1. - I will talk about my thesis:

    "Simulation of Simple Evolutionary System"

2. - I will give background, then address four questions, and make concluding remarks.

3. - Population is a collection of length l binary strings

    - A population can be represented by a vector; the jth 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 $u' = um + v\bar{m}, v' = u\bar{m} + vm$

    - Mutation flips bits in x using mutation mask m according to rule x to $x + m$.

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'.

1

- The sequence shows finite population evolution which forms Markov chain

- However, τ is stochastic function of crossover and mutation, and crossover masks and mutation masks are randomly chosen. So, $\tau(p)$ can not be predicted with certainty

- $\boldsymbol{\chi}_i$ is probability of choosing crossover mask i

- $\boldsymbol{\mu}_i$ is probability of choosing mutation mask i

7.
- In infinite population model, population is modeled as by a vector $\boldsymbol{p}$

- $\mathcal{G}$ is a function that maps $\boldsymbol{p}$ to the next generation where $\mathcal{G}(p)_j$ is probability that string j occurs in the next generation.

- The sequence shows evolution of p under infinite population model.

- The variance between next generation finite population and infinite population is given by expression where N is population size

8.
- Our first question concerns about convergence of finite population to infinite population

- Chebyshev's inequality, Jensen's inequality and geometric point of view, all, suggests the distance might decrease as $1/\sqrt{r}$

- But all of 3 implications are from inequalities. So the distance might decrease much smaller than $1/\sqrt{r}$.

- We investigate if the distance infact decreases like $1/\sqrt{r}$ in practice. For this we set up simple diploid model.

9.
- It is infinite population model.

- We consider diploid genome α with genome length l.

- q is current generation population

- 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.
- Evolution equation can be expressed in terms of haploid distributions as

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

- In the matrix form, evolution equations can be expressed as
- 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 in it, which is expensive

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

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

- And evolution eqn in Walsh basis is

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

- 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 + cardinal value of S_f terms

- S_f is set of diploids that are in finite populations

15.
- We ran 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

16.
- Regression shows slope m is nearly equal to -0.5

- This proves distance does decrease as $1/\sqrt{N}$

17.
- $1/\sqrt{N}$ was theoretical upper bound but our simulaion shows it is really very good bound and the distance decreases as $1/\sqrt{N}$.

18. • Our 2nd question concerns about oscillation in finite populations.

19. • First I want to introduce term limit; we also call it fixed point.

    • The infinite population sequence $\boldsymbol{p}$, $\mathcal{G}(\boldsymbol{p})$, $\mathcal{G}^2(\boldsymbol{p})$, $\cdots$ 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 p* and q*

20. • And the conditions are: For some g

$$
\begin{aligned}
-1 &= \sum_j (-1)^{g^T j} \boldsymbol{\mu}_j \\
1 &= \sum_{k \in \bar{g}\mathcal{R}} \boldsymbol{\chi}_{k+g} + \boldsymbol{\chi}_k
\end{aligned}
$$

    • So our 2nd question is do finite populations also exhibit oscillations from random intial populations when infinite population oscillate?

21. • Akin, Hasting, Wright, Bidwell and Agapie have studied oscillations in the past.

22. • Akin considered - continuous time model, we consider - discrete time model

    • Hastings' study - limited to two bits length, includes only crossover, not 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 used dynamic mutation that depends upon where population is in the population space while we use static mutation

    • We study oscillation for

    • fixed fitness function and random: initial population, mutation and crossover distribution

- higher bit length (up to 14), and actually oscillation conditions for infinite populations are independent of string length

- both haploid and diploid populations, and for both finite and infinite populations

- We also visualize oscillation

23.
- Simulations were run for both haploid and diploid populations

- for different string lengths l

- and for population size 4096, 40960, 81920

- To visualize oscillations, distances between fixed points (p* , q*) and population are plotted

24.
- Graphs show the results for haploid and diploid population of length 8

- top haploid, bottom - 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.

25.
- Graphs show the results for haploid and diploid population of length 12

- We observe that increase in l degrades oscillation (particularly in diploids)

26.
- Our conclusions from this simulation are

- Finite populations exhibit approximate oscillation when infinite populations oscillate

- As population size increases, randomness decreases

27. • Question 3 concerns the robustness of finite population oscillation under violation in mutation

28. • We introduce violation ϵ as following
    • The modification in $\boldsymbol{\mu}$ 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

29. • Simulations were run for different values of ϵ, ℓ and N
    • Distances of population to limits p and q without violation are plotted

30. • 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.
    • As value of epsilon grows, oscillation diminishes

31. • 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.

32.
- 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

33.
- 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 violation in crossover

34.
- Violation in crossover means no periodic orbit exists for infinite population
- But we don't know if Markov chain is regular in this case
- We investigate if finite population can exhibit approximate oscillations in this case

35.
- We introduce violation ϵ as following

36.
- Simulations were run for different values of $\epsilon, \ell and N$
- 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