Modeling Long-Term Dynamics in Population Genetics

Diffusion Approximation and Maximum Entropy Principle

Stephen J. Liu— Department of Statistics, University of Auckland, Auckland, New Zealand

Introduction

Population genetics is a field of biology that studies the genetic composition of biological populations, and the changes in genetic composition that result from the force of various factors, including mutation.

In a wild population, this change in genetic composition over time usually form a interesting dynamic system which is random in nature. To study several quantities of evolutionary interest and the impact of various forces on the ultimate future of a population, it is often desirable to obtain the stationary or the limiting distribution for the system.

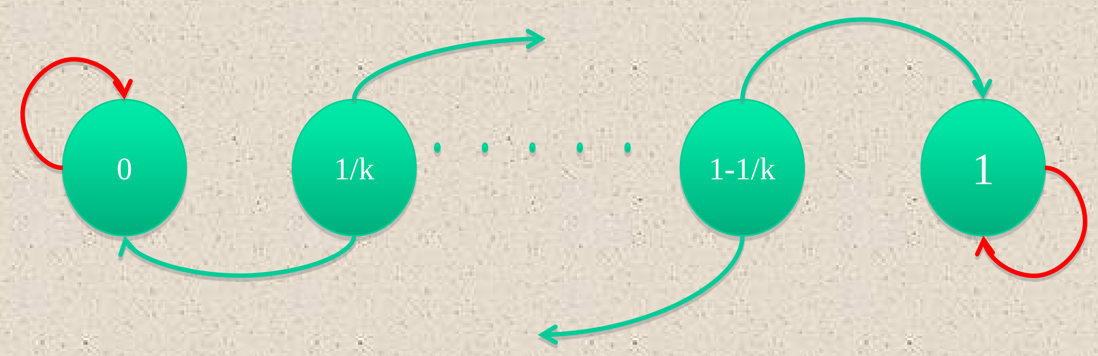

Fig. 1. In the absence of other forces, random mating ultimately causes alleles to be fixed in a finite population, the fluctuation of the allele frequencies can be considered as a Markov Chain with absorbing states at the two ends

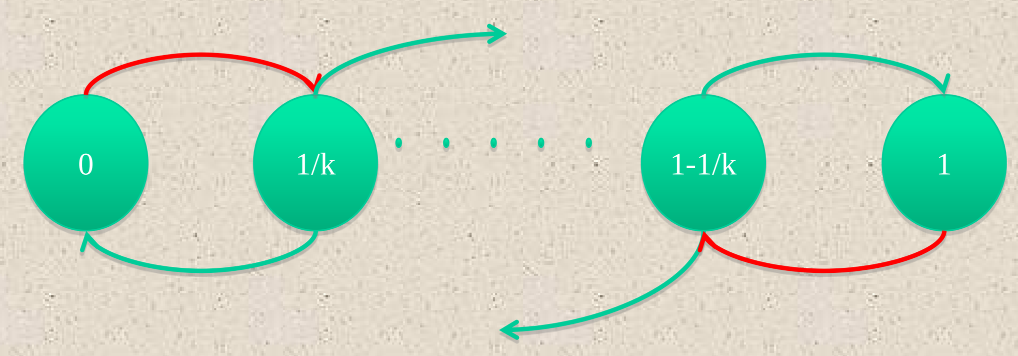

Fig. 2. With various other forces, such as two way mutation, the Markov Chain takes a form without the absorbing states, therefore an equilibrium point exists for such a system.

Diffusion Approximation

The fluctuation of the allele frequencies from generation to generation can be approximated by a continuous-time continuous-space diffusion process, this approximation is called the diffusion approximation in population genetics.

It is often written in the form of a second partial differential operator. For any well behaved function of the random variable in the system, the expectation of the operator acting on the function always vanishes.

$$E\{\ell[\Phi(X)]\} = 0$$

For two alleles Wright-Fisher Model with two way mutation, the corresponding diffusion operator has the following form :

$$\ell = \left(\beta_1(1-x) - \beta_2 x\right)\frac{\partial}{\partial x} + \frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2}$$

where β_1 and β_2 are the scaled mutation rate.

Maximum Entropy Principle

Given a series of moments of an unknown distribution, which distribution is the most appropriate choice to be used to describe the unknown distribution?

The Maximum Entropy Principle (MEP) provides a solution to the above question. It involves maximizing Shannon's entropy, which can be understood as a measure of "uncertainty", it is widely used in information theory. According to MEP, the distribution shall be chosen such that Shannon's entropy is maximized while satisfying the set of moment constraints.

$$A[f] = -\int f \log(f)\,dx$$

s.t. $\int f\,dx = 1$ and $\int \phi_i f\,dx = \mu_i$ For $\forall i$

Results

The stationary distribution or the limiting distribution can be derived by applying MEP with the set of moment constraints obtained from the diffusion approximation:

- The shape of the unknown distribution greatly affects the convergence rate.
- Non-standard or fractional moment constraints tend to give a better convergence rate.
- Only a few non-standard moment constraints are needed to achieve an excellent accuracy.

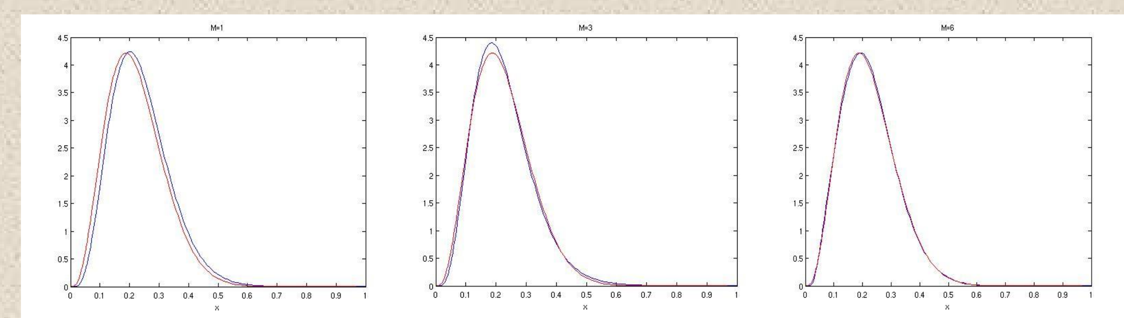

Fig. 3(a-c). The Wright-Fisher model with two way mutation is considered, the true density for the allele frequency is represented by the red curve, MEP estimates by the blue curve, the number of moment constraints used are 1 , 3 and 6 respectively.

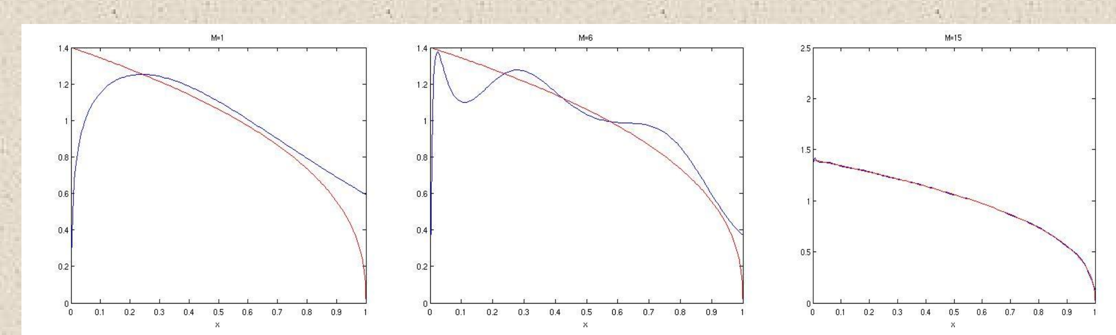

Fig. 4(a-c). The same Wright-Fisher model is considered but with much smaller scaled mutation rates, again the true density for the allele frequency is represented by the red curve, MEP estimates by the blue curve, the number of moment constraints used are 1 , 6 and 15 respectively.

It has been found that a symmetric and smooth distribution can be obtained with high accuracy using only 2 moment constraints. Even a distribution with singularities can be approximated with tolerable errors using a reasonable number of moment constraints.

A number of basis functions have been considered, shifted Chebyshev polynomials of the right support seem to be the most robust basis for a wide range of distributions.

Conclusions

The diffusion approximation and maximum entropy principle together provide a powerful procedure for obtaining a stationary or limiting distribution for stochastic systems.

It has been shown by others that using fractional moment constraints greatly increases the convergence rate for maximum entropy procedure. Often 10 times as many integer moment constraints are required to achieve the same amount of accuracy that fractional moment constraints can achieve.

The diffusion approximation provides a way of obtaining values not only for fractional moments but for some non-standard moments as well. Hence by establishing a suitable diffusion approximation, the underlying stationary or limiting distribution can be obtained with very few moment constraints.

Literature cited

Kimura, M. 1964. Diffusion models in population genetics. *Journal of Applied Probability* 1:177--232.

Ethier, SN and Nagylaki, T. Diffusion approximations of the two-locus Wright-Fisher model. *Journal of mathematical biology* 27:17--28.

Hartl, D.L. and Clark, A.G. and Clark, A.G. 1997. *Principles of population genetics*, Sinauer associates Sunderland, MA.

Bandyopadhyay, K. and Bhattacharya, AK and Biswas, P. and Drabold, DA. 2005. Maximum entropy and the problem of moments: A stable algorithm. *Physical Review E*, 71

Novi Inverardi, P. and Pontuale, G. and Petri, A. and Tagliani, A. 2003. Hausdorff moment problem via fractional. *Applied Mathematics and Computation*. 144:61--74

Acknowledgments

I thank Rachel Fewster for her supervision and guidance, Lisa and Benny for their useful discussions. and I thank my department for providing its support. Funding for this project was provided by The New Zealand Institute of Mathematics and its Applications.

For further information

Please contact *jliu070@aucklanduni.ac.nz*