

1.   • I will talk about my thesis:  
           “Simulation of Simple Evolutionary System”
2. Part I
3.   • I will give background, and computations involving our model in the simulations we did.  
       • Then I will address these four questions, and make concluding remarks.
4.   • Population is a collection of length  $\ell$  binary strings  
       • Population can be represented by a vector; the  $j$ th component is the proportion of string  $j$  in the population.  
       •  $\mathcal{R}$  is the set of length  $\ell$  binary strings and operations under  $\mathcal{R}$  are bitwise modulo 2 operations.
5.   • Crossover and mutation operators are defined using the bitwise operations in  $\mathcal{R}$ .  
       • Crossover exchanges bits in parents  $u$  and  $v$  using crossover mask to produce children  $u'$  and  $v'$  using the rule  $u' = um + v\bar{m}, v' = u\bar{m} + vm$   
       •  $\chi_m$  is probability of using crossover mask  $m$   
       •  
       • Mutation flips bits in  $x$  using mutation mask  $m$ .  
       •  $\mu_m$  is probability of using mutation mask  $m$
6.   • This flowchart illustrates finite population genetic algorithm  
       • Start from random initial population  $p$ .  
       • Randomly select parents  $u$  and  $v$   
       • Crossover  $u$  and  $v$  to produce  $u'$  and  $v'$  according to randomly chosen crossover mask

- Keep one of  $u'$ ,  $v'$  and mutate using randomly chosen mutation mask to produce gamete  $g$
  - Repeat to form next generation  $\tau(p)$
7. • In Random Heuristic Search,
- Given population  $p$ ,  $\tau$  is transition rule that maps  $p$  to the next generation  $p'$  where  $p$  and  $p'$  both belong to population space  $\Lambda_N$ .  $N$  is population size.
  - Finite population evolution forms Markov chain
  - However,  $\tau$  is stochastic function of crossover and mutation and  $\tau(p)$  can not be predicted with certainty
8. • In infinite population model, population is modeled by vector  $p$  where  $p$  belongs to population space  $\Lambda$ .
- $\mathcal{G}$  is a function that maps  $p$  to the next generation  $p'$  where  $j$ th component is proportion of string  $j$  occurs in the next generation.
  - The sequence shows evolution of  $p$  under infinite population model.
  - The variance is expectation of square of distance between finite population and infinite population.
9. • This is our model setup for diploid population.
- We consider diploid genome  $\alpha$  with genome length  $l$ .
  - Population is modeled by vector  $q$
  - $q_\alpha$  is prevalence of diploid  $\alpha$
  - $t_\alpha(g)$  is transmission function which is probability of gamete  $g$  being produced from parent  $\alpha$
  - $q'$  is next generation
10. • Diploids can be determined in terms of haploid distributions.
- Haploids can be determined in terms of diploid distributions.

- And evolution equation can be expressed in terms of haploid distributions as

$$\mathbf{p}'_{\gamma_0} = \sum_{\alpha_0, \alpha_1} \mathbf{p}_{\alpha_0} \mathbf{p}_{\alpha_1} t_{\langle \alpha_0, \alpha_1 \rangle}(\gamma_0)$$

- In the matrix form, evolution equation can be expressed as  $\mathbf{p}'_g = \mathbf{p}^T M_g \mathbf{p}$
  - Where  $M(g)$  is mixing matrix
- 11.
- This slide shows computation of crossover and mutation distributions, and transmission function
  - Transmission function is computed with this expression
  - There are 3 sum terms in it, which is expensive to compute
- 12.
- $W$  is walsh matrix
  - $\hat{A}$  is walsh transform of matrix  $A$ , and  $\hat{w}$  is walsh transform of column vector  $w$ .
  - Mixing matrix is given by the expression

$$\widehat{M}_{u,v} = 2^{\ell-1} [uv = \mathbf{0}] \hat{\mu}_u \hat{\mu}_v \sum_{k \in \overline{u+v} \mathcal{R}} \chi_{k+u} + \chi_{k+v}$$

- This is evolution eqn in Walsh basis

$$\hat{\mathbf{p}}'_g = 2^{\ell/2} \sum_{i \in g \mathcal{R}} \hat{\mathbf{p}}_i \hat{\mathbf{p}}_{i+g} \widehat{M}_{i,i+g}$$

- 13.
- Now if we compare evolution eqn in Walsh basis to what we had before, we got rid of matrix multiplication to compute next generation.
  - We don't need to refer to  $2^\ell$  mixing matrices either, only one mixing matrix is required in walsh basis computation.
  - Calculating each mixing matrix in previous eqn required 3 sum terms, which is reduced to 1 sum term in walsh basis.
  - Consider  $\ell = 14$ ,  $2^{14}$  mixing matrices would require 32 TB of memory, while one mixing matrix would require only 2 GB of memory

- We think that is great optimization in resource
14.
    - With naive computation as shown in equation (1), there would be  $2^l \cdot 2^l$  multiplication terms to be computed
    - where  $f$  is finite population and  $q$  is infinite population
    - We simplify it as shown in eqn (2) where it needs only  $2^l + \text{cardinal value of } S_f$  terms
    - $S_f$  is set of diploids that are in finite populations
  15. Writing code to compute in the Walsh basis, and running simulations using the Walsh basis is a significant part of my thesis. It represents the "Efficient" in the thesis title "Efficient Simulation of a Simple Evolutionary System".
  - 16.
  17.
    - Our first question concerns about convergence of finite population to infinite population.
    - Chebyshev's inequality, Jensen's inequality and geometric point of view, all, suggest the distance might decrease as  $1/\sqrt{N}$
    - But all of 3 implications are from inequalities. The distance might decrease much smaller than  $1/\sqrt{N}$ .
    - We investigate if the distance infact decreases like  $1/\sqrt{N}$  in practice.
  18.
    - We ran some simulations with  $\chi = 0.1$  and  $\mu = 0.001$  for different values of  $l$
    - These graphs show the results in logarithmic scale.
    - (point to axis)  $d$  is the distance, (point to axis)  $N$  is population size, (point to axis)  $n$  is number of generations
    - Data shows as population size increases, distance decreases and converge to infinite population
    - Graphs show flat surface so we model the data by linear equation

19.
  - Regression shows slope  $m$  is nearly equal to  $-0.5$
  - This proves distance does decrease as  $1/\sqrt{N}$
20.
  - $1/\sqrt{N}$  was theoretical upper bound but our simulation shows it is really very good bound and the distance decreases as  $1/\sqrt{N}$ .
21.
  - Our 2nd question concerns about oscillation in finite populations.
22.
  - The infinite population sequence  $\mathbf{p}, \mathcal{G}(\mathbf{p}), \mathcal{G}^2(\mathbf{p}), \dots$  may converge to a point, and we call that point a fixed point, we also call it limit.
  - But under certain conditions, the sequence converges to a periodic orbit between two fixed points, say  $\mathbf{p}^*$  and  $\mathbf{q}^*$
23.
  - And the conditions are: For some  $g$

$$\begin{aligned}
 -1 &= \sum_j (-1)^{g^T j} \mu_j \\
 1 &= \sum_{k \in \bar{g}\mathcal{R}} \chi_{k+g} + \chi_k
 \end{aligned}$$

- 1st one is for mutation distribution and 2nd one is for crossover distribution
  - We try to answer our 2nd question through simulations; that is do finite populations exhibit oscillations from random initial populations when infinite populations oscillate?
24.
    - Akin, Hastings, Wright, Bidwell and Agapie have studied oscillations in the past.
  25.
    - Akin considers - continuous time model, we consider - discrete time model
    - Hastings' study is limited to two bits length, includes only crossover, but no mutation
    - Wright and Bidwell compute a specific fitness function and a specific initial population for randomly generated mutation and crossover distributions in an attempt to find cyclic behavior

- Wright and Agapie use dynamic mutation that depends upon where population is in the population space while we use static mutation
- 26.
- Simulations were run for both haploid and diploid populations
  - To visualize oscillations, distance to fixed points ( $p^*$ ,  $q^*$ ) are plotted
- 27.
- These are the results for haploid and diploid population of length 8
  - top row shows results for haploid, bottom row for diploids
  - Population size is in increasing order in columns
  - As population size increases, oscillation approaches the behavior exhibited by infinite population
  - We observe more randomness in diploids than in haploids for same string length and population size.
- 28.
- Graphs show the results for haploid and diploid population of length 12
  - We observe that increase in  $l$  degrades oscillation (particularly in diploids)
- 29.
- So our answer to the 2nd question is
  - Finite populations exhibit approximate oscillation when infinite populations oscillate
- 30.
- Question 3 concerns the robustness of finite population oscillation under mutation-violation
- 31.
- We introduce violation  $\epsilon$  as following
  - This means no periodic orbits for finite population
  - The modification in  $\mu$  makes the Markov chain regular
  - This means no periodic orbits for infinite population
  - We study if finite population can exhibit approximate oscillations in that case

32.
  - Simulations were run for different values of  $\epsilon$ .
  - Distance of population to limits p and q without violation are plotted
33.
  - These figures show results for haploid population behavior of string length 8
  - Graphs are arranged in rows for increasing  $\epsilon$  and columns for increasing population size
  - Graphs show convergence of finite population behavior to infinite population behavior as population increases
  - Results show oscillating behavior of population for smaller values of epsilon that diminishes with time.
  - As value of  $\epsilon$  grows, oscillation diminishes
  - Oscillation dies out for infinite population.
  - Even though finite population also appears to be dying out, since Markov chain is regular, finite population will oscillate infinitely often (for  $\epsilon = 0.1$ ).
34.
  - Graphs show results for diploid population behavior of string length 12
  - We observed oscillation also degrades as string length increases. It is seen in both haploid and diploid cases, but particularly noticeable in diploids.
35.
  - Our conclusions from this simulation are
  - Finite populations exhibit approximate oscillation even if Markov chain is regular when violation is small
  - If violation becomes larger, finite population oscillation decreases
  - As string length increases, oscillation degrades
36.
  - Our previous question concerns with the robustness of finite population oscillation under mutation-violation, our next question also

concerns with the robustness of finite population but under crossover-violation

37.
  - We introduce violation  $\epsilon$  in crossover distribution
  - Crossover-violation means no periodic orbit exists for infinite population
  - We investigate if finite population can exhibit approximate oscillations in this case
38.
  - Simulations were run for different values of  $\epsilon$
  - Distances of population to limits  $p$  and  $q$  without violation are plotted
39.
  - Graphs show results for haploid population behavior of string length 8
  - Results show similar behavior to that of violation in mutation case
40.
  - Graphs show results for diploid population behavior of string length 12
  - However, we noticed that rate of damping of amplitudes of oscillation is slower than in mutation violation
  - And, we see more randomness in population behavior than in mutation violation, especially for diploid case
41.
  - Conclusions from experiment on violation in crossover are
  - Finite populations exhibit approximate oscillation if violation is small
  - If violation becomes larger, finite population oscillation decreases
42.
  - Overall conclusions from this research are
  - Vose's haploid model makes computation efficient in diploid case by reducing to haploid case
  - Distance between finite population and infinite population can decrease like  $1/\sqrt{N}$



- When infinite populations oscillate, finite populations exhibit approximate oscillation
- Finite populations exhibit approximate oscillation for small mutation-violation
- Finite populations exhibit approximate oscillation for small crossover-violation