INFORMATION GEOMETRY AND THE WRIGHT-FISHER MODEL OF MATHEMATICAL POPULATION GENETICS

Der Fakultät für Mathematik und Informatik
der Universität Leipzig
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM

(DR. RER. NAT.)

IM FACHGEBIET

MATHEMATIK

VORGELEGT VON

Master-Mathematiker Tat Dat Tran

geboren am 04. März 1982
in Hanoi - Vietnam

Die Annahme der Dissertation wurde empfohlen von:

1. **Prof. Dr. Jürgen Jost (MPI-MiS, Leipzig, Deutschland)**
2. **Prof. Dr. Shun-ichi Amari (RIKEN-BSI, Saitama, Japan)**

Die Verleihung des akademischen Grades erfolgt auf Beschluss des Rates der Fakultät für Mathematik und Informatik vom 04.07.2012 mit dem Gesamtprädikat magna cum laude.

Leipzig, den July 04, 2012

Acknowledgment

I owe a major debt to my advisor, Prof. Dr. Jürgen Jost, for giving me the opportunity to join his research groups, for giving me this interesting and challenging research project as well as for his friendship, constant availability, enthusiasm. I also would like to thank Prof. Dr. Nguyen Van Minh, for recommending me to the IMPRS fellowship program.

I am indebted to the Max Planck Institute for Mathematics in the Sciences who provided with the IMPRS an environment of hospitality and the possibility to get in touch with many other interesting PhD-students of various fields.

I wish also to thank the IMPRS and project VARIOGEO (GA-No 267087) for finance support, which allow me take up doctoral studies at University of Leipzig, Germany.

My gratitude also goes to Julian Hofrichter, Arun Maiti, Johannes Rauh, Bobo Hua, Chiranjib Mukherjee, etc., my colleagues in MPI MiS for many valuable suggestion and discussion hours.

Many thanks also go to A. Maiti, C. Mukherjee, J. Hofrichter for helping me handing in my thesis with correct grammar.

Grateful appreciation is also due to Dr. Hayk Mikayelyan, Ms. Antje Vandenberg, Ms. Valeria Hünniger, Ms. Katja Heid, Ms. Heike Rackwitz, Ms. Elke Herrmann, and staffs from the guest program, library, computer, etc., for their warm friendship and constant help during my stay in Leipzig.

Last but not least, I want to sincerely thank my parents and my sister, without their sacrifices and encouragement the thesis would never have been produced.

Dedicated to my parents and my sister

Abstract

This thesis is about a systematic approach to stochastic models in population genetics. We considered Markov chain models called the Wright-Fisher models, and their continuous versions, diffusion models for the evolution of finite monoecious diploid populations of non-overlapping random mating.

An important question in population genetics is that how do genetic change factors (random genetic drift, selection, mutation, migration, random environment, etc.) affect the behavior of gene frequencies or genotype frequencies in generations?

As we know, in a Hardy-Weinberg model, the Mendelian population model of a very large number of individuals without genetic change factors, gene frequencies remain unchanged from generation to generation, and genotype frequencies from the second generation onward remain also unchanged (see for example [33, 20]). With directional genetic change factors (selection, mutation, migration), we will have a deterministic dynamics of gene frequencies, which has been studied in detail in [33, 20]. With non-directional genetic change factors (random genetic drift, random environment), we will have a stochastic dynamics of gene frequencies, which has been studied with much more interests ([21, 24]). A combination of these factors has also been considered there.

In this thesis, I will only focus on affects of random genetic drift to the WF-models and the corresponding diffusion models without other factors (such as selection, mutation, migration, or random environment, etc.). This thesis is organized as follows.

Chapter 1 is an introduction to Wright-Fisher models and problems of interest.

Chapter 2 briefly discusses probabilistic aspects of the Wright-Fisher models. Results given here will be compared with results in Chapters 4, 5.

Chapter 3 focuses on the rigorous approach to the convergence of the Wright-Fisher models to diffusion models. We will see that diffusion model is a good approximation of the corresponding WF model. We also recall the method of Kolmogorov equations to diffusion models. The classification of boundaries has been considered.

In Chapter 4, we will systematically consider the diffusion model in one dimension. We recall results obtained by Kimura and others concerning local solutions, then we define a new solution (global solution) as a probability density function of diffusion model on the whole state space (Problem 4.3.5-4.3.7). We will prove the existence and uniqueness of this global solution (Theorem 4.9). Then we apply this global solution to calculate some genetic quantities such as probabilities of fixation, coexistence, heterogeneity, $k^{th}-$moments, and the absorption time as well as distribution at the absorption time. Results have been checked with results known.

In Chapter 5, we will systematically consider the diffusion model in higher dimension. We recall local results due to different methods by Kimura, Littler and Fackerell, Baxter, Blythe, McKane. We also define a global solution for this general diffusion model ([65]). With this new definition, we can prove that there is such a unique solution (Theorem 5.19). J. Hofrichter's hierarchichal product and boundary then allow the more precise definition of the global solution and yield a more transparent proof (Definition 5.7 and Section 5.3.5.). We then apply this global solution to calculate some genetic quantities (Section 5.4.).

Some genetic quantities have been known to satisfy singular elliptic linear second order equations with given boundary values. A lot of papers, textbooks have used this property to find those quantities ([16, 24]). However, the uniqueness of these problems have not been proved. Littler, in his PhD thesis in 1975 ([52]), took up the uniqueness problem but his proof, in my view, is not rigorous. In joint work with J. Hofrichter, we showed two different ways to prove the uniqueness rigorously. The first way is the approximation method (Lemma 5.25). The second way is the blow-up method which is conducted by J. Hofrichter ([34]).

In Chapter 6, we will consider geometric structures lying in the existing biological phenomena to get a deeper understanding of them. Firstly, we note that the state space is an $n-$dimensional smooth statistical manifold, an Einstein space, and also a dually flat manifold with the Fisher metric. We then will see that the Fisher metric is nothing but a the standard metric on the positive part of sphere of radius two. Next, we consider the affine Laplacian as well as its behavior in various coordinates and on various spaces. Finally, we deal with dynamics on the whole state space.

Appendix 1 is a brief introduction to hypergeometric functions used in Chapter 4. These functions are very useful tools for solving singular linear second order ODEs.

Appendix 2 is an overview of generalized hypergeometric functions used in Chapter 5. These functions are very useful tools for solving singular linear second order PDEs.

Information geometry is a bridge connecting non-Euclidean geometry and probability theory which reached maturity through the work of Amari in 1980 ([2]). The main idea is to find out the correspondence between structure of the families of distributions and that of manifolds. Formally, we can consider a distribution as a point, the score as a tangent vector, a family of distributions as a Riemannian manifold with the Riemannian metric as the Fisher information metric, etc. Then geometrical and probability theoretical results are interchangeable. Appendix 3 summaries the basis of Information Geometry used in Chapter 6.

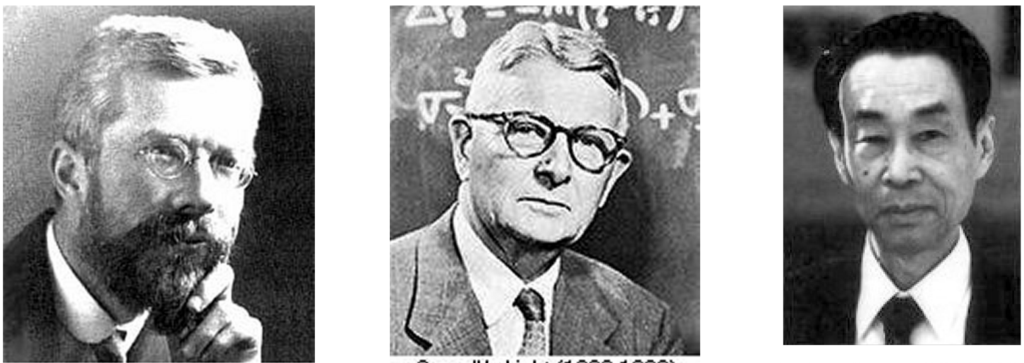
Contents

Chapter 1

Wright-Fisher models

In this chapter, we shall introduce one of the most popular stochastic models in population genetics, the Wright-Fisher models. We shall then give brief overview of their generalized versions. We also discuss some interesting questions about population genetics.

1.1. Introduction



Ronald Aylmer Fisher (1890-1962) Sewall Wright (1889-1988) Motoo Kimura (1924-1994)

Figure 1.1 Fisher, Wright and Kimura

In a population of finite size, changes in gene frequencies must be viewed as part of a stochastic (rather than a deterministic) process. So, it is necessary to set up a stochastic model which reasonably describes the behavior of a population in the stochastic case and arrive at theoretical estimates keeping track of the stochastic factor.

One of the stochastic models, caused by random genetic drift, was first introduced implicitly by Fisher[1] and explicitly by Wright[2] in 1920s, 1930s and was called the Wright-Fisher model (the WF model for short).

The effect of genetic drift is larger in small populations, and smaller in large populations. Many debates occurred over the relative importance of natural selection versus neutral processes, including genetic drift. In 1968, Kimura[3] rekindled the debate with his neutral theory of molecular evolution which claims that most of the changes in the genetic material (although not necessarily changes in phenotypes) are caused by random genetic drift.

For convenience, we describe some terminologies we have used in this thesis, for detail, we refer the readers to [20, 32, 33]:

- Diploid population means that we consider the somatic cells whose chromosome occur in pairs;
- Haploid population means that we consider the germ cells having only half the number of chromosome, one from each pair;
- Phenotype is the set of manifested attributes of an organism, for instance eye color or blood group;
- Locus (in a diploid population) is the specific location of a pair of genes on a chromosome $(A, B, \ldots)$;
- Alleles are different types of genes which may occupy a locus $(A_1, \ldots, A_n)$;
- Genotype is determined by a pair of alleles which actually occurs;
- The genotype is homozygous if the same allele appears twice (A_iA_i);
- The genotype is heterozygous if the different alleles appear (A_iA_j);
- An individual is monoecious if it has both male and female reproductive units;
- If A_iA_j manifests itself as A_iA_i then A_i is dominant and A_j is recessive.

1.2. The simplest WF model

We consider here a diploid population of fixed size N. Suppose that the individuals in this population are monoecious, that no selective difference between two possible alleles A_1, A_2 at a given locus A, and that there is no mutation. There are $2N$ alleles in the population in any generation, and it is sufficient to center our attention on the number

[1] Fisher R. A., *On the dominance ratio*, Proc. Roy. Soc. Edinb., **42** (1922), 321-341.
[2] S. Wright, *Evolution in Mendelian populations*, Genetics, **16** (1931), 97-159.
[3] Kimura, M., *Evolutionary rate at the molecular level*, Nature, **217**, (1968) 624-626.

X of A_1 alleles. Clearly in any generation X takes one of the values $\{0, 1, \ldots, 2N\}$, and we denote the value X in generation t by X_t. The model which we consider assumes that the alleles in generation $t+1$ are derived by sampling with replacement from the alleles in generation t (multinomial sampling). This means that the number X_{t+1} is a binomial random variable with index $2N$ and parameter $\frac{X_t}{2N}$. In another word, the transition probability function is

$$\begin{aligned} p_{i,j} =& \mathbb{P}\left(X_{t+1} = j \middle| X_t = i\right) \\ =& \binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{2N-j} \quad \text{for } i, j = 0, \ldots, 2N. \end{aligned} \tag{1.2.1}$$

1.3. The general WF models

By dropping out some assumptions of the simplest WF model, we will have more general WF models with more interesting problems. Below are brief introductions about some general WF models.

- The WF models of multi-alleles: Which will be considered in detail in this thesis;
- The WF models of the finite size - The effective population size: When the population size is finite, it is considered as a special constant, called the effective population size (usually smaller than the absolute population size). The concept of effective population size N_e was first introduced by Wright in two landmark papers [69, 70]. Effective population size may be defined by $\frac{1}{N_e} = \frac{1}{t}\sum_{i=1}^{t}\frac{1}{N_i}$.
- The WF models of multilocus: When population has more than one locus, perhaps there are genetic linkages between loci, i.e. there the cross combine of alleles on other loci. To consider this problem, the recombination frequency has been introduced. See, for example, [20, 31, 32, 52].
- The WF models with Selection: When the affect of selection is considerable, the fitness function should be introduced. We can see that the affect of selection is as a directional force. See, for example, [20, 21, 31, 32].
- The WF models with Mutations: When the affect of mutation is considerable, the mutation rate should be introduced. We can see that the affect of mutation is as a directional force. See, for example, [20, 21, 31, 32].

In this thesis, to emphasize the mathematical methods, we will focus on only the WF model with multi-alleles, one locus, no selection, no mutation, and call it the general WF model.

Chapter 2

Probability aspects

In this chapter we will discuss some properties of WF-models. The moment results will be applied in Chapter 3 as the conditions to approximate the diffusion models. Moreover, we demonstrate the behavior of gene frequencies in these WF-models which will be compared with the corresponding ones in diffusion models in Chapter 4 and Chapter 5.

2.1. Motivation

The problem of WF models as a problem in finite Markov chains was first considered by Malécot [56]. Based on the largest non-unit eigenvalue of the transition probability matrix $\mathbf{M}$, Malécot obtained the asymptotic rate of decrease of heterozygosity. Feller in 1951 [27] succeeded in finding a general expression for all eigenvalues of $\mathbf{M}$.

2.2. The simplest WF model

Let X_τ be the relative frequency of alleles A_1 in population at generation τ. By rescaling time axis $t = \frac{\tau}{2N}$, we consider Markov chain $\{X_t\}_{t \in \frac{1}{2N}\mathbb{N}_0}$ taking values in $\{0, \frac{1}{2N} \ldots, 1\}$, with the transition probability matrix

$$\mathbf{M} = (p_{ij})_{i,j}\,, \qquad \text{for } i,j = 0,\dots,2N,$$

where

$$\begin{aligned} p_{i,j} =& \mathbb{P}\left(X_{t+\delta t} = \frac{j}{2N} \middle| X_t = \frac{i}{2N} \right) \\ =& \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} . \end{aligned} \tag{2.2.1}$$

2.2.1. The probability mass function

Denote by $\mathbf{p}_t$ the probability mass function in population at time t of the simplest WF model. Then we have the following result:

Theorem 2.1. *If $\mathbf{p}_0$ is the initial probability mass function in population, then the probability mass function at time t is*

$$\boxed{\mathbf{p}_t = \mathbf{M}^t \mathbf{p}_0,}$$

where

$$M_{ij} = p_{i,j} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \quad \textit{for } i,j = 0,\dots,2N. \tag{2.2.2}$$

By direct calculation, we see that the genetic frequency could be accumulated at boundary and after a large number of generations, the genetic frequency concentrate on the boundary. It means that the population eventually becomes homozygous (one allele is extinct).

Below is the Mathematica code to plot the behavior of the population in discrete case.

```
a[n_, i_, j_] :=
 If[(i <= 0 && j <= 0) || (i >= n && j >= n), 1,
  Binomial[n, j] (i/n)^j (1 - i/n)^(n - j)]
Matran[n_] := Table[a[n - 1, j - 1, i - 1], {i, n}, {j, n}]
MP[n_, k_] := MatrixPower[Matran[n], k]
X0[n_, p_] := Table[If[i < p || i > p, 0, 1], {i, n}, {j, 1}]
X[n_, k_, p_] := MP[n, k].X0[n, p]
G[n_, k_, p_] := BarChart[Table[X[n, k, p][[i]], {i, 1, n}]]
Table[G[9, k, 5], {k, 0, 10}]
Table[G[20, k, 5], {k, 0, 50}]
```

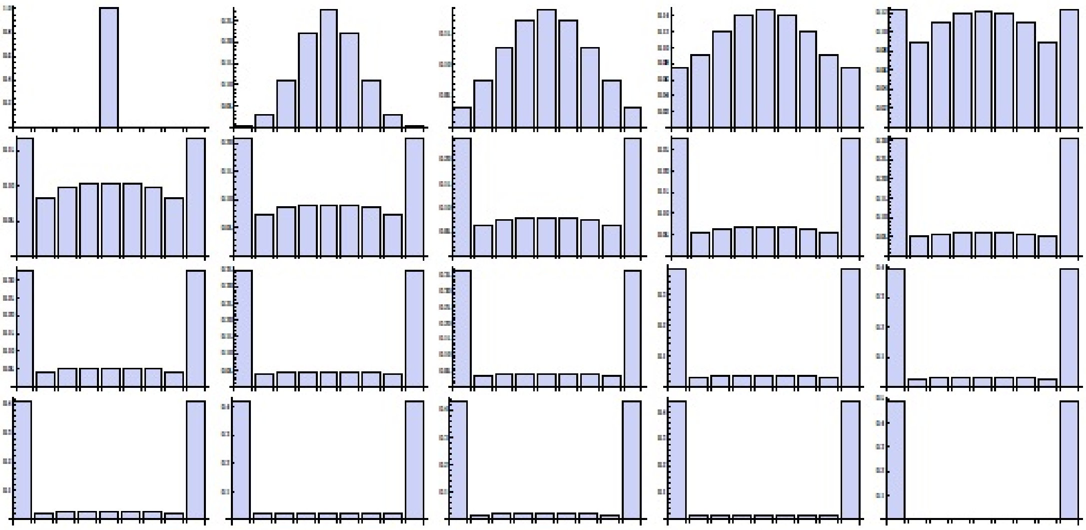


Figure 2.1 Behavior of the probability mass function of $2N = 9, p = 0.5$ at times $t = \{0, 1, \cdots 18\}$ and $t = 32$

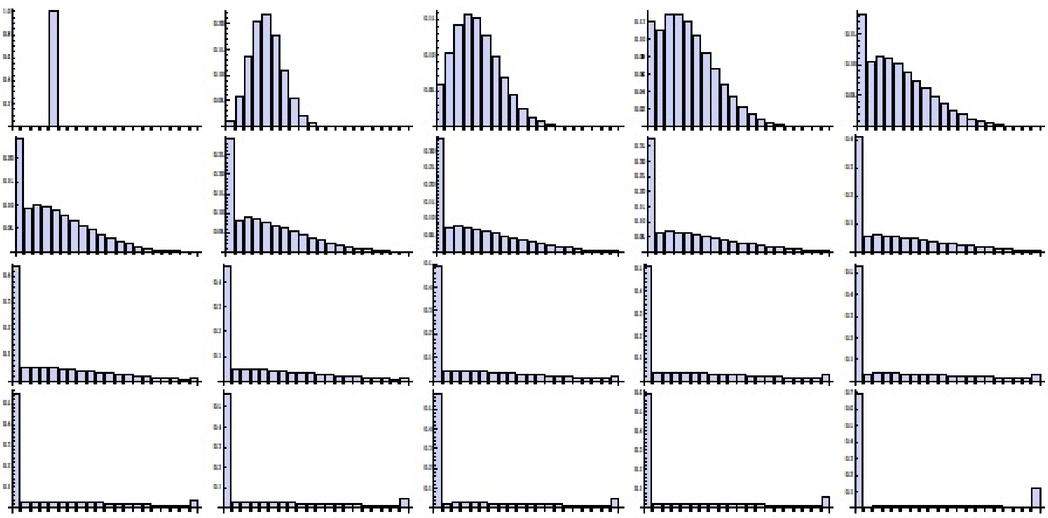


Figure 2.2 Behavior of the probability mass function of $2N = 20, p = 0.25$ at times $t = \{0, 1, \cdots 18\}$ and $t = 30$

2.2.2. Moments

By direct calculation, we immediately have (where $\delta t = \frac{1}{2N}$) the following proposition:

Proposition 2.2. *We consider a Markov chain X_t as above. Assume that x is the initial state, then we obtain*

$$\mathbb{E}(\delta x|x) = 0; \tag{2.2.3}$$

$$\mathbb{E}((\delta x)^2|x) = x(1-x)\delta t; \tag{2.2.4}$$

$$\mathbb{E}((\delta x)^k|x) = o(\delta t) \text{ with } k \geq 3. \tag{2.2.5}$$

Proof. In fact,

- $\sum\limits_{j=0}^{2N} p_{i,j} = \left(\frac{i}{2N} + \left(1 - \frac{i}{2N}\right)\right)^{2N} = 1;$

-
$$\begin{aligned}
\sum_{j=0}^{2N} j p_{i,j} &= 2N \sum_{j=0}^{2N} \binom{2N-1}{j-1} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \\
&= i \sum_{j=1}^{2N} \binom{2N-1}{j-1} \left(\frac{i}{2N}\right)^{j-1} \left(1 - \frac{i}{2N}\right)^{2N-j} \\
&= i \left(\frac{i}{2N} + \left(1 - \frac{i}{2N}\right)\right)^{2N-1} \\
&= i.
\end{aligned}$$

-
$$\begin{aligned}
\sum_{j=0}^{2N} j(j-1) p_{i,j} &= 2N(2N-1) \sum_{j=2}^{2N} \binom{2N-2}{j-2} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \\
&= \frac{2N-1}{2N} i^2 \sum_{j=2}^{2N} \binom{2N-2}{j-2} \left(\frac{i}{2N}\right)^{j-2} \left(1 - \frac{i}{2N}\right)^{2N-j} \\
&= \frac{2N-1}{2N} i^2 \left(\frac{i}{2N} + \left(1 - \frac{i}{2N}\right)\right)^{2N-2} \\
&= \frac{2N-1}{2N} i^2.
\end{aligned}$$

- $$\begin{aligned}\sum_{j=0}^{2N} j(j-1)(j-2)p_{i,j} &= 2N(2N-1)(2N-2)\sum_{j=3}^{2N}\binom{2N-3}{j-3}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{2N-j}\\ &= \frac{(2N-1)(2N-2)}{(2N)^2}i^3\sum_{j=3}^{2N}\binom{2N-3}{j-3}\left(\frac{i}{2N}\right)^{j-3}\left(1-\frac{i}{2N}\right)^{2N-j}\\ &= \frac{(2N-1)(2N-2)}{(2N)^2}i^3\left(\frac{i}{2N}+\left(1-\frac{i}{2N}\right)\right)^{2N-3}\\ &= \frac{(2N-1)(2N-2)}{(2N)^2}i^3.\end{aligned}$$

- By induction, we have

$$\sum_{j=0}^{2N} j(j-1)\dots(j-k)p_{ij} = \frac{(2N-1)\cdots(2N-k)}{(2N)^k}i^{k+1}.$$

Thus, by setting

$$x=\frac{i}{2N}, \delta x = \frac{j-i}{2N}, \delta t = \frac{1}{2N}$$

we conclude that the conditional expectation of the change is

$$\mathbb{E}(\delta x|x) = \sum_{j=0}^{2N}\left(\frac{j-i}{2N}\right)p_{i,j} = \frac{1}{2N}\left(\sum_{j=0}^{2N} jp_{i,j} - i\sum_{j=0}^{2N}p_{i,j}\right) = \frac{1}{2N}(i-i)=0;$$

the conditional variance of the change is

$$\begin{aligned}\mathbb{E}((\delta x)^2|x) &= \sum_{j=0}^{2N}\left(\frac{j-i}{2N}\right)^2 p_{i,j}\\ &= \frac{1}{4N^2}\left(\sum_{j=0}^{2N} j^2p_{i,j} - 2i\sum_{j=0}^{2N} jp_{i,j} + i^2\sum_{j=0}^{2N}p_{i,j}\right)\\ &= \frac{1}{4N^2}\left(\sum_{j=0}^{2N} j(j-1)p_{i,j} + \sum_{j=0}^{2N} jp_{i,j} - 2i^2 + i^2\right)\\ &= \frac{1}{4N^2}\left(\frac{2N-1}{2N}i^2 + i - i^2\right)\\ &= \frac{1}{4N^2}\left(i - \frac{1}{2N}i^2\right)\\ &= \left(\frac{i}{2N} - \left(\frac{i}{2N}\right)^2\right)\frac{1}{2N}\\ &= x(1-x)(\delta t);\end{aligned}$$

and the conditional higher moments of the change are

$$\begin{aligned}
\mathbb{E}((\delta x)^3|x) =& \frac{1}{8N^3}\left(\sum_{j=0}^{2N} j^3 p_{i,j} - 3i\sum_{j=0}^{2N} j^2 p_{i,j} + 3i^2\sum_{j=0}^{2N} j p_{i,j} - i^3\sum_{j=0}^{2N} p_{i,j}\right)\\
=& \frac{1}{8N^3}(\sum_{j=0}^{2N} j(j-1)(j-2)p_{i,j} + 3\sum_{j=0}^{2N} j(j-1)p_{i,j} + \sum_{j=0}^{2N} jp_{i,j}\\
&- 3i\left(\frac{2N-1}{2N}i^2 + i\right) + 3i^3 - i^3)\\
=& \frac{1}{8N^3}(\frac{(2N-1)(2N-2)}{4N^2}i^3 + 3\frac{2N-1}{2N}i^2 + i\\
&- 3i\left(\frac{2N-1}{2N}i^2 + i\right) + 3i^3 - i^3)\\
=& \frac{1}{8N^3}\left(\frac{2}{4N^2}i^3 - \frac{3}{2N}i^2 + i\right)\\
=& \left(2\left(\frac{i}{2N}\right)^3 - 3\left(\frac{i}{2N}\right)^2 + \frac{i}{2N}\right)\left(\frac{1}{2N}\right)^2\\
=& (2x^3 - 3x^2 + x)(\delta t)^2\\
=& o(\delta t),
\end{aligned}$$

and similarly, we have

$$\mathbb{E}((\delta x)^4|x) = o(\delta t),$$

as well as for $k > 4$

$$\begin{aligned}
\mathbb{E}((\delta x)^k|x) =& \sum_{j=0}^{2N}\left(\frac{j-i}{2N}\right)^k p_{i,j} \quad \text{for} \quad k > 4\\
\leq& \sum_{j=0}^{2N}\left(\frac{j-i}{2N}\right)^4 p_{i,j} \quad \text{because of} \quad \left|\frac{j-i}{2N}\right| \leq 1\\
=& o(\delta t).
\end{aligned}$$

This completes the proof. □

From (2.2.3) it follows the expected number of alleles is constant in time.

2.2.3. Eigenvalues of M

We present here all eigenvalues of **M** proved first by Feller in 1951

Lemma 2.3. *The transition probability matrix* **M** *has* $2N+1$ *eigenvalues as*

$$\lambda_0 = 1, \qquad \lambda_j = \binom{2N}{j}\frac{j!}{(2N)^j} \quad \textit{for } j = 1, \cdots, 2N$$

Proof. It is easy to see that $\lambda_0 = \lambda_1 = 1$ is the eigenvalue with the corresponding eigenvectors $\mathbf{u}_0 = (1, 1, \cdots, 1)$ and $\mathbf{u}_1 = (0, \frac{1}{2N}, \cdots, 1)$. We now find out other eigenvalues λ_r and their corresponding eigenvectors $\mathbf{u}_r$ by solving the system of linear equations

$$\sum_{k=0}^{2N} p_{jk} u_r^k = \lambda_r u_r^j, \quad j = 0, 1, \ldots, 2N.$$

Put for abbreviation

$$k_{[\nu]} = k(k-1)\ldots(k-\nu+1) = \frac{\Gamma(k)}{\Gamma(k-\nu)}$$

and note that

$$\sum_{k=0}^{2N} p_{jk} k_{[\nu]} = \frac{d^\nu}{dx^\nu}\Big(1 - \frac{1}{2N} + \frac{1}{2N}x\Big)^{2N}|_{x=1} = (2N)_{[\nu]}\Big(\frac{1}{2N}\Big)^\nu. \tag{2.2.6}$$

We will find the eigenvector u_r^k as a polynomial in k of degree r in the following form

$$u_r^k = a_r k_{[r]} + a_{r-1} k_{[r-1]} + \cdots + a_1 k + a_0.$$

In fact, a_i can be solved as solution of system

$$\begin{aligned}
\lambda_r \sum_{\nu=0}^{r} j_{[\nu]} a_\nu &= \sum_{k=0}^{2N} p_{jk} \sum_{\nu=0}^{r} k_{[\nu]} a_\nu \\
&= \sum_{\nu=0}^{r} a_\nu \sum_{k=0}^{2N} p_{jk} k_{[\nu]} \\
&= \sum_{\nu=0}^{r} a_\nu (2N)_{[\nu]} \Big(\frac{1}{2N}\Big)^\nu, \quad \text{due to (2.2.6)}
\end{aligned}$$

for $r = 0, 1, \ldots, 2N$.

We expand once again $\left(\frac{1}{2N}\right)^\nu$ by $j_{[s]}$ in the form

$$\Big(\frac{1}{2N}\Big)^\nu = \sum_{s=0}^{\nu} c_{s,\nu} j_{[s]}, \tag{2.2.7}$$

then we obtain

$$\lambda_r \sum_{s=0}^{r} j_{[s]} a_s = \lambda_r \sum_{\nu=0}^{r} j_{[\nu]} a_\nu \tag{2.2.8}$$

$$= \sum_{\nu=0}^{r} (2N)_{[\nu]} \sum_{s=0}^{\nu} c_{s,\nu} j_{[s]} a_\nu \tag{2.2.9}$$

$$= \sum_{s=0}^{r} j_{[s]} \sum_{\nu=s}^{r} (2N)_{[\nu]} c_{s,\nu} a_\nu. \tag{2.2.10}$$

Equating the coefficients in (2.2.10) we get

$$\sum_{\nu=s}^{r}(2N)_{[\nu]}c_{s,\nu}a_{\nu} = \lambda_r a_s, \quad s = 0, \ldots, r. \tag{2.2.11}$$

Note that in (2.2.7)

$$c_{\nu,\nu} = \frac{1}{(2N)^{\nu}}.$$

Then for $s = \nu = r$ in (2.2.11), we get

$$\lambda_r = c_{r,r}(2N)_{[r]} = \frac{(2N)_{[r]}}{(2N)^r} = \binom{2N}{r}\frac{r!}{(2N)^r},$$

and a_r is arbitrary. Put $a_r = 1$, we can calculate in succession $a_{r-1}, a_{r-2}, \ldots, a_0$ to get the eigenvector $\mathbf{u}_r$. □

Remark 2.4. Some first eigenvectors are

$$u_2^k = \frac{k}{2N}\Big(1 - \frac{k}{2N}\Big)$$

and

$$u_3^k = \frac{k}{2N}\Big(1 - \frac{k}{2N}\Big)\Big(\frac{1}{2} - \frac{k}{2N}\Big).$$

2.2.4. Probability of fixation

By denoting

$p_{i,j}^n$ is the probability of starting at i alleles and having j alleles after n generations

we have the following lemma

Lemma 2.5.

$$\lim_{n\to\infty} p_{i,0}^n = 1 - \frac{i}{2N}$$

$$\lim_{n\to\infty} p_{i,2N}^n = \frac{i}{2N}$$

$$\lim_{n\to\infty} p_{i,j}^n = 0 \text{ } \mathit{for} \text{ } j \neq 0 \text{ } \mathit{or} \text{ } 2N$$

Proof.

$$p_{i,j}^n = \binom{2N}{j}\sum_{t=0}^{2N-j}(-1)^t\binom{2N-j}{t}(2N)^{-t-j}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}\lambda_s^{n-1}u_{j+t}^s v_s^{\beta} i^{\beta}$$

$$p_{i,2N}^n = \frac{i}{2N} + (2N)^{-2N}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}\lambda_s^{n-1}u_{2N}^s v_s^{\beta} i^{\beta}$$

$$p_{i,0}^n = 1 - \frac{i}{2N} + \sum_{t=1}^{2N}(-1)^t\binom{2N}{t}(2N)^{-t}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}\lambda_s^{n-1}u_t^s v_s^\beta i^\beta$$

See [41, 42] □

Remark 2.6. It is easy to see that

$$p_{i,0}^n > 0, \quad p_{i,2N}^n > 0, \quad p_{0,j}^n = p_{2N,j}^n = 0 \quad \text{for } j \neq 0 \text{ or } 2N.$$

It means that boundaries 0 and $2N$ are exit boundaries.

2.2.5. Absorption times and distribution at absorption time

In the previous subsection we know that extinction or fixation of the allele A_1 occurs with probability one. This conclusion suggests the interesting problem of determining the time required for the population to reach one of the homozygous conditions. To do that, we denote by τ_i the first time taken to reach 0 or $2N$ given that initially there were i allele A_1

Denote by

$$C(i,s,\beta) = (\lambda_s - 1)[i^\beta + (2N-i)^\beta]u_{2N}^s v_s^\beta.$$

Then we have

Lemma 2.7.

$$\mathcal{P}(\tau_i = n) = (2N)^{-2N}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}C(i,s,\beta)\lambda_s^{n-2};$$

$$\mathbb{E}(\tau_i) = \Big(1 - \frac{i}{2N}\Big)^{2N} + \Big(\frac{i}{2N}\Big)^{2N} + (2N)^{-2N}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}C(i,s,\beta)\frac{2-\lambda_s}{(1-\lambda_s)^2};$$

$$\begin{aligned}\mathbb{E}(\tau_i^2) =& \Big(1 - \frac{i}{2N}\Big)^{2N} + \Big(\frac{i}{2N}\Big)^{2N} + (2N)^{-2N}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}C(i,s,\beta)\frac{2-\lambda_s}{(1-\lambda_s)^2}\\ &+ (2N)^{-2N}\sum_{s=2}^{2N}\sum_{\beta=1}^{s}C(i,s,\beta)\frac{2}{(1-\lambda_s)^3}.\end{aligned}$$

Proof. See [41, 42]. □

2.3. The general WF model

Similar to the simplest WF model case, with X_τ^i the related frequency of alleles A_i $(i = 1, \ldots, n)$ in population at generation τ, by rescaling time axis $t = \frac{\tau}{2N}$ we receive a discrete time discrete space Markov chain $\{\mathbf{X}_t\}_{t\in\frac{1}{2N}\mathbb{N}_0}$ in

$$S_n^{(2N)} = \left\{\mathbf{x} = (x^1,\ldots,x^n) : x^i \in \left\{0, \frac{1}{2N},\ldots,1\right\}, \sum_{i=1}^n x^i \le 1\right\},$$

with the transition probability function

$$\begin{aligned} p_{\boldsymbol{\alpha},\boldsymbol{\beta}} =& \mathbb{P}\left(\mathbf{X}_{t+\delta t} = \frac{\boldsymbol{\beta}}{2N} | \mathbf{X}_t = \frac{\boldsymbol{\alpha}}{2N}\right) \\ =& \frac{2N!}{\beta^1!\ldots\beta^{n+1}!}\left(\frac{\alpha^1}{2N}\right)^{\beta^1}\cdots\left(\frac{\alpha^{n+1}}{2N}\right)^{\beta^{n+1}} \quad \text{for } \boldsymbol{\alpha},\boldsymbol{\beta}\in\Omega. \end{aligned} \tag{2.3.1}$$

where $\alpha^{n+1} = 2N - \alpha^1 - \ldots - \alpha^n; \beta^{n+1} = 2N - \beta^1 - \ldots - \beta^n$.

2.3.1. Moments

By directly calculating we will immediately have (where $\delta t = \frac{1}{2N}$) the following proposition

Proposition 2.8.
- *$\mathbb{E}(\delta x^i|\mathbf{x}) = 0$;*
- *$\mathbb{E}(\delta x^i \delta x^j|\mathbf{x}) = b^{ij}(\mathbf{x})\delta t$, where $b^{ij}(\mathbf{x}) = x^i(\delta_{ij} - x^j)$;*
- *$\mathbb{E}((\delta x)^\delta|\mathbf{x}) = o(\delta t)$ with $\delta \in \mathbb{N}_0^n : |\delta| = \delta^1 + \ldots + \delta^n \ge 3$.*

Proof. Prove similarly as in the simplest WF model, we note that

- $\sum\limits_{\boldsymbol{\beta}\in\Omega} p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \left(\frac{\alpha^1}{2N} + \cdots + \frac{\alpha^n}{2N}\right)^{2N} = 1$;
- $\sum\limits_{\boldsymbol{\beta}\in\Omega} \beta^i p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \alpha^i$;
- $\sum\limits_{\boldsymbol{\beta}\in\Omega} \beta^i\beta^j p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{2N-1}{2N}\alpha^i\alpha^j$;
- $\sum\limits_{\boldsymbol{\beta}\in\Omega} (\beta^i)^2 p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{2N-1}{2N}(\alpha^i)^2 + \alpha^i$;
- $\sum\limits_{\boldsymbol{\beta}\in\Omega} \beta^i\beta^j\beta^k p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{(2N-1)(2N-2)}{4N^2}\alpha^i\alpha^j\alpha^k$;
- $\sum\limits_{\boldsymbol{\beta}\in\Omega} (\beta^i)^2\beta^j p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{(2N-1)(2N-2)}{4N^2}(\alpha^i)^2\alpha^j + \frac{2N-1}{2N}\alpha^i\alpha^j$;
- $\sum\limits_{\boldsymbol{\beta}\in\Omega} (\beta^i)^3 p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \frac{(2N-1)(2N-2)}{4N^2}(\alpha^i)^3 + 3\frac{2N-1}{2N}(\alpha^i)^2 + \alpha^i$;

Thus, by setting

$$\mathbf{x} = \frac{\boldsymbol{\alpha}}{2N}, \delta x^i = \frac{\beta^i - \alpha^i}{2N}, \delta t = \frac{1}{2N},$$

we follow the conditional expectation of the change is

$$\mathbb{E}(\delta x^i|\mathbf{x}) = \sum_{\boldsymbol{\beta}\in\Omega} \left(\frac{\beta^i - \alpha^i}{2N}\right) p_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 0;$$

the conditional covariance of the change is

$$\begin{aligned}
\mathbb{E}((\delta x^i)(\delta x^j)|\mathbf{x}) &= \sum_{\boldsymbol{\beta}\in\Omega} \left(\frac{\beta^i - \alpha^i}{2N}\right)\left(\frac{\beta^j - \alpha^j}{2N}\right) p_{\boldsymbol{\alpha},\boldsymbol{\beta}} \\
&= \frac{1}{4N^2}\left(\sum_{\boldsymbol{\beta}\in\Omega} \beta^i\beta^j p_{\boldsymbol{\alpha},\boldsymbol{\beta}} - \alpha^i \sum_{\boldsymbol{\beta}\in\Omega} \beta^j p_{\boldsymbol{\alpha},\boldsymbol{\beta}} - \alpha^j \sum_{\boldsymbol{\beta}\in\Omega} \beta^i p_{\boldsymbol{\alpha},\boldsymbol{\beta}} + \alpha^i\alpha^j \sum_{\boldsymbol{\beta}\in\Omega} p_{\boldsymbol{\alpha},\boldsymbol{\beta}}\right) \\
&= \frac{1}{4N^2}\left(\frac{2N-1}{2N}\alpha^i\alpha^j - \alpha^i\alpha^j\right) \\
&= -x^i x^j (\delta t);
\end{aligned}$$

the conditional variance of the change is

$$\begin{aligned}
\mathbb{E}((\delta x^i)^2|\mathbf{x}) &= \sum_{\boldsymbol{\beta}\in\Omega} \left(\frac{\beta^i - \alpha^i}{2N}\right)^2 p_{\boldsymbol{\alpha},\boldsymbol{\beta}} \\
&= \frac{1}{4N^2}\left(\sum_{\boldsymbol{\beta}\in\Omega} (\beta^i)^2 p_{\boldsymbol{\alpha},\boldsymbol{\beta}} - 2\alpha^i \sum_{\boldsymbol{\beta}\in\Omega} \beta^i p_{\boldsymbol{\alpha},\boldsymbol{\beta}} + (\alpha^i)^2 \sum_{\boldsymbol{\beta}\in\Omega} p_{\boldsymbol{\alpha},\boldsymbol{\beta}}\right) \\
&= \frac{1}{4N^2}\left(\frac{2N-1}{2N}(\alpha^i)^2 + \alpha^i - (\alpha^i)^2\right) \\
&= x^i(1 - x^i)(\delta t);
\end{aligned}$$

and the conditional moments of the change are

$$\begin{aligned}
\mathbb{E}((\delta x^i)^3|\mathbf{x}) &= \left(2(x^i)^3 - 3(x^i)^2 + x^i\right)(\delta t)^2 \\
\mathbb{E}((\delta x^i)^2(\delta x^j)|\mathbf{x}) &= \left(2(x^i)^2 x^j - x^i x^j\right)(\delta t)^2 \\
\mathbb{E}((\delta x^i)(\delta x^j)(\delta x^k)|\mathbf{x}) &= \left(2x^i x^j x^k\right)(\delta t)^2,
\end{aligned}$$

Moreover, we easily have by the same way as the simplest WF model

$$\mathbb{E}((\delta x)^\delta|\mathbf{x}) = o(\delta t),\ \forall|\delta| \geq 3.$$

This completes the proof. □

2.3.2. Eigenvalues of M

Proposition 2.9. *The transition probability matrix M has eigenvalues*

$$\lambda_j = \binom{2N}{j} \frac{j!}{(2N)^j} \qquad \textit{for } j = 2, \cdots, 2N$$

with multiplicity

$$\binom{n-1+j}{n-1}$$

and $\lambda_1 = 1$ has multiplicity $n+1$.

Proof. See [24]. □

Chapter 3

Diffusion approximation and Kolmogorov equations

In this chapter we give rigorous proof for diffusion approximations to make sure that making limits is acceptable and manageable and derive Kolmogorov equations of these diffusion models.

3.1. Diffusion approximation

3.1.1. The simplest case

We start from the simplest WF model with N diploid individuals of type A_1, A_2. X_t is the frequency of allele A_1 is a Markov chain on the state space $\{0, 1, \cdots, 2N\}$ with the transition probability operator

$$P_N f(i) = \sum_{j=0}^{2N} p_{ij} f(j)$$

and the associated continuous time pure jump process has generator

$$L_N f(i) = (P_N - I) f(i) = \sum_{j=0}^{2N} p_{ij} \Big(f(j) - f(i) \Big)$$

One then considers the re-scaled process

$$X_N(t) = \frac{1}{2N} X_{2Nt}$$

which is a continuous-time Markov (pure jump) process on $\{0, \frac{1}{2N}, \ldots, 1\} \subset [0,1]$ with generator

$$\mathcal{L}_N f(\frac{i}{2N}) = 2N \sum_{j=0}^{2N} p_{ij} \Big(f(\frac{j}{2N}) - f(\frac{i}{2N}) \Big).$$

Expanding this expression up to second order, we obtain

$$\begin{aligned} \mathcal{L}_N f\Big(\frac{i}{2N}\Big) =& \sum_{j=0}^{2N} p_{ij}(j-i) f'\Big(\frac{i}{2N}\Big) + \frac{1}{2}\sum_{j=0}^{2N} p_{ij} \frac{(j-i)^2}{2N} f''\Big(\frac{i}{2N}\Big) \\ &+ \frac{1}{6}\sum_{j=0}^{2N} p_{ij} \frac{(j-i)^3}{4N^2} f'''(\theta), \end{aligned}$$

where $\theta \in (0,1)$. Now, in the limit $N \to \infty$, due to Proposition 2.2 we obtain a second order differential operator which we can associate to a diffusion process, where the first part corresponds to the "deterministic component" and the second order part corresponds to the "random component". We choose $i = i_N$ such that $\frac{i_N}{2N} \to x \in [0,1]$ as $N \to \infty$, and thus the limiting generator becomes

$$\mathcal{L} f(x) = \frac{1}{2} x(1-x) f''(x).$$

Definition 3.1. We say that a sequence of processes $\{X_n(t)\}_{t\geq 0}$ *converge weakly in path space* to the process $\{X(t)\}_{t\geq 0}$ if for all continuous functions f on trajectories.

$$\lim_{n\to\infty} \int f(\omega) \mathbb{P}_x^n(d\omega) = \int f(\omega) \mathbb{P}_x(d\omega),$$

where $\mathbb{P}_x^n, \mathbb{P}_x$ are unique distributions starting from x of the processes X_n, X correspondingly.

Using the Trotter-Kato theorem (see [18, 62, 74]) we obtain the weak convergence of the processes, i.e., the discrete processes will *converge weakly in path space* to the continuous process with the generator $\mathcal{L}$.

3.1.2. The general case

We denote by

$$S_n^{(M)} = \Big\{ \mathbf{x} = (x^1, \ldots, x^n) : \sum_{i=1}^n x^i \leq 1, x^i = \frac{j^i}{M} \geq 0 \quad j^i \text{ integer } , i = 1, \ldots, n \Big\},$$

$$R_n^{(M)} = \Big\{ \mathbf{x} = (x^1, \ldots, x^n) : \sum_{i=1}^n x^i < 1, x^i = \frac{j^i}{M} > 0 \quad j^i \text{ integer } , i = 1, \ldots, n \Big\},$$

$$S_n = \Big\{ \mathbf{x} = (x^1, \ldots, x^n) : \sum_{i=1}^n x^i \leq 1, x^i \geq 0, i = 1, \ldots, n \Big\} = \overline{V_n},$$

$$R_n = \Big\{\mathbf{x} = (x^1, \ldots, x^n) : \sum_{i=1}^{n} x^i < 1, x^i > 0, i = 1, \ldots, n\Big\} = V_n,$$

Suppose we have a sequence of finite Markov chains indexed by M, $\{\mathbf{X}^{(M)}(t)\}_{t\geq 0}$, having state space $S_n^{(M)}$ with discrete time t and state variable

$$\mathbf{X}^{(M)}(t) = (X_1^{(M)}(t), \ldots, X_n^{(M)}(t)).$$

We define, for each chain in the sequence, the transition distributions by

$$P^{(M)}(t, \mathbf{x}, B) = \mathcal{P}(X^{(M)}(t) \in B | X^{(M)}(0) = \mathbf{x}),$$

where we assume that $\mathbf{x} \in S_n^{(M)}$ and that B is a Borel subset of S_n.

Definition 3.2. We say that a sequence of Markov chains $\{X^{(M)}(t)\}_{M\geq 1}$ satisfies condition I if each chain satisfies

- $\mathbb{E}\Big(\delta\mathbf{X}_i^{(M)}(t)|\mathbf{X}^{(M)}(t) = \mathbf{x}(t)\Big) = a^i(\mathbf{x}(t))\delta t + o(\delta t), \quad i = 1, \ldots, n;$
- $\mathbb{E}\Big(\delta\mathbf{X}_i^{(M)}(t)\delta\mathbf{X}_j^{(M)}(t)|\mathbf{X}^{(M)}(t) = \mathbf{x}(t)\Big) = b^{ij}(\mathbf{x})\delta t + o(\delta t), \quad i, j = 1, \ldots, n;$
- $\mathbb{E}\Big((\delta\mathbf{X}^{(M)}(t))^\alpha|\mathbf{X}^{(M)}(t) = \mathbf{x}(t)\Big) = o(\delta t), \quad$ with $\alpha \in \mathbb{N}_0^n : |\alpha| = \alpha^1 + \ldots + \alpha^n \geq 3.$

Definition 3.3. We say that a sequence of Markov chains $\{X^{(M)}(t)\}_{M\geq 1}$ satisfies condition II if to any given polynomial $h_1(\mathbf{x})$, there corresponds a finite number of further polynomials $h_2(\mathbf{x}), \ldots, h_m(\mathbf{x})$ such that for each M and $\mathbf{x}(t) \in S_n$ we have

$$\mathbb{E}(h_i(\mathbf{X}^{(M)})(t+1)|\mathbf{X}^{(M)}(t) = \mathbf{x}(t)) = \sum_{j=1}^{m} w_{ij} h_j(\mathbf{x}(t)),$$

where $W = (w_{ij})_{i,j}$ is an matrix whose elements may depend on M.

Theorem 3.4. *Suppose the sequence of Markov chains $\{X^{(M)}(t)\}_{M\geq 1}$ satisfies conditions I and II, so that $\mathbf{x} \in S_n^{(M)}$ and that t is a nonnegative integer. Then the sequence of transition distributions*

$$\{P^{(kM)}(kt, \mathbf{x}, B)\}_{k\geq 1}$$

converges weakly to a unique proper distribution, $P(\tau, \mathbf{x}, B)$, on S_n with $\tau = \frac{t}{M}$.

Proof. We refer the reader to [53]. □

3.2. Kolmogorov equations for the general WF model

Andrey N. Kolmogorov (1903-1987)

Figure 3.1 Kolmogorov

The Kolmogorov backward equation (KBE) and its adjoint known as the Kolmogorov forward equation (KFE) are second order partial differential equations (PDE) that arise in the theory of continuous-time continuous-state Markov processes. Both were published by Andrey Kolmogorov[1] in 1931. Later it was realized that the forward equation was already known to physicists under the name Fokker-Planck equation.

The following proposition is due to Kolmogorov

Proposition 3.5. *Let $\{X(t)\}_{t\geq 0}$ be a $N-$dimensional state continuous time continuous Markov process with the conditional probability density function $f(x,t|y,s)$. Then the function f satisfies the Kolmogorov backward equation (KBE)*

$$-\frac{\partial f}{\partial s} = \sum_{i=1}^{N} a^i(\mathbf{y},s)\frac{\partial}{\partial y^i}\,[f] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} b^{ij}(\mathbf{y},s)\frac{\partial^2}{\partial y^i\,\partial y^j}\,[f]\,, \tag{3.2.1}$$

and the Kolmogorov forward equation (KFE)

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x^i}\left[a^i(\mathbf{x},t) f\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2}{\partial x^i\,\partial x^j}\left[b^{ij}(\mathbf{x},t) f\right]\,, \tag{3.2.2}$$

where,

$$a^i(\mathbf{x},t) = \lim_{\delta t\to 0}\frac{1}{\delta t}E(\delta x^i|\mathbf{x})$$

are the drift terms,

$$b^{ij}(\mathbf{x},t) = \lim_{\delta t\to 0}\frac{1}{\delta t}E(\delta x^i\delta x^j|\mathbf{x})$$

[1]Kolmogorov, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Mathematische Annalen, Vol. 104, Nr. 1, (1931) 415-458.

are diffusion terms.

Proof. We will prove for the forward equation. The backward equation is similar. In fact, for $h \in C_c^\infty(\mathbb{R}^N)$ a smooth function with a compact support, we have

$$\begin{aligned}
\int_{\mathbb{R}^N} h(\mathbf{x}) \frac{\partial f(\mathbf{x},t|\mathbf{y},s)}{\partial t} d\mathbf{x} &= \lim_{\delta t \to 0} \int_{\mathbb{R}^N} h(\mathbf{x}) \frac{f(\mathbf{x},t+\delta t|\mathbf{y},s) - f(\mathbf{x},t|\mathbf{y},s)}{\delta t} d\mathbf{x} \\
&= \lim_{\delta t \to 0} \int_{\mathbb{R}^N} h(\mathbf{x}) \frac{\int_{\mathbb{R}^N} f(\mathbf{x},t+\delta t|\mathbf{z},t) f(\mathbf{z},t|\mathbf{y},s) d\mathbf{z} - f(\mathbf{x},t|\mathbf{y},s)}{\delta t} d\mathbf{x} \\
&\quad \text{(using the Chapman-Kolmogorov equation)} \\
&= \lim_{\delta t \to 0} \frac{1}{\delta t} \Big[\int_{\mathbb{R}^N} f(\mathbf{z},t|\mathbf{y},s) \int_{\mathbb{R}^N} h(\mathbf{x}) f(\mathbf{x},t+\delta t|\mathbf{z},t) d\mathbf{x} d\mathbf{z} - \int_{\mathbb{R}^N} h(\mathbf{x}) f(\mathbf{x},t|\mathbf{y},s) d\mathbf{x} \Big] \\
&\quad \text{(using the Fubini theorem)} \\
&= \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}^N} f(\mathbf{z},t|\mathbf{y},s) \Big[\int_{\mathbb{R}^N} h(\mathbf{x}) f(\mathbf{x},t+\delta t|\mathbf{z},t) d\mathbf{x} - h(\mathbf{z}) \Big] d\mathbf{z} \\
&\quad \text{(transforming } \mathbf{x} \to \mathbf{z} \text{ in the second term)} \\
&= \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}^N} f(\mathbf{z},t|\mathbf{y},s) \Big\{ \int_{\mathbb{R}^N} f(\mathbf{x},t+\delta t|\mathbf{z},t) \Big[h(\mathbf{x}) - h(\mathbf{z}) \Big] d\mathbf{x} \Big\} d\mathbf{z} \\
&= \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}^N} f(\mathbf{z},t|\mathbf{y},s) \Big\{ \int_{\mathbb{R}^N} f(\mathbf{x},t+\delta t|\mathbf{z},t) \Big[\sum_{|\alpha|\geq 1} h^{(\alpha)}(\mathbf{z}) \frac{(\mathbf{x}-\mathbf{z})^\alpha}{\alpha!} \Big] d\mathbf{x} \Big\} d\mathbf{z} \\
&= \int_{\mathbb{R}^N} f(\mathbf{z},t|\mathbf{y},s) \sum_{|\alpha|\geq 1} \frac{1}{\alpha!} D_\alpha(\mathbf{z}) h^{(\alpha)}(\mathbf{z}) d\mathbf{z} \\
&\quad \text{where } D_\alpha(\mathbf{z}) = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}^N} f(\mathbf{x},t+\delta t|\mathbf{z},t)(\mathbf{x}-\mathbf{z})^\alpha d\mathbf{x} = \lim_{\delta t \to 0} \frac{1}{\delta t} E\Big((\delta \mathbf{x})^\alpha | \mathbf{x} \Big) \\
&= \int_{\mathbb{R}^N} h(\mathbf{z}) \Big\{ \sum_{|\alpha|\geq 1} \frac{(-1)^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial \mathbf{z}^\alpha} \Big(D_\alpha(\mathbf{z}) f(\mathbf{z},t|\mathbf{y},s) \Big) \Big\} d\mathbf{z}. \\
&\quad \text{integration by parts.}
\end{aligned}$$

Therefore we conclude

$$\int_{\mathbb{R}^N} h(\mathbf{x}) \Big\{ \frac{\partial f(\mathbf{x},t|\mathbf{y},s)}{\partial t} - \sum_{|\alpha|\geq 1} \frac{(-1)^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial \mathbf{x}^\alpha} \Big(D_\alpha(\mathbf{x}) f(\mathbf{x},t|\mathbf{y},s) \Big) \Big\} d\mathbf{x} = 0.$$

Because h is arbitrary, this follows that

$$\begin{aligned}\frac{\partial f(\mathbf{x},t|\mathbf{y},s)}{\partial t} &= \sum_{|\alpha|\geq 1} \frac{(-1)^\alpha}{\alpha!}\frac{\partial^\alpha}{\partial \mathbf{x}^\alpha}\Big(D_\alpha(\mathbf{x})f(\mathbf{x},t|\mathbf{y},s)\Big)\\ &= \sum_{1\leq|\alpha|\leq 2} \frac{(-1)^\alpha}{\alpha!}\frac{\partial^\alpha}{\partial \mathbf{x}^\alpha}\Big(D_\alpha(\mathbf{x})f(\mathbf{x},t|\mathbf{y},s)\Big)\\ &\quad \text{due to the continuity of process (diffusion)}\\ &= -\sum_{i=1}^{N}\frac{\partial}{\partial x^i}\left[a^i(\mathbf{x},t)f\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial x^i\,\partial x^j}\left[b^{ij}(\mathbf{x},t)f\right],\end{aligned}$$

The proof is finished. □

By applying for our general WF model, we have the drift terms

$$a^i(\mathbf{x}) = \lim_{\delta t\to 0}\frac{1}{\delta t}E(\delta x^i|\mathbf{x}) = 0,$$

and diffusion terms

$$b^{ij}(\mathbf{x}) = \lim_{\delta t\to 0}\frac{1}{\delta t}E(\delta x^i\delta x^j|\mathbf{x}) = x^i(\delta_{ij} - x^j)$$

Therefore, the Kolmogorov forward equation of the general WF model is

$$\frac{\partial u(\mathbf{x},\mathbf{p},t)}{\partial t} = \sum_{i,j=1}^{n-1}\frac{1}{2}\frac{\partial^2(b^{ij}(\mathbf{x})u(\mathbf{x},\mathbf{p},t))}{\partial x^i\partial x^j},$$

and the Kolmogorov backward equations is

$$\frac{\partial u(\mathbf{x},\mathbf{p},t)}{\partial t} = \sum_{i,j=1}^{n-1}\frac{1}{2}b^{ij}(\mathbf{p})\frac{\partial^2 u(\mathbf{x},\mathbf{p},t)}{\partial p^i\partial p^j},$$

where

$$b^{ij}(\mathbf{x}) = x^i(\delta_{ij} - x^j)$$

In particular, in the simplest WF model, we have also the Kolmogorov forward equation

$$\frac{\partial u(x,p,t)}{\partial t} = \frac{1}{2}\frac{\partial^2\Big(x(1-x)u(x,p,t)\Big)}{\partial x^2},$$

and the Kolmogorov backward equations

$$\frac{\partial u(x,p,t)}{\partial t} = \frac{1}{2}p(1-p)\frac{\partial^2 u(x,p,t)}{\partial p^2}.$$

3.3. Classification of boundaries

This section is to classify boundaries of a diffusion equation. In practical problems, due to some simple estimations we can learn the properties of the boundary and make some conclusions on the behavior of solutions at the boundary. This work was started by William Feller in 1952[2]. In the paper, he showed the result for a differential equation

$$u_t(t,x) = a(x)u_{xx}(t,x) + b(x)u_x(t,x),$$

and its adjoint

$$v_t(t,x) = \Big(\big(a(x)v(t,x)\big)_x - b(x)v_x(t,x)\Big)_x,$$

where $a'(x), b(x)$ are continuous, but not necessarily bounded, in the interval (r_1, r_2), $a(x) > 0$, $a \not\equiv 1$, $a''(x)$ need not exit. Set a function

$$W(x) = e^{-\int_{x_0}^{x} b(s)a^{-1}(s)ds},$$

where $x_0 \in (r_1, r_2)$ is fixed. Denote $L(r_1, r_2)$ by the space of integrable function on (r_1, r_2).

Then we have

Proposition 3.6.

*(i) The boundary r_j is **regular** if for all $x_0 \in (r_1, r_2)$*

$$W(x) \in L(x_0, r_j) \text{ and } a^{-1}(x)W^{-1}(x) \in L(x_0, r_j),$$

in this case, the trajectory can reach r_j from the interior of (r_1, r_2) and can return to the interior within a finite length of time;

*(ii) The boundary r_j is **exit** if for all $x_0 \in (r_1, r_2)$*

$$a^{-1}(x)W^{-1}(x) \notin L(x_0, r_j) \text{ and } W(x)\int_{x_0}^{x} a^{-1}(s)W^{-1}(s)ds \in L(x_0, r_j),$$

in this case, the trajectory can reach r_j from the interior of (r_1, r_2) within a finite length of time but can not return to the interior;

*(iii) The boundary r_j is **entrance** if for all $x_0 \in (r_1, r_2)$*

$$a^{-1}(x)W^{-1}(x) \in L(x_0, r_j) \text{ and } a^{-1}(x)W^{-1}(x)\int_{x_0}^{x} W(s)ds \in L(x_0, r_j),$$

in this case, the trajectory can enter from r_j to the interior of (r_1, r_2) but can not return from the interior to r_j;

[2] William Feller, *The parabolic differential equations and the associated semi-groups of transformations*, Annals of Mathematics, Vol **55**, No. **3** (1952) 468-519.

*(iv) The boundary r_j is **natural** in all other cases, in this case, the trajectory can neither reach r_j from the interior of (r_1, r_2) to r_j, nor from r_j to the interior.*

Some signification of these boundaries are (also see [57], Chapter 3)

1. If r_j is a regular or exit boundary, a trajectory has a probability to reach it from the interior within a finite time, thus they are also called "Accessible";
2. If r_j is a entrance or natural boundary, a trajectory will never reach it from the interior within a finite time, thus they are also called "Inaccessible";
3. If both r_1 and r_2 are inaccessible, there exists one and only one process in (r_1, r_2) obeying a given KBE.
4. If both r_1 and r_2 are accessible, there exists an "absorbing barrier process" in (r_1, r_2) obeying a given KBE. By absorbing we mean that at the first arrival at a boundary (i.e. as soon as it arrives at boundary) it is terminated.
5. Suppose that ρ_1, ρ_2 is a proper subinterval of (r_1, r_2) where r_1, r_2 are inaccessible and ρ_1, ρ_2 are regular. We can construct a solution for the absorbing barrier process in (ρ_1, ρ_2). Then as we let $\rho_1 \to r_1$ and $\rho_2 \to r_2$, the solution corresponding to the absorbing process converges to the unique solution describe in (iii).

Now, we apply the proposition to our simplest WF model. We have $a(x) = \frac{1}{2}x(1-x)$ and $b(x) = 0$, $r_1 = 0, r_2 = 1$. Then $W(x) = 1$. Obviously, at $r_1 = 0$, for all $x_0 \in (0,1)$ we have $W(x) \in L(0, x_0)$ but

$$a^{-1}(x)W^{-1}(x) = \frac{2}{x(1-x)} \notin L(0, x_0)$$

Hence $r_1 = 0$ is not regular. We also have

$$W(x)\int_{x_0}^{x} a^{-1}(s)W^{-1}(s)ds = 2\log\left(\frac{x(1-x_0)}{x_0(1-x)}\right) \in L(0, x_0).$$

Therefore $r_1 = 0$ is an exit boundary. Similarly we can show that $r_2 = 1$ is also an exit boundary.

We will prove that, the boundary of KBE corresponding to the general WF model are also exit as this following proposition

Proposition 3.7. *∂V_n is an exit boundary.*

Proof. In fact, it is easy to see that the $\mathbf{X}(t)$ is also a vector valued random variable which satisfies a stochastic dynamics

$$dX^i(t) = \sum_{j=1}^{n} \sigma_{ij}(\mathbf{X}(t))dW^j(t),$$

where

$$\sigma_{ij}(\mathbf{x}) = \sqrt{x^i}\left(\delta_{ij} - \frac{\sqrt{x^i x^j}}{1+\sqrt{x^{n+1}(t)}}\right).$$

Therefore, when $\mathbf{X}(t)$ goes to the boundary ∂V_n, for example at $X^i(t) = 0$, we have

$$dX^i(t) = \sum_{j=1}^{n} \sqrt{X^i(t)}\left(\delta_{ij} - \frac{\sqrt{X^i(t)X^j(t)}}{1+\sqrt{X^{n+1}(t)}}\right) dW^j(t) = 0,$$

i.e., $X^i(s) = 0$ for all $s \geq t$. $\square$

Chapter 4

Analytical aspects of the simplest diffusion model

In this chapter, we describe a systematic approach to solve the Kolmogorov equations of simplest diffusion models (derived from the simplest WF models). We shall define a probability density function on the entire state space $[0, 1]$ (the global solution). Qualitative results such as existence, uniqueness, behavior at boundary and behavior as $t \to \infty$ of the global solution gives us a clear picture about the evolution of initial WF model as well as other evolutionary quantities.

4.1. Motivation

In 1945, Wright[1] first used the Fokker-Planck equation to study these WF models. His approach led us to use Partial Differential Equations to solve the diffusion equations

[1]S. Wright, *The differential equation of the distribution of gene frequencies*, Proc. Nat. Acad. Sci., **31** (1945), 382-389.

corresponding to the original WF models. In 1955, Kimura[2] used Fokker-Planck equations and moments of distribution to obtain an exact local solution for the simplest diffusion model. That gave us more information about the WF model as well as the corresponding diffusion model. However, Kimura considered this continuous process locally, by dividing it into fixed probabilities and the probability distribution of unfixed classes. Therefore, his solution (the improper probability density function of unfixed classes $(0,1)$ which was assumed to be bounded at boundary) satisfies only the corresponding Fokker-Planck equation, but has no clear connection with its condition on moments. For example, we can see that the integrated probability density function on its defined domain is not equal to 1.

We will construct a new space H and a generalized integration $[\cdot,\cdot]$ with respect to a special measure μ on $([0,1],\mathcal{B}([0,1]))$. On this new space H, we define a new solution which is not necessarily bounded at the boundary but satisfies both Fokker-Planck equation and the condition on moments (4.3.7). We will then prove that there is such a unique solution (see Theorem 4.9). We will see that many results can be calculated easily and quickly due to this global solution (see Section 4.4).

4.2. Kimura's local solution to the simplest diffusion model

In 1955, Kimura found the local solution (probability density function in the unfixed class) of the problem

$$\begin{cases} \frac{\partial v(x,t)}{\partial t} & = Lv(x,t), \quad \text{in } (0,1)\times(0,\infty) \\ v(0,t), v(1,t) & < \infty \text{ for } 0 \le t < \infty, \\ v(x,0) & = \delta(x-p), \end{cases} \tag{4.2.1}$$

where

$$Lv(x,t) := \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big(x(1-x)v(x,t)\Big).$$

Theorem 4.1 (Kimura, 1955). *The problem (4.2.1) has a local solution as follows*

$$u_{loc}(x,p,t) := \sum_{m\ge 0} c_m(p)X_m(p)X_m(x)e^{-\lambda_m t},$$

where,

$$\lambda_m = \frac{(m+1)(m+2)}{2}, \quad c_m(p) = \frac{8p(1-p)(m+3/2)}{(m+1)(m+2)};$$

$X_m(x) = C_m^{(\frac{3}{2})}(1-2x)$ *is the Gegenbauer polynomial (see Appendix A.1).*

[2]Kimura M., *Solution of a Process of Random Genetic Drift with a Continuous Model*, PNAS–USA, Vol. 41, No. 3, (1955), 144-150.

Proof of Kimura. Due to the form of moments, the solution is of the form

$$u(x,t) = \sum_{m\geq 0} c_m X_m(x) e^{-\lambda_m t}$$

Then, due to the definition of the Gegenbauer polynomials and the boundedness at boundary we have

$$X_m(x) = C_m^{(\frac{3}{2})}(1-2x).$$

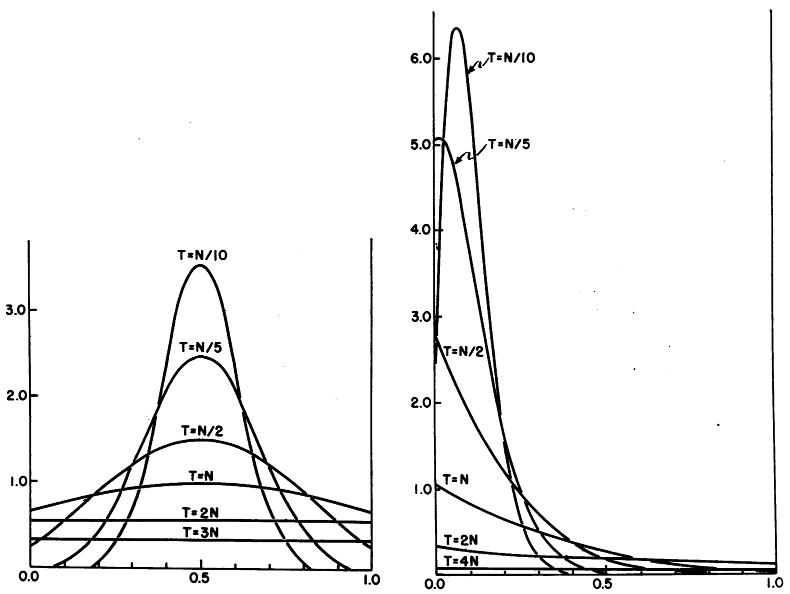
Finally, the coefficients c_m can be determined from the initial condition

$$c_m = 4pq\frac{2m+3}{(m+1)(m+2)}X_m(p).$$

Thus, the local solution is

$$u_{loc}(x,p,t) := \sum_{m\geq 0} 4pq\frac{2m+3}{(m+1)(m+2)}X_m(p)X_m(x)e^{-\lambda_m t}.$$

This completes the proof. □



Figs. 1–2.—The processes of the change in the probability distribution of heterallelic classes, due to random sampling of gametes in reproduction. It is assumed that the population starts from the gene frequency 0.5 in Fig. 1 (left) and 0.1 in Fig. 2 (right). t = time in generation; N = effective size of the population; abscissa is gene frequency; ordinate is probability density.

Figure 4.1 Kimura's description of the behavior of the local solution in the simplest diffusion model

4.3. The global solution to the simplest diffusion model

In this section, we will construct the new space of global solutions and a generalized integration. Then, we will find out eigenvalues and eigenvectors of the operators L and L^*, solve the local problem using the method of separation of variables. Next, we determine the probability density function for the simplest diffusion model on the whole state space $[0, 1]$. The existence and uniqueness result of this solution will be proved (Theorem 4.9). We will see that behavior of this global solution as its corollary.

4.3.1. Preliminaries

We denote by $H_0 = C^\infty([0, 1])$,

$$H = \Big\{u = u_1\chi_{(0,1)} + f\delta_0 + g\delta_1, \text{ for } u_1(\cdot, t), f(\cdot, t), g(\cdot, t) \in H_0\Big\}, \forall t \geq 0$$

where χ is the indicator function and δ_0, δ_1 are the Dirac functions. We define a measure μ on the measurable space $([0, 1], \mathcal{B}([0, 1]))$ as

$$\mu(A) = dx(A \cap (0, 1)) + \delta_0(A \cap \{0\}) + \delta_1(A \cap \{1\}) \quad \forall A \in \mathcal{B}([0, 1]),$$

and define the integration (the action from H into H_0) as

$$\int_A u(x)d\mu(x) = \int_{A\cap(0,1)} u_1(x)dx + \chi_A(0)f(0) + \chi_A(1)g(1) \quad \forall A \in \mathcal{B}([0, 1]).$$

We also denote by

$$[u, \phi] = \int_{[0,1]} f\phi d\mu = \int_0^1 u_1(x)\phi(x)dx + f(0)\phi(0) + g(1)\phi(1), \quad \forall u \in H, \phi \in H_0.$$

We will prove that this is well-defined, i.e. $[u, \phi] = 0\, \forall\phi$ if and only if $u = 0$. In fact, we have the following lemma.

Lemma 4.2. $[u, \phi] = 0$, $\forall\phi$ *if and only if* $u = 0$.

Proof. It is easy to see that $u = 0$ implies $[u, \phi] = 0$, $\forall\phi$. Now, we assume that $[u, \phi] = 0$, $\forall\phi$. Expanding u_1 by the $X_m(x)$ polynomials we have $u_1(x) = \sum_{m\geq 0} a_m X_m(x)$. From the orthogonality of $X_m(x)$ polynomials with the weighted function $w(x) = x(1 - x)$ we have

$$\begin{aligned} a_m &= \int_0^1 u_1(x)w(x)X_m(x)dx \\ &= [u, wX_m] = 0 \quad \forall m \end{aligned}$$

It follows that $u_1 = 0$. Then, by choosing $\phi = x$ and $\phi = 1 - x$ we have $g(1) = f(0) = 0$. Therefore we have $u = 0$. This completes the proof. ☐

Remark 4.3. We will see that the k^{th}-moment corresponding to the probability density function u will be $[u, x^k]$.

4.3.2. Eigenvalues and eigenvectors of the operators L and L^*

Below, we formulate some results on the eigenvalues and eigenvectors of the operator L and its dual L^*, which will be used in the proof of the existence and the uniqueness of the global solution.

Proposition 4.4. *For all $m \geq 0$ we have*

$$LX_m = -\lambda_m X_m, \ \ in\ H_0.$$

Proof. Putting $z = 1 - 2x$ implies $Y_m(z) = X_m(x) = C_m^{(\frac{3}{2})}(z)$, it means Y_m is a solution of the Gegenbauer equation

$$(1-z^2)\frac{\partial^2}{\partial z^2}Y_m(z) - 4z\frac{\partial}{\partial z}Y_m(z) + m(m+3)Y_m(z) = 0.$$

This is equivalent to

$$\begin{aligned} x(1-x)\frac{\partial^2}{\partial x^2}X_m(x) - 2(1-2x)\frac{\partial}{\partial x}X_m(x) + m(m+3)X_m(x) = 0, \\ \iff \frac{\partial^2}{\partial x^2}\left(x(1-x)X_m(x)\right) = -(m+1)(m+2)X_m, \\ \iff LX_m = -\lambda_m X_m. \end{aligned}$$

This completes the proof. □

Proposition 4.5. *Let $w(x) = x(1-x)$ be the weighted function. If X is an eigenvector of L corresponding to the eigenvalue λ then wX is also a eigenvector of L^* corresponding to the eigenvalue λ.*

Proof. Assume that X is an eigenvector of L corresponding to the eigenvalue λ, i.e.,

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}(wX) = -\lambda X,$$

multiplying two sides by w follows

$$L^*(wX) = \frac{1}{2}w\frac{\partial^2}{\partial x^2}(wX) = -\lambda(wX).$$

This completes the proof. □

Proposition 4.6. *The spectrum of the operator L is*

$$Spec(L) = \bigcup_{m\geq 0} \left\{ \lambda_m = \frac{(m+1)(m+2)}{2} \right\} := \Lambda,$$

and the eigenvectors of L corresponding to λ_m are the Gegenbauer polynomials $X_m(x)$ (up to a constant).

Proof. From Proposition (4.4) we have $L(X_m) = -\lambda_m X_m$ in H_0. So, $\Lambda \subseteq Spec(L)$. Inversely, for $\lambda \notin \Lambda$, we will prove that λ is not an eigenvalue of L. In fact, assume that $X \in H_0$ such that $LX = -\lambda X \in H_0$. Because $\{X_m\}_{m\geq 0}$ is a orthogonal basis of H_0 with respect to weighted w, we can write $X = \sum_{m=0}^{\infty} d_m X_m$. It follows that

$$\sum_{m=0}^{\infty} d_m(-\lambda_m X_m) = \sum_{m=0}^{\infty} d_m L(X_m) = L(\sum_{m=0}^{\infty} d_m X_m) = -\lambda \sum_{m=0}^{\infty} d_m X_m.$$

For any $n \geq 0$, multiplying them by wX_n and then integrating on $[0,1]$ we have

$$d_n \lambda_n (X_n, wX_n) = d_n \lambda (X_n, wX_n).$$

Since $(X_n, wX_n) \neq 0$ and $\lambda \neq \lambda_n$, $d_n = 0$, $\forall n \geq 0$. It follows $X = 0$, i.e., λ is not an eigenvalue of L. It follows

$$Spec(L) = \Lambda.$$

Moreover, assume that X is an eigenvector of L corresponding to the eigenvalue λ_m, we will prove that $X = cX_m$. In fact, writing $X = \sum_{n=0}^{\infty} d_n X_n$ follows that

$$\sum_{n=0}^{\infty} d_n(-\lambda_n X_n) = \sum_{n=0}^{\infty} d_n L(X_n) = L(\sum_{n=0}^{\infty} d_n X_n) = -\lambda_m \sum_{n=0}^{\infty} d_n X_n.$$

For any $k \geq 0$, multiplying them by wX_k and then integrating on $[0,1]$ we have

$$d_k \lambda_k (X_k, wX_k) = d_k \lambda_m (X_k, wX_k).$$

Since $(X_k, wX_k) \neq 0$ and $\lambda_m \neq \lambda_k$ for all $k \neq m$, $d_k = 0$, $\forall k \neq m$. It follows $X = d_m X_m$. This completes the proof. □

4.3.3. The global solution

In this subsection, we formulate the definition of solutions. We consider a diploid population of fixed size N with two possible alleles A_1, A_2 at a given locus A. Suppose that the individuals in the population are monoecious, that there is no selection between the two alleles and is no mutation. There are $2N$ alleles in the population in any generation, so it is sufficient to focus on the number Y_τ of alleles A_1 at generation time τ. Assume that $Y_0 = i_0$ and the alleles in generation $\tau + 1$ are derived by sampling with the replacement from the alleles of generation τ, it means that the transition probability is

$$\mathbb{P}(Y_{\tau+1} = j | Y_\tau = i) = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \quad \text{for } i, j = 0, \dots, 2N.$$

With the rescaled process,

$$t = \frac{\tau}{2N}; \ \ , X_t = \frac{Y_t}{2N},$$

we have a discrete Markov chain X_t taking values in $\left\{0, \frac{1}{2N}, \dots, 1\right\}$ with $t = 1$ means $2N$ generations. In Chapter 2 Proposition (2.2), we proved that

$$\begin{aligned} X_0 =& p = \frac{i_0}{2N}; \\ \mathbb{E}(\delta X_t) =& 0; \\ \mathbb{E}(\delta X_t)^2 =& X_t(1 - X_t)\delta t + o(\delta t); \\ \mathbb{E}(\delta X_t)^k =& o(\delta t) \text{ for } k \geq 3. \end{aligned} \tag{4.3.1}$$

On the other hand, denoting by $m_k(t)$ the $k^{th}-$moment of distribution about zero at the $(2Nt)^{th}$ generation, i.e.,

$$m_k(t) = \mathbb{E}(X_t)^k.$$

We have

$$m_k(t + \delta t) = \mathbb{E}(X_t + \delta X_t)^k.$$

We expand the right hand side and note (4.3.1) to obtain the following recurrence formula, under the assumption that the population number N is so large that the terms of higher order $\frac{1}{N^2}, \frac{1}{N^3}$ etc., can be neglected without serious error:

$$\begin{aligned} m_k(t + \delta t) =& \mathbb{E}(X_t + \delta X_t)^k \\ =& \mathbb{E}(X_t)^k + \binom{k}{1} \mathbb{E}\big((X_t)^{k-1} \mathbb{E}(\delta X_t)\big) + \binom{k}{2} \mathbb{E}\big((X_t)^{k-2} \mathbb{E}(\delta X_t)^2\big) \\ & + \binom{k}{3} \mathbb{E}\big((X_t)^{k-3} \mathbb{E}(\delta X_t)^3\big) + \ldots + \binom{k}{k} \mathbb{E}\big(\mathbb{E}(\delta X_t)^k\big) \\ \sim& \mathbb{E}(X_t)^k + \frac{k(k-1)}{2} \mathbb{E}\big((X_t)^{k-2} X_t(1 - X_t)(\delta t)\big) \quad \text{due to (4.3.1)} \\ =& \left\{1 - \frac{k(k-1)}{2}(\delta t)\right\} \mathbb{E}(X_t)^k + \frac{k(k-1)}{2} \mathbb{E}(X_t)^{k-1}(\delta t) \\ =& \left\{1 - \frac{k(k-1)}{2}(\delta t)\right\} m_k(t) + \frac{k(k-1)}{2} m_{k-1}(t)(\delta t). \end{aligned}$$

It follows that

$$\dot{m}_k(t) = -\frac{k(k-1)}{2}m_k(t) + \frac{k(k-1)}{2}m_{k-1}(t). \tag{4.3.2}$$

We are aiming at finding a continuous process which is a good approximation for the above discrete process. Hence, we should look for a continuous $[0,1]$-valued Markov process $\{X_t\}_{t\geq 0}$ with the same conditions as (4.3.1) and (4.3.2). Especially, if we call $u(x,t)$ the probability density function of this continuous process, the condition (4.3.1) implies (see for example [24], p. 137) that u is a solution of the Fokker-Planck (Kolmogorov forward) equation

$$\begin{cases} u_t & = Lu \text{ in } (0,1)\times(0,\infty), \\ u(x,0) & = \delta_p(x) \text{ in } (0,1); \end{cases} \tag{4.3.3}$$

and the condition (4.3.2) implies

$$\begin{aligned}
[u_t, x^k] &= \dot{m}_k(t) \\
&= -\frac{k(k-1)}{2}m_k(t) + \frac{k(k-1)}{2}m_{k-1}(t) \\
&= [u, -\frac{k(k-1)}{2}x^k + \frac{k(k-1)}{2}x^{k-1}] \\
&= [u, L^*(x^k)]
\end{aligned}$$

i.e.,

$$[u_t, \phi] = [u, L^*\phi],\ \forall \phi \in H_0. \tag{4.3.4}$$

This leads to a definition of solutions as follows

Definition 4.7. We call $u \in H$ a global solution of the Fokker-Planck equation associated to the simplest diffusion model if

$$u_t = Lu \text{ in } (0,1)\times(0,\infty), \tag{4.3.5}$$

$$u(x,0) = \delta_p(x) \text{ in } (0,1); \tag{4.3.6}$$

$$[u_t, \phi] = [u, L^*\phi],\ \forall \phi \in H_0. \tag{4.3.7}$$

Remark 4.8. We note that the conditions 4.3.6-4.3.7 imply the condition 4.3.5. But because of the usefulness in proof, we keep it as a part of the definition of the global solution.

We prove that

Theorem 4.9. *The Problem (4.3.5-4.3.7) always possesses a unique global solution.*

4.3.4. Proof of Theorem 4.9

This subsection is devoted to the construction of the global solution and proves its existence as well as uniqueness. The process of finding the solution goes as follows: First, we find the local solution of the Fokker-Planck equation (4.3.5) by the separation of variables method. Then we construct a solution due to parameters at boundaries. Furthermore, we use some conditions of (4.3.6, 4.3.7) to determine the parameters. Finally, we check the properness of the solution.

Step 1: Assume that $u_1(x,t) = X(x)T(t)$ is a solution of the Fokker-Planck equation (4.3.5). Then we have

$$\frac{T_t}{T} = \frac{LX}{X} = -\lambda.$$

Clearly λ is a constant which is independent of T, X. Proposition (4.6) implies that the local solution of the equation (4.3.5) is of form

$$u_1(x,t) = \sum_{m\geq 0} c_m X_m(x) e^{-\lambda_m t}.$$

Remark 4.10. Note that u_1 is the same solution as in Kimura (see for example [43, 44]).

Step 2: The solution $u \in H$ satisfying (4.3.5) has the following form

$$u(x,t) = \sum_{m\geq 0} c_m \left(X_m(x) + a_{m,0}\delta_0(x) + a_{m,1}\delta_1(x)\right) e^{-\lambda_m t} + b_0\delta_0(x) + b_1\delta_1(x). \qquad (4.3.8)$$

Step 3: We check condition (4.3.7) with $\phi = 1, \phi = x, \phi = wX_n$ respectively, it follows

$$[u_t, 1] = [u, L^*(1)] = 0,$$

and

$$[u_t, x] = [u, L^*(x)] = 0,$$

and

$$[u_t, wX_n] = [u, L^*(wX_n)] = -\lambda_n[u, wX_n].$$

Combining with the condition (4.3.6) follows that

$$1 = [u(\cdot,0), 1] = [u(\cdot,\infty), 1] = b_0 + b_1,$$

$$p = [u(\cdot,0), x] = [u(\cdot,\infty), x] = b_1,$$

$$1 = [u(\cdot,0), 1] = [u, 1] = \sum_{m\geq 0} c_m \left(\int\limits_{[0,1]} X_m(x)dx + a_{m,0} + a_{m,1} \right) e^{-\lambda_m t} + b_0 + b_1,$$

$$p = [u(\cdot,0), x] = [u, x] = \sum_{m\geq 0} c_m \left(\int\limits_{[0,1]} xX_m(x)dx + a_{m,1} \right) e^{-\lambda_m t} + b_1,$$

and

$$[u, wX_n] = [u(\cdot, 0), wX_n]e^{-\lambda_n t} = w(p)X_n(p)e^{-\lambda_n t}. \tag{4.3.9}$$

Therefore, we have founded all the parameters

$$\begin{aligned} b_1 &= p;\ b_0 = 1 - p \\ a_{m,1} &= -\int_{[0,1]} xX_m(x)dx;\ a_{m,0} = -\int_{[0,1]} (1-x)X_m(x)dx \\ c_n &= \frac{w(p)X_n(p)}{(X_n, wX_n)}. \end{aligned} \tag{4.3.10}$$

It follows that the solution should be of the form

$$\begin{aligned} u(x,t) = &\sum_{m\geq 0} c_m X_m(x)e^{-\lambda_m t} + \left\{ 1 - p + \sum_{m\geq 0} c_m a_{m,0} e^{-\lambda_m t} \right\} \delta_0(x) \\ &+ \left\{ p + \sum_{m\geq 0} c_m a_{m,1} e^{-\lambda_m t} \right\} \delta_1(x) \end{aligned} \tag{4.3.11}$$

where

$$\begin{aligned} \lambda_m =& \frac{(m+1)(m+2)}{2}, \\ X_m(x) =& GegenbauerC(m, 3/2, 1-2x), \\ a_{m,0} =& -\int_{\Omega_1} (1-x)X_m(x)dx = -\frac{1}{2}, \\ a_{m,1} =& -\int_{\Omega_1} xX_m(x)dx = (-1)^{m+1}\frac{1}{2}, \\ c_m =& \frac{w(p)X_m(p)}{(X_m, wX_m)} = \frac{8w(p)X_m(p)(m+3/2)}{(m+1)(m+2)}. \end{aligned} \tag{4.3.12}$$

Step 4: We will prove that the solution u found above satisfies (4.3.5, 4.3.6, and 4.3.7). In fact, $u = u_1$ in $(0, 1)$, therefore u satisfies the Fokker-Planck equation (4.3.5).

Moreover, from the representation of this new solution, we have

$$\begin{aligned} [u, 1] &= \sum_{m=0}^{\infty} \left(\int_{[0,1]} X_m(x)dx + a_{m,0} + a_{m,1} \right) e^{-\lambda_m t} + 1 - p + p = 1, \\ [u, x] &= \sum_{m=0}^{\infty} \left(\int_{[0,1]} xX_m(x)dx + a_{m,1} \right) e^{-\lambda_m t} + p = p, \\ [u, wX_n] &= c_n(X_n, wX_n)e^{-\lambda_n t} = w(p)X_n(p)e^{-\lambda_n t}. \end{aligned} \tag{4.3.13}$$

It follows

$$\begin{aligned} [u(\cdot, 0), 1] &= 1 = [\delta_p, 1] \\ [u(\cdot, 0), x] &= p = [\delta_p, x] \\ [u(\cdot, 0), wX_n] &= w(p)X_n(p) = [\delta_p, wX_n]. \end{aligned} \tag{4.3.14}$$

Because $\{1, x, \{wX_n\}_{n\geq 0}\}$ is also a basis of H_0 it follows

$$[u(\cdot, 0), \phi] = [\delta_p, \phi], \ \ \forall \phi \in H_0,$$

i.e. $u(\cdot, 0) = \delta_p \in H$, i.e. u satisfies the condition (4.3.6).

Finally, from (4.3.13) we have

$$\begin{aligned}
[u_t, 1] &= 0 = [u, L^*(1)] \\
[u_t, x] &= 0 = [u, L^*(x)] \\
[u_t, wX_n] &= w(p)X_n(p)(-\lambda_n)e^{-\lambda_n t} = -\lambda_n[u, wX_n] = [u, L^*(wX_n)].
\end{aligned} \tag{4.3.15}$$

Because L^* is linear and $\{1, x, \{wX_n\}_{n\geq 0}\}$ is also a basis of H_0 it follows

$$[u_t, \phi] = [u, L^*(\phi)], \ \ \forall \phi \in H_0,$$

i.e. u satisfies the condition (4.3.7).

Therefore, u is a solution of the Fokker-Planck equation associated with the WF model.

We can easily see that this solution is unique. In fact, assume that u_1, u_2 are two solutions of the Fokker- Planck equation associated with the WF model. Then $u = u_1 - u_2$ satisfies

$$\begin{aligned}
u_t &= Lu \text{ in } (0,1) \times (0, \infty), \\
u(x, 0) &= 0 \text{ in } (0,1); \\
[u_t, \phi] &= [u, L^*\phi], \ \forall \phi \in H_0.
\end{aligned}$$

It follows

$$\begin{aligned}
[u_t, 1] &= [u, L^*(1)] = 0, \\
[u_t, x] &= [u, L^*(x)] = 0, \\
[u_t, wX_n] &= [u, L^*(wX_n)] = -\lambda_n[u, wX_n].
\end{aligned}$$

Therefore

$$\begin{aligned}
[u, 1] &= [u(\cdot, 0), 1] = 0, \\
[u, x] &= [u(\cdot, 0), x] = 0, \\
[u, wX_n] &= [u(\cdot, 0), wX_n]e^{-\lambda_n t} = 0.
\end{aligned}$$

Because $\{1, x, \{wX_n\}_{n\geq 0}\}$ is also a basis of H_0 it follows $u = 0 \in H$.

Corollary 4.11. *1. The global solution accumulates at boundaries* $\{0, 1\}$.

2. As $t \to \infty$ *the global solution becomes flat in* $(0, 1)$ *but blows up on the boundaries and*

$$\lim_{t \to \infty} u(x, t) = (1 - p)\delta_0(x) + p\delta_1(x).$$

Remark 4.12. It is easy to see that

$$f(0,t) = \frac{1}{2}\int_0^t u_1(0,s)ds$$

and

$$g(1,t) = \frac{1}{2}\int_0^t u_1(1,s)ds$$

These show that values at boundaries can be calculated by the boundary flux, which will be made precise in Chapter 5.

The above solution is a continuous deformation from $\delta_p(x)$ to $p\delta_1(x) + (1-p)\delta_0(x)$ in time from 0 to ∞. For having a imagination of behavior of the new solution in time t we construct a sequence $\{u_m(x,t)\}_{m\geq 0}$ satisfying

$$\begin{aligned} u_m(x,t) = &\sum_{i=0}^{m} c_i X_i(x) e^{-\lambda_i t} \\ &+ \left\{1 - p + \sum_{i=0}^{m} c_i \left(-\frac{1}{2}\right) e^{-\lambda_i t}\right\} \frac{m}{\sqrt{2\pi}} e^{-x^2 m^2/2} \\ &+ \left\{p + \sum_{i=0}^{m} c_i \left(\frac{(-1)^{i+1}}{2}\right) e^{-\lambda_i t}\right\} \frac{m}{\sqrt{2\pi}} e^{-(1-x)^2 m^2/2} \end{aligned} \tag{4.3.16}$$

It is easy to prove that u_m pointwisely converges to u as $m \to \infty$. Therefore, we can figure this behavior by, for example, Mathematica as follows

```
w[x_]:= x (1 - x)
X[m_, x_]:= GegenbauerC[m, 3/2, 1 - 2 x]
c[m_]:= 8 w[p] X[m, p] (m + 3/2)/{(m + 1) (m + 2)}
ld[m_]:= (m + 1) (m + 2)/2
u1[m_,x_,t_]:=Sum[c[i] X[i, x] Exp[-(i+1)(i+2) t/2],{i,0,m}]
f[n_, x_] := n/Sqrt[2 Pi] Exp[-x^2 n^2/2]
g[n_, x_] := n/Sqrt[2 Pi] Exp[-(x - 1)^2 n^2/2]
u2[m_,x_,t_]:={1-p+Sum[c[i](-1/2)Exp[-(i+1)(i+2) t/2],{i,0,m}]}f[m, x]
u3[m_,x_,t_]:={p+Sum[c[i](-1)^{i+1}1/2 Exp[-(i+1)(i+2)t/2],{i,0,m}]}g[m,x]
u[m_, x_, t_] := u1[m, x, t] + u2[m, x, t] + u3[m, x, t]
Plot3D[u[50, x, t] /. p -> .4, {x, 0, 1}, {t, .01, 20}]
```

4.4. Applications of the global solution

In this section we spell out some applications of this global solution. With the solution received, we can immediately obtain almost all information of the evolution of the

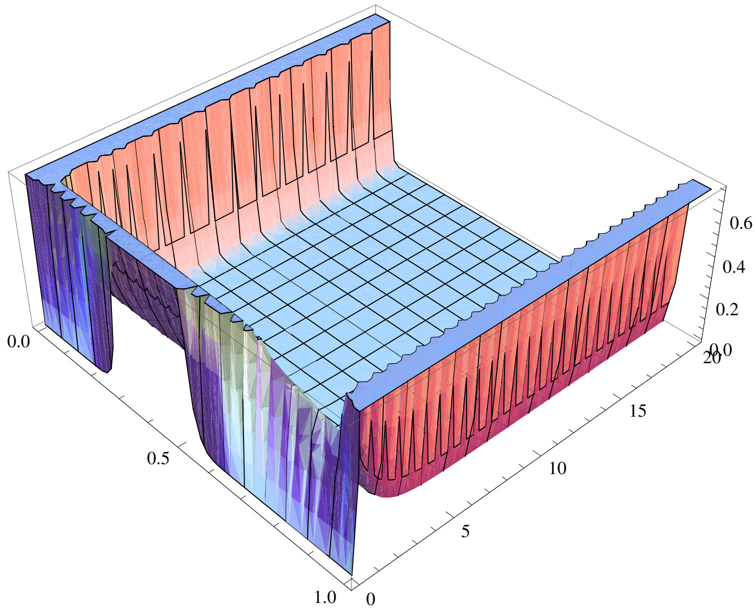


Figure 4.2 Behavior of the global solution from δ_p to $p\delta_1 + (1-p)\delta_0$ in time with $p = 0.4$.

process $(X_t)_{t\geq 0}$ such as probabilities of fixation, probability of coexistence, probability of heterogeneity, $k^{th}-$moments, the absorption time as well as the distribution at absorption time. We detail them here.

4.4.1. Probabilities of fixation and Probability of coexistence of two alleles

The probability of fixation of A_2 allele (the loss probability of A_1 allele) is

$$\begin{aligned}\mathbb{P}(X_t = 0|X_0 = p) &= \int\limits_{\{0\}} u(x,t)d\mu(x) = u_0(0,p,t) \\ &= 1 - p + \sum_{m=0}^{\infty} c_m a_{m,0} e^{-\lambda_m t} \\ &= 1 - p - \frac{1}{2}\sum_{m=0}^{\infty} \frac{8w(p)X_m(p)(m+3/2)}{(m+1)(m+2)} e^{-\lambda_m t}.\end{aligned}$$

The probability of fixation of A_1 allele (the lost probability of A_2 allele) is

$$\begin{aligned}\mathbb{P}(X_t = 1|X_0 = p) &= \int\limits_{\{1\}} u(x,t)d\mu(x) = u_0(1,p,t) \\ &= p + \sum_{m=0}^{\infty} c_m a_{m,1} e^{-\lambda_m t} \\ &= p - \frac{1}{2}\sum_{m=0}^{\infty} (-1)^m \frac{8w(p)X_m(p)(m+3/2)}{(m+1)(m+2)} e^{-\lambda_m t}.\end{aligned}$$

The probability of coexistence of two alleles A_1, A_2 is

$$\begin{aligned}
\mathbb{P}(X_t \in (0,1)|X_0 = p) &= \int\limits_{(0,1)} u(x,t)d\mu(x) \\
&= \sum_{m=0}^{\infty} c_m \int\limits_0^1 X_m(x)dx e^{-\lambda_m t} \\
&= \sum_{m=0}^{\infty} c_{2m} e^{-\lambda_{2m} t} \\
&= \sum_{m=0}^{\infty} \frac{8w(p)X_{2m}(p)(2m+3/2)}{(2m+1)(2m+2)} e^{-\lambda_{2m} t}.
\end{aligned}$$

Remark 4.13. We note that after a large enough number of generation ($t \to \infty$), one allele will disappear, the other will remain fixed, and the population becomes homozygous with $\mathbb{P}(X_\infty = 1|X_0 = p) = p$ and $\mathbb{P}(X_\infty = 0|X_0 = p) = 1 - p$. Below, we will prove that, this phenomenon of homozygosity happens in a finite number of generations.

For more detail, we consider the behavior of these probabilities with different initial frequency distributions $p = 0.3$ and $p = 0.5$ as follows

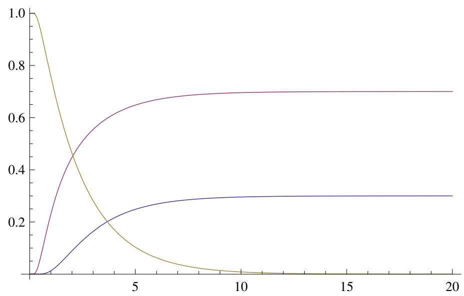


Figure 4.3 Behavior of the probability of fixation and coexistence with $p = 0.3$

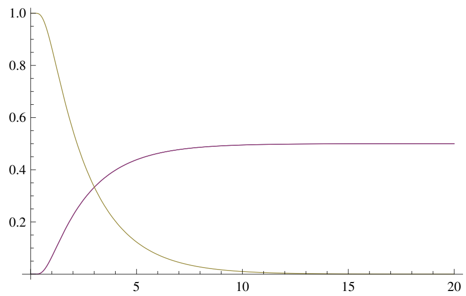


Figure 4.4 Behavior of the probability of fixation and coexistence with $p = 0.5$

Remark 4.14. (i) $\mathbb{P}(X_t \in [0,1]|X_0 = p) = \mathbb{P}(X_t = 0|X_0 = p) + \mathbb{P}(X_t = 1|X_0 = p) + \mathbb{P}(X_t \in (0,1)|X_0 = p) = 1$;

(ii) $\mathbb{P}(X_t = 0|X_0 = p)$ increase quickly in $t \in (0,5)$ ($10N$ generations) from 0 and then asymptotically slowly to $1 - p$;

(iii) $\mathbb{P}(X_t = 1|X_0 = p)$ increase quickly in $t \in (0,5)$ ($10N$ generations) from 0 and then asymptotically slowly to p;

(iv) When $p = 0.5$, $\mathbb{P}(X_t = 0|X_0 = p) = \mathbb{P}(X_t = 1|X_0 = p)$.

4.4.2. Loss of heterozygosity

The probability of heterozygosity is

$$\begin{aligned}
H_t &= \int\limits_{[0,1]} 2x(1-x)u(x,t)d\mu(x) \\
&= 2(u_{loc}, wX_0) \quad \text{because } w \text{ vanishes on boundary} \\
&= 2\Big(\sum_{m\geq 0} c_m X_m e^{-\lambda_m t}, wX_0\Big) \\
&= 2(c_0 X_0 e^{-\lambda_0 t}, wX_0) \quad \text{because } X_m \text{ are orthogonal} \\
&= 2w(p)X_0(p)e^{-t} \\
&= H_0 e^{-t}.
\end{aligned}$$

This shows that heterozygosity decreases exactly at the rate $\frac{1}{2N}$ per generation.

4.4.3. k^{th}−moments

By induction, it is easy to prove that

$$\int\limits_0^1 x^k X_{m-1}(x)d\mu(x) = (-1)^m \frac{1}{2}\left\{\frac{(k-1)\dots(k-m)}{(n+1)\dots(k+m)} - 1\right\}.$$

Therefore, the k^{th}−moment is

$$\begin{aligned}
m_k(t) &= [u, x^k]_1 \\
&= \sum_{m=0}^{\infty} c_m \left(\int\limits_0^1 x^k X_m(x)dx\right) e^{-\lambda_m t} + \left(p + \sum_{m=0}^{\infty} c_m a_{m,1} e^{-\lambda_m t}\right) \\
&= p + \sum_{i=1}^{\infty} c_{i-1}\left(\int\limits_0^1 x^k X_{i-1}(x)dx + a_{i-1,1}\right) e^{-\lambda_{i-1} t} \\
&= p + \sum_{i=1}^{\infty} \frac{2(2i+1)}{i(i+1)} p(1-p)(-1)^i X_{i-1}(p) \frac{(k-1)\dots(k-i)}{(k+1)\dots(k+i)} e^{-\frac{i(i+1)}{2}t}.
\end{aligned}$$

This result of the k^{th}−moment is consistent with the Kimura's one in ([43]).

4.4.4. Absorption time and the distribution at absorption time

We denote by $T_2^1(p) = \inf\{t > 0 : X_t \in V_0 | X_0 = p\}$ the first time of population having 1 allele. $T_2^1(p)$ is a continuous random variable valued in $[0,\infty)$ and we call $\phi(t,p)$ its probability density function. It is easy to see that V_0 is invariant (absorption set) under the process X_t, i.e. if $X_s \in V_0$ then $X_t \in V_0$ for all $t \geq s$. We have the identity

$$\mathbb{P}(T_2^1(p) \leq t) = \mathbb{P}(X_t \in V_0 | X_0 = p) = u_0(0,p,t) + u_0(1,p,t).$$

It follows that

$$\phi(t,p) = \frac{\partial}{\partial t}\Big(u_0(0,p,t) + u_0(1,p,t)\Big).$$

Therefore the expectation of absorption time having 1 allele is

$$\begin{aligned}
\mathbb{E}(T_2^1(p)) &= \int_0^\infty t\phi(t,p)dt \\
&= \int_0^\infty t\frac{\partial}{\partial t}\Big(u_0(0,p,t) + u_0(1,p,t)\Big)dt \\
&= \sum_{m=0}^\infty (-\lambda_m)c_m\Big(a_{m,0} + a_{m,1}\Big)\left(\int_0^\infty te^{-\lambda_m t}dt\right) \\
&= -\sum_{m=0}^\infty \frac{1}{\lambda_m}c_m\left(a_{m,0} + a_{m,1}\right) \\
&= \sum_{m=0}^\infty \frac{1}{\lambda_{2m}}c_{2m} \\
&= 16p(1-p)\sum_{m=0}^\infty \frac{(2m+3/2)}{(2m+1)^2(2m+2)^2}X_{2m}(p).
\end{aligned}$$

and the second moment of absorption time having 1 allele is

$$\begin{aligned}
\mathbb{E}(T_2^1(p))^2 &= \int_0^\infty t^2\phi(t,p)dt \\
&= \int_0^\infty t^2\frac{\partial}{\partial t}\Big(u_0(0,p,t) + u_0(1,p,t)\Big)dt \\
&= \sum_{m=0}^\infty (-\lambda_m)c_m\Big(a_{m,0} + a_{m,1}\Big)\left(\int_0^\infty t^2e^{-\lambda_m t}dt\right) \\
&= -\sum_{m=0}^\infty \frac{2}{\lambda_m^2}c_m\left(a_{m,0} + a_{m,1}\right) \\
&= \sum_{m=0}^\infty \frac{2}{\lambda_{2m}^2}c_{2m} \\
&= 64p(1-p)\sum_{m=0}^\infty \frac{(2m+3/2)}{(2m+1)^3(2m+2)^3}X_{2m}(p)
\end{aligned} \tag{4.4.1}$$

We also know that $\mathbb{E}(T_2^1(p))$ is the unique solution of the following boundary value problem in one dimension:

$$\begin{cases} L_1^* v &= -1, \text{ in } (0,1) \\ v(0) &= v(1) = 0. \end{cases}$$

This can be solved easily to obtain

$$\mathbb{E}(T_2^1(p)) = -2\left\{pln(p) + (1-p)ln(1-p)\right\}.$$

It also show that the phenomenon of homozygosity happen in a finite number of generations.

Remark 4.15. To check that two expectation results are the same, we use Mathematica as follows

```
X[m_,x_]:=GegenbauerC[m,3/2,1-2 x]
c[m_,p_]:=16 p(1-p)(2 m +3/2)/{(2 m+1)^2(2 m+2)^2}
f[p_] := -2 p Log[p] - 2 (1 - p) Log[1 - p]
Plot[{Sum[c[m,p] X[2 m,p],{m,0,200}],f[p]},{p,0.1,0.9},AxesOrigin->{0,0}]
```

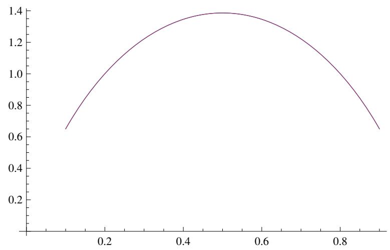


Figure 4.5 Comparison results of expectation of the absorption time.

We note that $X_{T_2^1(p)}$ is a $\{0,1\}$ valued random variable. Due to the attraction of boundary $\{0,1\}$, it is easy to show that its distribution is the same as the distribution of X_∞:

$$\mathbb{P}(X_{T_2^1(p)} = 1) = \mathbb{P}(X_\infty = 1|X_0 = p) = p,$$

and

$$\mathbb{P}(X_{T_2^1(p)} = 0) = \mathbb{P}(X_\infty = 0|X_0 = p) = 1 - p.$$

Chapter 5

Analytical aspects of the general diffusion model

In this chapter, we systematically consider methods for solving the Kolmogorov equations of the general diffusion models, derived from general WF models. We shall construct a proper probability density function on S_n (the global solution). Results are also given for a wide range of other quantities of interest.

5.1. Motivation

Until the publication of the paper by Kimura in 1955 ([43]), the theoretical treatment of the problem of the random genetic drift was confined so far to the case of two-allele. This restriction, however, seems to be unnatural and undesirable in some cases.

In 1955, Kimura[1] found out the asymptotic local solution for n alleles at time having exact k alleles. In 1956, he found out[2] the exact local solution for 3-alleles, however his method is not easy to generalize to the case of n-alleles.

By using a system of polynomials bi-orthogonal to Appel's polynomials, in 1975, Littler and Fackerell[3] solved locally the problem of KBE to the general WF model.

Another approach is due to Baxter, Blythe, McKane[4] in 2007 by using the method of change of variables to KFE. They obtained the local solution as product of Jacobi polynomials.

From the essential properties of the diffusion process, we define a new solution (global solution) which is not necessary bounded at boundary but satisfies both Fokker-Planck equation and the condition of moments (5.3.10) ([65]). With this new definition, we can prove that there is such a unique solution (see Theorem 5.19). J. Hofrichter's hierarchichal product and boundary then allow the more precise definition of the global solution and yield a more transparent proof (Definition 5.7 and Section 5.3.5.). Finally, we can easily calculate information of the continuous process due to this global solution.

One interesting property is that some statistical quantities of interest are solutions of a singular elliptic second order linear equation with discontinuous (or incomplete) boundary values. A lot of papers, textbooks have used this property to find those quantities ([16, 24]). However, the uniqueness of these problems have not been proved. Littler, in his PhD thesis in 1975 ([52]), took up the uniqueness problem but his proof, in my view, is not rigorous. We shall formulate two different ways to prove the uniqueness rigorously. The first way is the approximation method (Lemma 5.25). The second way is the blow-up method which is conducted by J. Hofrichter ([34]).

[1] Kimura M., *Random genetic drift in multi-allele locus*, Evolution, **9** (1955), 419-435.

[2] Kimura M., *Random genetic drift in a triallelic locus; exact solution with a continuous model*, Biometrics, **12** (1956), 57-66.

[3] R.A.Littler, E.D.Fackerell, *Transition Densities for Neutral Multi-Allele Diffusion Models*, Biometrics, Vol. 31, No. **1**, (1975) 117-123.

[4] G. J. Baxter, R. A. Blythe, A. J. McKane, *Exact solution of the multi-allelic diffusion model*, Math. Biosciences **209** (2007) 124-170.

5.2. The local solution to the general diffusion model

In this section, we review some methods used to find out the local solutions of the Kolmogorov equations derived from the general diffusion model. These local solutions defined until the first absorption time are "improper" probability density functions of a diffusion process on V_n.

The Kolmogorov forward equation (Fokker-Planck equation) reads as

$$\frac{\partial u(\mathbf{x},t)}{\partial t} = \sum_{i,j=1}^{n} \frac{1}{2} \frac{\partial^2}{\partial x^i \partial x^j} \Big(x^i(\delta_{ij} - x^j) u(\mathbf{x},t) \Big), \quad \mathbf{x} \in V_n,\ 0 < t < \infty, \tag{5.2.1}$$

with the initial condition

$$u(\mathbf{x},0) = \delta_{\mathbf{p}}(\mathbf{x})$$

and the bounded boundary condition.

The Kolmogorov backward equation is of the form

$$\frac{\partial u(\mathbf{p},t)}{\partial t} = \sum_{i,j=1}^{n} \frac{1}{2} p^i(\delta_{ij} - p^j) \frac{\partial^2 u(\mathbf{p},t)}{\partial p^i \partial p^j}, \quad \mathbf{p} \in V_n,\ 0 < t < \infty, \tag{5.2.2}$$

with the initial condition

$$u(\mathbf{p},0) = \delta_{\mathbf{x}}(\mathbf{p})$$

and the vanishing boundary condition.

5.2.1. Kimura's local solutions

In 1955, Kimura found out an asymptotic estimate for local solution. Moreover, he proved the following important asymptotic result:

Theorem 5.1 (Kimura, 1955)**.** *Assume we start from a population which contains m alleles, say $A_1, A_2, \cdots, A_m$ with frequencies $p^1, p^2, \cdots, p^m$ respectively $\Big(\sum_{i=1}^{m} p^i = 1\Big)$, the probability density that it contains k of them, say $A_1, A_2, \cdots, A_k$ with respective frequencies $x^1, x^2, \cdots, x^k$ $\Big(\sum_{i=1}^{k} x^i = 1\Big)$ in the t^{th} generation is given asymptotically by*

$$\phi_{1,2,\ldots,k}(x^1, x^2, \cdots, x^{k-1} | p^1, p^2, \cdots, p^{m-1}, t) \sim (2k-1)! \left(\prod_{j=1}^{k} p^{i_j} \right) e^{-\frac{k(k-1)t}{4N}},$$

where $k \le m$. In the case $k = m = n+1$ we have the local asymptotic solution

$$u_{loc}(\mathbf{x}|\mathbf{p},t) \sim (2n+1)! \left(\prod_{j=1}^{n+1} p^j \right) e^{-\frac{n(n+1)t}{4N}},$$

Remark 5.2. This indicates that as the number of coexisting alleles increases, the rate by which that state is eliminated increases with rapid progression. In this sense, random drift might be effective in keeping down the number of coexisting alleles in the population.

Next, in 1956, Kimura gave out the local exact solution for three alleles by using a transformation of variables

$$x = \rho(1-\xi) \quad \text{and } y = \rho\xi \qquad (0 \leq \rho, \xi \leq 1)$$

which makes the problem KFE become separable.

Theorem 5.3 (Kimura, 1956). *In the case of three alleles, the local solution can be represented by*

$$\begin{aligned} u_{loc}(x, y|p, q, t) &= \sum_{m\geq 0}\sum_{j\geq 0} c(m,j)(1-z)^m T_m^1\left(\frac{x-y}{1-z}\right) \\ &\quad \times J_j(2m+5, 2m+4, 1-z)e^{-\frac{(j+m+2)(j+m+3)}{4N}t}, \end{aligned}$$

where

$$z = 1-x-y, \quad T_m^1, J_j \textit{ are Gegenbauer and Jacobi polynomials},$$

and

$$\begin{aligned} C(m,j) &= \frac{4(j+2m+3)!(j+2m+4)!(2j+2m+5)}{j!(j+1)!(m+1)(m+2)(2m+2)!(2m+3)!} pqr(1-r)^m \\ &\times \quad T_m^1\left(\frac{p-q}{1-r}\right) J_j(2m+5, 2m+4, 1-r), \end{aligned}$$

with $r = 1-p-q$.

Remark 5.4. He stated in this paper that his transformation can solve the general case.

5.2.2. Littler and Fackerell's local solutions

In 1975, Littler and Fackerell solved the problem of KBE.

$$\begin{cases} \frac{\partial u(\mathbf{x},\mathbf{p},t)}{\partial t} &= L_n^* u(\mathbf{x},\mathbf{p},t), \quad \text{in } V_n, \\ u(\mathbf{x},\mathbf{p},t) &= 0 \text{ for } \mathbf{p} \in \partial V_n, \\ u(\mathbf{x},\mathbf{p},0) &= \prod_{i=1}^n \delta(x^i - p^i). \end{cases} \tag{5.2.3}$$

Theorem 5.5 (Littler & Fackerell, 1975). *The problem (5.2.3) has a unique local solution given by*

$$u_{loc}(\mathbf{x},\mathbf{p},t) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^{n-1}} c_{n,\boldsymbol{\alpha}} p^1 \dots p^n \times \mathcal{F}_{\boldsymbol{\alpha}}(2n-1;2,\dots,2;\mathbf{p})$$
$$\times \mathcal{E}_{\boldsymbol{\alpha}}(2n-1;2,\dots,2;\mathbf{x}) e^{-\frac{1}{2}(|\boldsymbol{\alpha}|+n-1)(|\boldsymbol{\alpha}|+n)t},$$

where

$$c_{n,\boldsymbol{\alpha}} = ((2n-1)+2|\boldsymbol{\alpha}|)(\alpha^1+1)\dots(\alpha^{n-1}+1)(|\boldsymbol{\alpha}|+2)_{2n-3},$$

and $\mathcal{F}_{\boldsymbol{\alpha}}, \mathcal{E}_{\boldsymbol{\alpha}}$ *are a bi-orthogonal system on* V_{n-1}

On V_k, we may find a bi-orthogonal system for the weighted function

$$w(x^1,\dots,x^k) = (x^1)^{\gamma^1-1}\dots(x^k)^{\gamma^k-1}(1-x^1-\dots-x^k)^{\alpha-\gamma^1-\dots-\gamma^k}$$

by defining

$$\mathcal{F}_{m^1,\dots,m^k}(\alpha,\gamma^1,\dots,\gamma^k;x^1,\dots,x^k) = \frac{[w(x^1,\dots,x^k)]^{-1}}{(\gamma^1)_{m^1}\dots(\gamma^k)_{m^k}}\times$$
$$\frac{\partial^{m^1+\dots+m^k}}{\partial(x^1)^{m^1}\dots\partial(x^k)^{m^k}}\left\{w(x^1,\dots,x^k)(x^1)^{m^1}\dots(x^k)^{m^k}(1-x^1-\dots-x^k)^{m^1+\dots+m^k}\right\}$$

and

$$\mathcal{E}_{m^1,\dots,m^k}(\alpha,\gamma^1,\dots,\gamma^k;x^1,\dots,x^k) =$$
$$F_A(\alpha+m^1+\dots+m^k,-m^1,\dots,-m^k,\gamma^1,\dots,\gamma^k;x^1,\dots,x^k),$$

is the Lauricella's generalized hypergeometric function.

5.2.3. Baxter, Blythe, and McKane's local solutions

In 2007, Baxter, Blythe, and McKane ([16]) used the method of change of variables to find out the local solutions. This method makes the equation separable, and therefore reduces the local problem with an arbitrary number of alleles to the two-allele case. One advantage of this change is that it can be applied to the Kolmogorov equation with the effect of mutation satisfying the condition that the mutation matrix has equal entries in each row.

The change of variables are as follows

$$y^1 = x^1, \qquad y^i = \frac{x^i}{1-\sum\limits_{j<i} x^j}, \quad i = 2,\cdots,n-1, \tag{5.2.4}$$

with the inverse transformation

$$x^1 = y^1, \qquad x^i = y^i \prod_{j<i}(1-x^i), \quad i = 2,\cdots,n-1. \tag{5.2.5}$$

Theorem 5.6 (Baxter, Blythe & McKane, 2007). *The problem to KFE has a unique local solution given by*

$$u_{loc}(\mathbf{x}, \mathbf{p}, t) = \prod_{i=1}^{n-1} (1 - x^1 - \cdots - x^i)^{-1} v(\mathbf{y}, \mathbf{q}, t),$$

where $\mathbf{y}, \mathbf{q}$ *are changed variables of* $\mathbf{x}, \mathbf{p}$ *after the transformation of coordinates, and* v *is the new probability density function with respect to these new coordinates.*

$$v(\mathbf{y}, \mathbf{q}, t) = \sum_{\lambda^{(n)}} w(\mathbf{q}) \phi_{\lambda^{(n)}}(\mathbf{q}) \phi_{\lambda^{(n)}}(\mathbf{y}) e^{-\lambda^{(n)} t}$$

where the sum is for every $(l_1, \ldots, l_n) \in \mathbb{N}_0^n$

$L_i = \sum_{j \leq i} (l_j + 1)$

$$\lambda^{(i)} = \frac{L_i(L_i + 1)}{2}$$

$$\begin{aligned} \phi_{\lambda^{(n)}}(\mathbf{y}) &= \prod_{i=1}^{n} \psi_{\lambda^{(n+1-i)}, \lambda^{(n-i)}}(y^i) \\ &= \prod_{i=1}^{n} (1 - y^i)^{L_{i-1}} P_l^{(1, 2\gamma_i + 1)}(1 - 2y^i) \end{aligned}$$

5.3. The global solution to the general diffusion model

In this section, we construct the global solution of the general WF model. The existence and uniqueness of this solution will be proved (Theorem 5.19). From this solution, we will have a lot of information of the considered process.

5.3.1. Preliminaries

This subsection is devoted to construct the space on which the global solutions lie and the way to integrate these solutions.

We denote by $H_n = C^\infty(\overline{V_n}) = C^\infty(S_n)$,

$$H = \left\{ u = \sum_{k=0}^{n} u_k V_k = \sum_{k=0}^{n} \sum_{(i_0, \ldots, i_k) \in I_k} u_k^{(i_0, \ldots, i_k)} \chi_{V_k^{(i_0, \ldots, i_k)}}, \text{ for } u_k^{(i_0, \ldots, i_k)}(\cdot, t) \in H_n \right\}$$

We define a measure μ on the measurable space $(S_n, \mathcal{B}(S_n))$ and the integration as follows

$$\int_A u(\mathbf{x}, t) d\mu(\mathbf{x}) = \sum_{k=1}^{n} \sum_{(i_0, \ldots, i_k) \in I_k} \int_{A \cap V_k^{(i_0, \ldots, i_k)}} u_k^{(i_0, \ldots, i_k)}(\mathbf{x}, t) d\mu_k(\mathbf{x}) + \sum_{i=0}^{n} \chi_{A \cap e^i} u_0(e^i, t),$$

where A is a Borel subset in S_n. We also denote by

Definition 5.7. The global solution will be defined by the hierarchical product, which is given by J. Hofrichter ([34], Chapter 4).

$$\begin{aligned}[u,\phi]_n &= \int_{S_n} u(\mathbf{x},t)d\mu(\mathbf{x})\\ &= \sum_{k=1}^{n}\sum_{(i_0,\dots,i_k)\in I_k}\int_{V_k^{(i_0,\dots,i_k)}} u_k^{(i_0,\dots,i_k)}(\mathbf{x},t)\phi_{|V_k^{(i_0,\dots,i_k)}}(\mathbf{x})d\mu_k(\mathbf{x}) + \sum_{i=0}^{n} u_0(e^i,t)\phi(e^i).\end{aligned}$$

We will prove this is a well-defined, i.e. $[u,\phi]_n = 0$, $\forall\phi$ if and only if $u = 0$. In fact, we have the following lemma

Lemma 5.8. *$[u,\phi]_n = 0$, $\forall\phi$ if and only if $u = 0$.*

Proof. Do the same as in Chapter 4, using the complete basis of eigenvectors we will have $u_k = 0$ with k decreases from n to 0. □

Therefore we can define the integration

$$\int_A u(\mathbf{x})\phi(\mathbf{x})\mu(d\mathbf{x}) := (u,\chi_A\phi) = \sum_{k=0}^{n}\sum_{(i_0,\dots,i_k)\in I_k}\int_{A\cap V_k^{(i_0,\dots,i_k)}} u_k^{(i_0,\dots,i_k)}(\mathbf{x})\phi(x)\mu(d\mathbf{x}),$$

where A is a Borel subset in Ω.

Remark 5.9. We will see that the α^{th}-moment corresponding to the probability density function u will be $[u,\mathbf{x}^\alpha]_n$.

5.3.2. Eigenvalues and eigenvectors of operators L_n and L_n^*

Below, we formulate some facts on eigenvalues and eigenvectors of the operator L_n and its dual L_n^*, which will be used in the proof of the existence and uniqueness of the global solution.

Proposition 5.10. *For each $1 \le k \le n, m \ge 0, |\alpha| = \alpha^1 + \dots + \alpha^k = m$, the $m-$degree polynomial of $k-$variables $x = (x^{i_1},\dots,x^{i_k})$ in $\overline{V_k^{(i_0,\dots,i_k)}}$*

$$X_{m,\alpha}^{(k)}(x) = x^\alpha + \sum_{|\beta|<m} a_{m,\beta}^{(k)} x^\beta, \tag{5.3.1}$$

where $a_{m,\beta}^{(k)}$ is defined inductively as

$$a_{m,\beta}^{(k)} = -\frac{\sum_{i=1}^{k}(\beta_i+2)(\beta_i+1)a_{m,\beta+e_i}^{(k)}}{(m-|\beta|)(m+\beta+2k+1)}, \forall|\beta|<m,$$

is the eigenvector of $L_k^{(i_0,\dots,i_k)}$ corresponding to the eigenvalue $\lambda_m^{(k)}$.

Proof. This is easy to see by putting (5.3.1) into

$$L_k^{(i_0,\ldots,i_k)} X_{m,\alpha}^{(k)}(x) = -\lambda_m^{(k)} X_{m,\alpha}^{(k)}(x),$$

and equalization of coefficients. □

Remark 5.11. When $k = 1$, it is easy to see that $X_{m,m}^{(1)}(x^1)$ is the $m^{th}-$Gegenbauer polynomial (up to a constant). Thus, these polynomials can be understood as a generalization of the Gegenbauer polynomials to higher dimensions.

Proposition 5.12. *Denote $w_k^{(i_0,\ldots,i_k)} = x^{i_0} x^{i_1} \cdots x^{i_k}$ by the weighted functions in the space $V_k^{(i_0,\ldots,i_k)}$. Then, if $X \in \overline{V_k^{(i_0,\ldots,i_k)}}$ is an eigenvector of $L_k^{(i_0,\ldots,i_k)}$ corresponding to λ then $w_k^{(i_0,\ldots,i_k)} X$ is also an eigenvector of $(L_k^{(i_0,\ldots,i_k)})^*$ corresponding to λ.*

Proof. If $X \in \overline{V_k^{(i_0,\ldots,i_k)}}$ is an eigenvector of $L_k^{(i_0,\ldots,i_k)}$ corresponding to λ, it follows

$$
\begin{aligned}
-\lambda(w_k^{(i_0,\ldots,i_k)}(x)X) =& \frac{1}{2} w_k^{(i_0,\ldots,i_k)}(x) \sum_{i,j\in\{i_1,\ldots,i_k\}} \frac{\partial^2}{\partial x^i \partial x^j} \left(x^i(\delta_{ij} - x^j)X\right) \\
=& \frac{1}{2} w_k^{(i_0,\ldots,i_k)}(x) \sum_{i,j\in\{i_1,\ldots,i_k\}} \left(x^i(\delta_{ij} - x^j)\right) \frac{\partial^2 X}{\partial x^i \partial x^j} \\
&+ \frac{1}{2} w_k^{(i_0,\ldots,i_k)}(x) \sum_{i,j\in\{i_1,\ldots,i_k\}} \frac{\partial \left(x^i(\delta_{ij} - x^j)\right)}{\partial x^i} \frac{\partial X}{\partial x^j} \\
&+ \frac{1}{2} w_k^{(i_0,\ldots,i_k)}(x) \sum_{i,j=1}^{k} \frac{\partial \left(x^i(\delta_{ij} - x^j)\right)}{\partial x^j} \frac{\partial X}{\partial x^i} \\
&+ \frac{1}{2} w_k^{(i_0,\ldots,i_k)}(x) \sum_{i,j\in\{i_1,\ldots,i_k\}} \frac{\partial^2 \left(x^i(\delta_{ij} - x^j)\right)}{\partial x^i \partial x^j} X \\
=& \frac{1}{2} \sum_{i,j=1}^{k} \left(x^i(\delta_{ij} - x^j)\right) \left(w_k^{(i_0,\ldots,i_k)}(x) \frac{\partial^2 X}{\partial x^i \partial x^j} \right) \\
&+ \frac{1}{2} \sum_{j\in\{i_1,\ldots,i_k\}} w_k^{(i_0,\ldots,i_k)}(x) \left(1 - (k-1)x^j\right) \frac{\partial X}{\partial x^j} \\
&+ \frac{1}{2} \sum_{i\in\{i_1,\ldots,i_k\}} w_k^{(i_0,\ldots,i_k)}(x) \left(1 - (k-1)x^i\right) \frac{\partial X}{\partial x^i} \\
&- \frac{k(k+1)}{2} w_k^{(i_0,\ldots,i_k)}(x) X \\
=& \frac{1}{2} \sum_{i,j\in\{i_1,\ldots,i_k\}} \left(x^i(\delta_{ij} - x^j)\right) \left(w_k^{(i_0,\ldots,i_k)}(x) \frac{\partial^2 X}{\partial x^i \partial x^j} \right) \\
&+ \frac{1}{2} \sum_{i,j\in\{i_1,\ldots,i_k\}} \left(x^i(\delta_{ij} - x^j)\right) \frac{\partial w_k^{(i_0,\ldots,i_k)}(x)}{\partial x^i} \frac{\partial X}{\partial x^j} \\
&+ \frac{1}{2} \sum_{i,j\in\{i_1,\ldots,i_k\}} \left(x^i(\delta_{ij} - x^j)\right) \frac{\partial w_k^{(i_0,\ldots,i_k)}(x)}{\partial x^j} \frac{\partial X}{\partial x^i} \\
&+ \frac{1}{2} \sum_{i,j\in\{i_1,\ldots,i_k\}} \left(x^i(\delta_{ij} - x^j)\right) \frac{\partial^2 w_k^{(i_0,\ldots,i_k)}(x)}{\partial x^i \partial x^j} X \\
=& \frac{1}{2} \sum_{i,j\in\{i_1,\ldots,i_k\}} \left(x^i(\delta_{ij} - x^j)\right) \frac{\partial^2 (w_k^{(i_0,\ldots,i_k)} X)(x)}{\partial x^i \partial x^j} \\
=& (L_k^{(i_0,\ldots,i_k)})^* (w_k^{(i_0,\ldots,i_k)}(x) X).
\end{aligned}
$$

This completes the proof. □

Proposition 5.13. *Denoting ν by the exterior unit normal vector of the domain $V_k^{(i_0,\ldots,i_k)}$,* *we have*

$$\sum_{j\in\{i_1,\dots,i_k\}} a_{ij}\nu^j = 0, \qquad on\ \partial V_k^{(i_0,\dots,i_k)}, \quad \forall i \in \{i_1,\dots,i_k\}. \tag{5.3.2}$$

Proof. In fact, on the surface $(x^s = 0)$, for some $s \in \{i_1,\dots,i_k\}$ we have $\nu = -e_s$, it follows $\sum\limits_{j\in\{i_1,\dots,i_k\}} a_{ij}\nu^j = a_{is} = x^s(\delta_{si} - x^i) = 0$. On the surface $(x^{i_0} = 0)$ we have $\nu = \frac{1}{\sqrt{k}}(e_{i_1} + \dots + e_{i_k})$, it follows $\sum\limits_{j\in\{i_1,\dots,i_k\}} a_{ij}\nu^j = \frac{1}{\sqrt{k}} \sum\limits_{j\in\{i_1,\dots,i_k\}} a_{ij} = \frac{1}{\sqrt{k}} x^i x^{i_0} = 0$. This completes the proof. This proof was conducted by J. Hofrichter, see also ([34]). □

Proposition 5.14. $L_k^{(i_0,\dots,i_k)}$ *and* $(L_k^{(i_0,\dots,i_k)})^*$ *are weighted adjoins in* $H_k^{(i_0,\dots,i_k)}$*, i.e.*

$$(L_k^{(i_0,\dots,i_k)}X, w_k^{(i_0,\dots,i_k)}Y) = (X, (L_k^{(i_0,\dots,i_k)})^*(w_k^{(i_0,\dots,i_k)}Y)), \quad \forall X, Y \in H_k^{(i_0,\dots,i_k)}.$$

Proof. We denote by $F_i^{(k)}(x) = \sum\limits_{j\in\{i_1,\dots,i_k\}} \frac{\partial(a_{ij}(x)X(x))}{\partial x^j}$. Because $w_k^{(i_0,\dots,i_k)}Y \in C_0^\infty(\overline{V}_k^{(i_0,\dots,i_k)})$ and the second Green formula, and the proposition (5.13) we have

$$\begin{aligned}
(L_k^{(i_0,\dots,i_k)}X, w_k^{(i_0,\dots,i_k)}Y) =& \frac{1}{2} \sum_{i,j\in\{i_1,\dots,i_k\}} \int\limits_{\overline{V}_k^{(i_0,\dots,i_k)}} \frac{\partial^2(a_{ij}(x)X(x))}{\partial x^i \partial x^j} w_k^{(i_0,\dots,i_k)}(x)Y(x)dx \\
=& \frac{1}{2} \sum_{i\in\{i_1,\dots,i_k\}} \int\limits_{\overline{V}_k^{(i_0,\dots,i_k)}} \frac{\partial F_i^{(k)}(x)}{\partial x^i} w_k^{(i_0,\dots,i_k)}(x)Y(x)dx \\
=& \frac{1}{2} \sum_{i\in\{i_1,\dots,i_k\}} \int\limits_{\partial V_k^{(i_0,\dots,i_k)}} F_i^{(k)}(x)\nu_i w_k^{(i_0,\dots,i_k)}(x)Y(x)do(x) \\
&- \frac{1}{2} \sum_{i\in\{i_1,\dots,i_k\}} \int\limits_{\overline{V}_k^{(i_0,\dots,i_k)}} F_i^{(k)}(x) \frac{\partial(w_k^{(i_0,\dots,i_k)}(x)Y(x))}{\partial x^i} dx \\
=& - \frac{1}{2} \sum_{i\in\{i_1,\dots,i_k\}} \int\limits_{\overline{V}_k^{(i_0,\dots,i_k)}} F_i^{(k)}(x) \frac{\partial(w_k^{(i_0,\dots,i_k)}(x)Y(x))}{\partial x^i} dx \\
=& - \frac{1}{2} \sum_{i,j\in\{i_1,\dots,i_k\}} \int\limits_{\overline{V}_k^{(i_0,\dots,i_k)}} \frac{\partial(a_{ij}(x)X(x))}{\partial x^j} \frac{\partial(w_k^{(i_0,\dots,i_k)}(x)Y(x))}{\partial x^i} dx \\
=& - \frac{1}{2} \sum_{i,j\in\{i_1,\dots,i_k\}} \int\limits_{\partial V_k^{(i_0,\dots,i_k)}} a_{ij}(x)\nu_j X(x) \frac{\partial(w_k^{(i_0,\dots,i_k)}(x)Y(x))}{\partial x^i} do(x) \\
&+ \left(X, L_k^*(w_k^{(i_0,\dots,i_k)}Y)\right) \\
=& \left(X, L_k^*(w_k Y)\right).
\end{aligned}$$

□

Proposition 5.15. *In $\overline{V}_k^{(i_0,\ldots,i_k)}$ we have $\left\{X_{m,\alpha}^{(k)}\right\}_{m\geq 0,|\alpha|=m}$ as a basis of $H_k^{(i_0,\ldots,i_k)}$ which is degree orthogonal with respect to weighted $w_k^{(i_0,\ldots,i_k)}$, i.e.,*

$$\left(X_{m,\alpha}^{(k)}, w_k^{(i_0,\ldots,i_k)} X_{j,\beta}^{(k)}\right) = 0, \quad \forall j \neq m, |\alpha| = m, |\beta| = j.$$

Proof. It is easy to see $\left\{X_{m,\alpha}^{(k)}\right\}_{m\geq 0,|\alpha|=m}$ is a basis of $H_k^{(i_0,\ldots,i_k)}$ because $\{x^\alpha\}_\alpha$ is a basis of $H_k^{(i_0,\ldots,i_k)}$. To prove the degree orthogonality we apply the propositions (5.10, 5.12, 5.16)

$$\begin{aligned}
-\lambda_m^{(k)} \left(X_{m,\alpha}^{(k)}, w_k^{(i_0,\ldots,i_k)} X_{j,\beta}^{(k)}\right) &= \left(L_k^{(i_0,\ldots,i_k)} X_{m,\alpha}^{(k)}, w_k^{(i_0,\ldots,i_k)} X_{j,\beta}^{(k)}\right) \\
&= \left(X_{m,\alpha}^{(k)}, (L_k^{(i_0,\ldots,i_k)})^*(w_k^{(i_0,\ldots,i_k)} X_{j,\beta}^{(k)})\right) \\
&= -\lambda_j^{(k)} \left(X_{m,\alpha}^{(k)}, w_k^{(i_0,\ldots,i_k)} X_{j,\beta}^{(k)}\right)
\end{aligned}$$

Because $\lambda_m^{(k)} \neq \lambda_j^{(k)}$, it completes the proof. □

Proposition 5.16. *(i) The spectrum of the operator $L_k^{(i_0,\ldots,i_k)}$ is*

$$Spec(L_k^{(i_0,\ldots,i_k)}) = \bigcup_{m\geq 0} \left\{\lambda_m^{(k)} = \frac{(m+k)(m+k+1)}{2}\right\} = \Lambda_k$$

and the eigenvectors of $L_k^{(i_0,\ldots,i_k)}$ corresponding to $\lambda_m^{(k)}$ are of form

$$X = \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} X_{m,\alpha}^{(k)},$$

i.e., the eigenspaces corresponding to $\lambda_m^{(k)}$ is $\binom{k+m-1}{k-1}$ dimensional;

(ii) The spectrum of the operator L_k is the same.

Proof. (i) Proposition (5.10) implies that $\Lambda_k \subseteq Spec(L_k^{(i_0,\ldots,i_k)})$. Inversely, for $\lambda \notin \Lambda_k$, we will prove that λ is not an eigenvalue of $L_k^{(i_0,\ldots,i_k)}$. In fact, assume that $X \in H_k^{(i_0,\ldots,i_k)}$ such that $L_k^{(i_0,\ldots,i_k)} X = -\lambda X$ in $H_k^{(i_0,\ldots,i_k)}$. Because $\left\{X_{m,\alpha}^{(k)}\right\}_{m,\alpha}$ is a degree orthogonal basis of $H_k^{(i_0,\ldots,i_k)}$ with respect to weighted $w_k^{(i_0,\ldots,i_k)}$ (proposition 5.13), we can represent X by $X = \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} X_{m,\alpha}^{(k)}$. It follows

$$\begin{aligned}
\sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} (-\lambda_m^{(k)}) X_{m,\alpha}^{(n)} =& \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} L_k^{(i_0,\ldots,i_k)} X_{m,\alpha}^{(k)} \\
=& L_k^{(i_0,\ldots,i_k)} X \\
=& -\lambda \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} X_{m,\alpha}^{(k)}.
\end{aligned}$$

For any $j \geq 0$, $|\beta| = j$, multiplying them by $w_k X_{j,\beta}^{(k)}$ and then integrating on $\overline{V}_n$ we have

$$\sum_{|\alpha|=j} d_{j,\alpha}^{(k)} \lambda_j^{(k)} \left(X_{j,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{j,\beta}^{(k)} \right) = \sum_{|\alpha|=j} d_{j,\alpha}^{(k)} \lambda \left(X_{j,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{j,\beta}^{(k)} \right), \forall j \geq 0, |\beta| = j,$$

$$\Rightarrow \left(X_{j,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{j,\beta}^{(k)} \right)_{\beta,\alpha} (d_{j,\alpha}^{(k)} \lambda_j^{(k)})_\alpha = \left(X_{j,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{j,\beta}^{(k)} \right)_{\beta,\alpha} (d_{j,\alpha}^{(k)} \lambda)_\alpha, \forall j \geq 0, |\beta| = j,$$

$$\Rightarrow d_{j,\alpha}^{(k)} \lambda_j^{(k)} = d_{j,\alpha}^{(k)} \lambda, \quad \forall j \geq 0, |\beta| = j, \text{ because } det \left(X_{j,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{j,\beta}^{(k)} \right)_{\beta,\alpha} \neq 0$$

$$\Rightarrow d_{j,\alpha}^{(k)} = 0, \quad \forall j \geq 0, |\alpha| = j, \text{ because } \lambda \neq \lambda_j^{(k)}.$$

It follows $X = 0$ in $H_k^{(i_0,\dots,i_k)}$. Therefore

$$Spec(L_k^{(i_0,\dots,i_k)}) = \bigcup_{m \geq 0} \left\{ \lambda_m^{(k)} = \frac{(m+k)(m+k+1)}{2} \right\} = \Lambda_k.$$

Moreover, assume that $X \in H_k^{(i_0,\dots,i_k)}$ is a eigenvector of $L_k^{(i_0,\dots,i_k)}$ corresponding to $\lambda_j^{(k)}$, i.e., $L_k^{(i_0,\dots,i_k)} X = -\lambda_j X$. We represent X by

$$X = \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} X_{m,\alpha}^{(k)}.$$

It follows

$$\begin{aligned} \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} (-\lambda_m^{(k)}) X_{m,\alpha}^{(k)} &= \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} L_k^{(i_0,\dots,i_k)} X_{m,\alpha}^{(k)} \\ &= L_k^{(i_0,\dots,i_k)} X \\ &= -\lambda_j^{(k)} \sum_{m=0}^{\infty} \sum_{|\alpha|=m} d_{m,\alpha}^{(k)} X_{m,\alpha}^{(k)}. \end{aligned}$$

For any $i \neq j$, $|\beta| = i$, multiplying them by $w_k X_{i,\beta}^{(k)}$ and then integrating on $\overline{V}_n$ we have

$$\sum_{|\alpha|=i} d_{i,\alpha}^{(k)} \lambda_i^{(k)} \left(X_{i,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{i,\beta}^{(k)} \right) = \sum_{|\alpha|=i} d_{i,\alpha}^{(k)} \lambda_j^{(k)} \left(X_{i,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{i,\beta}^{(k)} \right), \forall i \neq j, |\beta| = i,$$

$$\Rightarrow \left(X_{i,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{i,\beta}^{(k)} \right)_{\beta,\alpha} (d_{i,\alpha}^{(k)} \lambda_i^{(k)})_\alpha = \left(X_{i,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{i,\beta}^{(k)} \right)_{\beta,\alpha} (d_{i,\alpha}^{(k)} \lambda_j^{(k)})_\alpha, \forall i \neq j, |\beta| = i,$$

$$\Rightarrow d_{i,\alpha}^{(k)} \lambda_i^{(k)} = d_{i,\alpha}^{(k)} \lambda_j^{(k)}, \quad \forall i \neq j, |\beta| = i, \text{ because } det \left(X_{i,\alpha}^{(k)}, w_k^{(i_0,\dots,i_k)} X_{i,\beta}^{(k)} \right)_{\beta,\alpha} \neq 0$$

$$\Rightarrow d_{i,\alpha}^{(k)} = 0, \quad \forall i \neq j, |\alpha| = i, \text{ because } \lambda_i^{(k)} \neq \lambda_j^{(k)}.$$

It follows

$$X = \sum_{|\alpha|=j} d_{j,\alpha}^{(k)} X_{j,\alpha}^{(k)}.$$

This completes the proof.

(ii) We leave this for the reader to check.

□

5.3.3. The global solution

In this subsection, we describe a definition of solutions. We consider a diploid population of fixed size N with $n+1$ possible alleles $A_0, \ldots, A_n$, at a given locus A. Suppose that the individuals in the population are monoecious, that there is no selection between these alleles and there are no mutations. There are $2N$ alleles in the population in any generation, so it is sufficient to focus on the number $\mathbf{Y}_m = (Y_m^1, \ldots, Y_m^n)$ of alleles $A_1, \ldots, A_n$ at generation time m. Assume that $Y_0 = i_0 = (i_0^1, \ldots, i_0^n)$ and the alleles in generation $m+1$ are derived by sampling with the replacement from the alleles of generation m. This means that the transition probability can be written as

$$\mathbb{P}(Y_{m+1} = j | Y_m = i) = \frac{(2N)!}{(j^0)!(j^1)! \ldots (j^n)!} \prod_{k=0}^{n} \left(\frac{i^k}{2N} \right)^{j^k}, \quad \text{for } i^k, j^k \in \{0, \ldots, 2N\} \ \forall k.$$

where

$$i^0 = 1 - |i| = 1 - i^1 - \ldots - i^n; \qquad j^0 = 1 - |j| = 1 - j^1 - \ldots - j^n.$$

With the re-scaled process,

$$t = \frac{m}{2N}; \quad , X_t = \frac{Y_t}{2N},$$

we have a discrete Markov chain $\mathbf{X}_t$ in $\left\{0, \frac{1}{2N}, \ldots, 1\right\}^n$ (for $t = 1$ we have $2N$ generations). In Chapter 2 Proposition (2.8), we proved that

$$\begin{aligned} \mathbf{X}_0 =& p = \frac{i_0}{2N}; \\ \mathbb{E}(\delta X_t^i) =& 0; \\ \mathbb{E}(\delta X_t^i . \delta X_t^j) =& X_t^i (\delta_{ij} - X_t^j)(\delta t); \\ \mathbb{E}(\delta \mathbf{X}_t)^\alpha =& o(\delta t) \text{ for } |\alpha| \geq 3. \end{aligned} \tag{5.3.3}$$

On the other hand, denoting by $m_\alpha(t)$ the $\alpha^{th}-$moment of distribution about zero at the t^{th} generation, i.e.,

$$m_\alpha(t) = \mathbb{E}(\mathbf{X}_t)^\alpha$$

Then

$$m_\alpha(t + \delta t) = \mathbb{E}(\mathbf{X}_t + \delta \mathbf{X}_t)^\alpha$$

Expanding the right hand side and recalling (5.3.3), we obtain the following recurrence formula, under the assumption that the population number N so large that the terms of higher order $\frac{1}{N^2}, \frac{1}{N^3}$ etc., can be neglected without serious error:

$$\begin{aligned}
m_\alpha(t+\delta t) =& \mathbb{E}(\mathbf{X}_t+\delta\mathbf{X}_t)^\alpha \\
=& \mathbb{E}\Bigg(\Bigg(\mathbb{E}(X_t^1)^{\alpha_1} + \binom{\alpha_1}{1}\mathbb{E}\big((X_t^1)^{\alpha_1-1}\mathbb{E}(\delta X_t^1)\big) + \binom{\alpha_1}{2}\mathbb{E}\big((X_t^1)^{\alpha_1-2}\mathbb{E}(\delta X_t^1)^2\big) \\
&+ \binom{\alpha_1}{3}\mathbb{E}\big((X_t^1)^{\alpha_1-3}\mathbb{E}(\delta X_t^1)^3\big) + \ldots + \binom{\alpha_1}{\alpha_1}\mathbb{E}\big(\mathbb{E}(\delta X_t^1)^{\alpha_1}\big)\Bigg)\times \\
&\times\Bigg(\mathbb{E}(X_t^2)^{\alpha_2} + \binom{\alpha_2}{1}\mathbb{E}\big((X_t^2)^{\alpha_2-1}\mathbb{E}(\delta X_t^2)\big) + \binom{\alpha_2}{2}\mathbb{E}\big((X_t^2)^{\alpha_2-2}\mathbb{E}(\delta X_t^2)^2\big) \\
&+ \binom{\alpha_2}{3}\mathbb{E}\big((X_t^2)^{\alpha_2-3}\mathbb{E}(\delta X_t^2)^3\big) + \ldots + \binom{\alpha_2}{\alpha_2}\mathbb{E}\big(\mathbb{E}(\delta X_t^2)^{\alpha_2}\big)\Bigg)\times\ldots \\
&\times\Bigg(\mathbb{E}(X_t^n)^{\alpha_n} + \binom{\alpha_n}{1}\mathbb{E}\big((X_t^n)^{\alpha_n-1}\mathbb{E}(\delta X_t^n)\big) + \binom{\alpha_n}{2}\mathbb{E}\big((X_t^n)^{\alpha_n-2}\mathbb{E}(\delta X_t^n)^2\big) \\
&+ \binom{\alpha_n}{3}\mathbb{E}\big((X_t^n)^{\alpha_n-3}\mathbb{E}(\delta X_t^n)^3\big) + \ldots + \binom{\alpha_n}{\alpha_n}\mathbb{E}\big(\mathbb{E}(\delta X_t^n)^{\alpha_n}\big)\Bigg)\Bigg) \\
=& \mathbb{E}(\mathbf{X}_t)^\alpha + \sum_{i=1}^n \mathbb{E}\big((\mathbf{X}_t)^{\alpha-e_i}\mathbb{E}(\delta\mathbf{X}_t)^{e_i}\big) + \sum_{i=1}^n\binom{\alpha_i}{2}\mathbb{E}\big((\mathbf{X}_t)^{\alpha-2e_i}\mathbb{E}(\delta\mathbf{X}_t)^{2e_i}\big) \\
&+ \sum_{i\neq j}\alpha_i\alpha_j\mathbb{E}\big((\mathbf{X}_t)^{\alpha-e_i-e_j}\mathbb{E}(\delta\mathbf{X}_t)^{e_i+e_j}\big) + O(\mathbb{E}(\delta\mathbf{X}_t)^\beta) \quad (|\beta|\geq 3) \\
\sim& \left\{1-\frac{|\alpha|(|\alpha|-1)}{2}(\delta t)\right\}\mathbb{E}(\mathbf{X}_t)^\alpha + \sum_{i=1}^n\frac{\alpha_i(\alpha_i-1)}{2}\mathbb{E}(\mathbf{X}_t)^{\alpha-e_i}(\delta t) \quad \text{due to (5.3.3)} \\
=& \left\{1-\frac{|\alpha|(|\alpha|-1)}{2}(\delta t)\right\}m_\alpha(t) + \sum_{i=1}^n\frac{\alpha_i(\alpha_i-1)}{2}m_{\alpha-e_i}(t)(\delta t).
\end{aligned} \tag{5.3.4}$$

It follows that

$$\dot{m}_\alpha(t) = -\frac{|\alpha|(|\alpha|-1)}{2}m_\alpha(t) + \sum_{i=1}^n\frac{\alpha_i(\alpha_i-1)}{2}m_{\alpha-e_i}(t). \tag{5.3.5}$$

With the aim to find a continuous process which is a good approximation to the above discrete process, we should look for a continuous Markov process $\{X_t\}_{t\geq 0}$ in $[0,1]^n$ with the same conditions as (5.3.3) and (5.3.5). Specially, if we call $u(x,t)$ the probability density function of this continuous process, the condition (5.3.3) implies (see for example [24], p. 137) that u is a solution of the Fokker-Planck (Kolmogorov forward) equation

$$\begin{cases} u_t &= L_n u \text{ in } V_n\times(0,\infty), \\ u(x,0) &= \delta_p(x) \text{ in } V_n. \end{cases} \tag{5.3.6}$$

Moreover, the condition (5.3.5) and the above definition of the global solution of J. Hofrichter imply

$$[u_t,\mathbf{x}^\alpha]_n = \left[u, -\frac{|\alpha|(|\alpha|-1)}{2}\mathbf{x}^\alpha + \sum_{i=1}^n\frac{\alpha_i(\alpha_i-1)}{2}\mathbf{x}^{\alpha-e_i}\right]_n = [u, L^*(\mathbf{x}^\alpha)]_n, \forall\alpha,$$

i.e.,

$$[u_t, \phi]_n = [u, L_n^* \phi]_n, \forall \phi \in H_n. \tag{5.3.7}$$

This makes us formulate the following definition.

Definition 5.17. We call $u \in H$ a global solution of the Fokker-Planck equation associated with the general WF model if

$$u_t = L_n u \text{ in } V_n \times (0, \infty); \tag{5.3.8}$$

$$u(\mathbf{x}, 0) = \delta_{\mathbf{p}}(\mathbf{x}) \text{ in } V_n; \tag{5.3.9}$$

$$[u_t, \phi]_n = [u, L_n^* \phi]_n, \ \forall \phi \in H_n. \tag{5.3.10}$$

Remark 5.18. Similar to the definition of the global solution in Chapter 4, here the condition 5.3.8 is only for the usefulness of proof.

We shall prove that

Theorem 5.19. *The Fokker-Planck equation associated with WF model of* $(n+1)-$*alleles has always a unique global solution.*

5.3.4. Proof of Theorem 5.19

This subsection will be devoted to construct the solution as well as the uniqueness of the solution. The process of finding the solution goes as follows:

Firstly, we solve the local solution u_n of the problem (5.3.8-5.3.10) by the separation of variables method. Then we find gradually sub-solutions u_k from $k = n-1$ to $k = 0$ due to the moment condition and known sub-solutions. The global solution will be sum of all these sub-solutions.

$$u = \sum_{k=1}^{n} u_k \chi_{V_k} + \sum_{i=0}^{n} u_0^i \delta_{e_i}.$$

Finally, we check the uniqueness of this global solution.

Step 1: Consider on V_n, assume that $u_n(\mathbf{x}, t) = X(\mathbf{x})T(t)$ is a solution of the Fokker-Planck equation (5.3.8). Then we have

$$\frac{T_t}{T} = \frac{L_n X}{X} = -\lambda$$

Clearly λ is a constant which is independent on T, X. From the Proposition (5.16) we obtain the local solution of the equation (5.3.8) of the form

$$u_n(\mathbf{x}, t) = \sum_{m=0}^{\infty} \sum_{|\alpha|=m} c_{m,\alpha}^{(n)} X_{m,\alpha}^{(n)}(\mathbf{x}) e^{-\lambda_m^{(n)} t},$$

where

$$\lambda_m^{(n)} = \frac{(n+m)(n+m+1)}{2}$$

is the eigenvalue of L_n and

$$X_{m,\alpha}^{(n)}(\mathbf{x}), \quad |\alpha| = m$$

are the corresponding eigenvectors of L_n.

For $m \geq 0, |\beta| = m$, we conclude from Proposition (5.12) that

$$L_n^*\Big(w_n X_{m,\beta}^{(n)}\Big) = -\lambda_m^{(n)} w_n X_{m,\beta}^{(n)}.$$

It follows that

$$\begin{aligned}
[u_t, w_n X_{m,\beta}^{(n)}]_n &= \Big[u, L_n^*\Big(w_n X_{m,\beta}^{(n)}\Big)\Big]_n \quad \text{(the moment condition)}\\
&= -\lambda_m^{(n)}\Big[u, w_n X_{m,\beta}^{(n)}\Big]_n.
\end{aligned}$$

Therefore

$$\begin{aligned}
[u, w_n X_{m,\beta}^{(n)}]_n &= [u(\cdot,0), w_n X_{m,\beta}^{(n)}]_n e^{-\lambda_m^{(n)} t}\\
&= w_n(\mathbf{p}) X_{m,\beta}^{(n)}(\mathbf{p}) e^{-\lambda_m^{(n)} t}.
\end{aligned}$$

Thus,

$$\begin{aligned}
w_n(\mathbf{p}) X_{m,\beta}^{(n)}(\mathbf{p}) e^{-\lambda_m^{(n)} t} &= [u, w_n X_{m,\beta}^{(n)}]_n\\
&= (u_n, w_n X_{m,\beta}^{(n)})_n \quad \text{(because } w_n \text{ vanishes on boundary)}\\
&= \sum_{|\alpha|=m} c_{m,\alpha}^{(n)} (X_{m,\alpha}^{(n)}, w_n X_{m,\beta}^{(n)})_n e^{-\lambda_m^{(n)} t}.
\end{aligned}$$

It follows that

$$\Big(c_{m,\alpha}^{(n)}\Big)_\alpha = \left[\Big((X_{m,\alpha}^{(n)}, w_n X_{m,\beta}^{(n)})_n\Big)_{\alpha,\beta}\right]^{-1} \Big(w_n(\mathbf{p}) X_{m,\beta}^{(n)}(\mathbf{p})\Big)_\beta.$$

Step 2: The solution $u \in H$ satisfying (5.3.8) will be found in the following form

$$u(\mathbf{x},t) = \sum_{k=1}^n u_k(\mathbf{x},t)\chi_{V_k}(x) + \sum_{i=0}^n u_0^i(\mathbf{x},t)\delta_{e^i}(\mathbf{x}). \tag{5.3.11}$$

We use the condition (5.3.10) to obtain gradually values of u_k, $k = n-1, \ldots, 0$. In fact, assume that we want to calculate $u_{n-1}^{(0,\ldots,n-1)}(x^1, \cdots, x^{n-1}, 0, t)$.

We note that, if we choose

$$\phi(\mathbf{x}) = x^1 \cdots x^n X_{k,\beta}^{(n-1)}(x^1, \ldots, x^{n-1}), \quad |\beta| = k.$$

then $\phi(\mathbf{x})$ vanishes on faces of dimension at most $n-1$ except the face $V_{n-1}^{0,\dots,n-1}$. Therefore, the expectation of ϕ will be

$$[u,\phi]_n = (u_n,\phi)_n + (u_{n-1}^{(0,\dots,n-1)},\phi)_{n-1}.$$

The left hand side can be calculated easily by the condition (5.3.10)

$$[u_t,\phi]_n = [u, L_n^*(\phi)]_n = -\lambda_k^{(n-1)}[u,\phi]_n. \tag{5.3.12}$$

It follows

$$[u,\phi]_n = \phi(\mathbf{p})e^{-\lambda_k^{(n-1)}t}.$$

The first part of the right hand side is known as

$$(u_n,\phi)_n = \sum_{m,\alpha} c_{m,\alpha}^{(n)}\left(\int_{V_n} X_{m,\alpha}^{(n)}(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}\right)e^{-\lambda_m^{(n)}t}.$$

Therefore we can expand $u_{n-1}^{(0,\dots,n-1)}(x^1,\cdots,x^{n-1},0,t)$ as follows

$$\begin{aligned}
u_{n-1}^{(0,\dots,n-1)}(x^1,\cdots,x^{n-1},0,t) &= \sum_{m\geq 0} c_m^{(n-1)}(\mathbf{x})e^{-\lambda_m^{(n-1)}t}\\
&= \sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l} c_{m,l,\alpha}^{(n-1)}X_{l,\alpha}^{(n-1)}(x^1,\dots,x^{n-1})e^{-\lambda_m^{(n-1)}t}.
\end{aligned}$$

Put this formula into Equation (5.3.12) we will obtain all the coefficients $c_{m,l,\alpha}^{(n-1)}$. It means that we will obtain $u_{n-1}^{(0,\dots,n-1)}(x^1,\cdots,x^{n-1},0,t)$. Similarly we will obtain u_{n-1}. And finally we will obtain all u_k, $\quad k = n-1,\dots,0$. It means we obtain the global solution in form

$$\begin{aligned}
u(\mathbf{x},t) =& \sum_{k=1}^{n} u_k\chi_{V_k}(\mathbf{x}) + \sum_{i=0}^{n} u_0^i(\mathbf{x},t)\delta_{e_i}(\mathbf{x}).\\
=& \sum_{k=1}^{n}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l} c_{m,l,\alpha}^{(k)}X_{l,\alpha}^{(k)}(\mathbf{x})e^{-\lambda_m^{(k)}t}\chi_{V_k}(\mathbf{x}) + \sum_{i=0}^{n} u_0^i(\mathbf{x},t)\delta_{e_i}(\mathbf{x}).
\end{aligned} \tag{5.3.13}$$

It is not difficult to show that u is a solution of the Fokker-Planck equation associated with WF model.

Step 3: We can easy see that this solution is unique. In fact, assume that u_1,u_2 are two solutions of the Fokker- Planck equation associated with WF model. Then $u = u_1 - u_2$ will satisfy

$$\begin{aligned}
u_t &= L_n u \text{ in } V_n\times(0,\infty),\\
u(x,0) &= 0 \text{ in } \overline{V}_n;\\
[u_t,\phi]_n &= [u, L^*\phi]_n,\ \forall\phi\in H_n.
\end{aligned}$$

It follows

$$
\begin{aligned}
[u_t, 1]_n &= [u, L_n^*(1)]_n = 0,\\
[u_t, x^i]_n &= [u, L_n^*(x^i)]_n = 0,\\
[u_t, w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k^{(i_0,\dots,i_k)}}]_n &= [u, L_n^*(w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k^{(i_0,\dots,i_k)}})]_n\\
&= [u, L_k^*(w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k^{(i_0,\dots,i_k)}})]_n\\
&= -\lambda_j^{(k)} [u, w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k^{(i_0,\dots,i_k)}}]_n.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
[u, 1]_n &= [u(\cdot,0), 1]_n = 0,\\
[u, x^i]_n &= [u(\cdot,0), x^i]_n = 0,\\
[u, w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k^{(i_0,\dots,i_k)}}]_n &= [u(\cdot,0), w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k^{(i_0,\dots,i_k)}}]_n e^{-\lambda_j^{(k)} t} = 0.
\end{aligned}
$$

Since $\left\{1, \left\{x^i\right\}_i, \{w_k^{(i_0,\dots,i_k)} X_{j,\alpha}^{(k)} \chi_{V_k(i_0,\dots,i_k)}\}_{1\le k\le n,(i_0,\dots,i_k)\in I_k, j\ge 0, |\alpha|=j}\right\}$ is also a basis of H_n it follows $u = 0 \in H$.

Example 5.20. To illustrate this process, we consider the case of three alleles.

We will construct the global solution for the problem

$$
\begin{cases}
\frac{\partial u}{\partial t} &= L_2 u, \quad \text{in } V_2 \times (0,\infty),\\
u(\mathbf{x}, 0) &= \delta_{\mathbf{p}}(\mathbf{x}), \quad \mathbf{x} \in V_2,\\
[u_t, \phi]_2 &= [u, L_2^* \phi]_2, \quad \text{for all } \phi \in H_2,
\end{cases}
$$

where the global solution of the form

$$
u = u_2 \chi_{V_2} + u_1^{0,1} \chi_{V_1^{0,1}} + u_1^{0,2} \chi_{V_1^{0,2}} + u_1^{0,0} \chi_{V_1^{0,0}} + u_0^1 \chi_{V_0^1} + u_0^2 \chi_{V_0^2} + u_0^0 \chi_{V_0^0}.
$$

and the product is

$$
\begin{aligned}
[u, \phi]_2 = &\int_{V_2} u_2 \phi_{|V_2} d\mathbf{x} + \int_0^1 u_1^{0,1}(x^1, 0, t) \phi(x^1, 0) dx^1 + \int_0^1 u_1^{0,2}(0, x^2, t) \phi(0, x^2) dx^2\\
&+ \frac{1}{\sqrt{2}} \int_0^1 u_1^{1,2}(x^1, 1-x^1, t) \phi(x^1, 1-x^1) dx^1\\
&+ u_0^1(1,0,t)\phi(1,0) + u_0^2(0,1,t)\phi(0,1) + u_0^0(0,0,t)\phi(0,0).
\end{aligned}
$$

Step 1: We find out the local solution u_2 as follows

$$
u_2(\mathbf{x}, t) = \sum_{m\ge 0} \sum_{\alpha^1+\alpha^2=m} c_{m,\alpha^1,\alpha^2}^{(2)} X_{m,\alpha^1,\alpha^2}^{(2)}(\mathbf{x}) e^{-\lambda_m^{(2)} t}.
$$

To define coefficients $c^{(2)}_{m,\alpha^1,\alpha^2}$ we use the initial condition and the orthogonality of eigenvectors $X^{(2)}_{m,\alpha^1,\alpha^2}$

$$\begin{aligned} w_2(\mathbf{p})X^{(2)}_{m,\beta^1,\beta^2}(\mathbf{p}) &= \big[u(0), w_2X^{(2)}_{m,\beta^1,\beta^2}\big]_2 \\ &= \big(u_2(0), w_2X^{(2)}_{m,\beta^1,\beta^2}\big)_2 \quad \text{because } w_2 \text{ vanishes on the boundary} \\ &= \sum_{\alpha^1+\alpha^2=m} c^{(2)}_{m,\alpha^1,\alpha^2}\Big(X^{(2)}_{m,\alpha^1,\alpha^2}, w_2X^{(2)}_{m,\beta^1,\beta^2}\Big) \quad \text{for all } \beta^1+\beta^2=m. \end{aligned}$$

Because the matrix

$$\Big(X^{(2)}_{m,\alpha^1,\alpha^2}, w_2X^{(2)}_{m,\beta^1,\beta^2}\Big)_{(\alpha^1,\alpha^2),(\beta^1,\beta^2)}$$

is positive definite then we have unique values of $c^{(2)}_{m,\alpha^1,\alpha^2}$. It follows that we have a unique local solution u_2.

Step 2: We will use the moment condition to define all other coefficients of the global solution.

Firstly, we define the coefficients of $u_1^{1,2}$ as follows

$$u_1^{1,2}(x^1, 1-x^1, t) = \sum_{m\geq 0} c_m(x^1)e^{-\lambda^{(1)}_m t} \tag{5.3.14}$$

$$= \sum_{m,l\geq 0} c_{m,l}X^{(1)}_l(x^1)e^{-\lambda^{(1)}_m t}. \tag{5.3.15}$$

We note that

$$L_2^*\Big(x^1x^2X^{(1)}_k(x^1)\Big) = -\lambda^{(1)}_k x^1x^2X^{(1)}_k(x^1).$$

Therefore

$$\big[u_t, x^1x^2X^{(1)}_k(x^1)\big]_2 = \Big[u, L_2^*\Big(x^1x^2X^{(1)}_k(x^1)\Big)\Big]_2 = -\lambda^{(1)}_k[u, x^1x^2X^{(1)}_k(x^1)]_2.$$

It follows that

$$\big[u, x^1x^2X^{(1)}_k(x^1)\big]_2 = p^1p^2X^{(1)}_k(p^1)e^{-\lambda^{(1)}_k t}.$$

Thus we have

$$\begin{aligned}
p^1p^2X_k^{(1)}(p^1)e^{-\lambda_k^{(1)}t} &= \left[u, x^1x^2X_k^{(1)}(x^1)\right]_2 \\
&= \left(u_2, x^1x^2X_k^{(1)}(x^1)\right)_2 + \left(u_1^{1,2}, x^1(1-x^1)X_k^{(1)}(x^1)\right)_1 \\
&\quad \text{because } x^1x^2 \text{ vanish on the other boundaries} \\
&= \sum_{m\geq 0}\left(\sum_{|\alpha|=m} c_{m,\alpha}^{(2)}\left(\int_{V_2} x^1x^2X_{m,\alpha}^{(2)}(x^1,x^2)X_k^{(1)}(x^1)d\mathbf{x}\right)\right)e^{-\lambda_m^{(2)}t} \\
&\quad + \sum_{m\geq 0} c_{m,k}\left(X_k^{(1)}, w_1X_k^{(1)}\right)e^{-\lambda_m^{(1)}t} \\
&\quad \text{because of the orthogonality of } (\cdot,\cdot)_1 \text{ with respect to } w_1 \\
&= \sum_{m\geq 0} r_m e^{-\lambda_m^{(2)}t} + \sum_{m\geq 0} c_{m,k}d_k e^{-\lambda_m^{(1)}t}
\end{aligned}$$

By equating of coefficients of $e^{\alpha t}$ we obtain $u_1^{1,2}$. Similarly we obtain u_1. Then, we define the coefficients of u_0^1 from the 1−th moment.

Note that when $\phi = x^i$, $L_2^*(\phi) = 0$, therefore $[u_t, \phi]_2 = 0$ or

$$[u, x^i]_2 = [u(0), x^i] = p^i.$$

It follows

$$p^1 = [u, x^1] = (u_2, x^1)_2 + (u_1^{0,1}, x^1)_1 + (u_1^{1,2}, x^1)_1 + u_0^1(1, 0, t).$$

Thus we obtain $u_0^1(1, 0, t)$. Similarly we have all u_0. Therefore we obtain the global solution u.

It is easy to check that u is a global solution. To prove the uniqueness we proceed as follows Assume that u is the difference of any two global solutions, i.e. u satisfies

$$\begin{cases}
u_t &= L_2u, \quad \text{in } V_2\times(0,\infty), \\
u(\mathbf{x},0) &= 0, \quad \text{in } V_2 \\
[u_t,\phi]_2 &= [u, L_2^*\phi]_2, \quad \text{for all } \phi\in H_2.
\end{cases}$$

We will prove that

$$[u, \phi]_2 = 0 \quad \forall\phi\in H_2. \tag{5.3.16}$$

In fact,

$$
\begin{aligned}
[u_t, 1]_2 &= [u, L_2^*(1)]_2 = 0 \Rightarrow [u, 1]_2 = [u(0), 1]_2 = 0, \\
[u_t, x^i]_2 &= [u, L_2^*(x^i)]_2 = 0 \Rightarrow [u, x^i]_2 = [u(0), x^i]_2 = 0, \\
[u_t, w_1(x^i) X_m^{(1)}(x^i)]_2 &= [u, L_2^*(w_1(x^i) X_m^{(1)}(x^i))]_2 = -\lambda_m^{(1)} [u, w_1(x^i) X_m^{(1)}(x^i)]_2 \\
&\Rightarrow [u, w_1(x^i) X_m^{(1)}(x^i)]_2 = [u(0), w_1(x^i) X_m^{(1)}(x^i)]_2 e^{-\lambda_m^{(1)} t} = 0, \\
[u_t, w_2(x^1, x^2) X_{m,\alpha}^{(2)}(x^1, x^2)]_2 &= [u, L_2^*(w_2(x^1, x^2) X_{m,\alpha}^{(2)}(x^1, x^2))]_2 = -\lambda_m^{(2)} [u, w_2(x^1, x^2) X_{m,\alpha}^{(2)}(x^1, x^2)]_2 \\
&\Rightarrow [u, w_2(x^1, x^2) X_{m,\alpha}^{(2)}(x^1, x^2)]_2 = [u(0), w_2(x^1, x^2) X_{m,\alpha}^{(2)}(x^1, x^2)]_2 e^{-\lambda_m^{(2)} t} = 0.
\end{aligned}
$$

We need only to prove that Eq. (5.3.16) holds for all

$$\phi(x^1, x^2) = (x^1)^m (x^2)^n, \quad \forall m, n \geq 0.$$

1. If $n = 0, m \geq 0$, we see that ϕ can be generated from $\{1, x^1, w_1(x^1) X_m^{(1)}(x^1)\}$, therefore $[u, \phi]_2 = 0$

2. If $m = 0, n \geq 0$, we see that ϕ can be generated from $\{1, x^2, w_1(x^2) X_m^{(1)}(x^2)\}$, therefore $[u, \phi]_2 = 0$

3. If $n = 1, m \geq 1$, we expand $(x^1)^{m-1}$ by

   $$(x^1)^{m-1} = \sum_{k \geq 0} c_k X_k^{(1)}(x^1).$$

   Note that

   $$L_2^*\Big(x^1 x^2 X_k^{(1)}(x^1)\Big) = -\lambda_k^{(1)} x^1 x^2 X_k^{(1)}(x^1)$$

   Therefore

   $$[u_t, x^1 x^2 X_k^{(1)}(x^1)]_2 = [u, L_2^*\Big(x^1 x^2 X_k^{(1)}(x^1)\Big)]_2 = -\lambda_k^{(1)} [u, x^1 x^2 X_k^{(1)}(x^1)]_2.$$

   It follows

   $$[u, x^1 x^2 X_k^{(1)}(x^1)]_2 = [u(0), x^1 x^2 X_k^{(1)}(x^1)]_2 e^{-\lambda_k^{(1)}} = 0.$$

   Therefore

   $$[u, \phi]_2 = \sum_{k \geq 0} c_k [u, x^1 x^2 X_k^{(1)}(x^1)]_2 = 0.$$

4. If $n \geq 2, m \geq 1$ we use the inductive method in n. We have

$$\begin{aligned}(x^1)^m(x^2)^n &= x^1x^2(x^1+x^2-1)(x^1)^{m-1}(x^2)^{n-2} + (x^1)^m(1-x^1)(x^2)^{n-1} \\ &= -w_2(x^1,x^2)(x^1)^{m-1}(x^2)^{n-2} + (x^1)^m(1-x^1)(x^2)^{n-1}.\end{aligned}$$

In the assumption of induction, we have

$$[u,(x^1)^m(1-x^1)(x^2)^{n-1}]_2 = 0$$

Then, we expand $(x^1)^{m-1}(x^2)^{n-2}$ by

$$(x^1)^{m-1}(x^2)^{n-2} = \sum_{m,\alpha} c^{(2)}_{m,\alpha} X^{(2)}_{m,\alpha}(x^1,x^2).$$

Therefore

$$[u, w_2(x^1,x^2)(x^1)^{m-1}(x^2)^{n-2}]_2 = \sum_{m,\alpha} c^{(2)}_{m,\alpha}[u, w_2(x^1,x^2)X^{(2)}_{m,\alpha}(x^1,x^2)]_2 = 0.$$

It follows $[u,(x^1)^m(x^2)^n]_2 = 0$.

Thus, $u = 0$.

5.3.5. Boundary flux

In this section, we will show that the global solution can be constructed by using the boundary flux (the probability flux at boundary). This work was conducted by J. Hofrichter ([34]).

Definition 5.21. Let $v(\mathbf{x},t)$ be a smooth function (probability density function of a diffusion process) in an open domain $\Omega \times [0,T) \subset \mathbb{R}^{m+1}$. We call $G : \overline{\Omega} \times [0,T) \to \mathbb{R}^n$ a probability flux of v if G is smooth and satisfies the continuity equation

$$\frac{\partial v}{\partial t}(\mathbf{x},t) = -\nabla \cdot G(\mathbf{x},t), \quad (\mathbf{x},t) \in \Omega \times [0,T),$$

and call $G_{|\partial\Omega}$ a boundary flux. We note that if v is improper, i.e. $\int_\Omega v d\mathbf{x} < 1$ then the boundary flux will be nontrivial.

In our case, $u(\mathbf{x},t)$ is a solution of the KFE $u_t = L_n u$ in $\Omega_n \times [0,T)$ and $G = (G^1,\cdots,G^n)$ with

$$G^i(\mathbf{x},t) = -\frac{1}{2}\sum_{j=1}^n \frac{\partial}{\partial x^j}\Big(x^i(\delta_{ij}-x^j)u(\mathbf{x},t)\Big)$$

is a probability flux of u. We will show the relation between boundary flux and difference of the L_n and its adjoint as follows

Proposition 5.22. *If $u(\mathbf{x},t)$ is a solution of the KFE $u_t = L_n u$ in $\Omega_n \times [0,T)$ and $\phi \in H_n$, G is the probability flux of u, then*

$$(L_n u, \phi)_n = -\int_{\partial\Omega_n} G \cdot \nu \phi d\mu_{n-1} + (u, L_n^* \phi)_n$$

where ν is the outward unit normal to $\partial\Omega_n$.

Proof. See [34]. □

The theorem below will show that when we represent global solution in the form

$$u = \sum_{k=0}^{n} u_k \chi_{V_k} = \sum_{k=0}^{n} \sum_{l=1}^{\binom{n+1}{k+1}} u_{k,l} \chi_{V_k^{(l)}},$$

the k^{th} components can be calculated from $(k+1)^{th}$ components due to boundary fluxes. Starting from a local solution u_n, for $0 \le k \le n-1$ decreasing, we construct

$$u_{k,l}(\mathbf{x},t) = \int_0^t v_{k,l}^{\tau}(\mathbf{x}, t-\tau) d\tau,$$

where $v_{k,l}^{\tau} \in H_n$ being the unique solution of the problem

$$\begin{cases} \frac{\partial}{\partial t} v_{k,l}^{\tau} & = L_k v_{k,l}^{\tau} \\ v_{k,l}^{\tau}(\mathbf{x}, 0) & = \sum\limits_{m: \partial V_{k+1}^{(m)} \supset V_k^{(l)}} G_{k+1,m}^{\perp}(\mathbf{x}, \tau) \end{cases}$$

and

$$u_k(\mathbf{x},t) = \sum_l u_{k,l}(\mathbf{x},t) \chi_{V_k^{(l)}}$$

Then, we obtain all u_k for $k = 0, \cdots, n$ and construct

$$u = \sum_k u_k$$

Theorem 5.23. *u constructed by the above procedure is the global solution for the problem.*

Proof. See [34]. □

Example 5.24. We consider the case of two alleles, i.e. $n = 1$. We know that the local solution is

$$u_1(x,p,t) = \sum_{m \ge 0} \frac{4p(1-p)(2m+3)}{(m+1)(m+2)} X_m(p) X_m(x) e^{-\frac{(m+1)(m+2)}{2}t}$$

The boundary flux of u_1 is

$$G_1(x,t) = -\frac{1}{2}\frac{\partial}{\partial x}\Big(x(1-x)u_1(x,t)\Big) = \frac{1}{2}u_1(x,t), \quad x \in \{0,1\},$$

The unique solution of the problem

$$\begin{cases} \frac{\partial}{\partial t} v_0^\tau & = 0, \quad x \in \{0,1\} \\ v_0^\tau(x,0) & = G_1^\perp(x,\tau), \quad x \in \{0,1\} \end{cases}$$

is $v_0^\tau(x,t) = \frac{1}{2} u_1(x,\tau)$ at boundary $x \in \{0,1\}$. Therefore the 0^{th} components of the global solution are

$$u_0(x,t) = \int_0^t v_0^\tau(x, t-\tau) d\tau = \frac{1}{2} \int_0^t u_1(x,\tau) d\tau, \qquad \text{at boundary } x \in \{0,1\}$$

5.4. Applications of the global solution

We mention some applications of this new solution. With the solution received, we can calculate some information of the evolution of the process $(X_t)_{t\geq 0}$ such as the expectation and the second moment of the absorption time, $\alpha^{th}-$moments, fix probabilities, probability of coexistence of $(k+1)-$alleles, and probability of heterogeneity. Some of these quantities have been calculated as solutions to an elliptic equation with appropriate boundary values. However, the proof of uniqueness has not been done. We will also give a rigorous proof for this generalized result (Lemma 5.25). Another approach, which is the blow-up method, has been done by J. Hofrichter ([34]) and applied for a below example (Example 5.29).

Lemma 5.25. *The following generalized Dirichlet problem has at most one solution.*

$$\begin{cases} Au & = 0, \ \text{in } \Omega \subset \mathbb{R}^n \\ \lim\limits_{\mathbf{x} \to \mathbf{x}_0} u(\mathbf{x}) & = f(\mathbf{x}_0), \quad \mathbf{x}_0 \in \Gamma \subset \partial\Omega, \end{cases} \tag{5.4.1}$$

where Ω is a domain in $\mathbb{R}^n$, Γ is the set of regular points of boundary $\partial\Omega$ satisfying $|\partial\Omega - \Gamma| = 0$, A is an elliptic operator, f is of class T_1, i.e. f is the limit of a sequence of continuous functions.

Proof. Step 1: If f_n is a continuous function on $\partial\Omega_n$ then there is a unique solution u_n of the Lemma 5.25.

Step 2: Because f is of class T_1, it is a limit of a sequence of continuous functions $(f_n)_n$ on $\partial\Omega_n$, therefore we obtain the sequence of corresponding unique solutions $(u_n)_n$. This sequence has the limit u. It is not difficult to prove that u is the unique solution to the generalized Dirichlet problem.

See also [27, 28, 40, 60, 63]. □

5.4.1. Probability of having $(k+1)$ alleles

Probability of having only 1 allele A_i (allele A_i is fix) is

$$\begin{aligned}\mathbb{P}(X_t \in V_0^{(i)} | X_0 = \mathbf{p}) &= \int\limits_{V_0^{(i)}} u_0^{(i)}(\mathbf{x}, t) d\mu_0^{(i)}(\mathbf{x}) \\ &= u_0^{(i)}(e_i, t) \\ &= p^i - \sum_{k=1}^{n} \sum_{m^{(k)} \geq 0} \sum_{l^{(k)} \geq 0} \sum_{|\alpha^{(k)}| = l^{(k)}} c^{(k)}_{m^{(k)}, l^{(k)}, \alpha^{(k)}} \left(x^i, X^{(k)}_{l^{(k)}, \alpha^{(k)}} \right)_k e^{-\lambda^{(k)}_{m^{(k)}} t}.\end{aligned}$$

Probability of having exactly $(k+1)$ allele $\{A_0, \dots, A_k\}$ (coexistence probability of alleles $\{A_0, \dots, A_k\}$) is

$$\begin{aligned}\mathbb{P}(X_t \in V_k^{(i_0, \dots, i_k)} | X_0 = \mathbf{p}) &= \int\limits_{V_k^{(i_0, \dots, i_k)}} u_k^{(i_0, \dots, i_k)}(\mathbf{x}, t) d\mu_k^{(i_0, \dots, i_k)}(\mathbf{x}) \\ &= \sum_{m \geq 0} \sum_{l \geq 0} \sum_{|\alpha| = l} c^{(k)}_{m,l,\alpha} \left(\int\limits_{V_k^{(i_0, \dots, i_k)}} X^{(k)}_{m,\alpha}(\mathbf{x}) d\mu_k^{(i_0, \dots, i_k)}(\mathbf{x}) \right) e^{-\lambda^{(k)}_m t}.\end{aligned}$$

5.4.2. Loss of heterozygosity

Probability of heterogeneity is

$$\begin{aligned}H_t =&(n+1)!\, [u, w_n]_n \\ =&(n+1)!\, (u_n, w_n)_n \quad \text{(because } w_n \text{ vanishes on boundary)} \\ =&(n+1)! \Big(\sum_{m \geq 0} \sum_{|\boldsymbol{\alpha}| = m} c^{(n)}_{m,\boldsymbol{\alpha}} X^{(n)}_{m,\boldsymbol{\alpha}} e^{-\lambda^{(n)}_{m,\boldsymbol{\alpha}} t}, w_n X^{(n)}_{0,\mathbf{0}} \Big)_n \\ =&(n+1)! \left(c^{(n)}_{0,\mathbf{0}} X^{(n)}_{0,\mathbf{0}}, w_n X^{(n)}_{0,\mathbf{0}} \right)_n e^{-\lambda^{(n)}_{0,\mathbf{0}} t} \quad \text{(because the orthogonality of eigenvectors } X^{(n)}_{m,\boldsymbol{\alpha}}) \\ =&H_0\, e^{-\frac{(n+1)(n+2)}{2} t}\end{aligned}$$

This means that the heterogeneity decreases with the rate $\frac{(n+1)(n+2)}{4N}$ in generation.

5.4.3. α^{th}-moments

The α^{th}-moments are

$$\begin{aligned}m_\alpha(t) =&[u, \mathbf{x}^\alpha]_n \\ =&\int\limits_{\overline{V_n}} x^\alpha u(\mathbf{x}, t) d\mu(\mathbf{x}) \\ =&\sum_{k=0}^{n} \sum_{(i_0, \dots, i_k) \in I_k} \int\limits_{V_k^{(i_0, \dots, i_k)}} \mathbf{x}^\alpha u_k^{(i_0, \dots, i_k)}(\mathbf{x}, t) d\mu_k^{(i_0, \dots, i_k)}(\mathbf{x}).\end{aligned}$$

5.4.4. Rate of loss of one allele in a population having $(k+1)$ alleles

We have the solution of form

$$u=\sum_{k=0}^{n}u_k(\mathbf{x},t)\chi_{V_k}(\mathbf{x})$$

Then we immediately have that the rate of lost one allele at population having $(k+1)$ alleles is the rate of decrease of

$$u_k(\mathbf{x},t)=\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l}c^{(k)}_{m,l,\alpha}X^{(k)}_{l,\alpha}(x)\chi_{V_k}(x)e^{-\lambda^{(k)}_m t}.$$

which is $\lambda^{(k)}_0=\frac{k(k+1)}{2}$. This means the rate of loss of alleles in the population decreases.

5.4.5. Absorbtion time of having $(k+1)$ alleles

We denote by $T^{k+1}_{n+1}(\mathbf{p})=\inf\left\{t>0:X_t\in\overline{V}_k|X_0=\mathbf{p}\right\}$, the first time the population has at most $(k+1)$ alleles. $T^{k+1}_{n+1}(\mathbf{p})$ is a continuous random variable valued in $[0,\infty)$ and we call $\phi(t,\mathbf{p})$ its probability density function. It is easy to see that $\overline{V}_k$ is invariant with the process $(X_t)_{t\geq 0}$, i.e. if $X_s\in\overline{V}_k$ then $X_t\in\overline{V}_k$ for all $t\geq s$. We have the equality

$$\mathbb{P}(T^{k+1}_{n+1}(\mathbf{p})\leq t)=\mathbb{P}(X_t\in\overline{V}_k|X_0=\mathbf{p})=\int_{\overline{V}_k}u(\mathbf{x},\mathbf{p},t)d\mu(\mathbf{x}).$$

It follows that

$$\phi(t,\mathbf{p})=\int_{\overline{V}_k}\frac{\partial}{\partial t}u(\mathbf{x},\mathbf{p},t)d\mu(\mathbf{x})$$

Therefore the expectation of absorbtion time having $(k+1)$ alleles is

$$\begin{aligned}
\mathbb{E}(T^{k+1}_{n+1}(\mathbf{p}))&=\int_0^\infty t\phi(t,\mathbf{p})dt\\
&=\int_{\overline{V}_k}\int_0^\infty t\frac{\partial}{\partial t}u(\mathbf{x},\mathbf{p},t)dtd\mu(\mathbf{x})\\
&=\sum_{j=1}^{k}\sum_{(i_0,\ldots,i_j)\in I_j}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l}c^{(j)}_{m,l,\alpha}\int_{V_j^{(i_0,\ldots,i_j)}}X^{(j)}_{l,\alpha}(\mathbf{x})\left(\int_0^\infty t\frac{\partial}{\partial t}e^{-\lambda^{(j)}_m t}dt\right)d\mu_j^{(i_0,\ldots,i_j)}(\mathbf{x})\\
&\quad-\sum_{i=0}^{n}\sum_{j=1}^{n}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l}c^{(j)}_{m,l,\alpha}(x^i,X^{(j)}_{j,\alpha})_j\left(\int_0^\infty t\frac{\partial}{\partial t}e^{-\lambda^{(k)}_m t}dt\right),\\
&=\sum_{j=1}^{k}\sum_{(i_0,\ldots,i_j)\in I_j}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l}c^{(j)}_{m,l,\alpha}\int_{V_j^{(i_0,\ldots,i_j)}}X^{(j)}_{l,\alpha}(\mathbf{x})\left(-\frac{1}{\lambda^{(j)}_m}\right)d\mu_j^{(i_0,\ldots,i_j)}(x)\\
&\quad-\sum_{i=0}^{n}\sum_{j=1}^{n}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l}c^{(j)}_{m,l,\alpha}(x^i,X^{(j)}_{j,\alpha})_j\left(-\frac{1}{\lambda^{(k)}_m}\right).
\end{aligned}$$

and the second moment of absorbtion time having $(k+1)$ allele is

$$\begin{aligned}
\mathbb{E}(T_{n+1}^{k+1}(\mathbf{p}))^2 &= \int_0^\infty t^2\phi(t,\mathbf{p})dt \\
&= \int_{\overline{V}_k}\int_0^\infty t^2\frac{\partial}{\partial t}u(\mathbf{x},\mathbf{p},t)dtd\mu(\mathbf{x}) \\
&= \sum_{j=1}^{k}\sum_{(i_0,\ldots,i_j)\in I_j}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l} c_{m,l,\alpha}^{(j)}\int_{V_j^{(i_0,\ldots,i_j)}} X_{m,\alpha}^{(j)}(\mathbf{x})\left(\int_0^\infty t^2\frac{\partial}{\partial t}e^{-\lambda_m^{(j)}t}dt\right)d\mu_j^{(i_0,\ldots,i_j)}(x) \\
&\quad -\sum_{i=0}^{n}\sum_{j=1}^{n}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l} c_{m,l,\alpha}^{(j)}(x^i,X_{j,\alpha}^{(j)})_j\left(\int_0^\infty t^2\frac{\partial}{\partial t}e^{-\lambda_m^{(k)}t}dt\right), \\
&= \sum_{j=1}^{k}\sum_{(i_0,\ldots,i_j)\in I_j}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l} c_{m,l,\alpha}^{(j)}\int_{V_j^{(i_0,\ldots,i_j)}} X_{m,\alpha}^{(j)}(\mathbf{x})\left(-\frac{2}{(\lambda_m^{(j)})^2}\right)d\mu_j^{(i_0,\ldots,i_j)}(\mathbf{x}) \\
&\quad -\sum_{i=0}^{n}\sum_{j=1}^{n}\sum_{m\geq 0}\sum_{l\geq 0}\sum_{|\alpha|=l} c_{m,l,\alpha}^{(j)}(x^i,X_{j,\alpha}^{(j)})_j\left(-\frac{2}{(\lambda_m^{(k)})^2}\right).
\end{aligned}$$

Definition 5.26. The entropy of level one of the variable Y valued in $\{y_1,\ldots,y_M\}$ is

$$H_1(Y) = -\sum_{i=1}^{M}(1-\mathbb{P}(Y=y_i))\log(1-\mathbb{P}(Y=y_i)).$$

The entropy of level k, $\quad (k\leq M)$ of the variable Y valued in $\{y_1,\ldots,y_M\}$ is

$$\begin{aligned}
H_k(Y) = &\sum_{s=1}^{k}(-1)^{k-s}\binom{n-1-s}{k-s}\sum(1-\mathbb{P}(Y\in\{y_{i_1},\ldots,y_{i_s}\})) \\
&\log(1-\mathbb{P}(Y\in\{y_{i_1},\ldots,y_{i_s}\}))
\end{aligned}$$

In 1975, Little ([52]) has been shown that the expectation of the first time to have at most $k-$ alleles in a population of $M-$alleles is

$$E(T_M^k(\mathbf{p})) = -2\sum_{s=1}^{k}(-1)^{k-s}\binom{M-1-s}{k-s}\sum(1-p^{i_1}-\ldots-p^{i_s})\log(1-p^{i_1}-\ldots-p^{i_s})$$

Then we have immediately the following theorem

Theorem 5.27.

$$E(T_M^k(\mathbf{p})) = 2H_k(X_{T_M^1(\mathbf{p})}).$$

5.4.6. Probability distribution at the absorbtion time of having $(k+1)$ alleles

We note that $X_{T_{n+1}^{k+1}(p)}$ is a random variable valued in $\overline{V_k}$. We consider what probability of this random variable valued in $V_k^{(i_0,\ldots,i_k)}$ is, i.e., the probability of the population at the first time to have at most $(k+1)-$alleles consists of $(k+1)-$alleles $\{A_{i_0}, \ldots, A_{i_k}\}$. To make the representation simpler, we denote by g_k is a function of $k-$variables defined by induction as follows

$$\begin{aligned} g_1(p^1) =& p^1; \\ g_2(p^1, p^2) =& \frac{p^1}{1-p^2} g_1(p^2) + \frac{p^2}{1-p^1} g_1(p^1); \\ g_{k+1}(p^1, \ldots, p^{k+1}) =& \sum_{i=1}^{k+1} \frac{p^i}{1 - \sum\limits_{j \neq i} p^j} g_k(p^1, \ldots, p^{i-1}, p^{i+1}, \ldots, p^{k+1}) \end{aligned}$$

Then we have the following result

Theorem 5.28.

$$\mathbb{P}\left(X_{T^{k+1}_{n+1}(p)} \in \overline{V_k^{(i_0,\ldots,i_k)}} \right) = g_{k+1}(p^{i_0}, \ldots, p^{i_k}).$$

Proof. Method 1: By applying the probability of a particular sequence of extinctions and elementary combinatorial arguments, we have immediately the result.

Method 2: By applying the Lemma (5.25) we obtain this probability as the unique solution of the Dirichlet problem

$$\begin{cases} (L_k^{(i_0,\ldots,i_k)})^* v(p) & = 0 \text{ in } V_k \\ \lim\limits_{p \to q} v(p) & = 1, q \in V_k^{(i_0,\ldots,i_k)}, \\ \lim\limits_{p \to q} v(p) & = 0, q \in \partial V_k \backslash V_k^{(i_0,\ldots,i_k)} \backslash V_{k-1}. \end{cases}$$

□

5.4.7. Probability of a particular sequence of extinctions

Let $Q_{M,M-1,\ldots,2}(p^1, \ldots, p^M)$ be the probability with given initial allele frequencies $(p^1, \ldots, p^M)$, allele A_M becomes extinct first, followed by allele A_{M-1}, A_{M-2} and so on, ending with fixation of allele A_1. Littler [55] in 1978, found the result

$$Q_{M,M-1,\ldots,2}(\mathbf{p}) = p^1 \frac{p^2}{1-p^1} \cdots \frac{p^{M-1}}{1 - p^1 - \cdots - p^{M-2}},$$

using known two allele results and elementary combinatorial arguments. Later, in 2007, Baxter, Blythe, and McKane proved above formula by solving the generalized Dirichlet problem

$$\begin{cases} L^*_{M-1} Q_{M,M-1,\ldots,2}(p^1, \ldots, p^M) & = 0, \quad (p^1, \ldots, p^M) \in V_{M-1}, \\ \lim_{(p^1,\ldots,p^M) \to (q^1,\ldots,q^M)} Q_{M,M-1,\ldots,2}(p^1, \ldots, p^M) & = f(q^1, \ldots, q^M), \end{cases} \tag{5.4.2}$$

where

$$f(q^1,\dots,q^M) = \begin{cases} 0, & \text{if } q^i = 0 \text{ for } i = 1,\dots M-1, \\ Q_{M-1,\dots,2}(q^1,\dots,q^{M-1}), & \text{if } q^M = 0. \end{cases}$$

Unfortunately, they did not consider the uniqueness problem, therefore their proof is not rigorous. To prove the uniqueness, note that the boundary value is incomplete and can not be extended continuously to all boundary, so we can not apply the maximum principle as usual. J. Hofrichter has developed a scheme of the blow-up method to solve this problem as follows: First, change of coordinates to decrease the discontinuity (incomplete) of boundary values. Then, after this transformation, the problem becomes a Dirichlet problem with incomplete boundary values but can be extended to all boundary. It means that it will be a classical Dirichlet problem which has the uniqueness of solution. From that we have the uniqueness of the initial solution.

Example 5.29. (See also [34], Chapter 4)

1. We consider the case of 3 alleles A_1, A_2, A_3 and want to calculate the probability $Q_{3,2}(p^1,p^2,p^3)$ which allele A_3 extinct first, then allele A_2 extinct, allele A_1 is fixed. We know that $Q_{3,2}(p^1,p^2,p^3) = u(p^1,p^2)$ satisfies

$$\begin{cases} L_2^* u(p^1,p^2) &= 0, \quad \mathbf{p} \in R_2, \\ u(0,p^2) &= 0, \quad 0 \le p^2 \le 1, \\ u(p^1,0) &= 0, \quad 0 \le p^1 < 1 \\ u(p^1,1-p^1) &= p^1, \quad 0 \le p^1 < 1. \end{cases} \tag{5.4.3}$$

Note that $u(\mathbf{p})$ is not defined at $(1,0)$ because at this point, alleles A_2, A_3 were disappeared without the order of extinction. By blowing-up with the transformation

$$q^1 = p^1, \quad q^2 = \frac{p^2}{1-p^1}, \quad v(\mathbf{q}) = u(\mathbf{p}),$$

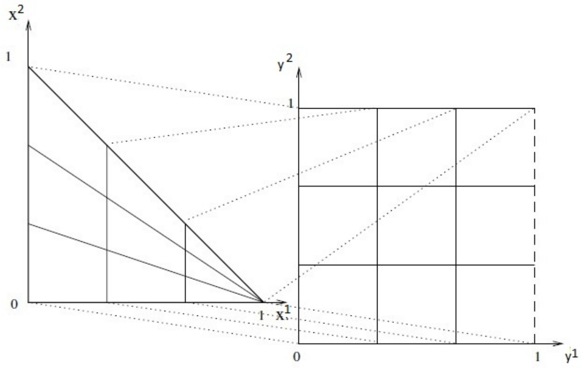


Figure 5.1 Blowup by changing of variables from the simplex R_2 to the square D_2

we have

$$\begin{cases} \frac{1}{2}q^1(1-q^1)^2\frac{\partial^2}{\partial(q^1)^2}v(q^1,q^2)+\frac{1}{2}q^2(1-q^2)\frac{\partial^2}{\partial(q^2)^2}v(q^1,q^2) & =0, \quad \mathbf{q}\in D_2 \\ v(0,q^2) & =0, \quad 0\le q^2\le 1 \\ v(q^1,0) & =0, \quad 0\le q^1<1 \\ v(q^1,1) & =q^1, \quad 0\le q^1<1. \end{cases} \tag{5.4.4}$$

By the continuity of v, as $q^1 \to 1$, on the part of boundary $0 \le q^2 \le 1, q^1 = 1$, we have $v(1, q^2)$ satisfying

$$\begin{cases} \frac{1}{2}q^2(1-q^2)\frac{\partial^2}{\partial(q^2)^2}v(1,q^2) & =0, \quad 0<q^2<1 \\ v(1,0) & =0, \\ v(1,1) & =1, \end{cases} \tag{5.4.5}$$

Therefore $v(1, q^2) = q^2$ on this part of boundary. Now Problem 5.4.4 becomes the classical Dirichlet problem with continuous boundary values. Its unique solution is $v(q^1, q^2) = q^1 q^2$ thus $Q_{3,2}(p^1, p^2, p^3) = p^1 \frac{p^2}{1-p^1}$ is the unique solution of this initial problem (5.4.3).

2. In the case of four alleles A_1, A_2, A_3, A_4, the probability $Q_{4,3,2}(p^1, p^2, p^3, p^4)$ which allele A_4 extinct first, then allele A_3 extinct, next allele A_2 extinct, last allele A_1 is

fixed. We know that $Q_{4,3,2}(p^1,p^2,p^3,p^4) = u(p^1,p^2,p^3)$ satisfies

$$\begin{cases} L_3^* u(p^1,p^2,p^3) & = 0, \quad \mathbf{p} \in R_3, \\ u(0,p^2,p^3) & = 0, \quad 0 \le p^2 \le p^2+p^3 \le 1, p^2 \ne 1, \\ u(p^1,0,p^3) & = 0, \quad 0 \le p^1 \le p^1+p^3 \le 1, p^1 \ne 1, \\ u(p^1,p^2,0) & = 0, \quad 0 \le p^1 \le p^1+p^2 \le 1, \\ u(p^1,p^2,1-p^1-p^2) & = Q_{3,2}(p^1,p^2,p^3) = p^1 \frac{p^2}{1-p^1}, \quad 0 \le p^1 \le p^1+p^2 < 1. \end{cases} \tag{5.4.6}$$

By blowing-up with the transformation

$$q^1 = p^1, \quad q^2 = \frac{p^2}{1-p^1}, \quad q^3 = \frac{p^3}{1-p^1-p^2}, v(\mathbf{q}) = u(\mathbf{p}),$$

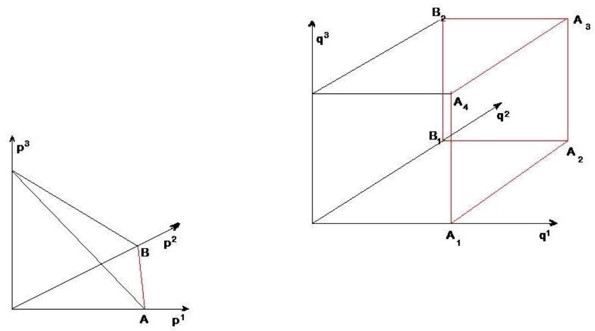


Figure 5.2 Blowup by changing of variables from the simplex R_3 to the cube D_3

the transformed equation is

$$\frac{1}{2}q^1(1-q^1)\frac{\partial^2 v}{\partial(q^1)^2} + \frac{1}{2}\frac{q^2(1-q^2)}{1-q^1}\frac{\partial^2 v}{\partial(q^2)^2} + \frac{1}{2}\frac{q^3(1-q^3)}{(1-q^1)(1-q^2)}\frac{\partial^2 v}{\partial(q^3)^2} = 0, \quad \mathbf{q} \in D_3, \tag{5.4.7}$$

with incomplete boundary values

$$v(\mathbf{q}) = \begin{cases} 0, & \{\mathbf{q} \in \partial D_3 : q^1 = 0\} - [B_1, B_2], \\ 0, & \{\mathbf{q} \in \partial D_3 : q^2 = 0\} - [A_1, A_4], \\ 0, & \{\mathbf{q} \in \partial D_3 : q^3 = 0\} - [A_1, A_2] - [B_1, A_2], \\ q^1 q^2, & \{\mathbf{q} \in \partial D_3 : q^3 = 1\} - [A_3, A_4], -[A_3, B_2] \end{cases}$$

By the continuity of v on $\overline{D_3}$, it follows

$$v(\mathbf{q}) = \begin{cases} 0, & \{\mathbf{q} \in \partial D_3 : q^1 = 0\}, \\ 0, & \{\mathbf{q} \in \partial D_3 : q^2 = 0\}, \\ 0, & \{\mathbf{q} \in \partial D_3 : q^3 = 0\}, \\ q^1 q^2, & \{\mathbf{q} \in \partial D_3 : q^3 = 1\}. \end{cases}$$

We note that, missing boundary values are

$$int\{q^1 = 1\} \cup int\{q^2 = 1\} \cup (A_2, A_3).$$

But due to the continuity of v, we will obtain these values as follows.

By multiplying the equation (5.4.7) with $(1 - q^1)(1 - q^2)$, then taking the limit as $q^1 \to 1, q^2 \to 1$, we obtain

$$\frac{1}{2} q^3 (1 - q^3) \frac{\partial^2}{\partial (q^3)^2} v(1, 1, q^3) = 0,$$

with boundary values $v(A_2) = 0, v(A_3) = 1$. Follows that $v(1, 1, q^3) = q^3$.

Then, by multiplying the equation (5.4.7) with $(1 - q^1)$, then taking the limit as $q^1 \to 1$, we obtain classical Dirichlet problem on the face $\{q^1 = 1\}$ whose solution can be seen easily as

$$v(1, q^2, q^3) = q^2 q^3.$$

Similarly, by multiplying the equation (5.4.7) with $(1 - q^2)$, then taking the limit as $q^2 \to 1$, we obtain a classical Dirichlet problem on the face $\{q^2 = 1\}$ whose solution can be seen easily as

$$v(q^1, 1, q^3) = q^1 q^3.$$

Therefore, we obtain all the boundary values of v. It is easy to see that this boundary value is continuous, it follows that $v = q^1 q^2 q^3$ is the unique solution to the equation (5.4.7) with continuous complete boundary values. Thus $Q_{4,3,2}(p^1, p^2, p^3) = p^1 \frac{p^2}{1-p^1} \frac{p^3}{1-p^1-p^2}$ is the unique solution of this initial problem (5.4.6).

Chapter 6

Geometric structures

In Chapters 4 and 5, we obtained the global solutions. These solutions enabled us to calculate and understand quantitatively as well as qualitatively some biological phenomena of the original WF models. In this chapter, we will consider geometric structures behind these biological phenomena to get a deeper understanding of them. Firstly, we note that the state space is an $n-$dimensional smooth statistical manifold, an Einstein space, and also a dually flat manifold with the Fisher metric. We then will see that the Fisher metric is nothing but the standard metric on the positive part of the sphere of radius two. Next, we consider the affine Laplacian A on this state space. We shall study the behavior of A in various coordinates as well as on various spaces. Finally, we deal with dynamics on the whole state space.

6.1. The geometry of the state space

6.1.1. The Fisher information metric on the state space

Consider a finite set $\Omega = \{1, \dots, n+1\}$ with $\sigma-$algebra $\mathcal{F} = 2^\Omega$ on it. By convention

$$x(\{i\}) := x^i \quad \text{ and } x(A) := \sum_{i\in A} x^i, \ \forall A \in \mathcal{F},$$

it follows that the probability measure x is defined by $\mathbf{x} = (x^1, \dots, x^{n+1})$ and vice versa, moreover the state space

$$\Omega_n = \left\{ \mathbf{x} = (x^1, \dots, x^{n+1}) \ : \ x_i > 0 \text{ for } i = 1, \dots, n+1, \text{ and } \sum_{i=1}^{n+1} x^i = 1 \right\}$$

with the metric

$$g_{ij}(\mathbf{x}) = \frac{\delta_{ij}}{x^i} + \frac{1}{x^{n+1}}, \quad \forall i, j = 1, \dots, n$$

is an $n-$dimensional smooth statistical manifold, i.e., an $n-$dimensional smooth manifold whose points are probability measures defined on $(\Omega, \mathcal{F})$.

Definition 6.1. The Riemannian manifold M is called a *space of constant sectional curvature*, or a *space form* if

$$K(X \wedge Y) := \frac{R_{ijkl}\xi^i \eta^j \xi^k \eta^l}{\Big(g_{ik}g_{jl} - g_{ij}g_{kl}\Big)\xi^i \eta^j \xi^k \eta^l}$$

is a constant K for all independently linear tangent vectors $X = \xi^i \frac{\partial}{\partial x^i}, Y = \eta^i \frac{\partial}{\partial x^i}$. A space form is called *spherical, flat* or *hyperbolic*, depending on whether $K > 0, = 0$ or < 0. M is called an *Einstein manifold* if

$$R_{ik} = c g_{ik}, \quad \text{where } c \text{ is a constant.}$$

We can prove the following lemma.

Lemma 6.2. *1. The sectional curvature of Ω_n is constant and equal to $\frac{1}{4}$;*

2. Ω_n is an Einstein manifold with the Ricci tensor

$$R_{ik} = \frac{n-1}{4} g_{ik};$$

3. The scalar curvature of Ω_n is constant and equal to $R = \frac{n(n-1)}{4}$.

Proof. By some simple calculation we obtain

$$R_{ijkl} = \frac{1}{4}\Big(\frac{\delta_{ik}\delta_{jl} - \delta_{ij}\delta_{kl}}{x^i x^l} + \frac{\delta_{ik} - \delta_{ij}}{x^i x^{n+1}} + \frac{\delta_{jl} - \delta_{kl}}{x^l x^{n+1}}\Big),$$

$$g_{ik}g_{jl} - g_{ij}g_{kl} = \frac{\delta_{ik}\delta_{jl} - \delta_{ij}\delta_{kl}}{x^i x^l} + \frac{\delta_{ik} - \delta_{ij}}{x^i x^{n+1}} + \frac{\delta_{jl} - \delta_{kl}}{x^l x^{n+1}},$$

$$R_{ik} = g^{jl} R_{ijkl} = \frac{n-1}{4} g_{ik},$$

and

$$R = g^{ik} R_{ik} = \frac{n(n-1)}{4}.$$

This completes the proof. □

We will see that the above metric is nothing but the Fisher metric. In fact, we consider a mixture family of probability distributions with parameter η of the form (see also Appendix C.2.2)

$$q_\alpha(\eta) := g^0_\alpha + \sum_{i=1}^{I} g^i_\alpha \eta_i = \sum_{i=1}^{I}\left(g^i_\alpha + g^0_\alpha\right)\eta_i + g^0_\alpha\left(1 - \sum_{i=1}^{I}\eta_i\right). \tag{6.1.1}$$

The important assumption here is that the quantities $q_\alpha(\eta)$ are *linear* functions of the expectation values η_i.
Since $g^i_\alpha + g^0_\alpha$ and g^0_α are probability distributions, we have

$$\sum_\alpha \left(g^i_\alpha + g^0_\alpha\right) = 1, \quad \sum_\alpha g^0_\alpha = 1 \text{ for } i = 1, \dots, I,$$

or simply

$$\sum_\alpha g^0_\alpha = 1, \quad \sum_\alpha g^i_\alpha = 0 \text{ for } i = 1, \dots, I, \tag{6.1.2}$$

and

$$\sum_\alpha q_\alpha(\eta) = 1, \tag{6.1.3}$$

i.e., q is indeed a probability distribution, and

$$\sum_\alpha f^\alpha_i q_\alpha(\eta) = \eta_i \text{ for } i = 1, \dots, I, \tag{6.1.4}$$

that is, η_i is the expectation value of f_i for the probability distribution q.
We consider

$$\varphi(\eta) := \sum_\alpha q_\alpha(\eta) \log q_\alpha(\eta) \tag{6.1.5}$$

and compute

$$\frac{\partial^2}{\partial\eta_i \partial\eta_j}\varphi(\eta) = \sum_\alpha \frac{1}{q_\alpha}\frac{\partial q_\alpha(\eta)}{\partial\eta_i}\frac{\partial q_\alpha(\eta)}{\partial\eta_j} = \sum_\alpha \frac{1}{q_\alpha} g^i_\alpha g^j_\alpha. \tag{6.1.6}$$

Thus, φ is a strictly convex function. Since strict convexity is invariant under *affine linear* coordinate transformations, the particular form of (6.1.1) is not important, as long as we represent our family of probability distributions as a linear family.
We can therefore consider (6.1.6) as a metric on an affine space. That is, we put

$$g_{ij}(\eta) := \frac{\partial^2\varphi(\eta)}{\partial\eta_i\eta_j} \tag{6.1.7}$$

The metric is called the *Fisher metric*, and it will play an important role below.

Since $\varphi(\eta)$ is a convex function, we can perform a Legendre transformation to obtain

$$\vartheta^i(\eta) := \frac{\partial\varphi(\eta)}{\partial\eta_i} = \sum_\alpha g^i_\alpha \log q_\alpha \text{ for } i = 1, \dots, I \tag{6.1.8}$$

and

$$\begin{aligned}
\psi(\vartheta) : &= \max_{\eta} \left(\sum_{i=1}^{I} \vartheta^i \eta_i - \varphi(\eta) \right) \\
&= \sum_{i=1}^{I} \sum_{\alpha} (g_\alpha^i \eta_i - q_\alpha) \log q_\alpha \\
&= - \sum_{\alpha} g_\alpha^0 \log q_\alpha,
\end{aligned} \tag{6.1.9}$$

since the maximum is realized when (6.1.8) holds.

From the properties of the Legendre transformation, we also obtain

$$\Big(\frac{\partial^2 \psi(\vartheta)}{\partial \vartheta^i \vartheta^j} \Big)_{i,j=1,\dots,I} = \left(\Big(\frac{\partial^2 \varphi(\eta)}{\partial \eta_i \eta_j} \Big)_{i,j=1,\dots,I} \right)^{-1} = \left(\sum_{\alpha} \frac{1}{q_\alpha} g_\alpha^i g_\alpha^j \right)^{-1}, \tag{6.1.10}$$

see (6.1.6).
We then have

$$g^{ij}(\eta) = \frac{\partial^2 \psi(\vartheta)}{\partial \vartheta^i \vartheta^j} = \frac{\partial \eta_j}{\partial \vartheta^i}, \tag{6.1.11}$$

again by the properties of the Legendre transform, i.e. (6.1.11) is the inverse of the Fisher metric.
We recall (6.1.7)

$$g_{ij}(\eta) = \frac{\partial^2 \varphi(\eta)}{\partial \eta_i \eta_j} = \frac{\partial \vartheta^i}{\partial \eta_j}. \tag{6.1.12}$$

In fact, we have

$$\begin{aligned}
\sum_{i,j} g^{ij}(\eta) d\vartheta^i d\vartheta^j &= \sum_{i,j} g^{ij}(\eta) \sum_{k,l} \frac{\partial \vartheta^i}{\partial \eta^k} \frac{\partial \vartheta^j}{\partial \eta^l} d\eta^k d\eta^l \\
&= \sum_{i,j} g^{ij}(\eta) \sum_{k,l} g_{ik}(\eta) g_{jl}(\eta) d\eta^k d\eta^l \\
&= \sum_{k,l} g_{kl}(\eta) d\eta^k d\eta^l,
\end{aligned} \tag{6.1.13}$$

i.e., the inverse metric tensor $g^{ij}(\eta)$ in the ϑ-coordinates is the same as the tensor $g_{ij}(\eta)$ in the η-coordinates.
When we also put $\vartheta^0 := \sum_\alpha g_\alpha^0 \log q_\alpha$, we can invert (6.1.8) and obtain

$$\log q_\alpha(\eta) = \sum_{j=0}^{I} f_j^\alpha \vartheta^j, \tag{6.1.14}$$

i.e.,

$$q_\alpha(\eta) = \exp\Big(\sum_{j=0}^{I} f_j^\alpha \vartheta^j \Big). \tag{6.1.15}$$

Recalling $f_0^\alpha = 1$ for all α, (6.1.9) and the normalization $\sum_\alpha q_\alpha(\eta) = 1$ yield

$$\psi(\vartheta) = -\vartheta^0 = \log \sum_\alpha \exp\Big(\sum_{i=1}^I f_i^\alpha \vartheta^i\Big) \tag{6.1.16}$$

so that we can rewrite (6.1.15) as

$$q_\alpha(\eta) = \exp\Big(\sum_{i=1}^I f_i^\alpha \vartheta^i - \psi(\vartheta)\Big) =: p^\alpha(\vartheta). \tag{6.1.17}$$

This is the so-called *Gibbs distribution*.
We now compute the expectation value η_{ij} of the product $f_i f_j$ for the distribution q. We obtain, using (6.1.4),

$$\begin{aligned}
\eta_{ij} &= \sum_\alpha f_i^\alpha f_j^\alpha p^\alpha(\vartheta) \\
&= \exp(-\psi(\vartheta)) \frac{\partial}{\partial \vartheta^i} \sum_\alpha f_j^\alpha \exp\Big(\sum_{k=1}^I f_k^\alpha \vartheta^k\Big) \\
&= \exp(-\psi(\vartheta)) \frac{\partial}{\partial \vartheta^i}\Big(\exp(\psi(\vartheta))\eta_j\Big) \\
&= \frac{\partial \psi(\vartheta)}{\partial \vartheta^i}\eta_j + \frac{\partial \eta_j}{\partial \vartheta^i} \\
&= \eta_i \eta_j + g^{ij}(\eta).
\end{aligned}$$

We now look at an important special case that will appear below in our discussion of the Wright-Fisher model. We simply take the probabilities $p^1, \dots, p^{k-1}$ as our observables, and the expectation value η_i of the ith observable then is the p^i, so that we get the special case $q_\alpha = \eta_\alpha$ of (6.1.1) and can apply our result that the covariance is the inverse of the Fisher metric.
From (6.1.1), we then have

$$\begin{aligned}
q_\alpha(\eta) &= \eta_\alpha \text{ for } \alpha = 1, \dots, k-1 && (6.1.18)\\
q_k &= \eta_k = 1 - \sum_{\beta=1}^{k-1} \eta_\beta, && (6.1.19)
\end{aligned}$$

that is,

$$g_\alpha^i = \delta_{i\alpha}, \quad g_k^i = -1 \quad g_\alpha^0 = 0 \text{ for } \alpha = 1, \dots, k-1, \qquad g_k^0 = 1 \tag{6.1.20}$$

and from (6.1.8), (6.1.9)

$$\vartheta^i = \log \frac{\eta_i}{\eta_k} \text{ for } i = 1, \dots, k-1, \qquad \psi(\vartheta) = -\log p^k. \tag{6.1.21}$$

6.1.2. The Fisher metric as the standard metric on the sphere

For simplicity, we show the computations for the special case where $\eta_\alpha = p_\alpha$, as considered at the end of the previous section. Then, recalling (6.1.20), the metric tensor (6.1.6) in the coordinates $p_1, \dots, p_k$ becomes

$$\begin{pmatrix} \frac{1}{p_1} & 0 & \dots & 0 \\ 0 & \frac{1}{p_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{p_k} \end{pmatrix}. \tag{6.1.22}$$

However, this is not yet the expression for a Riemannian metric because we have k coordinates $p_1, \dots, p_k$ on a $(k-1)$-dimensional space. This can easily be corrected by expressing

$$p_k = 1 - \sum_{j=1}^{k-1} p_j. \tag{6.1.23}$$

Therefore we can consider the Fisher metric as a metric on the $(k-1)$-dimensional simplex

$$\Sigma^{k-1} = \{(p_1, \dots p_k) : p_i \geq 0, \sum p_i = 1\}. \tag{6.1.24}$$

We know that the transformation behavior of a Riemannian metric,

$$g_{ij}(x) = \sum_{\alpha,\beta} \gamma_{\alpha\beta}(y) \frac{\partial y^\alpha}{\partial x^i} \frac{\partial y^\beta}{\partial x^j}, \tag{6.1.25}$$

where x and $y = y(x)$ are two different coordinates.

Applying for our case with $x = (p_1, \dots, p_{k-1})$, $y = (p_1, \dots, p_k)$, we obtain

$$\frac{\partial p_k}{\partial p_j} = -1 \text{ for } j = 1, \dots, k-1, \tag{6.1.26}$$

and the metric tensor $g_{ij}(p)$ in the coordinates $p_1, \dots p_{k-1}$ as

$$\begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_k} & \frac{1}{p_k} & \dots & \frac{1}{p_k} \\ \frac{1}{p_k} & \frac{1}{p_2} + \frac{1}{p_k} & \dots & \frac{1}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p_k} & \frac{1}{p_k} & \dots & \frac{1}{p_{k-1}} + \frac{1}{p_k} \end{pmatrix}, \tag{6.1.27}$$

with p_k given by (6.1.23). The inverse metric tensor $g^{ij}(p)$ then becomes

$$\begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_{k-1} \\ -p_1p_2 & p_2(1-p_2) & \dots & -p_2p_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_{k-1} & -p_2p_{k-1} & \dots & p_{k-1}(1-p_{k-1}) \end{pmatrix}. \tag{6.1.28}$$

We can also rewrite the metric in spherical coordinates, by simply putting

$$q_i := \sqrt{p_i}. \tag{6.1.29}$$

Applying the transformation rule (6.1.25) with $\frac{\partial p_\alpha}{\partial q_j} = 2\delta_\alpha^j q_j$, we obtain $g_{ij} = 4\delta_{ij}$ in the q-coordinates, which is, simply the Euclidean metric. However, as before, we need to satisfy the normalization constraint (6.1.23) which now becomes $q_k = \sqrt{1 - \sum_{j=1}^{k-1}(q_j)^2}$. Using $\frac{\partial q_k}{\partial q_j} = -\frac{q_j}{q_k}$ for $j = 1, \ldots, k-1$, we obtain

$$\left(g^{ij}(q)\right)_{ij} = \frac{4}{q_k^2}\begin{pmatrix} q_k^2 + q_1^2 & q_1 q_2 & \cdots & q_1 q_{k-1} \\ q_1 q_2 & q_k^2 + q_2^2 & \cdots & q_2 q_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ q_1 q_{k-1} & q_2 q_{k-1} & \cdots & q_k^2 + q_{k-1}^2 \end{pmatrix}. \tag{6.1.30}$$

Since this has been obtained by restricting the Euclidean metric to the unit sphere, it must be the standard metric on the unit sphere S^{k-1}, up to the factor 4 that emerged in our computations. Since the standard metric on the sphere has sectional curvature $\equiv 1$, and since the curvature of a Riemannian metric scales with the inverse of a scaling factor, our factor 4 leads to the following lemma.

Lemma 6.3. *The Fisher metric on the standard simplex Σ^{k-1} is four times the standard metric on the unit sphere S^{k-1}, and its sectional curvature is equal to $\frac{1}{4}$.*

6.2. The affine Laplacian on the state space

Given an affine structure with coordinates $\eta_i, i = 1, \ldots, k$ as above and a metric g_{ij}, we formulate

Definition 6.4. The operator given by

$$A := \sum_{i,j} g^{ij} \frac{\partial^2}{\partial \eta_i \partial \eta_j} \tag{6.2.1}$$

is called the affine Laplacian.

A solution of

$$Af = 0 \tag{6.2.2}$$

is called an *affine harmonic function.*

Lemma 6.5. *For the function*

$$\varphi(\eta) = \sum_\alpha q_\alpha(\eta) \log q_\alpha(\eta) \tag{6.2.3}$$

of (6.1.5), we have

$$A\varphi = k \equiv \text{const.} \tag{6.2.4}$$

The *proof* is obvious from the definition (6.1.7) of the Fisher metric.

This result will play an useful role below when we want to understand exit times of genetic drift.

6.2.1. The affine Laplacian in dual coordinates

We can transform the affine Laplacian from the coordinates η_i to the dual coordinates ϑ^i. Even though those dual coordinates also define an affine structure that is dual to the original one, this transformation is not affine itself, simply because the two affine structures are different. Therefore, in the ϑ-coordinates, A will acquire an additional first order term. We have

$$\frac{\partial^2}{\partial\eta_i\partial\eta_j} = \sum_{\ell,m} \frac{\partial\vartheta^\ell}{\partial\eta_i}\frac{\partial\vartheta^m}{\partial\eta_j}\frac{\partial^2}{\partial\vartheta^\ell\partial\vartheta^m} + \sum_\ell \frac{\partial^2\vartheta^\ell}{\partial\eta_i\partial\eta_j}\frac{\partial}{\partial\vartheta^\ell}. \tag{6.2.5}$$

Here and in the sequel, all sums range from 1 to $k-1$.
We recall (6.1.11), i.e.,

$$g_{ij}(\eta) = \frac{\partial^2\varphi(\eta)}{\partial\eta_i\eta_j} = \frac{\partial\vartheta^i}{\partial\eta_j} \tag{6.2.6}$$

and therefore obtain

$$A = \sum_{i,j} g^{ij}(\eta)\frac{\partial^2}{\partial\eta_i\partial\eta_j} = \sum_{\ell,m} g_{\ell m}\frac{\partial^2}{\partial\vartheta^\ell\partial\vartheta^m} + \sum_{i,j,\ell} g^{ij}(\eta)\frac{\partial^3\varphi(\eta)}{\partial\eta_i\partial\eta_j\partial\eta_\ell}\frac{\partial}{\partial\vartheta^\ell}. \tag{6.2.7}$$

Since $g^{ij}(\eta) = \eta_i(\delta_{ij} - \eta_j)$ and $\frac{\partial^3\varphi(\eta)}{\partial\eta_i\partial\eta_j\partial\eta_\ell} = -\frac{1}{(\eta_i)^2}\delta_{ij}\delta_{il} + \frac{1}{(\eta_k)^2}$ by (6.1.6), we obtain

$$\begin{aligned} A &= \sum_{\ell,m}\Big(\frac{\delta_{\ell m}}{\eta_\ell} + \frac{1}{\eta_k}\Big)\frac{\partial^2}{\partial\vartheta^\ell\partial\vartheta^m} + \sum_\ell\Big(-\frac{1}{\eta_\ell} + \frac{1}{\eta_k} + k - 2\Big)\frac{\partial}{\partial\vartheta^\ell}. \\ &= \sum_{\ell,m}\Big(\frac{\delta_{\ell m}}{e^{(\vartheta_\ell)}} + 1\Big)\Big(1 - \sum_i e^{\vartheta^i}\Big)\frac{\partial^2}{\partial\vartheta^\ell\partial\vartheta^m} + \sum_\ell\Big(-\frac{1-\sum_i e^{\vartheta^i}}{e^{\vartheta^\ell}} + k - 1 - \sum_i e^{\vartheta^i}\Big)\frac{\partial}{\partial\vartheta^\ell}. \end{aligned} \tag{6.2.8}$$

Here, ϑ^i ranges between $-\infty$ and $\log\left(\frac{1}{\eta_k}\right)$. Thus, we have transformed the singularity at the boundary, where some η_i become 0, to $-\infty$.

6.2.2. The affine and the Beltrami Laplacian on the sphere

Under the coordinate transformations from p^i on the simplex to the coordinates $q^i = \sqrt{p^i}$ on the sphere, the affine Laplacian becomes

$$\frac{\partial^2}{\partial p^i\partial p^j} = \sum_{\ell,m}\frac{\partial q^\ell}{\partial p^i}\frac{\partial q^m}{\partial p^j}\frac{\partial^2}{\partial q^\ell\partial q^m} + \sum_\ell\frac{\partial^2 q^\ell}{\partial p^i\partial p^j}\frac{\partial}{\partial q^\ell}. \tag{6.2.9}$$

Then on the sphere, the affine Laplacian is given by the form

$$A = \bar{g}^{\ell m}(q)\frac{\partial^2}{\partial q^\ell\partial q^m} + \frac{1-(q^\ell)^2}{4q^\ell}\frac{\partial}{\partial q^\ell},$$

where $\bar{g}^{\ell m}(q) = \frac{1}{4}\Big(\delta_{\ell m} - q^\ell q^m\Big)$ is the inverse metric tensor on the sphere.

We know that on the simplex the difference of the affine Laplacian from Beltrami Laplacian is twice of the Christoffel force, when we change the coordinates to the sphere, it adds some additional first order terms. To detail on this, let us consider the Beltrami Laplacian on the sphere

$$\Delta_{\bar{g}}(q) = \bar{g}^{\ell m}(q)\frac{\partial^2}{\partial q^\ell \partial q^m} + \frac{(1-n)q^\ell}{4}\frac{\partial}{\partial q^\ell}.$$

6.2.3. Eigenvalues and eigenfunctions of A and A^* and orthogonality relations

In this subsection, we will summarize some properties of the operator A defined from $K = C_0^\infty(\Omega)$ into itself and its dual A^* defined from $C^\infty(\Omega)$ into itself. First, we recall

$$w(x^1, \ldots, x^n) = x^1 x^2 \ldots x^n(1 - x^1 - \ldots - x^n)$$

the weight function, we have the following propositions

Proposition 6.6. *The spectra of A and A^* are the same, and are given by*

$$Spec(A) = Spec(A^*) = \bigcup_{m\geq 0}\left\{\lambda_m = \frac{(m+n)(m+n+1)}{2}\right\} = \Lambda_n.$$

Proposition 6.7. *The eigenvectors of A corresponding to the eigenvalue λ_m are*

$$w(x^1, \ldots, x^n)\mathcal{E}_{m_1,\ldots,m_n}(2n+1, 2, \ldots, 2; x^1, \ldots, x^n).$$

It follows that the eigenspace corresponding to the eigenvalue λ_m is of dimension $\binom{n+m-1}{n-1}$.

Proposition 6.8. *The eigenvectors of A^* corresponding to the eigenvalue λ_m are*

$$\mathcal{F}_{m_1,\ldots,m_n}(2n+1, 2, \ldots, 2; x^1, \ldots, x^n).$$

It follows the eigenspace corresponding to the eigenvalue λ_m is of dimension $\binom{n+m-1}{n-1}$.

Proposition 6.9. *The eigenvetors of A and A^* form a biorthogonal system with the weight function w, i.e.*

$$\begin{aligned}&\int_\Omega \mathcal{F}_{m_1,\ldots,m_n}(2n+1, 2, \ldots, 2; x^1, \ldots, x^n)\\ &\quad\times w(x^1, \ldots, x^n)\mathcal{E}_{m_1,\ldots,m_n}(2n+1, 2, \ldots, 2; x^1, \ldots, x^n)dx^1 \ldots dx^n\\ &= K_{m_1,\ldots,m_n}\delta_{m_1,m_1'} \ldots \delta_{m_n,m_n'}.\end{aligned}$$

where

$$K_{m_1,\ldots m_n} = \frac{1}{(m_1+1)\ldots(m_n+1)(m_1+\ldots+m_n+2)_{2n-1}(2m_1+\ldots+2m_n+2n+1)}.$$

6.3. Dynamics on the whole state space

In Chapter 3, we learned that the diffusion equation approximation to the general WF models is given in backward form by the equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} g^{ij}(p) \frac{\partial^2 u}{\partial p^i \partial p^j}.$$

where $(g^{ij}(p))$ is the covariance matrix of the multinomial distribution. In the point of view of Principle of Optimal Randomness (see for example [5]) we will represent this as follows

$$\frac{\partial u}{\partial t} = \frac{1}{2} \Delta_g u + ch^i(p) \frac{\partial u}{\partial p^i},$$

where $ch^k(p) = \frac{1}{2} g^{ij}(p) \Gamma^k_{ij}(p)$ is the Christoffel force and Δ_g is the Laplace-Beltrami operator with respect to the metric g_{ij}. The Christoffel forces can be obtained easily by the following lemma

Lemma 6.10. *The Christoffel forces (velocities) of the BKE of the general WF model are*

$$ch^k(p) = -\frac{1}{4}(1 - np^k).$$

Proof. The covariance matrix of the multinomial distribution is given by $g^{ij}(p) = p^i(\delta_{ij} - p^j)$ and its inverse is given by the Fisher metric $g_{ij}(p) = \frac{\delta_{ij}}{p^i} + \frac{1}{p^n}$. Then the Christoffel symbols are

$$\Gamma_{ij,k}(p) = \frac{1}{2} \left(g_{ik,j}(p) + g_{jk,i}(p) - g_{ij,k}(p) \right) = -\frac{1}{2} \left(\frac{\delta_{ijk}}{(p^i)^2} + \frac{1}{(p^n)^2} \right),$$

$$\Gamma^k_{ij}(p) = g^{kl}(p) \Gamma_{ij,l}(p) = \frac{1}{2} \left(-\frac{\delta_{ijk}}{p^k} + \frac{\delta_{ij} p^k}{p^i} + \frac{p^k}{p^n} \right).$$

Therefore the Christoffel force are

$$ch^k(p) = \frac{1}{2} g^{ij}(p) \Gamma^k_{ij}(p) = -\frac{1}{4}(1 - np^k).$$

□

Here, we give some immediate interpretations (see [5]) of the above results

(i) The Christoffel velocities vanish at the centroid of the frequency space, so that near the centroid, the diffusion equation for the general WF model is very well approximated by that of spherical Brownian motion.

(ii) The Christoffel velocities drive the populations toward the vertices of the frequency space. These forces should not be confused with the biologically real forces of mutation or selection.

(iii) The diffusion equation for n-allele random drift is invariant under the group of isometries of frequency space. This can be shown to be isomorphic to the group of permutations S, of the n allelic frequencies among themselves. The diffusion equation for random drift is optimally random relative to this permutation group but is not relative to the full group of isometries of an $(n-1)$-sphere.

(iv) Since it is a property of Brownian motion on an $(n-1)$-sphere that equal distances have equal mean transit times, this is definitely not true for the diffusion equation of the general WF model. Therefore, the exact (rather than relative) positions of populations are necessary in order to compute mean transit times precisely.

Appendix A

Introduction to hypergeometric functions

In this appendix, we briefly introduce hypergeometric functions used in Chapter 4. These functions are most useful tools for solving singular linear second order ODEs. We refer readers to [1],[72],[73] for further details.

A.1. Gegenbauer polynomials

Definition A.1. Gegenbauer polynomials or ultraspherical polynomials (named for Leopold Gegenbauer) $C_n^\alpha(x)$ are defined in terms of their generating function ([72, § IV.2]):

$$\frac{1}{(1-2xt+t^2)^\alpha} = \sum_{n\geq 0} C_n^\alpha(x) t^n.$$

They generalize the Legendre polynomials and the Chebyshev polynomials, and are special cases of the Jacobi polynomials.

Proposition A.2. *[73]*

- *The Gegenbauer polynomials satisfy the recurrence relation*

$$C_0^\alpha(x) = 1 \tag{A.1.1}$$

$$C_1^\alpha(x) = 2\alpha x \tag{A.1.2}$$

$$C_n^\alpha(x) = \frac{1}{n}[2x(n+\alpha-1)C_{n-1}^\alpha(x) - (n+2\alpha-2)C_{n-2}^\alpha(x)]. \tag{A.1.3}$$

- *The Gegenbauer polynomials are particular solutions of the Gegenbauer differential equation*

$$(1-x^2)y'' - (2\alpha+1)xy' + n(n+2\alpha)y = 0. \tag{A.1.4}$$

- *The Gegenbauer polynomials are special cases of the Jacobi polynomials*

$$C_n^{(\alpha)}(x) = \frac{(2\alpha)_n}{(\alpha+\frac{1}{2})_n} P_n^{(\alpha-1/2,\alpha-1/2)}(x).$$

Proposition A.3. *[1, p. 774] For a fixed α, the Gegenbauer polynomials are orthogonal polynomials on the interval $[-1,1]$ with respect to the weight function $(1-x^2)^{\alpha-\frac{1}{2}}$:*

$$\int_{-1}^{1} (1-x^2)^{\alpha-\frac{1}{2}} \left[C_n^{(\alpha)}(x)\right] \left[C_m^{(\alpha)}(x)\right] dx = \delta_{nm} \frac{\pi 2^{1-2\alpha}\Gamma(n+2\alpha)}{n!(n+\alpha)[\Gamma(\alpha)]^2}. \tag{A.1.5}$$

A.2. Jacobi polynomials

Definition A.4.

$$P_n^{(\alpha,\beta)}(z) = \frac{\Gamma(\alpha+n+1)}{n!\,\Gamma(\alpha+\beta+n+1)} \sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\alpha+\beta+n+m+1)}{\Gamma(\alpha+m+1)} \left(\frac{z-1}{2}\right)^m,$$

Proposition A.5.

- *The Jacobi polynomials show the symmetry relation*

$$P_n^{(\alpha,\beta)}(-z) = (-1)^n P_n^{(\beta,\alpha)}(z);$$

- *The k^{th} derivative of the explicit expression leads to*

$$\frac{\mathrm{d}^k}{\mathrm{d}z^k} P_n^{(\alpha,\beta)}(z) = \frac{\Gamma(\alpha+\beta+n+1+k)}{2^k\Gamma(\alpha+\beta+n+1)} P_{n-k}^{(\alpha+k,\beta+k)}(z);$$

- *The Jacobi polynomial $P_n^{(\alpha,\beta)}$ is a solution of the second order linear homogeneous differential equation*

$$(1-x^2)y'' + (\beta-\alpha-(\alpha+\beta+2)x)y' + n(n+\alpha+\beta+1)y = 0; \tag{A.2.1}$$

- *The Jacobi polynomials are special cases of the hypergeometric polynomials*

$$P_n^{(\alpha,\beta)}(z) = \frac{(\alpha+1)_n}{n!}\, {}_2F_1\left(-n, 1+\alpha+\beta+n; \alpha+1; \frac{1-z}{2}\right).$$

Proposition A.6. *For fixed $\alpha > -1$ and $\beta > -1$, the Jacobi polynomials are orthogonal polynomials on the interval $[-1,1]$ with respect to the weight function $(1-x)^\alpha(1+x)^\beta$:*

$$\int_{-1}^{1} (1-x)^\alpha (1+x)^\beta P_m^{(\alpha,\beta)}(x) P_n^{(\alpha,\beta)}(x)\, dx = \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1} \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)n!} \delta_{nm}$$

A.3. Hypergeometric functions

Definition A.7. The Gaussian or ordinary hypergeometric function ${}_2F_1(a,b;c;z)$ is a special function represented by the hypergeometric series,

$${}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n} \frac{z^n}{n!} \tag{A.3.1}$$

provided c is not $0, -1, -2, \ldots$, where the Pochhammer symbol is given by

$$(a)_n = a(a+1)(a+2)\cdots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)},\ (a)_0 = 1. \tag{A.3.2}$$

For complex values of z it can be analytically continued along any path that avoids the branch points 0 and 1.

Proposition A.8. *The hypergeometric function is a solution of the Euler's hypergeometric differential equation*

$$z(1-z)\frac{d^2w}{dz^2} + [c-(a+b+1)z]\frac{dw}{dz} - abw = 0. \tag{A.3.3}$$

which has three regular singular points: $0, 1$ and ∞. The generalization of this equation to three arbitrary regular singular points is given by Riemann's differential equation. Any second order differential equation with three regular singular points can be converted to the hypergeometric differential equation by a change of variables.

Proposition A.9.

- *The Jacobi polynomials $P_n^{(\alpha,\beta)}$ and their special cases the Legendre polynomials, the Chebyshev polynomials, the Gegenbauer polynomials can be written in terms of hypergeometric functions using the following*

$$_2F_1(-n, \alpha+1+\beta+n; \alpha+1; x) = \frac{n!}{(\alpha+1)_n} P_n^{(\alpha,\beta)}(1-2x)$$

- *The Gegenbauer polynomials are given as Gaussian hypergeometric series where the series is finite*

$$C_n^{(\alpha)}(z) = \frac{(2\alpha)_n}{n!}\, {}_2F_1\left(-n, 2\alpha+n; \alpha+\frac{1}{2}; \frac{1-z}{2}\right).$$

Appendix B

Introduction to generalized hypergeometric functions

In this appendix, we give an overview of generalized hypergeometric functions used in Chapter 5. These functions are very useful tools for solving singular linear second order equations of multivariables. We refer readers to [11], [12], [13], [15], [26], [54] for further details.

B.1. Appell's generalized hypergeometric functions

Definition B.1. In 1880, Appell [11, 12] introduced the concept of a generalized hypergeometric function

$$ {}_2F_2(a,b,b';c,c';x,y)=\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(a)_{j+k}\frac{(b)_j(b')_k}{(c)_j(c')_k}\frac{x^jy^k}{j!k!} $$

Proposition B.2. *(i)* ${}_2F_2(a,b,b';c,c';x,y)$ *is a solution to the equation*

$$ x(1-x)z_{xx}-xyz_{xy}+(c-(a+b+1)x)z_x-byz_y-abz=0; $$

(ii) ${}_2F_2(a,b,b';c,c';x,y)$ *is a solution to the equation*

$$ y(1-y)z_{yy}-xyz_{xy}+(c'-(a+b'+1)x)z_y-b'xz_x-ab'z=0; $$

(iii) ${}_2F_2(a,b,b';c,c';x,y)$ *is a solution to the equation*

$$ \begin{aligned} x(1-x)z_{xx}-&2xyz_{xy}+y(1-y)z_{yy}+(c-(a+b+b'+1)x)z_x \\ &+(c'-(a+b+b'+1)y)z_y-a(b+b')z=0. \end{aligned} $$

Proof. Firstly, note that we need only to prove the first assertion, the second follows immediately by interchanging x and y, b and b', c and c' respectively, the last assertion can be obtained when we add the first two assertions.

We use the method of equality of coefficients to prove the first assertion as follows. We denote by $[v][j,k]$ the coefficient of $x^j y^k$ in the Taylor's expansion at $(0,0)$ of v. Then for

$$v=\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(a)_{j+k}\frac{(b)_j(b')_k}{(c)_j(c')_k}\frac{x^jy^k}{j!k!},$$

we have

- $[v][j,k]=\frac{(a)_{j+k}(b)_j(b')_k}{(c)_j(c')_kj!k!}$;
- $[v_x][j,k]=\frac{(a)_{j+k+1}(b)_{j+1}(b')_k}{(c)_{j+1}(c')_kj!k!}$;
- $[v_y][j,k]=\frac{(a)_{j+k+1}(b)_j(b')_{k+1}}{(c)_{j+1}(c')_{k+1}j!k!}$;
- $[v_{xx}][j,k]=\frac{(a)_{j+k+2}(b)_{j+2}(b')_k}{(c)_{j+2}(c')_kj!k!}$;
- $[v_{xy}][j,k]=\frac{(a)_{j+k+2}(b)_{j+1}(b')_{k+1}}{(c)_{j+1}(c')_{k+1}j!k!}$;
- $[xv_x][j,k]=\frac{(a)_{j+k}(b)_j(b')_k}{(c)_j(c')_k(j-1)!k!}$;
- $[yv_y][j,k]=\frac{(a)_{j+k}(b)_j(b')_k}{(c)_j(c')_kj!(k-1)!}$;
- $[xv_{xx}][j,k]=\frac{(a)_{j+k+1}(b)_{j+1}(b')_k}{(c)_{j+1}(c')_k(j-1)!k!}$;
- $[x^2v_{xx}][j,k]=\frac{(a)_{j+k}(b)_j(b')_k}{(c)_j(c')_k(j-2)!k!}$;

Therefore,

$$\begin{aligned}
&[x(1-x)v_{xx}-xyv_{xy}+(c-(a+b+1)x)v_x-byv_y-abv][j,k]\\
&\quad=[xv_{xx}][j,k]-[xv_{xx}][j,k]-[xyv_{xy}][j,k]\\
&\quad+c[v_x][j,k]-(a+b+1)[xv_x][j,k]-b[yv_y][j,k]-ab[v][j,k]\\
&\quad=\frac{(a)_{j+k}(b)_j(b')_k}{(c)_j(c')_kj!k!}\left\{\frac{(a+j+k)(b+j)}{c+j}j-j(j-1)-jk\right.\\
&\quad\left.+c\frac{(a+j+k)(b+j)}{c+j}-(a+b+1)j-bk-ab\right\}\\
&\quad=0.
\end{aligned}\tag{B.1.1}$$

This holds for all $j,k\geq 0$, thus

$$x(1-x)v_{xx}-xyv_{xy}+(c-(a+b+1)x)v_x-byv_y-abv=0;$$

i.e. v is the solution of the first equation. This completes the proof. $\square$

B.2. Lauricella′s generalized hypergeometric functions

Definition B.3.

$$\begin{aligned}&{}_2F_n(a, b_1, \dots, b_n; c_1, \dots c_n; x^1, \dots, x^n)\\&\qquad= \sum_{m_1,\dots,m_n=0}^{\infty} \frac{(a)_{m_1+\ldots+m_n}(b_1)_{m_1}\dots(b_n)_{m_n}}{(c_1)_{m_1}\dots(c_n)_{m_n}} \frac{(x^1)^{m_1}}{m_1!}\cdots\frac{(x^n)^{m_n}}{m_n!}\end{aligned}$$

Proposition B.4. *(i)* ${}_2F_n(a, b_1, \dots, b_n; c_1, \dots c_n; x^1, \dots, x^n)$ *is a solution to the equation*

$$\mathcal{A}_i z = \sum_{j=1}^{n} x^i(\delta_{ij} - x^j) z_{x^i x^j} + \Big(c_i - (a+1)x^i\Big) z_{x^i} - b_i \sum_{j=1}^{n} x^j z_{x^j} - ab_i z = 0;\ \forall i = \overline{1,n}$$

(ii) ${}_2F_n(a, b_1, \dots, b_n; c_1, \dots c_n; x^1, \dots, x^n)$ *is a solution to the equation*

$$\sum_{i,j=1}^{n} x^i(\delta_{ij} - x^j) z_{x^i x^j} + \sum_{i=1}^{n} \Big(c_i - (a + b_1 + \dots + b_n + 1)x^i\Big) z_{x^i} - a(b_1 + \ldots + b_n) z = 0.$$

Proof. It is easy to see that the second assertion follows by adding the equalities in the first assertion for varying i. So we need only show the first assertion.

Similar to the case of two variables, for any given $(m_1, \dots, m_n)$, we denote by $[v] = [v][m_1, \dots, m_n]$ the coefficient of $(x^1)^{m_1}\dots(x^n)^{m_n}$ in the Taylor's expansion at $(\underbrace{0, \dots, 0}_{n})$. Then we have

- $[v] = \frac{(a)_{m_1+\ldots m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_n)_{m_n} m_1!\ldots m_n!}$;
- $[v_{x^i}] = \frac{(a)_{m_1+\ldots m_n+1}(b_1)_{m_1}\ldots(b_i)_{m_i+1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_i)_{m_i+1}\ldots(c_n)_{m_n} m_1!\ldots(m_i-1)!\ldots m_n!}$;
- $[v_{x^i x^j}] = \frac{(a)_{m_1+\ldots m_n+2}(b_1)_{m_1}\ldots(b_i)_{m_i+1}\ldots(b_j)_{m_j+1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_i)_{m_i+1}\ldots(c_j)_{m_j+1}\ldots(c_n)_{m_n} m_1!\ldots(m_i-1)!\ldots(m_j-1)!\ldots m_n!}$; for $i \neq j$
- $[v_{x^i x^i}] = \frac{(a)_{m_1+\ldots m_n+2}(b_1)_{m_1}\ldots(b_i)_{m_i+2}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_i)_{m_i+1}\ldots(c_n)_{m_n} m_1!\ldots(m_i-2)!\ldots m_n!}$;
- $[x^i v_{x^i x^i}] = \frac{(a)_{m_1,\ldots,m_n+1}(b_1)_{m_1}\ldots(b_i)_{m_i+1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_i)_{m_i+1}\ldots(c_n)_{m_n} m_1!\ldots(m_i-1)!\ldots m_n!}$;
- $[(x^i)^2 v_{x^i x^i}] = \frac{(a)_{m_1,\ldots,m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_n)_{m_n} m_1!\ldots(m_i-2)!\ldots m_n!}$;
- $[x^i x^j v_{x^i x^j}] = \frac{(a)_{m_1,\ldots,m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_n)_{m_n} m_1!\ldots(m_i-1)!\ldots(m_j-1)!\ldots m_n!}$; for $i \neq j$
- $[x^i v_{x^i}] = \frac{(a)_{m_1,\ldots,m_n+1}(b_1)_{m_1}\ldots(b_i)_{m_i+1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_i)_{m_i+1}\ldots(c_n)_{m_n} m_1!\ldots m_n!}$;

Therefore

$$\begin{aligned}
[\mathcal{A}_i v] &= [x^i v_{x^i x^i}] - [(x^i)^2 v_{x^i x^i}] - \sum_{j\neq i}[x^i x^j v_{x^i x^j}] + c_i[v_{x^i}] \\
&\quad - (a+1)[x^i v_{x^i}] - b_i \sum_{j=1}^{n}[x^j v_{x^j}] - ab_i[v] \\
&= K\Bigg\{ \frac{(a+m_1,\dots,m_n)(b_i+m_i)}{c_i+m_i} m_i - m_i(m_i-1) - \sum_{j\neq i} m_i m_j \\
&\quad + c_i \frac{(a+m_1,\dots,m_n)(b_i+m_i)}{c_i+m_i} - (a+1)m_i - b_i \sum_{j=1}^{n} m_j - ab_i \Bigg\} \\
&= 0
\end{aligned}$$

This holds for all $(m_1,\dots,m_n)$ so

$$\mathcal{A}_i v = 0,$$

i.e. v is the solution of the first equation. This completes the proof. $\square$

B.3. Biorthogonal systems

In 1881, Appell [13] introduced the polynomials

$$F_{m,n}(\alpha,\gamma,\gamma';x,y) = \frac{[w(x,y)]^{-1}}{(\gamma)_m(\gamma')_n}\frac{\partial^{m+n}}{\partial x^m \partial y^n}\left(w(x,y)x^m y^n(1-x-y)^{m+n}\right)$$

in connection with analysis of polynomials orthogonal with respect to the weight function

$$w(x,y) = x^{\gamma-1}y^{\gamma'-1}(1-x-y)^{\alpha-\gamma-\gamma'}$$

in the triangle $T = \overline{V_2}$.

In 1882, he also proved that in the special case $\alpha = \gamma+\gamma'$, two families of functions

$$F_{m,n}(\gamma,\gamma';x,y) = \frac{[t(x,y)]^{-1}}{(\gamma)_m(\gamma')_n}\frac{\partial^{m+n}}{\partial x^m \partial y^n}\left(t(x,y)x^m y^n(1-x-y)^{m+n}\right)$$

and

$$E_{m,n}(\gamma,\gamma';x,y) = {}_2F_2(\gamma+\gamma'+m+n,-m,-n;\gamma,\gamma';x,y)$$

form a biorthogonal system with the weight function

$$t(x,y) = x^{\gamma-1}y^{\gamma'-1},$$

i.e.

$$\begin{aligned}
&\iint_T t(x,y)F_{m,n}(\gamma,\gamma';x,y)E_{k,l}(\gamma,\gamma';x,y)dxdy \\
&\qquad = \frac{\delta_{mk}\delta_{nl}\Gamma(\gamma)\Gamma(\gamma')m!n!(m+n)!}{(\gamma+\gamma'+2m+2n)\Gamma(\gamma+\gamma'+m+n)(\gamma)_m(\gamma')_n}
\end{aligned}$$

Definition B.5. In n dimensions, we call

$$w(x^1,\ldots,x^n) = (x^1)^{\gamma_1-1}\ldots(x^n)^{\gamma_n-1}(1-x^1-\ldots-x^n)^{\alpha-\gamma_1-\ldots-\gamma_n}$$

the weight function on V_n and

$$\begin{aligned}&\mathcal{E}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n)\\ &\qquad = \quad {}_2F_n(\alpha+m_1+\ldots+m_n,-m_1,\ldots,-m_n;\gamma_1,\ldots\gamma_n;x^1,\ldots,x^n)\end{aligned}$$

and

$$\begin{aligned}\mathcal{F}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n) &= \frac{[w(x^1,\ldots,x^n)]^{-1}}{(\gamma_1)_{m_1}\ldots(\gamma_n)_{m_n}}\left\{\frac{\partial^{m_1+\ldots+m_n}}{\partial(x^1)^{m_1}\ldots\partial(x^n)^{m_n}}\right.\\ &\left.\left(w(x^1,\ldots,x^n)(x^1)^{m_1}\ldots(x^n)^{m_n}(1-x^1-\ldots-x^n)^{m_1+\ldots+m_n}\right)\right\}\end{aligned}$$

corresponding biorthogonal systems.

We have some properties of two families of hypergeometric functions as follows

Proposition B.6. *Each*

$$\mathcal{F}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n)$$

and

$$\mathcal{E}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n)$$

is a solution of the equation

$$\sum_{i,j=1}^{n} x^i(\delta_{ij}-x^j)z_{x^ix^j}+\sum_{i=1}^{n}\Big(\gamma_i-(\alpha+1)x^i\Big)z_{x^i}+\Big(m_1+\ldots+m_n\Big)\Big(\alpha+m_1+\ldots+m_n\Big)z=0.$$

Proof. From Proposition B.4, it is easy to see that

$$\mathcal{E}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n)$$

is the solution to the equation

$$\sum_{i,j=1}^{n} x^i(\delta_{ij}-x^j)z_{x^ix^j}+\sum_{i=1}^{n}\Big(\gamma_i-(\alpha+1)x^i\Big)z_{x^i}+\Big(m_1+\ldots+m_n\Big)\Big(\alpha+m_1+\ldots+m_n\Big)z=0.$$

To prove the other assertion we proceed as follows. First, note that, due to the Taylor's expansion at $\underbrace{(0,\ldots,0)}_{n}$ we have

$$(1-x_1-\ldots-x_n)^a = \sum_{i_1,\ldots,i_n=0}^{\infty}(-a)_{i_1+\ldots+i_n}\frac{(x^1)^{i_1}\ldots(x^n)^{i_n}}{i_1!\ldots i_n!}$$

Therefore

$$\begin{aligned}
&\mathcal{F}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n)\\
&\quad=\frac{[w(x^1,\ldots,x^n)]^{-1}}{(\gamma_1)_{m_1}\ldots(\gamma_n)_{m_n}}\Bigg\{\frac{\partial^{m_1+\ldots+m_n}}{\partial(x^1)^{m_1}\ldots\partial(x^n)^{m_n}}\\
&\qquad\Bigg(w(x^1,\ldots,x^n)(x^1)^{m_1}\ldots(x^n)^{m_n}(1-x^1-\ldots-x^n)^{m_1+\ldots+m_n}\Bigg)\Bigg\}\\
&\quad=\frac{[w(x^1,\ldots,x^n)]^{-1}}{(\gamma_1)_{m_1}\ldots(\gamma_n)_{m_n}}\Bigg\{\frac{\partial^{m_1+\ldots+m_n}}{\partial(x^1)^{m_1}\ldots\partial(x^n)^{m_n}}\\
&\qquad\Bigg((x^1)^{m_1+\gamma_1-1}\ldots(x^n)^{m_n+\gamma_n-1}(1-x^1-\ldots-x^n)^{m_1+\ldots+m_n+\alpha-\gamma_1-\ldots-\gamma_n}\Bigg)\Bigg\}\\
&\quad=\frac{[w(x^1,\ldots,x^n)]^{-1}}{(\gamma_1)_{m_1}\ldots(\gamma_n)_{m_n}}\Bigg\{\frac{\partial^{m_1+\ldots+m_n}}{\partial(x^1)^{m_1}\ldots\partial(x^n)^{m_n}}\\
&\qquad\Bigg(\sum_{i_1,\ldots,i_n=0}^{\infty}(\gamma_1+\ldots+\gamma_n-m_1-\ldots-m_n-\alpha)_{i_1+\ldots+i_n}\\
&\qquad\times\frac{(x^1)^{m_1+\gamma_1-1+i_1}\ldots(x^n)^{m_n+\gamma_n-1+i_n}}{i_1!\ldots i_n!}\Bigg)\Bigg\}\\
&\quad=\frac{[w(x^1,\ldots,x^n)]^{-1}}{(\gamma_1)_{m_1}\ldots(\gamma_n)_{m_n}}\Bigg(\sum_{i_1,\ldots,i_n=0}^{\infty}(\gamma_1+\ldots+\gamma_n-m_1-\ldots-m_n-\alpha)_{i_1+\ldots+i_n}\\
&\qquad\times(\gamma_1+i_1)_{m_1}\ldots(\gamma_n+i_n)_{m_n}\frac{(x^1)^{\gamma_1-1+i_1}\ldots(x^n)^{\gamma_n-1+i_n}}{i_1!\ldots i_n!}\Bigg)\\
&\quad=\sum_{i_1,\ldots,i_n=0}^{\infty}(\gamma_1+\ldots+\gamma_n-m_1-\ldots-m_n-\alpha)_{i_1+\ldots+i_n}\\
&\qquad\times\frac{(\gamma_1+i_1)_{m_1}\ldots(\gamma_n+i_n)_{m_n}}{(\gamma_1)_{m_1}\ldots(\gamma_n)_{m_n}}\frac{(x^1)^{i_1}\ldots(x^n)^{i_n}}{i_1!\ldots i_n!}(1-x_1-\ldots-x_n)^{\gamma_1+\ldots+\gamma_n-\alpha}\\
&\quad=\sum_{i_1,\ldots,i_n=0}^{\infty}(\gamma_1+\ldots+\gamma_n-m_1-\ldots-m_n-\alpha)_{i_1+\ldots+i_n}\\
&\qquad\times\frac{(\gamma_1+m_1)_{i_1}\ldots(\gamma_n+m_n)_{i_n}}{(\gamma_1)_{i_1}\ldots(\gamma_n)_{i_n}}\frac{(x^1)^{i_1}\ldots(x^n)^{i_n}}{i_1!\ldots i_n!}(1-x_1-\ldots-x_n)^{\gamma_1+\ldots+\gamma_n-\alpha}\\
&\quad=(1-x_1-\ldots-x_n)^{\gamma_1+\ldots+\gamma_n-\alpha}\,{}_2F_n\Big(\gamma_1+\ldots+\gamma_n-m_1-\ldots-m_n-\alpha,\\
&\qquad\gamma_1+m_1,\ldots,\gamma_n+m_n;\gamma_1,\ldots\gamma_n;x^1,\ldots,x^n\Big).
\end{aligned}$$

To simplify the notations, we denote by

$$f(x)=\mathcal{F}_{m_1,\ldots,m_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n),$$

$$g(x)={}_2F_n(\gamma_1+\ldots+\gamma_n-m_1-\ldots-m_n-\alpha,\gamma_1+m_1,\ldots,\gamma_n+m_n;\gamma_1,\ldots\gamma_n;x^1,\ldots,x^n),$$

and

$$\phi(x) = (1 - x_1 - \ldots - x_n)^{\gamma_1 + \ldots + \gamma_n - \alpha}.$$

Then we have

$$f(x) = \phi(x) g(x).$$

It follows

$$\begin{aligned}
&\sum_{i,j=1}^{n} x^i(\delta_{ij} - x^j) f_{x^i x^j} + \sum_{i=1}^{n} \Big(\gamma_i - (\alpha+1)x^i\Big) f_{x^i} + \Big(m_1 + \ldots + m_n\Big)\Big(\alpha + m_1 + \ldots + m_n\Big) f \\
&\quad = \sum_{i,j=1}^{n} x^i(\delta_{ij} - x^j)\Big(\phi g_{x^i x^j} + \phi_{x^i} g_{x^j} + \phi_{x^j} g_{x^i} + \phi_{x^i x^j} g\Big) \\
&\quad + \sum_{i=1}^{n} \Big(\gamma_i - (\alpha+1)x^i\Big)\Big(\phi g_{x^i} + \phi_{x^i} g\Big) + \Big(m_1 + \ldots + m_n\Big)\Big(\alpha + m_1 + \ldots + m_n\Big)(\phi g) \\
&\quad = \sum_{i,j=1}^{n} x^i(\delta_{ij} - x^j)\phi g_{x^i x^j} + \sum_{i=1}^{n} \Big((\gamma_i - (\alpha+1)x^i)\phi + 2\sum_{j=1}^{n} x^i(\delta_{ij} - x^j)\phi_{x^j}\Big) g_{x^i} \\
&\quad + \Bigg(\sum_{i,j=1}^{n} x^i(\delta_{ij} - x^j)\phi_{x^i x^j} + \sum_{i=1}^{n} \Big((\gamma_i - (\alpha+1)x^i\Big)\phi_{x^i} \\
&\quad + \Big(m_1 + \ldots + m_n\Big)\Big(\alpha + m_1 + \ldots + m_n\Big)\phi\Bigg) g \\
&\quad = \phi\Bigg(\sum_{i,j=1}^{n} x^i(\delta_{ij} - x^j) g_{x^i x^j} + \sum_{i=1}^{n} \Big(\gamma_i - \big(2\gamma_1 + \ldots + 2\gamma_n - \alpha + 1\big)x^i\Big) g_{x^i} \\
&\quad - \Big(\gamma_1 + \ldots + \gamma_n - m_1 - \ldots - m_n - \alpha\Big)\Big(\gamma_1 + \ldots + \gamma_n + m_1 + \ldots + m_n\Big) g\Bigg) \\
&\quad = 0 \quad \text{by Proposition B.4}
\end{aligned}$$

This completes the proof. □

Proposition B.7. *The generalized hypergeometric function*

$$\mathcal{E}_{m'_1, \ldots, m'_n}(\alpha, \gamma_1, \ldots, \gamma_n; x^1, \ldots, x^n)$$

is a mono polynomial of degree $m'_1 + \ldots + m'_n$ of the form

$$(-1)^{m'_1 + \ldots + m'_n} \frac{(\alpha + m'_1 + \ldots + m'_n)_{m'_1 + \ldots + m'_n}}{(\gamma_1)_{m'_1} \cdots (\gamma_n)_{m'_n}} (x^1)^{m'_1} \ldots (x^n)^{m'_n} \\
+ \Big\{ \textit{polynomial of degree } < m'_1 + \ldots + m'_n \Big\}.$$

Proof. In fact, from the definition of $\mathcal{E}$ we have

$$
\begin{aligned}
&\mathcal{E}_{m'_1,\ldots,m'_n}(\alpha,\gamma_1,\ldots,\gamma_n;x^1,\ldots,x^n)\\
&= \quad {}_2F_n(\alpha+m'_1+\ldots+m'_n,-m'_1,\ldots,-m'_n;\gamma_1,\ldots\gamma_n;x^1,\ldots,x^n)\\
&= \sum_{i_1,\ldots,i_n=0}^{\infty}(\alpha+m'_1+\ldots+m'_n)_{i_1+\ldots+i_n}\frac{(-m'_1)_{i_1}\ldots(-m'_n)_{i_n}}{(\gamma_1)_{i_1}\ldots(\gamma_n)_{i_n}i_1!\ldots i_n!}(x^1)^{i_1}\ldots(x^n)^{i_n}\\
&= \sum_{i_1=0}^{m'_1}\ldots\sum_{i_n=0}^{m'_n}(\alpha+m'_1+\ldots+m'_n)_{i_1+\ldots+i_n}\frac{(-m'_1)_{i_1}\ldots(-m'_n)_{i_n}}{(\gamma_1)_{i_1}\ldots(\gamma_n)_{i_n}i_1!\ldots i_n!}(x^1)^{i_1}\ldots(x^n)^{i_n}.\\
&\quad\Big(\text{because if } i_k>m'_k \text{ then } (-m'_k)_{i_k}=0\Big)\\
&= (\alpha+m'_1+\ldots+m'_n)_{m'_1+\ldots+m'_n}\frac{(-m'_1)_{m'_1}\ldots(-m'_n)_{m'_n}}{(\gamma_1)_{m'_1}\ldots(\gamma_n)_{m'_n}m'_1!\ldots m'_n!}(x^1)^{m'_1}\ldots(x^n)^{m'_n}\\
&+\Big\{\text{polynomial of degree } <m'_1+\ldots+m'_n\Big\}\\
&= (-1)^{m'_1+\ldots+m'_n}\frac{(\alpha+m'_1+\ldots+m'_n)_{m'_1+\ldots+m'_n}}{(\gamma_1)_{m'_1}\ldots(\gamma_n)_{m'_n}}(x^1)^{m'_1}\ldots(x^n)^{m'_n}\\
&+\Big\{\text{polynomial of degree } <m'_1+\ldots+m'_n\Big\}.
\end{aligned}
$$

This completes the proof. $\square$

Lemma B.8. *With $a_1,\ldots,a_{n+1}$ given positive numbers, we have*

$$\int_{V_n}(x^1)^{a_1-1}\ldots(x^n)^{a_n-1}(1-x_1-\ldots-x_n)^{a_{n+1}-1}d\mathbf{x}=\frac{\Gamma(a_1)\ldots\Gamma(a_{n+1})}{\Gamma(a_1+\ldots+a_{n+1})}.$$

Proof. It is obvious to see that when $a,b>0$

$$\int_0^r x^{a-1}(r-x)^{b-1}dx=\int_0^1(ry)^{a-1}[r(1-y)]^{b-1}rdy=r^{a+b-1}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

So we have

$$\int_{V_n} (x^1)^{a_1-1} \dots (x^n)^{a_n-1}(1-x^1-\dots-x^n)^{a_{n+1}-1} d\mathbf{x}$$

$$= \int_0^1 (x^1)^{a_1-1} \left\{ \int_0^{1-x^1} (x^2)^{a_2-1} \right.$$

$$\left. \dots \left(\int_0^{1-x^1-\dots-x^{n-1}} \left(x^n\right)^{a_n-1} \left(1-x^1-\dots-x^{n-1}-x^n\right)^{a_{n+1}-1} dx^n \right) \dots dx^2 \right\} dx^1$$

$$= \int_0^1 (x^1)^{a_1-1} \left\{ \int_0^{1-x^1} (x^2)^{a_2-1} \right.$$

$$\left. \dots \left(\left(1-x^1-\dots-x^{n-1}\right)^{a_n+a_{n+1}-1} \frac{\Gamma(a_n)\Gamma(a_{n+1})}{\Gamma(a_n+a_{n+1})} \right) \dots dx^2 \right\} dx^1$$

$$= \cdots$$

$$= \int_0^1 (x^1)^{a_1-1} \left\{ \int_0^{1-x^1} (x^2)^{a_2-1}(1-x^1-x^2)^{a_3+\dots+a_{n+1}-1} \frac{\Gamma(a_3)\dots\Gamma(a_{n+1})}{\Gamma(a_3+\dots+a_{n+1})} dx^2 \right\} dx^1$$

$$= \int_0^1 (x^1)^{a_1-1}(1-x^1)^{a_2+\dots+a_{n+1}-1} \frac{\Gamma(a_2)\Gamma(a_3+\dots+a_{n+1})}{\Gamma(a_2+\dots+a_{n+1})} \frac{\Gamma(a_3)\dots\Gamma(a_{n+1})}{\Gamma(a_3+\dots+a_{n+1})} dx^1$$

$$= \int_0^1 (x^1)^{a_1-1}(1-x^1)^{a_2+\dots+a_{n+1}-1} \frac{\Gamma(a_2)\dots\Gamma(a_{n+1})}{\Gamma(a_2+\dots+a_{n+1})} dx^1$$

$$= \frac{\Gamma(a_1)\Gamma(a_2+\dots+a_{n+1})}{\Gamma(a_1+\dots+a_{n+1})} \frac{\Gamma(a_2)\dots\Gamma(a_{n+1})}{\Gamma(a_2+\dots+a_{n+1})}$$

$$= \frac{\Gamma(a_1)\dots\Gamma(a_{n+1})}{\Gamma(a_1+\dots+a_{n+1})}.$$

This completes the proof. □

Proposition B.9. *The system of*

$$\mathcal{F}_{m_1,\dots,m_n}(\alpha,\gamma_1,\dots,\gamma_n;x^1,\dots x^n)$$

and

$$\mathcal{E}_{m_1,\dots,m_n}(\alpha,\gamma_1,\dots,\gamma_n;x^1,\dots x^n)$$

is a biorthogonal system with the weight function w, i.e.

$$\int_{V_n} w(x^1,\dots,x^n)\mathcal{F}_{m_1,\dots,m_n}\mathcal{E}_{m'_1,\dots,m'_n} dx^1 \dots dx^n = K_{m_1,\dots,m_n}\delta_{m_1,m'_1}\dots\delta_{m_n,m'_n}.$$

where

$$K_{m_1,\dots m_n} = \frac{(\alpha + m_1 + \ldots + m_n)_{m_1+\ldots+m_n} m_1! \ldots m_n!}{[(\gamma_1)_{m_1}]^2 \ldots [(\gamma_n)_{m_n}]^2} \times \frac{\Gamma(m_1+\gamma_1)\ldots\Gamma(m_n+\gamma_n)\Gamma(m_1+\ldots+m_n+\alpha-\gamma_1-\ldots-\gamma_n+1)}{\Gamma(2m_1+\ldots+2m_n+\alpha+1)}.$$

Proof. From Proposition B.6 we have

$$\sum_{i,j=1}^{n} x^i(\delta_{ij}-x^j)u_{x^ix^j} + \sum_{i=1}^{n}\Big(\gamma_i-(\alpha+1)x^i\Big)u_{x^i} + \Big(m_1+\ldots+m_n\Big)\Big(\alpha+m_1+\ldots+m_n\Big)u = 0. \tag{B.3.1}$$

where

$$u = \mathcal{F}_{m_1,\dots,m_n}(\alpha,\gamma_1,\dots,\gamma_n;x^1,\dots,x^n)$$

and

$$\sum_{i,j=1}^{n} x^i(\delta_{ij}-x^j)v_{x^ix^j} + \sum_{i=1}^{n}\Big(\gamma_i-(\alpha+1)x^i\Big)v_{x^i} + \Big(m_1'+\ldots+m_n'\Big)\Big(\alpha+m_1'+\ldots+m_n'\Big)v = 0. \tag{B.3.2}$$

where

$$v = \mathcal{E}_{m_1',\dots,m_n'}(\alpha,\gamma_1,\dots,\gamma_n;x^1,\dots,x^n)$$

Multiplying equation B.3.2 by u and subtracting equation B.3.1 multiplied by v, we obtain

$$\sum_{i,j=1}^{n} x^i(\delta_{ij}-x^j)\Big(uv_{x^ix^j}-u_{x^ix^j}v\Big) + \sum_{i=1}^{n}\Big(\gamma_i-(\alpha+1)x^i\Big)\Big(uv_{x^i}-u_{x^i}v\Big)$$
$$= \Big(\alpha+m_1+\ldots+m_n+m_1'+\ldots+m_n'\Big)\Big(m_1+\ldots+m_n-m_1'-\ldots-m_n'\Big)uv.$$

Multiplying both sides of above equations by w and integrating over V_n we obtain

$$\int\limits_{V_n}\big(\alpha+m_1+\ldots+m_n+m_1'+\ldots+m_n'\big)\big(m_1+\ldots+m_n-m_1'-\ldots-m_n'\big)wuv\,d\mathbf{x}$$
$$= \int\limits_{V_n}\Big(\sum_{i,j=1}^{n} x^i(\delta_{ij}-x^j)\big(wuv_{x^ix^j}-wu_{x^ix^j}v\big) + \sum_{i=1}^{n}\big(\gamma_i-(\alpha+1)x^i\big)\big(wuv_{x^i}-wu_{x^i}v\big)\Big)d\mathbf{x}$$
$$= \int\limits_{V_n}\sum_{j=1}^{n}\frac{\partial}{\partial x^j}\left(\sum_{i=1}^{n} x^i(\delta_{ij}-x^j)w(uv_{x^i}-u_{x^i}v)\right)d\mathbf{x}$$
$$= \int\limits_{V_n}\operatorname{div} F\,d\mathbf{x}, \qquad \text{where } F^j = \sum_{i=1}^{n} x^i\Big(\delta_{ij}-x^j\Big)w\Big(uv_{x^i}-u_{x^i}v\Big)$$
$$= \int\limits_{\partial V_n} F\cdot\nu\,do(\sigma)$$
$$= 0, \qquad \text{since } F_{|\partial V_n} = 0 \quad \text{follows from } w_{|\partial V_n} = 0.$$

It follows that if

$$m_1 + \ldots + m_n \neq m_1' + \ldots + m_n'$$

then

$$\int\limits_{V_n} wuv d\mathbf{x} = 0.$$

Now we consider the case $m_1 + \ldots + m_n = m_1' + \ldots + m_n'$. Applying the integration by parts to the Proposition B.7 and Lemma B.8, we obtain

$$\begin{aligned}
&\int\limits_{V_n} wuv d\mathbf{x} \\
&= \int\limits_{V_n} w \frac{w^{-1}}{(\gamma_1)_{m_1} \ldots (\gamma_n)_{m_n}} \left\{ \frac{\partial^{m_1+\ldots+m_n}}{\partial (x^1)^{m_1} \ldots \partial (x^n)^{m_n}} \left((x^1)^{m_1+\gamma_1-1} \ldots (x^n)^{m_n+\gamma_n-1} \right.\right. \\
&\quad \left.\left. \times (1 - x^1 - \ldots - x^n)^{m_1+\ldots+m_n+\alpha-\gamma_1-\ldots-\gamma_n} \right) \right\} v d\mathbf{x} \\
&= \frac{(-1)^{m_1+\ldots+m_n}}{(\gamma_1)_{m_1} \ldots (\gamma_n)_{m_n}} \int\limits_{V_n} \left((x^1)^{m_1+\gamma_1-1} \ldots (x^n)^{m_n+\gamma_n-1} \right. \\
&\quad \left. \times (1 - x^1 - \ldots - x^n)^{m_1+\ldots+m_n+\alpha-\gamma_1-\ldots-\gamma_n} \right) \frac{\partial^{m_1+\ldots+m_n} v}{\partial (x^1)^{m_1} \ldots \partial (x^n)^{m_n}} d\mathbf{x} \\
&= \frac{(-1)^{m_1+\ldots+m_n}}{(\gamma_1)_{m_1} \ldots (\gamma_n)_{m_n}} \int\limits_{V_n} \left((x^1)^{m_1+\gamma_1-1} \ldots (x^n)^{m_n+\gamma_n-1} \right. \\
&\quad \left. \times (1 - x^1 - \ldots - x^n)^{m_1+\ldots+m_n+\alpha-\gamma_1-\ldots-\gamma_n} \right) \\
&\delta_{m_1,m_1'} \ldots \delta_{m_n,m_n'} \frac{(-1)^{m_1+\ldots+m_n} (\alpha + m_1 + \ldots + m_n)_{m_1+\ldots+m_n} m_1! \ldots m_n!}{(\gamma_1)_{m_1} \ldots (\gamma_n)_{m_n}} d\mathbf{x} \\
&= \delta_{m_1,m_1'} \ldots \delta_{m_n,m_n'} \frac{(\alpha + m_1 + \ldots + m_n)_{m_1+\ldots+m_n} m_1! \ldots m_n!}{[(\gamma_1)_{m_1}]^2 \ldots [(\gamma_n)_{m_n}]^2} \\
&\quad \times \frac{\Gamma(m_1+\gamma_1) \ldots \Gamma(m_n+\gamma_n) \Gamma(m_1 + \ldots + m_n + \alpha - \gamma_1 - \ldots - \gamma_n + 1)}{\Gamma(2m_1 + \ldots + 2m_n + \alpha + 1)}.
\end{aligned}$$

This completes the proof. □

Corollary B.10. *When $\alpha = 2n+1, \gamma_1 = \ldots = \gamma_n = 2$ we have the result obtained in [54]* *by Littler and Fackerell in 1975*

$$
\begin{aligned}
K_{m_1,\ldots,m_n} = & \frac{(2n+1+m_1+\ldots+m_n)_{m_1+\ldots+m_n} m_1!\ldots m_n!}{[(2)_{m_1}]^2\ldots[(2)_{m_n}]^2} \\
& \times \frac{\Gamma(m_1+2)\ldots\Gamma(m_n+2)\Gamma(m_1+\ldots+m_n+2)}{\Gamma(2m_1+\ldots+2m_n+2n+2)}. \\
= & \frac{(2n+1+m_1+\ldots+m_n)_{m_1+\ldots+m_n} m_1!\ldots m_n!}{[(m_1+1)!]^2\ldots[(m_n+1)!]^2} \\
& \times \frac{(m_1+1)!\ldots(m_n+1)!(m_1+\ldots+m_n+1)!}{(2m_1+\ldots+2m_n+2n+1)!} \\
= & \frac{1}{(m_1+1)\ldots(m_n+1)} \\
& \times \frac{(2n+1+m_1+\ldots+m_n)_{m_1+\ldots+m_n}(m_1+\ldots+m_n+1)!}{(2m_1+\ldots+2m_n+2n+1)!} \\
= & \frac{1}{(m_1+1)\ldots(m_n+1)} \\
& \times \frac{1}{(m_1+\ldots+m_n+2)\ldots(m_1+\ldots+m_n+2n)(2m_1+\ldots+2m_n+2n+1)} \\
= & \frac{1}{(m_1+1)\ldots(m_n+1)(m_1+\ldots+m_n+2)_{2n-1}(2m_1+\ldots+2m_n+2n+1)}.
\end{aligned}
$$

Appendix C

Introduction to Information Geometry

This appendix gives the basics of Information Geometry used in Chapter 6.

Information geometry is a bridge connecting between non-Euclidean geometry and probability theory which reached maturity through the work of Amari in 1980 (see [2]). The main idea is to find out the correspondence between structure of the families of distributions and that of manifolds. Formally, we can consider a distribution as a point, the score as a tangent vector, a family of distributions as a Riemannian manifold with the Riemannian metric is the Fisher information metric, etc., with thinking that their results are interchangeable to each other. For further details, we refer readers to [2], [3], [4], [37], [66].

C.1. Family of probability distributions as a Riemannian manifold

C.1.1. Probability distributions as points

Note that almost all of the popular probability distributions depend on some parameters which constitute a domain in an Euclidean space. For example, writing a Gaussian distribution in one dimension $\mathcal{N}(\mu, \sigma)$ in the general form

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

implies parameters $\boldsymbol{\theta} = (\mu, \sigma)$ here are in $\mathcal{M} = \mathbb{R} \times \mathbb{R}_{(>0)}$. We then can consider each Gaussian distribution as a point in $\mathcal{M}$ with coordinates μ, σ.

Remark C.1. Note that the parametrization of a Gaussian distribution is not unique. For example, we can choose the parameters $\boldsymbol{\theta} = \left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right)$ for the above Gaussian distribution. This is also understood as one point can have many local systems of coordinates (local charts).

C.1.2. Families of probability distributions as manifolds

We extend a family of distributions $\mathcal{P} = \{p(\mathbf{x};\boldsymbol{\theta})\}$ to a manifold $\mathcal{M}$ such that the points $\mathbf{p} \in \mathcal{M}$ are in a one-to-one relation with the distributions p. The parameters θ^μ of $\mathcal{P}$ can thus also be used as coordinates on $\mathcal{M}$. We hope to gain some insight into the structure of such family. For example, we want to discover a reasonable measure of nearness of two distributions in the family. To simplify the notations, we denote by (in coordinate system $\boldsymbol{\theta}$)

$$\ell(\boldsymbol{\theta}) := \log p(\mathbf{x};\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}), \quad \partial_\mu := \frac{\partial}{\partial \theta^\mu}$$

and call $\partial_\mu \ell(\boldsymbol{\theta})$ *the score* of family with respect to θ^μ.
Then the tangent space $T_{\boldsymbol{\theta}}$ of $\mathcal{M}$ is seen to be isomorphic to the vector space spanned by the scores

$$T^{(1)}_{\boldsymbol{\theta}} = \mathrm{span}\{\partial_\mu \ell(\boldsymbol{\theta})\}_\mu.$$

A vector field $\mathbf{X} \in T(\mathcal{M})$,

$$\mathbf{X}(\boldsymbol{\theta}) = X^\mu(\boldsymbol{\theta}) e_\mu,$$

thus is equivalent to a random variable $\mathbf{X}^{(1)} \in T^{(1)}_{\boldsymbol{\theta}}$,

$$\mathbf{X}^{(1)}(\boldsymbol{\theta}) = X^\mu(\boldsymbol{\theta}) \partial_\mu \ell(\boldsymbol{\theta}),$$

which is called the $1-$*representation* of the vector field $\mathbf{X}$.

Now we go to define a metric

Definition C.2. The Fisher metric on a manifold of probability distributions is defined as

$$g_{\mu\nu}(\boldsymbol{\theta}) = \mathbb{E}(\partial_\mu \ell(\boldsymbol{\theta}) \partial_\nu \ell(\boldsymbol{\theta})).$$

Remark C.3. Because p is a probability distribution, it follows

$$\begin{aligned}
\mathbb{E}(\partial_\mu \partial_\nu \ell) &= \int_\Omega d\mathbf{x} p \partial_\mu \Big(\frac{1}{p}\partial_\nu p\Big) \\
&= \int_\Omega d\mathbf{x} \Big(\partial_\mu \partial_\nu p - \frac{1}{p}\partial_\mu p \partial_\nu p\Big) \\
&= \partial_\mu \partial_\nu 1 - \int_\Omega d\mathbf{x} p \partial_\mu \ell \partial_\nu \ell \\
&= - \mathbb{E}(\partial_\mu \ell(\boldsymbol{\theta}) \partial_\nu \ell(\boldsymbol{\theta})) = -g_{\mu\nu}(\boldsymbol{\theta}).
\end{aligned} \tag{C.1.1}$$

We can prove easily that the Fisher metric is invariant under transformations of the random variable and covariant under reparametrizations as follows

Proposition C.4. *The Fisher metric is invariant under transformations of the random variable.*

Proof. Suppose that the probability distributions are defined by a vector random variable $\mathbf{X}$ valued in $\Omega \subset \mathbb{R}^n$. Then

$$g_{ij}(\boldsymbol{\theta}) = \int_\Omega d\mathbf{x} \frac{1}{\mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x})} \partial_i \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \partial_j \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}).$$

Let $\mathbf{f}$ be a (bijective) transformation from the random variable $\mathbf{X}$ to the random variable $\mathbf{Y}$ valued in Ω' then we have

$$\bar{\mathbf{p}}_{\boldsymbol{\theta}}(\mathbf{y}) = \int_\Omega d\mathbf{x} \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \delta(\mathbf{y} - \mathbf{f}(\mathbf{x})).$$

Because $\mathbf{f}$ is bijective, we can use

$$\delta(\mathbf{y} - \mathbf{f}(\mathbf{x})) = \frac{1}{|\frac{\partial \mathbf{f}}{\partial \mathbf{x}}|} \delta(\mathbf{f}^{-1}(\mathbf{y}) - \mathbf{x})$$

to find that

$$\bar{\mathbf{p}}_{\boldsymbol{\theta}}(\mathbf{y}) = \int_\Omega d\mathbf{x} \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \frac{1}{|\frac{\partial \mathbf{f}}{\partial \mathbf{x}}|} \delta(\mathbf{f}^{-1}(\mathbf{y}) - \mathbf{x}) = \left[\frac{1}{|\frac{\partial \mathbf{f}}{\partial \mathbf{x}}|} \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \right]_{\mathbf{x}=\mathbf{f}^{-1}(\mathbf{y})}$$

Note that $|\frac{\partial \mathbf{f}}{\partial \mathbf{x}}|$ does not depend on $\boldsymbol{\theta}$, we obtain

$$\begin{aligned}
\int_{\Omega'} d\mathbf{y} \frac{1}{\bar{\mathbf{p}}_{\boldsymbol{\theta}}(\mathbf{y})} \partial_i \bar{\mathbf{p}}_{\boldsymbol{\theta}}(\mathbf{y}) \partial_j \bar{\mathbf{p}}_{\boldsymbol{\theta}}(\mathbf{y}) &= \int_{\Omega'} d\mathbf{y} \left[\frac{1}{\mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x})} \partial_i \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \partial_j \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \right]_{\mathbf{x}=\mathbf{f}^{-1}(\mathbf{y})} \\
&= \int_\Omega d\mathbf{x} \frac{1}{\mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x})} \partial_i \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}) \partial_j \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x}).
\end{aligned} \tag{C.1.2}$$

□

Proposition C.5. *The Fisher metric is covariant under reparametrization.*

Proof. Suppose that, we have the reparametrization $\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}(\boldsymbol{\theta})$. We define by $\bar{\mathbf{p}}_{\bar{\boldsymbol{\theta}}}(\mathbf{x}) = \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x})$ then we have

$$\begin{aligned}
\bar{g}_{ij}(\bar{\boldsymbol{\theta}}) &= \int_\Omega d\mathbf{x} \frac{1}{\bar{\mathbf{p}}_{\bar{\boldsymbol{\theta}}}(\mathbf{x})} \frac{\partial}{\partial \bar{\theta}^i} \bar{\mathbf{p}}_{\bar{\boldsymbol{\theta}}}(\mathbf{x}) \frac{\partial}{\partial \bar{\theta}^j} \bar{\mathbf{p}}_{\bar{\boldsymbol{\theta}}}(\mathbf{x}) \\
&= \left[\frac{\partial \theta^k}{\partial \bar{\theta}^i} \frac{\partial \theta^l}{\partial \bar{\theta}^j} g_{kl}(\boldsymbol{\theta}) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}(\bar{\boldsymbol{\theta}})}.
\end{aligned} \tag{C.1.3}$$

This completes the proof. □

C.1.3. Affine connections and their dual connections

To work with a statistical manifold, as usual, we need to construct a connection on it. Here we shall briefly introduce a family of affine connections based on the $1-$representation of vectors on a statistical manifold. These connections have been named $\alpha-$connections by Amari in [2].

Definition C.6. The $\alpha-$connection $\nabla^{(\alpha)}$ on a statistical manifold is defined by its coefficients (Christoffel symbols)

$$\Gamma^{(\alpha)\lambda}_{\mu\nu} = \mathbb{E}\Big(\partial_\mu\partial_\nu\ell\partial_\rho\ell + \frac{1-\alpha}{2}\partial_\mu\ell\partial_\nu\ell\partial_\rho\ell\Big)g^{\rho\lambda}.$$

or equivalent

$$\Gamma^{(\alpha)}_{\mu\nu\rho} = \mathbb{E}\Big(\partial_\mu\partial_\nu\ell\partial_\rho\ell + \frac{1-\alpha}{2}\partial_\mu\ell\partial_\nu\ell\partial_\rho\ell\Big).$$

Remark C.7. It is easy to see the metric connection (the Levi-Civita connection for the Fisher metric) is same as the $0-$connection, so we can rewrite the above expression as follows

$$\Gamma^{(\alpha)}_{\mu\nu\rho} = \Gamma^{(metric)}_{\mu\nu\rho} + \alpha T_{\mu\nu\rho},$$

where

$$T_{\mu\nu\rho} = -\frac{1}{2}\mathbb{E}(\partial_\mu\ell\partial_\nu\ell\partial_\rho\ell).$$

Definition C.8. Two connections ∇ and ∇^* are said to be dual to each other if for all vector fields $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, we have

$$\mathbf{X} < \mathbf{Y}, \mathbf{Z} > = < \nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z} > + < \mathbf{Y}, \nabla^*_{\mathbf{X}}\mathbf{Z} >$$

We have some properties

Proposition C.9. *There exists always exactly one dual ∇^* to any affine connection ∇, and $\nabla^{**} = \nabla$.*

Proposition C.10. $\left(\nabla^{(\alpha)}\right)^* = \nabla^{(-\alpha)}$.

Definition C.11. A manifold is flat with respect to an affine connection when there is a coordinate system such that

$$\Gamma^\rho_{\mu\nu} = 0.$$

The following lemma links flatness with respect to a connection and flatness with respect to its dual:

Lemma C.12. *When a manifold is flat with respect to an affine connection, it is also flat with respect to its dual. We call it dually flat manifold.*

On a dually flat manifold, there exist two special coordinates systems which are dual to each other.

Definition C.13. Two coordinate systems (θ^μ) and $(\bar{\theta}_\nu)$ are said to be dual to one another when their coordinate basis vectors satisfy

$$< e_\mu, \bar{e}^\nu >= \delta_{\mu\nu},$$

where e_μ and $\bar{e}^\nu$ are the coordinate basis vectors for the $\boldsymbol{\theta}$ and $\bar{\boldsymbol{\theta}}$ systems respectively.

C.2. Some special families of probability distributions

C.2.1. Exponential families

An exponential family of probability distributions is a family of distributions that can be written as

$$\mathcal{M} = \left\{ p(\mathbf{x};\boldsymbol{\theta}) | p(\mathbf{x};\boldsymbol{\theta}) = \exp\Big\{ \sum_\mu \theta^\mu f_\mu(\mathbf{x}) + c_0(\boldsymbol{\theta}) \Big\} P_0(\mathbf{x}) \right\},$$

where $P_0(\mathbf{x})$ is some fixed measure, c_0 ensures that the distribution is normalized.

Then the score is given by

$$\partial_\mu \ell(\theta) = f_\mu(\mathbf{x}) + \partial_\mu c_0(\theta).$$

Therefore the metric will be

$$g_{\mu\nu}(\boldsymbol{\theta}) = -\partial_\mu \partial_\nu c_0(\boldsymbol{\theta}).$$

It follows the connection $\nabla^{(\alpha)}$ is given by

$$\Gamma^{(\alpha)}_{\mu\nu\rho}(\boldsymbol{\theta}) = \frac{\alpha - 1}{2} \partial_\mu \partial_\nu \partial_\rho c_0(\boldsymbol{\theta}).$$

Therefore the exponential family is $1-$flat.

C.2.2. Mixture families

A mixture family of probability distributions is a family of distributions that can be written in the form

$$\mathcal{M} = \Big\{ p(\mathbf{x};\boldsymbol{\theta}) | p(\mathbf{x};\boldsymbol{\theta}) = \sum_\mu \theta^\mu P_\mu(\mathbf{x}) + \Big(1 - \sum_\mu \theta^\mu\Big) P_0(\mathbf{x}) \Big\},$$

where P_0, $P_\mu's$ are probability distributions, and $\theta^\mu \in (0,1)$ such that $\sum_\mu \theta^\mu < 1$.

Then the score is given by

$$\partial_\mu \ell(\theta) = \frac{1}{p}\Big(P_\mu - P_0\Big).$$

Therefore the metric will be

$$g_{\mu\nu}(\boldsymbol{\theta}) = \int d\mathbf{x}\frac{1}{p}\Big(P_\mu - P_0\Big)\Big(P_\nu - P_0\Big).$$

It follows the connection $\nabla^{(\alpha)}$ is given by

$$\Gamma^{(\alpha)}_{\mu\nu\rho}(\boldsymbol{\theta}) = -\frac{\alpha+1}{2}\int d\mathbf{x}\frac{1}{p^2}\partial_\mu p \partial_\nu p \partial_\rho p.$$

Therefore the exponential family is $(-1)-$flat.

In the special case, when $P_\mu = \delta_{a_\mu}$, $P_0 = \delta_{a_0}$. The score is given by

$$\partial_\mu \ell(\theta) = \frac{1}{p}\Big(\delta_{a_\mu} - \delta_{a_0}\Big).$$

Therefore the metric will be

$$g_{\mu\nu}(\boldsymbol{\theta}) = \frac{\delta_{\mu\nu}}{\theta^\mu} + \frac{1}{1-\sum_\lambda \theta^\lambda}.$$

It follows the connection $\nabla^{(\alpha)}$ is given by

$$\Gamma^{(\alpha)}_{\mu\nu\rho}(\boldsymbol{\theta}) = \frac{\alpha+1}{2}\left\{\frac{\delta_{\mu\nu}\delta_{\mu\rho}}{\big(\theta^\mu\big)^2} + \frac{1}{\big(1-\sum_\lambda \theta^\lambda\big)^2}\right\}.$$

Therefore the mixture family is $(-1)-$flat.

List of Figures

List of Notations

Spaces

Δ_n	$\left\{(x^0, x^1, \cdots, x^n) \in \mathbb{R}^{n+1} : x^i \geq 0 \quad i = 0, \cdots, n \text{ and } \sum\limits_{i=0}^{n} x^i = 1\right\}$
	the standard simplex in $\mathbb{R}^{n+1}$
e_0	$(0, \ldots, 0) \in \mathbb{R}^n$
e_k	$(0, \ldots, \underbrace{1}_{k^{th}}, \ldots, 0) \in \mathbb{R}^n, \quad k = 1, \cdots, n$
Ω_n	$intco\{e_0, \ldots, e_n\}$
	$= \left\{\sum\limits_{k=0}^{n} x^k e_k,\ (x^0, x) = \left(1 - \sum\limits_{k=1}^{n} x^k, x^1, \ldots, x^n\right) \in int\Delta_n\right\}$
	the image of projection of the standard simplex onto hyper-surface $x^0 = 0$
$I_k^{i_0,\ldots,i_k}$	$\{i_0, \ldots, i_k\}$
I_k	$\{\{i_0, \ldots, i_k\},\ 0 \leq i_0 < \ldots < i_k \leq n\}$
Ω_k	$intco\{e_0, \ldots, e_k\}, \quad k \in \{1, \ldots, n\}$
	the domain representing a population of alleles $\{A_0, \ldots, A_k\}$
$V_k^{i_0,\ldots,i_k}$	$intco\{e_{i_0}, \ldots, e_{i_k}\}$
	the domain representing a population of alleles $\{A_{i_0}, \ldots, A_{i_k}\}$
$V_k^{(l)}$	the l^{th} component of V_k, $l = 1, \cdots, \binom{n+1}{k+1}$
V_k	$\left\{intco\{e_{i_0}, \ldots, e_{i_k}\} \text{ for some } i_0 < \ldots < i_k \in \overline{0, n}\right\}$
	$= \bigsqcup\limits_{(i_0,\ldots,i_k) \in I_k} V_k^{i_0,\ldots,i_k} = \bigsqcup\limits_{l=1}^{\binom{n+1}{k+1}} V_k^{(l)}$
	the domain representing a population of $(k+1)$ alleles
$\overline{V}_k$	$\bigsqcup\limits_{i=0}^{k} V_i,\ k = 0, \ldots, n$
	the domain representing a population of as most $(k+1)$ alleles
$w_k^{i_0,\ldots,i_k}(x)$	$\prod\limits_{i \in I_k^{i_0,\ldots,i_k}} x^i$
H_k	$C^\infty(\overline{\Omega}_k), \quad k \in \{1, \ldots, n\}$
$H_k^{i_0,\ldots,i_k}$	$C^\infty\left(\overline{V_k^{i_0,\ldots,i_k}}\right)$
H	$\left\{f : \overline{V}_n \to [0, \infty] \text{ measurable such that } \int\limits_{\overline{V}_n} f(x)g(x)dx < \infty, \forall g \in H_n\right\}$

Operators

L_k	$H_k \to H_k, \quad L_k f(x) = \frac{1}{2} \sum_{i,j=1}^{k} \frac{\partial^2}{\partial x^i \partial x^j}(x^i(\delta_{ij} - x^j) f(x)), \quad k \in \{1, \ldots, n\}$
L_k^*	$H_k \to H_k, \quad L_k^* g(x) = \frac{1}{2} \sum_{i,j=1}^{k} (x^i(\delta_{ij} - x^j)) \frac{\partial^2}{\partial x^i \partial x^j} g(x), \quad k \in \{1, \ldots, n\}$
L_0	$= L_0^* = 0$
L	L_1
L^*	L_1^*
$L_k^{i_0,\ldots,i_k}$	$H_k^{i_0,\ldots,i_k} \to H_k^{i_0,\ldots,i_k}, \quad L_k^{i_0,\ldots,i_k} f(x) = \frac{1}{2} \sum_{i,j \in I_k^{i_0,\ldots,i_k}} \frac{\partial^2}{\partial x^i \partial x^j}(x^i(\delta_{ij} - x^j) f)$
$(L_k^{i_0,\ldots,i_k})^*$	$H_k^{i_0,\ldots,i_k} \to H_k^{i_0,\ldots,i_k}, \quad (L_k^{i_0,\ldots,i_k})^* g(x) = \frac{1}{2} \sum_{i,j \in I_k^{i_0,\ldots,i_k}} (x^i(\delta_{ij} - x^j)) \frac{\partial^2}{\partial x^i \partial x^j} g(x)$

Others

$[f,g]_n$	$\int_{\overline{V}_n} f(x)g(x)d\mu, \forall f \in H, g \in H_n$
	$= \sum_{k=0}^{n} \int_{V_k} f_k(\mathbf{x}) g(\mathbf{x}) d\mu_k$
	$= \sum_{k=0}^{n} (f_k, g)_k$
	where $f_{\mid V_k} = f_k, \quad k = 0, \cdots, n$
$[f,g]$	$[f,g]_1$
$(a)_n$	$a(a+1)(a+2)\cdots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)}, \quad (a)_0 = 1$
$a_{[n]}$	$a(a-1)(a-2)\cdots(a-n+1) = \frac{\Gamma(a)}{\Gamma(a-n)}, \quad a_{[0]} = 1$

References

[1] M. Abramowitz, I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, 1965.

[2] S. Amari, *Differential geometrical methods in statistics*, Lecture Notes in Statistics, Springer-Verlag, Berlin, 1985.

[3] S. Amari, *Information geometry on hierarchy of probability distributions*, IEEE Transactions on Information Theory, Vol. **47**, No. **5**, (2001), 1701-1711.

[4] S. Amari, H. Nagaoka, *Methods of information geometry*, AMS Oxford University Press, Translations of mathematical monographs Vol. **191**, 2000.

[5] P. L. Antonelli, C. Strobeck, *The Geometry of random drift I. Stochastic Distance and Diffusion*, Advances in Applied Probability, Vol. **9**, No. **2**, (1977), 238-249.

[6] P. L. Antonelli, J. Chapin, G. M. Lathrop, K. Morgan, *The Geometry of Random Drift II. The Symmetry of Random Genetic Drift*, Advances in Applied Probability, Vol. **9**, No. **2** (1977), 250-259.

[7] P. L. Antonelli, K. Morgan, G. M. Lathrop, *The Geometry of random drift III. Stochastic Distance and Diffusion*, Advances in Applied Probability, Vol. **9**, No. **2** (1977), 260-267.

[8] P. L. Antonelli, *The Geometry of Random Drift IV. Random Time Substitutions and Stationary Densities*, Advances in Applied Probability, Vol. **10**, No. **3** (1978), 563-569.

[9] P. L. Antonelli, C. Strobeck, *The Geometry of Random Drift V. Axiomatic Derivation of the WFK Diffusion from a Variational Principle*, Advances in Applied Probability, Vol. **11**, No. **3** (1979), 502-509.

[10] P. L. Antonelli, J. Chapin, B. H. Voorhees, *The Geometry of Random Genetic Drift VI. A Random Selection Diffusion Model*, Advances in Applied Probability, Vol. **12**, No. **1** (1980), 50-58.

[11] P. Appell, *Sur les series hypergeometriques de deux variables, et sur des equations differentielles lineaires aux derivees partielles*, C. R. Acad. Sci. Paris **90** (1880), 296-299.

[12] P. Appell, *Sur les series hypergeometriques de deux variables, et sur des equations differentielles lineaires simultanees aux derivees partielles*, C. R. Acad. Sci. Paris **90** (1880), 731-734.

[13] P. Appell, *Sur des polynomes de deux variables analogues aux polynomes de Jacobi*, Arch. Math. Phys. **66** (1881), 238-245.

[14] D. G. Aronson, H. F. Weinberger, *Multidimensional nonlinear diffusion arising in population genetics*, Adv. Math. **30** (1978) 33-76.

[15] H. Bateman, *Higher transcendental functions*, Volume I. McGraw-Hill, New York, Toronto, London, 1953.

[16] G. J. Baxter, R. A. Blythe, A. J. McKane, *Exac solution of the multi-allelic diffusion model*, Math. Biosciences **209** (2007) 124-170.

[17] J. H. Bennett, *On the theory of random mating*, Ann. Eugen, **18** (1954), 311-317.

[18] P. Billingsley, *Convergence of probability measures*, Willey N.Y., 1954.

[19] A. A. Borokov, *Theorems on the convergence to Markov diffusion processes*, Z. Wahrs., **16** (1970), 47-76.

[20] R. Bürger, *The mathematical theory of selection, recombination, and mutation*, Willey Series in Mathematical and Computational Biology, John Willey & Sons, LTD, 2000.

[21] J. F. Crow, M. Kimura, *An introduction to population genetics theory*, Harper and Row, N.Y., 1970.

[22] E. B. Dynkin, *Markov Processes*, Volume I, Band **121**, Springer-Verlag, Berlin, 1965.

[23] E. B. Dynkin, *Markov Processes*, Volume II. Band **122**, Springer-Verlag, Berlin, 1965.

[24] Warren J. Ewens, *Mathematical Population Genetics I. Theoretical Introduction*, Springer-Verlag New York Inc., Interdisciplinary Applied Mathematics, 2nd ed., 2004.

[25] R. A. Fisher, *On the dominance ratio*, Proc. Roy. Soc. Edinb., **42** (1922), 321-341.

[26] E. D. Fackerell, R. A. Littler, *Polynomials biorthogonal to Appell's polynomials*, Bull. Austral. Math. Soc., Vol **11** (1974), 181-195.

[27] W. Feller, *Diffusion processes in genetics*, Proc. Second Berkeley Symposium on Math. Stat. and Prob. (1951) 227-246.

[28] W. Feller, *The parabolic differential equations and the associated semi-groups of transformations*, Ann. Math., **Vol 55**, No. 3 (1952) 468-519.

[29] W. Feller, *Diffusion processes in one dimension*, Trans. Amer. Math. Soc., **77** (1954) 1-31.

[30] J. Felsenstein, *The rate of loss of multiple alleles in finite haploid populations*, Theor. Pop. Biol., **2** (1971) 391-403.

[31] J. H. Gillespie, *Population genetics: A concise guide*, The Johns Hopskins University Press, 2nd ed., 2004.

[32] D. L. Hartl, *A Primer of population genetics*, Sinauer Associates, Inc, 3rd ed., 2000.

[33] J. Hofbauer, K. Sigmund, *The theory of evolution and dynamical systems*, London Mathematical Society Student Texts, Cambridge University Press, Vol. **7**, 1988.

[34] J. Hofrichter, PhD thesis, *On the diffusion approximation of Wrigh-Fisher models with several alleles and loci and its geometry*, to be publiced.

[35] K. Ito, Jr. H. P. McKean, *Diffustion processes and their sample paths*, Springer-Verlag, Berlin, 1965.

[36] J. Jost, *Partial Differential Equations*, Springer - Verlag Berlin, Series: Graduate Texts in Mathematics , Vol. 214, 2nd ed., 2007.

[37] J. Jost, *Information geometry*, Lecture Notes, IMPRS, 2007.

[38] J. Jost, *Riemannian geometry and geometric analysis*, Universitext, Springer, 5th ed., 2008.

[39] S. Karlin, J. McGregor, *The number of mutant forms maintained in a population*, Proc. Fifth Berkeley Symp. Math. Stat. Prob., **4** (1967) 415-438.

[40] R. Z. Khasminskii, *Diffusion processes and elliptic differential equations degenerating at the boundary of the domain*, Theor. Prob. Appl., **3** (1958) 400-419.

[41] R. G. Khazanie, H. E. McKean, *A Mendenlian Markov process with binomial transition probabilities*, Biometrika, **53** No. **1/2** (1966) 37-48.

[42] R. G. Khazanie, H. E. McKean, *A Mendenlian Markov process with multinomial transition probabilities*, J. Appl. Prob., **3** No. **2** (1966) 353-364.

[43] M. Kimura, *Solution of a Process of Random Genetic Drift with a Continuous Model*, PNAS–USA, Vol. 41, No. 3, (1954), 144-150.

[44] M. Kimura, *Random genetic drift in multi-allele locus*, Evolution, **9** (1955), 419-435.

[45] M. Kimura, *Random genetic drift in a tri-allelic locus; exact solution with a continuous model*, Biometrics, **12** (1956), 57-66.

[46] M. Kimura, *Evolutionary rate at the molecular level*, Nature, **217**, (1968) 624-626.

[47] M. Kimura, *Diffusion models in population genetics*, J. Appl. Prob., Vol. **1** No. **2**, (1964) 177-232.

[48] M. Kimura, T. Ohta, *The average number of generations until fixation of a mutant gene in a finite population*, Genetics, **61**, (1969) 763-771.

[49] M. Kimura, T. Ohta, *The average number of generations until extinction of an individual mutant gene in a finite population*, Genetics, **63**, (1969) 701-709.

[50] Kolmogorov, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Mathematische Annalen, Vol. 104, Nr. 1, (1931) 415-458.

[51] H. J. Kushner, *Probability limit theorems and the convergence of finite defference approximations of partial differential equations*, J. Math. Anal. Appl., **32**, (1970) 77-103.

[52] R. A. Littler, *Loss of variability at one locus in a finite population*, Math. Bio. **25**, (1975) 151-163.

[53] R. A. Littler, *Multidimensional stochastic models in genetics*, PhD Thesis, Monash University, 1975.

[54] R. A. Littler, E. D. Fackerell, *Transition densities for neutral multi-allele diffusion models*, Biometrics, Vol **31** No. 1 (1975), 117-123.

[55] R. A. Littler, A. J. Good, *Ages, extinction times, and first passage probabilities for a multiallele diffusion model with irreversible mutation*, Theor. Popul. Biol., Vol **13** (1978), 214-225.

[56] G. Malécot, *Sur une probleme de probabilité en chaine que pose la genetique*, C. R. Acad. Sci., Paris **219** (1944), 279-281.

[57] T. Maruyama, *Stochastic problems in population genetics*, Springer-Verlag, 1977.

[58] S. J. Michell, *Random mating and random union of gametes for finite dioecious populations*, Theor. Pop. Biol., **4**, (1973) 222-247.

[59] M. Pinsky, *A note on degenerate diffusion processes*, Theor. Propb. Appl., **14**, (1969) 502-507.

[60] S. Rempel, B. W. Schulze, *A theory of pseudo-differential boundary value problems with discontinuous conditions I*, Ann. Glob. Analysis and Geometry, Vol. **2**, No. **2** (1984) 163-251.

[61] H. Risken, *The Fokker-Planck-Equation. Methods of Solution and Applications*, 2nd ed., Berlin etc., Springer-Verlag, 1989.

[62] F. Redig, *Diffusion limits in population models*, Unpublished, 2006.

[63] D. W. Stroock, S. R. S. Varadhan, *Diffusion processes with boundary conditions*, Comm. Pure. Appl. Math. **24** (1971) 147-225.

[64] T. D. Tran, J. Jost, J. Hofrichter, *A new point of view of solutions of Fokker-Planck equations associated with the Wright-Fisher models of* 2*-alleles*, unpublished.

[65] T. D. Tran, J. Jost, J. Hofrichter, *A new point of view of solutions of Fokker-Planck equations associated with the Wright-Fisher models of multi-alleles*, unpublished.

[66] D. Wagenaar, Information geometry for neutral networks, Master Thesis, 1998.

[67] G. A. Watterson, *Markov chains with absorbing states, a genetic example*, Ann. Math. Stat., **32** (1961), 716-729.

[68] G. A. Watterson, *Some theoretical aspects of diffusion theory in population genetics*, Ann. Math. Stat., **33** (1962), 939-957.

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

[70] S. Wright, *Size of population and breeding structure in relation to evolution.* Science 87 (1938) 430-431.

[71] S. Wright, *The differential equation of the distribution of gene frequencies*, Proc. Nat. Acad. Sci., **31** (1945), 382-389.

[72] E. Stein, G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton, N.J.: Princeton University Press, 1971.

[73] P. K. Suetin, *Ultraspherical polynomials*, in Hazewinkel, Michiel, Encyclopaedia of Mathematics, Springer, 2001.

[74] H. F. Trotter, *Approximation of semi-groups of operators*, Pac. J. Math., **8** (1958), 887-919.

[75] T. Ueno, *A survey on the Markov process on the boundary of multidimensional diffusion*, 5th Berkeley Simp. Math. Stat. Prob., **2** (1969), 111-130.

[76] A. D. Venttsel, *On Boundary Conditions For Multidimensional Diffusion Processes*, Theor. Prob. Appl., **4** (1959), 164-177.

Wissenschaftlicher Werdegang

- 2000-2004 Studium an der Universität Hanoi University of Science, Vietnam, Bachelor Mathematik.

- 2004-2006 Studium an der Universität Hanoi University of Science, Vietnam, Master Mathematik

- seit Oktober 2007 Promotionsstudium an der Universität Leipzig, gefördert durch die International Max Planck Research School am Max-Planck-Institut für Mathematik in den Naturwissenschaften.

....

Tat Dat Tran

Bibiliographische Daten

Tat Dat Tran
Information Geometry and the Wright-Fisher model of mathematical population genetics (Informationen Geometrie und der Wright-Fisher-Modell der mathematischen Populationsgenetik)
Dissertation, Universität Leipzig, 2012

viii + 123 Seiten, 76 Referenzen
Mathematics Subject Classification (2000):
Keyword: Random genetic drift, Wright-Fisher models, Fokker-Planck equations, Kolmogorov equations, affine Laplacian operators, exact solutions, hierarchical solutions, Information Geometry.

Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichen oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, den July 6, 2012

. . . .
(Tran Tat Dat)