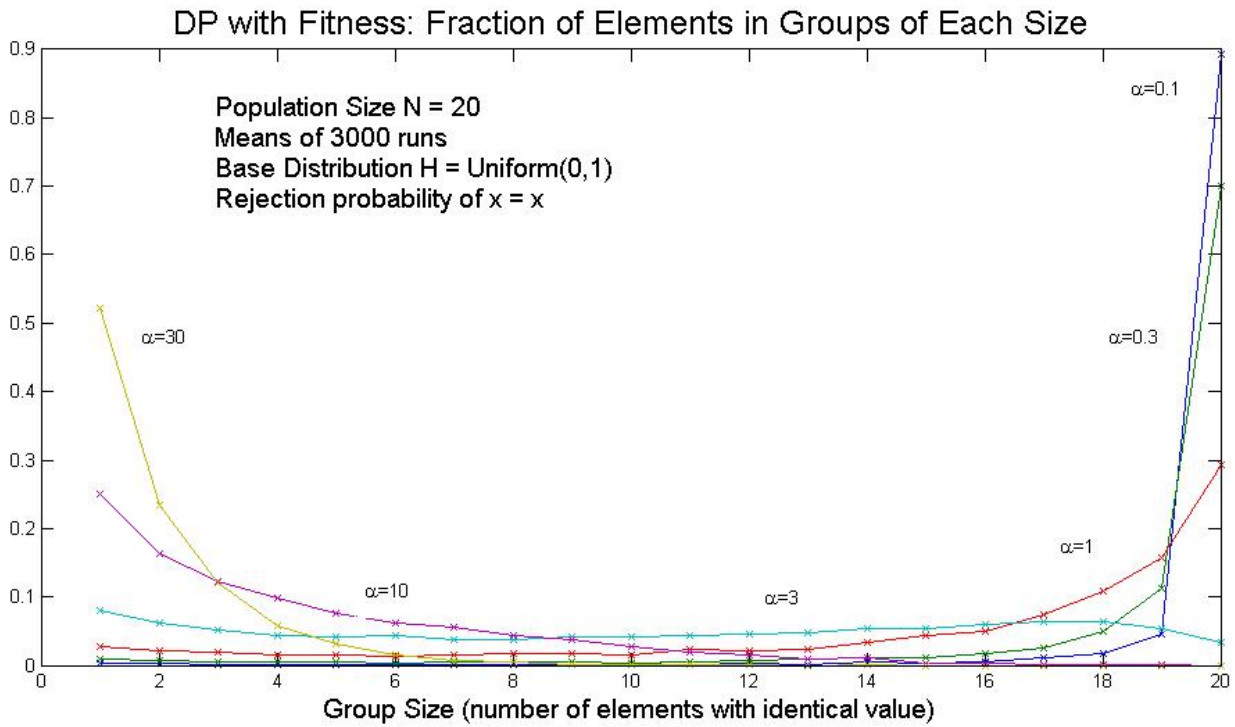
For intermediate α ... do MCMC

Marginals of a Dirichlet Process

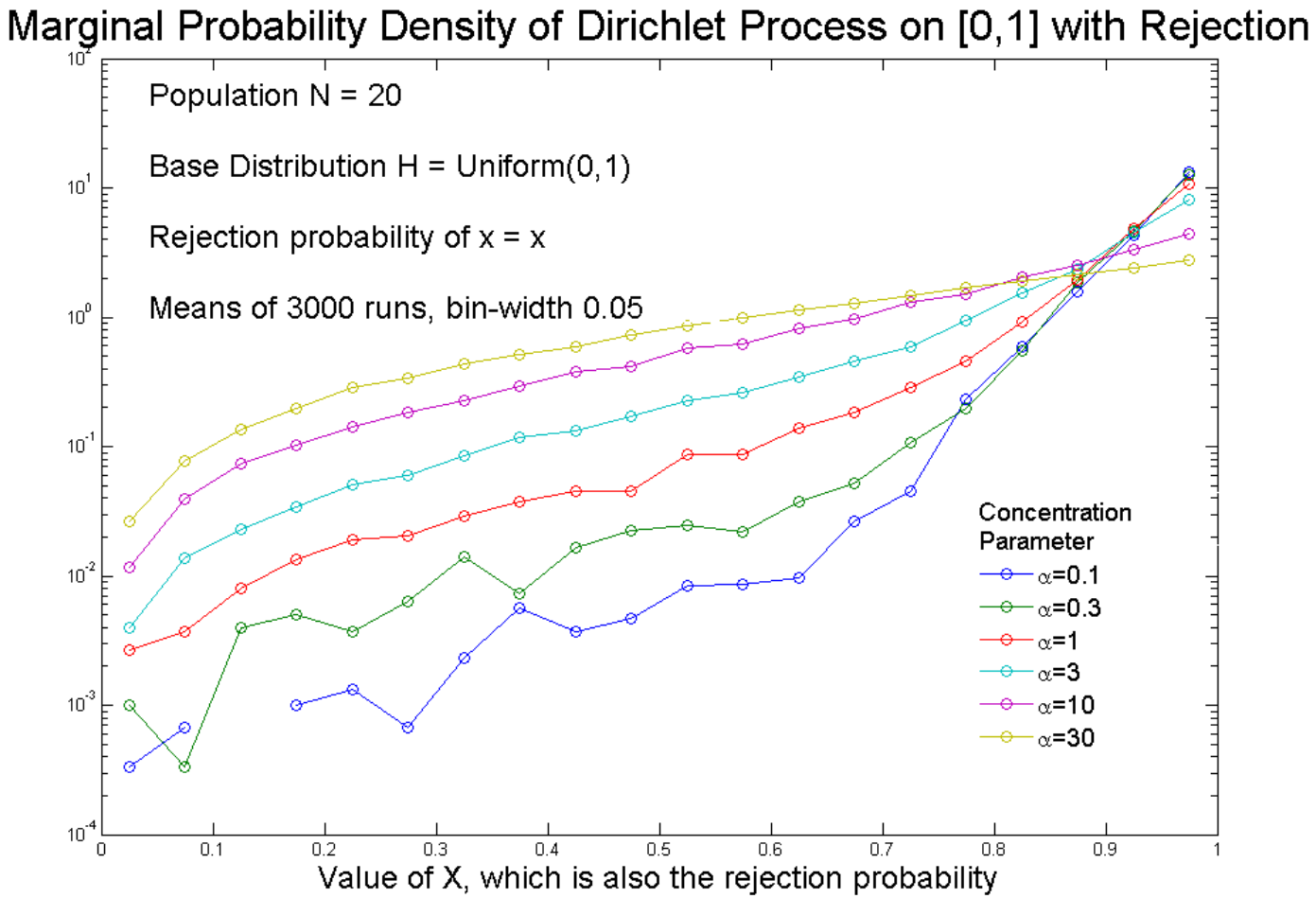


$x \in [0, 1]$, and $f(x) = x$. Marginal histograms for samples of size 20, for values of α between 0.1 and 30.

Cluster sizes of a Dirichlet Process



Marginals of a Dirichlet Process

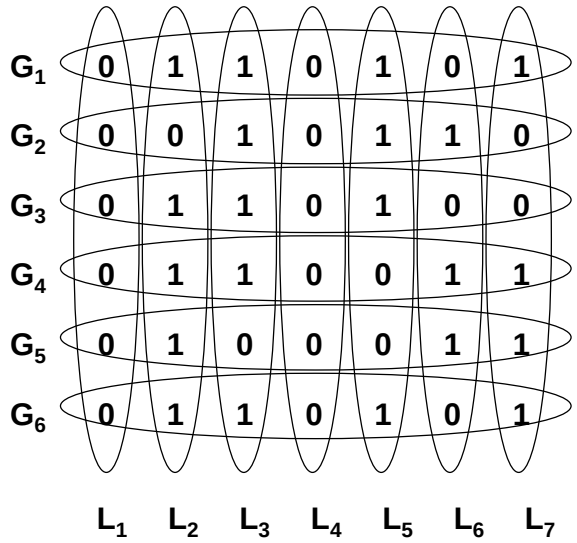


Putting it all together: relational Gibbs Breeding Models

$$
\begin{array}{ccccccc}
\theta_{11} & \theta_{12} & \cdots & \theta_{1k} & \cdots & \theta_{1L} & \rightarrow & f_1 \\
\theta_{21} & \theta_{22} & \cdots & \theta_{2k} & \cdots & \theta_{2L} & \rightarrow & f_2 \\
\vdots & & \ddots & \vdots & & \vdots & & \vdots \\
\theta_{m1} & \theta_{m2} & \cdots & \theta_{mk} & \cdots & \theta_{mL} & \rightarrow & f_m \\
\vdots & & & \vdots & \ddots & \vdots & & \vdots \\
\theta_{N1} & \theta_{N2} & \cdots & \theta_{Nk} & \cdots & \theta_{NL} & \rightarrow & f_N
\end{array}
$$

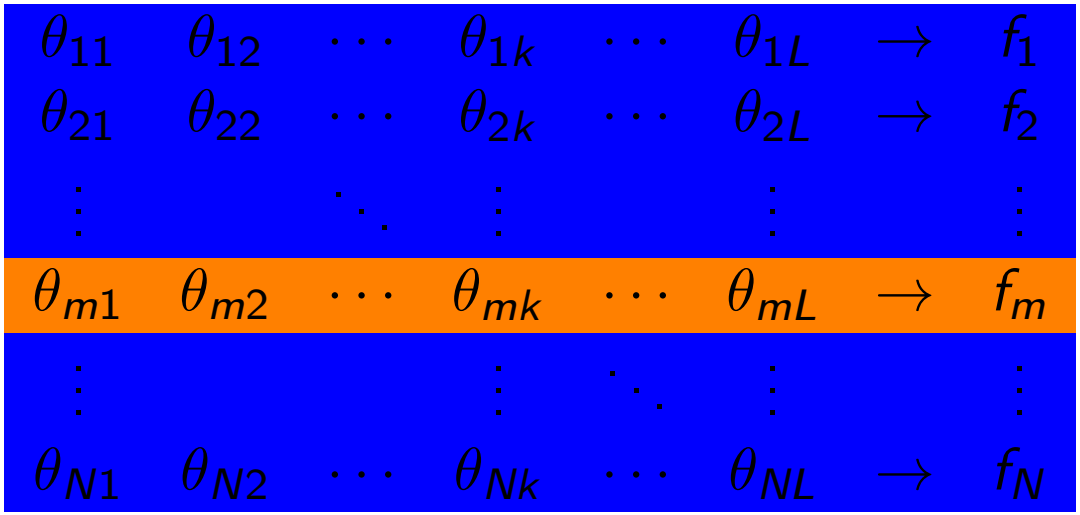
Each column is an independent tiny GBM.
Each fitness function applies to a whole row, thus linking the columns.

Factor Graph



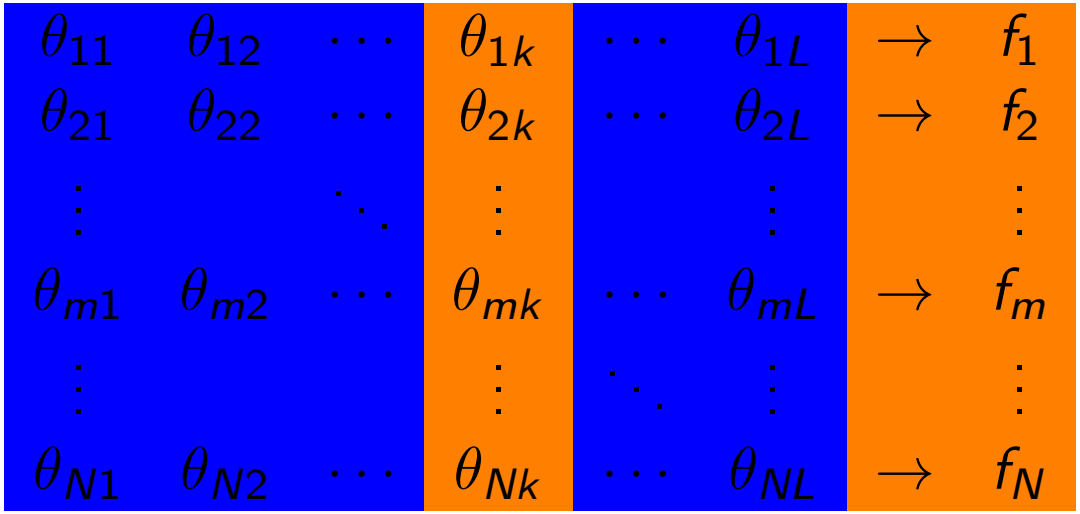
$$\text{Joint probability} \propto \prod_i P(L_i) \prod_j F(G_j)$$

Gibbs Breeding with Selection



repeat propose entire row $\tilde{\theta}_{k1}, \ldots, \tilde{\theta}_{kN}$
until new row accepted with probability $f_k(\tilde{\theta}_{k1}, \ldots, \tilde{\theta}_{kN})$

Column view of GBM



Conditional on the other columns, column k is a Dirichlet mixture model, with likelihoods given by the fitnesses.

Any of the MCMC methods for DP mixtures may be used on each column in turn.

We have a 'light-weight' slice-sampling approach.

Column Orientated Sampling

$p(x)$ distribution we want to sample from
$\mathcal{C} \subseteq \mathcal{P}(\mathcal{X})$ a set of subsets
$n_{\mathcal{C}}(x) := |\{c \in \mathcal{C} : x \in c\}|$ number of subsets that contain x

$$q(x, c) := \begin{cases} p(x) n_{\mathcal{C}}(x)^{-1} & \text{if } x \in c \\ 0 & \text{otherwise.} \end{cases}$$

- q has the right marginals

$$\sum_{c \in \mathcal{C}} q(x, c) = \sum_{\substack{c \in \mathcal{C} \\ x \in c}} q(x, c) = \sum_{\substack{c \in \mathcal{C} \\ x \in c}} p(x) n_{\mathcal{C}}(x)^{-1} = p(x)$$

- q is a proper probability distribution

$$\sum_{x \in \mathcal{X}} (\sum_{c \in \mathcal{C}} q(x, c)) = \sum_{x \in \mathcal{X}} p(x) = 1$$

Column Orientated Sampling

Sample alternately from the conditional distributions

$$q(x|c) = \frac{q(x,c)}{\sum_{y\in\mathcal{X}} q(y,c)} = \frac{p(x)n_{\mathcal{C}}(x)^{-1}}{\sum_{y\in c} p(y)n_{\mathcal{C}}(y)^{-1}},$$

and

$$q(c|x) = \frac{q(x,c)}{\sum_{b\in\mathcal{C}} q(x,b)} = \frac{p(x)n_{\mathcal{C}}(x)^{-1}}{n_{\mathcal{C}}(x)\cdot p(x)n_{\mathcal{C}}(x)^{-1}} = n_{\mathcal{C}}(x)^{-1}.$$

The update step $x \to x'$ leaves $q(x,c)$ invariant !

First test: 3-SAT problems

Standard collection of satisfiable 3-SAT constraint problems from DIMACS

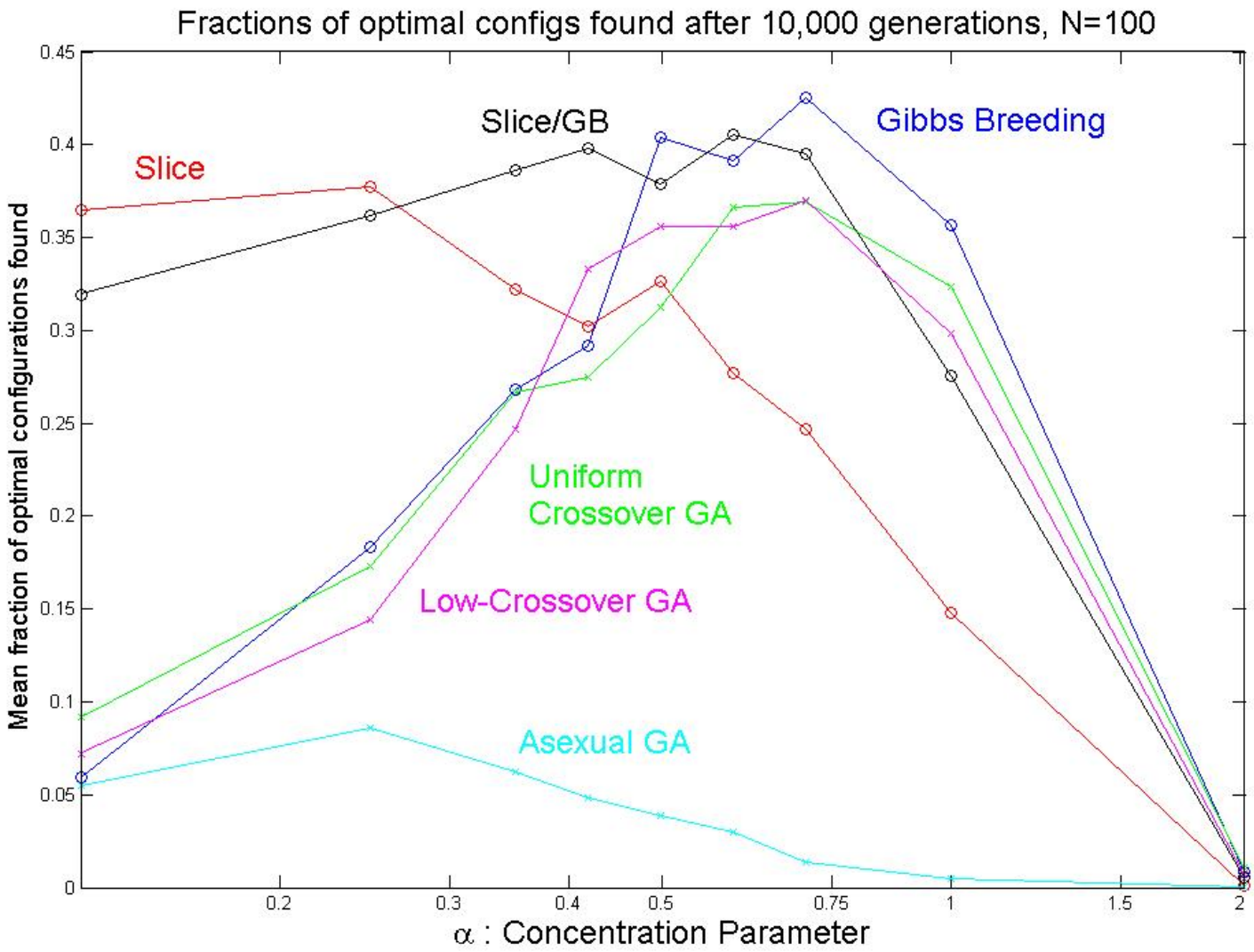
50 logical variables; 491 disjunctions of random sets of 3 literals

'Fitness' proportional to number of satisfied disjunctions

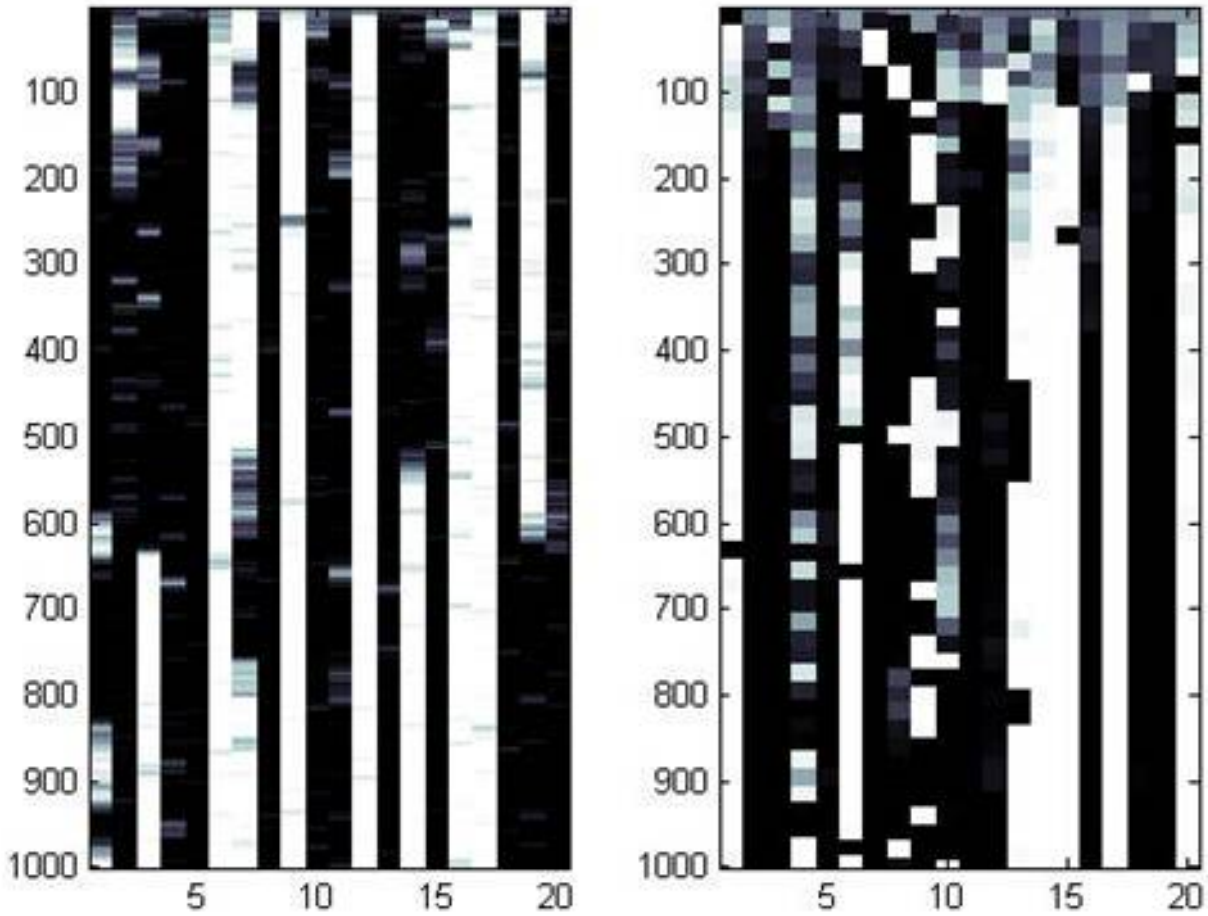
Comparison of 'standard GA' approach and GBM with row and column-oriented sampling.

Rate of mixing assessed by fraction of solutions found after 10^6 fitness evaluations

Performance of GBM and GAs on MaxSat problems



Population means by generation



Numbers of unique optimas found

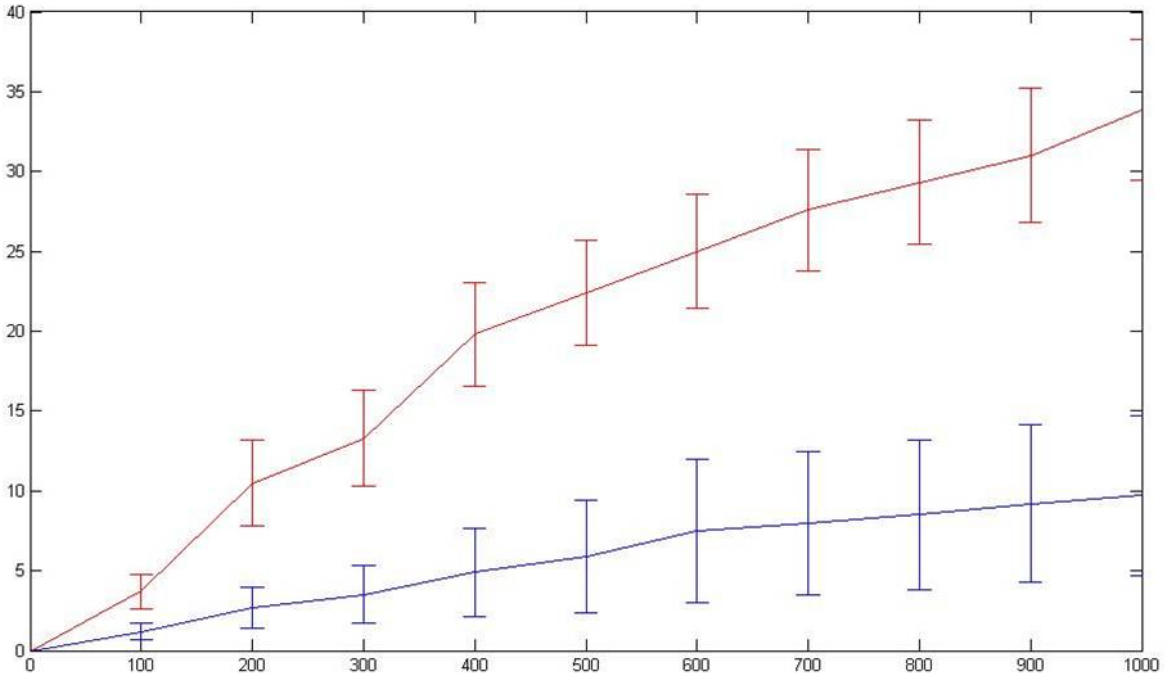


Figure : numbers of unique optimas found

The Moment of Embarrassment

- GBA novel?

The Moment of Embarrassment

- GBA novel?
- Surely GBA too simple to explain effectiveness of evolution? Is that it?

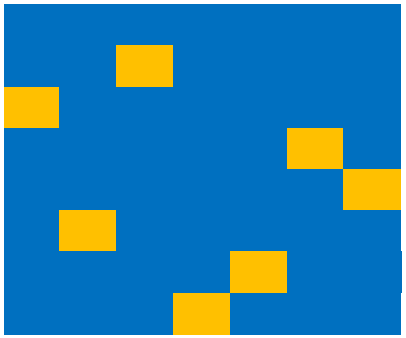
Hypotheses for effectiveness of sexual evolution?

Any reasonable hypothesis needs to apply to nearly all sexual populations...

1. GBA model too simple: crossover / 2 parents / some other genetic details necessary?
2. Mechanisms for gene expression and morphogenesis provide rich search space for possible organisms?
3. Coevolution: ecology and symbiosis/parasitism are important effects? (BUT: does not seem to explain individual complexity...)
4. GBA prior distribution has strong effects...

Sexual prior prefers combinatorial allele codes

Separate mutation for each F



F1
F2
F3
F4
F5
F6
F7
F8

Combinatorial code of mutants

8 polymorphic loci

3 polymorphic loci

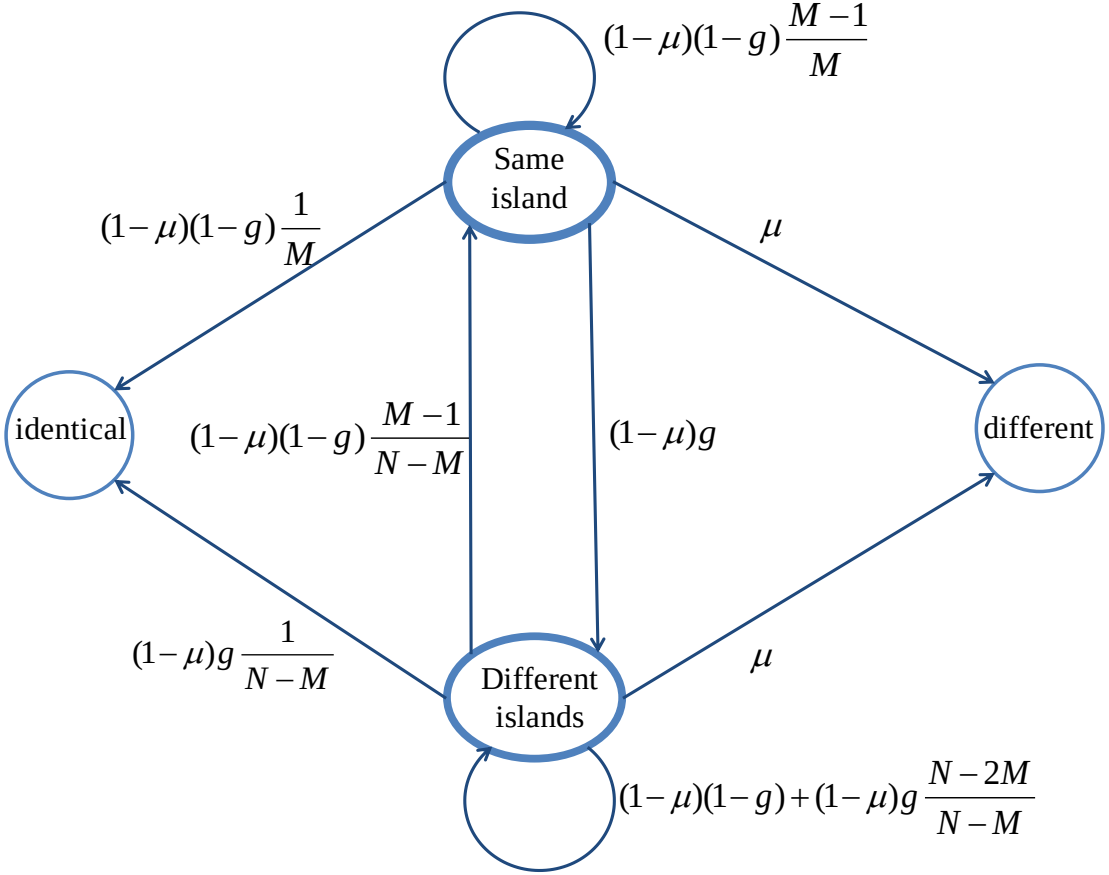
Factor Model of Evolutionary Computation (GBM)

Condition 1	θ_{11}^1	θ_{12}^1	$\cdots$	θ_{1j}^1	$\cdots$	θ_{1L}^1	$\rightarrow$ f_1^1
	$\vdots$					$\vdots$	$\vdots$
	θ_{M1}^1	θ_{M2}^1	$\cdots$	θ_{Mj}^1	$\cdots$	θ_{ML}^1	$\rightarrow$ f_2^1
	$\vdots$						$\vdots$
Condition K	θ_{11}^K	θ_{12}^K	$\cdots$	θ_{1j}^K	$\cdots$	θ_{1L}^K	$\rightarrow$ f_1^K
	$\vdots$					$\vdots$	$\vdots$
	θ_{M1}^K	θ_{M2}^K	$\cdots$	θ_{Mj}^K	$\cdots$	θ_{ML}^K	$\rightarrow$ f_M^K

- Joint probability factorises as column scores $\times$ row scores.
- Various MCMC approaches (both column and row based)
- Columns have discrete value sets; conditions must share values.
- Develops compromise solutions to multiple conditions using common components.

Markov Chain for Island Migration Model

N # genomes, M # genomes on 1 island, μ mutation rate, g migration rate



$$D = \frac{1+\lambda\frac{M-1}{M}}{1+\lambda}S \quad \text{where} \quad \lambda = (1-\mu)(\frac{N}{N-M}g - 1)$$

Summary

"Gibbs Breeding Model" (GBM) : evolution as reversible Markov Chain, with energy function and detailed balance

Superevolution : alternative MCMC methods for GBM with DP, with accelerated mixing in low-α regime.

Sexual evolution does *not* perform pure fitness optimisation: a strongly concentrated prior forces compromise solutions with many shared elements for different fitness conditions.

Quasi-Gibbs Breeding : implementation with representations of marginal distributions, not explicit populations

Hierarchical models : Island process vs hierarchical DP