

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.
 - The terms here have technical meaning which will be explained in the following slides.
3.
 - Population is a collection of length l binary strings
 - A population is represented by a vector; the j th component is the proportion of string j in the population.
 - Bitwise modulo 2 operations are defined on the set \mathcal{R} of length ℓ binary strings.
4.
 - Crossover and mutation operators are defined using the Bitwise operations in \mathcal{R} .
 - Given parents u and v , crossover produces children u' and v' .
 - x mutates to $x + m$.
5.
 - This flowchart illustrates finite population genetic algorithm
 - We start from random initial population P .
 - Randomly select parents u and v
 - Crossover u and v to produce u' and v'
 - Keep one of u' , v' and mutate
 - Repeat to form next generation and when next generation completes, replace it with new generation
 - The process is repeated until some threshold is met
6.
 - In infinite population model, population is modeled as by a vector \mathbf{p}
 - \mathcal{G} is a function that maps \mathbf{p} to the next generation where $\mathcal{G}(\mathbf{p})_j$ is probability that string j occurs in the next generation.

- The sequence shows evolution of p under infinite population model.
- 7.
- Random Heuristic Search was presented in his book by Vose in 1999. Our work is largely based on this model
 - Given population p , τ is transition rule that maps p to the next generation p' .
 - The sequence shows finite population evolution which forms Markov chain
 - However, τ is stochastic function of crossover and mutation, and hence, τp can not be predicted with certainty.
 - $\mathcal{G}p$ gives the expected next generation
 - The variance between finite population and infinite population is given by expression where r is population size
- 8.
- Our first question concerns about distance between finite population and 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 what happens 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^n is population at generation n
 - q_α^n is prevalence of diploid α at generation n
 - $t_\alpha(g)$ is transmission function which is probability of gamete g being produced from parent α

- $q_\gamma^{n+1} = \sum_\alpha q_\alpha^n t_\alpha(\gamma_0) \sum_\beta q_\beta^n t_\beta(\gamma_1)$
gives us next generation population.
- 10.
- Diploids are determined in terms of haploid distributions.
 - Haploids are determined in terms of diploid distributions.
 - Evolution equations can be expressed in terms of haploid distributions as
 - 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
 - Instead of using this transmission function, we want to compute mixing matrix in walsh basis.
- 12.
- W is walsh matrix
 - (hat) represents walsh transform and for matrix A , \hat{A} is given by the expression $\hat{A} = WAW$, and for vector w , \hat{w} is by the expression $\hat{w} = wW$.given
 - Mixing matrix is given by the expression

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

- And evolution eqn in Walsh basis is

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

- 13.
- Reduction to haploid model and computing in Walsh basis simplified computations and made it efficient
 - 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 $2l$ x $2l$ 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 $2l + \text{cardinal value of } S_f$ terms
 - S_f is set of diploids that are in finite populations
- 15.
- With naive computation as shown in equation (1), there would be $2^l x 2^l$ 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 $2l + \text{cardinal value of } S_f$ terms
 - S_f is set of diploids that are in finite populations
- 16.
- We ran simulations with $\chi = 0.1$ and $\mu = 0.001$ for different values of l
 - These graphs show the results.
 - d is the distance, N is population size, n is number of generations
 - Graphs show as population size increases, distance decreases, converge to infinite population
 - Graphs also show smoothing as genome length increases
 - Data shows a near linear dependence of $\log d$ on $\log N$ and that is what we are particularly interested in
- 17.
- We ran simulations with $\chi = 0.1$ and $\mu = 0.001$ for different values of l

- These graphs are the results.
 - d is the distance, N is population size, n is number of generations
 - Graphs show as population size increases, distance decreases, converge to infinite population
 - Graphs also show smoothing as genome length increases
 - Data shows a near linear dependence of $\log d$ on $\log N$ and that is what we are particularly interested in
- 18.
- So we tried to fit data in the linear equation
 - m is slope and b is y intercept
 - Regression shows m nearly equal to -0.5
 - This proves distance does decrease as $1/\sqrt{N}$
 - Also graph for y intercept b shows, $k = e^b$ decreases monotonically with genome length l , and increases monotonically with generation n
 - Increase in k for larger n seems to be a manifestation of the growing nonlinearity likely that the nonlinearity results partly from genetic drift experienced by finite populations
 - and partly because predicted distance is for single step case as generation increases difference between finite population and infinite population accumulates non linearly
- 19.
- Conclusion from this simulation is that
 - Vose's infinite population model makes computation in diploid case efficient by reducing to the haploid case
 - Distance between finite diploid population and infinite diploid population can decrease like $1/N$
- 20.
- We move on to our 2nd question which concerns oscillation in finite populations.

21.
 - First I want to introduce term limit; we also call it fixed point interchangeably.
 - The infinite population sequence (point to sequence) may converge to a point, that we call a fixed point.
 - But under certain conditions, the sequence converges to a periodic orbit between two fixed points, say p^* and q^*
22.
 - And the conditions are (point to the expressions)
 - So our 2nd question is do finite populations also exhibit oscillations from random initial populations?
23.
 - Other people in the past also have been interested in studying oscillations in populations.
 - Akin (1982) proved existence of cycling for continuous-time 2-bit diploid model
 - Hasting (1981) studied cycling in populations with infinite 2-bit diploid population model
 - Wright and Bidwell (1997) provided examples when cycles in an infinite haploid population model occur with crossover and mutation for 3 bit and 4 bit populations
 - Wright and Agapie (2001) described cyclings in infinite populations for up to 4 bits, and also presented data for cyclings in finite populations
24.
 - 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
- 25.
- 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
- 26.
- Graphs show the results for haploid and diploid population of length 10
 - left side haploid, right side diploids
 - and for population size 4096, 40960, 81920
 - last row graphs show results for infinite population
 - As population size increases, oscillation approaches the behavior exhibited by infinite population
 - We observe more randomness in diploids than in haploids for same l and N
- 27.
- Graphs show the results for haploid and diploid population of length 12

- We observe that Increase in l degrades oscillation, and it is particularly noticeable in diploid case
 - We see an interesting behavior in this top right graph,
- 28.
- Here is zoomed in picture
 - We see oscillation between different levels
 - we observe this kind of behavior only for higher values of l in diploids and for small population size
 - It might be because there could be fixed points for other distributions in the vicinity of finite population trajectory and get attracted to those fixed points
 - And smaller populations do not follow infinite population path as closely as larger populations do; so as for larger populations, these behaviors were not observed
- 29.
- We also plotted average amplitudes of oscillation in haploids and diploids
 - Graphs show
 - Oscillation amplitude increases with increase in population size
 - Amplitude of oscillation decreases with increase in l
- 30.
- Our conclusions from this simulation are
 - Finite populations exhibit approximate oscillation when infinite populations oscillate
 - As l increases, oscillation amplitude decreases
 - As population size increases, oscillation amplitude increases and randomness decreases
 - Finite population also oscillates between different pairs of fixed points for diploid population of smaller size and larger l