Random Mating with Variable Population Size

Eric M Downes[*]
MIT Department of Physics
(Dated: May 18, 2005)

1. INTRODUCTION

A recursive equation in discrete time for the allele-frequency within a randomly mating population of fixed size with non-overlapping generations was developed by Fisher[1] and Wright[2] to account for *genetic drift*, the finite-size effects of random selection on allele frequency.

A large-population diffusion approximation to the Fisher-Wright theory was studied extensively by Kimura[3]. Kingman incorporated observations about ancestral relations to develop an alternate *Coalescent* approach[4]. Griffiths and Tavar*é* extended this Coalescent model to the case of a deterministicly varying population[5].

The sizes of real biological populations vary in time in a fashion that is not obviously deterministic. Thus it is desirable to extend the Fisher-Wright model to cover the case of a stochasticly varying population size in order to study the effects of variability in population on genetic drift.

We begin by generalizing the Fisher-Wright model of genetic drift to include a time varying population size. We incorporate this into a model which considers the general case where a population will evolve according to a conditional probability distribution of the population at the previous generation. By noting the impulse response of this model, we propose an expansion of the update equation as a sum of delta functions, deriving a general expression, from which properties analogous to the work of Kimura could be calculated, in principle. We examine the first two cumulants of a simple deterministic case, in order to demonstrate compatibility with previous work, and close by proposing future work.

2. BACKGROUND

As observed by Fisher and Wright, if a fixed population of diploids is mating randomly, then the concentration $j/2N = [A_1]/([A_1]+[A_2])$ of an allele A_1 can be considered without calculating the frequency of the genotypes in the population (e.g. independent of $[A_1 A_2]$.) This is done by selecting (and then replacing) $2N$ individuals from the

current generation (with allele-frequency $i/2N$) to determine the genetic makeup of the next generation, and thus follows the binomial distribution:

$$\pi(j|i,N) = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \quad (1)$$

It can be easily shown, from the properties of the binomial distribution that the first two cumulants of $\pi(j|i,N)$ are

$$< j(i) >_c = i; \; < j(i)^2 >_c = i\left(1 - \frac{i}{2N}\right). \quad (2)$$

Where we use $< j(i) >$ instead of $< j|i >$ for the conditional mean to avoid confusion with energy eigenstates in quantum mechanics.

From here, one can construct a diffusion approximation[3] (requiring that the population be large) and address questions about the probability of fixation, the steady-state population, etc. This is ultimately the goal of our work, but first we must build an expression for $p(i|j)$.

3. VARIABLE POPULATION

3.1. Genetic Drift

It is straightforward to generalize (1) by selecting N' individuals to form the next generation (instead of N):

$$\pi(j|i,N,N') = \binom{2N'}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N'-j}. \quad (3)$$

Note that $\pi(j|i,N,N')$ says nothing about the update rule from N to N' (the time evolution), only the mating rule. If one wished to relax the random-mating assumption, one could substitute another expression for $\pi(j|i,N,N')$ into our work without difficulty. We now turn to the update rule.

3.2. Probabilistic Update

Allow the population to evolve according to some time-independent conditional probability density $\rho(N|N')$. Make the expectation value of

----

[*]Electronic address: edown@mit.edu

the population

$$< N'(N) > \equiv f(N) = \sum_{N'=0}^{\infty} N' \sum_{N=1}^{\infty} \rho(N'|N). \quad (4)$$

For example, $f(N) = rN$ would correspond to exponential growth, $f(N) = rN(1 - N/K)$ to logistic growth, etc.

We many now write an expression for $p(j|i)$:

$$p(j|i) = \sum_{N=1}^{\infty} \sum_{N'=0}^{\infty} \pi(j|i, N, N') \rho(N'|N) \quad (5)$$

3.3. Decomposition of Update Rule

We observe that if $\rho(N'|N) = \delta(N' - f(N))$ (deterministic population growth), $p(j|i)$ becomes

$$p(j|i) = \sum_{N=1}^{\infty} \pi(j|i, N, f(N)) \quad (6)$$

suggesting that it may be helpful to view the function $p(j|i)$ as the response of a filter, the kernel of which is $\pi(j|i, N, N')$, taking the input $\rho(N'|N)$.

The filter is linear by inspection, and so we expand $\rho(N'|N)$ as a sum over delta functions, one for each (potential) population value, the height of delta function being the probability of having population N', given the previous population N.

Thus, we have:

$$
\begin{aligned}
p(j|i) &= \sum_{N=1}^{\infty} \sum_{N'=0}^{\infty} \pi(j|i, N, N') \\
&\quad \times \left(\sum_{k} A_k(N) \delta(N' - f(N) - k) \right) \quad (7)
\end{aligned}
$$

As an example, for non-deterministic exponential population growth $f(N) = rN$ with a Gaussian-like variance $< N'(N)^2 >_c = D\sqrt{N}$, the full expression reads:

$$
\begin{aligned}
p(j|i) &= \sum_{N=1}^{\infty} \frac{1}{\sqrt{2\pi D\sqrt{N}}} \sum_{k} e^{-\frac{-k^2}{2D\sqrt{N}}} \\
&\quad \times \binom{2rN + 2k}{j} \left(\frac{i}{2N} \right)^j \left(1 - \frac{j}{2N} \right)^{2rN + 2k - j} \quad (8)
\end{aligned}
$$

3.4. Deterministic Growth

In the case where $N' = rN$, (deterministic exponential growth), we see from (6) that $p(j|i, N')$

is

$$p(j|i, N') = \binom{2N'}{j} \left(\frac{ir}{2N'} \right)^j \left(1 - \frac{ir}{2N'} \right)^{2N' - j} \quad (9)$$

which is a function of N' only. Thus, from the properties of the binomial distribution, we deduce:

$$< j(i) >_c = ir; \quad < j(i)^2 >_c = ir \left(1 - \frac{ir}{2N'} \right) \quad (10)$$

which is just (2) with the allele population in the previous generation i scaled by the growth rate: growth and genetic drift are independent *in this case*. This provides anecdotal confirmation of our intuition the allele-frequencies of a randomly-mating population with deterministicly varying population size will not behave differently than a randomly-mating population with fixed population size. Further work would take mutation and selection into account.

4. CONCLUSIONS

A general expression (7) for the conditional gene-frequency distribution was derived. Unfortunately this equation is not in a readily-usable state yet. Future work will attempt to resolve the expression into a closed mathematical form (without sums). Once such an expression is available, following the work of Kimura[3] we plan to incorporate the effects of mutation and selection (both assumed independent of population size) into the model, and derive the probability of fixation, time to fixation, and steady-state population for a number of simple cases of population update rules.

There are many subtleties in (7) to catch the unwary. Aside from the sums involved being unwieldy, it is not so simple to calculate the the change in allele-frequencies $< \Delta(jN) >$ as both i and N vary stochasticly with time. (This is the usual method by which to remove the conditional dependence of j on i.[3]) However, now the problem of the genetic drift of a stochasticly-varying population size is a well-posed one. Hopefully it will prove to be tractable in addition to being well-posed.

5. REFERENCES

[1] Fisher, R.A., On the Dominance Ratio *Proc. Royal Soc. Edinborough* **42** 321-341 (1930)

[2] Wright S., Evolution in Mendelian Populations *Genetics* **16** 97-159 (1931)

[3] Kimura, M., Diffusion Models in Population Genetics *J Applied Prob.* **1** 177-232 (1964)

[4] Kingman J.F.C. On the genealology of large populations Applied Prob. **19A** 27-43 (1982)

[5] Griffiths R.C. and Tavar*é* S. Sampling theory for neutral alleles in a varying environment *Phil. Trans. Royal Soc. London B* **334** 403-410 (1994)