

1.
 - I will talk about my thesis:
“Simulation of Simple Evolutionary System”
2.
 - I will give background, then address these four questions, and make concluding remarks.
3.
 - Population is a collection of length ℓ 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 ℓ binary strings and operations under \mathcal{R} are bitwise modulo 2 operations.
4.
 - 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 $u' = um + v\bar{m}, v' = u\bar{m} + vm$
 - χ_m is probability of using crossover mask m
 -
 - Mutation flips bits in x using mutation mask m .
 - μ_m is probability of using mutation mask m
5.
 - 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 $\tau(p)$

6.
 - In Random Heuristic Search,
 - Given population p , τ is transition rule that maps p to the next generation p' where p and p' both belong to population space Λ_N . N is population size.
 - Finite population evolution forms Markov chain
 - However, τ is stochastic function of crossover and mutation and $\tau(p)$ can not be predicted with certainty
7.
 - In infinite population model, population is modeled by vector p where p belongs to population space Λ .
 - \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.
8.
 - 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. For this we set up simple diploid model.
9.
 - We consider diploid genome α with genome length l .
 - Population is modeled by vector q
 - q_α is prevalence of diploid α
 - $t_\alpha(g)$ is transmission function which is probability of gamete g being produced from parent α
 - 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

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

- This is evolution eqn in Walsh basis

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

- 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^ℓ 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.

13.
 - Reduction to haploid model and computing in Walsh basis simplified computations and made it efficient
 - 3 sum terms in transmission function is reduced to only one sum terms for computing mixing matrix in walsh basis
 - We only need one mixing matrix as opposed to 2^l
 - Consider $l = 14$, 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.
 - 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
17.
 - Regression shows slope m is nearly equal to -0.5

- This proves distance does decrease as $1/\sqrt{N}$
- 18. • $1/\sqrt{N}$ was theoretical upper bound but our simulation shows it is really very good bound and the distance decreases as $1/\sqrt{N}$.
- 19. • Our 2nd question concerns about oscillation in finite populations.
- 20. • First I want to introduce term limit; we also call it fixed point.
 - 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.
 - But under certain conditions, the sequence converges to a periodic orbit between two fixed points, say \mathbf{p}^* and \mathbf{q}^*
- 21. • 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
- So our 2nd question is do finite populations also exhibit oscillations from random initial populations when infinite populations oscillate?
- 22. • Akin, Hastings, Wright, Bidwell and Agapie have studied oscillations in the past.
- 23. • Akin considers - continuous time model, we consider - discrete time model
 - Hastings' study is limited to two bits length, includes only crossover, 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
- 24.
- Simulations were run for both haploid and diploid populations
 - for different string lengths l and population sizes N
 - To visualize oscillations, distance to fixed points (p^*, q^*) are plotted
- 25.
- Graphs show the results for haploid and diploid population of length 8
 - top row shows results for haploid, bottom row for diploids
 - Population size 4096, 40960, 81920 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.
- 26.
- Graphs show the results for haploid and diploid population of length 12
 - We observe that increase in l degrades oscillation (particularly in diploids)
- 27.
- Our conclusion from this simulation is
 - Finite populations exhibit approximate oscillation when infinite populations oscillate
- 28.
- Question 3 concerns the robustness of finite population oscillation under mutation-violation
- 29.
- We introduce violation ϵ as following
 - The modification in μ makes the Markov chain regular
 - This means no periodic orbits for finite population and no periodic orbits for infinite population

- We study if finite population can exhibit approximate oscillations in that case
- 30.
- Simulations were run for different values of ϵ , ℓ and N
 - Distance of population to limits p and q without violation are plotted
- 31.
- Graphs show results for haploid population behavior of string length 8
 - Graphs are arranged in rows for increasing ϵ 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.
 - 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 $\epsilon = 0.1$).
 - As value of ϵ grows, oscillation diminishes
- 32.
- 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.
- 33.
- 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

34.
 - Our last 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
35.
 - We introduce violation ϵ 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
36.
 - Simulations were run for different values of ϵ, ℓ and N
 - Distances of population to limits p and q without violation are plotted
37.
 - Graphs show results for haploid population behavior of string length 8
 - Results show similar behavior to that of violation in mutation case
38.
 - 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
39.
 - 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
40.
 - 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 population exhibits approximate oscillation for small mutation violation
- Finite populations exhibit approximate oscillation for small crossover violation