

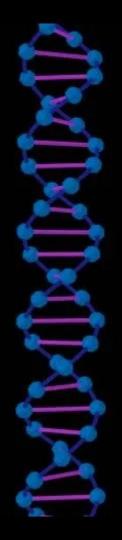
STOCHASTIC PROCESSES AND APPLICATIONS



https://docs.google.com/presentation/d/1fm7JiUtfZ95spsfAUV4w9dh Ft8Rtxy2kVEwoLlbjpOg/edit?usp=sharing

Reshmi Chatterjee 2019441 Ishani Laha 2019247 Mudit Balooja 2019258

Chintan Puggalok 2019239 Yashasvi Chhabra 2019290



STOCHASTIC PROCESSES AND APPLICATIONS

MODELLING GENETIC DRIFT

Reshmi Chatterjee 2019441 Ishani Laha 2019247 Mudit Balooja 2019258

Chintan Puggalok 2019239 Yashasvi Chhabra 2019290

Genetic Drift

- RANDOM FLUCTUATIONS IN THE NUMBERS OF GENE VARIANTS IN A POPULATION
- FLUCTUATIONS IN THE PRESENCE OF ALLELES ARE MEASURED AS CHANGES IN ALLELE FREQUENCIES

Hardy-Weinberg Equilibrium

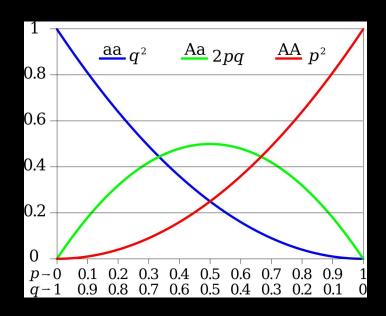
- Null model
- Assuming evolution never occured
- Allele frequencies would be at Hardy Weinberg equilibrium heterozygous

$$p^2 + 2pq + q^2 = 1$$

homozygous dominant

homozygous recessive

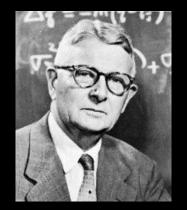
- Idealized situation
- External factors can change the equilibrium
- Why useful?



Credits:

https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle#/media/File:Hardy-Weinberg.svg

Wright-Fisher Model





Assumptions



- The generations of a population do not overlap.
- The population must be diploid.
- They have a constant size of N, which does not change between generations.



Assumptions

- If m alleles are present in one generation, then 2N chromosomes in the next generation select the allele with probability p = m/2N
- Parents of the next generation of N genes are picked randomly with replacement from the preceding generation.

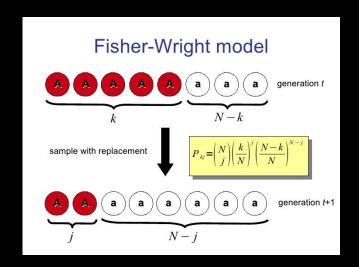


Definition

- Let 'A' and 'a' be two alleles.
- Let X_t be the count of A allele at generation t.
- The state-space => $Xt \in \{0,1,...,2N\}$.
- At generation t+1, alleles are sampled with replacement from generation t to a new population.

Definition

- This happens by binomial sampling of alleles at every generation.
- The probability transition matrix of the Markov Chain is:
 P_{ij}=(²NC_i) (i/2N)^j (1-i/2N)²N-j
- $X_t \mid X_{t-1} = X_{t-1} \sim \text{Binomial (n=2N, p= } X_{t-1}/2N)$



Credits

https://pos-darwinista.blogspot.com/2016/09/inferencia-estatistica-no-mo

Mean

- The expected value of the Wright-Fisher process at any time is just the count of the A allele in the first generation.
- Hence, if Yt is the frequency of the A allele at time t then: $E(Y_t) = E(Y_0)$

Variance

The formula for variance in relation to Expectation is as follows:

$$egin{aligned} Var(X_t) &= E(X_t^2) - E(X_t)^2 \ \ \Rightarrow Var(X_t) &= Var(X_{t-1})(1-rac{1}{2N}) + X_0(1-rac{X_0}{2N}) \end{aligned}$$

The above recurrence results in:

$$Var(X_t) = 2NX_0(1-rac{X_0}{2N})(1-(1-rac{1}{2N})^t)$$

In terms of the frequency of any one of the 2 alleles:

$$Var(Y_t) = Y_0(1-Y_0)(1-(1-rac{1}{2N})^t)$$

Fixation probabilities

d%20Tutorials/pol2e at 1502 genetic drift simulation/pol2e at 1502 genetic drift simulation.html

if X_t^* is an absorbing state then:

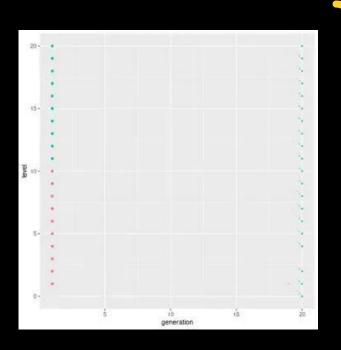
$$P_{\{i=X_i^*\}_j}=0$$

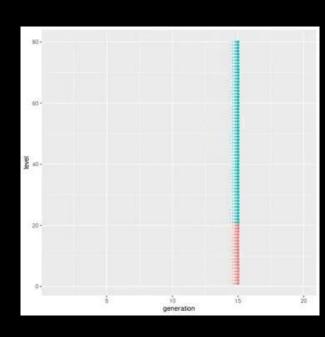
The absorbing states in the Wright-Fisher model are $X_t^* \in \overline{\{0,2N\}}$

$$\Rightarrow P(X_{t+1} = 2N \mid X_t = i) = \frac{i}{2N}$$

The probability of fixation or loss of an allele in a pure drift Wright-Fisher model only depends on the previous count of the allele and the population size

Simulation of Wright-Fisher Model





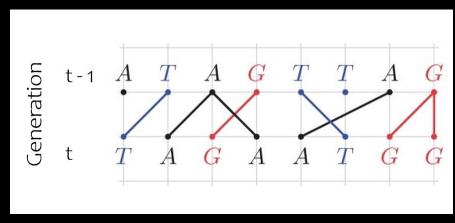
K-ALLELE POPULATION & MULTINOMIAL SAMPLING

Instead of only two alleles, we now consider a population with alleles $\{a_i, ..., a_k\}$. A Markov Chain X_n gives the Wright-Fisher model for this k-allele population with state-space $\{0, 1, 2, ..., 2N\}$ for each individual allele. This case employs the use of multinomial sampling in place of binomial sampling.

K-allele Wright Fisher Model

The random mating of the finite population is alone responsible for stochastic variations in allele frequencies.

Let X_t denote the allele frequency at generation t. Then, we have:



Credits: https://academic.oup.com/sysbio/article/66/1/e30/2670014#112688514

$$X_t|X_{t-1} \sim Multinomial(2N, p_1, ..., p_K)$$

K-allele Wright Fisher Model

• Probability Distribution

$$P(X_t = (\beta_1, ..., \beta_K) | X_{t-1} = (\alpha_1, ..., \alpha_K)) = \frac{(2N)!}{\beta_1! ... \beta_K!} \left(\frac{\alpha_1}{2N}\right)^{\beta_1} ... \left(\frac{\alpha_K}{2N}\right)^{\beta_K}$$

Expectation

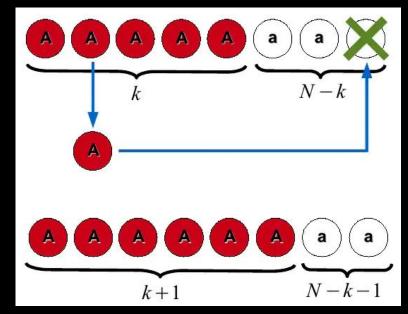
$$E(X_t) = (2Np_1, 2Np_2, ..., 2Np_K)$$

Variance

$$Var(X_{i,t}) = 2Np_i(1 - p_i)$$

"Moran model is a simple stochastic process used in biology to describe finite populations."

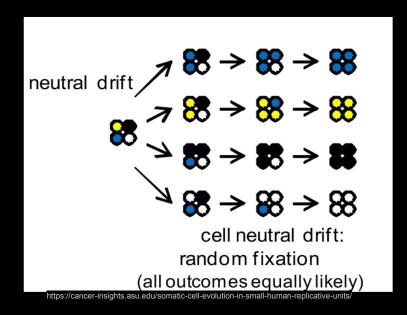
- Proposed by Patrick Moran in 1958
- Finite, constant-size population
- Generations overlap
- One individual reproduces while one individual dies at each step
- Results qualitatively similar to the Wright-Fisher model



Credits: https://www.slideshare.net/fer.reche/mathematical-models-of-evolution

Moran process describes the phenomenon of neutral drift.

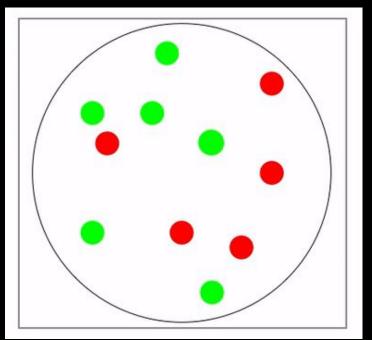
Neutral drift is the inclusion of mutations which have little or no effect on the organism, ie it gives no advantage in its current environment



Let the 2 competing alleles be A and B State space i=0, 1, 2,, N i counts the number of A allele N=Total population

$$egin{aligned} P_{i,i-1} &= rac{N-i}{N}rac{i}{N} \ P_{i,i} &= 1-P_{i,i-1}-P_{i,i+1} \ P_{i,i+1} &= rac{i}{N}rac{N-i}{N} \end{aligned}$$

Absorbing states are 0 and N

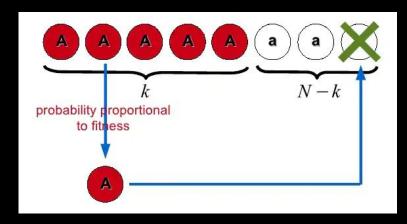


https://constabl3.wixsite.com/gwaconstable/models-of-population-genetics

One allele may have a "fitness advantage" over the other. This is a consequence of the Selection process.

Assuming individuals with alleles A and B have fitness f_i and g_i respectively:

$$egin{aligned} P_{i,i-1} &= rac{g_i(N-i)}{f_i \cdot i + g_i(N-i)} \cdot rac{i}{N} \ P_{i,i} &= 1 - P_{i,i-1} - P_{i,i+1} \ P_{i,i+1} &= rac{f_i \cdot i}{f_i \cdot i + g_i(N-i)} \cdot rac{N-i}{N} \end{aligned}$$



Credits: https://www.slideshare.net/fer.reche/mathematical-models-of-evolution

Ethier and Nagylaki Model

- Proposed by S. N. Ethier and T. Nagylaki in 1980
- A multinomial sampling model for selection, mutation, and random genetic drift at a single multiallelic locus.
- Panmictic, diploid, monoecious population
- Discrete, non-overlapping generations
- It posits that for frequent alleles, the standard diffusion approximation holds, whilst for rare alleles, the approximation holds only for moderate number of alleles.

REFERENCES

- https://www.youtube.com/watch?v=vVgdjiLoV5U
- https://stephens999.github.io/fiveMinuteStats/wright-fisher-model.html
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3118342/
- https://personal.math.ubc.ca/~db5d/SummerSchool09/lectures-dd/lectures6-7.
 pdf
- https://en.wikipedia.org/wiki/Moran process
- https://academic.oup.com/sysbio/article/66/1/e30/2670014

THANK YOU!