THE LIMITS OF CERTAIN PROBABILITY DISTRIBUTIONS ASSOCIATED WITH THE WRIGHT-FISHER MODEL

THE LIMITS OF CERTAIN PROBABILITY DISTRIBUTIONS ASSOCIATED WITH THE WRIGHT-FISHER MODEL

By

YOUZHOU ZHOU, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2010) McMaster University
(Mathematics) Hamilton, Ontario

TITLE: THE LIMITS OF CERTAIN PROBABILITY DISTRIBUTIONS ASSOCIATED WITH THE WRIGHT-FISHER MODEL

AUTHOR: Youzhou Zhou, B.Sc.
(Huazhong Normal University)

SUPERVISOR: Dr. Shui Feng

NUMBER OF PAGES: LIII, 53

To my dearest parents and wife

Acknowledgements

I wouldn't let go the chance to express my appreciation to people from whom I received so many helps. Undoubtedly, my acknowledgement first should go to Dr. Shui Feng. Without his introduction, I can hardly have the opportunity to work on this fantastic problem. Secondly, I owe too much to people I know, such as Dr. Lia Bronsard, Dr. Hoppe and Chris Cappadocia, and even some persons personaly unknown to me as well. Last but not least, My parents and my wife deserve anything I can give, their support and encouragement constantly make me energetic while I was working on this problem.

Abstract

In this thesis, I endeavor to solve the remaining problem in Dr.Feng's paper[8], where Dr.Feng obtain the Large Deviation Principle of the following distribution

$$\Pi_{\alpha,\lambda}(dx) = C_\alpha \exp(\lambda\theta(\alpha)\sum_{i=1}^{+\infty} x_i^2)PD(\alpha)(dx), \alpha > 0.$$

Generally speaking, the Large Deviation Principle can yield the limit distribution if its rate function has only one zero point. Unfortunately, however, the rate function in [8] involves another parameter λ. When $\theta(\alpha) = -\log\alpha, \lambda = -k(k+1), k \geq 1$, the rate function has exactly two zero points, thus by way of the Large Deviation Principle, we can hardly know its limit distribution. Therefore, I try to figure out another way to find it. Since $PD(\alpha)(dx)$ is the limit of the ordered Dirichlet distribution $\hat{D}(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})$ as $K \to +\infty$, then $\Pi_{\alpha,\lambda}(dx)$ is the limit of

$$C_{\alpha,K} \exp(\lambda\theta(\alpha)\sum_{i=1}^{+\infty} x_i^2)\hat{D}(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})(dx_1 \cdots dx_{K-1})$$

on $\Delta = \{(x_1, x_2, \cdots)|x_1 \geq x_2 \geq \cdots, \sum_{i=1}^{+\infty} x_i = 1\}$, as $K \to +\infty$. Hence, we could first try to find the limit of

$$C_{\alpha,K} \exp(\lambda\theta(\alpha)\sum_{i=1}^{+\infty} x_i^2)\hat{D}(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})(dx_1 \cdots dx_{K-1}),$$

as $\alpha \to 0$, then manage to deal with $\Pi_{\alpha,\lambda}$. In this thesis , I only find the limit of $C_{\alpha,2} \exp(\lambda \log\alpha(x_1^2 + x_2^2))\hat{B}(\alpha, \alpha)(dx_1)$, as $\alpha \to 0, \lambda > 0$, which is the case when $K = 2, \theta(\alpha) = -\log\alpha, \lambda < 0$. The result is quite unexpected!

Contents

Chapter 1

Introduction

Unlike any other species, which once dominated the earth as dinosaurs did, we human beings constantly explore the origin of life, and even, once in a while, reflect on their own and ask who we are, where we come from and where should we go. Among various fantastic theories as to these questions, Charles Darwin's Evolution Theory[4] seems to have given us a promising path toward answer in 1859. Ever since then, evolution theory and natural selection were commonly accepted and studied by biologists.

But why do organic creatures evolve in this way? Mendel, one of Darwin's contemporaries, came up with an idea which explains the mode of inheritance. Based on his famous breeding experiment with pea plants, he postulated there is an inherent mechanism governing inheritance, with some basic units playing a vital role. Moreover, thanks to the breakthrough in molecule biology, these basic units are proved to be DNA molecule with beautiful double helix structure[15]. Within cells, DNA is organized orderly in line into long structure called chromosome, they are duplicated before cell division, then each of these cells inherits a copy of similar chromosomes from the previous cell. Each chromosome consists of a DNA molecule, a RNA molecule and some other giant protein molecules, among which DNA actually governs the formation of almost all protein molecules which are the basic components of living creatures. Thus, this process guarantees that off-springs always resemble their parental species in some way. Usually only one piece of the linear structure of a chromosome is responsible for the growth of a particular property, and they are the so-called genes. The locations where these genes are located are called locus. On each of these locus there is usually some alternatives called alleles. In the process of chromosome duplication, a particular gene may be wrongly chosen as one of its alternatives, hence mutation occurs. This is the original impetus of the evolution according to Charles Darwin's opinions in [4]. In addition, for some species, their chromosomes within each cell can be paired, or grouped in some fixed style, whereas, for some species, their chromosomes may be not capable of doing this. It is proved that these pairs or groups in a locus together are responsible for formation of some characters. Species with these structures are usually called diploid, polyploid, or possibly called haploid, of which, diploid is the most commonly found to exist on earth so far, while the others are also widely applied to breeding which may possibly give us new species, such as rice, wheat and corn, and finally it may aid the treatment of food crisis in this world.

All these topics have been thoroughly studied by biologists. Mathematicians, however, especially statisticians, like Fisher, also make substantial contribution to this subject. Wright-Fisher Model is one of the famous genetic models originally proposed by Fisher implicitly in his research works, while it is Wright who explicitly constructed this model in his paper later on, and that's why the model is given such a name. In math, it is actually a discrete-time Markov chain. With the assumed gene pool, totality of all possible alleles in one specific locus, becoming bigger, this Markov chain is also complexified. J.F.C.Kingman introduced a coalescent process [13], based on which we can find the most frequent gene type finally, that means there is a common ancestor, hence, it is, at least partially, an answer to the problem where we come from.

In this model, each generation need to take an equal amount of time to evolve and the number of individuals remains the same. Suppose, however, the evolving time is shrinking down to zero in a particular way, then this Markov chain can be approximated by a diffusion process in the sense of probability, please refer to [7]. Though the diffusion process loss some information, it is easy to handle mathematically. Therefore, research papers have sprung up in late 20th century. In the meantime, the gene pool was also assumed to be countably infinite. And the diffusion process turns out to be an infinite dimensional diffusion process. Please refer to [6], where the diffusion process was rigorously constructed by Ethier and Kurtz. To our surprise, the stationary distribution of this diffusion is the Poisson-Dirichlet distribution which will be briefly introduced in chapter 3. In fact, J.F.C.Kingman first introduced the distribution in [13], where he tended to find the limit distribution of ordered Dirichlet distribution.

In the frame work of this infinite dimensional diffusion process with stationary distribution- Poisson-Dirichlet distribution, the asymptotic behaviors under various conditions have been studied thoroughly by Dr.Shui Feng, with the results (some with collaborators) collected in [9]. Among all these results, the large deviation results with small mutation rate [8] particularly captured my attention because of Dr.Feng's introduction. He pointed out that this result, surprisingly, has an intimate connection with J.F.C.Kingman's Coalescent structure. Generally speaking, if the large deviation principle holds and its rate function has a unique zero point, then the law of large numbers will be readily obtained, but there is no such luck when rate function has more than one zero point, this will be clarified in Chapter 2. Unfortunately, the rate function in Dr.Feng's paper is also determined by another variable λ, when $\lambda = -k(k+1), k \geq 1$ the rate function exactly has two zero points. On the contrary, for $\lambda \neq -k(k+1)$, the rate function has only one zero point, hence its limit distribution is the Dirac measure at the zero point. One aspect of this that is not satisfactory is that it doesn't give us the actual limit distribution as mutation rate gradually vanishes for $\lambda = -k(k+1)$.

Therefore, this may not shed light on the evolution of the remote ancient creatures at the very beginning of life on earth, when the effect of mutations is likely to have been quite weak.

Although I can not solve this problem completely, I was able to make some progress, with unexpected results. In this thesis, I endeavour to find the limit distribution of a certain distribution associated with Wright-Fisher Model with selection.

Chapter 2

Preliminaries

In order to make this thesis friendly readable, some basic probability concept and biology ideas seems to be quite necessary to be introduced here. We only give a brief introduction, for more details, please refer to [5], [2], [1], and the references therein.

2.1 Probability Space, Random Variable and Conditional Expectation

2.1.1 Probability Space, Random Variable

Let Ω be a given set, consisting of outcomes of an experiment, we call it sample space, and let $\mathcal{F}$ be a σ-field in Ω, the elements of $\mathcal{F}$ are called events in the context of probability. Together with the probability measure P, the triple $(\Omega, \mathcal{F}, P)$ is called probability space.

A map $X : \Omega \mapsto S$, where S is some topological space, is called $\mathcal{F}$-measurable if

$$X^{-1}(U) = \{\omega \in \Omega : X(\omega) \in U, U \text{ is any open set in } S\} \in \mathcal{F}.$$

In probability space $(\Omega, \mathcal{F}, P)$, any measurable map is called random variable. We denote $\sigma(X)$ as the σ-field generated by random variable X, it is the smallest σ-field containing all the sets $X^{-1}(U)$, where U is open, that is

$$\sigma(X) = \cap\{\mathcal{H} : \mathcal{H} \quad \sigma - \text{field containing } X^{-1}(U), U \subset S \text{ open}\}.$$

2.1.2 Conditional Expectation

If $\int_{\Omega} |X(\omega)| dP(\omega) < \infty$, then the number

$$E(X) := \int_{\Omega} X(\omega) dP(\omega) = \int_{S} x d\mu_X(x)$$

is called the expectation of X (w.r.t. P), where $\mu_X(\cdot) := P(X^{-1}(\cdot))$, and it is called the distribution of X. Furthermore, consider a sub σ-field of $\mathcal{F}$, say $\mathcal{G}$, define $\tilde{P}$ as

$$\tilde{P}(B) = \int_{B} X dP, B \in \mathcal{G}.$$

Then $\tilde{P} << P$($\tilde{P}$ is absolutely continous with respect to P), by Radon-Nikodym theorem, $\frac{d\tilde{P}}{dP}$ exists a.e. in P, such that $\forall B \in \mathcal{G}$,

$$\int_{B} X dP = \int_{B} \frac{d\tilde{P}}{dP} dP,$$

thus, $E(X|\mathcal{G}) := \frac{d\tilde{P}}{dP}$ is called the conditional expectation. Specifically, when we take X to be $I_B(Y)$, then $E(I_B(Y)|\mathcal{G})$, denoted by $P(Y \in B|\mathcal{G})$, is called the conditional probability.

Proposition 1 (Properties of Conditional Expectation). *[5] Suppose X, Y are random variables; a, b, c are constants; $\mathcal{G}$ is a sub σ-field, and $E(X|\mathcal{G}), E(Y|\mathcal{G})$ both exist.*

1. *If $E(aX + bY + c|\mathcal{G})$ exists, then $E(aX + bY + c|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G}) + c$.*

2. *If X and $\mathcal{G}$ are independent, then $E(X|\mathcal{G}) = E(X)$.*

3. If Y is $\mathcal{G}$ measurable, then $E(Y|\mathcal{G}) = Y$.

4. If $X \geq Y$, then $E(X|\mathcal{G}) \geq E(Y|\mathcal{G})$, especially, $E(X|\mathcal{G}) \leq E(|X||\mathcal{G})$.

5. If φ is convex, and $E(\varphi(X)|\mathcal{G})$ exists, then $\varphi(E(X|\mathcal{G})) \leq E(\varphi(X)|\mathcal{G})$.

6. If $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ and $\mathcal{G}_1 \subset \mathcal{G}_2$, then $E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(E(X|\mathcal{G}_1)|\mathcal{G}_2) = E(X|\mathcal{G}_1)$.

Remark 1. *Proposition (1) is linearity of conditional expectation, (2) and (3) are both properties different from expectation, (4) and (5) are monotonicity of conditional expectation and Jensen Inequality respectively. (6) is a very important property which is widely used in Martingale and Markov Process. Actually, if $\mathcal{G}$ is finite σ-field, then the definition in this thesis is identical with that in elementary textbook of probability theory.*

2.2 Markov Chain and Diffusion Process

Markov chain and diffusion process are both Markov process, having Markovian Property and named after it. By Markovian property, loosely speaking, that is, the current statistical law of the process depends only on the current state, independent of its history; or precisely,

$$E(X_t|\sigma(X_u, u \leq s)) = E(X_t|\sigma(X_s)). \tag{2.1}$$

Markov process is a rather widely studied probability model and is abundant in research results; on the contrary, there are a few non-Markovian models which, unfortunately, hardly give birth to rich research results like Markov process.

2.2.1 Markov Chain

Let $(S, \mathcal{S})$ be a measurable space. A function $p : S \times \mathcal{S} \mapsto \mathbb{R}$ is said to be a transition probability if

1. For each $x \in S, A \mapsto p(x, A)$ is a probability measure on $(S, \mathcal{S})$;

2. For each $A \in \mathcal{S}, x \mapsto p(x, A)$ is a measurable function.

X_n is called Markov chain with transition probability p if

$$P(X_{n+1} \in B|\mathcal{F}_n) = p(X_n, B),$$

where $\mathcal{F}_n = \sigma(X_k, 0 \leq k \leq n)$.

Given a transition probability p and initial distribution μ on $(S, \mathcal{S})$, a set of finite dimensional distributions can be defined as follows

$$P(X_j, 0 \leq j \leq n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots \int_{B_n} p(x_{n-1}, dx_n).$$

If we suppose that $(S, \mathcal{S})$ is nice, because of Kolmogorov's extension theorem [16], we can construct a probability measure P_μ on sequence space $(S^{\{0,1,\cdots\}}, \mathcal{S}^{\{0,1,\cdots\}})$, such that the coordinate maps $X_n(\omega) = \omega_n$, and have the desired distribution P_μ.

Remark 2. *When $\mu = \delta_x$, the Dirac measure at x, then P_{δ_x} is abbreviated as P_x. For any probability measure $\nu, P_\nu(A) = \int \nu(dx) P_x(A)$.*

Remark 3. *When S is a countable space, we only need to consider one-step transition probability matrix P, whose (i, j) element P_{ij} is*

$$P(X_{n+1} = j|X_n = i).$$

The two-step transition probability $P_{ij}(2) = \sum_{k \in S} P_{ik} P_{kj}$, according to the Chapman -Kolmogorov equation. Hence, two-step transition probability matrix $P(2) = P^2$.

Example 1. *Suppose $S = \{0, 1\}$, and*

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

then

$$P_\mu(X_n = 0) = \sum_{i\in S} P_{i0}(n)\mu(\{i\}) = P_{00}(n)\mu(0) + P_{10}(n)(1-\mu(0)).$$

Since

$$\begin{aligned} P(n) &= \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}^n \\ &= \frac{1}{\alpha+\beta}\begin{pmatrix} \beta+\alpha(1-\alpha-\beta)^n & \alpha-\alpha(1-\alpha-\beta)^n \\ \beta-\beta(1-\alpha-\beta)^n & \alpha+\beta(1-\alpha-\beta)^n \end{pmatrix}, \end{aligned}$$

thus,

$$P_\mu(X_n = 0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n(\mu(0) - \frac{\beta}{\alpha+\beta})$$

and

$$P_\mu(X_n = 1) = \frac{\alpha}{\alpha+\beta} - (1-\alpha-\beta)^n(\mu(0) - \frac{\beta}{\alpha+\beta}).$$

Letting $n \to \infty$

$$\lim_{n\to+\infty} P_\mu(X_n = 0) = \frac{\beta}{\alpha+\beta}$$

and

$$\lim_{n\to+\infty} P_\mu(X_n = 1) = \frac{\alpha}{\alpha+\beta},$$

independent of μ*; moreover, put* $\nu(0) = \frac{\beta}{\alpha+\beta}, \nu(1) = \frac{\alpha}{\alpha+\beta}$*, we get*

$$P_\nu(X_n = 0) = \frac{\beta}{\alpha+\beta}$$

and

$$P_\nu(X_n = 1) = \frac{\alpha}{\alpha+\beta},$$

which suggests that the distribution of X_n *remains the same with the initial distribution, hence we call* ν *the stationary distribution.*

2.2.2 Diffusion Process

Diffusion process , unlike Markov chain, is a Markov process with continuous sample paths (i.e.$\{X_t(\omega)|t \geq 0$, for ω fixed$\}$). The existence and construction of diffusion process produce several beautiful probability theories: firstly, stochastic integration was originally came up with by a Japanese mathematician Itô, who tried to solve Kolmogorov's equation and construct a Markov process in sample path space [12], ever since stochastic integration was introduced , he found a diffusion process defined by stochastic differential equation. Secondly, the family of transition probabilities determine a semigroup, which under some conditions can be uniquely generated by an operator; and the converse is also true. This theory was constructed by Hille and Yosida. The operator is often called generator or infinitesimal operator. Finally, Dirichlet form on a function space can also completely determine a diffusion process. In this thesis, we focus on the second theory [7]. $\{X_t, t \geq 0\}$ is a diffusion process if it satisfies (2.1) and its sample paths are continuous.

A function $P(t, x, A)$ defined on $[0, \infty) \times S \times \mathcal{B}(S)$ is called time-homogeneous transition function if

1. $P(t, x, \cdot)$ is a probability measure on $\mathcal{B}(S)$, $(t, x) \in [0, \infty) \times S$;
2. $P(0, x, \cdot) = \delta_x(\cdot), \quad x \in S$;
3. $P(\cdot, \cdot, A)$ is a measurable function on $[0, \infty) \times S, A \in \mathcal{B}(S)$;
4. $P(t + s, x, A) = \int P(s, y, A)P(t, x, dy), s, t \geq 0, x \in S, A \in \mathcal{B}(S),$.

Probability measure μ given by $\mu(A) = P(X(0) \in A)$ is called the initial distribution of X.

A transition function for X and the initial distribution μ determine the finite dimensional distribution of X by

$$\begin{aligned}
&P(X(0) \in A_0, X(t_1) \in A_1, \cdots, X(t_n) \in A_n) \\
&= \int_{A_0} \cdots \int_{A_{n-1}} P(t_n - t_{n-1}, y_{n-1}, A_n) P(t_{n-1} - t_{n-2}, y_{n-2}, dy_{n-1}) \\
&\cdots P(t_1, y_0, dy_1) \mu(dy_0).
\end{aligned}$$

Conversely, given a transition function, we can define the finite dimensional distribution as above. Under some consistent conditions, we can construct a measure P_μ on $(S^T, \mathcal{S}^T)$, owing to Kolmogorov's extension theorem, so that the coordinate maps $X_t(\omega) = \omega_t$, and have the desired distribution. That is, under P_μ, $X_t(\omega)$ is a Markov process starting with distribution μ; but the sample path properties are still unknown, and may be not continuous.

Let S be a metric space, denote $M(S)$ as the collection of all real-valued, Borel measurable functions on S. Moreover, $B(S) \subset M(S)$ is the Banach space consisting of bounded functions with norm $||f|| = \sup_{x \in S} |f(x)|$; and $C(S) \subset B(S)$ is the space of bounded continuous function.

Let the initial distribution $\mu = \delta_x, x \in S$, and denote P_{δ_x} as P_x. For $f \in B(S)$, we define

$$P_t f(x) = \int f(y) P(t, x, dy).$$

Applying (4), we have,

$$\begin{aligned}
P_{t+s} f(x) &= \int f(y) P(t+s, x, dy) \\
&= \int [\int f(y) P(s, z, dy)] P(t, x, dz) \\
&= \int P_s f(z) P(t, x, dz) \\
&= P_t(P_s)(f)(x) \\
&= P_t \circ P_s f(x).
\end{aligned}$$

Hence, $P_{t+s} = P_t \circ P_s$; then $\{P_t, t \geq 0\}$ is a semigroup under composition.

If

$$\lim_{t \to 0} ||P_t f - f|| = 0, \quad \forall f \in C(S),$$

then $\{P_t, t \geq 0\}$ is said to be strongly continuous. If $||P_t|| \leq 1, \forall t > 0$, then $\{P_t, t \geq 0\}$ is a contraction semigroup. If $\forall f \in C(S), \forall t > 0$, and $P_t f$ is positive whenever $f(x)$ is, then $\{P_t, t \geq 0\}$ is said to be positive. A semigroup with these three properties is is called a Feller semigroup.

The generator of a semigroup $\{P_t, t \geq 0\}$ on $C(S)$ is the linear operator G defined by

$$Gf = \lim_{t \to 0} \frac{1}{t}\{P_t f - f\}.$$

Remark 4. *G is usually a differential operator. It is generally defined on $C^2(S)$ or $C^\infty(S)$, called core of G. Then the domain $\mathcal{D}(G)$ of G is the closure of its core in $C(S)$. Furthermore, we say G satisfies the positive maximum principle if $Gf(x_0) \leq 0$, whenever $\sup_{x \in S} f(x) = f(x_0) > 0, f \in \mathcal{D}(G), x_0 \in S$.*

Theorem 1 (Hille-Yosida Theorem). *[7] Let S be locally compact, the closure $\bar{G}$ of a linear operator G on $C(S)$ is single-valued and generates a Feller semigroup if and only if*

1. *$\mathcal{D}(G)$ is dense on $C(S)$;*

2. *G satisfies the positive maximum principle;*

3. *range $\mathcal{R}(\lambda - G)$ is dense in $C(S)$ for some $\lambda > 0$.*

Example 2 (Ornstein-Uhlenbeck Process). *[17] The transition probability function of O-U process is*

$$P(x, t, dy) = \frac{1}{\sqrt{2\pi(1 - e^{-2t})}} \exp\left(-\frac{(y - xe^{-t})^2}{2(1 - e^{-2t})}\right) dy.$$

By the definition of generator, the generator of G turn out to be

$$Gf(x) = f''(x) - xf'(x).$$

Indeed,

$$\begin{aligned}
P_t f(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\left(-\frac{(y-xe^{-t})^2}{2(1-e^{-2t})}\right) f(y)dy \\
&(\text{set } z = \frac{y - xe^{-t}}{\sqrt{1-e^{-2t}}}) \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) f(z\sqrt{1-e^{-2t}} + xe^{-t})dz.
\end{aligned}$$

If $f \in C_0^2$ (second order continuous differentiable with compact support), by Taylor's expansion, we have

$$f(z\sqrt{1-e^{-2t}} + xe^{-t}) = f(xe^{-t}) + \sqrt{1-e^{-t}} z f'(xe^{-t}) \tag{2.2}$$

$$+\frac{1}{2}(1-e^{-t})z^2 f''(xe^{-t}) + \frac{1}{2}(1-e^{-t})z^2(f''(\theta_1) - f''(xe^{-t})), \tag{2.3}$$

where $\theta_1 \in [xe^{-t}, z\sqrt{1-e^{-2t}} + xe^{-t}]$. Likewise,

$$f(xe^{-t}) = f(x) + (e^{-t} - 1)xf'(x) + \frac{1}{2}(e^{-t}-1)^2 x^2 f''(\theta_2), \tag{2.4}$$

where $\theta_2 \in [x, e^{-t}x]$. Therefore if we substitute 2.3 and 2.4 into $P_t f$, it ends up with

$$\begin{aligned}
&P_t f(x) = f(x) + (e^{-t}-1)xf'(x) + \frac{1}{2}(e^{-t}-1)^2 x^2 f''(\theta_2) + \\
&\frac{1}{2}(1-e^{-2t}) f''(xe^{-t}) + \frac{1}{2}(1-e^{-2t}) \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} z^2 (f''(\theta_1) - f''(xe^{-t}))dz.
\end{aligned}$$

Then

$$\begin{aligned}
&\|\frac{P_t f(x) - f(x)}{t} - (f''(x) - xf'(x))\| \\
&\le |\frac{e^{-t}-1}{t} + 1| \|xf'(x)\| + |\frac{1-e^{-2t}}{2t}| \|f''(xe^{-t}) - f''(x)\| \\
&+ |\frac{1-e^{-2t}}{2t} - 1| \|f''(x)\| + \frac{(e^{-t}-1)^2}{2t} \|x^2 f''(\theta_2)\| \\
&+ |\frac{1-e^{-2t}}{2t}| \|f''(\theta_1) - f''(xe^{-t})\| \\
&\to 0, \quad \text{as } t \to 0.
\end{aligned}$$

Thus, the generator is $Gf(x) = f''(x) - xf'(x)$. Let $t \to +\infty$, then

$$\lim_{t\to+\infty} P_t f(x) = \gamma(f) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) f(z) dz,$$

where $\gamma(dx)$ is the standard normal distribution. Actually, put $\mu = \gamma$, then

$$\begin{aligned}
&P_\mu(X_t \in A) = \int_{\mathbb{R}} P(t,x,A)\mu(dx) \\
&= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp(-\frac{(y-xe^{-t})^2}{2(1-e^{-2t})}) I_A(y) dy \Big) \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx \\
&(\text{ set } z = \frac{y-e^{-t}}{\sqrt{1-e^{-2t}}}) \\
&= \int_{\mathbb{R}^2} I_A(z\sqrt{1-e^{-2t}} + xe^{-t}) \frac{1}{2\pi} e^{-\frac{x^2+z^2}{2}} dx dz \\
&\left(\text{put } (x_1, x_2) = (z, x) \begin{pmatrix} \sqrt{1-e^{-2t}} & -e^{-t} \\ e^{-t} & \sqrt{1-e^{-2t}} \end{pmatrix} \right) \\
&= \int_{\mathbb{R}^2} I_A(x_1) \frac{1}{2\pi} e^{-\frac{x_1^2+x_2^2}{2}} dx_1 dx_2 \\
&= \int_{\mathbb{R}} I_A(x_1) \frac{1}{\sqrt{2\pi}} \exp(-\frac{x_1^2}{2}) dx_1 \\
&= \gamma(A).
\end{aligned}$$

It suggests that the distribution of X_t will remain the same if X_t starts with γ, hence γ is the stationary distribution of O-U process. In fact, μ is the stationary distribution of X, if and only if $\int Gf d\mu = 0, f \in \mathcal{D}(G)$ [7].

2.3 Asymptotic Theory in Probability Theory

Definition 1. *Given $\{X_n, n \geq 1\}$ and X,*

1. *Let $F_n(y) = P(X_n \leq y), F(y) = P(X \leq y)$, if $F_n(y)$ converges to $F(y)$ at the continuous point of $F(y)$, then we say X_n converges to X in distribution.*

2. *If $\lim_{n\to+\infty} P(|X_n - X| \geq \epsilon) = 0, \forall \epsilon > 0$, then X_n converges to X in probability.*

3. *If* $P(\lim_{n\to+\infty} X_n(\omega) = X(\omega)) = 1$, *then* X_n *converges to* X *almost surely, abbreviated as a.s..*

Definition 2. *Suppose* $\{\mu_n, \mu : n \geq 1\}$ *is a family of probability measures. We say* μ_n *converges to* μ *weakly, which is denoted as* $\mu_n \xrightarrow{w} \mu$, *if*

$$\lim_{n\to+\infty} \int_S f(x)\mu_n(dx) = \int_S f(x)\mu(dx), \quad \forall f \in C(S),$$

where S *is metric space,* $C(S)$ *is bounded continuous function space.*

Proposition 2. *Suppose* $\{\mu_n, \mu : n \geq 1\}$ *is a family of probability measure on metric space* S, *then the following statements are equivalent:*

1. $\mu_n \xrightarrow{w} \mu$.

2. *For any uniformly continuous function* $f(x)$, *we have*

$$\lim_{n\to+\infty} \int_S f(x)\mu_n(dx) = \int_S f(x)\mu(dx).$$

3. *For any closed set* F, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.

4. *For any open set* G, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$.

5. *For any set* A *with* $\mu(\partial A) = 0$, *then* $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.

Remark 5. *(5) is equivalent to the convergence in distribution.*

Definition 3. *Given* $\{X_n, X, n \geq 1\}$, *define*

$$\phi_n(t) = Ee^{-tX_n}, \phi(t) = Ee^{-tX}.$$

If $\phi_n, \phi < \infty$, *we call* ϕ_n, ϕ *the moment generating function of* X_n *and* X *respectively.*

Theorem 2. *Let* $\phi_n(t), \phi(t)$ *be the moment generating function of* X_n, X *respectively, if* $\phi_n(t), \phi(t)$ *both exists for* $|t| \leq h, h > 0$, *and*

$$\phi_n(t) \to \phi(t), \quad \forall |t| \leq h,$$

then X_n *converges to* X *in distribution.*

Proof: please refer to [11] [3]. In probability theory, we say two events A and B are independent if and only if

$$P(AB) = P(A)P(B).$$

Inductively, n events $A_1, \cdots, A_n$ are independent if

$$P(A_{i_1} \cdots A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}), \quad 2 \leq k \leq n.$$

A family of events $\{A_t, t \in T\}$ is said to be independent if any finite members of them are independent.

Definition 4. *Two σ-field $\mathcal{F}$ and $\mathcal{G}$ are independent if*

$$P(AB) = P(A)P(B), \quad \forall A \in \mathcal{F}, B \in \mathcal{G}.$$

Similarly, n σ-field are independent if $A_1, \cdots, A_n$ are independent for all $A_1 \in \mathcal{F}_1, \cdots, A_n \in \mathcal{F}_n$. Therefore, two random variables X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

Remark 6. *By the definition of expectation, for two independent random variable X and Y the following must hold,*

$$EXY = EXEY.$$

Furthermore, $E(X|Y) = EX$.

Theorem 3 (Borel Cantelli Lemma). *Suppose $\{A_n, n \geq 1\}$ is a sequence of events.*

1. *If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then*

$$P(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n}^{\infty} A_n) = 0.$$

2. If, additionally, $\{A_n, n \geq 1\}$ are independent, then $\sum_{n=1}^{\infty} P(A_n) = \infty$ can yield

$$P(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n}^{\infty} A_n) = 1.$$

In reality, when we toss a coin consecutively, those coin tosses can be regarded as independent events. Now, similarly, we consider a sequence of independent random variables $X_n, n \geq 1$, each of them follows Bernoulli distribution

$$P(X_1 = 1) = p, P(X_1 = 0) = 1 - p, 0 < p < 1.$$

Then $\frac{S_n}{n} = \frac{\sum_{k=1}^{n} X_k}{n}$ is the frequency of taking 1 during n times experiments. Naturally, we ask what does the frequency finally approach as n goes to infinity? The next theorem is a beautiful answer.

Theorem 4 (The Law of Large Numbers). *If $\{X_n, n \geq 1\}$ are independent with identical distribution and $E|X_1| < \infty$, then*

$$\lim_{n \to +\infty} \frac{\sum_{k=1}^{n} X_k}{n} = EX_1 \quad a.s..$$

Proof: Please refer to [3], [11].

Remark 7. *Thanks to the theorem stated above, $\lim_{n \to \infty} \frac{S_n}{n} = EX_1 = p$; thus, when we toss a coin consecutively, the frequency of appearing head would be approximately $\frac{1}{2}$.*

Since $\lim_{n \to \infty} \frac{S_n}{n} = EX_1 = p$ *a.s.*, then $\forall \epsilon > 0$,

$$P(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n}^{\infty} \{|\frac{S_k}{k} - p| \geq \epsilon\}) = 0.$$

Or equivalently,

$$\lim_{n \to \infty} P(\bigcup_{k \geq n}^{\infty} \{|\frac{S_k}{k} - p| \geq \epsilon\}) = 0,$$

thus $\lim_{n\to\infty} P(|\frac{S_n}{n} - p| \geq \epsilon) = 0$. Usually $\{|\frac{S_n}{n} - p| \geq \epsilon\}$ is called rare event for its probability of occurrence is tiny. Under some specific conditions, however, the estimation of its probability is quite indispensable, such as the frequency of earthquake, or the bankruptcy in industry. Therefore, the more accurate the more beneficial. It's obviously insufficient to know only the limit of $\{|\frac{S_n}{n} - p| \geq \epsilon\}$. Fortunately, we can obtain a more accurate result.

On one hand, by Chebyshev's inequality, $\forall \alpha > 0, x \in (p, 1]$, we have

$$\begin{aligned}
&P(\frac{S_n}{n} \geq x) = P(\alpha S_n \geq \alpha n x) \\
&\leq e^{-\alpha n x} E e^{\alpha S_n} = e^{-\alpha n x} E e^{\alpha \sum_{k=1}^{n} X_k} = e^{-\alpha n x} (E e^{\alpha X_1})^n \\
&= e^{-n[\alpha x - \log E e^{\alpha X_1}]},
\end{aligned}$$

so

$$\frac{1}{n} \log P(\frac{S_n}{n} \geq x) \leq -(\alpha x - \log E e^{\alpha X_1}), \quad \alpha > 0.$$

Thus

$$\begin{aligned}
\frac{1}{n} \log P(\frac{S_n}{n} \geq x) &\leq \inf_{\alpha > 0} [-(\alpha x - \log E e^{\alpha X_1})] \\
&= -\sup_{\alpha > 0} (\alpha x - \log E e^{\alpha X_1}).
\end{aligned}$$

Let $f(\alpha) = \alpha x - \log E e^{\alpha X_1} = \alpha x - \log(p e^{\alpha} + 1 - p)$, then

$$f'(\alpha) = x - \frac{p e^{\alpha}}{p e^{\alpha} + 1 - p}.$$

Solve the equation in α,

$$0 = f'(\alpha) = x - \frac{p e^{\alpha}}{p e^{\alpha} + 1 - p}.$$

We have

$$\alpha = \log \frac{1-p}{1-x} + \log \frac{p}{x}.$$

Since f' is decreasing, then

- $f'(\alpha) > 0$ when $\alpha < \log\frac{1-p}{1-x} + \log\frac{p}{x}$;
- $f'(\alpha) < 0$ when $\alpha > \log\frac{1-p}{1-x} + \log\frac{p}{x}$.

Then $f(\alpha)$ attains its maximum at $\log\frac{1-p}{1-x} + \log\frac{p}{x}$. Hence

$$\frac{1}{n}\log P(\frac{S_n}{n} \geq x) \leq -I(x),$$

where

$$I(x) = \begin{cases} x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}, & x \in [0,1] \\ +\infty, & \text{otherwise} \end{cases}.$$

On the other hand,

$$P(\frac{S_n}{n} \geq x) = P(X_1 + \cdots + X_n \geq nx) = P(Y \geq nx),$$

Where Y follows binomial distribution $B(n,p)$. So

$$P(\frac{S_n}{n} \geq x) = \sum_{k \geq [nx]+1} \binom{n}{k} p^k (1-p)^{n-k} \geq \binom{n}{[nx]+1} p^{[nx]+1}(1-p)^{n-[nx]-1}.$$

Owing to Stirling's formula,

$$n! = (\frac{n}{e})^n \sqrt{2\pi n} e^{\theta(n)}, \quad \frac{1}{12(n+\frac{1}{2})} \leq \theta(n) \leq \frac{1}{12n}.$$

We have

$$\begin{aligned} P(\frac{S_n}{n} \geq x) &\geq \frac{n!}{(n-[nx]-1)!([nx]+1)!} p^{[nx]+1}(1-p)^{n-[nx]-1} \\ &= (\frac{p}{x_n})^{nx_n} (\frac{1-p}{1-x_n})^{n(1-x_n)} \sqrt{\frac{1}{2\pi x_n(1-x_n)n}}, \end{aligned}$$

where $x_n = \frac{[nx]+1}{n} \to x$, as $n \to +\infty$. Letting $n \to +\infty$, we obtain that

$$\begin{aligned} &\liminf_{n\to+\infty} \frac{1}{n}\log P(\frac{S_n}{n} \geq x) \\ &\geq \lim_{n\to+\infty} (x_n \log\frac{p}{x_n} + (1-x_n)\log\frac{1-p}{1-x_n}) - \lim_{n\to+\infty} (\frac{1}{2n}\log 2\pi x_n(1-x_n)n) \\ &= -I(x), \end{aligned}$$

therefore,

$$\lim_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\geq x) = -I(x).$$

Similarly, $\forall x \in [0,p]$, consider $Y_k = -X_k$, then

$$P(Y_k = -1) = p, P(Y_k = 0) = 1-p.$$

We have

$$\lim_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\leq x) = \lim_{n\to+\infty}\frac{1}{n}\log P(\frac{\sum_{k=1}^n Y_k}{n}\geq -x).$$

Following the above argument, we have

$$\lim_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\leq x) = -I(x).$$

Thus, on one hand

$$\begin{aligned}
&\limsup_{n\to+\infty}\frac{1}{n}\log P(|\frac{S_n}{n}-p|\geq\epsilon)\\
&=\limsup_{n\to+\infty}\frac{1}{n}\log[P(\frac{S_n}{n}\geq p+\epsilon)+P(\frac{S_n}{n}\leq p-\epsilon)]\\
&=\limsup_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\geq p+\epsilon)\bigvee\limsup_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\leq p-\epsilon)\\
&=-\min(I(p+\epsilon),I(p-\epsilon));
\end{aligned}$$

on the other hand

$$\begin{aligned}
&\liminf_{n\to+\infty}\frac{1}{n}\log P(|\frac{S_n}{n}-p|\geq\epsilon)\\
&=\liminf_{n\to+\infty}\frac{1}{n}\log[P(\frac{S_n}{n}\geq p+\epsilon)+P(\frac{S_n}{n}\leq p-\epsilon)]\\
&\geq\liminf_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\geq p+\epsilon)\bigvee\liminf_{n\to+\infty}\frac{1}{n}\log P(\frac{S_n}{n}\leq p-\epsilon)\\
&=-\min(I(p+\epsilon),I(p-\epsilon)).
\end{aligned}$$

Therefore,

$$\lim_{n\to+\infty}\frac{1}{n}\log P(|\frac{S_n}{n}-p|\geq\epsilon) = -\min(I(p+\epsilon),I(p-\epsilon)).$$

Then $\forall\eta > 0, \exists N > 0$, such that for all $n > N$

$$e^{-n(\min(I(p+\epsilon),I(p-\epsilon))+\eta)} \leq P(|\frac{S_n}{n} - p| \geq \epsilon) \leq e^{-n(\min(I(p+\epsilon),I(p-\epsilon))-\eta)}.$$

Compared with the law of large number, this result gives us a more accurate estimation.

In fact, $I(x)$ is a nonnegative convex function, and $I(x) = 0$ if and only if $x = p$. Furthermore, $I(x)$ is strictly increasing in $[p, +\infty)$, and strictly decreasing in $(-\infty, p]$, thus $\min(I(p+\epsilon), I(p-\epsilon)) > 0$, and

$$\sum_{n=1}^{+\infty} P(|\frac{S_n}{n} - p| \geq \epsilon) < \infty.$$

By theorem 3,

$$P(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n}^{\infty} |\frac{S_k}{k} - p| \geq \epsilon) = 0, \forall\epsilon > 0.$$

That is, $\frac{S_n}{n} \to p \quad a.s.$ as $n \to +\infty$. In words, this estimation is much more powerful than the law of large number; because, apart from giving us the limit, it also give us the convergent speed, which is approximately

$$\exp\{-n(\min(I(p+\epsilon), I(p-\epsilon)))\}.$$

Unfortunately, however, the limit of $\frac{1}{n}\log P(A_n)$ may not exist. We therefore need to consider its upper limit and lower limit, which at least give us an upper bound and a lower bound of its convergent speed. This is exactly what the large deviation principle does.

Definition 5 (The large deviation principle). $\{\mu_\alpha, \alpha > 0\}$ *is said to satisfy the large deviation principle, if* $\forall B \in \mathcal{B}$,

$$-\inf_{x\in B^o} I(x) \leq \liminf_{\alpha\to 0} \alpha \log \mu_\alpha(B) \leq \limsup_{\alpha\to 0} \alpha \log \mu_\alpha(B) \leq -\inf_{x\in \bar{B}} I(x),$$

where $I(x)$ *is a lower semi-continuous and nonnegative convex function, called rate function.* B^o *and* $\bar{B}$ *is the interior and closure of* B *respectively.*

Remark 8. *If $\{x|I(x) \leq y\}$ is a compact set, then we say I is a good rate function. By the way, following the above argument, we can obtain the following fact.*

Fact: If x_0 is the only zero point of $I(x)$, then $\mu_\alpha \xrightarrow{w} \delta_{x_0}$, as $\alpha \to 0$

Indeed, by (5) of Proposition 2, $\forall B \in \mathcal{B}$, satisfying $\delta_{x_0}(\partial B) = 0$,

if B does not contain x_0, then $x_0 \notin \bar{B}$, thus $\inf_{x\in\bar{B}} I(x) > 0$.

So take $\inf_{x\in\bar{B}} I(x) > \epsilon > 0, \exists\delta > 0$, such that $\forall 0 < \alpha < \delta$, we have

$$\exp\{-\frac{1}{\alpha}(inf_{x\in B^o}I(x) + \epsilon)\} \leq \mu_\alpha(B) \leq \exp\{-\frac{1}{\alpha}(inf_{x\in\bar{B}}I(x) - \epsilon)\}. \qquad (2.5)$$

Let $\alpha \to 0$, then $\mu_\alpha(B) \to 0$

If B contains x_0, then $x_0 \in B^o$ and $x_0 \notin \bar{B}^c$; since $\mu_\alpha(B) = 1-\mu_\alpha(B^c), \mu_\alpha(B^c) \to 0$, then $\mu_\alpha(B) \to 1$. Thus, $\mu_\alpha \xrightarrow{w} \delta_{x_0}$.

Remark 9. *If the rate function has more than one zero points, then we are pretty sure that its limit distribution only has support on those zero points. but we are incapable of knowing how the probability is distributed among them, anyway its limit distribution is a discrete distribution.*

Chapter 3

The Wright-Fisher Model, Diffusion Approximation, and the Poisson-Dirichlet Distribution

In this chapter, we focus on diploid species [7], most of which are produced through sexual reproduction. Each individual has a large number of germ cells. They first split into gametes, containing one chromosome from each homologous pair in the original cell. When two gametes fuse, they form a zygote that has two complete sets of chromosomes and therefore is called diploid. For a given locus, if the two alleles on this locus are the same, then it is called homozygote; otherwise, heterozygote. Here we assume random mating; thus, Hardy-Weinberg Principle can be applied.

Each generation may go through complicated procedures to produce the next generation. For the sake of simplicity, we assume that each generation goes through three stages. Firstly, random mating is the starting point of the next generation. Then a large number of zygotes are produced. Secondly, some species are better fitted to the environment, hence they are more likely to survive to reproductive age. And all species must face natural selection and mutation as well. Finally, random sampling will occur, and the population remains constant in the Wright-Fisher model. In the following sections, I will proceed to introduce the model in detail.

3.1 Two-allele Wright-Fisher Model and Diffusion Approximation

3.1.1 Two-allele Markov Chain

Let A, a be the two alleles at a particular locus in a population of N, and x be the frequency of A before random mating, and there are three genotypes: AA, Aa, aa. After random mating, $x^2, 2x(1-x)$ and $(1-x)^2$ are the frequencies of genotypes AA, Aa and aa respectively, because of Hardy-Weinberg principle. Suppose $A \rightarrow a$ denotes the event that A mutates into a, and $a \rightarrow A$ denotes the event that a mutates into A. Similarly, we have two extra events $A \rightarrow A$ and $a \rightarrow a$. Let $u_1 = P(A \rightarrow a), u_2 = P(a \rightarrow A)$. Let $x^{'}$ be the frequency of A after mutation, then

$$x^{'} = x(1-\mu_1) + (1-x)\mu_2. \tag{3.1}$$

If we assume that each allele mutates independently, the frequencies of its genotypes after mutation are:

$$\begin{aligned}\tilde{P}_{AA} &= P_{AA}P(A \rightarrow A)P(A \rightarrow A) + P_{Aa}P(A \rightarrow A)P(a \rightarrow A)\\ &+ P_{aa}P(a \rightarrow A)P(a \rightarrow A),\end{aligned}$$

$$\begin{aligned}\tilde{P}_{Aa} &= 2P_{AA}P(A \rightarrow A)P(A \rightarrow a) + P_{Aa}(P(A \rightarrow A)P(a \rightarrow a)\\ &+ P(A \rightarrow a)P(a \rightarrow A)) + 2P_{aa}P(a \rightarrow A)P(a \rightarrow a),\end{aligned}$$

$$\tilde{P}_{aa} = P_{AA}P(A \to a)P(A \to a) + P_{Aa}P(A \to a)P(a \to a)$$
$$+ P_{aa}P(a \to a)P(a \to a).$$

Then

$$\tilde{P}_{AA} = (x^{'})^2,$$
$$\tilde{P}_{Aa} = 2x^{'}(1-x^{'}),$$
$$\tilde{P}_{aa} = (1-x^{'})^2.$$

We then take into account selection, rewriting their frequencies as weighted mean

$$P^{''}_{AA} = \frac{w_1(x^{'})^2}{w_1(x^{'})^2 + 2x^{'}(1-x^{'}) + w_2(1-x^{'})^2},$$
$$P^{''}_{Aa} = \frac{2x^{'}(1-x^{'})}{w_1(x^{'})^2 + 2x^{'}(1-x^{'}) + w_2(1-x^{'})^2},$$
$$P^{''}_{aa} = \frac{w_2(1-x^{'})^2}{w_1(x^{'})^2 + 2x^{'}(1-x^{'}) + w_2(1-x^{'})^2},$$

where $w_1, w_2 > 0$. Obviously, if $w_1 > 1$, then the frequency of AA is relatively higher than its previous one, we say genotype AA is favored during natural selection. Let $x^{''}$ be the frequency of A after selection, then

$$x^{''} = \frac{2NP^{''}_{AA} + NP^{''}_{Aa}}{2N} = P^{''}_{AA} + \frac{1}{2}P^{''}_{Aa} = \frac{w_1(x^{'})^2 + x^{'}(1-x^{'})}{w_1(x^{'})^2 + 2x^{'}(1-x^{'}) + w_2(1-x^{'})^2}. \quad (3.2)$$

Finally, the population will go through random sampling. the frequencies of AA, Aa, aa are $P^{''}_{AA}, P^{''}_{Aa}, P^{''}_{aa}$, and

$$(NP^{'}_{AA}, NP^{'}_{Aa}, NP^{'}_{aa}) \sim Multinomial(N, (\tilde{P}_{AA}, \tilde{P}_{Aa}, \tilde{P}_{aa})).$$

Let $\tilde{x}$ be the frequency of A after random sampling, but $2N\tilde{x}$ may not necessarily follow Binomial(2N,$x^{''}$) unless we assume

$$\tilde{P}_{AA} = (x^{''})^2,$$
$$\tilde{P}_{Aa} = 2x^{''}(1-x^{''}),$$
$$\tilde{P}_{aa} = (1-x^{''})^2,$$

please refer to [7] for explanation. Therefore, if we assume that the above condition is satisfied, then

$$2N\tilde{x} \sim Binomial(2N, x'').$$

Given $\{X(n), n \geq 0\}$, if

$$P(X(n+1) = l | X(n) = k) = \binom{2N}{l}(x'')^l(1 - x'')^{2N-l},$$

then we call $\{X(n), n \geq 0\}$ two-allele Wright-Fisher model. Here x'' is obtained as above by (3.1) and (3.2) when we take $x = \frac{k}{2N}$.

Remark 10. *The original Wright-Fisher model does not take into account mutation and selection, hence its transition function P_{ij} is determined by*

$$\binom{2N}{j}(\frac{i}{2N})^j(1 - \frac{i}{2N})^{2N-j}.$$

3.1.2 Two-alleles Diffusion Approximation

Let $Y_N(t) = X([2Nt])$, define $P_N(t) = \frac{Y_N(t)}{2N}$. Since one unit of $P_N(t)$ corresponds to $2N$ units of X, then take $\Delta t = \frac{1}{2N}$, during $[t, t + \Delta t]$, $P_N(t) = x$ is changed into x'' because of selection and mutation. Next, we try to find the limit.

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E(P_N(t + \Delta t) - P_N(t) | P_N(t) = x) := b(x),$$
$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E((P_N(t + \Delta t) - P_N(t))^2 | P_N(t) = x) := a(x).$$

If we find them, from the theorem 1.1 of Chapter 10 in [7], then $P_N(t)$ approaches a diffusion process with generator

$$G = \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}.$$

But before we do the calculation, we need to scale the mutation and selection like

$$u_i = \frac{\mu_i}{2N} \wedge \frac{1}{2}, w_i = [1 + \frac{\sigma_i}{2N}] \vee \frac{1}{2}, i = 1, 2.$$

Then on one hand,

$$\frac{1}{\Delta t}E(P_N(t+\Delta t)-P_N(t)|P_N(t)=x)$$
$$=2NE(P_N(t+\Delta t)-x^{''})+2N(x^{''}-x)$$
$$=2N(x^{''}-x)$$
$$=2N\Big[\frac{w_1(x^{'})^2+x^{'}(1-x^{'})}{w_1(x^{'})^2+2x^{'}(1-x^{'})+w_2(1-x^{'})^2}(1-x)$$
$$-\frac{w_2(1-x^{'})^2+x^{'}(1-x^{'})}{w_1(x^{'})^2+2x^{'}(1-x^{'})+w_2(1-x^{'})^2}x\Big]$$

(Take N to be sufficiently large, then

$$u_i=\frac{\mu_i}{2N},w_i=1+\frac{\sigma_i}{2N},i=1,2\Big)$$

$$=\frac{2N(x^{'})^2(1-x)+\sigma_1(x^{'})^2(1-x)+2Nx^{'}(1-x^{'})(1-x)}{w_1(x^{'})^2+2x^{'}(1-x^{'})+w_2(1-x^{'})^2}$$
$$-\frac{2Nx(1-x^{'})^2+\sigma_2(1-x^{'})^2x+2Nx^{'}(1-x^{'})x}{w_1(x^{'})^2+2x^{'}(1-x^{'})+w_2(1-x^{'})^2}$$
$$=\frac{\sigma_1(x^{'})^2(1-x)-\sigma_2(1-x^{'})^2x+2N[x^{'}(1-x)-x(1-x^{'})]}{w_1(x^{'})^2+2x^{'}(1-x^{'})+w_2(1-x^{'})^2}$$
$$=\frac{\sigma_1(x^{'})^2(1-x)-\sigma_2(1-x^{'})^2x+(\mu_2-x(\mu_1+\mu_2))}{w_1(x^{'})^2+2x^{'}(1-x^{'})+w_2(1-x^{'})^2}$$

(Since $x^{'}=x(1-\frac{\mu_1}{2N})+(1-x)\frac{\mu_2}{2N},\omega_i=1+\frac{\sigma_i}{2N},i=1,2$, then $x^{'}\to x,\omega_i\to 1$)

$$\to -\mu_1x+\mu_2(1-x)+x(1-x)[x\sigma_1-(1-x)\sigma_2]$$

On the other hand,

$$\frac{1}{\Delta t}E((P_N(t+\Delta t)-P_N(t))^2|P_N(t)=x)$$
$$=2NE(P_N(t+\Delta t)-x^{''}+x^{''}-x)^2$$
$$=2NE(P_N(t+\Delta t)-x^{''})^2+2N(x^{''}-x)^2$$
$$=2N(\frac{1}{2N})^2E(2NP_N(t+\Delta t)-2Nx^{''})^2+2N(x^{''}-x)^2$$

$$= \frac{1}{2N} 2Nx''(1-x'') + 2N(x''-x)^2$$
$$= x''(1-x'') + \frac{1}{2N}(2N(x''-x))^2.$$

Since

$$x'' = \frac{w_1(x')^2 + x'(1-x')}{w_1(x')^2 + 2x'(1-x') + w_2(1-x')^2}$$
$$\to x^2 + x(1-x) = x,$$

and $2N(x''-x) \to -\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]$, therefore

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E((P_N(t+\Delta t) - P_N(t))^2 | P_N(t) = x) = x(1-x).$$

Hence,

$$a(x) = x(1-x), b(x) = -\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2].$$

We have completed the diffusion approximation and the limit process is the diffusion process with generator

$$Gf(x) = \frac{1}{2}x(1-x)\frac{d^2 f}{dx^2} + \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}\frac{df}{dx}.$$

Additionally, we are going to prove that it has a stationary distribution.

Proposition 3. *Suppose $\{X_t, t \geq 0\}$ is a diffusion process with generator stated above, then*

$$v(dx) = C(1-x)^{2\mu_2 - 1} x^{2\mu_1 - 1} e^{\sigma_1 x^2 + \sigma_2 (1-x)^2} dx$$

is its stationary distribution.

Proof: It is sufficient to show that

$$\int_0^1 Gf(x) v(dx) = 0, \quad \forall f \in C^2(\mathbb{R}).$$

By integration by parts formula, we have

$$
\begin{aligned}
&\int_0^1 Gf(x)v(dx) \\
&= C\int_0^1 \frac{1}{2}x(1-x)f''(x)v(dx) \\
&+ C\int_0^1 \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}f'(x)v(dx)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{C}{2}\int_0^1 f''(x)(1-x)^{2\mu_2}x^{2\mu_1}e^{\sigma_1 x^2+\sigma_2(1-x)^2}dx \\
&+ C\int_0^1 \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}f'(x)v(dx) \\
&= \frac{C}{2}\int_0^1 (1-x)^{2\mu_2}x^{2\mu_1}e^{\sigma_1 x^2+\sigma_2(1-x)^2}df'(x) \\
&+ C\int_0^1 \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}f'(x)v(dx) \\
&= -\frac{C}{2}\int_0^1 f'(x)[2\mu_2(1-x)^{2\mu_2-1}x^{2\mu_1}e^{\sigma_1 x^2+\sigma_2(1-x)^2} \\
&- 2\mu_1(1-x)^{2\mu_2}x^{2\mu_1-1}e^{\sigma_1 x^2+\sigma_2(1-x)^2} \\
&+ (1-x)^{2\mu_2}x^{2\mu_1}(2\sigma_1 x - 2\sigma_2(1-x))e^{\sigma_1 x^2+\sigma_2(1-x)^2}]dx \\
&+ C\int_0^1 \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}f'(x)v(dx) \\
&= -C\int_0^1 \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}f'(x)v(dx) \\
&+ C\int_0^1 \{-\mu_1 x + \mu_2(1-x) + x(1-x)[x\sigma_1 - (1-x)\sigma_2]\}f'(x)v(dx) \qquad = 0.
\end{aligned}
$$

3.2 K-allele Wright-Fisher Model and Diffusion Approximation

Let $A_1, \cdots, A_K$ be all the possible alleles at a particular locus in a population of N, then there are $\frac{K(K+1)}{2}$ genotypes. Let x_{ij} be the frequency of A_iA_j just before reproduction, $1 \le i \le j \le K$, then

$$x_i = \frac{2Nx_{ii} + \sum_{i \neq j} Nx_{ij}}{2N} = x_{ii} + \frac{1}{2}\sum_{i \neq j} x_{ij} \tag{3.3}$$

is the frequency of the allele $A_i, 1 \le i \le K$.

Similarly, three stages are assumed to go through before the population is totally replaced by the new generation. By the way, Hardy-Weinberg principle is also assumed to apply here. Firstly, after random mating, the frequency of genotype $A_iA_j, 1 \le i \le j \le K$ is

$$(2 - \delta_{ij})x_ix_j.$$

Secondly, owing to mutation, the frequency of A_iA_j is

$$x^*_{ij} = (1 - \frac{1}{2}\delta_{ij})\sum_{k \le l}(u^*_{ki}u^*_{lj} + u^*_{kj}u^*_{li})(2 - \delta_{kl})x_kx_l,$$

where u^*_{ij} is defined to be($u_{ii} = 0$)

$$u^*_{ij} = (1 - \sum_k u_{ik})\delta_{ij} + u_{ij}.$$

Next, it becomes

$$x^{**}_{ij} = \frac{w_{ij}x^*_{ij}}{\sum_{k \le l} w_{kl}x^*_{kl}}.$$

because of selection. Like we did in two-allele model, by direct computation, we have

$$x^*_i = \sum_{j=1}^{K} u^*_{ji}x_j,$$

where x^*_i is the frequency of type A_i after mutation.

Finally, the population need to go through random sampling. The frequency of A_iA_j becomes x'_{ij}, which is a random variable; and

$$(Nx'_{ij})_{1 \le i \le j \le K} \sim \text{ Multinomial}(N, (x^{**}_{ij})_{i \le j}).$$

Furthermore, let $x_i^{'}$ be the frequency of A_i after random sampling; and let x_i^{**} be the frequency of A_i after selection and mutation. If we assume that $x_{ij}^{**} = (2-\delta_{ij})x_i^{**}x_j^{**}$ is still true, then following the argument in page 413 [7], we have

$$(2Nx_i^{'})_{1\le i\le K} \sim \text{ Multinomial}(2N, (x_i^{**})_{1\le i\le K}).$$

Put $\bar{X}(n) = (X_1(n), \cdots, X_{K-1}(n)), n \ge 0$, if

$$\begin{aligned}&P(X_1(n+1) = j_1, \cdots, X_{K-1}(n+1) = j_{K-1}|X_1(n) = i_1, \cdots, X_{K-1}(n) = i_{K-1})\\&= \frac{(2N)!}{j_1! \cdots j_K!}(\frac{i_1}{2N})^{j_1} \cdots (\frac{i_K}{2N})^{j_K},\end{aligned}$$

where $i_K = 2N - \sum_{l=1}^{K-1} i_l, j_K = 2N - \sum_{l=1}^{K-1} j_l$, then $\{\bar{X}(n), n \ge 0\}$ is called K-allele Wright-Fisher model. Then scale the mutation and selection as follows

$$u_{ij} = \frac{\mu_{ij}}{2N} \wedge \frac{1}{K}, w_i = [1 + \frac{\sigma_{ij}}{2N}] \vee \frac{1}{2}, i, j = 1, \cdots, K.$$

Define $\bar{Y}_N(t) = \bar{X}([2Nt]), P^N(t) = \frac{\bar{Y}_N(t)}{2N}$, by direct computation as we did in two-alleles model, we have

$$\begin{aligned}a_{ij}(x) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E((P_i^N(t+\Delta t) - P_i^N(t))(P_j^N(t+\Delta t) - P_j^N(t))|P^N(t) = x),\\b_i(x) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E(P_i^N(t+\Delta t) - P_i^N(t)|P^N(t) = x),\end{aligned}$$

where

$$a_{ij}(x) = x_i(\delta_{ij} - x_j),$$

and

$$b_i(x) = -\sum_{j=1}^{K} \mu_{ij}x_i + \sum_{j=1}^{K} \mu_{ji}x_j + x_i(\sum_{j=1}^{K} \sigma_{ij}x_j - \sum_{k,l=1}^{K} \sigma_{kl}x_k x_l).$$

Hence, $P^N(t)$ approaches diffusion process $P_K(t)$ with generator

$$G^K = \frac{1}{2}\sum_{i=1}^{K-1} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{K-1} b_i(x)\frac{\partial}{\partial x_i}. \tag{3.4}$$

Remark 11. *The diffusion process $P_K(t)$ with generator (3.4) is the diffusion approximation of K-allele Wight-Fisher Model. We can choose special mutation rates μ_{ij} and selection variabilities σ_{ij}, say $\sigma_{ij} = 0, i \neq j$, and $\sigma_{ii} < 0$. It means that this model favors heterozygotes. Hence, as the diffusion evolves, the genotypes should spread out, rather than being absorbed into one species. By the way, from the construction of Wright-Fisher model, we know that $P_K(t)$ is a diffusion process on $\Delta_K = \{x = (x_1, \cdots, x_{K-1}) | \sum_{l=1}^{K-1} x_l \leq 1\}$.*

If the mutation rates $\mu_{ij} = \frac{\alpha}{K-1}, i \neq j, \alpha > 0$, then the mutation is symmetric; if, moreover, the selection is absent, then the allele is neutral. The stationary distribution of diffusion process with K neutral alleles would be the Dirichlet distribution $D(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})$,[9] with density

$$\Pi_K(dx) = \frac{\Gamma(\frac{K}{K-1}\alpha)}{\Gamma(\frac{\alpha}{K-1}) \cdots \Gamma(\frac{\alpha}{K-1})} x_1^{\frac{\alpha}{K-1}-1} \cdots x_K^{\frac{\alpha}{K-1}-1} dx_1 \cdots dx_{K-1}, \tag{3.5}$$

which is a multivariate generalization of the Beta distribution($K = 2$).

If the selection is added, then the stationary distribution is given by the following distribution

$$\Pi_\sigma(dx) = C \exp\Big(\sum_{i,j=1}^{K} \sigma_{ij} x_i x_j\Big) \Pi_K(dx_1 \cdots dx_{K-1}). \tag{3.6}$$

If we take $\sigma_{ij} = 0, i \neq j$, and $\sigma_{ii} = \lambda\theta(\alpha)$, and put $H^K(x) = \sum_{i=1}^{K} x_i^2$, which is the so-called homozygosity [1] . Then, (3.6) becomes

$$\Pi_\sigma(dx) = C \exp\Big(\lambda\theta(\alpha) H^K(x)\Big) \Pi_K(dx_1 \cdots dx_{K-1}). \tag{3.7}$$

[1]Generally speaking, homozygosity is defined to be the probability of random sample of size r sharing the same allele type. That is, $H_r(x) = \sum_{i=1}^{K} x_i^r$ or $H_r(x) = \sum_{i=1}^{\infty} x_i^r$ if the gene pool is infinite.

3.3 The Infinitely-Many-Neutral-Alleles Diffusion and the Poisson-Dirichlet Distribution

If we order $D(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})$, and let $K \to +\infty$, then the ordered Dirichlet Distribution $\hat{D}(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})$ has a limit $PD(\alpha)(dx)$, called Poisson-Dirichlet distribution [9]. It is originally put up by K.F.C.Kingman in the context of Poisson process, since $\frac{\alpha}{K-1}K \to \alpha$ (as $K \to +\infty$) is in the style of Poisson limit, thus it is given such a name by Kingman. In the case of neutral alleles, as $K \to +\infty$, the diffusion $P_K(t)$ goes to $P(t)$ which is an infinite dimensional diffusion on

$$\nabla = \{x = (x_1, \cdots, x_n, \cdots) | x_1 \geq x_2 \geq \cdots, \sum_{i=1}^{\infty} x_i = 1\},$$

with generator

$$Gf(x) = \frac{1}{2} \sum_{i,j=1}^{\infty} x_i(\delta_{ij} - x_j) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} - \frac{\alpha}{2} \sum_{i=1}^{\infty} \frac{\partial f(x)}{\partial x_i}. \tag{3.8}$$

Its stationary distribution is the Poisson-Dirichlet distribution.

Remark 12. *(3.8) is firstly defined on its core*

$$\mathcal{C} = \{1, \varphi_k(x) | \varphi_k(x) = \sum_{i=1}^{\infty} x_i^k, k \geq 2\}$$

then extended to the closure of $\mathcal{C}$.

If the selection is added, then the diffusion becomes more complicated. For the sake of simplicity, we omit the introduction of redundant concepts, but if we take the selection as follows

$$\sigma_{ij} = 0, i \neq j, \text{and } \sigma_{ii} = \lambda\theta(\alpha),$$

then its stationary distribution is

$$\Pi_{\alpha,\lambda}(dx) = C \exp\Big(\lambda\theta(\alpha) H_2(x)\Big) PD(\alpha)(dx). \tag{3.9}$$

Remark 13. *(3.9) is the limit of*

$$C\exp\Big(\lambda\theta(\alpha)H_2(x)\Big)\hat{D}(\frac{\alpha}{K-1},\cdots,\frac{\alpha}{K-1})(dx_1\cdots dx_{K-1})$$

in weak sense, as $K\to\infty$*, where* $\hat{D}(\frac{\alpha}{K-1},\cdots,\frac{\alpha}{K-1})(dx_1\cdots dx_{K-1})$ *is the ordered Dirichlet distribution and supports only on*

$$\Delta_K=\{(x_1,x_2,\cdots,x_K,0,\cdots)|\sum_{i=1}^{K}x_i=1,x_1\geq x_2\geq\cdots x_K\geq 0\}.$$

3.4 Feng's Result

In this section, I will briefly introduce Dr.Feng's result in [8], therefore show my research motivation. Dr.Feng's result is concerned with distribution

$$C\exp\big(\lambda\theta(\alpha)H_r(x)\big)PD(\alpha)(dx). \tag{3.10}$$

However, when $r=2$, it is the same with the distribution (3.9). In addition, as the distribution (3.9) has biological background to some extent, it is quite necessary to mention here.

Theorem 5. *Let* $\{\Pi_{\alpha,\lambda}(dx),\alpha>0,\lambda\in\mathbb{R}\}$ *be a family of distribution as (3.9), for fixed* λ*,* $\{\Pi_{\alpha,\lambda}(dx),\alpha>0\}$ *satisfies the large deviation principle on* ∇ *with speed* $\lambda(\alpha)=-\frac{1}{\log\alpha}$*, and the rate function is*

$$\tilde{S}(x)=\begin{cases}S(x), & \lim_{\alpha\to 0}\theta(\alpha)\lambda(\alpha)=0,\\ S(x)+\lambda(1-H_2(x)), & \theta(\alpha)\lambda(\alpha)=1,\lambda\geq 0,\\ S(x)+|\lambda|H_2(x)-\inf\{\frac{|\lambda|}{n}+n-1:n\geq 1\}, & \theta(\alpha)\lambda(\alpha)=1,\lambda<0,\end{cases}$$

where $S(x)$ *is defined by*

$$S(x)=\begin{cases}0, & x\in L_1,\\ n-1, & x\in L_n,x_n>0,n\geq 2,\\ +\infty, & x\notin L.\end{cases}$$

And $L_n = \{(x_1, \cdots, x_n, 0, \cdots) \in \nabla : \sum_{i=1}^{n} x_i = 1\}, L = \bigcup_{n=1}^{\infty} L_n$. *Obviously,* $L_n \subset L_{n+1}$.

Proof: Please refer to [8].

Note that there is only one zero point for $S(x)$, it is $(1, 0, \cdots)$, so when $\theta(\alpha)$ and $\lambda(\alpha)$ fail to be comparable, then $\tilde{S}(x) = S(x)$, thus they share the same zero point. Moreover, when $\theta(\alpha)\lambda(\alpha) = 1, \lambda > 0$, since $S(x) \geq 0$ and $1 - H_2(x) \geq 0$, then $\tilde{S}(x) = 0$ if and only if $S(x) = 0$ and $1 - H_2(x) = 0$, therefore, $\tilde{S}(x)$ has only one zero point $(1, 0, \cdots)$. Finally, when $\theta(\alpha)\lambda(\alpha) = 1, \lambda < 0$, it can be shown that $\tilde{S}(x)$ might have more than one zero points.

Case 1: $\lambda \leq -1$,

$$\begin{aligned}
&\inf\{\frac{|\lambda|}{n} + n - 1, n \geq 1\} \\
&= \min\{\frac{|\lambda|}{[\sqrt{|\lambda|}]} + [\sqrt{|\lambda|}] - 1, \frac{|\lambda|}{[\sqrt{|\lambda|}] + 1} + [\sqrt{|\lambda|}]\}.
\end{aligned}$$

If

$$\frac{|\lambda|}{[\sqrt{|\lambda|}]} + [\sqrt{|\lambda|}] - 1 = \frac{|\lambda|}{[\sqrt{|\lambda|}] + 1} + [\sqrt{|\lambda|}], \tag{3.11}$$

then

$$|\lambda| = [\sqrt{|\lambda|}]([\sqrt{|\lambda|}] + 1).$$

Therefore, only when $\lambda = -k(k+1)$, $k \geq 1$, can (3.11) hold. Take λ to be $-k(k+1)$, since $k < [\sqrt{|\lambda|}] < k+1$, for fixed k, then

$$\inf\{\frac{|\lambda|}{n} + n - 1, n \geq 1\} = 2k.$$

Suppose $P_0 \in L_l$ is a zero point of $\tilde{S}(x)$, hence is a minimum point of $\tilde{S}(x)$, then

$$0 = \tilde{S}(P_0) = S(P_0) + k(k+1)H_2(P_0) - 2k = l - 1 + k(k+1)\frac{1}{l} - 2k.$$

the second equality is due to the fact that $S(x)$ remains constant on L_l, and $H_2(x)$ on L_l attains its minimum only if $x_1 = x_2 = \cdots = x_l = \frac{1}{l}$. Thus the minimum of $\tilde{S}(x)$ on

L_l is

$$l-1+k(k+1)\frac{l}{l^2}-2k=l-1+\frac{k(k+1)}{l}-2k=0.$$

Solve $l-1+\frac{k(k+1)}{l}-2k=0$, we have $l=k,k+1$, then $\tilde{S}(x)$ has exactly two zero points: $(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)$ and $(\frac{1}{k+1},\cdots,\frac{1}{k+1},0,\cdots)$. However, if $\lambda\neq -k(k+1),k\geq 1$, then

- for $|\lambda|>[\sqrt{|\lambda|}]([\sqrt{|\lambda|}]+1)$, we have

$$\frac{|\lambda|}{[\sqrt{|\lambda|}]}+[\sqrt{|\lambda|}]-1>\frac{|\lambda|}{[\sqrt{|\lambda|}]+1}+[\sqrt{|\lambda|}].$$

So $\tilde{S}(x)=S(x)+|\lambda|H_2(x)-(\frac{|\lambda|}{[\sqrt{|\lambda|}]+1}+[\sqrt{|\lambda|}])$. Similarly, assume $P_0\in L_l$ is the zero point of $\tilde{S}(x)$, then

$$0=\tilde{S}(P_0)=l-1+|\lambda|\frac{1}{l}-(\frac{|\lambda|}{[\sqrt{|\lambda|}]+1}+[\sqrt{|\lambda|}]),$$

solving this equation, we get $l=[\sqrt{|\lambda|}]+1$ or $l=\frac{|\lambda|}{[\sqrt{|\lambda|}]+1}$. But $l=\frac{|\lambda|}{[\sqrt{|\lambda|}]+1}$ is not a valid solution; otherwise, let $k=[\sqrt{|\lambda|}]+1$, then $|\lambda|=kl$, hence $k=[\sqrt{kl}]+1$, then $k-1=[\sqrt{kl}]$, and

$$k-1\leq\sqrt{kl}<k\Longrightarrow k-(2-\frac{1}{k})\leq l<k.$$

Thus $l=k-1=[\sqrt{|\lambda|}]$, $|\lambda|=[\sqrt{|\lambda|}]([\sqrt{|\lambda|}]+1)$, contradiction thus is produced! So $P_0=(\frac{1}{[\sqrt{|\lambda|}]+1},\cdots,\frac{1}{[\sqrt{|\lambda|}]+1},0,\cdots)$ is the unique zero point!

- For $|\lambda|<[\sqrt{|\lambda|}]([\sqrt{|\lambda|}]+1)$, assume $P_0\in L_l$ is the zero point of $\tilde{S}(x)$, then

$$0=\tilde{S}(P_0)=l-1+|\lambda|\frac{1}{l}-(\frac{|\lambda|}{[\sqrt{|\lambda|}]}+[\sqrt{|\lambda|}]-1). \tag{3.12}$$

Solving the equation, we have $l=[\sqrt{|\lambda|}]$, or $l=\frac{|\lambda|}{[\sqrt{|\lambda|}]}$. If $|\lambda|=n^2$, then $l=n=[\sqrt{|\lambda|}]$, so $l=[\sqrt{|\lambda|}]$ is the unique solution of (3.12). If $|\lambda|\neq n^2,n\geq 1$, then $l=\frac{|\lambda|}{[\sqrt{|\lambda|}]}$ is not a valid solution of (3.12); otherwise, let $k=[\sqrt{|\lambda|}]$, so $|\lambda|=kl$ and

$$k<\sqrt{kl}<k+1\Longrightarrow k<l<k+2+\frac{1}{k},$$

then $l \geq k+1 = [\sqrt{|\lambda|}]+1$, and $|\lambda| = kl \geq [\sqrt{|\lambda|}]([\sqrt{|\lambda|}]+1)$, which is a contradiction! Therefore, we have the unique zero point $P_0 = (\frac{1}{[\sqrt{|\lambda|}]}, \cdots, \frac{1}{[\sqrt{|\lambda|}]}, 0, \cdots)$.

Case 2: $-1 < \lambda < 0$, then

$$\inf\{\frac{|\lambda|}{n} + n - 1, n \geq 1\} = |\lambda|,$$

so $\tilde{S}(x) = S(x) + |\lambda| H_2(x) - |\lambda| = S(x) - |\lambda|(1 - H_2(x))$. Since $0 \leq 1 - H_2(x) \leq 1, 0 < |\lambda| < 1$, then $0 \leq |\lambda|(1 - H_2(x)) < 1$. If $P_0 \in L_l, l \geq 2$ is the zero point of $\tilde{S}(x)$, then $0 = l - 1 - |\lambda|(1 - H_2(P_0))$; but

$$1 \leq l - 1 = |\lambda|(1 - H_2(P_0)) < 1,$$

a contradiction! Hence, $P_0 \in L_1$, and it is the unique zero point of $\tilde{S}(x)$.

From the discussion stated above, the rate function $\tilde{S}(x)$ has only one zero point unless

$$\theta(\alpha)\lambda(\alpha) = 1, \lambda = -k(k+1), k \geq 1.$$

Under this condition, $\tilde{S}(x)$ has exactly two zero points rather than one. By the results in Remark 8, we know that large deviation principle can deduce the limit distribution only if the rate function has unique zero point, and the limit distribution is the Dirac measure at the zero point. Unfortunately, when the rate function has more than one zero points, we know only that the limit distribution supports on these zero points, but we don't know how the probability is distributed among them. Therefore, when $\theta(\alpha)\lambda(\alpha) = 1, \lambda = -k(k+1), k \geq 1$, large deviation obtained by Dr.Feng fails to give the exact limit distribution. That's why I try to figure out another way to find its limit distribution. Since $\Pi_{\alpha,\lambda}(dx)$ is the limit of

$$C_{\alpha,K} \exp\Big(\lambda\theta(\alpha)H_2(x)\Big)\hat{D}(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})(dx_1 \cdots dx_{K-1})$$

as $K \to \infty$, thus, we first try to find the limit of

$$C_{\alpha,K} \exp\Big(\lambda\theta(\alpha)H_2(x)\Big)\hat{D}(\frac{\alpha}{K-1}, \cdots, \frac{\alpha}{K-1})(dx_1 \cdots dx_{K-1})$$

as $\alpha \to 0$, then we manage to deal with $\Pi_{\alpha,\lambda}(dx)$. In this thesis, I only find the limit for $K = 2$, which turns out to be unexpected. Finally, $\lambda = -k(k+1)$ seems to be very special in Dr.Feng's result, it might be because of its connection with coalescent process.

Chapter 4

Main Results and Proof

As we discuss at the end of last chapter, for $K = 2$, the distribution in question is

$$\mu_\alpha(dx) = C_{\alpha,2}x^{\alpha-1}(1-x)^{\alpha-1}e^{\lambda\log\alpha(x^2+(1-x)^2)}dx,$$

where $C_{\alpha,2}$ is a normalized constant. It is apparently a measure on $[\frac{1}{2}, 1]$. In this chapter, we proceed to find the limit of $\mu_\alpha(dx)$ as $\alpha \to 0$.

Theorem 6. *Suppose $\mu_\alpha(dx)$ is a probability measure on $[\frac{1}{2}, 1]$, which is defined as*

$$\mu_\alpha(dx) = C_\alpha x^{\alpha-1}(1-x)^{\alpha-1}e^{\lambda\log\alpha(x^2+(1-x)^2)}dx.$$

Let $\alpha \to 0$, then the limit of this distribution is

$$(1 - \frac{1}{A(\lambda)})\delta_1(dx) + \frac{1}{A(\lambda)}\delta_{\frac{1}{2}}(dx),$$

where $A(\lambda) = +\infty$, for $\lambda \leq 2$. Otherwise, $A(\lambda) = 1$.

The approach we take here is to compute the limit of Laplace transform of measure μ_α. Before we go to the details, some useful lemmas should be mentioned here.

Lemma 1. *[18] Suppose*

$$\lim_{n\to+\infty}\frac{b_n}{a_n} = C \in \bar{\mathbb{R}},$$

where a_n, b_n are both positive, then

$$\lim_{x\to+\infty}\frac{\sum_{n=0}^{+\infty} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} = C.$$

Remark 14. *If, however, a_n, b_n are also function of x, say $a_n(x), b_n(x)$, and*

$$\lim_{n\to+\infty}\sup_{x\geq 0}\frac{b_n(x)}{a_n(x)} = C,$$

then the same conclusion is also followed by the similar argument!

Lemma 2. *[10] If $a_j \in \mathbb{C}, |a_j| < 1$, then the partial product $\prod_{j=1}^{N}(1+|a_j|)$ satisfies*

$$\exp\Big(\frac{1}{2}\sum_{j=1}^{N}|a_j|\Big) \leq \prod_{j=1}^{N}(1+|a_j|) \leq \exp\Big(\sum_{j=1}^{N}|a_j|\Big).$$

[Proof of the theorem]:

$$E_\alpha e^{-tX} = \frac{\int_{\frac{1}{2}}^{1} x^{\alpha-1}(1-x)^{\alpha-1}e^{\lambda(\log\alpha)(x^2+(1-x)^2)}e^{-tx}dx}{\int_{\frac{1}{2}}^{1} x^{\alpha-1}(1-x)^{\alpha-1}e^{\lambda(\log\alpha)(x^2+(1-x)^2)}dx}$$

$$= e^{-t}\frac{\int_{\frac{1}{2}}^{1} x^{\alpha-1}(1-x)^{\alpha-1}e^{\lambda(\log\alpha)(x^2+(1-x)^2)}e^{(1-x)t}dx}{\int_{\frac{1}{2}}^{1} x^{\alpha-1}(1-x)^{\alpha-1}e^{\lambda(\log\alpha)(x^2+(1-x)^2)}dx}$$

$\Big($By substitution, put $u = 1 - x\Big)$

$$= e^{-t}\frac{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{\lambda(\log\alpha)(u^2+(1-u)^2)}e^{ut}du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{\lambda(\log\alpha)(u^2+(1-u)^2)}du}$$

$\Big($Cancel the common term $e^{\lambda\log\alpha}\Big)$

$$= e^{-t}\frac{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{2\lambda(\log\frac{1}{\alpha})u(1-u)}e^{ut}du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{2\lambda(\log\frac{1}{\alpha})u(1-u)}du}$$

$$= e^{-t}\frac{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{2\lambda(\log\frac{1}{\alpha})u(1-u)}\sum_{n=0}^{+\infty}(\frac{u^nt^n}{n!})du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{2\lambda(\log\frac{1}{\alpha})u(1-u)}du}$$

$\Big($Since power series can be regarded as integration with respect to counting measure, and both integrands are positive, by Fubini's theorem, switch the integration and summation$\Big)$

$$= e^{-t}\sum_{n=0}^{+\infty}\frac{t^n}{n!}\frac{\int_0^{\frac{1}{2}} u^{\alpha+n-1}(1-u)^{\alpha-1}e^{2\lambda(\log\frac{1}{\alpha})u(1-u)}du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}e^{2\lambda(\log\frac{1}{\alpha})u(1-u)}du}$$

$$= e^{-t}\sum_{n=0}^{+\infty}\frac{t^n}{n!}\frac{\int_0^{\frac{1}{2}} u^{\alpha+n-1}(1-u)^{\alpha-1}\sum_{k=0}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}u^k(1-u)^k du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}\sum_{k=0}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}u^k(1-u)^k du}$$

$\Big($Similarly, apply Fubini's theorem, switch the integration and summation$\Big)$

$$= e^{-t}\sum_{n=0}^{+\infty}\frac{t^n}{n!}\frac{\sum_{k=0}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+n+k-1}(1-u)^{\alpha+k-1}du}{\sum_{k=0}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}.$$

Put

$$a_n(\alpha) = \frac{\sum_{k=0}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+n+k-1}(1-u)^{\alpha+k-1}du}{\sum_{k=0}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}, \quad n \geq 0,$$

then $a_0(\alpha) = 1$. For $n \geq 1, a_n(\alpha) = \Omega_1 + \Omega_2$, where

$$\Omega_1 = \frac{\int_0^{\frac{1}{2}} u^{\alpha+n-1}(1-u)^{\alpha-1}du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}du + \sum_{k=1}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}$$

and

$$\Omega_2 = \frac{\sum_{k=1}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}du + \sum_{k=1}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}.$$

Actually,

$$\Omega_1 = \frac{\frac{\int_0^{\frac{1}{2}} u^{\alpha+n-1}(1-u)^{\alpha-1}du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1}du}}{1+D(\alpha)},$$

and

$$\Omega_2 = \frac{\frac{\sum_{k=1}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+n+k-1}(1-u)^{\alpha+k-1}du}{\sum_{k=1}^{+\infty}\frac{(2\lambda\log\frac{1}{\alpha})^k}{k!}\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}}{1+\frac{1}{D(\alpha)}},$$

where

$$D(\alpha) = \sum_{k=1}^{+\infty} \frac{(2\lambda \log \frac{1}{\alpha})^k}{k!} \frac{\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1} du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1} du}.$$

Put $A(\lambda) = 1 + \frac{1}{\lim_{\alpha\to 0} D(\alpha)}$, and we claim that

$$\lim_{\alpha\to 0} \frac{\sum_{k=1}^{+\infty} \frac{(2\lambda \log \frac{1}{\alpha})^k}{k!} \int_0^{\frac{1}{2}} u^{\alpha+n+k-1}(1-u)^{\alpha+k-1} du}{\sum_{k=1}^{+\infty} \frac{(2\lambda \log \frac{1}{\alpha})^k}{k!} \int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1} du} = \frac{1}{2^n}, n \geq 1, \tag{4.1}$$

and

$$\lim_{\alpha\to 0} \Omega_1 = 0.$$

Then $\lim_{\alpha\to 0} \Omega_2 = \frac{1}{A(\lambda)} \frac{1}{2^n}, \quad n \geq 1$. Define

$$\phi(t) = \lim_{\alpha\to 0} E_\alpha(e^{-tX}).$$

Since $\frac{\int_0^{\frac{1}{2}} u^{\alpha+n-1}(1-u)^{\alpha-1} e^{2\lambda \log \frac{1}{\alpha} u(1-u)} du}{\int_0^{\frac{1}{2}} u^{\alpha-1}(1-u)^{\alpha-1} e^{2\lambda \log \frac{1}{\alpha} u(1-u)} du} \leq 1, \forall n \geq 1$, then by Lebesgue Convergent Theorem, the above limit is

$$e^{-t}(1 + \sum_{n=1}^{+\infty} \frac{t^n}{n!} \lim_{\alpha\to 0} a_n(\alpha)),$$

where

$$\lim_{\alpha\to 0} a_n(\alpha) = \lim_{\alpha\to 0} \Omega_1 + \lim_{\alpha\to 0} \Omega_2$$
$$= \frac{1}{A(\lambda)} \frac{1}{2^n}.$$

Therefore,

$$\phi(t) = e^{-t}(1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \frac{1}{A(\lambda)} \frac{1}{2^n})$$
$$= e^{-t}[1 + \frac{1}{A(\lambda)}(e^{\frac{t}{2}} - 1)]$$
$$= (1 - \frac{1}{A(\lambda)}) e^{-t} + \frac{1}{A(\lambda)} e^{-\frac{t}{2}}.$$

Then by Theorem 2, we know that

$$\mu_\alpha(dx) \to (1 - \frac{1}{A(\lambda)})\delta_1(dx) + \frac{1}{A(\lambda)} \delta_{\frac{1}{2}}(dx).$$

Now it remains to show the claims and find $A(\lambda)$. Firstly, in order to show $\lim_{\alpha\to 0}\Omega_1 = 0$, we consider the moment generating function of $Beta(\alpha+n,\alpha)$, which is

$$1+\sum_{k=1}^{+\infty}(\prod_{r=0}^{k-1}\frac{\alpha+n+r}{2\alpha+n+r})\frac{(-t)^k}{k!}. \tag{4.2}$$

Let $\alpha\to 0$, it approaches $1+\sum_{k=1}^{+\infty}\frac{(-t)^k}{k!}=e^{-t}$. Hence, by Theorem 2,

$$Beta(\alpha+n,\alpha)\to\delta_1,\ \text{as } \alpha\to 0. \tag{4.3}$$

And

$$\frac{\int_0^{\frac{1}{2}}u^{\alpha+n-1}(1-u)^{\alpha-1}du}{B(\alpha+n,\alpha)}\to 0,\quad \text{as } \alpha\to 0,$$

hence, $\lim_{\alpha\to 0}\Omega_1=0$.

Secondly, consider the remaining claim (4.1), we have, on one hand,

$$\frac{\int_0^{\frac{1}{2}}u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}\le(\frac{1}{2})^n.$$

On the other hand, $\forall\epsilon>0$, we have

$$\begin{aligned}
&\frac{\int_0^{\frac{1}{2}}u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}\\
&=\frac{\int_0^{\frac{1}{2}-\epsilon}u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}+\frac{\int_{\frac{1}{2}-\epsilon}^{\frac{1}{2}}u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}\\
&\ge\frac{\int_0^{\frac{1}{2}-\epsilon}u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}+(\frac{1}{2}-\epsilon)^n\frac{\int_{\frac{1}{2}-\epsilon}^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}\\
&\ge(\frac{1}{2}-\epsilon)^n-(\frac{1}{2}-\epsilon)^n\frac{\int_0^{\frac{1}{2}-\epsilon}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}.
\end{aligned}$$

Since

$$\lim_{k\to+\infty}\sup_{\alpha\in[0,1]}\frac{\int_0^{\frac{1}{2}-\epsilon}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}}u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}=0, \tag{4.4}$$

in fact, let $0 < \alpha < 1$

$$\frac{\int_0^{\frac{1}{2}-\epsilon} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du} \leq \frac{\int_0^{\frac{1}{2}-\epsilon} u^{k-1}(1-u)^{k-1}du}{\int_0^{\frac{1}{2}} u^{k}(1-u)^{k}du}$$
$$= \frac{\int_0^{\frac{1}{2}-\epsilon} u^{k-1}(1-u)^{k-1}du}{B(k,k)} \cdot \frac{2B(k,k)}{B(k+1,k+1)}$$

follow the argument in (4.2) and (4.3), we have $B(k,k) \to \delta_{\frac{1}{2}}$, as $k \to +\infty$, in distribution, then

$$\frac{\int_0^{\frac{1}{2}-\epsilon} u^{k-1}(1-u)^{k-1}du}{B(k,k)} \to 0, \text{as } k \to +\infty$$

then (4.4) is true. Thus

$$(\frac{1}{2}-\epsilon)^n \leq \liminf_{k\to+\infty} \frac{\int_0^{\frac{1}{2}} u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du}$$
$$\leq \limsup_{k\to+\infty} \frac{\int_0^{\frac{1}{2}} u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du} \leq (\frac{1}{2})^n.$$

Letting $\epsilon \to 0$, we have

$$\lim_{k\to+\infty} \frac{\int_0^{\frac{1}{2}} u^{\alpha+k+n-1}(1-u)^{\alpha+k-1}du}{\int_0^{\frac{1}{2}} u^{\alpha+k-1}(1-u)^{\alpha+k-1}du} = \frac{1}{2^n}.$$

If we apply lemma 1 and the remark(14), then the remaining claim (4.4) is verified!

Finally, we endeavor to find $A(\lambda)$. Firstly,

$$D(\alpha) = \sum_{k=1}^{+\infty} \frac{(2\lambda\log\frac{1}{\alpha})^k}{k!} \frac{B(\alpha+k,\alpha+k)}{B(\alpha,\alpha)}$$
$$= \sum_{k=2}^{+\infty} \frac{(2\lambda\log\frac{1}{\alpha})^k}{k!} \frac{B(\alpha+k,\alpha+k)}{B(\alpha,\alpha)} - \frac{2\lambda}{\alpha+\frac{1}{2}}\alpha\log\alpha.$$

Since $\Gamma(2x)\Gamma(\frac{1}{2}) = 2^{2x-1}\Gamma(x)\Gamma(x+\frac{1}{2})$, then for $k \geq 2$

$$\frac{B(\alpha+k,\alpha+k)}{B(\alpha,\alpha)} = \frac{\Gamma^2(\alpha+k)}{\Gamma^2(\alpha)} \frac{\Gamma(2\alpha)}{\Gamma(2(\alpha+k))}$$
$$= \frac{\Gamma^2(\alpha+k)}{\Gamma^2(\alpha)} \frac{2^{2\alpha-1}\Gamma(\alpha)\Gamma(\alpha+\frac{1}{2})}{2^{(\alpha+k)-1}\Gamma(\alpha+k)\Gamma(\alpha+k+\frac{1}{2})}$$

$$= \frac{1}{4^k}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha+k+\frac{1}{2})}$$
$$= \frac{1}{4^k}\frac{\alpha}{\alpha+\frac{1}{2}}\frac{1}{\prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l})}.$$

By lemma 2, we have

$$\exp(\frac{1}{4}\sum_{l=1}^{k-1}\frac{1}{\alpha+l}) \le \prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l}) \le \exp(\frac{1}{2}\sum_{l=1}^{k-1}\frac{1}{\alpha+l}), k \ge 2;$$

but

$$\exp(\frac{1}{4}\sum_{l=2}^{k}\frac{1}{l}) \le \exp(\frac{1}{4}\sum_{l=1}^{k-1}\frac{1}{\alpha+l}) \le \exp(\frac{1}{4}\sum_{l=1}^{k-1}\frac{1}{l}), k \ge 2,$$

and

$$\exp(\frac{1}{2}\sum_{l=2}^{k}\frac{1}{l}) \le \exp(\frac{1}{2}\sum_{l=1}^{k-1}\frac{1}{\alpha+l}) \le \exp(\frac{1}{2}\sum_{l=1}^{k-1}\frac{1}{l}), k \ge 2.$$

Since $\lim_{k\to+\infty}(\sum_{l=1}^{k}\frac{1}{l} - \log k) = C$ (Euler's Constant), then

$$\sum_{l=1}^{k}\frac{1}{l} = \log k + C + \alpha_k, \quad \text{where } \lim_{k\to+\infty}\alpha_k = 0,$$

hence, for $k \ge 2$

$$\exp(\frac{1}{4}\sum_{l=2}^{k}\frac{1}{l}) = k^{\frac{1}{4}}\exp(\frac{1}{4}(C+\alpha_k-1)),$$
$$\exp(\frac{1}{4}\sum_{l=1}^{k-1}\frac{1}{l}) = (k-1)^{\frac{1}{4}}\exp(\frac{1}{4}(C+\alpha_{k-1})),$$
$$\exp(\frac{1}{2}\sum_{l=2}^{k}\frac{1}{l}) = k^{\frac{1}{2}}\exp(\frac{1}{2}(C+\alpha_k-1)),$$
$$\exp(\frac{1}{2}\sum_{l=1}^{k-1}\frac{1}{l}) = (k-1)^{\frac{1}{2}}\exp(\frac{1}{2}(C+\alpha_{k-1})).$$

Therefore,

$$\frac{1}{\exp(\frac{1}{2}\sum_{l=1}^{k-1}\frac{1}{l})} \le \frac{1}{\prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l})} \le \frac{1}{\exp(\frac{1}{4}\sum_{l=2}^{k}\frac{1}{l})}, k \ge 2,$$

and

$$m\frac{1}{(k-1)^{\frac{1}{2}}} \leq \frac{1}{\prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l})} \leq M\frac{1}{(k)^{\frac{1}{4}}}, k \geq 2,$$

where m, M are both positive constants, independent of α. Moreover, since

$$D(\alpha) = \sum_{k=2}^{+\infty} \frac{(2\lambda \log \frac{1}{\alpha})^k}{k!} \frac{1}{4^k} \frac{\alpha}{\alpha+\frac{1}{2}} \frac{1}{\prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l})} - \frac{2\lambda}{\alpha+\frac{1}{2}}\alpha\log\alpha$$

$$= \frac{\alpha}{\alpha+\frac{1}{2}} \sum_{k=2}^{+\infty} \frac{(\lambda \log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{\prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l})} - \frac{2\lambda}{\alpha+\frac{1}{2}}\alpha\log\alpha$$

$$= \frac{1}{\alpha+\frac{1}{2}} \frac{\sum_{k=2}^{+\infty} \frac{(\lambda \log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{\prod_{l=1}^{k-1}(1+\frac{\frac{1}{2}}{\alpha+l})}}{1+\sum_{k=1}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} - \frac{2\lambda}{\alpha+\frac{1}{2}}\alpha\log\alpha,$$

then

$$\frac{m}{\alpha+\frac{1}{2}} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k-1)^{\frac{1}{2}}}}{1+\sum_{k=1}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} - \frac{2\lambda}{\alpha+\frac{1}{2}}\alpha\log\alpha$$

$$\leq D(\alpha) \leq \frac{M}{\alpha+\frac{1}{2}} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k-1)^{\frac{1}{4}}}}{1+\sum_{k=1}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} - \frac{2\lambda}{\alpha+\frac{1}{2}}\alpha\log\alpha.$$

Moreover,

$$\lim_{\alpha\to 0} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k-1)^{\frac{1}{2}}}}{1+\sum_{k=1}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} = \lim_{\alpha\to 0} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k-1)^{\frac{1}{2}}}}{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}}$$

and

$$\lim_{\alpha\to 0} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k)^{\frac{1}{4}}}}{1+\sum_{k=1}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} = \lim_{\alpha\to 0} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k)^{\frac{1}{4}}}}{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}}.$$

By lemma 1,

for $\lambda \leq 2$,

$$\lim_{\alpha\to 0} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k-1)^{\frac{1}{2}}}}{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} = 0, \lim_{\alpha\to 0} \frac{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!} \frac{(\frac{\lambda}{2})^k}{(k)^{\frac{1}{4}}}}{\sum_{k=2}^{+\infty} \frac{(\log \frac{1}{\alpha})^k}{k!}} = 0;$$

for $\lambda > 2$,

$$\lim_{\alpha\to 0}\frac{\sum_{k=2}^{+\infty}\frac{(\log\frac{1}{\alpha})^k}{k!}\frac{(\frac{\lambda}{2})^k}{(k-1)^{\frac{1}{2}}}}{\sum_{k=2}^{+\infty}\frac{(\log\frac{1}{\alpha})^k}{k!}} = +\infty,\ \lim_{\alpha\to 0}\frac{\sum_{k=2}^{+\infty}\frac{(\log\frac{1}{\alpha})^k}{k!}\frac{(\frac{\lambda}{2})^k}{(k)^{\frac{1}{4}}}}{\sum_{k=2}^{+\infty}\frac{(\log\frac{1}{\alpha})^k}{k!}} = +\infty.$$

Therefore,

for $\lambda \leq 2$,

$$\lim_{\alpha\to 0} D(\alpha) = 0;$$

for $\lambda > 2$,

$$\lim_{\alpha\to 0} D(\alpha) = +\infty.$$

Hence $A(\lambda) = 1$, for $\lambda > 2$, otherwise, $A(\lambda) = +\infty$.

[Proof of lemma 1]:

Case 1: $C = +\infty$.

$\forall L > 0, \exists N \in \mathbb{N}^+, s.t.\ \ \forall n > N, \frac{b_n}{a_n} > L$, then

$$\begin{aligned}
\frac{\sum_{n=0}^{+\infty} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} &= \frac{\sum_{n=0}^{N} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} + \frac{\sum_{n=N+1}^{+\infty} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}}\\
&= \frac{\sum_{n=0}^{N} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} + \frac{\sum_{n=N+1}^{+\infty} a_n\frac{x^n}{n!}\frac{b_n}{a_n}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}}\\
&\geq \frac{\sum_{n=0}^{N} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} + L\frac{\sum_{n=N+1}^{+\infty} a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}},
\end{aligned}$$

thus

$$\begin{aligned}
&\lim_{x\to+\infty}\frac{\sum_{n=0}^{+\infty} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}}\\
\geq\ &\lim_{x\to+\infty}\frac{\sum_{n=0}^{N} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} + L\cdot\lim_{x\to+\infty}\frac{\sum_{n=N+1}^{+\infty} a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}}.
\end{aligned}$$

Since $\lim_{x\to+\infty}\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!} = +\infty$, then

$$0 \leq \lim_{x\to+\infty}\frac{\sum_{n=0}^{N} b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty} a_n\frac{x^n}{n!}} \leq \lim_{x\to+\infty}\frac{\sum_{n=0}^{N} b_n\frac{x^n}{n!}}{\sum_{n=0}^{N+1} a_n\frac{x^n}{n!}} = 0,$$

$$\lim_{x\to+\infty}\frac{\sum_{n=N+1}^{+\infty}a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}=\lim_{x\to+\infty}\frac{\sum_{n=N+1}^{+\infty}a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}+\lim_{x\to+\infty}\frac{\sum_{n=0}^{N}a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}=1.$$

Thus,

$$\lim_{x\to+\infty}\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}\geq L,\forall L>0.$$

Hence,

$$\lim_{x\to+\infty}\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}=+\infty=C.$$

Case 2: $0\leq C<+\infty$,

$\forall\epsilon>0,\exists N\in\mathbb{N}^+, s.t.\quad\forall n>N,|\frac{b_n}{a_n}-C|<\epsilon$, then

$$\begin{aligned}&\left|\frac{\sum_{n=0}^{+\infty}b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}-C\right|\leq\frac{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}\left|\frac{b_n}{a_n}-C\right|}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}\\&\leq\frac{\sum_{n=0}^{N}a_n\frac{x^n}{n!}\left|\frac{b_n}{a_n}-C\right|}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}+\epsilon\frac{\sum_{n=N+1}^{+\infty}a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}\\&\leq(\max_{0\leq k\leq N}\frac{b_k}{a_k}+C)\frac{\sum_{n=0}^{N}a_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}+\epsilon,\end{aligned}$$

then

$$\limsup_{x\to+\infty}\left|\frac{\sum_{n=0}^{+\infty}b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}-C\right|\leq\epsilon.$$

Letting $\epsilon\to 0_+$, yields

$$\lim_{x\to+\infty}\frac{\sum_{n=0}^{+\infty}b_n\frac{x^n}{n!}}{\sum_{n=0}^{+\infty}a_n\frac{x^n}{n!}}=C.$$

[Proof of lemma 2]:

On one hand, recall that $|a_j|<1$, we have $1+|a_j|\leq e^{|a_j|}$, then

$$\prod_{j=1}^{N}(1+|a_j|)\leq\exp(\sum_{i=1}^{N}|a_j|).$$

On the other hand, since $\frac{|a_j|}{2}<\frac{1}{2}$, then

$$1+|a_j|=1+2(\frac{1}{2}|a_j|)\geq e^{\frac{|a_j|}{2}},$$

therefore

$$\prod_{j=1}^{N}(1+|a_j|) \geq \exp(\frac{1}{2}\sum_{i=1}^{N}|a_j|).$$

Chapter 5

Conclusion and Further Discussion

From theorem 6, we know that the limit distribution is either δ_1 or $\delta_{\frac{1}{2}}$ and is uniquely determined by λ, where λ can be regarded as a selection factor. Owing to selection favoring heterozygotes, as λ becomes bigger the species tends to spread out, until finally two allele types coexist in identical proportions. This makes sense because the mutation rate is symmetric.

Interestingly, however, when $\lambda \leq 2$ there is only one allele type in the end, and unexpectedly there is no intermediate state. For $\lambda = 2$ we previously expected an intermediate state in which two configurations would coexist

$$p\delta_1 + (1-p)\delta_{\frac{1}{2}}, \qquad 0 < p < 1.$$

Hence, even though the species is quite weakly affected by mutation, the configuration of the species fails to change until the selection factor is strong enough. Simply put, even though the possibility of mutating into other species is very small, whenever species become well accustomed to their surrounding natural environments, they may go through alterations and accumulate such alterations by way of inheritance. Nevertheless, it is quite unexpected that species change suddenly without any intermediate state as the environment varies. The only sound account in my opinion would be that species is very sensitive to the environment , hence the sudden change of environment results in the sudden change of species. For example, great natural disasters may suddenly occur, with the result that most species become, like the dinosaurs, extinct, and only very few survive. This may be why the configuration of species undergoes sudden changes when the selection factor is strong enough. The frequency of great natural disasters is relatively low compared with the lifespan of any given organism, but is actually fairly high compared with the entire evolutionary history of a species. Hence, many unknown species may just begin their debuts and quickly exit during the play of natural history!

Finally, in this thesis only the case $K = 2$ is considered. The associated problems of other cases are still unsolved. However, I think the method used in this case might be useful for the other cases, with possibly a little adjustment. There are, however, indeed some technical difficulties. For example, when $K = 3$, then the measure in question would be

$$\mu_\alpha(dx_1, dx_2) = C_\alpha e^{\lambda \log \alpha (x_1^2+x_2^2+x_3^2)} x_1^{\frac{\alpha}{2}-1} x_2^{\frac{\alpha}{2}-1} x_3^{\frac{\alpha}{2}-1} dx_1 dx_2$$

defined on Δ_3, where

$$\Delta_3 = \{(x_1, x_2, x_3) | x_1 + x_2 + x_3 =, x_1 \geq x_2 \geq x_3 \geq 0\}.$$

If we try to compute the limit of Laplace transform, we might need to reduce it to the case $K = 2$. But consider the x_1 cross section

$$(\Delta_3)_{x_1} = \{(x_2, x_3) | x_2 + x_3 = 1 - x_1, \min(1 - x_1, x_1) \geq x_2 \geq x_3 \geq 0\},$$

when $x_1 < \frac{1}{2}$, then $\min(1 - x_1, x_1) = x_1 \geq x_2 \geq x_3 \geq 0$, and

$$1 > \frac{x_1}{1 - x_1} \geq \frac{x_2}{1 - x_1} \geq \frac{x_3}{1 - x_1} \geq 0,$$

so $(\Delta_3)_{x_1}$ is not $\Delta_2 = \{(x_1, x_2) | x_1 + x_2 = 1, 1 \geq x_1 \geq x_2 \geq 0\}$, hence, the reduction might not work here. Some adjustments are definitely needed.

Bibliography

[1] P Blingsley, *Convergence of Probability Measure.* Wiley,New York, [1968].

[2] P Blingsley, *Probability and Measure.* Wiley,New York, [1979].

[3] Casella, George and Berger, Roger L., *Statistical Inference.* Thomson Learning, [2002].

[4] Charles Darwin, *On Origin of Species:by Means of Natural Selection.* Dover Publication, [2006].

[5] R Durrett, *Probability Theory and Examples*, Duxbury Press, [1995].

[6] S.N.Ethier and T.G.Kurtz, *The Infinitely-many-neutral-alleles Diffusion Model.* Adv.Appl.Prob., 13 429-452, [1981].

[7] S.N.Ethier and T.G.Kurtz, *Markov Process: Characterization and Convergence.* John Wiley&Sons, Inc.,New Jersey, [2005].

[8] S Feng, *Poisson-Dirichlet Distribution With Small Mutation.* Stochastic Process and Their Application, Vol.18,No.5,1794-1824, [2009].

[9] S Feng, *The Poisson-Dirichlet Distribution and Related Topics: Models and Asymptotic Behaviors.* Springer,Berlin, [2010].

[10] R.E. Green and S.G. Krantz, *Function Theory of One Complex Variable.* John Wiley&Sons, Inc. New York, [1997].

[11] R.V.H Hogg and A.T. Craig, *Introduction to Mathematical Statistics.*Fourth Edition, Macmillan Publishing Co.,Inc., [1978].

[12] K. Itô, On Stochastic Differential Equations, Mem.Amer.Math.Soc., 4, [1951]

[13] J.C.F. Kingman, *The coalescent.* J.Stoch.Proc.Appl.,13, 235-248.[1982].

[14] J.C.F. Kingman, *Random Discrete Distribution.* J.R.Statist.Soc. 1337, 1-22.

[15] M Morange, *A History of Molecular Biology*, Harvard University Press, [2000].

[16] B Øksendal, *Stochastic Differential Equation: An introduction and Application.* Fifth Edition,Springer, [2000].

[17] D Revuz and M Yor, *Continuous Martingale and Brownian Motion*, Third Edition, Springer,[2005].

[18] W Rudin. *Principle of Mathematical Analysis.* Third Edition, McGrawHill,Inc., [1976].

[19] G.A. Watterson, *The Stationary Distribution of Infinite-many Netural Alleles Diffusion*, J.Appl.Probab.13,639-651,[1976].