ASYMPTOTIC THEORY FOR THREE INFINITE DIMENSIONAL DIFFUSION PROCESSES

ASYMPTOTIC THEORY FOR THREE INFINITE DIMENSIONAL DIFFUSION PROCESSES

by YOUZHOU ZHOU, M.Sc. B.Sc.

A Thesis Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy

McMaster University

McMaster University; DOCTOR OF PHILOSOPHY (2013);
Hamilton, Ontario (Department of Mathematics & Statistics)

TITLE:	Asymptotic Theory for Three Infinite Dimensional Diffusion Processes
AUTHOR:	Youzhou Zhou, M.Sc. B.Sc.
SUPERVISOR:	Professor Shui Feng

NUMBER OF PAGES: vi, 125

Abstract

This thesis is centered around three infinite dimensional diffusion processes:

(i). the infinitely-many-neutral-alleles diffusion model [Ethier and Kurtz, 1981],

(ii). the two-parameter infinite dimensional diffusion model [Petrov, 2009] and [Feng and Sun, 2010],

(iii). the infinitely-many-alleles diffusion with symmetric dominance [Ethier and Kurtz, 1998].

The partition structures, the ergodic inequalities and the asymptotic theory of these three models are discussed. In particular, the asymptotic theory turns out to be the major contribution of this thesis.

In Chapter 2, a slightly altered version of Kingman's one-to-one correspondence theorem on partition structures is provided, which in turn becomes a handy tool for obtaining the asymptotic result on the partition structures associated with models (i) and (ii).

In Chapter 3, the three diffusion models are briefly introduced. New representations of the transition densities of models (i) and (ii) are obtained simply by rearranging the previous representations obtained in [Ethier, 1992] and [Feng et al., 2011] respectively. These two new representations have their own advantages, by making use of which the corresponding ergodic inequalities easily follow. Furthermore, thanks to the functional inequalities in [Feng et al., 2011], the ergodic inequality for model (iii) becomes available as well.

In Chapter 4, the asymptotic properties of models (i) and (ii) are thoroughly studied. Various asymptotic results are obtained, such as the weak limits of models (i) and (ii) at different time scales when the mutation rate approaches ∞, and the large deviation principle for models (i) and (ii) at a fixed time, and that of the transient partition structures of models (i) and (ii). Of all these results, the weak limit and the large deviation principle of the transient partition structures are of particular interest.

In Chapter 5, the asymptotic results on the stationary distribution and the transient distribution of model (iii) are both obtained. The weak limit of the infinitely-many-alleles diffusion with symmetric overdominance at fixed time t serves as an alternative answer to Gillespie's conjecture [Gillespie, 1999]. The weak limit of the stationary distribution of the infinitely-many-alleles diffusion with symmetric overdominance provides a complete solution to the remaining problem in [Feng, 2009].

Acknowledgements

I am greatly indebted to various people during my PhD studies. Firstly, I would like to thank my PhD supervisor, professor Shui Feng, for his wise mentoring. This unforgettable mentorship will definitely becomes an everlasting source of encouragement and perseverance for me. Secondly, I also want to thank Professor Fred M. Hoppe and Professor Romän Viveros-Aguilera for their amazing suggestions, professional and personal as well. Moreover, my appreciation also goes to McMaster University, whose financial support made this thesis possible. Thirdly, my unspeakable gratitude shall belong to my beautiful and kind wife, whose understanding and love are a miraculous potion, washing away all my apprehensions. Last but not least, I must thank my parents and uncle for their nurturing. My journey of study as a student finally comes to an end. Hopefully, this thesis and my future achievements can make my parents proud in heaven!

Contents

Chapter 1

Introduction

This chapter serves as a road map of the thesis. As it proceeds, we will introduce some background information concerned with population genetics and the three infinite dimensional diffusion models mentioned in the title of this thesis. Next, the major contributions of this thesis will also be outlined briefly. Finally, the structure of this thesis will be briefly mentioned as well.

1.1 Background Information

Population genetics is a branch of biology. It studies the evolution of a whole population by considering the development of its gene frequencies. Thanks to modern technological achievements in biology, we now know that the genetic material of the majority of organisms is DNA, a double helix structure. It can copy itself almost exactly but sometimes commits copying errors as well. These copying errors are usually called mutations in biology. A segment at a certain position of a DNA sequence usually determines the formation of some proteins. Such a segment is then called an allele and the position where it resides is called its locus. Alleles can have various alternatives. The totality of all alleles is called a gene pool. Some species need two alternative alleles to determine the formation of its proteins, some, only one. We call the former diploid species and the latter haploid species. A diploid individual with two identical alleles is called a homozygote, otherwise a heterozygote.

In 1866, Gregor Mendel postulated an inheritance pattern of peas. It is a random pattern, whereas an ecosystem usually exhibits a directional evolution. Charles Darwin's natural selection theory explains this directional evolution at least to some extent. To reconcile the random pattern and the directional evolutionary phenomena, R. Fisher, S. Wright and J. Haldane used mathematical tools to build population genetic models consecutively which provide theoretical explanations for various evolutionary phenomenon. The Wright-Fisher model is one of the most influential population genetic models. It considers the allele frequencies at a single locus. Presumably the allele frequencies in a given population can be influenced by various evolutionary forces, such as mutations, selections, recombinations, immigrations and, of course, some random forces. In the Wright-Fisher model, random forces, mutations and selections are usually included. In particular, random forces are called random sampling in the Wright-Fisher model. If we only include mutation and random sampling, this type of model is called a neutral model. Moreover, if selection is also involved, we call it a selective model.

We can construct a Wright-Fisher model for both haploid species and diploid species. Neutral models for haploid species and neutral models for diploid species are mathematically equivalent, the selective models for these two species, however, are essentially different. In this thesis, we will only focus on models for diploid species.

Next the construction of the Wright-Fisher model will be briefly reviewed.

Suppose that the gene pool consists of K alternative alleles, $A_1, A_2, \cdots, A_K$. For diploid species, each pair $A_iA_j, 1 \leq i, j \leq K$, is called a genotype and can control one specific genetic characteristic. Therefore there are $\binom{K}{2} + K$ genotypes. In the Wright-Fisher model, the population size N is fixed. Thus, for each generation, there are always $2N$ genes. The development of each generation goes through three stages. For the sake of mathematical simplicity, we assume that they are mutation, selection and random sampling. Let $u_{ij}, 1 \leq i, j \leq K, i \neq j$ be the probability that A_i mutates to A_j and $u_{ii} = 1 - \sum_{j \neq i} u_{ij}$ be the probability that no mutation occurs to type A_i. Let $\omega_{ij}, 1 \leq i, j, \leq K$ be the relative fitness of genotype A_iA_j and $\omega_{ij} = \omega_{ji}$. Define $Y_i(n), 1 \leq i \leq K$, to be the frequency of A_i's in the nth generation; then $Y(n) = (Y_1(n), Y_2(n), \cdots, Y_K(n))$ is a K-dimensional Markov chain, the so-called Wright-Fisher model.

For the neutral model, the one step transition probability is

$$P(y, x) = \frac{(2N)!}{x_1! x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K},$$

where $x = (x_1, x_2, \cdots, x_K), y = (y_1, y_2, \cdots, y_K)$, and $x_i, y_i, 1 \leq i \leq K$, are non-negative integers such that

$$\sum_{i=1}^{K} x_i = \sum_{i=1}^{K} y_i = 2N,$$

and $p_i, 1 \leq i \leq K$, is the relative frequency of A_i after mutations; thus,

$$p_i = \left(1 - \sum_{j \neq i} u_{ij}\right) \frac{y_i}{2N} + \sum_{j \neq i} u_{ji} \frac{y_j}{2N} = \sum_{j} u_{ji} \frac{y_j}{2N}.$$

For selective models, the one step transition probability is

$$P'(y, x) = \frac{(2N)!}{x_1! x_2! \cdots x_K!} q_1^{x_1} q_2^{x_2} \cdots q_K^{x_K},$$

where $q_i, 1 \le i \le K$, is the relative frequency of A_i after mutations and selections. Thus,

$$q_i = \frac{\sum_{j=1}^K \omega_{ij} p_i p_j}{\sum_{k,l=1}^K \omega_{kl} p_k p_l}.$$

We can also consider the evolution of the relative frequencies of alleles. Letting $X_i^{2N}(n) = \frac{Y_i(n)}{2N}, 1 \le i \le K$, and $X^{2N}(n) = (X_1^{2N}(n), X_2^{2N}(n), \cdots, X_K^{2N}(n))$, we know $X^{2N}(n)$ is also a Markov chain with transition probability similar to that of $Y(n)$. Under an appropriate scaling transformation, such as,

$$u_{ij} = \frac{\mu_{ij}}{2N}(i \neq j), \quad \omega_{ij} = 1 + \frac{\sigma_{ij}}{2N},$$

$X^{2N}([2Nt])$, as $N \to +\infty$, will converge to a diffusion process X_t^K, the so-called Wright-Fisher diffusion, characterized by the generator

$$G_K = \frac{1}{2} \sum_{i,j=1}^K x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_i} + \sum_{i=1}^K b_i(x) \frac{\partial}{\partial x_i}.$$

Here

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

The first part of $b_i(x)$ comes from mutations, and the second part of $b_i(x)$ comes from selections. Furthermore, μ_{ij} and σ_{ij} are respectively called population mutation rate and selection intensity. If $\sigma_{ij} > 0$, then the genotype A_iA_j is favoured, otherwise A_iA_j is not favoured. The diffusion process X_t^K lives on a $(K-1)$-dimensional simplex

$$\Delta_K := \left\{ x = (x_1, x_2, \cdots, x_K) \middle| x_1, x_2, \cdots, x_K \ge 0, \sum_{i=1}^K x_i = 1 \right\}.$$

X_t^K is reversible if and only if the mutation rate is of parent independent type (refer to [Overbeck and Röckner, 1997], [Li et al., 1999]), that is, $\mu_{ij} = \frac{\theta_j}{2} > 0, \quad i \neq j$, in which case the unique stationary distribution is

$$\Pi(dx) = C x_1^{\theta_1 - 1} x_2^{\theta_2 - 1} \cdots x_K^{\theta_K - 1} \exp\left\{ \sum_{i,j=1}^K \sigma_{ij} x_i x_j \right\} dx_1 \cdots dx_{K-1}.$$

When $\sigma_{ij} = 0$, this model reduces to the neutral model, the stationary distribution of which is just Dirichlet distribution, $Dir(\theta_1, \theta_2, \cdots, \theta_K)$.

Moreover, instead of considering the relative frequencies of alleles, we can also consider the empirical distribution of alleles in the gene pool. Thus, we have a measure-valued Markov chain. Through the same diffusion approximation procedure, we will end up with a measure-valued diffusion process, called Fleming-Viot process. Please refer to [Ethier and Kurtz, 1993] for a comprehensive introduction.

1.2 Three Types of Diffusion Processes

Since the gene pool could consist of infinitely many alternative alleles, we need to consider infinite dimensional models. There are three infinite dimensional diffusion processes considered in this thesis. They are:

(i). The infinitely-many-neutral-alleles diffusion model

(ii). The infinitely-many-alleles diffusion model with symmetric dominance

(iii). The two-parameter infinite dimensional diffusion

For short, we will call them one-parameter neutral model, one-parameter selective model and two-parameter model in the subsequent chapters.

1.2.1 The Infinitely-many-neutral-alleles Diffusion

For the K-dimensional neutral model, if we consider symmetric mutation rates, say $\mu_{ij} = \frac{\theta}{2(K-1)}, i \neq j$, then

$$b_i(x) = \frac{1}{2}\left(-\theta x_i + \frac{\theta}{K-1}(1-x_i)\right).$$

When $K \to +\infty$, the decreasingly ordered X_t^K will approach the infinitely-many-neutral-alleles diffusion X_t, the generator of which is

$$G = \frac{1}{2}\sum_{i,j=1}^{\infty} x_i(\delta_{ij} - x_j)\frac{\partial^2}{\partial x_i \partial x_j} - \frac{\theta}{2}\sum_{i=1}^{\infty} x_i \frac{\partial}{\partial x_i}.$$

This process X_t was first constructed in [Ethier and Kurtz, 1981] and is a reversible diffusion process which lies in the Kingman simplex

$$\bar{\nabla}_\infty = \left\{ x = (x_1, x_2, \cdots,) \in [0,1]^\infty \middle| x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} x_i \leq 1 \right\}.$$

The unique stationary distribution of X_t is the Poisson-Dirichlet distribution PD(θ) first introduced in [Kingman et al., 1975]. Please refer to Chapter 2 of this thesis for a comprehensive introduction to both the Kingman simplex and the Poisson-Dirichlet distribution.

Moreover, the neutral Fleming-Viot process with parent independent mutation operator almost surely lives on a space consisting of purely atomic probability measures. Define a map

$$\rho : \mu \to (x_1, x_2, \cdots),$$

where x_i is the mass of the ith greatest atom of μ. Under the map ρ, the image of the neutral Fleming-Viot process with parent independent mutation operator is X_t.

1.2.2 The Infinitely-many-alleles Diffusion Model with Symmetric Dominance

Fleming-Viot processes can be quite general; they can include various evolutionary forces. We usually call Fleming-Viot processes labelled models and their atomic projections unlabelled models. In [Ethier and Kurtz, 1998], an unlabelled model with symmetric overdominance X_t^σ is obtained. The generator of X_t^σ is

$$G_\sigma = G + \sigma \sum_{i=1}^{\infty} x_i \left(x_i - \sum_{j=1}^{\infty} x_j^2 \right) \frac{\partial}{\partial x_j}.$$

The image of Fleming-Viot processes with selection under the map ρ is usually non-Markovian, but, under the map ρ, the image of Fleming-Viot processes with symmetric dominance is Markovian. By symmetric dominance, we mean $\sigma_{ij} = \sigma\delta_{ij}$. More specifically, when $\sigma < 0$, heterozygotes are favoured, corresponding to overdominance. When $\sigma > 0$, homozygotes are favoured, corresponding to underdominance. When $\sigma = 0$, this process reduces to the infinitely-many-neutral-alleles diffusion. We call X_t^σ the infinitely-many-alleles diffusion with symmetric dominance. X_t^σ is a reversible diffusion and its unique stationary distribution is

$$\pi_\sigma(dx) = C_\sigma \exp\left\{ \sigma \sum_{i=1}^{\infty} x_i^2 \right\} \mathrm{PD}(\theta)(dx).$$

1.2.3 Two-parameter Infinite Dimensional Diffusion

The Poisson-Dirichlet distribution has a two-parameter extension called two-parameter Poisson-Dirichlet distribution $\mathrm{PD}(\theta, \alpha)$, where $\theta + \alpha > 0, \alpha \in [0, 1)$. In [Petrov, 2009] and [Feng and Sun, 2010], an unlabelled infinite dimensional diffusion $X_t^{\theta,\alpha}$ was constructed; and we call $X_t^{\theta,\alpha}$ a two-parameter infinite dimensional diffusion. The generator of $X_t^{\theta,\alpha}$ is

$$G^{\theta,\alpha} = \frac{1}{2} \sum_{i,j=1}^{\infty} x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} - \frac{1}{2} \sum_{i=1}^{\infty} (\theta x_i + \alpha) \frac{\partial}{\partial x_i}.$$

$X_t^{\theta,\alpha}$ is also a reversible diffusion process, and its unique stationary distribution is the two-parameter Poisson-Dirichlet distribution PD(θ, α). Though we still call $X_t^{\theta,\alpha}$ an unlabelled model, its labelled model is still missing, and how to construct it still remains an open problem.

1.3 Main Results

In this thesis, several results associated with population genetic models are obtained. They mainly belongs to two categories: ergodic inequalities and asymptotic theory.

1.3.1 Sampling Formula

Among all results in population genetics, Ewens sampling formula is a very important result. It was generalized to random partition structures by J.F.C. Kingman, and just like allele frequencies it characterizes the population configuration in some sense.

Let $x = (x_1, x_2, \cdots) \in \bar{\nabla}_\infty$ be the allele spectrum. Define

$$\nabla_\infty = \left\{ x \in \bar{\nabla}_\infty \Big| \sum_{i=1}^{\infty} x_i = 1 \right\}.$$

If $x \in \nabla_\infty$, then we say there are countably many alleles in the gene pool; if, however, $\sum_{i=1}^{\infty} x_i < 1$, then we say there are uncountably many alleles in the gene pool. In the latter case, the gene pool consists of a discrete allele spectrum and a continuous allele spectrum. The probability weights of alleles in the discrete allele spectrum give rise to x, and the weight of the continuous allele spectrum is just $1 - \sum_{i=1}^{\infty} x_i$. This idea was communicated to J.F.C. Kingman by G.A. Watterson in [Kingman, 1978a].

Suppose that we want to draw a random sample S of size n from a population with the allele spectrum $x = (x_1, x_2, \cdots)$. If $\sum_{i=1}^{\infty} x_i = 1$, then the sample distribution is

$$P(S = (\alpha_1, \alpha_2, \cdots, \alpha_n)) = \frac{n!}{\prod_{i=1}^{n} (i!)^{\alpha_i}} \sum x_1^{n_1} x_2^{n_2} \cdots, \tag{1.3.1}$$

where, for $1 \le i \le n$, n_i means there are n_i alleles of weight x_i in the sample, and α_i means there are α_i types of alleles that show up exactly i times in the sample.

A sample with only one allele type has probability $\varphi_n(x) = \sum_{i=1}^{\infty} x_i^n, n \ge 2$. $\varphi_2(x)$ is called homozygosity in population genetics; $1 - \varphi_2(x)$ is called heterozygosity.

If $\sum_{i=1}^{\infty} x_i < 1$, the Kingman paintbox idea [Kingman, 1978b] enables us to have an expression similar to (1.3.1) for the sample distribution. If the spectrum x is random, then the sampling formula is

$$P(S = (\alpha_1, \alpha_2, \cdots, \alpha_n)) = E\Big(\frac{n!}{\prod_{i=1}^{n} (i!)^{\alpha_i}} \sum x_1^{n_1} x_2^{n_2} \cdots \Big), \tag{1.3.2}$$

where x should be changed associatively by the paintbox idea if $P(\sum_{i=1}^{\infty} x_i < 1) > 0$. As stated in [Watterson, 1977a], the Ewens sampling formula can be obtained from (1.3.2) if x follows the Poisson-Dirichlet distribution PD(θ). Kingman's one-to-one correspondence asserts that every partition structure can be uniquely expressed as (1.3.2).

In this thesis, we rewrite the sample distribution (1.3.1) as

$$\begin{aligned} p_\eta(x) =& \frac{n!}{\eta_1! \cdots \eta_l! \alpha_1! \cdots \alpha_n!} \sum_{d=1}^{l} (-1)^{l-d} \sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \\ & \varphi_{\sum_{i \in \beta_1} \eta_i}(x) \cdots \varphi_{\sum_{i \in \beta_d} \eta_i}(x), \end{aligned}$$

where $\eta = (\eta_1, \cdots, \eta_l)$ is an integer partition of n, $\alpha_i = \#\{j | \eta_j = i, 1 \leq j \leq l\}$, β is the partition of the set $\{1, 2, \cdots, l\}$ into d components, satisfying

$$\min \beta_1 < \min \beta_2 < \cdots < \min \beta_d,$$

$|\beta_i|$ denotes the number of elements in β_i, and $\varphi_1(x) = 1, \varphi_k(x) = \sum_{i=1}^{\infty} x_i^k, k \geq 2$. Since $\varphi_k(x), k \geq 2$, are continuous on $\bar{\nabla}_\infty$, clearly $p_\eta(x)$ is a continuous extension of (1.3.1). Thus, the Kingman one-to-one correspondence can also have the representation

$$P(S = \eta) = \int_{\bar{\nabla}_\infty} p_\eta(y) \mu(dy),$$

where μ is the distribution of the spectrum x. This new representation should be a well-known fact, but the author has not found the explicit $p_\eta(x)$ in any reference. Therefore, we include it here for the sake of completeness. Since p_η is continuous, the weak convergence of μ will lead to the pointwise convergence of the sampling formula.

1.3.2 Ergodic Inequalities

Surprisingly, the explicit transition density functions of the one-parameter neutral model and the two-parameter model can be obtained through the eigen expansions of their generators [Griffiths, 1979], [Ethier, 1992] and [Feng et al., 2011]. The structures of their transition density functions are quite similar.

In [Ethier, 1992], the transition density function of the one-parameter neutral model has the following representation:

$$\begin{aligned} p(t, x, y) &= 1 + \sum_{m=2}^{\infty} \exp\{-\lambda_m t\} Q_m(x, y), \\ Q_m(x, y) &= \frac{2m + \theta - 1}{m!} \sum_{n=0}^{m} (-1)^{m-n} \binom{m}{n} (n + \theta)_{(m-1)} p_n(x, y), \end{aligned}$$

$$p_n(x,y) = \sum_{|\eta|=n} \frac{p_\eta(x)p_\eta(y)}{\int_{\bar{\nabla}_\infty} p_\eta d\mathrm{PD}(\theta)}.$$

In [Feng et al., 2011], the transition density function of the two-parameter model has a similar representation:

$$\begin{aligned}
p^{\theta,\alpha}(t,x,y) &= 1 + \sum_{m=2}^{\infty} \exp\{-\lambda_m t\} Q_m^{\theta,\alpha}(x,y), \\
Q_m^{\theta,\alpha}(x,y) &= \frac{2m+\theta-1}{m!} \sum_{n=0}^{m} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n^{\theta,\alpha}(x,y), \\
p_n^{\theta,\alpha}(x,y) &= \sum_{|\eta|=n} \frac{p_\eta(x)p_\eta(y)}{\int_{\bar{\nabla}_\infty} p_\eta d\mathrm{PD}(\theta,\alpha)}.
\end{aligned}$$

In [Ethier, 1992], the following ergodic inequality of the one-parameter neutral model is also obtained

$$\sup_{x\in\bar{\nabla}_\infty} \left\| P(t,x,\cdot) - \mathrm{PD}(\theta)(\cdot) \right\|_{\mathrm{var}} \leq \frac{(\theta+2)(\theta+3)}{2} \exp\{-(\theta+1)t\}, \quad \forall t \geq 0. \tag{1.3.3}$$

In [Feng and Sun, 2010], the two-parameter model is shown to be an ergodic diffusion; but an ergodic inequality similar to (1.3.3) is not obtained.

In this thesis, we reorganize the expression for the transition density function of the one-parameter neutral model and the two-parameter model; then we have the following representation of the transition density functions for the one-parameter neutral model:

$$\begin{aligned}
p(t,x,y) &= d_0^\theta(t) + d_1^\theta(t) + \sum_{n=2}^{+\infty} d_n^\theta(t) p_n(x,y), \\
d_0^\theta(t) &= 1 - \sum_{m=1}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} (-1)^{m-1} \theta_{(m-1)} \\
d_n^\theta(t) &= \sum_{m=n}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} \binom{m}{n} (-1)^{m-n} (\theta+n)_{(m-1)}, n \geq 1.
\end{aligned}$$

Similarly, we have the following representation of transition density function for the two-parameter model:

$$p^{\theta,\alpha}(t,x,y) = d_0^\theta(t) + d_1^\theta(t) + \sum_{n=2}^{+\infty} d_n^\theta(t) p_n^{\theta,\alpha}(x,y). \tag{1.3.4}$$

By making use of (1.3.4), we have the ergodic inequality

$$\sup_{x\in\bar{\nabla}_\infty} \left\| P^{\theta,\alpha}(t,x,\cdot) - \mathrm{PD}(\theta,\alpha)(\cdot) \right\|_{\mathrm{var}} \leq \frac{(\theta+2)(\theta+3)}{2} \exp\{-(\theta+1)t\}, \quad \forall t \geq 0. \tag{1.3.5}$$

However, an explicit expression for the transition density function of the one-parameter selective model is still unknown. In [Barbour et al., 2000], an expression for the K-dimensional Wright-Fisher diffusion with selection is obtained; but some quantities in this expression are unlikely to have explicit forms. In [Ethier and Kurtz, 1998], the one-parameter selective model is shown to be uniformly strongly ergodic; whereas an ergodic inequality similar to (1.3.5) is still unknown.

In this thesis, by applying the methods in [Chen, 2005] and the functional inequalities in [Feng et al., 2011] , we can obtain an ergodic inequality analogous to (1.3.5):

$$\sup_{x\in\bar{\nabla}_\infty} \left\| P^{\sigma}(t,x,\cdot) - \pi_\sigma(\cdot) \right\|_{\mathrm{var}} \leq K(\theta,\sigma) \exp\{-t \text{ gap } (G_\sigma)\}, \quad \forall t \geq 0.$$

1.3.3 Asymptotic Results

As the title of this thesis indicates, asymptotic results are the main topic of this thesis. We focus on the asymptotic theory for the three infinite dimensional diffusions.

Asymptotic results for the one-parameter neutral model and the two-parameter model:

In population genetics, random sampling, mutation and selection are the three most commonly studied evolutionary forces. Mutations always increase gene variation by adding new types to the population; while random sampling washes away parts of these new types. The primary goal of population genetic model is to understand the interactions between these three forces. As Gillespie argued in [Gillespie, 2004], the time required for these evolutionary forces to have appreciable effects on the genetic configuration can quantify these forces. It is stated in [Gillespie, 2004] that the time scale of random sampling is $2N\log 2$, and the time scale of mutations is $\frac{4N}{\theta}$, where $\theta = 4Nu$, and u is the individual mutation rate. In all these quantities, N means the population size. Therefore, the time scale of random sampling in the diffusion models is $\log 2$ and the time scale of mutation is $\frac{2}{\theta}$. As $N \to +\infty$, $\theta \to +\infty$; then $\frac{2}{\theta} << \log 2$. Therefore, mutations will introduce many new types immediately, and the allele spectrum should be continuous. If, however, $\frac{2}{\theta} >> \log 2$, then the allele spectrum should be discrete. To verify the above intuition, we should look at the asymptotic distribution of X_t as $\theta \to +\infty$.

In this thesis, we have obtained the weak limit of $X_{t(\theta)}$, as $\theta \to +\infty$:

$$X_{t(\theta)} \to \begin{cases} \delta_{(0,0,\cdots)}, & \text{if } \lim_{\theta\to+\infty} \theta t(\theta) = +\infty \\ \delta_{e^{-\frac{c}{2}}x}, & \text{if } \lim_{\theta\to+\infty} \theta t(\theta) = c > 0 \\ \delta_x, & \text{if } \lim_{\theta\to+\infty} \theta t(\theta) = 0, \end{cases}$$

where $x \in \bar{\nabla}_\infty$ and is the starting point of X_t.

The above result finds the asymptotic distribution of $X_{\log 2}$ and $X_{\frac{2}{\theta}}$: the former is $\delta_{(0,0,\cdots)}$ and the latter is $\delta_{e^{-1}x}$. $\delta_{(0,0,\cdots)}$ means a purely continuous allele spectrum and $\delta_{e^{-1}x}$ means the mixture of a continuous spectrum and a discrete spectrum. From this result, we can see that mutations gradually suppress random sampling. Since mutations add more new types into the population, the allele spectrum starts to show the properties of continuous spectra. When $\lim_{\theta\to+\infty} \theta t(\theta) = \infty$, the population is completely controlled by mutations; therefore it only admits purely continuous allele spectrums.

Naturally, we would like to consider the large deviation principle associated with the above weak law of large numbers. I can not establish such a large deviation principle in this thesis. However, since partition structures provide an alternative description of allele spectrums, the large deviation principle of the associated partition structure shall show us some possible properties of the large deviation principle. In this thesis I successfully obtain the large deviation principles associated with the weak law of large numbers for the partition structures:

$$P_n^\theta(\eta) = Ep_\eta(X_{t(\theta)}).$$

Surprisingly, there is a phase transition in terms of the rate function.

- If $\lim_{\theta\to\infty} \frac{\theta t(\theta)}{\log\theta} = k \in (0,2)$, then P_n^θ has the LDP with the speed $\log\theta$ and the rate function
$$I_n(\eta) = \begin{cases} 0, & \text{if } \eta = (1,1,\cdots,1) \\ \frac{n-\alpha_1(\eta)}{2}k, & \text{if } \frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)} > \frac{2}{2-k} \\ n - l(\eta), & \text{if } \frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)} \le \frac{2}{2-k}. \end{cases}$$

- If $\lim_{\theta\to\infty} \frac{\theta t(\theta)}{\log\theta} = 0$, then P_n^θ has the LDP with the speed $\theta t(\theta)$ and the rate function $I_n(\eta) = \frac{n-\alpha_1(\eta)}{2}$.

- If $\lim_{\theta\to\infty} \frac{\theta t(\theta)}{\log\theta} = k \ge 2$, then P_n^θ has the LDP with the speed $\log\theta$ and the rate function $I_n(\eta) = n - l(\eta)$.

Here $l(\eta)$ is the number of components of the partition η, and $\alpha_i(\eta) = \#\{j|\eta_j = i, 1 \le i \le n\}$. As can be seen, when we increase the time scale, the large deviation principle for the partition structures shows the phase transition phenomenon, and the critical time scale is $\frac{2\log\theta}{\theta}$. Furthermore, all the above asymptotic results have a parallel generalization to the two-parameter model.

Asymptotic results for the one-parameter selective model (large mutation)

Mutations always introduce new types to the population. Due to the effect of random sampling, a small portion of those new alleles will be ultimately fixed. So the rate of substitution, denoted by k, is an important quantity for the evolutionary history. Presumably, k is a function of selection intensity, mutation rate and population size N. For neutral models, the rate of substitution is just the individual mutation rate u. In [Gillespie, 1999], Gillespie considered seven models: neutral, TIM, SAS-CFF, house of cards, exponential shift, normal shift and overdominance. Computer simulations of those models were conducted. A plot of $\frac{k}{u}$ and N indicates that these seven models fall into three categories. The Normal shift model is in the first category; because $\frac{k}{u}$ is an increasing function of N. TIM, the neutral model, SAS-CFF and the overdominance model are in the second category; because $\frac{k}{u}$ is almost a constant 1 as N increases. The Gamma shift, the exponential shift and the house of cards are in the third category; because $\frac{k}{u}$ is a decreasing function of N.

Therefore, Gillespie conjectured that the overdominance model converges to the neutral model as $N \to +\infty$. In the symmetric overdominance model, the selection strength is $\sigma = 2N\lambda$, where λ is the selection intensity, and the population mutation is $\theta = 4Nu$. As $N \to +\infty$, both θ and σ will explode. Thus, to verify Gillespie's conjecture, we should consider the asymptotic behaviour of the one-parameter selective model with symmetric overdominance as θ and σ both approach $+\infty$.

Gillespie's conjecture has been confirmed in [Joyce et al., 2003] and [Dawson and Feng, 2006] by studying the asymptotic behaviour of the stationary distribution π_σ. We can actually confirm this conjecture by the asymptotic behaviour of the dynamical system at a fixed time as well.

All the weak limits, which have been completed in this thesis, are collected in the charts.

overdominance	underdominance($\sigma > 0$)	
$\sigma < 0$	$\lim_{\theta\to\infty} \frac{\sigma}{\theta} \le \lambda_0$	$\lim_{\theta\to\infty} \frac{\sigma}{\theta} > \lambda_0$
$\pi_\sigma \to \delta_{(0,0,\cdots)}$	$\pi_\sigma \to \delta_{(0,0,\cdots)}$	$\pi_\sigma \to \delta_{(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{4\lambda}},0,0,\cdots)}$

Here λ_0 is the solution of the following transcendental equation:

$$\log\left(\frac{1-\sqrt{1-\frac{2}{\lambda}}}{2}\right) + \lambda\left(\frac{1+\sqrt{1-\frac{2}{\lambda}}}{2}\right)^2 = 0.$$

As a matter of fact, we can easily obtain the above results by the large deviation principle for π_σ in [Dawson and Feng, 2006] even though they are not stated there.

The next chart is about the weak limits of X_t^σ:

overdominance	underdominance
$\sigma < 0$	$\sigma > 0, \lim_{\theta\to\infty} \frac{\sigma}{\theta} < 2$
$X_t^\sigma \to \delta_{(0,0,\cdots)}, \forall t > 0.$	$X_t^\sigma \to \delta_{(0,0,\cdots)}, \forall t > 0.$

Notice that $\lambda_0 > 2$.

Asymptotic results for the one-parameter selective model (small mutation)

In [Feng, 2009], S. Feng considered the large deviation principle for PD(θ) as $\theta \to 0$. The rate function is

$$J(x) = \begin{cases} 0, & x \in L_1 \\ n-1, & x \in L_n, n \geq 2 \\ \infty, & x \in L, \end{cases}$$

where $L_n = \{(x_1, \cdots, x_n, 0, \cdots) \in \nabla_\infty \mid x_n > 0\}$ and $L = \cup_{n=1}^{\infty} L_n$.

As an application, the large deviation principle for $\pi_\sigma, \sigma = \lambda\alpha(\theta)$, is obtained as well. The rate function is

$$S_\lambda(x) = \begin{cases} J(x), & \lim_{\theta\to 0} \frac{\alpha(\theta)}{\log\theta} = 0 \\ J(x) + |\lambda|(1-\varphi_2(x)), & \lim_{\theta\to 0} \frac{\alpha(\theta)}{\log\theta} = 1, \lambda < 0 \\ J(x) + \lambda\varphi_2(x) - \inf\{\frac{\lambda}{n} + n - 1 : n \geq 1\}, & \lim_{\theta\to 0} \frac{\alpha(\theta)}{\log\theta} = 1, \lambda > 0 \end{cases}$$

As can be seen, apart from the case $\sigma = \lambda \log\theta, \lambda = k(k+1), k \geq 1$, all the rate functions have only one zero. The associated weak limit of π_σ as $\theta \to 0$ is just the point masses at the zero. For $\lambda = k(k+1), k \geq 1$, the rate function has exactly two zeros, $(\frac{1}{k}, \cdots, \frac{1}{k}, 0, \cdots)$ and $(\frac{1}{k+1}, \cdots, \frac{1}{k+1}, 0, \cdots)$. We cannot easily obtain the weak limit of π_σ from the large deviation principle in the critical case.

In this thesis, we have obtained the weak limit of π_σ when $\sigma = \lambda\log\theta, \lambda > 0$. It is

$$\sum_{k=1}^{\infty} \mathrm{I}_{(k(k-1),k(k+1)]}(\lambda)\delta_{(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)}.$$

From this result, we know that, for the critical case $\lambda = k(k+1), k \geq 1$, the weak limit of π_σ should be the point mass at $(\frac{1}{k}, \cdots, \frac{1}{k}, 0, \cdots)$.

1.4 Structure of the Thesis

The whole thesis is organized as follows.

In Chapter 2, we will introduce the concepts of partition structures and all the preliminary knowledge needed in this thesis.

In Chapter 3, we will introduce three types of infinite dimensional diffusion processes and ergodic inequalities for them.

In Chapter 4, we will mainly talk about the asymptotic theories of the one-parameter neutral model and the two-parameter model.

In Chapter 5, all asymptotic results associated with the one-parameter selective model will be discussed.

Chapter 2

Partition Structures

The Ewens sampling formula is one of the most influential results in population genetics. In order to characterize various sampling formulas, J.F.C. Kingman introduced the concept of partition structures in [Kingman, 1978a]. It turns out that every partition structure can be represented by a unique measure in an infinite dimensional simplex. This is usually known as the Kingman correspondence. Moreover, partition structures correspond to sampling formulas of certain population genetics models, and it thus sheds some light on the allele spectra. It is reasonable to expect that studying partition structures could be another way of understanding the evolution of a certain population.

In this chapter, we will first give the definition of integer partitions, and then introduce the concept of the Kingman simplex. The symmetric polynomials in the Kingman simplex will be introduced in detail. Next the definition of partition structure and the Kingman correspondence are presented. Finally, we will also recall the concept of weak convergence of measures and the large deviation principle.

2.1 Partitions of Integers

This section will focus on partitions of integers. For more comprehensive introductions, please refer to [Macdonald, 1995] and the references therein.

Definition 2.1. *For a given positive integer n, a vector $\eta = (\eta_1, \cdots, \eta_l)$ is called a partition of n, if $\eta_1 \geq \eta_2 \cdots \geq \eta_l \geq 1$ and $\sum_{i=1}^{l} \eta_i = n$. We usually use $l(\eta)$ to denote the number of components of the partition η and let $|\eta|$ denote the norm, i.e. $\sum_{i=1}^{l(\eta)} \eta_i$.*

Let $\mathcal{M}_n$ be the set of all partitions of n, then $\mathcal{M} = \cup_{n=1}^{\infty} \mathcal{M}_n$ will be the totality of partitions of all positive integers. Moreover, if we define

$$\alpha_i(\eta) = \#\{j \mid \eta_j = i, 1 \leq j \leq l(\eta)\}, 1 \leq i \leq n,$$

then $\alpha_i(\eta)$ denotes the number of components of size i in partition η. So $(\alpha_1(\eta), \cdots, \alpha_n(\eta))$

becomes another representation of the integer partition η. Naturally $l(\eta) = \sum_{i=1}^{n} \alpha_i(\eta)$ and $n = \sum_{i=1}^{n} i\alpha_i(\eta)$.

Definition 2.2 (A total order "$\leq$" in $\mathcal{M}$)**.** *For $\eta, \xi \in \mathcal{M}$, we say $\eta \leq \xi$ if one of the three conditions is satisfied:*

(i). $|\eta| < |\xi|$

(ii). the first non-vanishing difference $\eta_i - \xi_i$ is positive

(iii). there are no non-vanishing terms.

Remark 2.1. *The above order "$\leq$" can be easily shown to be a total order, and will be used as the order of partitions in this thesis.*

For a given integer n, a subset of $\mathcal{M}_n$, defined as

$$\tilde{\mathcal{M}}_n = \{\eta \in \mathcal{M}_n \mid \eta = (1) \text{ or } \eta_{l(\eta)} \geq 2\},$$

is also used in this thesis. Define $\tilde{\mathcal{M}} = \cup_{n \geq 1} \tilde{\mathcal{M}}_n$. Obviously, $\tilde{\mathcal{M}} \subset \mathcal{M}$. In this thesis, when we say a partition η we mean $\eta \in \mathcal{M}$ unless it is otherwise stated.

2.2 Kingman Simplex

The Kingman simplex is an important topological space, on which all the three diffusion models of this thesis are defined. It is therefore quite necessary to introduce it here in detail. Furthermore, typical symmetric polynomials and probability measures on the Kingman simplex will also be discussed.

Definition 2.3 (Kingman Simplex)**.** *The Kingman simplex, denoted by $\bar{\nabla}_\infty$, is the following simplex of infinite dimension:*

$$\bar{\nabla}_\infty = \left\{ x = (x_1, x_2, \cdots,) \in [0,1]^\infty \middle| x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} x_i \leq 1 \right\}.$$

Due to Tychonoff's theorem, $[0,1]^\infty$, equipped with the product topology, is a compact topological space. It is also metrizable and the product topology is generated by the following metric:

$$d(x,y) = \sum_{i=1}^{\infty} \frac{|x_i - y_i|}{2^i}, \quad x = (x_1, x_2, \cdots), y = (y_1, y_2, \cdots) \in [0,1]^\infty.$$

Restricting the metric $d(\cdot,\cdot)$ to the Kingman simplex $\bar{\nabla}_\infty$, there is no difficulty in showing $(\bar{\nabla}_\infty, d)$ is a closed compact subspace of $([0,1]^\infty, d)$.

Proposition 2.1. $(\bar{\nabla}_\infty, d(\cdot,\cdot))$ *is a closed compact metric space. Moreover,*

$$\nabla_\infty = \left\{ x = (x_1, x_2, \cdots) \in [0,1]^\infty \middle| x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} x_i = 1 \right\},$$

is dense in $\bar{\nabla}_\infty$.

Proof. First, the closedness is easy to prove. Indeed, assuming $\{x(n), n \geq 1\} \subset \bar{\nabla}_\infty$ and $x(n) \to x$, we have $x_i(n) \to x_i, n \geq 1$. By Fatou's lemma,

$$\sum_{i=1}^{\infty} x_i \leq \liminf_{n\to+\infty} \sum_{i=1}^{\infty} x_i(n) \leq 1,$$

so $x \in \bar{\nabla}_\infty$. Hence, $\bar{\nabla}_\infty$ is closed, and thereby compact as well. Next, we are going to show that ∇_∞ is dense in $(\bar{\nabla}_\infty, d(\cdot,\cdot))$. Indeed, $\forall x \in \bar{\nabla}_\infty$, consider

$$y(n) = x + \left(\frac{1 - \sum_{i=1}^{\infty} x_i}{n}, \cdots, \frac{1 - \sum_{i=1}^{\infty} x_i}{n}, 0, \cdots \right),$$

apparently $y(n) \in \nabla_\infty$, and

$$d(y(n), x) = \sum_{i=1}^{n} \frac{1 - \sum_{i=1}^{\infty} x_i}{n 2^i} = \frac{1 - \sum_{i=1}^{\infty} x_i}{n} \left(1 - \frac{1}{2^n}\right) \to 0, \text{ as } n \to \infty.$$

Therefore, ∇_∞ is dense in $\bar{\nabla}_\infty$. □

2.2.1 Continuous Function Space In Kingman Simplex

Continuous functions in the Kingman simplex play a very important role, not only because they are the test functions when we consider the weak convergence of measures in the Kingman simplex but also because they constitute the domain of the infinitesimal generator of the diffusion models in this thesis. Moreover, sampling probabilities are usually represented as symmetric polynomials in x as well. Therefore, we should introduce some facts about the continuous functions in the Kingman simplex.

Let $C(\bar{\nabla}_\infty)$ denote the space of continuous functions on $\bar{\nabla}_\infty$. Since $\bar{\nabla}_\infty$ is compact, then $C(\bar{\nabla}_\infty) = C_b(\bar{\nabla}_\infty)$, where $C_b(\bar{\nabla}_\infty)$ is the set of bounded continuous functions on $\bar{\nabla}_\infty$.

Definition 2.4. *Define* $\varphi_k(x), k \geq 1$, *as follows:*

$$\varphi_k(x) = \sum_{i=1}^{\infty} x_i^k, \quad k \geq 2, \varphi_1(x) = 1.$$

Note that $x \in \nabla_\infty$ often denotes the allele frequencies in population genetics models. Then $\varphi_2(x)$ is the probability that two randomly sampled individuals are of the same type. This probability is named homozygosity in population genetics. Additionally, $1-\varphi_2(x)$ corresponds to the probability that the two individuals are different, and is called heterozygosity in population genetics. It is not difficult to show that $\varphi_k(x), k \geq 2$, are continuous on $\bar{\nabla}_\infty$; $\sum_{i=1}^\infty x_i$, however, is not. Indeed, since $ix_i \leq \sum_{j=1}^i x_j \leq \sum_{i=1}^\infty x_i \leq 1$, we have $x_i \leq \frac{1}{i}, \quad \forall i \geq 1$. By Weierstrass's M-test, $\sum_{i=1}^\infty x_i^k$ is uniformly convergent $\forall k \geq 2$; hence $\varphi_k(x), \quad k \geq 2$ are continuous on $\bar{\nabla}_\infty$. However, $\sum_{i=1}^\infty x_i$ is not continuous on $\bar{\nabla}_\infty$. To this end, let us consider $x(n) = (\frac{1}{n}, \cdots, \frac{1}{n}, 0, \cdots)$; then $x(n) \to 0$. But

$$\sum_{i=1}^\infty \lim_{n\to+\infty} x_i(n) = 0 < 1 = \lim_{n\to+\infty} \sum_{i=1}^\infty x_i(n).$$

Nevertheless, the continuous extension of $\sum_{i=1}^\infty x_i$ is identical to $\varphi_1(x)$ for $\sum_{i=1}^\infty x_i = 1 = \varphi_1(x), \forall x \in \nabla_\infty$, and ∇_∞ is dense in $\bar{\nabla}_\infty$.

Definition 2.5. *For a given partition $\eta = (\eta_1, \cdots, \eta_l)$ of a positive integer n, we define*

$$p_\eta^o(x) = \sum_{i_1, i_2, \cdots, i_l \neq} x_{i_1}^{\eta_1} \cdots x_{i_l}^{\eta_l}. \tag{2.2.1}$$

Remark 2.2. *(i). In fact, we can rewrite (2.2.1) as follows:*

$$p_\eta^o(x) = \alpha_1! \cdots \alpha_n! \sum \prod_{i=1}^\infty x_i^{v_i}, \tag{2.2.2}$$

where the summation is over all sequences $(v_1, v_2, \cdots)$ such that $0 \leq v_i \leq n$, and, for $1 \leq j \leq n$, exactly α_j of v_i equal to j. We can easily prove this by rearranging (2.2.1).

(ii). For a fixed x, $p_\eta^o(x)$ is unchanged if we permute the components of η. Therefore, in this thesis, $p^o_{(\delta_1,\cdots,\delta_l)}$ also means p_η^o where η is defined to be the decreasing rearrangement of $(\delta_1, \cdots, \delta_l)$.

If $\eta_l = 1$, then $p_\eta^o(x)$ is not continuous either, as is shown in [Feng, 2007a]. Similarly to $\sum_{i=1}^\infty x_i$, however, we can explicitly obtain the continuous extension of $p_\eta^o(x)$.

Proposition 2.2. *For a given partition η of the positive integer n and $x \in \nabla_\infty$, $p_\eta^o(x)$ can be rewritten as*

$$\sum_{d=1}^{l(\eta)} (-1)^{l(\eta)-d} \sum_{\beta\in\pi(l(\eta),d)} (|\beta_1|-1)! \cdots (|\beta_d|-1)! \varphi_{\sum_{i\in\beta_1}\eta_i}(x) \cdots \varphi_{\sum_{i\in\beta_d}\eta_i}(x), \tag{2.2.3}$$

where β varies over partitions of the set $\{1, 2, \cdots, l(\eta)\}$ into d subsets, $\beta_1, \cdots, \beta_d$, satisfying

$$\min \beta_1 < \min \beta_2 < \cdots < \min \beta_d,$$

and $|\beta_i|$ denotes the cardinality of β_i.

Proof. We can prove this by mathematical induction on $l(\eta)$. When $l(\eta) = 1$, it is trivial; when $l(\eta) = 2$,

$$p^o_{(\eta_1,\eta_2)}(x) = \sum_{i \neq j} x_i^{\eta_1} x_j^{\eta_2} = \sum_{i=1}^{\infty} x_i^{\eta_1} (\varphi_{\eta_2}(x) - x_i^{\eta_2}) = \varphi_{\eta_1}(x)\varphi_{\eta_2}(x) - \varphi_{\eta_1+\eta_2}(x). \quad (2.2.4)$$

Now assuming that, for $l(\eta) = l$,

$$p^o_\eta = \sum_{d=1}^{l} (-1)^{l-d} \sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \varphi_{\sum_{i\in\beta_1} \eta_i} \cdots \varphi_{\sum_{i\in\beta_d} \eta_i},$$

we have, for $l(\eta) = l + 1$,

$$\begin{aligned} p^o_{(\eta_1,\cdots,\eta_l,\eta_{l+1})} &= p^o_{(\eta_1,\cdots,\eta_l)} \varphi_{\eta_{l+1}} - \sum_{u=1}^{l} p^o_{(\eta_1,\cdots,\eta_u+\eta_{l+1},\cdots,\eta_l)} \\ &= \sum_{d=1}^{l} (-1)^{l-d} \sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \varphi_{\sum_{i\in\beta_1} \eta_i} \cdots \varphi_{\sum_{i\in\beta_d} \eta_i} \varphi_{\eta_{l+1}} \\ &\quad - \sum_{u=1}^{l} \sum_{d=1}^{l} (-1)^{l-d} \sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \varphi_{\sum_{i\in\beta_1} \eta_i^u} \cdots \varphi_{\sum_{i\in\beta_d} \eta_i^u}, \end{aligned} \quad (2.2.5)$$

where

$$\eta_i^u = \begin{cases} \eta_i & \text{if } i \neq u \\ \eta_i + \eta_{l+1} & \text{if } i = u. \end{cases}$$

By switching the order of summation in (2.2.5), we have

$$\begin{aligned} &\sum_{u=1}^{l} \sum_{d=1}^{l} (-1)^{l-d} \sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \varphi_{\sum_{i\in\beta_1} \eta_i^u} \cdots \varphi_{\sum_{i\in\beta_d} \eta_i^u} \\ =& \sum_{d=1}^{l} (-1)^{l-d} \sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \sum_{u=1}^{l} \varphi_{\sum_{i\in\beta_1} \eta_i^u} \cdots \varphi_{\sum_{i\in\beta_d} \eta_i^u}, \end{aligned}$$

where, as a matter of fact,

$$\begin{aligned}\sum_{u=1}^{l} \varphi_{\sum_{i\in\beta_1}\eta_i^u}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i^u} &= \sum_{v=1}^{d}\sum_{u\in\beta_v} \varphi_{\sum_{i\in\beta_1}\eta_i^u}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i^u} \\ &= \sum_{v=1}^{d} |\beta_v| \varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_v}\eta_i+\eta_{l+1}}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i}.\end{aligned}$$

Therefore, (2.2.5) becomes

$$\begin{aligned}&\sum_{d=1}^{l}(-1)^{l-d}\sum_{v=1}^{d}\sum_{\beta\in\pi(l,d)}(|\beta_1|-1)!\cdots(|\beta_{v-1}|-1)!|\beta_v|!(|\beta_{v+1}|-1)!\cdots(|\beta_d|-1)! \\ &\varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_v}\eta_i+\eta_{l+1}}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i} \\ &=\sum_{d=1}^{l}(-1)^{l-d}\sum_{v=1}^{d}\sum_{\beta^v\in\pi(l+1,d)}(|\beta_1^v|-1)!\cdots(|\beta_d^v|-1)!\varphi_{\sum_{i\in\beta_1^v}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d^v}\eta_i}.\end{aligned}$$

Here $\beta^v = \beta_1^v \cup \cdots \cup \beta_d^v$,

$$\beta_i^v = \begin{cases}\beta_i & \text{if } i \neq v \\ \beta_i \cup \{l+1\} & \text{if } i = v,\end{cases}$$

and $\beta = \beta_1 \cup \cdots \cup \beta_d \in \pi(l,d)$. Thus,

$$\begin{aligned}&p^o_{(\eta_1,\cdots,\eta_l,\eta_{l+1})} \\ &=\sum_{d=2}^{l+1}(-1)^{l+1-d}\sum_{\beta\in\pi(l+1,d),\beta_d=\{l+1\}}(|\beta_1|-1)!\cdots(|\beta_d|-1)!\varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i}\varphi_{\eta_{l+1}} \qquad (2.2.6)\\ &+\sum_{d=1}^{l}(-1)^{l+1-d}\sum_{v=1}^{d}\sum_{\beta^v\in\pi(l+1,d)}(|\beta_1^v|-1)!\cdots(|\beta_d^v|-1)!\varphi_{\sum_{i\in\beta_1^v}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d^v}\eta_i}. \qquad (2.2.7)\end{aligned}$$

Let us separate the terms associated with $d = l + 1$ from (2.2.6) and separate the terms related to $d = 1$ from (2.2.7); then we combine other terms in (2.2.6) and (2.2.7). Therefore, we have

$$\begin{aligned}p^o_{(\eta_1,\cdots,\eta_l,\eta_{l+1})} =&\varphi_{\eta_1}\cdots\varphi_{\eta_l}\varphi_{\eta_{l+1}}+\sum_{d=2}^{l}(-1)^{l+1-d}\Big[\sum_{\beta\in\pi(l+1,d),\beta_d=\{l+1\}}(|\beta_1|-1)!\cdots(|\beta_d|-1)! \\ &\varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i}\varphi_{\eta_{l+1}}\end{aligned}$$

$$
\begin{aligned}
&+\sum_{v=1}^{d}\sum_{\beta^v\in\pi(l+1,d)}(|\beta_1^v|-1)!\cdots(|\beta_d^v|-1)!\varphi_{\sum_{i\in\beta_1^v}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d^v}\eta_i}\Big]\\
&+(-1)^{l+1-1}(l+1-1)!\varphi_{\sum_{i=1}^{l+1}\eta_i}\\
=&\sum_{d=1}^{l+1}(-1)^{l+1-d}\sum_{\beta\in\pi(l+1,d)}(|\beta_1|-1)!\cdots(|\beta_d|-1)!\varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d}}.
\end{aligned}
$$

The proof is thus completed. □

Clearly, the right hand side of (2.2.3) is the explicit continuous extension of $p^o_\eta(x)$. Also, because of Proposition 2.2, $\{p^o_\eta(x);\eta\in\mathcal{M}\}$ and $\{\varphi_\eta(x);\eta\in\tilde{\mathcal{M}}\}$ can be easily shown to be two sets of linearly independent vectors, as done in [Ethier, 1992]. Therefore, $\{p^o_\eta(x);\eta\in\mathcal{M}\}$ and $\{\varphi_\eta(x);\eta\in\tilde{\mathcal{M}}\}$ span the same subalgebra of $C(\bar\nabla_\infty)$, which we denote by $\mathcal{P}$. We can also show that $\mathcal{P}$ is dense in $C(\bar\nabla_\infty)$. The following argument is due to [Kingman, 1977].

Proposition 2.3. *$\mathcal{P}=\{f_n(x)\mid f_n(x)=\sum_{|\eta|\le n}a_\eta\varphi_\eta(x), n\ge 1, \alpha_\eta\in\mathbb{R}, \eta\in\tilde{\mathcal{M}}\}$ is a dense subalgebra of $(C(\bar\nabla_\infty),\|\cdot\|_\infty)$; i.e. $\forall f\in C(\bar\nabla_\infty), \exists f_n(x)\in\mathcal{P}, n\ge 1$, such that*

$$\|f-f_n\|_\infty\to 0,\ \textit{as } n\to\infty.$$

Proof. We can use the Stone-Weierstrass theorem to prove this. Thus we need to check the following three conditions:

(i). $\mathcal{P}$ contains a unit element.

(ii). $\forall f\in\mathcal{P}, \bar f\in\mathcal{P}$.

(iii). $\mathcal{P}$ separates points of $\bar\nabla_\infty$, i.e. $\forall x,y\in\bar\nabla_\infty, x\ne y, \exists f\in\mathcal{P}$, such that $f(x)\ne f(y)$.

The first two conditions are satisfied naturally. We only need to check the last condition. To each $x\in\bar\nabla_\infty$, we assign the probability measure

$$\nu_x(du)=\sum_{i=1}^\infty x_i\delta_{x_i}+\left(1-\sum_{i=1}^\infty x_i\right)\delta_0$$

on $[0,1]$. Then we have

$$\varphi_{k+1}(x)=\int_0^1 u^k\nu_x(du), k\ge 1.$$

As can be seen, $\nu_x(du)$ is uniquely determined by its moments $\varphi_k(x), k\ge 1$. Therefore, if $x\ne y$, then $\nu_x\ne\nu_y$. So $\exists\varphi_k(x)$, such that $\varphi_k(x)\ne\varphi_k(y)$. Hence, $\mathcal{P}$ separates points. Thus by the Stone-Weierstrass theorem $\mathcal{P}$ is dense in $(C(\bar\nabla_\infty),\|\cdot\|_\infty)$. □

2.2.2 Measures on $\bar{\nabla}_\infty$ and $[0,1]^\infty$

In this section, we will discuss five probability measures. The first four probability measures are the Poisson-Dirichlet distribution denoted by PD(θ), the two-parameter Poisson-Dirichlet distribution, PD(θ, α), the associated one parameter GEM distribution, GEM(θ), and the two parameter GEM distribution, GEM(θ, α). The last one is π_σ, the stationary distribution of the one-parameter selective model to be presented in Chapter 3. PD(θ), PD(θ, α) and π_σ are measures in $\bar{\nabla}_\infty$; GEM(θ) and GEM(θ, α) are measures in $[0,1]^\infty$.

Definition 2.6 (Poisson-Dirichlet Distribution)**.** *The Poisson-Dirichlet distribution,* PD*(θ)(dx), is defined to be the weak limit of the rearrangement of Dirichlet distribution in a decreasing order,*

$$
\begin{aligned}
&Dir(\frac{\theta}{n-1}, \cdots, \frac{\theta}{n-1})(dx)\\
&= \frac{\Gamma(\frac{n\theta}{n-1})}{\Gamma(\frac{\theta}{n-1})^n} x_1^{\frac{\theta}{n-1}-1} x_2^{\frac{\theta}{n-1}-1} \cdots x_{n-1}^{\frac{\theta}{n-1}-1} \left(1 - \sum_{i=1}^{n-1} x_i\right)^{\frac{\theta}{n-1}-1} dx_1 \cdots dx_{n-1}.
\end{aligned}
$$

The Poisson-Dirichlet distribution was first introduced by J.F.C. Kingman in [Kingman et al., 1975]. It turned out to be the stationary distribution of the infinitely-many-neutral-alleles diffusion constructed by S.N. Ethier and T.G. Kurtz in [Ethier and Kurtz, 1981]. Please refer to [Kingman et al., 1975], [Watterson, 1977a] and [Feng, 2010] for more detailed introductions. Though PD(θ) is very intractable, J.F.C. Kingman obtained a convenient representation for PD(θ) using Poisson point processes. For Poisson point processes, [Kingman, 1993] is a marvellous reference. The following proposition is due to J.F.C. Kingman.

Proposition 2.4. *For $\theta > 0$, let $\xi_1 \geq \xi_2 \geq \cdots$ be the random points of a Poisson process with intensity measure $\nu(dx) = \theta x^{-1} e^{-x} dx, x > 0$. Define $\sigma = \sum_{i=1}^\infty \xi_i$. Then $(\frac{\xi_1}{\sigma}, \frac{\xi_2}{\sigma}, \cdots)$ follows PD(θ); moreover, σ has the distribution of Gamma$(\theta, 1)$ and is independent $(\frac{\xi_1}{\sigma}, \frac{\xi_2}{\sigma}, \cdots)$.*

From this proposition, one can easily see that PD(θ) concentrates on the dense subset ∇_∞. Furhtermore, GEM(θ) is another possible representation of PD(θ). Please refer to [Feng, 2010] and the references therein.

Definition 2.7 (GEM Distribution)**.** *Let $\{U_i^\theta, i \geq 1\}$ be a sequence of i.i.d. Beta$(1, \theta)$ random variables. Define*

$$V_1^\theta = U_1^\theta, V_n^\theta = \Pi_{i=1}^{n-1}(1 - U_i^\theta) U_n^\theta, \quad n \geq 2.$$

Then the distribution of $(V_1^\theta, V_2^\theta, \cdots)$ is called a GEM distribution.

The size-bias sampling of PD(θ) gives rise to GEM(θ); conversely PD(θ) can be recovered from the decreasingly ordered GEM(θ). This fact is shown in detail in [Feng, 2010]. Thus, for any measurable symmetric function $f(x)$ defined in $\bar{\nabla}_\infty$, we have

$$\int_{\bar{\nabla}_\infty} f(x)\mathrm{PD}(\theta)(dx) = \int_{[0,1]^\infty} f(x)\mathrm{GEM}(\theta)(dx). \tag{2.2.8}$$

Due to (2.2.8), the integrals of all symmetric polynomials with respect to PD(θ) can be calculated easily.

Both PD(θ) and GEM(θ) have two-parameter extensions. They are the two-parameter Poisson-Dirichlet distribution, denoted as PD(θ, α) and the two-parameter GEM distribution, GEM(θ, α), respectively. Please refer to [Pitman and Yor, 1997] for more details.

Definition 2.8 (Two-parameter Poisson-Dirichlet Distribution). *Let $\{U_i^{\alpha,\theta}, i \geq 1\}$ be a sequence of independent random variables, where, for each $i \geq 1$, $U_i^{\alpha,\theta}$ follows Beta$(1-\alpha, \theta + i\alpha), \theta + \alpha > 0$. Define*

$$V_1^{\alpha,\theta} = U_1^{\alpha,\theta}, V_n^{\alpha,\theta} = \Pi_{i=1}^{n-1}(1 - U_i^{\alpha,\theta})U_n^{\alpha,\theta}, \quad n \geq 2.$$

Then the distribution of $\{V_1^{\alpha,\theta}, V_2^{\alpha,\theta}, \cdots\}$ is called a two-parameter GEM distribution.

Definition 2.9 (Two-parameter Poisson-Dirichlet Distribution). *The distribution of the decreasingly ordered GEM(θ, α) is called a two-parameter Poisson-Dirichlet distribution, denoted by PD(θ, α).*

Again PD(θ, α) also concentrates on the dense subset ∇_∞ due to the representation of the Poisson point process in [Pitman and Yor, 1997]. Finally, π_σ is a probability measure in $\bar{\nabla}_\infty$, and is absolutely continuous with respect to PD(θ). More precisely,

$$\pi_\sigma(dx) = C_\sigma \exp\{\sigma\varphi_2(x)\}\mathrm{PD}(\theta)(dx),$$

where C_σ is a normalized constant.

2.3 Kingman Partition Structures

J.F.C. Kingman introduces the concept of the partition structures in [Kingman, 1978a]. Such a concept actually grows out of sampling formulas. Suppose that we have an urn containing balls of infinitely many colours, and the colour frequencies are described by a point $x = (x_1, x_2, \cdots)$ in ∇_∞. Then a random sample S of size n has the sampling distribution,

$$P(S = (\alpha_1, \cdots, \alpha_n)) = \frac{n!}{\prod_{j=1}^n j!^{\alpha_j}} \sum \prod_{i=1}^\infty x_i^{v_i},$$

where the summation extends over all sequences $(v_1, v_2, \cdots)$ such that $0 \leq v_i \leq n$, and, for $1 \leq j \leq n$, exactly α_j of v_i equal to j.

Equivalently, due to (2.2.2) and Proposition 2.2, the sampling distribution can also be expressed as

$$p_\eta(x) := \frac{n!}{\eta_1! \cdots \eta_l! \alpha_1! \cdots \alpha_n!} \sum_{d=1}^{l} (-1)^{l-d}$$
$$\sum_{\beta \in \pi(l,d)} (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \varphi_{\sum_{i \in \beta_1} \eta_i}(x) \cdots \varphi_{\sum_{i \in \beta_d} \eta_i}(x),$$

where $\eta = (\eta_1, \cdots, \eta_l)$ is used to represent the configuration of the sample. l is the number of colors and $\eta_i, 1 \leq i \leq l$, is the number of times that one specific color shows. $\alpha_j, 1 \leq j \leq n$, denotes the number of colors which show up exactly j times.

If we draw a random subsample of size m from the sample of size $n (m < n)$ without replacement, and denote their sampling distribution by P_m and P_n respectively, then there is a natural projection σ_{mn} such that $P_m = \sigma_{mn} P_n$ and $\sigma_{pq} = \sigma_{pj} \sigma_{jq}, p < j < q$. More precisely,

$$\sigma_{n-1,n} P_n(\eta) = \sum_{|\xi| = n} P_n(\xi) S(\xi, \eta),$$

where

$$S(\xi, \eta) = \begin{cases} \frac{\alpha_1}{n}, & \text{if } \eta = (\alpha_1 - 1, \alpha_2, \cdots, \alpha_n) \\ \frac{r\alpha_r}{n}, & \text{if } \eta = (\alpha_1, \cdots, \alpha_{r-2}, \alpha_{r-1} + 1, \alpha_r - 1, \alpha_{r+1}, \cdots, \alpha_n) \\ 0, & \text{otherwise .} \end{cases}$$

Definition 2.10 (Partition Structures)**.** *A sequence of partition distributions $\{P_n, n \geq 1\}$ is called a partition structure if*

$$P_m = \sigma_{mn} P_n, \quad \textit{and} \quad \sigma_{nm} = \sigma_{nk} \sigma_{km}, \ \textit{for } m < n. \tag{2.3.1}$$

Remark 2.3.

- *If $x \in \nabla_\infty$, the sampling distribution $P_n(\eta) = p_\eta(x)$ defines a partition structure. So $\sum_{|\eta|=n} p_\eta(x) = 1$.*

- *If $x \in \bar{\nabla}_\infty - \nabla_\infty$, then $p_\eta(x)$ can also define a partition structure and $\sum_{|\eta|=n} p_\eta(x) = 1$. Indeed, because ∇_∞ is dense in $\bar{\nabla}_\infty$, then $\forall x \in \bar{\nabla}_\infty - \nabla_\infty$, $\exists x^k \in \nabla_\infty, k \geq 1$, such that $x^k \to x$. Naturally $P_n^k(\eta) = p_\eta(x^k)$ satisfies (2.3.1) and $\sum_{|\eta|=n} p_\eta(x^k) = 1$; therefore so does $P_n(\eta) = p_\eta(x)$ because of continuity of $p_\eta(x)$.*

- *Finally, if x is a random vector in $\bar{\nabla}_\infty$, then $P_n(\eta) = E p_\eta(x)$ also defines a partition structure and $\sum_{|\eta|=n} P_n(\eta) = 1$.*

As can be seen, a partition structure $\{P_n, n \geq 1\}$ can generate a probability distribution P on $\mathcal{M}$. Note that σ_{mn} is a projection from Π_n to Π_m, where both Π_m and Π_n

are totality of partition distributions on $\mathcal{M}_m$ and $\mathcal{M}_n$, rather than from $\mathcal{M}_n$ to $\mathcal{M}_m$. Of all partition structures, there are two typical ones, the Ewens partition structure and the Ewens-Pitman partition structure. The partition distribution of the Ewens partition structure is the famous Ewens sampling formula, and the partition distribution of the Ewens-Pitman partition structure is described by the Pitman sampling formula.

Example 2.1 (Ewens Partition Structure)**.** *The Ewens partition structure is the partition structure $\{P_n, n \geq 1\}$ where*

$$P_n(\eta) = \frac{n!}{\eta_1 \cdots \eta_l \alpha_1(\eta)! \cdots \alpha_n(\eta)!} \frac{\theta^l}{\theta_{(n)}}, \quad \eta = (\eta_1, \cdots, \eta_l) \in \mathcal{M}_n. \tag{2.3.2}$$

Example 2.2 (Ewens-Pitman Partition Structure)**.** *The Ewens-Pitman partition structure is the partition structure $\{P_n, n \geq 1\}$where*

$$\begin{aligned} P_n(\eta) =& \frac{n!}{\eta_1! \cdots \eta_l! \alpha_1(\eta)! \cdots \alpha_n(\eta)!} \frac{\Pi_{i=0}^{l-1}(\theta + i\alpha)}{\theta_{(n)}} \Pi_{i=1}^{l}(1-\alpha)_{(\eta_i - 1)}, \\ & \eta = (\eta_1, \cdots, \eta_l) \in \mathcal{M}_n. \end{aligned} \tag{2.3.3}$$

As we know, (2.3.2) and (2.3.3) are respectively the Ewens sampling formula (henceforth ESF) and the Pitman sampling formula (henceforth PSF). It turns out that the Ewens sampling formula and the Pitman sampling formula can be represented in terms of PD(θ) and PD(θ, α) respectively as follows:

$$ESF(\eta) = \int_{\bar{\nabla}_\infty} p_\eta(x) d\mathrm{PD}(\theta)(dx)$$

$$PSF(\eta) = \int_{\bar{\nabla}_\infty} p_\eta(x) d\mathrm{PD}(\theta, \alpha)(dx).$$

Likewise, any partition structure can be uniquely represented by a measure in $\bar{\nabla}_\infty$, due to Kingman's correspondence of partition structures.

Theorem 2.1 (Kingman Correspondence)**.** *Any partition structure $\{P_n, n \geq 1\}$ can be uniquely represented as follows by a measure μ in $\bar{\nabla}_\infty$:*

$$P_n(\eta) = \int_{\bar{\nabla}_\infty} p_\eta(x) \mu(dx); \tag{2.3.4}$$

conversely, suppose that we have a sequence of random partitions, $\{\eta(n), n \geq 1\}$, whose distributions constitute the partition structure $\{P_n, n \geq 1\}$. If we define

$$x = \lim_{n \to +\infty} \left(\frac{\eta_1(n)}{n}, \frac{\eta_2(n)}{n}, \cdots, \frac{\eta_{l(\eta)}(n)}{n}, 0, \cdots \right),$$

then the distribution of x *is* μ.

Remark 2.4. *Although we still call this theorem the Kingman correspondence, it is a little different from Kingman's original version, in which*

$$P_n(\eta) = \int_{\nabla_0} \mathcal{S}_\eta(x')\mu(dx'),$$

where

$$\nabla_0 = \left\{ x' = (x_0, x_1, x_2, \cdots) \middle| x = (x_1, x_2, \cdots) \in \bar{\nabla}_\infty, x_0 = 1 - \sum_{i=1}^{\infty} x_i \right\}.$$

For a given $x' \in \nabla_0$, $\mathcal{S}_\eta(x')$ *is defined to be the value distribution of* $(\xi_1, \cdots, \xi_n)$, *and* $\{\xi_r, r \geq 1\}$ *are independent integer-valued random variables with distributions*

$$P(\xi_r = n) = x_n, P(\xi_r = -r) = x_0, r \geq 1.$$

It is not hard to see that $p_\eta(x) = \mathcal{S}_\eta(x')$; *therefore, this representation is equivalent to Kingman's original representation. In this thesis, we would like to regard* x_0 *as the weight of the continuous spectrum. Thus, even degenerate measures, such as* $\delta_{(0,0,\cdots)}$ *and* $\delta_{(\frac{1}{2},0,\cdots)}$ *can determine partition structures through (2.3.4). Presumably, this new form should be a well-known fact, but the author has not found the explicit expression of* $p_\eta(x)$ *in (2.3.4) described in any reference.*

Proof. The proof of this theorem is analogous to Kingman's original proof, thereby omitted here. Please refer to [Kingman, 1978b]. □

$\delta_{(0,0,\cdots)}$ determines a trivial partition structure $\{P_n, n \geq 1\}$ such that

$$P_n(\eta) = \begin{cases} 1 & \text{if } \eta = (1, 1, \cdots, 1) \\ 0 & \text{if } \eta \neq (1, 1, \cdots, 1). \end{cases}$$

This partition structure tells us that we will always end up with a sample whose individuals are different from each other. Hence, all types in the allele spectrum are continuously distributed. This heuristic explanation is due to G. Watterson, communicated to J.F.C. Kingman in [Kingman, 1978b].

Beside the Ewens partition structure and the Ewens-Pitman structure, the partition structure represented by π_σ has been studied thoroughly in the literature, such as [Huillet, 2007], [Grote and Speed, 2002] and [Handa, 2005]. However, none of them provide as simple forms as the Ewens sampling formula. The next partition structure resembles the coupon collecting distribution (please refer to [Pitman, 2006]), and is related to a particular limiting partition structure represented by π_σ.

Example 2.3. *The following is a partition structure $\{P_n, n \geq 1\}$, represented by $\mu = \delta_{(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)}$,*

$$P_n(\eta) = \begin{cases} \frac{n!}{\eta_1!\cdots\eta_{l(\eta)}!\alpha_1(\eta)!\cdots\alpha_n(\eta)!}\frac{k(k-1)\cdots(k-l(\eta)+1)}{k^n} & \text{if } l(\eta) \leq k \\ 0 & \text{if } l(\eta) > k. \end{cases}$$

2.4 Weak Convergence and Large Deviation Principles

The convergence of partition distributions is a pointwise convergence. That is, by $P_n^k \to P_n$, as $k \to +\infty$, we mean

$$P_n^k(\eta) \to P_n(\eta), \text{ as } k \to +\infty, \quad \forall \eta \in \mathcal{M}_n.$$

Due to the Kingman correspondence, every partition structure is uniquely determined by a measure in $\bar{\nabla}_\infty$. The natural question is what topology of probability measures in $\bar{\nabla}_\infty$ makes the Kingman one-to-one correspondence a homeomorphism? The answer is the weak convergence topology of probability measures. At the end of this section, large deviation of principle (henceforth LDP) will be presented. An obvious fact on the relationship between the weak law of large numbers and LDPs is proved as well.

Definition 2.11 (Weak Convergence of Measures)**.** *A sequence of probability measures $\{\mu_n, n \geq 1\}$ in topological space E converge weakly to measure μ if*

$$\lim_{n\to+\infty} \int_{\bar{\nabla}_\infty} f(x)\mu_n(dx) = \int_{\bar{\nabla}_\infty} f(x)\mu(dx), \quad \forall f \in C_b(E).$$

Because $\mathcal{P}$ is a dense subalgebra of $C(\bar{\nabla}_\infty)$,

$$\lim_{n\to+\infty} \int_{\bar{\nabla}_\infty} f(x)\mu_n(dx) = \int_{\bar{\nabla}_\infty} f(x)\mu(dx), \quad \forall f \in \mathcal{P}, \tag{2.4.1}$$

is sufficient to guarantee the weak convergence of $\{\mu_n, n \geq 1\}$ on $\bar{\nabla}_\infty$. As a matter of fact, (2.4.1) can be deduced from the pointwise convergence of the partition structures. Therefore, it leads to the weak convergence of the representation probability measures in the Kingman correspondence; conversely the weak convergence of the representation probability measures guarantees the point-wise convergence of the partition structures as well. Hence the Kingman correspondence is a homeomorphism.

Moreover, the convergence of probability measures can also be characterized by the moment generating functions, that is, the Laplace transforms of the probability measures. Let us define the moment generating function of μ as $\phi(t) = \int_E e^{tx}\mu(dx)$. The following theorem is from [Hogg and Craig, 1978].

Theorem 2.2. *Let $\{X_n, n \geq 1\}$ be a sequence of random variables with moment generating function $\phi_n(t) = Ee^{tX_n}$ that exists for $|t| < h, \forall n \geq 1$. Let X be a random variable with moment generating function $\phi(t)$, which exists for $|t| \leq h_1 \leq h$. If $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ for $|t| \leq h$, then $X_n \to X$ in law; i.e. $P^{X_n} \to P^X$ weakly.*

If $X_n \to c$ in probability, as $n \to +\infty$, where c is a constant, then

$$\forall \epsilon > 0, P(|X_n - c| > \epsilon) \to 0, \text{ as } n \to \infty.$$

Such a conclusion is very weak, for, small as it may be, we can hardly know how small $P(|X_n - c| > \epsilon)$ is. The reason that we make such a big fuss about the magnitude of this small likelihood is because the consequences of some rare events, such as earthquakes, are catastrophic. Usually, applying some sophisticated estimation techniques, we may be able to find out that

$$P(|X_n - c| > \epsilon) \leq e^{-nI(\epsilon)}, \text{ where } I(\epsilon) \geq 0.$$

To describe its decaying speed, we should compare $\log P(|X_n - c| > \epsilon)$ with the speed n. Then $\frac{\log P(|X_n - c| > \epsilon)}{n}$ might have an exact limit as $n \to \infty$, or only have an upper limit and a lower limit, but both of which can be expressed in terms of a common function $I(x)$. This is the general idea of LDPs.

Definition 2.12. *We say $\{\mu_\theta, \theta > 0\}$ satisfies an LDP with speed $a(\theta)$, and lower semicontinuous rate function $I(x)$, if both of the following two conditions are satisfied*

- *$\forall$ closed set F, we have*

$$\limsup_{\theta\to 0} \frac{1}{a(\theta)} \log(\mu_\theta(F)) \leq -\inf_{x\in F} I(x); \tag{2.4.2}$$

- *$\forall$ open set O, we have*

$$\liminf_{\theta\to 0} \frac{1}{a(\theta)} \log(\mathrm{PD}(O)) \geq -\inf_{x\in O} I(x). \tag{2.4.3}$$

If $\forall a \in \mathbb{R}, \{x | I(x) \leq a\}$ is compact, then I is called a good rate function.

Remark 2.5. *Since $\mu_\theta(F)$ approaches either 0 or 1 as θ goes to 0, $\log(\mu_\theta(F))$ tends to either $-\infty$ or 0. Thus the speed $\alpha(\theta)$ must also go to ∞, and their ratio can be expressed in terms of a common function $I(x)$. If $\inf_{x\in F} I(x) > 0$, then $\exists \theta_0 > 0$, such that, $\forall \theta < \theta_0$,*

$$\mu_\theta(F) < \exp\left\{-\theta \frac{\inf_{x\in F} I(x)}{2}\right\}. \tag{2.4.4}$$

Similarly, if $\inf_{x\in O} I(x) > 0$*, then* $\exists\theta_1 > 0$*, such that,* $\forall\theta < \theta_1$*,*

$$\mu_\theta(F) > \exp\left\{-\theta\frac{3\inf_{x\in O} I(x)}{2}\right\}. \tag{2.4.5}$$

Example 2.4. *Let* $X_\alpha^k = \sum_{l=1}^k Y_l^\alpha$*, and* $U_\alpha^k = \sum_{l=1}^k V_l^\alpha$*, where* $\{Y_l^\alpha, 1 \le l \le k\}$ *and* $\{V_l^\alpha, 1 \le l \le k\}$ *are i.i.d. geometric random variables and Bernoulli random variables respectively; i.e.*

$$P(Y_l^\alpha = u) = (1-\alpha)^u\alpha, u \ge 0; \quad P(V_l^\alpha = v) = \alpha^v(1-\alpha)^{1-v}, v = 0, 1.$$

Then the distributions of X_α^k *and* U_α^k*, denoted by* μ_k *and* ν_k*, satisfy LDPs with speed* k *and rate function* $I_1(x)$ *and* $I_2(x)$ *respectively, where*

$$I_1(x) = x\log x - (x+1)\log(1+x) - [x\log(1-\alpha) + \log\alpha]$$

and

$$I_2(x) = x\log\left(\frac{x}{\alpha}\right) + (1-x)\log\left(\frac{1-x}{1-\alpha}\right).$$

Proof. This can be shown by the Cramér theorem. Please refer to [Dembo and Zeitouni, 2010]. □

Remark 2.6. *One can easily show that* $I_1(x)$ *and* $I_2(x)$ *have only one zero, and they are* $\frac{1-\alpha}{\alpha}$ *and* α *respectively.*

The rate function $I(x)$ is always nonnegative, and the zeros of $I(x)$ play an important role.

Proposition 2.5. *If* $\{\mu_\theta, \theta > 0\}$ *satisfies an LDP with the speed* $a(\theta)$ *and a good rate function* $I(x)$*, and the rate function* $I(x)$ *only has a single zero* x_0*, then* $\mu_\theta \to \delta_{x_0}$ *weakly as* $\theta \to 0$*.*

Proof. Fix a $f \in C_b(E)$. Then $\forall\epsilon > 0, \exists\delta > 0$, such that, $\forall x \in B(x_0, \delta)$,

$$|f(x) - f(x_0)| < \epsilon.$$

Then we have

$$\int_E f(x)\mu_\theta(dx) = \int_{B(x_0,\delta)} f(x)\mu_\theta(dx) + \int_{E-B(x_0,\delta)} f(x)\mu_\theta(dx),$$

and hence

$$\left|\int_E (f(x) - f(x_0))\mu_\theta(dx)\right| \le \epsilon + 2\sup_{x\in E}|f(x)|\mu_\theta(E - B(x_0,\delta)).$$

Therefore,

$$0 \leq \limsup_{\theta \to 0} \left| \int_E (f(x) - f(x_0))\mu_\theta(dx) \right| \leq \epsilon, \forall \epsilon > 0. \tag{2.4.6}$$

In (2.4.6), we have used the fact that $\lim_{\theta \to 0} \mu_\theta(E - B(x_0, \delta)) = 0$, which is due to

$$\inf_{x \in E - B(x_0,\delta)} I(x) > 0.$$

Indeed, otherwise, let $\{x_n, n \geq 1\}$ be a sequence on $E - B(x_0, \delta)$ such that $I(x_n)$ converges to

$$\inf_{x \in E - B(x_0,\delta)} I(x).$$

Since the level set of $I(x)$ is compact, there exists a convergent subsequence $\{x_{n_k}, k \geq 1\}$ such that $x_{n_k} \to x^{'}$ and $x^{'} \neq x_0$. Due to the lower semicontinuity of $I(x)$, we have $I(x^{'}) \leq \inf_{x \in E - B(x_0,\delta)} I(x)$. Thus, if $\inf_{x \in E - B(x_0,\delta)} I(x) = 0$, then, $\exists x^{'} \neq x_0$ such that $I(x^{'}) = 0$, which contradicts the uniqueness of zeros of $I(x)$. Finally, letting $\epsilon \to 0$ in (2.4.6), we have

$$\lim_{\theta \to 0} \int_E f(x)\mu_\theta(dx) = f(x_0).$$

The proof is completed. □

Remark 2.7. *If $I(x)$ has more than one zero, then by a similar argument, one can still show that the limiting measure concentrates on those zeros; but the allocation of the corresponding probability mass is uncertain.*

Chapter 3

Three Infinite Dimensional Diffusion Processes

In this chapter, we will first recall some preliminary knowledge of diffusion processes. Then we will successively introduce the one-parameter neutral model, the two-parameter model and the one-parameter selective model. For the one-parameter neutral model and the two-parameter model, we will provide the explicit representation of their transition density functions, which will be different from their previous spectral representations. Then, by making use of these representations, uniform ergodic inequalities for these two models are obtained as well. The uniform ergodic inequality for the two-parameter model is new. For the one-parameter selective model, we will also talk about its transition density function, though its explicit representation is still unavailable. By its transition kernel estimation, we can also prove a uniform ergodic inequality, which is stronger than the ergodic theorem obtained in [Ethier and Kurtz, 1998].

3.1 Preliminaries of Diffusion Theory

We will give a short background introduction to diffusion theory in this section. Please refer to [Revuz and Yor, 1999], [Ethier and Kurtz, 1986], [Karlin and Taylor, 1981] and [Stroock and Varadhan, 2006] for more comprehensive introductions.

Diffusion processes are usually used to model random dynamical systems, which may be subjected to deterministic forces and purely random forces. If there are no random forces, then deterministic laws will govern the evolution of this dynamical system. Therefore its future motions can be precisely predicted to some extent. If, however, random forces are involved, then the future motions are very unpredictable. We can only use a sequence of random variables $\{X_t, t \geq 0\}$ to describe the motions. In this case, we are not able to find any pattern if we only focus on the motion of a single individual. But we might see some patterns if we look at the whole macroscopic structure of the dynamical system. These macroscopic motions are usually characterized by a second-order differential operator. The partial differential equations associated with these operators are the main concerns of mathematicians working in the theory of partial differential equations. For probabilists, the primary goal is to figure out the statistical properties of random dynamical systems by studying the processes generated by these operators. Since at microscopic scale the motion is random, we can only use a family of transition probabilities $P(s, x, t, dy)$ to describe the motion. If the system is homogeneous, then a family of homogeneous transition probabilities $P(t, x, dy)$ should be applied. All the three diffusion models in this thesis are homogenous systems. Moreover, for all the trajectories are continuous, $P(t, x, dy)$ should satisfy certain extra conditions. Due to the Chapman-Kolmogorov equation, these transition probabilities usually generate a semigroup on a function space, which is often a Banach space.

Definition 3.1. *Let (E, d) be the a compact state space. Let $C(E)$ be the continuous function space on E. It is a Banach space under norm $\|f\| = \sup_{x\in E} |f(x)|$. The semigroup $\{P_t, t \geq 0\}$, associated with a family of transition probabilities $\{P(t, x, \cdot), t \geq 0, x \in E\}$, is defined as*

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

Usually, such a semigroup is a strongly continuous contraction semigroup, known as the Feller semigroup, which can determine a diffusion process. The generator of this diffusion process is defined as

$$Gf(x) = \lim_{t\to 0} \frac{P_t f(x) - f(x)}{t}, \text{ which exists under the norm } \|\cdot\|.$$

The domain $\mathcal{D}(G)$ is the set of functions such that the above limit exists. Due to Hille-Yosida theorem, a generator satisfying certain conditions can uniquely generate a Feller semigroup, and thereby determining a diffusion process.

3.2 One-Parameter Neutral Model

The One-parameter neutral model has two evolutionary forces involved: random sampling and mutations, where the random sampling removes genetic variation, thereby pushing the system toward $(1, 0, \cdots)$. Mutations, however, constantly create new species and hence drag the system toward $(0, 0, \cdots)$. Therefore the one-parameter neutral model reaches a dynamical equilibrium state, the Poisson-Dirichlet distribution PD(θ).

In this section, we will focus on the one-parameter neutral model, which is characterized by the following generator

$$G = \frac{1}{2}\sum_{i,j=1}^{\infty} x_i(\delta_{ij} - x_j)\frac{\partial^2}{\partial x_i \partial x_j} - \frac{\theta}{2}\sum_{i=1}^{\infty} x_i \frac{\partial}{\partial x_i}, \quad x \in \bar{\nabla}_\infty.$$

The core of G is $\mathcal{P} = \mathrm{span}\{\varphi_\eta(x) \mid \eta \in \tilde{\mathcal{M}}\}$, which is a dense sub-algebra of $C(\bar{\nabla}_\infty)$. The evaluation of $G\varphi_\eta(x)$ should be first on ∇_∞ and then continuously extended to $\bar{\nabla}_\infty$.

We can actually extend this generator through bounded pointwise convergence. Note that, however, we can not treat $\{x_n, n \geq 1\}$ as regular smooth functions in $\mathcal{D}(G)$, even though $\varphi_k^{\frac{1}{k}} \to x_1$ as $k \to +\infty$. Please refer to [Petrov, 2009] for more detailed comments.

3.2.1 Transition Density of the One-parameter Neutral Model

Explicit representations of the transition density functions of the one-parameter neutral model are available. R.C. Griffiths first obtained one representation in [Griffiths, 1979]; later on S.N. Ethier obtained another by the eigenfunction expansion in [Ethier, 1992]. Compared with the method of R.C. Griffiths, the method of S.N. Ethier is more direct and transparent. Our presentation in this part is mainly due to S.N. Ethier.

To find the eigenfunction expansion, we need to consider the Hilbert space $L^2(\mathrm{PD}(\theta))$. Let us define the associated inner product as follows:

$$\langle f, g\rangle_\theta = \int_{\bar{\nabla}_\infty} f(x)g(x)\mathrm{PD}(\theta)(dx), \quad f, g \in L^2(\mathrm{PD}(\theta)).$$

The norm induced by this inner product, which is denoted by $\|\cdot\|_{2,\theta}$, is defined to be $\|f\|_{2,\theta} = \sqrt{\langle f, f\rangle_\theta}$. In [Ethier, 1992] $\{\varphi_\eta \mid \eta \in \tilde{\mathcal{M}}\}$ has been shown to be linearly independent. Hence, to obtain an orthonormal basis, we can apply the Gram-Schmidt orthogonalization procedure to $\{\varphi_\eta \mid \eta \in \tilde{\mathcal{M}}\}$, the order of which is the order of their subindices. Thus,

$$\begin{aligned}
\psi_1^\theta(x) =& 1, \psi_2^\theta(x) = \varphi_2(x) - \frac{1}{1+\theta}, \\
\psi_3^\theta(x) =& \varphi_3(x) - \langle \varphi_3, \psi_2^\theta\rangle_\theta \psi_2^\theta(x) - \langle \varphi_3, 1\rangle_\theta \\
\psi_4^\theta(x) =& \varphi_4(x) - \langle \varphi_4, \psi_3^\theta\rangle_\theta \psi_3^\theta(x) - \langle \varphi_4, \psi_2^\theta(x)\rangle_\theta \psi_2^\theta(x) - \langle \varphi_4, 1\rangle_\theta \\
\psi_{2,2}^\theta(x) =& \varphi_{2,2}(x) - \langle \varphi_{2,2}, \psi_4^\theta\rangle_\theta \psi_4^\theta(x) - \langle \varphi_{2,2}, \psi_3^\theta\rangle_\theta \psi_3^\theta(x) - \langle \varphi_{2,2}, \psi_2^\theta\rangle_\theta \psi_2^\theta(x) - \langle \varphi_{2,2}, 1\rangle_\theta \\
& \cdots
\end{aligned}$$

and

$$\psi_\eta^\theta(x) = \varphi_\eta - \sum_{\xi<\eta} \langle \varphi_\eta, \psi_\xi^\theta\rangle_\theta \psi_\xi^\theta(x), \quad \eta \in \tilde{\mathcal{M}}. \tag{3.2.1}$$

Then we define $\chi_\eta^\theta(x) = \frac{\psi_\eta^\theta(x)}{\|\psi_\eta^\theta\|_{2,\theta}}, \eta \in \tilde{\mathcal{M}}$. Thus we get an orthonormal subset of $L^2(\mathrm{PD}(\theta))$. If we define $\mathcal{P}_n = \mathrm{span}\{\varphi_\eta \mid \eta \in \tilde{\mathcal{M}}_n\}$, then, obviously, $\mathcal{P}_n = \mathrm{span}\{\psi_\eta^\theta \mid \eta \in \tilde{\mathcal{M}}_n\}$ as well, and $\mathcal{P} = \cup_{n\geq 1}\mathcal{P}_n$. We can also prove that $\{\chi_\eta^\theta(x) \mid \eta \in \tilde{\mathcal{M}}\}$ is an orthonormal basis of the Hilbert space $L^2(\mathrm{PD}(\theta))$.

Proposition 3.1. $\{\chi_\eta^\theta(x) \mid \eta \in \tilde{\mathcal{M}}\}$ *is a complete orthonormal basis of* $L^2(\mathrm{PD}(\theta))$.

Proof. Since $\bar{\nabla}_\infty$ is a Polish space, we have that, $\forall f \in L^2(PD(\theta)), \exists f_n \in C(\bar{\nabla}_\infty)$ s.t.

$$\|f - f_n\|_{2,\theta} \to 0, \text{ as } n \to +\infty.$$

By Proposition 2.3, we know $\mathcal{P}$ is dense in $C(\bar{\nabla}_\infty)$; thus $\mathcal{P}$ is dense in $L^2(\mathrm{PD}(\theta))$ as well. Furthermore, $\mathcal{P}$ can be spanned by $\{\chi_\eta^\theta(x) | \eta \in \tilde{\mathcal{M}}\}$ too. Therefore, $\exists g_n(x) = \sum_{|\eta| \le n} a_\eta \chi_\eta^\theta(x), n \ge 1$, s.t.

$$\|f - g_n\|_{2,\theta} \to 0, \text{ as } n \to +\infty.$$

Thus, to show $\{\chi_\eta^\theta(x) | \eta \in \tilde{\mathcal{M}}\}$ is a complete orthonormal basis, we only need to show the following fact that

$$\text{if } \langle f, \chi_\eta^\theta \rangle_\theta = 0, \forall \eta \in \tilde{M}, \text{ then } f = 0.$$

Indeed, assuming $\langle f, \chi_\eta^\theta \rangle_\theta = 0, \forall \eta \in \tilde{M}$, we have $\langle f, g_n \rangle_\theta = 0, \forall n \ge 1$. Thus,

$$\|f - g_n\|_{2,\theta}^2 = \|f\|_{2,\theta}^2 + \|g_n\|_{2,\theta}^2 \ge \|f\|_{2,\theta}^2.$$

Letting $n \to +\infty$, we have $\|f\|_{2,\theta}^2 = 0$, thereby $f = 0$. This proposition is thus proved. $\square$

The one-parameter neutral model has already proved to be a reversible diffusion, the generator of which, therefore, is a self-adjoint operator in $L^2(\mathrm{PD}(\theta))$. By the theory of self-adjoint operators, the eigenfunctions of G associated with different eigenvalues are orthogonal to each other; moreover, due to [Ethier, 1992], all eigenvalues and some of eigenfunctions can be explicitly computed. It turns out that $\{-\lambda_m = -\frac{m(m+\theta-1)}{2}, \quad m \ge 2, \lambda_0 = 0\}$ is the complete set of the eigenvalues of G.

Proposition 3.2. *For each eigenvalue* $-\lambda_m, m \ge 2$, *we have one particular eigenfunction* $\phi_m^\theta(x)$, *which can be explicitly expressed as*

$$\varphi_m + \sum_{i=1}^{m-1} \left(\prod_{j=i}^{m-1} \frac{\binom{j+1}{2}}{\lambda_j - \lambda_m} \right) \varphi_i(x).$$

Proof. Notice that $G\varphi_m = -\lambda_m \varphi_m + \binom{m}{2} \varphi_{m-1}$, where $\lambda_m = \frac{m(m+\theta-1)}{2}$, and $-\lambda_m$ is an eigenvalue. Running S.N. Ethier's argument in [Ethier, 1992], one can express the eigenfunction associated with $-\lambda_m$ as $\varphi_m + g_m^\theta$, where $g_m^\theta \in \mathcal{P}_{m-1}$ and satisfies the equation

$$(G + \lambda_m) g_m^\theta = -\binom{m}{2} \varphi_{m-1}. \tag{3.2.2}$$

Suppose that $g_m^\theta = \sum_{i=1}^{m-1} a_i \varphi_i, \varphi_1 = 1$; then we can solve for $\{a_i, 1 \le i \le m-1\}$. Substituting g_m^θ into (3.2.2), we have

$$-\sum_{i=2}^{m-1} a_i \lambda_i \varphi_i + \sum_{i=2}^{m-1} a_i \binom{i}{2} \varphi_{i-1} + \sum_{i=2}^{m-1} \lambda_m a_i \varphi_i = -\binom{m}{2} \varphi_{m-1}.$$

Then

$$\begin{aligned}(\lambda_m - \lambda_{m-1}) a_{m-1} \varphi_{m-1} &+ \sum_{i=2}^{m-2} [a_{i+1} \binom{i+1}{2} + (\lambda_m - \lambda_i) a_i] \varphi_i + \lambda_m a_2 + a_1 \\ = -\binom{m}{2} \varphi_{m-1},\end{aligned}$$

and $a_{m-1} = \frac{\binom{m}{2}}{\lambda_{m-1} - \lambda_m}, \frac{a_i}{a_{i+1}} = \frac{\binom{i+1}{2}}{\lambda_i - \lambda_m}$. Therefore,

$$a_i = \frac{a_i}{a_{i+1}} \cdots \frac{a_{m-2}}{a_{m-1}} \cdot a_{m-1} = \prod_{j=i}^{m-1} \frac{\binom{j+1}{2}}{\lambda_j - \lambda_m}$$

□

$\forall \eta \in \tilde{\mathcal{M}}$, Lemma 2.2 in [Ethier, 1992] enables us to build one-to-one correspondence between φ_η and an unique eigenfunction denoted by ϕ_η^θ. Therefore, theoretically, we can get all the eigenfunctions. Now we can denote this set of eigenfunctions by $\{\phi_\eta^\theta, \eta \in \tilde{\mathcal{M}}\}$. If $|\eta| \neq |\xi|$, then ϕ_η^θ and ϕ_ξ^θ are orthogonal for they are the eigenfunctions of two different eigenvalues. If $|\eta| = |\xi|$, then ϕ_η^θ and ϕ_ξ^θ might not be orthogonal to each other. Therefore, applying the Gram-Schmidt orthogonalization procedure to the set of eigenfunctions of a specific eigenvalue, we will end up with a whole set of orthogonal eigenfunctions. Let us denote this set of orthogonal eigenfunctions by $\{\bar{\psi}_\eta^\theta \mid \eta \in \tilde{\mathcal{M}}\}$.

The next proposition specifically indicates that we can get eigenfunctions through Gram-Schmidt orthogonalization procedure.

Proposition 3.3. *For $\eta \in \tilde{\mathcal{M}}$, $\psi_\eta^\theta = \bar{\psi}_\eta^\theta$, If we define $\bar{\chi}_\eta^\theta(x)$ to be $\frac{\bar{\psi}_\eta^\theta(x)}{\|\bar{\psi}_\eta^\theta\|_{2,\theta}}$, then $\chi_\eta^\theta = \bar{\chi}_\eta^\theta$.*

Proof. We can use mathematical induction to prove this. Since $\psi_1^\theta = 1 = \bar{\psi}_1^\theta$ and $\psi_2^\theta = \varphi_2 - \frac{1}{1+\theta} = \bar{\psi}_2^\theta$, we assume that $\forall \xi < \eta$, $\bar{\psi}_\xi^\theta = \psi_\xi^\theta$. Then for η, we have

$$\bar{\psi}_\eta^\theta = \phi_\eta^\theta - \sum_{\xi < \eta} \langle \phi_\eta^\theta, \bar{\psi}_\xi^\theta \rangle_\theta \bar{\psi}_\xi^\theta,$$

$$\psi_\eta^\theta = \varphi_\eta - \sum_{\xi < \eta} \langle \varphi_\eta, \psi_\xi^\theta \rangle_\theta \psi_\xi^\theta.$$

Notice that $\phi_\eta^\theta \perp \phi_\xi^\theta, \forall |\xi| < |\eta|$; hence $\phi_\eta^\theta \perp \bar{\psi}_\xi^\theta, \forall |\xi| < |\eta|$. So

$$\bar{\psi}_\eta^\theta = \phi_\eta^\theta - \sum_{|\xi|=|\eta|, \xi<\eta} \langle \phi_\eta^\theta, \bar{\psi}_\xi^\theta \rangle_\theta \bar{\psi}_\xi^\theta, \tag{3.2.3}$$

Due to Lemma 2.2 in [Ethier, 1992], we have $\phi_\eta^\theta = \varphi_\eta + \sum_{|\xi|<|\eta|} a_\xi \varphi_\xi$. Since $\{\psi_\xi^\theta \mid \xi \in \tilde{\mathcal{M}}_{|\eta|-1}\}$ also span $\mathcal{P}_{|\eta|-1}$, $\phi_\eta^\theta = \varphi_\eta + \sum_{|\xi|<|\eta|} b_\xi \psi_\xi^\theta$. By the mathematical assumption, we have $\phi_\eta^\theta \perp \psi_\xi^\theta, \forall |\xi| < |\eta|$, hence

$$0 = \langle \phi_\eta^\theta, \psi_\xi^\theta \rangle_\theta = \langle \varphi_\eta, \psi_\xi^\theta \rangle_\theta + b_\xi.$$

Thus, $b_\xi = -\langle \varphi_\eta, \psi_\xi^\theta \rangle_\theta$ and

$$\phi_\eta^\theta = \varphi_\eta - \sum_{|\xi|<|\eta|} \langle \varphi_\eta, \psi_\xi^\theta \rangle_\theta \psi_\xi^\theta. \tag{3.2.4}$$

Substituting (3.2.4) into (3.2.3), we have

$$\bar{\psi}_\eta^\theta = \varphi_\eta - \sum_{\xi<\eta} \langle \varphi_\eta, \psi_\xi^\theta \rangle_\theta \psi_\xi^\theta = \psi_\eta^\theta.$$

Therefore, the theorem is thus proved. $\square$

Remark 3.1. *The above argument tells us that we can get eigenfunctions by applying the Gram-Schmidt orthogonalization procedure to* $\{\varphi_\eta \mid \eta \in \tilde{\mathcal{M}}\}$. *And the eigenfunctions can be expressed by* (3.2.4).

Making use of the eigenfunction expansion, S.N. Ethier obtained an explicit representation of the transition density function of the one-parameter neutral model. The following theorem is due to S.N. Ethier.

Theorem 3.1. *The transition probability* $P(t,x,y)$ *of the one-parameter neutral model has the density function denoted by* $p(t,x,y)$*; i.e.*

$$P_t f(x) = \int_{\bar{\nabla}_\infty} f(y) P(t,x,dy) = \int_{\bar{\nabla}_\infty} f(y) p(t,x,y) \mathrm{PD}(\theta)(dy),$$

where $p(t,x,y)$ *has the following explicit expression*

$$p(t,x,y) = 1 + \sum_{m=2}^{\infty} \exp\{-\lambda_m t\} Q_m(x,y),$$

$$Q_m(x,y) = \frac{2m+\theta-1}{m!} \sum_{n=0}^{m} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n(x,y),$$

and $p_n(x,y) = \sum_{|\eta|=n} \frac{p_\eta(x)p_\eta(y)}{\int_{\bar\nabla_\infty} p_\eta d\mathrm{PD}(\theta)}$.

Remark 3.2. *In [Ethier, 1992], $Q_m(x,y)$ was defined to be*

$$\sum_{|\eta|=m} \chi_\eta^\theta(x)\chi_\eta^\theta(y),$$

where $\{\chi_\eta^\theta, \eta \in \tilde{\mathcal{M}}\}$ is one specific orthonormal basis of the Hilbert space $L^2(\mathrm{PD}(\theta))$. For a given m, $\{\chi_\eta^\theta, |\eta| = m, \eta \in \tilde{\mathcal{M}}_m\}$ is an orthonormal basis of the eigenspace associated with the eigenvalue $-\lambda_m$. One can easily show that the definition of $Q_m(x,y)$ is independent of orthonormal basis. Furthermore, $p_n(x,y)$ is a probability kernel. All $p_\eta(x), \eta \in \mathcal{M}$ are the continuous extension of $p_\eta(x)$ restricted to ∇_∞. $\forall n \geq 2, x \in \bar\nabla_\infty$, there always exists $\eta(|\eta| = n)$ such that $p_\eta(x) > 0$. Because if $x = 0$, $p_{(1,1,\cdots,1)}(0) = 1 > 0$; otherwise $p_{(n)}(x) = \varphi_n(x) > 0$. Therefore, $p_n(x,y) > 0, \forall n \geq 2, x, y \in \bar\nabla_\infty$.

Proof. The proof is omitted, please refer to [Ethier, 1992] for the detailed proof. □

The next theorem provides another representation of the transition density function of the one-parameter neutral model. This result can also be obtained from the transition probability of the neutral Fleming-Viot process through the direct computation used in [Feng, 2010]. The explicit representation of the transition probability of the neutral Fleming-Viot process was obtained by R.C. Griffiths and S.N. Ethier in [Ethier and Griffiths, 1993]. Here our proof is more simple and transparent compared with the method in [Feng, 2010].

Theorem 3.2. *The transition density function $p(t,x,y)$ of the one-parameter neutral model is continuous, strictly positive and symmetric on $(0,+\infty)\times\bar\nabla_\infty\times\bar\nabla_\infty$. Moreover, $p(t,x,y)$ has the following explicit expression*

$$p(t,x,y) = d_0^\theta(t) + d_1^\theta(t) + \sum_{n=2}^{+\infty} d_n^\theta(t)p_n(x,y),$$

where

$$\begin{aligned} d_0^\theta(t) =& 1 - \sum_{m=1}^{\infty} e^{-\lambda_m t}\frac{2m+\theta-1}{m!}(-1)^{m-1}\theta_{(m-1)} \\ d_n^\theta(t) =& \sum_{m=n}^{\infty} e^{-\lambda_m t}\frac{2m+\theta-1}{m!}\binom{m}{n}(-1)^{m-n}(\theta+n)_{(m-1)}, n \geq 1. \end{aligned}$$

In particular, the coefficients $d_n^\theta(t), n \geq 1$, is the distribution of the ancestral process constructed by S. Tavaré in [Tavaré, 1984].

Proof. The continuity and symmetry of $p(t,x,y)$ is actually proved in [Ethier, 1992]; moreover, the eigenfunction expansion is absolutely convergent as well. Now we are going to show

$$p(t,x,y) = d_0^\theta(t) + d_1^\theta(t) + \sum_{n=2}^{+\infty} d_n^\theta(t) p_n(x,y). \tag{3.2.5}$$

By theorem 3.1, we have

$$p(t,x,y) = 1 + \sum_{m=2}^{\infty} \exp\{-\lambda_m t\} Q_m(x,y). \tag{3.2.6}$$

Due to the absolute convergence of (3.2.6) and Fubini's theorem, we can prove (3.2.5) by switching the order of summation in (3.2.6). That is,

$$\begin{aligned} p(t,x,y) = 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} \Big(&\sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n(x,y) \\ &+ \frac{2m+\theta-1}{m!} (-1)^{m-1} (\theta+1)_{(m-1)} m p_1(x,y) \\ &+ \frac{2m+\theta-1}{m!} (-1)^{m} \theta_{(m-1)} p_0(x,y) \Big). \end{aligned}$$

For $p_1(x,y), p_0(x,y) = 1$; we have

$$\begin{aligned} p(t,x,y) =& 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} \sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n(x,y) \\ &+ \sum_{m=2}^{\infty} e^{-\lambda_m t} \Big(\frac{2m+\theta-1}{m!} (-1)^{m-1} (\theta+1)_{(m-1)} m + \frac{2m+\theta-1}{m!} (-1)^{m} \theta_{(m-1)} \Big) \\ =& 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} \sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n(x,y) \\ &+ \sum_{m=2}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} (-1)^{m-1} [m(\theta+1)_{(m-1)} - \theta_{(m-1)}]. \end{aligned}$$

When $m = 1, m(\theta+1)_{(m-1)} - \theta_{(m-1)} = 0$. Then

$$\begin{aligned} p(t,x,y) =& 1 - \sum_{m=1}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} (-1)^{m-1} \theta_{(m-1)} \\ &+ \sum_{m=1}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} (-1)^{m-1} m(\theta+1)_{(m-1)} \\ &+ \sum_{m=2}^{\infty} e^{-\lambda_m t} \sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n(x,y) \end{aligned}$$

$$=d_0^\theta(t)+d_1^\theta(t)+\sum_{m=2}^{\infty}e^{-\lambda_m t}\sum_{n=2}^{m}\frac{2m+\theta-1}{m!}(-1)^{m-n}\binom{m}{n}(n+\theta)_{(m-1)}p_n(x,y)$$

Switching the order of summation, we have

$$p(t,x,y)=d_0^\theta(t)+d_1^\theta(t)+\sum_{n=2}^{+\infty}d_n^\theta(t)p_n(x,y).$$

Thus, (3.2.5) is proved. Lastly, since $\sum_{n=1}^{\infty}d_n^\theta(t)=1$, and $d_n^\theta(t)\geq 0$, obviously $\exists n_0>0$, such that $d_{n_0}^\theta(t)>0$, and $p_{n_0}(x,y)>0, \forall x,y\in\bar{\nabla}_\infty$. Hence,

$$p(t,x,y)\geq d_{n_0}^\theta(t)p_{n_0}(x,y)>0.$$

Therefore, $p(t,x,y)$ is strictly positive. □

Defining $\nu_n^\theta(x,dy)=p_n(x,y)\mathrm{PD}(\theta)(dy)$, we have a transition probability structure similar to that of the neutral Fleming-Viot process in [Ethier and Griffiths, 1993], i.e.

$$P(t,x,A)=(d_0^\theta(t)+d_1^\theta(t))\mathrm{PD}(\theta)(A)+\sum_{n=2}^{\infty}d_n^\theta(t)\nu_n^\theta(x,A). \tag{3.2.7}$$

Thus, the uniform ergodic inequality, obtained by R.C. Griffiths and S.N. Ethier in [Ethier and Griffiths, 1993], can also be derived from this transition probability structure.

Theorem 3.3. *Defining $K=\frac{1}{2}(2+\theta)(3+\theta)$, we have*

$$\sup_{x\in\bar{\nabla}_\infty}\|P(t,x,\cdot)-\mathrm{PD}(\theta)(\cdot)\|_{\mathrm{var}}\leq K\exp\{-(\theta+1)t\},\quad \forall t\geq 0.$$

Proof. By (3.2.7), we have

$$\begin{aligned}
&|P(t,x,A)-\mathrm{PD}(\theta)(A)|\\
=&\Big|\sum_{n=2}^{\infty}d_n^\theta(t)\Big(\nu_n^\theta(x,A)-\mathrm{PD}(\theta)(A)\Big)+(d_0^\theta(t)+d_1^\theta(t))\Big(\mathrm{PD}(\theta)(A)-\mathrm{PD}(\theta)(A)\Big)\Big|\\
\leq&\sum_{n=2}^{\infty}d_n^\theta(t)\Big|\nu_n^\theta(x,A)-\mathrm{PD}(\theta)(A)\Big|\\
\leq&\sum_{n=2}^{\infty}d_n^\theta(t)\\
=&1-d_0^\theta(t)-d_1^\theta(t)\leq\frac{(2+\theta)(3+\theta)}{2}e^{-(\theta+1)t}.
\end{aligned}$$

The last inequality is due to Tavaré's estimations in [Tavaré, 1984]. □

The semigroup of the one-parameter neutral model has an upper bound stated in the following proposition. It is due to [Feng et al., 2011].

Proposition 3.4. *There exists a constant $c > 1$, such that*

$$p(t, x, y) \leq ct^{c\frac{\log t}{t}}, t > 0.$$

This inequality guarantees that both the Poincaré inequality and the log-sobolev inequality hold in the one-parameter neutral model; whereas W. Stannat has shown that those inequalities do not hold in the neutral Fleming-Viot model in [Stannat, 2000].

3.3 Two-parameter Model

The two-parameter model is an extension of the one-parameter neutral model. The two-parameter model, however, does not have any biological interpretation. Moreover, the one-parameter neutral process is the corresponding atomic process of the neutral Fleming-Viot process; whereas the existence of the measure-valued process associated with the two-parameter model is still unknown. In this section, we will concentrate on the two-parameter model, and a new representation of its transition density function is obtained. Finally a uniform ergodic inequality is obtained as well.

3.3.1 Transition Density Function of the Two-parameter Model

The two-parameter model, denoted by $X_t^{\theta,\alpha}$, is a Feller diffusion generated by the closure of generator $G^{\theta,\alpha}$ on $C(\bar{\nabla}_\infty)$, where

$$G^{\theta,\alpha} = \frac{1}{2}\sum_{i,j=1}^{\infty} x_i(\delta_{ij} - x_j)\frac{\partial^2}{\partial x_i \partial x_j} - \frac{1}{2}\sum_{i=1}^{\infty}(\theta x_i + \alpha)\frac{\partial}{\partial x_i}, x \in \bar{\nabla}_\infty.$$

The core of $G^{\theta,\alpha}$ is $\mathcal{P}$. But, for $f \in \mathcal{P}$, $G^{\theta,\alpha} f$ is first evaluated on ∇_∞ and then continuously extended to $\bar{\nabla}_\infty$. We can extend the domain of $G^{\theta,\alpha}$ through the bounded pointwise convergence. However, we can not treat $\{x_i, i \geq 1\}$ as regular smooth functions in $D(\mathcal{G}^{\theta,\alpha})$. Please refer to [Petrov, 2009] for more detailed explanations.

The Two-parameter model is also a reversible model, the stationary distribution of which is depicted by the two-parameter Poisson-Dirichlet distribution, $\mathrm{PD}(\theta, \alpha)$. The generator $G^{\theta,\alpha}$ is also a self-adjoint operator in the Hilbert space $L^2(\mathrm{PD}(\theta, \alpha))$, the inner product of which, denoted by $\langle \cdot, \cdot \rangle_{\theta,\alpha}$, is defined to be $\int_{\bar{\nabla}_\infty} fgd\mathrm{PD}(\theta, \alpha)$. The complete set of the eigenvalues of $G^{\theta,\alpha}$ were found by L.A. Petrov in [Petrov, 2009], and were rederived in [Feng et al., 2011]. Surprisingly, the two-parameter model share the same set of eigenvalues with the one-parameter neutral model. They are $-\lambda_m = -\frac{(m+\theta-1)m}{2}, m \geq 2$ and $-\lambda_0 = 0$.

Proposition 3.5. *For each partition $\eta \in \tilde{\mathcal{M}}$, there is an eigenfunction $\phi_\eta^{\theta,\alpha}$ of $G^{\theta,\alpha}$, such that $\phi_\eta^{\theta,\alpha} = \varphi_\eta + g_\eta^{\theta,\alpha}$, where $g_\eta^{\theta,\alpha} \in \mathcal{P}_{|\eta|-1}$.*

Proof. This can be proved by the exactly the same argument used in Lemma 2.2 in [Ethier, 1992]. □

Proposition 3.6. *For each eigenvalue $-\lambda_m, m \geq 2$, we have one particular eigenfunction $\phi_m^{\theta,\alpha}(x)$, which can be explicitly expressed as*

$$\varphi_m + \sum_{i=1}^{m-1} \left(\prod_{j=i}^{m-1} \frac{\frac{(j+1)(j-\alpha)}{2}}{\lambda_j - \lambda_m} \right) \varphi_i.$$

Proof. Notice that $G^{\theta,\alpha}\varphi_m = -\lambda_m \varphi_m + \frac{m(m-\alpha-1)}{2}\varphi_{m-1}$, where $-\lambda_m$ is an eigenvalue. Then its corresponding eigenfunction can be written as $\varphi_m + g_m^{\theta,\alpha}$, where $g_m^{\theta,\alpha}$ satisfies the equation

$$(G^{\theta,\alpha} + \lambda_m) g_m^{\theta,\alpha} = -\frac{m(m-\alpha-1)}{2}\varphi_{m-1}. \tag{3.3.1}$$

Suppose that $g_m^{\theta,\alpha} = \sum_{i=1}^{m-1} a_i \varphi_i, \varphi_1 = 1$. Then we can actually solve for $\{a_i, 1 \leq i \leq m-1\}$. Substituting $g_m^{\theta,\alpha}$ into (3.3.1), we have

$$-\sum_{i=2}^{m-1} a_i \lambda_i \varphi_i + \sum_{i=2}^{m-1} a_i \frac{i(i-\alpha-1)}{2}\varphi_{i-1} + \sum_{i=1}^{m-1} \lambda_m a_i \varphi_i = -\frac{m(m-\alpha-1)}{2}\varphi_{m-1}.$$

Then

$$\begin{aligned}&(\lambda_m - \lambda_{m-1}) a_{m-1} \varphi_{m-1} + \sum_{i=2}^{m-2} \left[a_{i+1} \frac{(i+1)(i-\alpha)}{2} + (\lambda_m - \lambda_i) a_i \right] \varphi_i + \lambda_m a_1 \\ &+ a_2(1-\alpha) = -\frac{m(m-\alpha-1)}{2}\varphi_{m-1}.\end{aligned}$$

Thus, $a_{m-1} = \frac{\frac{m(m-\alpha-1)}{2}}{\lambda_{m-1}-\lambda_m}, \frac{a_i}{a_{i+1}} = \frac{\frac{(i+1)(i-\alpha)}{2}}{\lambda_i - \lambda_m}$. Therefore,

$$a_i = \frac{a_i}{a_{i+1}} \cdots \frac{a_{m-2}}{a_{m-1}} \cdot a_{m-1} = \prod_{j=i}^{m-1} \frac{\frac{(j+1)(j-\alpha)}{2}}{\lambda_j - \lambda_m}.$$

The conclusion follows. □

Therefore, theoretically, we can get all the eigenfunctions. Let us denote them by $\{\phi_\eta^{\theta,\alpha}, \eta \in \tilde{\mathcal{M}}\}$, where $\phi_\eta^{\theta,\alpha}$ is the eigenfunction of the eigenvalue $-\lambda_{|\eta|}$. If $|\eta| \neq |\xi|$, then $\phi_\eta^{\theta,\alpha}$ and $\phi_\xi^{\theta,\alpha}$ are orthogonal to each other; for they are the eigenfunctions of two different eigenvalues. If $|\eta| = |\xi|$, then $\phi_\eta^{\theta,\alpha}$ and $\phi_\xi^{\theta,\alpha}$ are not necessarily orthogonal to each other. Therefore, if we apply the Gram-Schmidt orthogonalization procedure to the set of eigenfunctions of each eigenvalue, then we will get a whole set of orthogonal functions. Let us denote the set of these eigenfunctions by $\{\bar{\psi}_\eta^{\theta,\alpha} \mid \eta \in \tilde{\mathcal{M}}\}$.

Proposition 3.7. *For $\eta \in \tilde{\mathcal{M}}, \psi_\eta^{\theta,\alpha} = \bar{\psi}_\eta^{\theta,\alpha}$. If we define $\bar{\chi}_\eta^{\theta,\alpha}(x)$ to be $\frac{\bar{\psi}_\eta^{\theta,\alpha}(x)}{\|\bar{\psi}_\eta^{\theta,\alpha}\|_{2,\theta,\alpha}}$, then $\chi_\eta^{\theta,\alpha} = \bar{\chi}_\eta^{\theta,\alpha}, \eta \in \tilde{\mathcal{M}}$.*

Proof. Notice that $\phi_1^{\theta,\alpha} = 1$ and $\phi_2^{\theta,\alpha} = \varphi_2 - \frac{1-\alpha}{1+\theta}$. Therefore, we can use mathematical induction to prove this proposition. The remaining arguments are quite similar to that of Proposition 3.3. □

Through eigenfunction expansion, an explicit representation of the transition density function of the two-parameter model is obtained in [Feng et al., 2011]. The following theorem is from [Feng et al., 2011].

Theorem 3.4. *The transition probability $P^{\theta,\alpha}(t,x,dy)$ of the two-parameter model has a density function, denoted by $p^{\theta,\alpha}(t,x,y)$ i.e.*

$$P_t^{\theta,\alpha} f(x) = \int_{\bar{\nabla}_\infty} f(y) P^{\theta,\alpha}(t,x,dy) = \int_{\bar{\nabla}_\infty} f(y) p^{\theta,\alpha}(t,x,y) \mathrm{PD}(\theta,\alpha)(dy),$$

where $p^{\theta,\alpha}(t,x,y)$ has the following explicit expression

$$p^{\theta,\alpha}(t,x,y) = 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} Q_m^{\theta,\alpha}(x,y)$$

$$Q_m^{\theta,\alpha}(x,y) = \frac{2m+\theta-1}{m!} \sum_{n=0}^{m} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n^{\theta,\alpha}(x,y),$$

and $p_n^{\theta,\alpha}(x,y) = \sum_{|\eta|=n} \frac{p_\eta(x) p_\eta(y)}{\int p_\eta d\mathrm{PD}(\theta,\alpha)}$.

Remark 3.3. *$Q_m^{\theta,\alpha}(x,y)$ was defined to be $\sum_{|\eta|=m} \bar{\chi}_\eta^{\theta,\alpha}(x)\bar{\chi}_\eta^{\theta,\alpha}(x)$, where $\{\bar{\chi}_\eta^{\theta,\alpha}, \eta \in \tilde{\mathcal{M}}\}$ is one specific orthogonal basis of the Hilbert space $L^2(\mathrm{PD}(\theta,\alpha))$. Clearly, the definition of $Q_m^{\theta,\alpha}(x,y)$ is independent of orthonormal basis of $L^2(\mathrm{PD}(\theta,\alpha))$. Moreover, $p_n^{\theta,\alpha}(x,y)$ is a probability kernel and all $p_\eta(x)$ are the continuous extensions of $p_\eta(x)$ restricted to ∇_∞. Therefore, $p_n^{\theta,\alpha}(x,y) > 0, \forall x,y \in \bar{\nabla}_\infty$.*

Likewise, we have another representation of the transition density function, which can not be obtained through direct computation used in [Feng, 2010]. Simply because the existence of the measure-valued process associated with the two-parameter model is still unclear, let alone its transition probability structure. Therefore, our method has its own advantage.

Theorem 3.5. *The transition density function $p^{\theta,\alpha}(t,x,y)$ of the two-parameter model is continuous, strictly positive and symmetric on $(0,+\infty)\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$. Moreover, $p^{\theta,\alpha}(t,x,y)$ has the following explicit expression*

$$p^{\theta,\alpha}(t,x,y)=d_0^\theta(t)+d_1^\theta(t)+\sum_{n=2}^{\infty}d_n^\theta(t)p_n^{\theta,\alpha}(x,y).$$

Proof. The transition density of the two-parameter model is

$$p^{\theta,\alpha}(t,x,y)=1+\sum_{m=2}^{\infty}e^{-\lambda_m t}Q_m^{\theta,\alpha}(x,y),\lambda_m=\frac{m(m+\theta-1)}{2}.$$

And for $\theta>-\alpha, 0<\alpha<1$,

$$Q_m^{\theta,\alpha}(x,y)=\frac{2m+\theta-1}{m!}\sum_{n=0}^{m}(-1)^{m-n}\binom{m}{n}(n+\theta)_{(m-1)}p_n^{\theta,\alpha}(x,y).$$

Now we are going to show that for any $t_0>0$, $\sum_{m=2}^{\infty}e^{-\lambda_m t}Q_m^{\theta,\alpha}(x,y)$ converges uniformly on $[t_0,+\infty)\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$. To this end, we first need to show

$$p_n^{\theta,\alpha}(x,y)\le\frac{(\theta+\alpha)_{(n)}}{((\theta+\alpha)^n\wedge 1)(1-\alpha)^n},$$

and there exists constant c and d, such that

$$Q_m^{\theta,\alpha}(x,y)\le(cm^d)^m.$$

Indeed,

$$\begin{aligned}
p_n^{\theta,\alpha}(x,y)&\le\max_{|\eta|=n}\frac{1}{\int p_\eta d\mathrm{PD}(\theta,\alpha)}\sum_{|\eta|=n}p_\eta(x)p_\eta(y)\\
&\le\max_{|\eta|=n}\frac{1}{\int p_\eta d\mathrm{PD}(\theta,\alpha)}\left(\sum_{|\eta|=n}p_\eta(x)\right)\left(\sum_{|\eta|=n}p_\eta(y)\right)\\
&=\max_{|\eta|=n}\frac{1}{\int p_\eta d\mathrm{PD}(\theta,\alpha)}.
\end{aligned}$$

By the Pitman sampling formula,

$$\int p_\eta d\mathrm{PD}(\theta,\alpha)=\frac{n!}{\eta_1!\cdots\eta_l!\alpha_1!\cdots\alpha_n!}\prod_{i=0}^{l-1}(\theta+i\alpha)\frac{\prod_{j=1}^{n}\left((1-\alpha)_{(j-1)}\right)^{a_j}}{\theta_{(n)}}$$

$$\geq \frac{(1-\alpha)^n\Big((\theta+\alpha)^n \wedge 1\Big)}{(\theta+\alpha)_{(n)}};$$

then

$$p_n^{\theta,\alpha}(x,y) \leq \frac{(\theta+\alpha)_{(n)}}{\Big((\theta+\alpha)^n \wedge 1\Big)(1-\alpha)^n}.$$

Thus,

$$\begin{aligned}||Q_m^{\theta,\alpha}||_\infty &\leq \frac{2m+\theta+\alpha-1}{m!}\sum_{n=0}^{m}\binom{m}{n}(n+\theta+\alpha)_{(m-1)}||p_n^{\theta,\alpha}||_\infty \\ &\leq \frac{2m+\theta+\alpha-1}{m!}\sum_{n=0}^{m}\binom{m}{n}(n+\theta+\alpha)_{(m-1)}\frac{(\theta+\alpha)_{(n)}}{\Big((\theta+\alpha)^n \wedge 1\Big)(1-\alpha)^n} \\ &\leq (2m+\theta+\alpha-1)\sum_{n=0}^{m}(n+\theta+\alpha)_{(m-1)}\frac{(\theta+\alpha)_{(n)}}{\Big((\theta+\alpha)^n \wedge 1\Big)(1-\alpha)^n}.\end{aligned}$$

Since $n+\theta+\alpha \leq 2m+\theta+\alpha-1, \forall 0 \leq n \leq m$, then $(\theta+\alpha+n)_{(m-1)} \leq (2m+\theta+\alpha-1)^{m-1}$. Hence, $(\theta+\alpha)_{(n)} \leq (\theta+\alpha+n)^n \leq (2m+\theta+\alpha-1)^m$. Moreover,

$$\frac{1}{(\theta+\alpha)^n \wedge 1} \leq \left(\frac{1}{(\theta+\alpha)\wedge 1}\right)^n \leq \left(\frac{1}{(\theta+\alpha)\wedge 1}\right)^m.$$

Therefore,

$$||Q_m^{\theta,\alpha}||_\infty \leq \frac{1}{(1-\alpha)^m}(m+1)(2m+\theta+\alpha-1)^{2m}\left(\frac{1}{(\theta+\alpha)\wedge 1}\right)^m.$$

There exists constant $c_1 > 0$, such that $m+1 \leq (c_1 m)^m$ and $2m+\theta+\alpha-1 \leq c_1 m^m$. Then

$$Q_m^{\theta,\alpha}(x,y) \leq (cm^d)^m,$$

where $c = \frac{c_1^3}{(1-\alpha)((\theta+\alpha)\wedge 1)}$, $d = 3$. Therefore, for $t_0 > 0$, and $\forall t \in [t_0, \infty)$, we have

$$\begin{aligned}\sum_{m=2}^{\infty}\exp\{-\lambda_m t\}Q_m^{\theta,\alpha}(x,y) &\leq \sum_{m=2}^{\infty}\exp\{-\lambda_m t\}(cm^d)^m \\ &= \sum_{m=2}^{\infty}\left(\frac{cm^d}{e^{\frac{t(m+\theta-1)}{2}}}\right)^m \leq \sum_{m=2}^{\infty}\left(\frac{cm^d}{e^{\frac{t_0(m+\theta-1)}{2}}}\right)^m.\end{aligned}$$

Since $\lim_{m\to+\infty} \frac{cm^d}{e^{\frac{t_0(m+\theta-1)}{2}}} = 0, \exists M > 0$ such that $\forall m > M$,

$$\frac{cm^d}{e^{\frac{t_0(m+\theta-1)}{2}}} < \frac{1}{2}.$$

Then

$$\sum_{m=2}^{\infty} \left(\frac{cm^d}{e^{\frac{t_0(m+\theta-1)}{2}}} \right)^m = \sum_{m=2}^{M} \left(\frac{cm^d}{e^{\frac{t_0(m+\theta-1)}{2}}} \right)^m + \sum_{m=M+1}^{\infty} \left(\frac{cm^d}{e^{\frac{t_0(m+\theta-1)}{2}}} \right)^m.$$

Because $\sum_{m\geq 1} \frac{1}{2^m}$ is convergent; then, by Weierstrass's M-test, $p^{\theta,\alpha}(t,x,y)$ is uniformly convergent on $[t_0, +\infty) \times \bar{\nabla}_\infty \times \bar{\nabla}_\infty$, and thus is continuous.
Next, by Fubini's theorem, we can rearrange $p^{\theta,\alpha}(t,x,y)$ by switching the order of summation. Then

$$\begin{aligned} p^{\theta,\alpha}(t,x,y) = 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} \Big(& \sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n^{\theta,\alpha}(x,y) \\ & + \frac{2m+\theta-1}{m!} (-1)^{m-1} (\theta+1)_{(m-1)} m p_1^{\theta,\alpha}(x,y) \\ & + \frac{2m+\theta-1}{m!} (-1)^m \theta_{(m-1)} p_0^{\theta,\alpha}(x,y) \Big). \end{aligned}$$

For $p_1^{\theta,\alpha}(x,y), p_0^{\theta,\alpha}(x,y) = 1$; we have

$$\begin{aligned} p^{\theta,\alpha}(t,x,y) =& 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} \sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n^{\theta,\alpha}(x,y) \\ & + \sum_{m=2}^{\infty} e^{-\lambda_m t} \Big(\frac{2m+\theta-1}{m!} (-1)^{m-1} (\theta+1)_{(m-1)} m \\ & + \frac{2m+\theta-1}{m!} (-1)^m \theta_{(m-1)} \Big) \\ =& 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t} \sum_{n=2}^{m} \frac{2m+\theta-1}{m!} (-1)^{m-n} \binom{m}{n} (n+\theta)_{(m-1)} p_n^{\theta,\alpha}(x,y) \\ & + \sum_{m=2}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} (-1)^{m-1} \Big[m(\theta+1)_{(m-1)} - \theta_{(m-1)} \Big]. \end{aligned}$$

When $m = 1, m(\theta+1)_{(m-1)} - \theta_{(m-1)} = 0$. Then we have

$$p^{\theta,\alpha}(t,x,y) = 1 - \sum_{m=1}^{\infty} e^{-\lambda_m t} \frac{2m+\theta-1}{m!} (-1)^{m-1} \theta_{(m-1)}$$

$$
\begin{aligned}
&+\sum_{m=1}^{\infty} e^{-\lambda_m t}\frac{2m+\theta-1}{m!}(-1)^{m-1}m(\theta+1)_{(m-1)}\\
&+\sum_{m=2}^{\infty} e^{-\lambda_m t}\sum_{n=2}^{m}\frac{2m+\theta-1}{m!}(-1)^{m-n}\binom{m}{n}(n+\theta)_{(m-1)}p_n(x,y)\\
=&d_0^{\theta}(t)+d_1^{\theta}(t)+\sum_{m=2}^{\infty} e^{-\lambda_m t}\sum_{n=2}^{m}\frac{2m+\theta-1}{m!}(-1)^{m-n}\binom{m}{n}(n+\theta)_{(m-1)}p_n^{\theta,\alpha}(x,y).
\end{aligned}
$$

Let us switch the order of summation. Then we have

$$
p^{\theta,\alpha}(t,x,y)=d_0^{\theta}(t)+d_1^{\theta}(t)+\sum_{n=2}^{+\infty}d_n^{\theta}(t)p_n^{\theta,\alpha}(x,y).
$$

□

Remark 3.4. *The the set of coefficients are identical to that of the one-parameter neutral model. If we define* $\nu_n^{\theta,\alpha}(x,dy)=p_n^{\theta,\alpha}(x,y)\mathrm{PD}(\theta,\alpha)(dy)$*, then all* $\nu_n^{\theta,\alpha}(x,\cdot)$ *concentrate on* ∇_∞*. Thus, we have the following transition probability structure*

$$
P^{\theta,\alpha}(t,x,A)=\Big[d_0^{\theta}(t)+d_1^{\theta}(t)\Big]\mathrm{PD}(\theta,\alpha)(A)+\sum_{n=2}^{\infty}d_n^{\theta}(t)\nu_n^{\theta,\alpha}(x,A). \tag{3.3.2}
$$

Therefore, from (3.3.2) we can see some rough structures of the transition probability of the two-parameter labelled model if it exists.

In theorem 3.5, the coefficients $d_n^{\theta}(t), n\geq 0$, still keep the same expression even though $d_0^{\theta}(t)$ could be negative when $\theta<0$. However, $d_0^{\theta}(t)+d_1^{\theta}(t)$ is always positive such that $d_0^{\theta}(t)+d_1^{\theta}(t), d_n^{\theta}(t), n\geq 2$, are still a probability mass function. What is more, Taveré's estimation still holds when $-1<\theta\leq 0$. This idea is suggested by S.N. Ethier, to whom I am greatly indebted.

Proposition 3.8. *For* $-1<\theta\leq 0$*, we have*

$$
e^{-\lambda_n t}\leq\sum_{k=n}^{\infty}d_k^{\theta}(t)\leq\frac{(n+\theta)_{(n)}}{n_{[n]}}e^{-\lambda_n t}.
$$

In particular, when $n=2$*, we know*

$$
\sum_{k=2}^{\infty}d_k^{\theta}(t)\leq\frac{(2+\theta)(3+\theta)}{2}e^{-(\theta+1)t}.
$$

Proof. Consider a pure-death Markov chain B_t in $\{1, 2, \cdots, m\}$ with Q matrix,

$$Q = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \lambda_2 & -\lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_3 & -\lambda_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_m & -\lambda_m, \end{pmatrix}$$

where $\lambda_k = \frac{k(k+\theta-1)}{2}, k \geq 2$. Running the similar arguments in Theorem 4.3 in [Feng, 2010], we will be able to find all the left eigenvectors and right eigenvectors of Q. Denote the matrix consisting of left eigenvectors by $U = (u_{ij})$ and the matrix consisting of right eigenvectors by $V = (v_{ij})$, where

$$u_{ij} = \begin{cases} \delta_{1j} & i = 1 \\ 0 & j > i > 1 \\ (-1)^{i-j} \binom{i}{j} \frac{(j+\theta)_{(i-1)}}{(i+\theta)_{(i-1)}}, & j \leq i, i > 1, \end{cases}$$

and

$$v_{ij} = \begin{cases} 1 & j = 1 \\ 0 & j > i \\ \binom{i}{j} \frac{(j+\theta)_{(j)}}{(i+\theta)_{(j)}}, & 1 < j \leq i. \end{cases}$$

Note that the row vectors of U are left eigenvectors of Q and the column vectors of V are the right eigenvectors of Q. Similarly, we can also show that $UV = I$ and Q is diagonlized as $V\Lambda U$, where $\Lambda = \mathrm{diag}\{0, -\lambda_2, \cdots, -\lambda_{\mathrm{m}}\}$. Therefore, the transition matrix P_t is

$$P_t = e^{tQ} = V e^{\Lambda t} U.$$

By direct computation, we know, for $2 \leq n \leq m$,

$$P_{mn}(t) = \sum_{k=n}^{m} (-1)^{k-n} \binom{m}{k} \binom{k}{n} \frac{(\theta+k)_{(k)}}{(\theta+m)_{(k)}} \frac{(\theta+n)_{(k-1)}}{(\theta+k)_{(k-1)}} e^{-\lambda_k t}.$$

Letting $m \to +\infty$, we have $d_n^\theta(t) = \lim_{m\to\infty} P_{mn}(t)$.

The remaining arguments are essentially due to Tavaré.

By the martingale argument in chapter 6 of [Karlin and Taylor, 1981], we know

$$Z_n(t) = \frac{e^{\lambda_n t} (B_t)_{[n]}}{(B_t + \theta)_{(n)}},$$

because $e^{-\lambda_n t}$ is one eigenvalue of P_t and

$$(0,0,\cdots,0,\frac{n_{[n]}}{(n+\theta)_{(n)}},\cdots,\frac{k_{[n]}}{(k+\theta)_{(n)}},\cdots,\frac{m_{[n]}}{(m+\theta)_{(n)}})^{\mathrm{T}}$$

is the corresponding eigenvector. So

$$EZ_n(t)=Z_n(0)=\frac{m_{[n]}}{(m+\theta)_{(n)}}.$$

Since, for $n\le k\le m$,

$$\frac{n_{[n]}}{(n+\theta)_{(n)}}\le\frac{k_{[n]}}{(k+\theta)_{(n)}}\le\frac{m_{[n]}}{(m+\theta)_{(n)}},$$

and

$$\frac{e^{-\lambda_n t}m_{[n]}}{(m+\theta)_{(n)}}=e^{-\lambda_n t}EZ_n(t)=\sum_{k=n}^{m}\frac{k_{[n]}}{(k+\theta)_{(n)}}P_{mk}(t),$$

we have

$$\frac{n_{[n]}}{(n+\theta)_{(n)}}P(B_t\ge n|B_0=m)\le\frac{e^{-\lambda_n t}m_{[n]}}{(m+\theta)_{(n)}}\le\frac{m_{[n]}}{(m+\theta)_{(n)}}P(B_t\ge n|B_0=m).$$

Thus, we have

$$e^{-\lambda_n t}\le P(B_t\ge n|B_0=m)\le\frac{(n+\theta)_{(n)}}{n_{[n]}}e^{-\lambda_n t}.$$

Letting $m\to\infty$, we have

$$e^{-\lambda_n t}\le\sum_{k=n}^{\infty}d_k^{\theta}(t)\le\frac{(n+\theta)_{(n)}}{n_{[n]}}e^{-\lambda_n t}.$$

□

3.3.2 An Ergodic Inequality of the Two-parameter Model

Now we have a transition probability structure similar to that of the one-parameter neutral model, and it is a convex combination of probability measures. The set of coefficients are as same as what are defined in the transition probability structure of the one-parameter neutral model. Even when $-\alpha<\theta<0$, this set of coefficients are the entrance law of the pure death Markov chain considered in proposition 3.8 . Thus, we can still apply the argument in the Theorem 3.3.

Theorem 3.6. *Defining* $K = \frac{1}{2}(2+\theta)(3+\theta)$, *we know*

$$\sup_{x\in\bar{\nabla}_\infty} \left\| P^{\theta,\alpha}(t,x,\cdot) - \mathrm{PD}(\theta,\alpha)(\cdot) \right\|_{\mathrm{var}} \leq K e^{-(\theta+1)t}.$$

Proof. Making use of (3.3.2) and proposition 3.8, this theorem can be proved easily. □

This theorem should be the strongest ergodic theorem known so far. Moreover, it is new.

3.4 One-parameter Selective Model

In Darwin's natural selection theory, selection plays a very important role. Furthermore, presumably, selective models are more realistic. However, among all selective models, models with symmetric dominance proposed by G.A. Watterson in [Watterson, 1977b] are comparatively tractable; because, as the atomic processes of the Fleming-Viot processes with symmetric dominance, they are diffusion processes. However, atomic processes of other measure-valued processes, if there is any, might not be Markovian, thereby intractable.

In the one-parameter selective model, three evolutionary forces are involved. They are random sampling, mutation and selection. Still, random sampling removes genetic variation and pushes the system toward $(1, 0, \cdots)$. By entering more new species into the system, mutations attract the system to $(0, 0, \cdots)$. But the effects of selection vary according to its different types. For instance, the selective model with symmetric dominance has selection $\sigma_{ij} = \sigma\delta_{ij}$, where σ represents the selection intensity. If $\sigma > 0$, then homozygotes are favoured. This selection is underdominant. If $\sigma < 0$ heterozygotes are favoured. This selection is overdominant. Obviously, when $\sigma = 0$, it corresponds to the one-parameter neutral model.

In this section, we will focus on the transition density and the ergodicity of the one-parameter selective model.

3.4.1 The Transition Density Function of the One-parameter Selective Model

The one-parameter selective model is first constructed in [Ethier and Kurtz, 1998] and characterized by the generator

$$G_\sigma = G + \sigma \sum_{i=1}^{\infty} x_i(x_i - \varphi_2(x)) \frac{\partial}{\partial x_i}, \quad x \in \bar{\nabla}_\infty.$$

The core of G_σ is also $\mathcal{P}$. $\forall f \in \mathcal{P}$, $G_\sigma f(x)$ is first evaluated in ∇_∞ and then extended continuously to $\bar{\nabla}_\infty$. The domain of G_σ can also be extended through the bounded pointwise convergence; but the natural coordinate functions, such as $x_i, i \geq 1$, are not included in the domain of G_σ. Furthermore, the $C_{\bar{\nabla}_\infty}([0,\infty))$ martingale problem for G_σ can also be shown to be well-posed. Please refer to [Ethier and Kurtz, 1998] for detailed proof. Let X_t^σ be the one-parameter model. Then we denote its transition probability by $P^\sigma(t,x,dy)$, which has a representation stated in the proof of theorem 4.3 in [Ethier and Kurtz, 1998].

Theorem 3.7.

$$\begin{aligned} P^\sigma(t,x,dy) = &\exp\left\{\frac{\sigma}{2}\varphi_2(y) - \frac{\sigma}{2}\varphi_2(x)\right\} \\ &E_x\left[\exp\left\{-\int_0^t g(X_s)ds\right\}\Big|X_t = y\right] P(t,x,dy), \end{aligned}$$

where

$$g(x) = \frac{\sigma}{2}(1-(1+\theta)\varphi_2(x)) + \frac{\sigma^2}{2}(\varphi_3(x) - \varphi_2^2(x)).$$

Proof. Since the $C_{\bar{\nabla}_\infty}([0,\infty))$ martingale problem for G is well-posed, $\forall f \in \mathcal{D}(\bar{G})$, for any $x \in \bar{\nabla}_\infty$, we know there is a solution P_x such that

$$M_t^x(f) = f(X_t) - f(x) - \int_0^t Gf(X_s)ds \tag{3.4.1}$$

is a continuous martingale under P_x, where X_t is the associated canonical process, i.e. for any $\omega \in C_{\bar{\nabla}_\infty}([0,\infty)), X_t(\omega) = \omega_t$.

Now consider

$$D_t = \exp\left\{M_t^x(\phi) - \frac{1}{2}\langle M^x(\phi), M^x(\phi)\rangle_t\right\},$$

where $\phi(x) = \frac{\sigma}{2}\varphi_2(x)$. By (3.4.1), $M_t^x(\phi)$ is a continuous martingale under P_x. More specifically,

$$\begin{aligned} D_t &= \exp\left\{\phi(X_t) - \phi(x) - \int_0^t G\phi(X_s)ds - \frac{1}{2}\int_0^t \Gamma(\phi,\phi)(X_s)ds\right\} \\ &= \exp\left\{\frac{\sigma}{2}\varphi_2(X_t) - \frac{\sigma}{2}\varphi_2(x) - \int_0^t \frac{\sigma}{2}(1-(1+\theta)\varphi_2)(X_s)ds \right. \\ &\quad \left. - \frac{\sigma^2}{2}\int_0^t (\varphi_3 - \varphi_2^2)(X_s)ds\right\} \\ &= \exp\left\{\frac{\sigma}{2}\varphi_2(X_t) - \frac{\sigma}{2}\varphi_2(x) - \int_0^t g(X_s)ds\right\}. \end{aligned}$$

Thus, D_t is bounded for every $t > 0$.

By Itô's formula, we have

$$\begin{aligned} dD_t =& D_t\Big(dM_t^x(\phi) - \frac{1}{2}d\langle M^x(\phi), M^x(\phi)\rangle_t\Big) + \frac{1}{2}D_t d\langle M^x(\phi), M^x(\phi)\rangle_t \\ =& D_t dM_t^x(\phi). \end{aligned}$$

Then D_t is a continuous local martingale under P_x as well. Since D_t is bounded for each $t > 0$, it is also a continuous martingale. In addition, $E_x D_t = 1$. Let $\mathcal{F} = \sigma(X_t, 0 < t < \infty)$, and $\mathcal{F}_t = \sigma(X_s, 0 < s \le t)$. Then we can define a new measure P_x^σ on $\mathcal{F}$ such that

$$P_x^\sigma\Big|_{\mathcal{F}_t} = D_t dP_x\Big|_{\mathcal{F}_t}, t \ge 0.$$

By Girsanov theorem,

$$M_t^x(f) - D_t^{-1}\langle M^x(f), D\rangle_t, \tag{3.4.2}$$

is a martingale under P_x^σ. Simplifying (3.4.2), we know

$$f(X_t) - f(x) - \int_0^t G_\sigma(f)(X_s)ds$$

is a martingale under P_x^σ. Since the martingale problem of G_σ is well-posed, P_x^σ should be the distribution of the one-parameter selective model X_t^σ. The theorem is thus proved. □

By this theorem, it is easy to know that the one-parameter selective model has a transition density function with respect to its stationary distribution $\pi_\sigma(dx)$ (refer to [Ethier and Kurtz, 1998]), where

$$\pi_\sigma(dx) = C_\sigma \exp\{\sigma\varphi_2(x)\}\mathrm{PD}(\theta)(dx),$$

and

$$C_\sigma = \frac{1}{\int_{\bar\nabla_\infty} \exp\{\sigma\varphi_2(x)\}\mathrm{PD}(\theta)(dx)}.$$

More specifically, let us denote the transition density function of the one-parameter selective model by $p_\sigma(t, x, y)$. Then $p_\sigma(t, x, y)$ is

$$\frac{1}{C_\sigma}\exp\Big\{-\frac{\sigma}{2}(\varphi_2(x)+\varphi_2(y))\Big\} E_x\Big[\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t = y\Big]p(t, x, y).$$

Moreover, we can easily obtain the following kernel estimation.

Proposition 3.9. *There exists a constant $c > 1$, such that*

$$p_\sigma(t, x, y) \le \frac{1}{C_\sigma}e^{|\sigma|(1+\theta)t+\sigma^2+3|\sigma|}ct^{c^{\frac{\log t}{t}}}, t > 0, x, y \in \bar\nabla_\infty.$$

Proof. This estimation can be easily obtained from the representation of $p_\sigma(t,x,y)$. □

Remark 3.5. *This estimation guarantees that the log-sobolev inequality and the Poincaré inequality both hold for the one-parameter selective model.*

Furthermore, if we can show

$$E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big)$$

is continuous and symmetric, then so is $p_\sigma(t,x,y)$.

Proposition 3.10. *The following function*

$$E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big)$$

is continuous in $(0,+\infty)\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$ *and symmetric in* x *and* y*; i.e.*

$$E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big)=E_y\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=x\Big).$$

Proof. In order to show the symmetry and continuity of

$$E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big),$$

we consider the pinned Markov processes, please refer to [Qian and Zheng, 2004]. We can use $P_x^{t,y}$ to denote the distribution of the pinned Markov process $X_s^{t,y}$, which starts with x and terminates at time t and state y. Then, for fixed t,y, $X_s^{t,y}$ is an inhomogeneous Markov process with the transition density function $p^{t,y}(r,x,s,z)$, which can be expressed as

$$p^{t,y}(r,x,s,z)=\frac{p(s-r,x,z)p(t-s,z,y)}{p(t-r,x,y)}.$$

Notice that $p(t,x,y)>0$; then $p^{t,y}(r,x,s,z)$ is well-defined everywhere.

Now

$$\begin{aligned}E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big)=&E_{P_x^{t,y}}\exp\Big\{-\int_0^t g(X_s)ds\Big\}\\=&E_{P_x^{t,y}}\sum_{k=0}^\infty\frac{(-1)^k}{k!}\Big(\int_0^t g(X_s)ds\Big)^k.\end{aligned}$$

Since

$$\Big|\frac{(-1)^k}{k!}\left(\int_0^t g(X_s)ds\right)^k\Big| \le \frac{(t||g||_\infty)^k}{k!},$$

where $g(x)=\frac{\sigma}{2}(1-(1+\theta)\varphi_2(x))+\frac{\sigma^2}{2}(\varphi_3-\varphi_2^2(x))$ and $||g||_\infty \le |\sigma|(\theta+1)+|\sigma|^2<\infty$, $\forall T>0$, and $(t,x,y)\in[0,T]\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$, we have

$$\Big|\frac{(-1)^k}{k!}\left(\int_0^t g(X_s)ds\right)^k\Big| \le \frac{T^k(|\sigma|(\theta+1)+|\sigma|^2)^k}{k!}.$$

By Fubini's theorem, we have

$$E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big)=\sum_{k=0}^\infty \frac{(-1)^k}{k!}E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k.$$

So, $\left|\frac{(-1)^k}{k!}E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k\right| \le \frac{T^k(|\sigma|(\theta+1)+|\sigma|^2)^k}{k!}$. Then by Weierstrass's M-test, we know $\sum_{k=0}^\infty \frac{(-1)^k}{k!}E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k$ is uniformly convergent in $(0,T]\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$. If we can show $E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k$ is continuous in $(0,T]\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$, and symmetric in x and y, then so is $E_x\Big(\exp\Big\{-\int_0^t g(X_s)ds\Big\}\Big|X_t=y\Big)$.

Now we only need to show that $E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k$ is continuous in $(0,T]\times\bar{\nabla}_\infty\times\bar{\nabla}_\infty$, and symmetric in x and y. To this end, we rewrite $\left(\int_0^t g(X_t)dt\right)^k$ as

$$k!\int\cdots\int_{0\le t_1<\cdots<t_k\le t}\prod_{i=1}^k g(X_{t_i})dt_i.$$

Then

$$\begin{aligned}&E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k\\ =&k!\int\cdots\int_{0\le t_1<\cdots<t_k\le t}\prod_{i=1}^k dt_i\int_{\bar{\nabla}_\infty}\cdots\int_{\bar{\nabla}_\infty}\prod_{i=1}^k g(x_i)p^{t,y}(t_{i-1},x_{i-1},t_i,x_i)\mathrm{PD}(\theta)(dx_i),\end{aligned}$$

where $t_0=0, x_0=x$. Recall that

$$p^{t,y}(t_{i-1},x_{i-1},t_i,x_i)=\frac{p(t_i-t_{i-1},x_{i-1},x_i)p(t-t_i,x_i,y)}{p(t-t_{i-1},x_{i-1},y)};$$

then

$$\begin{aligned}
\prod_{i=1}^{k} p^{t,y}(t_{i-1}, x_{i-1}, t_i, x_i) =& \prod_{i=1}^{k} p(t_i - t_{i-1}, x_{i-1}, x_i) \prod_{i=1}^{k} \frac{p(t - t_i, x_i, y)}{p(t - t_{i-1}, x_{i-1}, y)} \\
=& \prod_{i=1}^{k} p(t_i - t_{i-1}, x_{i-1}, x_i) \frac{p(t - t_k, x_k, y)}{p(t, x, y)} \\
=& \frac{\prod_{i=1}^{k+1} p(t_i - t_{i-1}, x_{i-1}, x_i)}{p(t, x, y)},
\end{aligned}$$

where $t_{k+1} = t, x_{k+1} = y$. Therefore,

$$\begin{aligned}
E_{P_x^{t,y}} \left(\int_0^t g(X_s) ds \right)^k =& k! \int \cdots \int_{0 \le t_1 < \cdots < t_k \le t} \prod_{i=1}^{k} dt_i \int_{\bar{\nabla}_\infty} \cdots \int_{\bar{\nabla}_\infty} \\
& \frac{\prod_{i=1}^{k} g(x_i) p(t_i - t_{i-1}, x_{i-1}, x_i) p(t - t_k, x_k, y)}{p(t, x, y)} \mathrm{PD}(\theta)(dx_i).
\end{aligned}$$

By the Feller property of the one-parameter neutral model, $E_{P_x^{t,y}} (\int_0^t g(X_s) ds)^k$ is continuous in $(0, \infty) \times \bar{\nabla}_\infty \times \bar{\nabla}_\infty$. Next, due to the symmetry of $p(t, x, y)$, we have

$$p(t_i - t_{i-1}, x_{i-1}, x_i) = p(t_i - t_{i-1}, x_i, x_{i-1}).$$

Then

$$\begin{aligned}
E_{P_x^{t,y}} \left(\int_0^t g(X_s) ds \right)^k =& k! \int \cdots \int_{0 \le t_1 < \cdots < t_k \le t} \prod_{i=1}^{k} dt_i \int_{\bar{\nabla}_\infty} \cdots \int_{\bar{\nabla}_\infty} \\
& \frac{\prod_{i=1}^{k} g(x_i) p(t_i - t_{i-1}, x_i, x_{i-1}) p(t - t_k, y, x_k)}{p(t, y, x)} \mathrm{PD}(\theta)(dx_i).
\end{aligned}$$

Consider the transformation $u_i = t - t_{k+1-i}, 1 \le i \le k+1$. Then $0 \le u_0 < u_1 < \cdots < u_k \le t$, and $u_0 = 0, u_{k+1} = t$. Therefore,

$$\begin{aligned}
E_{P_x^{t,y}} \left(\int_0^t g(X_s) ds \right)^k =& k! \int \cdots \int_{0 \le t_1 < \cdots < t_k \le t} \prod_{i=1}^{k} dt_i \int_{\bar{\nabla}_\infty} \cdots \int_{\bar{\nabla}_\infty} \\
& \frac{\prod_{i=1}^{k} g(x_i) p(t_i - t_{i-1}, x_i, x_{i-1}) p(t - t_k, y, x_k)}{p(t, y, x)} \mathrm{PD}(\theta)(dx_i) \\
=& k! \int \cdots \int_{0 \le u_1 < \cdots < u_k \le t} \prod_{i=1}^{k} du_i \int_{\bar{\nabla}_\infty} \cdots \int_{\bar{\nabla}_\infty}
\end{aligned}$$

$$\begin{aligned}&\frac{\prod_{i=1}^{k} g(x_i)p(u_i-u_{i-1},x_i,x_{i-1})p(t-u_k,y,x_k)}{p(t,y,x)}\mathrm{PD}(\theta)(dx_i)\\ =&E_{P_y^{t,x}}\left(\int_0^t g(X_s)ds\right)^k.\end{aligned}$$

Hence, $E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k$ is symmetric as well. □

Remark 3.6. *In one-parameter neutral model, $\bar{\nabla}_\infty - \nabla_\infty$ serves as an entrance law. Therefore, $\forall t > 0$, almost all the paths stay in ∇_∞. Since the one-parameter selective model is absolutely continuous with respect to one-parameter neutral model, $\bar{\nabla}_\infty - \nabla_\infty$ should also serve as an entrance boundary. Hence we can change the value of the density function $p_\sigma(t,x,y)$ when x or y is in $\bar{\nabla}_\infty - \nabla_\infty$. In this proof, $E_{P_x^{t,y}}(\int_0^t g(X_s)ds)^k, x, y \in \bar{\nabla}_\infty$ is chosen to be the continuous extension of*

$$E_{P_x^{t,y}}\left(\int_0^t g(X_s)ds\right)^k\Bigg|_{\nabla_\infty\times\nabla_\infty}.$$

Therefore, $p_\sigma(t,x,y)$ should be the continuous extension of $p_\sigma(t,x,y)|_{(x,y)\in\nabla_\infty\times\nabla_\infty}$. Moreover,

$$p_\sigma(t,x,y) = p_\sigma(t,y,x), \forall x, y \in \bar{\nabla}_\infty.$$

Although explicit representations of the transition density are still missing, $p_\sigma(t,x,y)$ can serve as an alternative representation of the transition density function of the one-parameter selective model. Much literature has tried to find the explicit representation of its transition density functions, which, however, still remains open. In [Arratia et al., 2003], a transition probability structure similar to that of the neutral models has been found. The explicit representation of the transition structure, however, is still unavailable; besides it only applies to finite dimensional models. Recently in [Steinrücken and Song, 2013] and [Steinrücken et al., 2013], the eigenfunction expansion was applied to the finite dimensional models; yet the explicit eigenfunction expansion is still unavailable. Furthermore, the legitimacy of the treatment of eigenfunction expansions in [Steinrücken and Song, 2013] and [Steinrücken et al., 2013] also need to be further verified. Thanks to [Wang, 2000], the generator G_σ indeed has a discrete spectrum and so do the models in [Steinrücken and Song, 2013] and [Steinrücken et al., 2013]. Therefore, the treatment of the eigenfunction expansion in [Steinrücken and Song, 2013] and [Steinrücken et al., 2013] is indeed valid. Such numerical simulation, unfortunately, seems not easy to be applied to the one-parameter selective model.

If we apply the method adopted in the Theorem 8.8 in [Chen, 2005], a uniform ergodicity inequality can be analogously obtained.

Theorem 3.8. *For the one-parameter selective model, $\exists K(\theta,\sigma) > 0$ such that the* *uniform ergodicity holds:*

$$\sup_{x\in\bar{\nabla}_\infty}\left\|P^\sigma(t,x,\cdot)-\pi_\sigma(\cdot)\right\|_{\mathrm{var}}\le K(\theta,\sigma)\exp\{-\mathrm{gap}(G_\sigma)t\},\quad \forall t\ge 0.$$

Proof. We are going to run the argument in theorem 8.8 in [Chen, 2005]. Since

$$P^\sigma(t,x,\cdot)=\int_{\bar{\nabla}_\infty}P^\sigma(t-s,z,\cdot)P^\sigma(s,x,dz),$$

and define $\mu^x(\cdot)=P_x(X_s^\sigma\in\cdot)$, we have

$$P^\sigma(t,x,\cdot)=\mu^x P_{t-s}^\sigma(\cdot).$$

Therefore,

$$\|P^\sigma(t,x,\cdot)-\pi(\cdot)\|_{\mathrm{var}}=\|\mu^x P_{t-s}^\sigma(\cdot)-\pi(\cdot)\|_{\mathrm{var}}.$$

By part (1) in theorem 8.8 in [Chen, 2005], we have, $\forall t\ge s$,

$$\begin{aligned}\|P^\sigma(t,x,\cdot)-\pi(\cdot)\|_{\mathrm{var}}\le&\left\|\frac{d\mu^x}{d\pi_\sigma}-1\right\|_2 e^{-(t-s)\mathrm{gap}(G_\sigma)}\\ &=\sqrt{\int_{\bar{\nabla}_\infty}p_\sigma(s,x,y)^2\pi_\sigma(dy)-1}e^{-(t-s)\mathrm{gap}(G_\sigma)}\end{aligned}$$

Therefore, for $t\ge\frac{1}{2}$, we have

$$\|P^\sigma(t,x,\cdot)-\pi(\cdot)\|_{\mathrm{var}}\le\sqrt{\int_{\bar{\nabla}_\infty}p_\sigma(1/2,x,y)^2\pi_\sigma(dy)-1}e^{\mathrm{gap}(G_\sigma)/2}\exp\{-\mathrm{gap}(G_\sigma)t\}.$$

By Proposition 3.9, there is a constant $K^{'}(\theta,\sigma)$ such that, $\forall x\in\bar{\nabla}_\infty$,

$$K^{'}(\theta,\sigma)\ge\sqrt{\int_{\bar{\nabla}_\infty}p_\sigma(1/2,x,y)^2\pi_\sigma(dy)-1}e^{\frac{1}{2}\mathrm{gap}(G_\sigma)}.$$

Then we have

$$\sup_{x\in\bar{\nabla}_\infty}\left\|P^\sigma(t,x,\cdot)-\pi_\sigma(\cdot)\right\|_{\mathrm{var}}\le K^{'}(\theta,\sigma)\exp\{-\mathrm{gap}(G_\sigma)t\},\forall t\ge\frac{1}{2}.$$

Moreover,

$$\sup_{x\in\bar{\nabla}_\infty}\left\|P^\sigma(t,x,\cdot)-\pi_\sigma(\cdot)\right\|_{\mathrm{var}}\le 1,\quad \forall t\ge 0.$$

Thus $\forall t \in [0, \frac{1}{2}]$, if we choose $K^{''}(\theta, \sigma)$ such that

$$K^{''}(\theta,\sigma)e^{-\mathrm{gap}(G_\sigma)/2} \geq 1,$$

then $\forall t \in [0, \frac{1}{2}]$,

$$\sup_{x \in \bar{\nabla}_\infty} \left\| P^\sigma(t,x,\cdot) - \pi_\sigma(\cdot) \right\|_{\mathrm{var}} \leq 1 \leq K^{''}(\theta,\sigma)e^{-\mathrm{gap}(G_\sigma)/2} \leq K^{''}(\theta,\sigma)\exp\{-\mathrm{gap}(G_\sigma)t\}.$$

Therefore, choosing $K(\theta, \sigma) = \max\{K^{'}(\theta,\sigma), K^{''}(\theta,\sigma)\}$, we have

$$\sup_{x \in \bar{\nabla}_\infty} \left\| P^\sigma(t,x,\cdot) - \pi_\sigma(\cdot) \right\|_{\mathrm{var}} \leq K(\theta,\sigma)\exp\{-\mathrm{gap}(G_\sigma)t\}.$$

□

Remark 3.7. *(i). In [Ethier and Kurtz, 1998], the following strong ergodicity is obtained:*

$$\lim_{t \to \infty} \left\| P^\sigma(t,x,\cdot) - \pi_\sigma(\cdot) \right\|_{\mathrm{var}} = 0, \text{ for each } x \in \bar{\nabla}_\infty.$$

Therefore, theorem 3.8 provides a sharp bound for the exponential convergence rate even though the constant $K(\theta, \sigma)$ and $\mathrm{gap}(G_\sigma)$ can not be explicitly evaluated.

(ii). The spectral gap $\mathrm{gap}(G_\sigma)$ could be another important quantity worthy of study.

All the above conclusions associated with the one-parameter selective model have two-parameter extensions concerned with the perturbed Dirichlet form (refer to [Feng et al., 2011]).

Chapter 4

Asymptotic Theory for the One-paramter Neutral Model and the Two-parameter Model

As we know, the one-parameter neutral model and the two-parameter model are both reversible diffusion processes. Their ergodic theorems have already been established in Chapter 3. We also know that their stationary distributions are $\mathrm{PD}(\theta)$ and $\mathrm{PD}(\theta, \alpha)$ respectively. Generally, once they start with their stationary distributions, all their transient distributions will be the same as their stationary distributions. Hence, the properties of $\mathrm{PD}(\theta)$ and $\mathrm{PD}(\theta, \alpha)$ are actually those of the one-parameter neutral model and the two-parameter model starting with their stationary distributions respectively. If, however, we want to consider the associated non-stationary model, then the effect of time is supposed to be taken into account.

A rather thorough research on $\mathrm{PD}(\theta)$ and $\mathrm{PD}(\theta, \alpha)$ has already been carried out in [Dawson and Feng, 2006], [Feng, 2007b], [Feng, 2007a], [Feng, 2009] and [Feng and Gao, 2010]. Thus, we will not focus on the asymptotic theories for $\mathrm{PD}(\theta)$ and $\mathrm{PD}(\theta, \alpha)$ in this chapter; on the contrary, the associated non-stationary model will be considered. Of course, we need to pay particular attention to the role of time. One can see that the asymptotic behaviour of stationary distribution, as mutation rate increases, is analogous to that of any transient distribution for fixed time $t > 0$. The small-time behaviour, therefore, ought to be able to explain the microscopic structures of those models. Even though the Wentzell-Freidlin type sample path LDP for the neutral Fleming-Viot process with the parent independent mutation was established in [Dawson and Feng, 1998], [Dawson and Feng, 2001] and [Feng and Xiong, 2002], the small-time behaviours of the one-parameter neutral model and the two-parameter model are untouched yet. Simply because the atomic projection is quite irregular and discontinuous, and the contraction principle of LDPs is not applicable here. So, to establish the small-time LDP, we need to start everything from scratch.

In this chapter, the LDPs for the one-parameter neutral model at fixed time are considered and similar result is also obtained for the two-parameter model. To fully understand the mechanism behind this, we need to understand the small-time behaviours. Instead of considering the generic sample path LDP, we consider the small-time LDP for the partition structures associated with the one-parameter neutral model and the two-parameter model. Thanks to Kingman's one-to-one correspondence, studying the behaviours of the associated partition structures could be another way of understanding the evolution of allele frequencies of these models.

4.1 Role of Time

In population genetics, one important goal is to understand the interactions between various evolutionary forces. As Gillespie argued in [Gillespie, 2004], the time required for evolutionary forces to exert appreciable effects to the genetic configuration can quantify these forces. It is also stated in [Gillespie, 2004] that the time scale of random sampling is $2N\log 2$, and the time scale of mutation is $\frac{4N}{\theta}$, where $\theta = 4Nu$, and u is the individual mutation rate. Therefore, the time scale of random sampling in the associated diffusion model is $\log 2$ and the time scale of mutation is $\frac{2}{\theta}$. As $N \to +\infty$, we have $\theta \to +\infty$; thus $\frac{2}{\theta} << \log 2$. Mutations will introduce many new types immediately; the allele spectrum, therefore, should be continuous. If, however, $\frac{2}{\theta} >> \log 2$, then the allele spectrum should be discrete. To verify the above intuition, we should look at the asymptotic distribution of $X_{t(\theta)}$ for different time scales $t(\theta)$.

4.2 Several Weak Laws of Large Numbers

The limiting distributions of the one-parameter neutral model and the two-parameter model under different time scales are obtained. The results are stated in the following theorem.

Theorem 4.1. *Let X_t be the one-parameter neutral model, starting at $x \in \bar{\nabla}_\infty$. As $\theta \to +\infty$, we have*

$$X_{t(\theta)} \to \begin{cases} \delta_{(0,0,\cdots)}, & \text{if } \lim_{\theta\to+\infty} \theta t(\theta) = +\infty \\ \delta_{e^{-\frac{c}{2}}x}, & \text{if } \lim_{\theta\to+\infty} \theta t(\theta) = c > 0 \\ \delta_x, & \text{if } \lim_{\theta\to+\infty} \theta t(\theta) = 0. \end{cases}$$

Remark 4.1.
- *As the time scale increases, the effect of mutations is enhanced gradually. Because the allele spectrum start to show some properties of continuous spectra. When $t(\theta) >> \frac{1}{\theta}$, mutations completely take over. Thus the allele spectrum is a purely continuous spectrum. In fact, this theorem unveils the fine structure of the interactions between mutations and random sampling.*

- *For fixed $t > 0$, $\lim_{\theta\to\infty} X_t = \delta_{(0,0,\cdots)}$, which can be seen from this theorem. This result can also be obtained from Theorem 3.3, thanks to a fact*

$$\lim_{\theta\to+\infty} \mathrm{PD}(\theta) = \delta_{(0,0,\cdots)}. \tag{4.2.1}$$

Now we are going to prove Theorem 4.1. To this end, we need the following lemmas.

Lemma 4.1. *For any given partition $\omega \in \tilde{\mathcal{M}}$, we have*

$$|\psi_\omega^\theta| \le M(\omega),$$

where $M(\omega)$ is positive constant independent of θ. Thus, we have

$$\lim_{\theta\to+\infty} \langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta = 0, \quad \forall \xi \le \omega,$$

and

$$\lim_{\theta\to+\infty} \psi_\omega^\theta = \varphi_\omega.$$

Proof. We can show the first statement by mathematical induction. Apparently, 1 and ψ_2^θ are bounded by 1 and 2. Thus we assume that for all $\xi < \omega$, we have

$$|\psi_\xi^\theta| \le M(\xi).$$

Then consider ψ_ω^θ. Due to (3.2.1), we have

$$|\psi_\omega^\theta| \le \varphi_\omega + \sum_{\xi<\omega} |\langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta| \cdot |\psi_\xi^\theta|$$

by the Cauchy-Schwarz inequality and $\varphi_\omega \le 1$, we have

$$\begin{aligned} &\le 1 + \sum_{\xi<\omega} \|\varphi_\omega\|_{2,\theta} \cdot \|\psi_\xi^\theta\|_{2,\theta} \cdot |\psi_\xi^\theta| \\ &\le 1 + \sum_{\xi<\omega} M(\xi)^2. \end{aligned}$$

Therefore, each ψ_ω^θ has a bound independent of θ. Next, by the Cauchy-Schwarz inequality, we have

$$0 \le |\langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta| \le \|\varphi_\omega\|_{2,\theta} \cdot \|\psi_\xi^\theta\|_{2,\theta} \le M(\xi)\|\varphi_\omega\|_{2,\theta} \to 0.$$

The limit is due to (4.2.1). Then we have $\lim_{\theta\to+\infty} \langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta = 0$, for $\xi \le \omega$. Thus it is easy to see that $\lim_{\theta\to+\infty} \psi_\omega^\theta = \varphi_\omega$ because of (3.2.1). $\square$

Lemma 4.2. *For any partition $\omega \in \tilde{\mathcal{M}}$, we have the following*

$$\int_{\bar{\nabla}_\infty} Q_m(x,y)\psi_\omega^\theta(y)\mathrm{PD}(\theta)(dy) = \begin{cases} \psi_\omega^\theta(x), & |\omega| = m \\ 0, & \textit{otherwise} \end{cases}.$$

Proof. Notice that

$$Q_m(x,y) = \sum_{|\eta|=m} \chi_\eta^\theta(x)\chi_\eta^\theta(y),$$

where $\chi_\eta^\theta(x)$ can be chosen as the normalized ψ_η^θ, i.e. $\chi_\eta^\theta(x) = \frac{\psi_\eta^\theta(x)}{\|\psi_\eta^\theta\|_{2,\theta}}$. Thus, we have

$$\begin{aligned}\int_{\bar{\nabla}_\infty} Q_m(x,y)\psi_\omega^\theta(y)\mathrm{PD}(\theta)(dy) =& \|\psi_\omega^\theta\|_{2,\theta} \int_{\bar{\nabla}_\infty} Q_m(x,y)\chi_\omega^\theta(y)\mathrm{PD}(\theta)(dy) \\ =& \|\psi_\omega^\theta\|_{2,\theta}\delta_{m,|\omega|}\chi_\omega^\theta(x) = \psi_\omega^\theta(x)\delta_{m,|\omega|}.\end{aligned}$$

□

Lemma 4.3. *For any given partition $\omega \in \tilde{\mathcal{M}}$, we have*

$$\begin{aligned}\lim_{\theta\to+\infty} E_x\Big(\varphi_\omega(X_{t(\theta)})\Big) &= \lim_{\theta\to+\infty}\int_{\bar{\nabla}_\infty} \varphi_\omega(y)p(t(\theta),x,y)\mathrm{PD}(\theta)(dy) \\ &= \begin{cases} 0, & \textit{if } \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ \varphi_\omega(e^{-\frac{c}{2}}x), & \textit{if } \lim_{\theta\to+\infty}\theta t(\theta) = c \\ \varphi_\omega(x), & \textit{if } \lim_{\theta\to+\infty}\theta t(\theta) = 0 \end{cases}.\end{aligned}$$

Proof. Since the transition density function of the one-parameter neutral model is

$$p(t,x,y) = 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t}Q_m(x,y),$$

we have

$$\begin{aligned}&\int_{\bar{\nabla}_\infty} \varphi_\omega(y)p(t(\theta),x,y)\mathrm{PD}(\theta)(dy) \\ =&\int_{\bar{\nabla}_\infty} \varphi_\omega(y)\mathrm{PD}(\theta)(dy) + \sum_{m=2}^{\infty} e^{-\lambda_m t(\theta)}\int_{\bar{\nabla}_\infty} \varphi_\omega(y)Q_m(x,y)\mathrm{PD}(\theta)(dy).\end{aligned}$$

Then

$$\varphi_\omega = \psi_\omega^\theta + \sum_{\xi<\omega}\langle\varphi_\omega,\psi_\xi^\theta\rangle_\theta\psi_\xi^\theta;$$

hence, by Lemma 4.2, we have

$$\int_{\bar{\nabla}_\infty} \varphi_\omega(y) Q_m(x,y)\mathrm{PD}(\theta)(dy) = \psi_\omega^\theta(x)\delta_{m,|\omega|} + \sum_{\xi<\omega} \langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta \psi_\xi^\theta(x)\delta_{m,|\xi|}.$$

Therefore,

$$\begin{aligned}\int_{\bar{\nabla}_\infty} \varphi_\omega p(t(\theta),x,y)\mathrm{PD}(\theta)(dy) &= \int_{\bar{\nabla}_\infty} \varphi_\omega(y)\mathrm{PD}(\theta)(dy) + \sum_{m=2}^{\infty} e^{-\lambda_m t(\theta)} \psi_\omega^\theta(x)\delta_{m,|\omega|} \\ &\quad + \sum_{\xi<\omega}\sum_{m=2}^{\infty} e^{-\lambda_m t(\theta)} \langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta \psi_\xi^\theta(x)\delta_{m,|\xi|} \\ &= \int_{\bar{\nabla}_\infty} \varphi_\omega(y)\mathrm{PD}(\theta)(y) + e^{-\lambda_{|\omega|} t(\theta)} \psi_\omega^\theta(x) \\ &\quad + \sum_{\xi<\omega} e^{-\lambda_{|\xi|} t(\theta)} \langle \varphi_\omega, \psi_\xi^\theta \rangle_\theta \psi_\xi^\theta(x).\end{aligned}$$

By Lemma 4.1, we have

$$\lim_{\theta\to+\infty} \int_{\bar{\nabla}_\infty} \varphi_\omega(x) p(t(\theta),x,y)\mathrm{PD}(\theta)(dy) = \begin{cases} 0, & \text{if } \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ \varphi_\omega(e^{-\frac{c}{2}}x), & \text{if } \lim_{\theta\to+\infty}\theta t(\theta) = c \\ \varphi_\omega(x), & \text{if } \lim_{\theta\to+\infty}\theta t(\theta) = 0 \end{cases}.$$

□

Now we are ready to show Theorem 4.1.

Proof. We only need to show that $\forall f \in C(\bar{\nabla}_\infty)$,

$$\lim_{\theta\to+\infty} Ef(X_{t(\theta)}) = \begin{cases} f(0), & \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ f(e^{-\frac{c}{2}}x), & \lim_{\theta\to+\infty}\theta t(\theta) = c \\ f(x), & \lim_{\theta\to+\infty}\theta t(\theta) = 0. \end{cases}$$

Notice that $\mathcal{P}$ is dense in $C(\bar{\nabla}_\infty)$. Therefore, $\forall \epsilon > 0$, there is $p \in \mathcal{P}$, such that

$$||f - p||_\infty < \epsilon.$$

By Lemma 4.3, we know

$$\lim_{\theta\to+\infty} E_x p(X_{t(\theta)}) = \begin{cases} p(0), & \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ p(e^{-\frac{c}{2}}x), & \lim_{\theta\to+\infty}\theta t(\theta) = c \\ p(x), & \lim_{\theta\to+\infty}\theta t(\theta) = 0. \end{cases}$$

Moreover,

$$|Ef(X_{t(\theta)}) - f(x)| \leq 2||f-p||_\infty + |Ep(X_{t(\theta)}) - p(x)| \leq 2\epsilon + |Ep(X_{t(\theta)}) - p(x)|.$$

If we replace x in the above by $0, e^{-\frac{c}{2}x}, x$, and let $n \to +\infty$, and then let $\epsilon \to 0$, we have

$$\lim_{\theta\to+\infty} E_x f(X_{t(\theta)}) = \begin{cases} f(0), & \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ f(e^{-\frac{c}{2}}x), & \lim_{\theta\to+\infty}\theta t(\theta) = c \\ f(x), & \lim_{\theta\to+\infty}\theta t(\theta) = 0. \end{cases}$$

Theorem 4.1 is thus proved. □

Theorem 4.2. *Let $X_t^{\theta,\alpha}$ be the two-parameter model, then as $\theta \to +\infty$, we have*

$$X_{t(\theta)}^{\theta,\alpha} \to \begin{cases} \delta_{(0,0,\cdots)}, & \textit{if } \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ \delta_{e^{-\frac{c}{2}}x}, & \textit{if } \lim_{\theta\to+\infty}\theta t(\theta) = c > 0 \\ \delta_x, & \textit{if } \lim_{\theta\to+\infty}\theta t(\theta) = 0. \end{cases}$$

Remark 4.2. *One can easily see that , for large θ, α has no influence at all. Moreover,*

$$\lim_{\theta\to+\infty} \mathrm{PD}(\theta,\alpha) = \delta_{(0,0,\cdots)}. \tag{4.2.2}$$

But the influence of α can be observed in the LDP rate functions for the corresponding Dirichlet process. Please refer to [Feng, 2007b].

Theorem 4.2 can also be shown analogously by the following three lemmas.

Lemma 4.4. *For any give partition $\omega \in \tilde{\mathcal{M}}$, we have*

$$|\psi_\omega^{\theta,\alpha}| \leq M(\omega),$$

where $M(\omega)$ is positive and independent of θ, α. Thus, we have

$$\lim_{\theta\to+\infty} \langle \varphi_\omega, \psi_\xi^{\theta,\alpha}\rangle_{\theta,\alpha} = 0, \quad \forall \xi \leq \omega,$$

and

$$\lim_{\theta\to+\infty} \psi_\omega^{\theta,\alpha} = \varphi_\omega.$$

Proof. Notice that $\psi_1^{\theta,\alpha} = 1$ and $\psi_2^{\theta,\alpha} = \varphi_2 - \frac{1-\alpha}{1+\theta}$, which are bounded respectively by 1 and 2. Then we can use mathematical induction to show this lemma. □

Lemma 4.5. *For any partition $\omega \in \tilde{\mathcal{M}}$, we have the following*

$$\int_{\bar{\nabla}_\infty} Q_m^{\theta,\alpha}(x,y)\psi_\omega^{\theta,\alpha}(y)\mathrm{PD}(\theta,\alpha)(dy) = \begin{cases} \psi_\omega^{\theta,\alpha}(x), & |\omega| = m \\ 0, & \textit{otherwise}\ . \end{cases}$$

Lemma 4.6. *For any given $\omega \in \tilde{\mathcal{M}}$, we have*

$$\lim_{\theta\to+\infty} E_x\Big(\varphi_\omega(X^{\theta,\alpha}_{t(\theta)})\Big) = \lim_{\theta\to+\infty}\int_{\bar{\nabla}_\infty}\varphi_\omega(y)p^{\theta,\alpha}(t(\theta),x,y)\mathrm{PD}(\theta,\alpha)(dy)$$
$$= \begin{cases} 0, & \text{if } \lim_{\theta\to+\infty}\theta t(\theta) = +\infty \\ \varphi_\omega(e^{-\frac{c}{2}}x), & \text{if } \lim_{\theta\to+\infty}\theta t(\theta) = c \\ \varphi_\omega(x), & \text{if } \lim_{\theta\to+\infty}\theta t(\theta) = 0 \end{cases}.$$

4.3 LDP for the One-parameter Neutral Model and the Two-parameter Model

Theorem 4.1 and Theorem 4.2 tell us that both the one-parameter neutral model and the two-parameter model approach $(0,0,\cdots)$ instantly. This phenomenon seems to be very odd at first glance; but it makes sense intuitively. As mutations get bigger, the acting time scale of mutations will be greatly decreased. Thus, large amount of new allele types enter the system immediately. Then the allele frequency will be more evenly spread, thereby much more closer to $(0,0,\cdots)$. It also stays close to $(0,0,\cdots)$ ever since. Thus, the system will move to $(0,0,\cdots)$ immediately and stay there.

The best way to verify the above heuristic explanation is to consider the small-time behaviour of $X_{t(\theta)}$ for various time scales $t(\theta)$. To uncover even finer structures of the interactions between evolutionary forces, we need to consider the LDP for $X_{t(\theta)}$. Of particular interest is the time scale $t(\theta) >> \frac{1}{\theta}$. Because even though the weak limits of $X_{t(\theta)}$ at this region are the same, the LDPs for the corresponding partition structures indicate that there is a critical time scale $t(\theta) = \frac{2\log\theta}{\theta}$.

In this section, the LDP for X_t is provided. The LDP for the associated partition structures are also obtained. Moreover, each of these results has a two-parameter generalization.

4.3.1 LDP for Transient Distributions

Since the one-parameter neutral model and the two-parameter model reach their equilibrium points instantly, the LDP for their transient distributions should tend to resemble that of their stationary distribution. This fact can be shown by the LDP for X_t and $X_t^{\theta,\alpha}$ in the next two theorems.

Theorem 4.3. *For fixed $t > 0$, X_t satisfies an LDP with the speed θ and rate function $I(x)$, where*

$$I(x) = \begin{cases} \log\frac{1}{1-\sum_{i=1}^\infty x_i}, & x\in\bar{\nabla}_\infty, \sum_{i=1}^\infty x_i < 1 \\ +\infty, & x\in\bar{\nabla}_\infty, \sum_{i=1}^\infty x_i = 1. \end{cases}$$

Proof. By the LDP for PD(θ), we know PD(θ) satisfies the LDP with rate $\log\theta$ and rate function $I(x)$. In the following, we are going to show upper bound and lower bound respectively.

[Proof of Upper Bound]:

Let F be a closed subset of $\bar{\nabla}_\infty$. Then

$$P(t,x,F)=\sum_{n=2}^{+\infty} d_n^\theta(t)\nu_n^\theta(x,F)+(d_0^\theta(t)+d_1^\theta(t))\mathrm{PD}(\theta)(F).$$

Notice that when $\theta>1$,

$$\nu_n^\theta(x,F)=\int_F p_n(x,y)\mathrm{PD}(\theta)(dy)\le\max_{|\eta|=n}\frac{1}{\int p_\eta d\mathrm{PD}(\theta)}\mathrm{PD}(\theta)(F)\le\theta_{(n)}\mathrm{PD}(\theta)(F).$$

So

$$\begin{aligned}\frac{1}{\theta}\log P(t,x,F)\le&\frac{1}{\theta}\log\Big[\sum_{n=2}^{\infty} d_n^\theta(t)\theta_{(n)}+d_0^\theta(t)+d_1^\theta(t)\Big]\mathrm{PD}(\theta)(F)\\ \le&\frac{1}{\theta}\log\Big[\sum_{n=2}^{\infty} d_n^\theta(t)\theta_{(n)}+d_0^\theta(t)+d_1^\theta(t)\Big]+\frac{1}{\theta}\log\mathrm{PD}(\theta)(F).\end{aligned}$$

We claim that, as $\theta\to+\infty$,

$$\frac{1}{\theta}\log\Big[\sum_{n=2}^{\infty} d_n^\theta(t)\theta_{(n)}+d_0^\theta(t)+d_1^\theta(t)\Big]\to 0. \tag{4.3.1}$$

Then

$$\limsup_{\theta\to+\infty}\frac{1}{\theta}\log P(t,x,F)\le\limsup_{\theta\to+\infty}\frac{1}{\theta}\log\mathrm{PD}(\theta)(F)=-\inf_{x\in F}I(x).$$

For the claim (4.3.1), we have

$$d_0^\theta(t)+d_1^\theta(t)+\sum_{n=2}^{\infty} d_n^\theta(t)\theta_{(n)}\le 1+\sum_{m=2}^{\infty}e^{-\lambda_m t}\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{(m-1)}\theta_{(n)}.$$

Since $(\theta+n)_{(m-1)}\le(\theta+2m-1)^{m-1},\theta_{(n)}\le(\theta+m)^m\le(2m+\theta-1)^m$, we have

$$\begin{aligned}\Big|\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{(m-1)}\theta_{(n)}\Big|\le&(2m+\theta-1)^{2m}\sum_{n=0}^{m}\frac{1}{n!(m-n)!}\\ \le&(2m+\theta-1)^{2m}\frac{2^m}{m!}\le 2(2m+\theta-1)^{2m}.\end{aligned}$$

Thus,

$$
\begin{aligned}
&\sum_{m=2}^{\infty} e^{-\lambda_m t} \sum_{n=0}^{m} \frac{2m+\theta-1}{m!} \binom{m}{n} (n+\theta)_{(m-1)} \theta_{(n)} \\
\leq & 2 \sum_{m=2}^{\infty} e^{-\frac{m(m+\theta-1)}{2} t} (2m+\theta-1)^{2m} = 2 \sum_{m=2}^{\infty} \left\{ e^{-\left[\frac{m+\theta-1}{2} t - 2\log(2m+\theta-1)\right]} \right\}^m \\
= & 2 \sum_{m=2}^{\infty} \left\{ \frac{1}{e^{\frac{m+\theta-1}{2} t - 2\log(2m+\theta-1)}} \right\}^m .
\end{aligned}
$$

Define $h_\theta^t(x) = \frac{x+\theta-1}{2} t - 2\log(2x+\theta-1)$. Then, for fixed t,

$$
h_\theta^t(x)' = \frac{t}{2} - \frac{4}{2x+\theta-1} \geq \frac{t}{2} - \frac{2}{x} > 0, \forall \theta > 1, \text{ and } x > \frac{4}{t}.
$$

We have, $\forall m > \frac{4}{t}, \theta > 1$,

$$
h_\theta^t(m) \geq h_\theta^t(\frac{4}{t}) = \frac{4+\theta t - t}{2} - 2\log(\frac{8}{t} + \theta - 1).
$$

Define $K_t(\theta) = \frac{4+\theta t - t}{2} - 2\log(\frac{8}{t} + \theta - 1), \theta > 1$. Then

$$
K_t(\theta)' = \frac{t}{2} - \frac{2}{\frac{8}{t} + \theta - 1} > \frac{t}{2} - \frac{t}{4} = \frac{t}{4} > 0.
$$

Therefore, for fixed t, $K_t(\theta)$ is increasing on $[1, \infty)$. Thus $\exists \theta_0 > 1$, such that $K_t(\theta) \geq K_t(\theta_0) > 0$. Then $\forall \theta \geq \theta_0 > 1, m > \frac{4}{t}$, we have

$$
h_\theta^t(m) \geq h_\theta^t(\frac{4}{t}) \geq K_t(\theta_0) > 0.
$$

Therefore

$$
\frac{1}{e^{\frac{m+\theta-1}{2} t - 2\log(2m+\theta-1)}} = \frac{1}{e^{h_\theta^t(m)}} \leq \frac{1}{e^{K_t(\theta_0)}} < 1.
$$

By Weierstrass M-test, for fixed t,

$$
\sum_{m=2}^{\infty} \left\{ \frac{1}{e^{\frac{m+\theta-1}{2} t - 2\log(2m+\theta-1)}} \right\}^m
$$

is uniformly convergent on $[\theta_0, +\infty)$. Then

$$
\lim_{\theta \to +\infty} \sum_{m=2}^{\infty} e^{-\frac{m(m+\theta-1)}{2} t} (2m+\theta-1)^{2m} = \sum_{m=2}^{\infty} \lim_{\theta \to +\infty} e^{-\frac{m(m+\theta-1)}{2} t} (2m+\theta-1)^{2m} = 0;
$$

thus

$$\lim_{\theta\to+\infty}\sum_{m=2}^{\infty}e^{-\lambda_m t}\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{(m-1)}\theta_{(n)}=0.$$

Hence

$$\begin{aligned}0 &\le \limsup_{\theta\to+\infty}\frac{1}{\theta}\log\Big[\sum_{n=2}^{\infty}d_n^{\theta}(t)\theta_{(n)}+d_0^{\theta}(t)+d_1^{\theta}(t)\Big]\\ &\le \limsup_{\theta\to+\infty}\frac{1}{\theta}\log\Big[1+\sum_{m=2}^{\infty}e^{-\lambda_m t}\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{(m-1)}\theta_{(n)}\Big]\\ &=\lim_{\theta\to+\infty}\frac{1}{\theta}\Big[\sum_{m=2}^{\infty}e^{-\lambda_m t}\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{(m-1)}.\theta_{(n)}\Big]=0\end{aligned}$$

The claim (4.3.1) is thus proved.

[Proof of Lower Bound]:

Suppose that O is an open subset of $\bar{\nabla}_{\infty}$. Then

$$\begin{aligned}\frac{1}{\theta}\log P(t,x,O)=&\frac{1}{\theta}\log\Big[(d_0^{\theta}(t)+d_1^{\theta}(t))\mathrm{PD}(\theta)(O)+\sum_{n=2}^{\infty}d_n^{\theta}(t)\nu_n^{\theta}(x,O)\Big]\\ \ge&\frac{1}{\theta}\log\Big[(d_0^{\theta}(t)+d_1^{\theta}(t))\mathrm{PD}(\theta)(O)\Big].\end{aligned}$$

But

$$\frac{1}{\theta}\log\Big[(d_0^{\theta}(t)+d_1^{\theta}(t))\mathrm{PD}(O)\Big]=\frac{1}{\theta}\log(d_0^{\theta}(t)+d_1^{\theta}(t))+\frac{1}{\theta}\log\mathrm{PD}(\theta)(O);$$

and

$$\begin{aligned}\frac{1}{\theta}\log(d_0^{\theta}(t)+d_1^{\theta}(t))=&\frac{1}{\theta}\log[1-(1-d_0^{\theta}(t)-d_1^{\theta}(t))]\\ \sim&\frac{1-d_0^{\theta}(t)-d_1^{\theta}(t)}{\theta}\to 0,\ \text{ as }\theta\to+\infty;\end{aligned}$$

for $|1-d_0^{\theta}(t)-d_1^{\theta}(t)|\le\frac{(\theta+2)(\theta+3)}{2}e^{-(\theta+1)t}\to 0$(refer to [Tavaré, 1984]). Then

$$\liminf_{\theta\to+\infty}\frac{1}{\theta}\log(d_0^{\theta}(t)+d_1^{\theta}(t))\mathrm{PD}(\theta)(O)=\liminf_{\theta\to+\infty}\frac{1}{\theta}\log\mathrm{PD}(\theta)(O)=-\inf_{x\in O}I(x).$$

Thus,

$$\liminf_{\theta\to+\infty}\frac{1}{\theta}\log P(t,x,O)\ge-\inf_{x\in O}I(x).$$

□

Furthermore, $\nu_n^\theta(x,dy) = p_n(x,y)\mathrm{PD}(\theta)(dy)$, $\nu_n^\theta(x,dy)$ also have an LDP similar to that for $\mathrm{PD}(\theta)$.

Proposition 4.1. *For $n \geq 2$, $\nu_n^\theta(x,dy)$ satisfies an LDP with speed θ and rate function $I(y)$, where*

$$I(y) = \begin{cases} \log \frac{1}{1-\sum_{i=1}^\infty y_i}, & \sum_{i=1}^\infty y_i < 1 \\ +\infty, & \sum_{i=1}^\infty y_i = 1. \end{cases}$$

Proof. Suppose that the topology of $\bar{\nabla}_\infty$ is induced by the metric

$$d(p,q) = \sum_{i=1}^\infty \frac{|p_i - q_i|}{2^i},$$

under which $\bar{\nabla}_\infty$ is a compact space. By the Puhalskii theorem (refer to [Dawson and Feng, 2006] or [Puhalskii, 1991]), we only need to show

$$\lim_{\delta\to 0}\liminf_{\theta\to+\infty}\frac{1}{\theta}\log\nu_n^\theta(x,B_\delta(y)) = \lim_{\delta\to 0}\limsup_{\theta\to+\infty}\frac{1}{\theta}\log\nu_n^\theta(x,\bar{B}_\delta(y)) = -I(y),$$

where $B_\delta(y), \bar{B}_\delta(y)$ are open balls and closed balls centered at y with radius δ respectively. By the LDP for $\mathrm{PD}(\theta)$ in [Dawson and Feng, 2006], we have

$$\lim_{\delta\to 0}\liminf_{\theta\to+\infty}\frac{1}{\theta}\log\mathrm{PD}(\theta)(B_\delta(y)) = \lim_{\delta\to 0}\limsup_{\theta\to+\infty}\frac{1}{\theta}\log\mathrm{PD}(\theta)(\bar{B}_\delta(y)) = -I(y).$$

Since

$$\lim_{\delta\to 0}\liminf_{\theta\to+\infty}\frac{1}{\theta}\log\nu_n^\theta(x,B_\delta(y)) \leq \lim_{\delta\to 0}\limsup_{\theta\to+\infty}\frac{1}{\theta}\log\nu_n^\theta(x,\bar{B}_\delta(y)),$$

we only need to show

$$\lim_{\delta\to 0}\liminf_{\theta\to+\infty}\frac{1}{\theta}\log\nu_n^\theta(x,B_\delta(y)) \geq -I(y), \tag{4.3.2}$$

and

$$\lim_{\delta\to 0}\limsup_{\theta\to+\infty}\frac{1}{\theta}\log\nu_n^\theta(x,\bar{B}_\delta(y)) \leq -I(y). \tag{4.3.3}$$

First, we are going to show the first inequality (4.3.2). Note that

$$\inf_{|\eta|=n}\inf_{z\in B_\delta(y)} p_\eta(z) > 0, \text{ for } p_\eta(z) > 0 \text{ and } \bar{\nabla}_\infty \text{ is compact.}$$

Thus

$$\nu_n^\theta(x,B_\delta(y)) = \int_{B_\delta(y)} p_n(x,z)\mathrm{PD}(\theta)(dz)$$

$$\geq \inf_{|\eta|=n} \inf_{z\in B_\delta(y)} p_\eta(z) \min_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta)} \mathrm{PD}(\theta)(B_\delta(y)),$$

and

$$\begin{aligned}
&\frac{1}{\theta}\log \nu_n^\theta(x, B_\delta(y)) \\
&\geq \frac{1}{\theta}\log \inf_{|\eta|=n} \inf_{z\in B_\delta(y)} p_\eta(z) + \frac{1}{\theta}\log \min_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta)} + \frac{1}{\theta}\log \mathrm{PD}(\theta)(B_\delta(y)).
\end{aligned}$$

Obviously, $\frac{1}{\theta}\log \inf_{|\eta|=n} \inf_{z\in B_\delta(y)} p_\eta(z) \to 0$, as $\theta \to +\infty$; and we claim that

$$\lim_{\theta\to+\infty} \frac{1}{\theta}\log \min_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta)} = 0. \tag{4.3.4}$$

Therefore, we have

$$\lim_{\delta\to 0} \liminf_{\theta\to+\infty} \frac{1}{\theta}\log \nu_n^\theta(x, B_\delta(y)) \geq \lim_{\delta\to 0} \liminf_{\theta\to+\infty} \frac{1}{\theta}\log \mathrm{PD}(\theta)(B_\delta(y)) = -I(y).$$

The claim is indeed true. By the Ewens sampling formula, we have, for $\theta > 1$,

$$\frac{1}{\int p_\eta d\mathrm{PD}(\theta)} = \frac{1}{\frac{n!}{n_1!\cdots n_{l(\eta)}!\alpha_1!\cdots\alpha_n!}\frac{\theta^{l(\eta)}}{\theta_{(n)}}} \geq \frac{1}{n!}\frac{\theta_{(n)}}{\theta} > \frac{1}{n!}\frac{1}{\theta},$$

and

$$\frac{1}{\int p_\eta d\mathrm{PD}(\theta)} \leq \theta_{(n)}. \tag{4.3.5}$$

Hence the claim (4.3.4) follows immediately. Second, we are going to show the second inequality (4.3.3). Since

$$\begin{aligned}
\nu_n^\theta(x, \bar{B}_\delta(y)) = \int_{\bar{B}_\delta(y)} p_n(x,z)\mathrm{PD}(\theta)(dz) &\leq \sum_{|\eta|=n} p_\eta(x) \frac{\mathrm{PD}(\theta)(\bar{B}_\delta(y))}{\int p_\eta d\mathrm{PD}(\theta)} \\
&\leq \max_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta)} \mathrm{PD}(\theta)(\bar{B}_\delta(y)),
\end{aligned}$$

we have

$$\begin{aligned}
\lim_{\delta\to 0} \limsup_{\theta\to+\infty} \frac{1}{\theta}\log \nu_n^\theta(x, \bar{B}_\delta(y)) \leq& \lim_{\theta\to+\infty} \frac{1}{\theta}\log \max_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta)} \\
&+ \lim_{\delta\to 0} \limsup_{\theta\to+\infty} \frac{1}{\theta}\log \mathrm{PD}(\theta)(\bar{B}_\delta(y)) = -I(y).
\end{aligned}$$

Because $\lim_{\theta\to+\infty} \frac{1}{\theta}\log \max_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta)} = 0$ due to (4.3.5); then it is easy to see the required result. $\square$

Theorem 4.4. *For fixed t, $X_t^{\theta,\alpha}$ satisfies an LDP with speed θ and rate function $I(x)$, where*

$$I(x) = \begin{cases} \log \frac{1}{1-\sum_{i=1}^{\infty} x_i}, & x \in \bar{\nabla}_{\infty}, \sum_{i=1}^{\infty} x_i < 1 \\ +\infty, & x \in \bar{\nabla}_{\infty}, \sum_{i=1}^{\infty} x_i = 1. \end{cases}$$

Proof. Notice that the transition density of $X_t^{\theta,\alpha}$ is

$$P^{\theta,\alpha}(t,x,F) = \sum_{n=2}^{+\infty} d_n^{\theta}(t)\nu_{\theta,\alpha}^n(x,F) + (d_0^{\theta}(t) + d_1^{\theta}(t))\mathrm{PD}(\theta,\alpha)(F).$$

When $\theta + \alpha > 1$, we have

$$\begin{aligned} \nu_{\theta,\alpha}^n(x,F) &= \int_F p_n^{\theta,\alpha}(x,y)\mathrm{PD}(\theta,\alpha)(dy) \\ &\leq \max_{|\eta|=n} \frac{1}{\int p_\eta d\mathrm{PD}(\theta,\alpha)}\mathrm{PD}(\theta,\alpha)(F) \leq \frac{(\theta+\alpha)_{(n)}}{(1-\alpha)^n}\mathrm{PD}(\theta,\alpha)(F); \end{aligned}$$

then

$$P^{\theta,\alpha}(t,x,F) \leq \Big[\sum_{n=2}^{+\infty} d_n^{\theta}(t)\frac{(\theta+\alpha)_{(n)}}{(1-\alpha)^n} + (d_0^{\theta}(t) + d_1^{\theta}(t))\Big]\mathrm{PD}(\theta,\alpha)(F).$$

Moreover,

$$\begin{aligned} &\sum_{n=2}^{+\infty} d_n^{\theta}(t)\frac{(\theta+\alpha)_{(n)}}{(1-\alpha)^n} + (d_0^{\theta}(t) + d_1^{\theta}(t)) \\ \leq& 1 + \sum_{m=2}^{\infty} e^{-\lambda_m t}\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{m-1}\frac{(\theta+\alpha)_{(n)}}{(1-\alpha)^n}. \end{aligned}$$

Since

$$\begin{aligned} &\sum_{n=0}^{m}\frac{2m+\theta-1}{m!}\binom{m}{n}(n+\theta)_{m-1}\frac{(\theta+\alpha)_{(n)}}{(1-\alpha)^n} \\ \leq& \frac{(2m+\theta+\alpha-1)^{2m}}{(1-\alpha)^m}\frac{2^m}{m!} \leq \frac{(2m+\theta+\alpha-1)^{2m}}{(1-\alpha)^m}, \end{aligned}$$

we have

$$\sum_{n=2}^{+\infty} d_n^{\theta}(t)\frac{(\theta+\alpha)_{(n)}}{(1-\alpha)^n} \leq \sum_{m=n}^{\infty} e^{-\lambda_m t}\frac{(2m+\theta+\alpha-1)^{2m}}{(1-\alpha)^m}$$

$$= \sum_{m=n}^{\infty} \left\{ e^{-\left[\frac{m+\theta-1}{2}t - 2\log(2m+\theta+\alpha-1) + \log\frac{1}{1-\alpha}\right]} \right\}^m.$$

Define $h^t_{\theta,\alpha}(x) = \frac{x+\theta-1}{2}t - 2\log(2x+\theta+\alpha-1) + \log\frac{1}{1-\alpha}$. Then, for fixed t,

$$h^t_{\theta,\alpha}(x)' = \frac{t}{2} - \frac{2}{x} > 0, \forall \theta+\alpha > 1, \text{ and } x > \frac{4}{t}.$$

We have $\forall m > \frac{4}{t}$,

$$h^t_{\theta,\alpha}(m) \geq h^t_{\theta,\alpha}(\frac{4}{t}) = \frac{4+(\theta-1)t}{2} - 2\log(\frac{8}{t} + \theta + \alpha - 1) + \log\frac{1}{1-\alpha}.$$

Define $K^\alpha_t(\theta) = \frac{4+(\theta-1)t}{2} - 2\log(\frac{8}{t}+\theta+\alpha-1) + \log\frac{1}{1-\alpha}, \theta+\alpha > 1$. Then, for fixed t, α, we have

$$K^\alpha_t(\theta)' = \frac{t}{2} - \frac{2}{\frac{8}{t}+\theta+\alpha-1} > \frac{t}{2} - \frac{t}{4} = \frac{t}{4} > 0.$$

Therefore, for fixed $t, K^\alpha_t(\theta)$ is increasing on $[1, +\infty)$. So $\exists \theta_0 > 1$ such that $K^\alpha_t(\theta_0) > 0$. Then $\forall \theta \geq \theta_0 > 1, m \geq \frac{4}{t}$, we have

$$h^t_{\theta,\alpha}(m) \geq h^t_{\theta,\alpha}(\frac{4}{t}) = K^\theta_t(\theta) \geq K^\alpha_t(\theta_0) > 0;$$

thus

$$\left\{ e^{-\left[\frac{m+\theta-1}{2}t - 2\log(2m+\theta+\alpha-1) + \log\frac{1}{1-\alpha}\right]} \right\}^m \leq \frac{1}{e^{h^t_{\theta,\alpha}(m)}} \leq \frac{1}{e^{K^\alpha_t(\theta_0)}} < 1.$$

By Weierstrass M-test, for fixed t,

$$\sum_{m=n}^{\infty} \left\{ e^{-\left[\frac{m+\theta-1}{2}t - 2\log(2m+\theta+\alpha-1) + \log\frac{1}{1-\alpha}\right]} \right\}^m$$

is uniformly convergent on $[\theta_0, +\infty)$. Then

$$\begin{aligned} &\lim_{\theta\to+\infty} \sum_{m=n}^{\infty} \left\{ e^{-\left[\frac{m+\theta-1}{2}t - 2\log(2m+\theta+\alpha-1) + \log\frac{1}{1-\alpha}\right]} \right\}^m \\ =&\sum_{m=n}^{\infty} \lim_{\theta\to+\infty} \left\{ e^{-\left[\frac{m+\theta-1}{2}t - 2\log(2m+\theta+\alpha-1) + \log\frac{1}{1-\alpha}\right]} \right\}^m = 0. \end{aligned}$$

Hence, we can easily show upper bound; for the lower bound, we can prove it much as the one-parameter neutral model. □

Remark 4.3. *Note that we can also show that* $\nu^n_{\theta,\alpha}(x, dy)$ *satisfies an* LDP *similar to* *that of* $\mathrm{PD}(\theta, \alpha)$.

4.3.2 Small-Time LDP for Transient Partition Structures

Since the associated transient partition structures are also an alternative description of the evolution of the allele frequency; in this section, we consider the small-time LDP for the transient partition structures associated with the one-parameter neutral model and the two-parameter model. Interestingly, one can also observe some phase transitions concerned with the LDP rate functions.

Lemma 4.7. *As* $\theta \to +\infty$*, for any given partition* $\eta = (\eta_1, \cdots, \eta_l)(\eta_l \geq 2)$*, we have the following estimation:*

$$\langle \varphi_\eta, 1\rangle_\theta \sim (\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{1}{\theta^{|\eta|-l}}. \tag{4.3.6}$$

For $\xi = (\xi_1, \cdots, \xi_p)(\xi_p \geq 2, \xi \leq \eta)$*, we have*

$$\begin{aligned}\langle \varphi_\eta, \psi_\xi^\theta\rangle_\theta \sim &\left[\sum_{i=1}^{l}\sum_{j=1}^{p}\frac{(\eta_i + \xi_j - 1)!}{(\eta_i - 1)!(\xi_j - 1)!} - |\eta||\xi|\right] \\ &(\eta_1 - 1)! \cdots (\eta_l - 1)!(\xi_1 - 1)! \cdots (\xi_p - 1)! \frac{1}{\theta^{|\eta|-l+|\xi|-p+1}}\end{aligned} \tag{4.3.7}$$

Proof. Since $\varphi_\eta = \sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)} p^o_{\sum_{i\beta_1}\eta_i, \cdots, \sum_{i\beta_d}\eta_i}$, where $\pi(l, d)$ is the set of partitions β of $\{1, \cdots, l\}$ into d subsets, $\beta_1, \cdots, \beta_d$, satisfying $\min \beta_1 < \cdots < \min \beta_d$, we have

$$\langle \varphi_\eta, 1\rangle_\theta = \sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)}\int_{\bar{\nabla}_\infty} p^o_{\sum_{i\in\beta_1}\eta_i, \cdots, \sum_{i\in\beta_d}\eta_i} d\mathrm{PD}(\theta).$$

By the Ewens sampling formula, we have

$$\langle \varphi_\eta, 1\rangle_\theta = \sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)}(\sum_{i\in\beta_1}\eta_i - 1)! \cdots (\sum_{i\in\beta_d}\eta_i - 1)! \frac{\theta^d}{\theta_{(|\eta|)}}; \tag{4.3.8}$$

therefore, the leading term in (4.3.8) is the term associated with partition $\beta = \{1\} \cup \{2\} \cup \cdots \{l\}$. Then

$$\langle \varphi_\eta, 1\rangle_\theta \sim (\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{1}{\theta^{|\eta|-l}}.$$

Now we can use mathematical induction on partition $\xi(\xi_p \geq 2)$ to show the second statement (4.3.7). Let us first check $\langle\varphi_\eta, \psi_2^\theta\rangle_\theta$. Since $\psi_2^\theta = \varphi_2 - \frac{1}{1+\theta}$, it yields that

$$\langle\varphi_\eta, \psi_2^\theta\rangle_\theta = \langle\varphi_\eta, \varphi_2\rangle_\theta - \frac{1}{1+\theta}\langle\varphi_\eta, 1\rangle_\theta. \tag{4.3.9}$$

Let us define $\tilde{\eta} = (\eta_1, \cdots, \eta_l, 2)$. Then $\langle\varphi_\eta, \varphi_2\rangle_\theta = \langle\varphi_{\tilde{\eta}}, 1\rangle_\theta$. By the Ewens sampling formula, we have

$$\langle\varphi_\eta, \varphi_2\rangle_\theta = \sum_{d=1}^{l+1} \sum_{\beta\in\pi(l+1,d)} (\sum_{i\in\beta_1} \tilde{\eta}_i - 1)! \cdots (\sum_{i\in\beta_d} \tilde{\eta}_i - 1)! \frac{\theta^d}{\theta_{(|\eta|+2)}}. \tag{4.3.10}$$

Substituting (4.3.10) and (4.3.8) into (4.3.9), we have

$$\begin{aligned}\langle\varphi_\eta, \psi_2^\theta\rangle_\theta =& \sum_{d=1}^{l+1} \sum_{\beta\in\pi(l+1,d)} (\sum_{i\in\beta_1} \tilde{\eta}_i - 1)! \cdots (\sum_{i\in\beta_d} \tilde{\eta}_i - 1)! \frac{\theta^d}{\theta_{(|\eta|+2)}} & (4.3.11)\\ & - \sum_{d=1}^{l} \sum_{\beta\in\pi(l,d)} (\sum_{i\in\beta_1} \eta_i - 1)! \cdots (\sum_{i\in\beta_d} \eta_i - 1)! \frac{\theta^d}{\theta_{(|\eta|)}} \frac{1}{1+\theta} & (4.3.12)\end{aligned}$$

In (4.3.11), for a given d, the corresponding term is of the order $\frac{1}{\theta^{|\eta|+2-d}}$; in (4.3.12), for a given d, the associated term is of the order $\frac{1}{\theta^{|\eta|+1-d}}$. Let us first check terms that are associated with $d = l$ or $l+1$ in (4.3.11) and with $d = l$ or $d = l-1$ in (4.3.12). Then the summation of those terms are

$$\begin{aligned}&(\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{\theta^{l+1}}{\theta_{(|\eta|+2)}} + (\eta_1 - 1)! \cdots (\eta_l - 1)! \sum_{u=1}^{l} (\eta_u + 1)\eta_u \frac{\theta^l}{\theta_{(|\eta|+2)}}\\ &+ (\eta_1 - 1)! \cdots (\eta_l - 1)! \sum_{1\le u<v\le l} \frac{(\eta_u + \eta_v - 1)!}{(\eta_u - 1)!(\eta_v - 1)!} \frac{\theta^l}{\theta_{(|\eta|+2)}}\\ &- (\eta_1 - 1)! \cdots (\eta_l - 1)! \sum_{1\le u<v\le l} \frac{(\eta_u + \eta_v - 1)!}{(\eta_u - 1)!(\eta_v - 1)!} \frac{\theta^{l-1}}{\theta_{(|\eta|)}(\theta+1)}\\ &- (\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{\theta^l}{\theta_{(|\eta|)}} \frac{1}{\theta+1}.\end{aligned} \tag{4.3.13}$$

Clearly, the above summation (4.3.13) can be rewritten as

$$\begin{aligned}&(\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{\theta^l}{\theta_{(|\eta|+2)}} \Bigg[\sum_{u=1}^{l} \eta_u(\eta_u + 1) + \theta - \frac{(|\eta| + \theta)(|\eta| + \theta + 1)}{\theta + 1}\\ &+ \sum_{1\le u<v\le l} \frac{(\eta_u + \eta_v - 1)!}{(\eta_u - 1)!(\eta_v - 1)!} \left(1 - \frac{(\theta + |\eta|)(\theta + |\eta| + 1)}{\theta(\theta + 1)}\right) \Bigg]\end{aligned}$$

$$
\begin{aligned}
=&(\eta_1-1)!\cdots(\eta_l-1)!\frac{\theta^l}{\theta_{(|\eta|+2)}}\left[\sum_{u=1}^{l}\eta_u(\eta_u+1)-\frac{2\theta+1}{\theta+1}|\eta|-\frac{|\eta|^2}{1+\theta}\right.\\
&\left.-\sum_{1\le u<v\le l}\frac{(\eta_u+\eta_v-1)!}{(\eta_u-1)!(\eta_v-1)!}\frac{|\eta|^2+|\eta|(2\theta+1)}{\theta(\theta+1)}\right]\\
\sim&(\eta_1-1)!\cdots(\eta_l-1)!(\sum_{u=1}^{l}\eta_u(\eta_u+1)-2|\eta|)\frac{1}{\theta^{|\eta|+2-l}}\quad \text{as } \theta\to+\infty.
\end{aligned}
$$

All the remaining terms in (4.3.11) and (4.3.12) are at least of the order $\frac{1}{\theta^{|\eta|+3-l}}$. Therefore,

$$
\langle\varphi_\eta,\psi_2^\theta\rangle_\theta\sim(\eta_1-1)!\cdots(\eta_l-1)!(\sum_{u=1}^{l}\eta_u(\eta_u+1)-2|\eta|)\frac{1}{\theta^{|\eta|+2-l}}.
$$

Now we assume that, for a given partition $\xi(\xi_p>2)$, $\forall\delta<\xi(\delta_{l(\delta)}\ge 2)$ and $\forall\eta\ge 2$, we have

$$
\begin{aligned}
\langle\varphi_\eta,\psi_\delta^\theta\rangle_\theta\sim&\left[\sum_{i=1}^{l}\sum_{j=1}^{l(\delta)}\frac{(\eta_i+\delta_j-1)!}{(\eta_i-1)!(\delta_j-1)!}-|\eta||\delta|\right]\\
&(\eta_1-1)!\cdots(\eta_l-1)!(\delta_1-1)!\cdots(\delta_{l(\delta)}-1)!\frac{1}{\theta^{|\eta|-l+|\delta|-l(\delta)+1}}.
\end{aligned}
\tag{4.3.14}
$$

Now we consider $\langle\varphi_\eta,\psi_\xi^\theta\rangle_\theta$. Since

$$
\psi_\xi^\theta=\varphi_\xi-\sum_{\delta<\xi}\langle\varphi_\xi,\psi_\delta^\theta\rangle_\theta\psi_\delta^\theta,
$$

$$
\langle\varphi_\eta,\psi_\xi^\theta\rangle_\theta=\langle\varphi_\eta,\varphi_\xi\rangle_\theta-\sum_{2\le\delta<\xi}\langle\varphi_\xi,\psi_\delta^\theta\rangle_\theta\langle\varphi_\eta,\psi_\delta^\theta\rangle_\theta-\langle\varphi_\xi,1\rangle_\theta\langle\varphi_\eta,1\rangle_\theta. \tag{4.3.15}
$$

By the above assumption (4.3.14), for each partition $\delta<\xi$, the associated summands in the second part of (4.3.15) are of the order $\frac{1}{\theta^{|\eta|-l+|\xi|-p+1}}\frac{1}{\theta^{2(|\delta|-l(\delta))+1}}$. Notice that $2(|\delta|-l(\delta))+1\ge 3$ for $\delta\ge(2)$. So the leading term in

$$
\sum_{2\le\delta<\xi}\langle\varphi_\xi,\psi_\delta^\theta\rangle_\theta\langle\varphi_\eta,\psi_\delta^\theta\rangle_\theta
$$

is of the order $\frac{1}{\theta^{|\eta|+|\xi|-l-p+4}}$; but the leading term in $\langle\varphi_\eta,\varphi_\xi\rangle_\theta$ is of the order $\frac{1}{\theta^{|\eta|+|\xi|-l-p}}$. Therefore, we should first check $\langle\varphi_\eta,\varphi_\xi\rangle_\theta-\langle\varphi_\xi,1\rangle_\theta\langle\varphi_\eta,1\rangle_\theta$. Let us define $\omega=(\eta_1,\cdots,\eta_l,\xi_1,\cdots,\xi_p)$. Then $\langle\varphi_\eta,\varphi_\xi\rangle_\theta=\langle\varphi_\omega,1\rangle_\theta$. By the Ewens sampling formula, we have

$$
\begin{aligned}
\langle\varphi_\eta,\varphi_\xi\rangle_\theta =& \sum_{d=1}^{l+p}\sum_{\beta\in\pi(l+p,d)}\int_{\bar\nabla_\infty} p^o_{\sum_{i\in\beta_1}\omega_i,\cdots,\sum_{i\in\beta_d}\omega_i} d\mathrm{PD}(\theta)\\
=& \sum_{d=1}^{l+p}\sum_{\beta\in\pi(l+p,d)}(\sum_{i\in\beta_1}\omega_i-1)!\cdots(\sum_{i\in\beta_d}\omega_i-1)!\frac{\theta^d}{\theta_{(|\eta|+|\xi|)}}.
\end{aligned}
$$

Thus, we should consider terms of the order $\frac{1}{\theta^{|\eta|+|\xi|-l-p+1}}$ and $\frac{1}{\theta^{|\eta|+|\xi|-l-p}}$ in

$$
\langle\varphi_\eta,\varphi_\xi\rangle_\theta-\langle\varphi_\xi,1\rangle_\theta\langle\varphi_\eta,1\rangle_\theta;
$$

and the summation of these terms is

$$
\begin{aligned}
&(\omega_1-1)!\cdots(\omega_{l+p}-1)!\frac{\theta^{l+p}}{\theta_{(|\eta|+|\xi|)}}\\
&+\sum_{1\le i<j\le l+p}\frac{(\omega_i+\omega_j-1)!}{(\omega_i-1)!(\omega_j-1)!}(\omega_1-1)!\cdots(\omega_{l+p}-1)!\frac{\theta^{l+p-1}}{\theta_{(|\eta|+|\xi|)}}
\end{aligned}
\tag{4.3.16}
$$

$$
\begin{aligned}
&-(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{\theta^{l+p}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}\\
&-\sum_{1\le i<j\le l}\frac{(\eta_i+\eta_j-1)!}{(\eta_i-1)!(\eta_j-1)!}(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{\theta^{l+p-1}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}\\
&-\sum_{1\le i<j\le p}\frac{(\xi_i+\xi_j-1)!}{(\xi_i-1)!(\xi_j-1)!}(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{\theta^{l+p-1}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}.
\end{aligned}
\tag{4.3.17}
$$

Recall that

$$
(\omega_1-1)!\cdots(\omega_{l+p}-1)!=(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!.
$$

Then the above summation, (4.3.16)+(4.3.17), is the following:

$$
\begin{aligned}
&(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{\theta^{l+p-1}}{\theta_{(|\eta|+|\xi|)}}\\
&\Bigg[\sum_{1\le i<j\le l+p}\frac{(\omega_i+\omega_j-1)!}{(\omega_i-1)!(\omega_j-1)!}+\theta\big(1-\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}\big)\\
&-\sum_{1\le i<j\le l}\frac{(\eta_i+\eta_j-1)!}{(\eta_i-1)!(\eta_j-1)!}\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta(|\xi|)}-\sum_{1\le i<j\le p}\frac{(\xi_i+\xi_j-1)!}{(\xi_i-1)!(\xi_j-1)!}\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
=&(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{\theta^{l+p-1}}{\theta_{(|\eta|+|\xi|)}}\\
&\left[\sum_{1\le i<j\le l+p}\frac{(\omega_i+\omega_j-1)!}{(\omega_i-1)!(\omega_j-1)!}+\theta\left(\frac{\prod_{i=0}^{|\eta|-1}(\theta+i)-\prod_{i=0}^{|\eta|-1}(|\xi|+\theta+i)}{\theta_{(|\eta|)}}\right)\right.\\
&\left.-\sum_{1\le i<j\le l}\frac{(\eta_i+\eta_j-1)!}{(\eta_i-1)!(\eta_j-1)!}\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}-\sum_{1\le i<j\le p}\frac{(\xi_i+\xi_j-1)!}{(\xi_i-1)!(\xi_j-1)!}\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}\right]\\
=&(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{\theta^{l+p-1}}{\theta_{(|\eta|+|\xi|)}}\\
&\left[\sum_{1\le i<j\le l+p}\frac{(\omega_i+\omega_j-1)!}{(\omega_i-1)!(\omega_j-1)!}-\theta\left(\frac{\sum_{u=1}^{|\eta|}|\xi|^u\sum_{0\le i_1<\cdots<i_{|\eta|-u}\le|\eta|-1}(\theta+i_1)\cdots(\theta+i_{|\eta|-u})}{\theta_{(|\eta|)}}\right)\right.\\
&\left.-\sum_{1\le i<j\le l}\frac{(\eta_i+\eta_j-1)!}{(\eta_i-1)!(\eta_j-1)!}\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}-\sum_{1\le i<j\le p}\frac{(\xi_i+\xi_j-1)!}{(\xi_i-1)!(\xi_j-1)!}\frac{\theta_{(|\eta|+|\xi|)}}{\theta_{(|\eta|)}\theta_{(|\xi|)}}\right]\\
\sim&(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\frac{1}{\theta^{|\eta|-l+|\xi|-p+1}}\\
&\left[\sum_{1\le i<j\le l+p}\frac{(\omega_i+\omega_j-1)!}{(\omega_i-1)!(\omega_j-1)!}-|\xi|\theta\sum_{i=0}^{|\eta|-1}\frac{1}{\theta+i}\right.\\
&\left.-\sum_{1\le i<j\le l}\frac{(\eta_i+\eta_j-1)!}{(\eta_i-1)!(\eta_j-1)!}-\sum_{1\le i<j\le p}\frac{(\xi_i+\xi_j-1)!}{(\xi_i-1)!(\xi_j-1)!}\right]\\
\sim&(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\\
&\left[\sum_{i=1}^{l}\sum_{j=1}^{p}\frac{(\eta_i+\xi_j-1)!}{(\eta_i-1)!(\xi_j-1)!}-|\eta||\xi|\right]\frac{1}{\theta^{|\eta|-l+|\xi|-p+1}}.
\end{aligned}
$$

Therefore, the leading term should be

$$
\begin{aligned}
&(\eta_1-1)!\cdots(\eta_l-1)!(\xi_1-1)!\cdots(\xi_p-1)!\\
&\left[\sum_{i=1}^{l}\sum_{j=1}^{p}\frac{(\eta_i+\xi_j-1)!}{(\eta_i-1)!(\xi_j-1)!}-|\eta||\xi|\right]\frac{1}{\theta^{|\eta|-l+|\xi|-p+1}}.
\end{aligned}
$$

This lemma is thus proved! □

Similarly, for the two-parameter model, we have the following lemma.

Lemma 4.8. *As $\theta\to+\infty$, for any given partition $\eta=(\eta_1,\cdots,\eta_l)(\eta_l\ge 2)$, we have* *the following estimations:*

$$\langle \varphi_\eta, 1\rangle_{\theta,\alpha} \sim \prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\frac{1}{\theta^{|\eta|-l}}.$$

For $\xi = (\xi_1, \cdots, \xi_p)(\xi_p \geq 2, \xi \leq \eta)$, *we have*

$$\begin{aligned}\langle \varphi_\eta, \psi_\xi^{\theta,\alpha}\rangle_{\theta,\alpha} \sim &\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\prod_{j=1}^{p}(1-\alpha)_{(\xi_j-1)}\\ &\left[\sum_{i=1}^{l}\sum_{j=1}^{p}\frac{(1-\alpha)_{(\eta_i+\xi_j-1)}}{(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(\xi_j-1)}} + l\alpha - |\eta||\xi|\right]\frac{1}{\theta^{|\eta|-l+|\xi|-p+1}}.\end{aligned}$$

Proof. First, we are going to show the first part. By the Pitman sampling formula, we have

$$\int_{\bar{\nabla}_\infty} p_\eta^o d\mathrm{PD}(\theta,\alpha) = \prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\frac{\prod_{j=0}^{l-1}(\theta+j\alpha)}{\theta_{(|\eta|)}}.$$

Thus,

$$\begin{aligned}\langle \varphi_\eta, 1\rangle_{\theta,\alpha} =&\sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)}\prod_{j=1}^{d}(1-\alpha)_{(\sum_{i\in\beta_j}\eta_i-1)}\frac{\prod_{u=0}^{d-1}(\theta+u\alpha)}{\theta_{(|\eta|)}}\\ \sim&\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\frac{1}{\theta^{|\eta|-l}}.\end{aligned}$$

Again, we are going to use mathematical induction on $\xi(\xi_p \geq 2)$ to show the second part. Let us first check $\langle \varphi_\eta, \psi_2^{\theta,\alpha}\rangle_{\theta,\alpha}$. Since $\psi_2^{\theta,\alpha} = \varphi_\eta - \frac{1-\alpha}{\theta+1}$, we have

$$\langle \varphi_\eta, \psi_2^{\theta,\alpha}\rangle_{\theta,\alpha} = \langle \varphi_2, \varphi_2\rangle_{\theta,\alpha} - \langle \varphi_\eta, 1\rangle_{\theta,\alpha}\frac{1-\alpha}{\theta+1},$$

where

$$\begin{aligned}\langle \varphi_\eta, \varphi_2\rangle_{\theta,\alpha} =&\sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)}\prod_{j=1}^{d}(1-\alpha)_{(\sum_{i\in\beta_j}\eta_i-1)}(1-\alpha)_{(2-1)}\frac{\prod_{u=0}^{d}(\theta+u\alpha)}{\theta_{(|\eta|+2)}}\\ &+\sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)}\prod_{j=1}^{d}(1-\alpha)_{(\sum_{i\in\beta_j}\eta_i-1)}(1-\alpha)_{(2-1)}\frac{\prod_{u=0}^{d-1}(\theta+u\alpha)}{\theta_{(|\eta|+2)}}\\ &\sum_{v=1}^{d}\frac{(1-\alpha)_{(\sum_{i\in\beta_v}\eta_i+2-1)}}{(1-\alpha)_{(\sum_{i\in\beta_v}\eta_i-1)}(1-\alpha)_{(2-1)}},\end{aligned}$$

and

$$\frac{1-\alpha}{\theta+1}\langle\varphi_\eta,1\rangle_{\theta,\alpha}=\sum_{d=1}^{l}\sum_{\beta\in\pi(l,d)}\prod_{j=1}^{d}(1-\alpha)_{(\sum_{i\in\beta_j}\eta_i-1)}\frac{\prod_{u=0}^{d-1}(\theta+u\alpha)}{\theta_{(|\eta|)}}\frac{1-\alpha}{\theta+1}.$$

Consider terms of the order $\frac{1}{\theta^{|\eta|-l+2}}$ or lower. We have, as $\theta\to+\infty$,

$$\begin{aligned}
&\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(2-1)}\frac{\prod_{u=0}^{l}(\theta+u\alpha)}{\theta_{(|\eta|+2)}}\\
&+\sum_{v=0}^{l}\frac{(1-\alpha)_{(\eta_v+2-1)}}{(1-\alpha)_{(\eta_v-1)}(1-\alpha)_{(2-1)}}\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(2-1)}\frac{\prod_{u=0}^{l-1}(\theta+u\alpha)}{\theta_{(|\eta|+2)}}\\
&+\sum_{1\le i<j\le l}\frac{(1-\alpha)_{(\eta_i+\eta_j-1)}}{(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(\eta_j-1)}}\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(2-1)}\frac{\prod_{u=0}^{l-1}(\theta+u\alpha)}{\theta_{(|\eta|+2)}}\\
&-\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\frac{\prod_{u=0}^{l-1}(\theta+u\alpha)}{\theta_{(|\eta|)}}\frac{1-\alpha}{\theta+1}\\
&-\sum_{1\le i<j\le l}\frac{(1-\alpha)_{(\eta_i+\eta_j-1)}}{(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(\eta_j-1)}}\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\frac{\prod_{u=0}^{l-2}(\theta+u\alpha)}{\theta_{(|\eta|)}}\frac{1-\alpha}{\theta+1}\\
&\sim(1-\alpha)\prod_{j=1}^{l}(1-\alpha)_{(\eta_j-1)}\left[\sum_{i=1}^{l}\frac{(1-\alpha)_{(\eta_i+2-1)}}{(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(2-1)}}+l\alpha-2|\eta|\right]\frac{1}{\theta^{|\eta|-l+2}}.
\end{aligned}$$

Now we assume that for a given partition $\xi(\xi_p\ge 2)$, $\forall\delta<\xi(\delta_{l(\delta)}\ge 2)$ and $\forall\eta\ge 2$, we have, as $\theta\to+\infty$,

$$\begin{aligned}
\langle\varphi_\eta,\psi_\delta^{\theta,\alpha}\rangle_{\theta,\alpha}\sim&\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\prod_{j=1}^{l(\delta)}(1-\alpha)_{(\delta_j-1)}\\
&\left[\sum_{i=1}^{l}\sum_{j=1}^{l(\delta)}\frac{(1-\alpha)_{(\eta_i+\delta_j-1)}}{(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(\delta_j-1)}}+l\alpha-|\eta||\delta|\right]\frac{1}{\theta^{|\eta|-l+|\delta|-l(\delta)+1}}.
\end{aligned}$$

The remaining part can be proved as analogously as Lemma (4.7). □

Let us use $\{P_n^\theta,n\ge 1\}$ to denote the partition structures associated with $X_{t(\theta)}$; and we use $\{P_n^{\theta,\alpha},n\ge 1\}$ to represent the partition structures associated with $X_{t(\theta)}^{\theta,\alpha}$. We have the following LDP for $\{P_n^\theta,\theta>0\}$ and $\{P_n^{\theta,\alpha},\theta>0\}$, for each fixed $n\ge 2$.

Theorem 4.5. *Assume that $\lim_{\theta\to+\infty}\theta t(\theta)=\infty$. As $\theta\to+\infty$, for any given integer $n\ge 2$, P_n^θ has the following LDPs:*

- *If* $\lim_{\theta\to\infty}\frac{\theta t(\theta)}{\log\theta}=k\geq 2$, *then* P_n^θ *has an LDP with speed* $\log\theta$ *and rate function* $I_n(\eta)=n-l(\eta)$.
- *If* $\lim_{\theta\to\infty}\frac{\theta t(\theta)}{\log\theta}=k\in(0,2)$, *then* P_n^θ *has an LDP with speed* $\log\theta$ *and rate function*

$$I_n(\eta)=\begin{cases}0, & \text{if } \eta=(1,1,\cdots,1)\\ \frac{n-\alpha_1(\eta)}{2}k, & \text{if } \frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)}>\frac{2}{2-k}\\ n-l(\eta), & \text{if } \frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)}\leq\frac{2}{2-k}.\end{cases}$$

- *If* $\lim_{\theta\to\infty}\frac{\theta t(\theta)}{\log\theta}=0$, *then* P_n^θ *has an LDP with speed* $\theta t(\theta)$ *and rate function* $I_n(\eta)=\frac{n-\alpha_1(\eta)}{2}$.

Remark 4.4. *This theorem indicates that each* $k\in(0,2)$ *serves as a critical value. But since all the partition distributions together determine the distribution of* $X_{t(\theta)}$, $k=2$ *should be a critical point for* $X_{t(\theta)}$. *Therefore,* $X_{t(\theta)}$ *has a critical time scale* $\frac{2\log\theta}{\theta}$. *Furthermore,* $\frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)}$ *is the average density of a given partition excluding singletons.*

Proof. Since

$$P_n^\theta(\eta)=\frac{n!}{\eta_1!\cdots\eta_{l(\eta)}!\alpha_1(\eta)!\cdots\alpha_n(\eta)!}Ep_\eta^o(X_{t(\theta)}),$$

and $X_{t(\theta)}\to\delta_{(0,0,\cdots)}$, one can easily have

$$P_n^\theta(\eta)\to\delta_{(1,1,\cdots,1)}(\eta).$$

Then $\log P_n^\theta(1,1,\cdots,1)\to 0$, as $\theta\to+\infty$. So we only need to consider the case $\eta\neq(1,1,\cdots,1)$. Now we assume $\eta=(\eta_1,\cdots,\eta_l)$ thereby $l(\eta)=l$. Thanks to Proposition 2.2, we know

$$p_\eta^o=\sum_{d=1}^{l}(-1)^{l-d}\sum_{\beta\in\pi(l,d)}(|\beta_1|-1)!\cdots(|\beta_d|-1)!\varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i}.$$

Let $(\sum_{i\in\beta_{(1)}}\eta_i,\cdots,\sum_{i\in\beta_{(d)}}\eta_i)$ be the decreasing rearrangement of $(\sum_{i\in\beta_1}\eta_i,\cdots,\sum_{i\in\beta_d}\eta_i)$. We define $\eta^\beta=(\sum_{i\in\beta_{(1)}}\eta_i,\cdots,\sum_{i\in\beta_{(d)}}\eta_i)$. If $\sum_{i\in\beta_{(j)}}\eta_i=1$, then we delete it such that $\eta^\beta\in\tilde{\mathcal{M}}$ for we will repeatedly apply lemma 4.2; but we always have $|\eta^\beta|-l(\eta^\beta)=|\eta|-d$ and $|\eta^\beta|\geq|\eta|-\alpha_1(\eta)$. Since

$$\begin{aligned}Ep_\eta^o(X_{t(\theta)})=&\int_{\bar{\nabla}_\infty}p_\eta^o(y)\mathrm{PD}(\theta)(dy)+\sum_{m=2}^{\infty}e^{-\lambda_m t(\theta)}\int_{\bar{\nabla}_\infty}p_\eta^o(y)Q_m(x,y)\mathrm{PD}(\theta)(dy)\\ =&(\eta_1-1)!\cdots(\eta_l-1)!\frac{\theta^l}{\theta_{(|\eta|)}}+\sum_{d=1}^{l}(-1)^{l-d}\sum_{\beta\in\pi(l,d)}\prod_{i=1}^{d}(|\beta_i|-1)!\end{aligned}$$

$$\sum_{m=2}^{\infty} e^{-\lambda_m t(\theta)} \int_{\bar{\nabla}_\infty} \varphi_{\eta^\beta}(y) Q_m(x,y) \mathrm{PD}(\theta)(dy),$$

and $\varphi_{\eta^\beta} = \psi^\theta_{\eta^\beta} + \sum_{\delta<\xi^\beta} \langle \varphi_{\eta^\beta}, \psi^\theta_\delta \rangle_\theta \psi^\theta_\delta$, one can easily have

$$\begin{aligned}
Ep^o_\eta(X_{t(\theta)}) =& (\eta_1-1)!\cdots(\eta_l-1)!\frac{\theta^l}{\theta_{(|\eta|)}} + \sum_{d=1}^{l}(-1)^{l-d} \sum_{\beta\in\pi(l,d)} \prod_{i=1}^{d} (|\beta_i|-1)! \\
& \sum_{m=2}^{\infty} e^{-\lambda_m t(\theta)} \left(\psi^\theta_{\eta^\beta}(x)\delta_{m,|\eta^\beta|} + \sum_{2\le\delta<\eta^\beta} \langle \varphi_{\eta^\beta}, \psi^\theta_\delta\rangle_\theta \psi^\theta_\delta(x)\delta_{m,|\delta|} \right) \\
=& (\eta_1-1)!\cdots(\eta_l-1)!\frac{\theta^l}{\theta_{(|\eta|)}} + \sum_{d=1}^{l}(-1)^{l-d} \sum_{\beta\in\pi(l,d)} \prod_{i=1}^{d} (|\beta_i|-1)! \\
& \left[\psi^\theta_{\eta^\beta}(x) e^{-\lambda_{|\eta^\beta|} t(\theta)} + \sum_{2\le\delta<\eta^\beta} \langle \varphi_{\eta^\beta}, \psi^\theta_\delta\rangle_\theta \psi^\theta_\delta(x) e^{-\lambda_{|\delta|} t(\theta)} \right]
\end{aligned}$$

(due to Lemma 4.2). If we rearrange $Ep^o_\eta(X_{t(\theta)})$, then it will be

$$(\eta_1-1)!\cdots(\eta_l-1)!\frac{\theta^l}{\theta_{(|\eta|)}} \tag{4.3.18}$$

$$+\sum_{d=1}^{l}(-1)^{l-d} \sum_{\beta\in\pi(l,d)} \prod_{i=1}^{d} (|\beta_i|-1)! \varphi_{\eta^\beta}(x) e^{-\lambda_{\eta^\beta} t(\theta)} \tag{4.3.19}$$

$$+\sum_{d=1}^{l}(-1)^{l-d} \sum_{\beta\in\pi(l,d)} \prod_{i=1}^{d} (|\beta_i|-1)! \sum_{2\le\delta<\eta^\beta} \langle \varphi_{\eta^\beta}, \psi^\theta_\delta\rangle_\theta \psi^\theta_\delta(x) (e^{-\lambda_{|\delta|} t(\theta)} - e^{-\lambda_{|\eta^\beta|} t(\theta)}) \tag{4.3.20}$$

$$-\sum_{d=1}^{l}(-1)^{l-d} \sum_{\beta\in\pi(l,d)} \prod_{i=1}^{d} (|\beta_i|-1)! \langle \varphi_{\eta^\beta}, 1\rangle_\theta e^{-\lambda_{|\eta^\beta|} t(\theta)}. \tag{4.3.21}$$

Recall that $|\eta^\beta| \ge n - \alpha_1(\eta)$; then

$$\begin{aligned}
(4.3.19) \sim & \left[\sum_{d=1}^{l-\alpha_1(\eta)} (-1)^{l-(d+\alpha_1(\eta))} \sum_{\beta\in\pi(l-\alpha_1(\eta),d)} \prod_{i=1}^{d} (|\beta_i|-1)! \varphi_{\eta^\beta}(x) \right] e^{-\lambda_{|\eta|-\alpha_1(\eta)} t(\theta)} \\
& \sim p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}\theta t(\theta)} \quad \text{due to Proposition 2.2.}
\end{aligned}$$

By Lemma 4.7, we know $\forall \delta < \eta^\beta$,

$$\langle \varphi_{\eta^\beta}, \psi^\theta_\delta \rangle_\theta \sim \left[\sum_{i=1}^{l(\eta^\beta)} \sum_{j=1}^{l(\delta)} \frac{(\eta^\beta_i + \delta_j - 1)!}{(\eta^\beta_i - 1)!(\delta_j - 1)!} - |\eta^\beta||\delta| \right]$$

$$(\eta^\beta_1 - 1)! \cdots (\eta^\beta_{l(\eta^\beta)} - 1)!(\delta_1 - 1)! \cdots (\delta_{l(\delta)} - 1)! \frac{1}{\theta^{|\eta^\beta| - l(\eta^\beta) + |\delta| - l(\delta) + 1}}.$$

Since $|\eta^\beta| - l(\eta^\beta) = n - d$ and $|\delta| - l(\delta) \geq 1$, for $\delta \geq (2)$, one can easily conclude that

$$|\eta^\beta| - l(\eta^\beta) + |\delta| - l(\delta) + 1 = n - d + |\delta| - l(\delta) + 1 \geq n - d + 2.$$

Moreover, $1 \leq d \leq l$; then $|\eta^\beta| - l(\eta^\beta) + |\delta| - l(\delta) + 1 \geq n - l + 2$, where the equality holds if and only if $d = l$ and $\delta = (2)$. So if $\eta^\beta = (2)$, then (4.3.20) = 0; if $\eta^\beta > (2)$, then

$$(4.3.20) \sim \left[\sum_{i=1}^{l-\alpha_1(\eta)} \frac{(\eta_i + 1)!}{(\eta_i - 1)!} - 2(|\eta| - \alpha_1(\eta)) \right] \prod_{i=1}^{l} (\eta_i - 1)! \varphi_2(x) \frac{1}{\theta^{|\eta| - l + 2}} e^{-(\theta+1)t(\theta)}$$

and

$$\begin{aligned} (4.3.21) &\sim -(\eta_1 - 1)! \cdots (\eta_{l-\alpha_1(\eta)} - 1)! \frac{1}{\theta^{|\eta| - l}} e^{-\frac{|\eta| - \alpha_1(\eta)}{2} \theta t(\theta)}, \\ (4.3.18) &\sim (\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{1}{\theta^{n-l}}. \end{aligned}$$

Obviously, (4.3.18) is the leading term among (4.3.18), (4.3.20) and (4.3.21). Then

$$(4.3.18) + (4.3.20) + (4.3.21) \sim (\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{1}{\theta^{n-l}};$$

and $(4.3.19) \sim p^o_{\eta_1, \cdots, \eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta| - \alpha_1(\eta)}{2} \theta t(\theta)}$. Hence,

$$Ep^o_\eta(X_{t(\theta)}) \sim (\eta_1 - 1)! \cdots (\eta_l - 1)! \frac{1}{\theta^{n-l}} + p^o_{\eta_1, \cdots, \eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta| - \alpha_1(\eta)}{2} \theta t(\theta)}.$$

Let us consider three cases: $\lim_{\theta \to +\infty} \frac{\theta t(\theta)}{\log \theta} \geq 2$, $\lim_{\theta \to +\infty} \frac{\theta t(\theta)}{\log \theta} = k \in (0, 2)$ and $\lim_{\theta \to +\infty} \frac{\theta t(\theta)}{\log \theta} = 0$.

Case I: $\lim_{\theta \to +\infty} \frac{\theta t(\theta)}{\log \theta} = 0$

We know

$$Ep^o_\eta(X_{t(\theta)}) \sim e^{-(\frac{|\eta| - \alpha_1(\eta)}{2}) \theta t(\theta)} \left[p^o_{\eta_1, \cdots, \eta_{l-\alpha_1(\eta)}}(x) + \prod_{i=1}^{l} (\eta_i - 1)! e^{-\log \theta (n - l - \frac{|\eta| - \alpha_1(\eta)}{2} \frac{\theta t(\theta)}{\log \theta})} \right].$$

Then

$$Ep_\eta^o(X_{t(\theta)}) \sim e^{-(\frac{|\eta|-\alpha_1(\eta)}{2})\theta t(\theta)} p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x).$$

Therefore,

$$\lim_{\theta\to\infty} \frac{1}{\theta t(\theta)} \log P_n^\theta(\eta) = -\frac{|\eta|-\alpha_1(\eta)}{2}.$$

Now for the remaining two cases, we have

$$\begin{aligned} Ep_\eta^o(X_{t(\theta)}) \sim & e^{-\frac{|\eta|-\alpha_1(\eta)}{2}k\log\theta}\Bigg[\prod_{i=1}^{l}(\eta_i-1)! e^{-(n-l-\frac{|\eta|-\alpha_1(\eta)}{2}k)\log\theta} \\ & + p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}(\frac{\theta t(\theta)}{\log\theta}-k)\log\theta}\Bigg]. \end{aligned}$$

Define

$$\begin{aligned} D(\theta) = & \frac{1}{\log\theta}\log\Bigg[\prod_{i=1}^{l}(\eta_i-1)! e^{-(n-l-\frac{|\eta|-\alpha_1(\eta)}{2}k)\log\theta} \\ & + p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}(\frac{\theta t(\theta)}{\log\theta}-k)\log\theta}\Bigg]. \end{aligned}$$

We claim that $\lim_{\theta\to\infty} D(\theta) = \max\left\{0, -[(n-l) - \frac{|\eta|-\alpha_1(\eta)}{2}k]\right\}$. Indeed, by lemma 1.2.15 in [Dembo and Zeitouni, 2010],

$$\begin{aligned} \limsup_{\theta\to+\infty} D(\theta) = & \max\Bigg\{\limsup_{\theta\to+\infty}\frac{1}{\log\theta}\log\left[\prod_{i=1}^{l}(\eta_i-1)! e^{-(n-l-\frac{|\eta|-\alpha_1(\eta)}{2}k)\log\theta}\right], \\ & \limsup_{\theta\to+\infty}\frac{1}{\log\theta}\log\left[p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}(\frac{\theta t(\theta)}{\log\theta}-k)\log\theta}\right]\Bigg\} \\ = & \max\left\{-[(n-l) - \frac{|\eta|-\alpha_1(\eta)}{2}k], 0\right\}. \end{aligned}$$

Moreover,

$$\begin{aligned} \liminf_{\theta\to+\infty} D(\theta) \geq & \max\Bigg\{\liminf_{\theta\to+\infty}\frac{1}{\log\theta}\log\left[\prod_{i=1}^{l}(\eta_i-1)! e^{-(n-l-\frac{|\eta|-\alpha_1(\eta)}{2}k)\log\theta}\right] \\ & \liminf_{\theta\to+\infty}\frac{1}{\log\theta}\log\left[p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}(\frac{\theta t(\theta)}{\log\theta}-k)\log\theta}\right]\Bigg\} \\ = & \max\left\{-[(n-l) - \frac{|\eta|-\alpha_1(\eta)}{2}k], 0\right\}. \end{aligned}$$

Therefore, the claim is true. Then we have

$$\lim_{\theta\to+\infty}\frac{1}{\log\theta}\log P_n^\theta(\eta) = -\frac{n-\alpha_1(\eta)}{2}k + \max\left\{0, -\left[(n-l) - \frac{|\eta|-\alpha_1(\eta)}{2}k\right]\right\}.$$

Case II: $\lim_{\theta\to+\infty}\frac{\theta t(\theta)}{\log\theta} = k \geq 2$
For fixed $k > 0$, if $k \geq 2$, then

$$\frac{n-\alpha_1(\eta)}{2}k \geq n-\alpha_1(\eta) \geq n-l, \text{ for } \alpha_1(\eta) \leq l.$$

Additionally $(n-l) - \frac{n-\alpha_1(\eta)}{2}k \leq 0$; thus,

$$\lim_{\theta\to+\infty}\frac{1}{\log\theta}\log P_n^\theta(\eta) = -(n-l).$$

Case III: $\lim_{\theta\to+\infty}\frac{\theta t(\theta)}{\log\theta} = k \in (0,2)$
When $0 < k < 2$, we have $n-l = n-\alpha_1(\eta) - (l-\alpha_1(\eta))$. Then

$$(n-l) - \frac{n-\alpha_1(\eta)}{2}k \geq 0 \text{ if and only if } \frac{n-\alpha_1(\eta)}{l-\alpha_1(\eta)} \geq \frac{2}{2-k}$$

and

$$(n-l) - \frac{n-\alpha_1(\eta)}{2}k < 0 \text{ if and only if } \frac{n-\alpha_1(\eta)}{l-\alpha_1(\eta)} < \frac{2}{2-k}.$$

Thus,

$$\lim_{\theta\to+\infty}\frac{1}{\log\theta}\log P_n^\theta(\eta) = \begin{cases} -\frac{n-\alpha_1(\eta)}{2}k, & \text{if } \frac{n-\alpha_1(\eta)}{l-\alpha_1(\eta)} \geq \frac{2}{2-k} \\ -(n-l), & \text{if } \frac{n-\alpha_1(\eta)}{l-\alpha_1(\eta)} < \frac{2}{2-k}. \end{cases}$$

$\square$

Theorem 4.6. *Assume that* $\lim_{\theta\to\infty}\theta t(\theta) = \infty$. *As* $\theta \to +\infty$, *for any given integer* $n \geq 2$, $P_n^{\theta,\alpha}$ *has the following LDPs*

- *If* $\lim_{\theta\to\infty}\frac{\theta t(\theta)}{\log\theta} = k \geq 2$, *then* $P_n^{\theta,\alpha}$ *has an LDP with speed* $\log\theta$ *and rate function* $I_n(\eta) = n - l(\eta)$.

- *If* $\lim_{\theta\to\infty}\frac{\theta t(\theta)}{\log\theta} = k \in (0,2)$, *then* $P_n^{\theta,\alpha}$ *has an LDP with speed* $\log\theta$ *and rate function*

$$I_n(\eta) = \begin{cases} 0, & \text{if } \eta = (1,1,\cdots,1) \\ \frac{n-\alpha_1(\eta)}{2}k, & \text{if } \frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)} \geq \frac{2}{2-k} \\ n-l(\eta), & \text{if } \frac{n-\alpha_1(\eta)}{l(\eta)-\alpha_1(\eta)} < \frac{2}{2-k}. \end{cases}$$

- *If* $\lim_{\theta\to\infty}\frac{\theta t(\theta)}{\log\theta}=0$, *then* $P_n^{\theta,\alpha}$ *has an LDP with speed* $\theta t(\theta)$ *and rate function*

$$I_n(\eta)=\frac{n-\alpha_1(\eta)}{2}.$$

Proof. Since

$$P_n^{\theta,\alpha}(\eta)=\frac{n!}{\eta_1!\cdots\eta_l!\alpha_1(\eta)!\cdots\alpha_n(\eta)!}Ep_\eta^o(X_{t(\theta)}^{\theta,\alpha})$$

and $X_{t(\theta)}^{\theta,\alpha}\to\delta_{(0,0,\cdots)}$, as $\theta\to+\infty$, obviously $P_n^{\theta,\alpha}(1,1,\cdots,1)\to 1$, as $\theta\to+\infty$. We only consider $\eta\neq(1,1,\cdots,1)$ and $\eta=(\eta_1,\cdots,\eta_l)$. By Proposition 2.2, we know

$$p_\eta^o=\sum_{d=1}^{l}(-1)^{l-d}\sum_{\beta\in\pi(l,d)}\varphi_{\sum_{i\in\beta_1}\eta_i}\cdots\varphi_{\sum_{i\in\beta_d}\eta_i}.$$

Let $(\sum_{i\in\beta_{(1)}}\eta_i,\cdots,\sum_{i\in\beta_{(d)}}\eta_i)$ be decreasing rearrangement of $(\sum_{i\in\beta_1}\eta_i,\cdots,\sum_{i\in\beta_d}\eta_i)$. We define $\eta^\beta=(\sum_{i\in\beta_{(1)}}\eta_i,\cdots,\sum_{i\in\beta_{(d)}}\eta_i)$, where if $\sum_{i\in\beta_j}\eta_i=1, 1\le j\le d$, we delete it, such that $\eta^\beta\in\tilde{\mathcal{M}}$. Then $Ep_\eta^o(X_{t(\theta)}^{\theta,\alpha})$ can be written as

$$\prod_{i=1}^{l}(1-\alpha)_{(\eta_i-1)}\frac{\prod_{u=0}^{l-1}(\theta+\alpha u)}{\theta_{(|\eta|)}} \tag{4.3.22}$$

$$+\sum_{d=1}^{l}(-1)^{l-d}\sum_{\beta\in\pi(l,d)}(|\beta_1|-1)!\cdots(|\beta_d|-1)!\varphi_{\eta^\beta}(x)e^{-\lambda_{|\eta^\beta|}t(\theta)} \tag{4.3.23}$$

$$+\sum_{d=1}^{l}(-1)^{l-d}\sum_{\beta\in\pi(l,d)}(|\beta_1|-1)!\cdots(|\beta_d|-1)!$$

$$\sum_{2\le\delta<\eta^\beta}\langle\varphi_{\eta^\beta},\psi_\delta^{\theta,\alpha}\rangle_{\theta,\alpha}\psi_\delta^{\theta,\alpha}(x)(e^{-\lambda_{|\delta|}t(\theta)}-e^{-\lambda_{|\eta^\beta|}t(\theta)}) \tag{4.3.24}$$

$$-\sum_{d=1}^{l}(-1)^{l-d}\sum_{\beta\in\pi(l,d)}(|\beta_1|-1)!\cdots(|\beta_d|-1)!\langle\varphi_{\eta^\beta},1\rangle_{\theta,\alpha}e^{-\lambda_{|\eta^\beta|}t(\theta)}. \tag{4.3.25}$$

By Lemma 4.8, we have

$$\langle\varphi_{\eta^\beta},\psi_\delta^{\theta,\alpha}\rangle_{\theta,\alpha}\sim\Bigg[\sum_{i=1}^{l(\eta^\beta)}\sum_{j=1}^{l(\delta)}\frac{(1-\alpha)_{(\eta_i^\beta+\delta_j-1)}}{(1-\alpha)_{(\eta_i^\beta-1)}(1-\alpha)_{(\delta_j-1)}}+l(\eta^\beta)\alpha-|\eta^\beta||\delta|$$

$$\prod_{i=1}^{l(\eta^\beta)}(1-\alpha)_{(\eta_i^\beta-1)}\prod_{j=1}^{l(\delta)}(1-\alpha)_{(\delta_j-1)}\frac{1}{\theta^{|\eta^\beta|-l(\eta^\beta)+|\delta|-l(\delta)+1}}\Bigg].$$

Since $|\eta^\beta| - l(\eta^\beta) = n - d$, and $|\delta| - l(\delta) \geq 1, \forall \delta \geq (2)$, we have

$$|\eta^\beta| - l(\eta^\beta) + |\delta| - l(\delta) + 1 \geq n - l + 2$$

and $|\eta^\beta| \geq n - \alpha_1(\eta)$. Then

$$\begin{aligned}
(4.3.23) \sim & p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}\theta t(\theta)}, \\
(4.3.24) \sim & \left[\sum_{d=1}^{l-\alpha_1(\eta)} \frac{(1-\alpha)_{(\eta_i+2-1)}}{(1-\alpha)_{(\eta_i-1)}(1-\alpha)_{(2-1)}} + (l-\alpha_1(\eta))\alpha - 2(|\eta| - \alpha_1(\eta)) \right] \\
& \prod_{i=1}^{l} (1-\alpha)_{(\eta_i-1)} \varphi_2(x) \frac{1}{\theta^{|\eta|-l+2}} e^{-\theta t(\theta)} \\
(4.3.25) \sim & - \prod_{i=1}^{l-\alpha_1(\eta)} (1-\alpha)_{(\eta_i-1)} \frac{1}{\theta^{|\eta|-l}} e^{-\frac{|\eta|-\alpha_1(\eta)}{2}\theta t(\theta)},
\end{aligned}$$

and

$$(4.3.22) \sim \prod_{i=1}^{l} (1-\alpha)_{(\eta_i-1)} \frac{1}{\theta^{n-l}}.$$

Similarly,

$$(4.3.22) + (4.3.24) + (4.3.25) \sim \prod_{i=1}^{l} (1-\alpha)_{(\eta_i-1)} \frac{1}{\theta^{n-l}};$$

therefore,

$$E p^o_\eta(X^{\theta,\alpha}_{t(\theta)}) \sim \prod_{i=1}^{l} (1-\alpha)_{(\eta_i-1)} \frac{1}{\theta^{n-l}} + p^o_{\eta_1,\cdots,\eta_{l-\alpha_1(\eta)}}(x) e^{-\frac{|\eta|-\alpha_1(\eta)}{2}\theta t(\theta)}.$$

The remaining argument is exactly the same as Theorem 4.5. □

In summary, to see how this system instantly evolves, we should consider the asymptotic behaviour at different time scale. As the time scale gets bigger, the system will get closer to state $(0, 0, \cdots)$. This process is very intricate. As stated in Theorem 4.5 and Theorem 4.6, this system will experience a phase transition, in which the critical scales should be $\frac{2\log\theta}{\theta}$. Once the time scale is greater than $\frac{2\log\theta}{\theta}$, this system will behave analogously as its stationary system.

Chapter 5

Asymptotic Theory for the One-parameter Selective Model

In Chapter 3, we have introduced the one-parameter neutral model, characterized by generator

$$G = \frac{1}{2}\sum_{i,j=1}^{\infty} x_i(\delta_{ij} - x_j)\frac{\partial^2}{\partial x_i \partial x_j} - \frac{\theta}{2}\sum_{i=1}^{\infty} x_i \frac{\partial}{\partial x_i},$$

and the one-parameter selective model, the generator of which is

$$G_\sigma = G + \sigma \sum_{i=1}^{\infty} x_i(x_i - \varphi_2(x))\frac{\partial}{\partial x_i}.$$

In the one-parameter neutral model, mutation and random sampling are considered. The former evolutionary force attracts the system towards its unique equilibrium point $(0, 0, \cdots)$; whereas the latter pushes the system away from it. Therefore, in this sense, the one-parameter neutral model resembles O-U processes although their stationary distributions are quite different from each other.

As we know, an infinite-dimensional O-U process is a more natural and convenient replacement of the Brownian motion in infinite-dimensional space. We can, therefore, formally regard the one-parameter neutral model as a "Brownian motion" on the Kingman simplex. Then the one-parameter selective model can be formally regarded as a model with a "Brownian noise" related to generator G and a "gradient system" associated with $\sum_{i=1}^{\infty} x_i(x_i - \varphi_2(x))\frac{\partial}{\partial x_i}$. It is not difficult to check that $\varphi_2(x)$ is a Lyapunov function of this "gradient system". Interestingly, $\varphi_2(x)$ has countably many local minima. If we decompose the Kingman simplex $\bar{\nabla}_\infty$ as $(\cup_{n=1}^{\infty} L_n) \cup L_0 \cup L_\infty$, where

$$L_0 = \left\{ x \in \bar{\nabla}_\infty \middle| \sum_{i=1}^{\infty} x_i < 1 \right\}, \quad L_\infty = \left\{ x \in \bar{\nabla}_\infty \middle| \sum_{i=1}^{\infty} x_i = 1, x_i > 0, \forall i \geq 1 \right\} \quad (5.0.1)$$

$$L_n = \left\{ x \in \bar{\nabla}_\infty \middle| \sum_{i=1}^{n} x_i = 1, x_n > 0, x_{n+1} = 0 \right\}, \quad n \geq 1, \tag{5.0.2}$$

then, for $n \geq 1$, the restriction of $\varphi_2(x)$ to L_n, denoted by $\varphi_2|_{L_n}(x)$, takes its minimum value at $(\frac{1}{n}, \cdots, \frac{1}{n}, 0, \cdots)$. Analogously $\varphi_2|_{L_0}(x)$ and $\varphi_2|_{L_\infty}(x)$ share a common minimum value 0, which is attained by $\varphi_2|_{L_0}(x)$ at $(0, 0, \cdots)$. Moreover, it is not hard to see that all these local minima are the equilibrium points of the "gradient system".

Due to the ergodicity of the one-parameter selective model, every neighbourhood of equilibrium point will be visited. It is naturally very tempting to ask how often each neighbourhood of equilibrium points will be visited.

The answer to the above question depends on the specific interactions between evolutionary forces. In [Gillespie, 1999], J.H. Gillespie conducted computer simulations of various selective models and neutral models; and conjectured that the neutral model and the selective model with symmetric overdominance ($\sigma < 0$) behave similarly when mutation and selection are both strong. This conjecture has been verified in [Joyce et al., 2003] and [Dawson and Feng, 2006] respectively.

We could have other types of interactions, such as the interaction between strong mutation and strong underdominant selection($\sigma > 0$), and the interaction between weak mutation and strong overdominant selection ($\sigma < 0$). In [Feng, 2009], the interaction between small mutation and large overdominant selection was considered. Much to our surprise, countably many phase transitions occurred under different selection intensities $\sigma = \lambda \log\theta, \lambda > 0, \theta \in [0, 1]$. The critical selection intensity is $\sigma = k(k+1)\log\theta, k \geq 1$. S. Feng observed that $\forall \lambda \in (k(k-1), k(k+1)), k \geq 1$, the system will most likely be around $(\frac{1}{k}, \cdots, \frac{1}{k}, 0, \cdots)$. But the case that $\lambda = k(k+1), k \geq 1$, is left open.

In this section, we will consider all sorts of interactions between mutation and selection. Especially, we will provide a positive answer to the critical case in [Feng, 2009]. Moreover, the transient distribution under different interactions between mutation and selection are also considered.

5.1 Asymptotic Results on Stationary Distributions

As is stated in Chapter 3, the stationary distribution of the one-parameter selective model is

$$\pi_\sigma = C_\sigma \exp\{\sigma\varphi_2(x)\}\mathrm{PD}(\theta)(dx), \quad x \in \bar{\nabla}_\infty.$$

Similarly to [Joyce et al., 2003] and [Dawson and Feng, 2006], we rewrite σ as $\lambda\alpha(\theta)$, where $\alpha(\theta) > 0$. If $\lambda > 0$, the associated model is termed a one-parameter selective model with symmetric underdominance; if $\lambda < 0$, the model is then called a one-parameter selective model with symmetric overdominance. When $\lambda = 0$, it reduces to the one-parameter neutral model.

5.1.1 Large Mutations and Large Selections

Let us take a look at the LDP associated with π_σ in [Dawson and Feng, 2006].

Proposition 5.1. *If $\sigma = \lambda\alpha(\theta), \lambda < 0$, then π_σ has an LDP with speed θ and rate function $I_\lambda(x)$. More specifically,*

$$I_\lambda(x) = \begin{cases} I(x), & \text{if } \lim_{\theta\to\infty} \frac{\alpha(\theta)}{\theta} = 0 \\ I^{'}(x), & \text{if } \lim_{\theta\to\infty} \frac{\alpha(\theta)}{\theta} = 1 \\ I^{''}(x), & \text{if } \lim_{\theta\to+\infty} \frac{\alpha(\theta)}{\theta} = +\infty, \end{cases}$$

where

$$I(x) = \begin{cases} \log \frac{1}{1-\sum_{i=1}^{\infty} x_i}, & \text{if } \sum_{i=1}^{\infty} x_i < 1 \\ +\infty, & \text{otherwise ,} \end{cases}$$

$$I^{'}(x) = \sup_{y\in\bar{\nabla}_\infty} \{\lambda\varphi_2(y) - I(y)\} - (\lambda\varphi_2(x) - I(x)),$$

$$I^{''}(x) = \begin{cases} 0, & \text{if } x \text{ is the maximum point of } -\varphi_2(x) \\ +\infty, & \text{otherwise.} \end{cases}$$

Remark 5.1. *This proposition corresponds to the one-parameter selective model with symmetric overdominance, in which the selection actually enhances the effects of mutations. Therefore the one-parameter selective model with symmetric overdominance behaves like the one-parameter neutral model. As a matter of fact, we can specify $I^{'}(x)$ by calculating $\sup_{y\in\bar{\nabla}_\infty}\{\lambda\varphi_2(y) - I(y)\}$. To this end, for $\lambda < 0$, we know*

$$\lambda\varphi_2(y) - I(y) \le 0, \text{ and } \lambda\varphi_2(0) - I(0) = 0.$$

So $\sup_{y\in\bar{\nabla}_\infty}\{\lambda\varphi_2(y) - I(y)\} = 0$. Thus $I^{'}(x) = I(x) - \lambda\varphi_2(x)$. Moreover, we can also specify $I^{''}(x)$ as

$$I^{''}(x) = \begin{cases} 0, & \text{if } x = (0, 0, \cdots) \\ +\infty, & \text{otherwise .} \end{cases}$$

For the one-parameter selective model with symmetric underdominance, the following result was communicated by Paul Joyce to the authors of [Dawson and Feng, 2006].

Proposition 5.2. *If $\sigma = \lambda\alpha(\theta), \lambda > 0$, then π_σ has an LDP with speed θ and rate* *function* $\tilde{I}_\lambda(x)$*. More specifically,*

$$\tilde{I}_\lambda(x) = \begin{cases} I(x), & \text{if } \lim_{\theta\to\infty}\frac{\alpha(\theta)}{\theta} = 0 \\ -\lambda\varphi_2(x) + I(x), & \text{if } \lim_{\theta\to\infty}\frac{\alpha(\theta)}{\theta} = 1, 0 < \lambda \le \lambda_0 \\ \log\left(\frac{1-\sqrt{1-\frac{2}{\lambda}}}{2}\right) + \lambda\left(\frac{1+\sqrt{1-\frac{2}{\lambda}}}{2}\right)^2 & \\ -\lambda\varphi_2(x) + I(x), & \text{if } \lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = 1, \lambda > \lambda_0, \\ \tilde{I}''(x), & \lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = +\infty, \end{cases}$$

where $\lambda_0 \approx 2.4554$ *and solves the transcendental equation*

$$\log\left(\frac{1-\sqrt{1-\frac{2}{\lambda}}}{2}\right) + \lambda\left(\frac{1+\sqrt{1-\frac{2}{\lambda}}}{2}\right)^2 = 0.$$

and

$$\tilde{I}''(x) = \begin{cases} 0, & x = (1, 0, \cdots) \\ \infty, & \textit{otherwise} \end{cases}$$

Proof. Applying Theorem 5.2 and Theorem 5.3 in [Dawson and Feng, 2006], we can easily see that π_σ has the LDP with speed θ and rate $\tilde{I}_\lambda(x), \lambda > 0$, where

$$\tilde{I}_\lambda(x) = \begin{cases} I(x), & \text{if } \lim_{\theta\to\infty}\frac{\alpha(\theta)}{\theta} = 0 \\ \tilde{I}'(x), & \text{if } \lim_{\theta\to\infty}\frac{\alpha(\theta)}{\theta} = 1 \\ \tilde{I}''(x), & \text{if } \lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = +\infty. \end{cases}$$

Furthermore, $\tilde{I}'(x) = \sup_{y\in\bar{\nabla}_\infty}\{\lambda\varphi_2(y) - I(y)\} - (\lambda\varphi_2(x) - I(x))$, we can specify it by calculating $\sup_{y\in\bar{\nabla}_\infty}\{\lambda\varphi_2(y) - I(y)\}$. One can rewrite it as

$$\sup_{0\le r<1}\{\lambda r^2 \sup_{y\in\nabla_\infty}\varphi_2(y) + \log(1-r)\} = \sup_{0\le r<1}\{\lambda r^2 + \log(1-r)\}.$$

Define $f_\lambda(r) = \lambda r^2 + \log(1-r), 0 \le r < 1$. Then $f_\lambda'(r) = 2\lambda r - \frac{1}{1-r} = -2\lambda\frac{r^2-r+\frac{1}{2\lambda}}{1-r}$. If $1 - 4\frac{1}{2\lambda} \le 0$, or equivalently $\lambda \le 2$, then $f_\lambda'(r) \le 0$. Thus, $\sup_{0\le r<1} f_\lambda(r) = f_\lambda(0) = 0$; then $I'(x) = -\lambda\varphi_2(x) + I(x)$. So $I'(x)$ thus has only one zero $(0, 0, \cdots)$.

If, however, $\lambda > 2$, then $f_\lambda'(r)$ should be

$$\begin{cases} \ge 0, & \text{if } \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{2\lambda}} \le r \le \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{2\lambda}} \\ < 0, & \text{if } r < \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{2\lambda}} \text{ or } r > \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{2\lambda}}; \end{cases}$$

therefore

$$\sup_{0\le r<1} f_\lambda(r) = \max\left\{f_\lambda(0), f_\lambda\left(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{2\lambda}}\right)\right\} = \max\left\{0, f_\lambda\left(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{2\lambda}}\right)\right\}.$$

Notice that

$$f_\lambda\left(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{2\lambda}}\right) = \log\left(\frac{1-\sqrt{1-\frac{2}{\lambda}}}{2}\right) + \lambda\left(\frac{1+\sqrt{1-\frac{2}{\lambda}}}{2}\right)^2,$$

which is a function of λ when $\lambda > 2$. Define $g(\lambda) = f_\lambda\left(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{2\lambda}}\right)$; and $v = 1 - \sqrt{1-\frac{2}{\lambda}} (\lambda > 2)$, where thus $v \in (0,1]$. Consider a function $h(v) = \log v + \frac{1}{v} - \log 2 - \frac{1}{2}$, which is decreasing in $(0,1]$. Since $h\left(1-\sqrt{1-\frac{2}{\lambda}}\right) = g(\lambda)$, we know $g(\lambda)$ is increasing when $\lambda > 2$. $\sup_{\lambda>2} g(\lambda) = g(+\infty) = +\infty$ and $\inf_{\lambda>2} = \log\frac{1}{2}+\frac{1}{2} \approxeq -0.193 < 0$. Therefore, When $\lambda > 2$, there exists a unique $\lambda_0 > 2$, such that $g(\lambda)$ is

$$\begin{cases} > 0, & \text{if } \lambda > \lambda_0 \\ = 0, & \text{if } \lambda = \lambda_0 \\ < 0, & \text{if } 2 < \lambda < \lambda_0, \end{cases}$$

where λ_0 solves the transcendental equation,

$$\log\left(\frac{1-\sqrt{1-\frac{2}{\lambda}}}{2}\right) + \lambda\left(\frac{1+\sqrt{1-\frac{2}{\lambda}}}{2}\right)^2 = 0.$$

Then

$$\sup_{0\le r<1} f_\lambda(r) = \begin{cases} 0, & \text{if } \lambda \le \lambda_0 \\ \log\left(\frac{1-\sqrt{1-\frac{2}{\lambda}}}{2}\right) + \lambda\left(\frac{1+\sqrt{1-\frac{2}{\lambda}}}{2}\right)^2, & \text{if } \lambda > \lambda_0. \end{cases}$$

Thus, the proof is completed. □

Remark 5.2. *From the above argument, we can easily see that $\tilde{I}_\lambda(x)$ has only one zero, which is*

$$\begin{cases} (0,0,\cdots), & \textit{if } \lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = 0, \\ & \textit{or } \lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = 1, 0 < \lambda \le \lambda_0; \\ \left(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{4\lambda}}, 0, \cdots\right), & \textit{if } \lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = 1, \lambda > \lambda_0 \\ (1,0,\cdots), & \textit{if } \lim_{\theta\to\infty}\frac{\alpha(\theta)}{\theta} = \infty \end{cases}$$

Now we know that, for large mutation and large selection, the LDP rate function for the one-parameter selective model, whether it has symmetric overdominance or underdominance, has a unique zero. By Proposition 2.5, the limiting distribution of π_σ can only be degenerate distribution concentrated at a single point.

Theorem 5.1. *When $\lambda > 0$, we have*

$$\pi_\sigma \to \begin{cases} \delta_{(0,0,\cdots)}, & \text{if } \lim_{\theta\to+\infty} \frac{\alpha(\theta)}{\theta} = 0, \\ & \text{or } \lim_{\theta\to+\infty} \frac{\alpha(\theta)}{\theta} = 1, 0 < \lambda \le \lambda_0 \\ \delta_{(\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{1}{4\lambda}},0,\cdots)}, & \text{if } \lim_{\theta\to+\infty} \frac{\alpha(\theta)}{\theta} = 1, \lambda > \lambda_0 \\ \delta_{(1,0,\cdots)}, & \text{if } \lim_{\theta\to\infty} \frac{\alpha(\theta)}{\theta} = \infty \end{cases}$$

when $\lambda < 0$, we have $\pi_\sigma \to \delta_{(0,0,\cdots)}$.

This theorem can be easily derived from the LDP for π_σ in [Dawson and Feng, 2006]. It also indicates that π_σ and PD(θ) share the same limit when $\sigma = \lambda\theta(\lambda < 0)$ and $\theta \to +\infty$. Gillespie's conjecture is true in this sense. For the one-parameter selective model with symmetric underdominance, however, the limit of its stationary distribution $\pi_\sigma(\sigma = \lambda\theta, \lambda > 0)$ exhibits a phase transition. If $\lambda \le \lambda_0$, π_σ and PD(θ) still share the same limit; but if $\lambda > \lambda_0$, asymptotically, there appears to be one dominant type. The reason that the weight of this dominant type is less than 1 is because there are still many other less dominant types continuously distributed. Their presence shall be detected in the labeled model.

5.1.2 Small Mutation and Large Selection

In [Feng, 2009], S. Feng considered the LDP for PD(θ) with small mutation. As an application, he considered the LDP for π_σ with small mutation and various selections. It turns out that the one-parameter selective model with symmetric overdominance exhibits a lot of interesting properties.

In this section, only the selection intensity $\sigma = \lambda \log\theta(\lambda > 0, 0 < \theta \le 1)$ is considered; and we denote π_σ by $\pi_{\lambda,\theta}$. The following two theorems are due to S. Feng.

Theorem 5.2. *The family $\{\text{PD}(\theta), \theta > 0\}$ satisfies an LDP with speed $(-\log\theta)$ and rate function*

$$J(x) = \begin{cases} 0, & x \in L_1 \\ n-1, & x \in L_n, n \ge 2 \\ \infty, & x \in L_0 \cup L_\infty, \end{cases}$$

where $L_n, n \ge 0$, are defined as in (5.0.1) and (5.0.2).

Theorem 5.3. *For fixed λ, $\{\pi_{\lambda,\theta}, \theta > 0\}$ satisfies an LDP with speed $(-\log\theta)$ and* *rate function*

$$S_\lambda(x) = J(x) + \lambda\varphi_2(x) - \inf\left\{\frac{\lambda}{n} + n - 1 : n \geq 1\right\}.$$

Remark 5.3. *The LDP for π_σ in [Feng, 2009] falls into three categories, only one of which is stated here. In this theorem, the LDP rate function $S_\lambda(x)$ has two zeros $(\frac{1}{k}, \cdots, \frac{1}{k}, 0, \cdots)$ and $(\frac{1}{k+1}, \cdots, \frac{1}{k+1}, 0, \cdots)$, when $\lambda = k(k+1), k \geq 1$; whereas it has only one zero, $(\frac{1}{k}, \cdots, \frac{1}{k}, 0, \cdots)$, when $(k-1)k < \lambda < k(k+1), k \geq 1$. Therefore, $\lambda = k(k+1), k \geq 1$ serves as critical selection intensities.*

Again, due to Proposition 2.5, the limiting distribution of $\pi_{\lambda,\theta}$ can be obtained except for the case $\lambda = k(k+1), k \geq 1$.

Proposition 5.3. *When $(k-1)k < \lambda < k(k+1), k \geq 1$, $\pi_{\lambda,\theta}$ converges weakly to $\delta_{(\frac{1}{k},\cdots,\frac{1}{k},0\cdots)}$, as $\theta \to 0$.*

In the critical case $\lambda = k(k+1), k \geq 1$, however, the limiting distribution of $\pi_{\lambda.\theta}$ can not be directly derived from its LDP; simply because its LDP rate function has two zeros, and Proposition 2.5 can not be applied. Luckily, theorem 5.5 provides a complete answer to the critical cases.

Theorem 5.4. *For $\lambda > 0$, the limiting distributions of homozygosity, H_2, under the distribution $\pi_{\theta,\lambda}$ is*

$$\sum_{k=1}^{+\infty} \mathrm{I}_{(k(k-1),k(k+1)]}(\lambda)\delta_{\frac{1}{k}}(dx).$$

Before we present the proof, we need the following lemmas, the proofs of which are postponed!

Lemma 5.1. *The moments of the heterozygosity $m_k = E(1-H_2)^k$ has the form*

$$m_k = \sum_{l=1}^{k} A_{k,l}(\theta)\theta^l,$$

where

$$A_{k,1}(\theta) = \frac{2k+\theta}{2k}\frac{2^k k!\Gamma(k+\theta)}{\Gamma(2k+1+\theta)}$$
$$A_{k,p}(\theta) = \sum_{l=p-1}^{k-1} \frac{2k+\theta}{2k}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l+\theta)}{\Gamma(2k+1+\theta)}A_{l,p-1}(\theta), \quad p \geq 2.$$

Let us define $A_{k,p} = A_{k,p}(0)$; then

$$A_{k,1} = \frac{2^k k!(k-1)!}{(2k)!},$$

and

$$A_{k,p} = \sum_{l=p-1}^{k-1} \frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)} A_{l,p-1}, \quad p \geq 2.$$

Thus $A_{k,p}$ does not depend on θ anymore; but it is an appropriate approximation of $A_{k,p}(\theta)$, as can be seen in the proof of Lemma 5.7.

Lemma 5.2. *If we fix integer $p \geq 1$, then, $\forall \theta \in [0,1]$, we have*

$$\begin{aligned} &\frac{1}{2^p} A_{k,p} \leq A_{k,p}(\theta) \leq A_{k,p} \leq \frac{1}{2^{p-2}}, && \forall k \geq p \geq 1; \\ &|A_{k,p}(\theta) - A_{k,p}| \leq \theta p A_{k,p}, && 1 \leq p \leq k. \end{aligned}$$

Lemma 5.3. *For $\lambda > 0$, we have*

$$\lim_{\theta \to 0} \sum_{l=[\lambda]+1}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta) = 0.$$

and

$$\lim_{\theta \to 0} \sum_{l=[\lambda]+1}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta) = 0.$$

Lemma 5.4. *Suppose that a_n, b_n are two positive sequences, and $\lim_{n\to\infty} \frac{a_n}{b_n} = c$. Then $\lim_{x \to +\infty} \frac{\sum_{n=0}^{\infty} a_n \frac{x^n}{n!}}{\sum_{n=0}^{\infty} b_n \frac{x^n}{n!}} = c$.*

Lemma 5.5. *For any fixed integer $p \geq 1$, we have, as $k \to +\infty$,*

$$A_{k,p} \sim C_p \frac{1}{k^{\frac{p}{2}}} (\frac{p}{p+1})^k,$$

where $C_1 = \sqrt{\pi}$, and $C_{p+1} = C_p \sqrt{\pi} (\frac{p+2}{p})^{\frac{p}{2}}$.

Lemma 5.6. *Define $C_{k,l} = \sum_{s=0}^{k-l} \binom{k}{s} (\frac{\lambda - l}{\lambda})^l A_{k-s,l}$. Then, as $k \to +\infty$,*

$$C_{k,l} \sim C_l (1 + \frac{(\lambda - l)(l+1)}{\lambda l})^{\frac{l}{2}} \frac{1}{k^{\frac{l}{2}}} (\frac{\lambda - l}{\lambda} + \frac{l}{l+1})^k.$$

Lemma 5.7. *For $\lambda > 2$, define*

$$K_n^{\lambda}(\theta) = \frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta)}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)}$$

and

$$\tilde{K}_n^{\lambda}(\theta) = \frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}}.$$

When $u(u-1) < \lambda \leq u(u+1), \quad u \geq 2$, *we have*

$$\lim_{\theta\to 0} K_n^\lambda(\theta) = \lim_{\theta\to 0} \tilde{K}_n^\lambda(\theta) = (\frac{u-1}{u})^n.$$

[PROOF OF THEOREM 5.4]:

Proof. Let us use ϕ_H to denote the moment generating function of the homozygosity H_2 under $\pi_{\lambda,\theta}$. Thus,

$$\phi_H(t) = \frac{E\exp\{tH_2\} \cdot \exp\{\lambda \log \theta H_2\}}{E\exp\{\lambda \log \theta H_2\}}.$$

For some technical reason, we need to multiply the numerator and denumerator in the above equation by the common term $\lambda \log \frac{1}{\theta}$; then

$$\begin{aligned}
\phi_H(t) =& e^t \frac{E\exp\{-t(1-H_2)\} \cdot \exp\{\lambda \log \frac{1}{\theta}(1-H_2)\}}{E\exp\{\lambda \log \frac{1}{\theta}(1-H_2)\}} \\
=& e^t \left(1 + \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \frac{E(1-H_2)^n \exp\{\lambda \log \frac{1}{\theta}(1-H_2)\}}{E\exp\{\lambda \log \frac{1}{\theta}(1-H_2)\}} \right) \\
=& e^t \left(1 + \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \frac{\sum_{m=0}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^m}{m!} E(1-H_2)^{n+m}}{\sum_{m=0}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^m}{m!} E(1-H_2)^m} \right).
\end{aligned}$$

In the above expansion, all terms are positive, which greatly facilitates our calculations. If we denote the limit of $\phi_H(t)$, as $\theta \to 0$, by $\psi_H(t)$, then we have

$$\begin{aligned}
\psi_H(t) = e^t \Bigg[1 + \lim_{\theta\to 0} \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \Bigg(& \frac{\sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=1}^{k} A_{k+n,l}(\theta)\theta^l}{1 + \sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=1}^{k} A_{k,l}(\theta)\theta^l} \\
& + \frac{m_n + \sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=k+1}^{k+n} A_{k+n,l}(\theta)\theta^l}{1 + \sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=1}^{k} A_{k,l}(\theta)\theta^l} \Bigg) \Bigg].
\end{aligned}$$

By the Lebesgue dominant convergence theorem, we can switch the order of summation and limit. Thus,

$$\begin{aligned}
\psi_H(t) = e^t \Bigg[1 + \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \Bigg(& \lim_{\theta\to 0} \frac{\sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=1}^{k} A_{k+n,l}(\theta)\theta^l}{1 + \sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=1}^{k} A_{k,l}(\theta)\theta^l} \\
& + \lim_{\theta\to 0} \frac{m_n + \sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=k+1}^{k+n} A_{k+n,l}(\theta)\theta^l}{1 + \sum_{k=1}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} \sum_{l=1}^{k} A_{k,l}(\theta)\theta^l} \Bigg) \Bigg]. \qquad (5.1.1)
\end{aligned}$$

Now we claim that

$$\lim_{\theta\to 0}\frac{m_n+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=1+k}^{k+n}A_{k+n,l}(\theta)\theta^l}{1+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=1}^{k}A_{k,l}(\theta)\theta^l}=0. \tag{5.1.2}$$

Indeed, we have

$$0\le\frac{m_n+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=1+k}^{k+n}A_{k+n,l}(\theta)\theta^l}{1+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=1}^{k}A_{k,l}(\theta)\theta^l}\le m_n+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=1+k}^{k+n}A_{k+n,l}(\theta)\theta^l.$$

By Lemma 5.2, we have

$$\begin{aligned}
m_n+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=1+k}^{k+n}A_{k+n,l}(\theta)\theta^l &\le m_n+\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\sum_{l=k+1}^{k+n}\frac{\theta^l}{2^{l-2}}\\
\le& m_n+4\sum_{k=1}^{\infty}\frac{(\lambda\frac{\theta}{2}\log\frac{1}{\theta})^k}{k!}(1-(\frac{\theta}{2})^n)\frac{\theta}{2-\theta}\\
=& m_n+4\frac{\theta}{2-\theta}(1-(\frac{\theta}{2})^n)(e^{\lambda\frac{\theta}{2}\log\frac{1}{\theta}}-1)\to 0,\ \text{as } \theta\to 0.
\end{aligned}$$

We have used the fact that $m_n\to 0$, as $\theta\to 0$, which is due to $\mathrm{PD}(\theta)(dx)\to\delta_{(1,0,\cdots)}(dx)$, as $\theta\to 0$. Thus, claim (5.1.2) is true. Therefore, by switching the summation order in (5.1.1), we have

$$\psi_H(t)=e^t\left(1+\sum_{n=1}^{\infty}\frac{(-t)^n}{n!}\lim_{\theta\to 0}\frac{\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)}{1+\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k,l}(\theta)}\right).$$

Now what we need to show is, for $u(u-1)<\lambda\le u(u+1), u\ge 1$,

$$\lim_{\theta\to 0}\frac{\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)}{1+\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k,l}(\theta)}=\left(\frac{u-1}{u}\right)^n. \tag{5.1.3}$$

Once we have got the above equation, then

$$\begin{aligned}
\psi_H(t)=&e^t\left(1+\sum_{n=1}^{\infty}\frac{(-t)^n}{n!}\Big(\frac{u-1}{u}\Big)^n\right)\\
=&e^t\left(1+\sum_{n=1}^{\infty}\frac{\Big(-\frac{t(u-1)}{u}\Big)^n}{n!}\right)=e^te^{-\frac{t(u-1)}{u}}=e^{\frac{t}{u}}.
\end{aligned}$$

Thus, $\psi_H(t)=\sum_{u=1}^{\infty}I_{(u(u-1),u(u+1)]}(\lambda)e^{\frac{t}{u}}$. Therefore, the limiting distribution of H_2

under $\pi_{\lambda,\theta}$ is $\sum_{u=1}^{\infty} I_{(u(u-1),u(u+1)]}(\lambda)\delta_{\frac{1}{u}}$.

Now we are going to verify the claim (5.1.3). Firstly, when $0<\lambda\leq 2$, we have

$$\begin{aligned}0\leq\lim_{\theta\to 0}&\frac{\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)}{1+\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k,l}(\theta)}\leq\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)\\ \leq&\theta\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,1}(\theta)+\theta^2\sum_{k=2}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,2}(\theta)\\ &+\sum_{l=3}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta).\end{aligned}$$

We can actually show that the above three terms approach 0 as $\theta\to 0$. Indeed, by Lemma 5.2, we have

$$\begin{aligned}0\leq&\sum_{l=3}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)\leq\sum_{l=3}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\frac{1}{2^{l-2}}\\ \leq&4\sum_{l=3}^{\infty}(\frac{\theta}{2})^l e^{\lambda\log\frac{1}{\theta}}=\frac{4}{\theta^{\lambda}}(\frac{\theta}{2})^3\frac{1}{1-\theta/2}=\frac{\theta^{3-\lambda}}{2-\theta}\to 0,\text{ as }\theta\to 0.\end{aligned}$$

And by Lemma 5.2, we have

$$\begin{aligned}0\leq&\theta\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,1}(\theta)\leq\theta\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,1}\\ \leq&\frac{\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,1}}{\sum_{k=0}^{\infty}\frac{(\log\frac{1}{\theta})^k}{k!}}\to 0,\qquad\text{as }\theta\to 0.\end{aligned}$$

The above limit is due to Lemma 5.5 and Lemma 5.4. Similarly,

$$\begin{aligned}0\leq&\theta^2\sum_{k=2}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,2}(\theta)\leq\theta^2\sum_{k=1}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,2}\\ \leq&\frac{\sum_{k=2}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,2}}{\sum_{k=0}^{\infty}\frac{(2\log\frac{1}{\theta})^k}{k!}}\to 0,\qquad\text{as }\theta\to 0.\end{aligned}$$

Thus, we have for $0<\lambda\leq 2$,

$$\lim_{\theta\to 0}\frac{\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)}{1+\sum_{l=1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k,l}(\theta)}=0=(\frac{1-1}{1})^n.$$

Secondly, for $u(u-1) < \lambda \leq u(u+1), u \geq 2$, then $\lambda > 2$. we can show that

$$\lim_{\theta \to 0} \frac{\sum_{l=1}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta)}{1 + \sum_{l=1}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)} = \lim_{\theta \to 0} K_n^{\lambda}(\theta); \tag{5.1.4}$$

then by Lemma 5.7, we have proved claim (5.1.3). Now we only need to verify (5.1.4). To this end, we rewrite

$$\frac{\sum_{l=1}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta)}{1 + \sum_{l=1}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)}$$

as

$$\frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta) + \sum_{l=1+[\lambda]}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta)}{1 + \sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta) + \sum_{l=1+[\lambda]}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)}.$$

By Lemma 5.3, we know, as $\theta \to 0$

$$\frac{\sum_{l=1+[\lambda]}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}(\theta)}{1 + \sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)} \to 0,$$

and

$$\frac{\sum_{l=1+[\lambda]}^{\infty} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)}{1 + \sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}(\theta)} \to 0.$$

Therefore, claim (5.1.4) is proved. Theorem 5.4 is thus proved! $\square$

Theorem 5.5. *For a given $\lambda > 0$, $\pi_{\lambda,\theta}$ converges weakly to the following*

$$\sum_{k=1}^{\infty} \mathrm{I}_{(k(k-1),k(k+1)]}(\lambda) \delta_{(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)}.$$

Proof. For integer $k \geq 1$, and $(k-1)k < \lambda < k(k+1)$, the limit of $\pi_{\lambda,\theta}$ has already been verified in Proposition 5.3. Therefore, we only need to consider the critical case $\lambda = (k+1)k, k \geq 1$. For a fixed $k \geq 1$, and $\lambda = (k+1)k$, we need to prove that $\pi_{\lambda,\theta}$ converges weakly to $\delta_{(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)}$.

For a given $f \in C(\bar{\nabla}_{\infty})$, one can conclude that, $\forall \epsilon > 0, \exists \delta < \frac{1}{k(k+1)+1}$, such that $\forall x \in B_{\delta}(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)$,

$$\left| f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right) \right| < \epsilon.$$

Thus,

$$
\begin{aligned}
&\Big| \int_{\nabla_\infty} f(x)\pi_{\lambda,\theta}(dx) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\Big| \\
=&\Big| \int_{\nabla_\infty} \left(f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right)\pi_{\lambda,\theta}(dx)\Big| \\
\le& \int_{\nabla_\infty} \left|f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right|\pi_{\lambda,\theta}(dx) \\
=& \int_{S_\lambda\ge\delta} \left|f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right|\pi_{\lambda,\theta}(dx) \\
&+ \int_{S_\lambda<\delta} \left|f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right|\pi_{\lambda,\theta}(dx) \\
\le& 2\|f\|_\infty \pi_{\lambda,\theta}(S_\lambda\ge\delta) + \int_{(S_\lambda<\delta)\cap(|\varphi_2-\frac{1}{k}|\ge\delta)} \left|f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right|\pi_{\lambda,\theta}(dx) \\
&+ \int_{(S_\lambda<\delta)\cap(|\varphi_2-\frac{1}{k}|<\delta)} \left|f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right|\pi_{\lambda,\theta}(dx) \\
\le& 2\|f\|_\infty \left[\pi_{\lambda,\theta}(S_\lambda\ge\delta) + \pi_{\lambda,\theta}\left(\left|\varphi_2-\frac{1}{k}\right|\ge\delta\right)\right] \\
&+ \int_{(S_\lambda<\delta)\cap(|\varphi_2-\frac{1}{k}|<\delta)} \left|f(x) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right|\pi_{\lambda,\theta}(dx)
\end{aligned}
$$

By the LDP for $\pi_{\lambda,\theta}$ and the weak convergence of H_2 under $\pi_{\lambda,\theta}$, we have

$$\lim_{\theta\to 0}\pi_{\lambda,\theta}(S_\lambda\ge\delta) = \lim_{\theta\to 0}\pi_{\lambda,\theta}\left(\left|\varphi_2-\frac{1}{k}\right|\ge\delta\right) = 0.$$

Moreover, we claim that

$$(S_\lambda<\delta)\cap\left(\left|\varphi_2-\frac{1}{k}\right|<\delta\right) \subset B_\delta\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right). \tag{5.1.5}$$

Then we have,

$$\limsup_{\theta\to 0}\left|\int_{\nabla_\infty} f(x)\pi_{\lambda,\theta}(dx) - f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right| \le \epsilon.$$

Letting $\epsilon\to 0$, we have

$$\lim_{\theta\to 0}\int_{\nabla_\infty} f(x)\pi_{\lambda,\theta}(dx) = f\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right).$$

Therefore, $\pi_{\lambda,\theta}$ converges weakly to $\delta_{(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)}$. Now we need to show the claim

(5.1.5). For $\lambda = (k+1)k$, since

$$S_\lambda(x) = J(x) + (k+1)k\varphi_2(x) - \inf_{n\geq 1}\left\{\frac{(k+1)k}{n} + n - 1\right\},$$

clearly

$$S_\lambda|_{L_n}(x) = n - 1 + k(k+1)\varphi_2|_{L_n}(x) - 2k.$$

Because $\varphi_2|_{L_n}(x)$ has a unique minimum point $(\frac{1}{n}, \cdots, \frac{1}{n}, 0, \cdots)$; then

$$S_\lambda\Big|_{L_n}(x) \geq n - 1 + \frac{k(k+1)}{n} - 2k.$$

By the monotonicity of the righthand function in n, we know it attains its minimum at k and $k+1$. Since $\delta < \frac{1}{k(k+1)+1} < \frac{2}{k+2}$, one can see, $\forall n \neq k, k+1$,

$$S_\lambda|_{L_n}(x) \geq \min\left\{k - 1 - 1 + \frac{k(k+1)}{k-1} - 2k, k + 2 - 1 + \frac{k(k+1)}{k+2} - 2k\right\} = \frac{2}{k+2} > \delta.$$

Then $(S_\lambda < \delta) = (S_\lambda < \delta) \cap (L_k \cup L_{k+1})$. Thus,

$$\begin{aligned}&(S_\lambda < \delta) \cap \left(\left|\varphi_2 - \frac{1}{k}\right| < \delta\right)\\ =&\left[(S_\lambda < \delta) \cap \left(\left|\varphi_2 - \frac{1}{k}\right| < \delta\right) \cap L_k\right] \cup \left[(S_\lambda < \delta) \cap \left(\left|\varphi_2 - \frac{1}{k}\right| < \delta\right) \cap L_{k+1}\right].\end{aligned}$$

But $S_\lambda|_{L_{k+1}} = k(k+1)(\varphi_2|_{L_{k+1}} - \frac{1}{k+1})$, then $(S_\lambda < \delta) \cap L_{k+1} = (\varphi_2 < \frac{\delta}{k(k+1)} + \frac{1}{k+1}) \cap L_{k+1}$. Since $\delta < \frac{1}{k(k+1)+1}$, then $\frac{\delta}{k(k+1)} + \frac{1}{k+1} < \frac{1}{k} - \delta$, thus $\forall x \in L_{k+1} \cap (|\varphi_2 - \frac{1}{k}| < \delta) \cap (S_\lambda < \delta)$, we have

$$\varphi_2(x) < \frac{\delta}{k(k+1)} + \frac{1}{k+1} < \frac{1}{k} - \delta < \varphi_2(x).$$

Therefore, $(S_\lambda < \delta) \cap (|\varphi_2 - \frac{1}{k}| < \delta) \cap L_{k+1} = \emptyset$, hence,

$$(S_\lambda < \delta) \cap \left(\left|\varphi_2 - \frac{1}{k}\right| < \delta\right) = (S_\lambda < \delta) \cap \left(\left|\varphi_2 - \frac{1}{k}\right| < \delta\right) \cap L_k.$$

Since $\forall x \in L_k \cap \left(|\varphi_2(x) - \frac{1}{k}| < \delta\right)$, we have $\frac{1}{k} - \delta < \sum_{i=1}^k x_i^2 < \frac{1}{k} + \delta$, and

$$\frac{1}{k^2} - \frac{\delta}{k} \leq \frac{\sum_{i=1}^k x_i^2}{k} < \frac{1}{k^2} + \frac{\delta}{k}.$$

Therefore, $\sqrt{\frac{1}{k^2}-\frac{\delta}{k}} < \min_{1\le i\le k} x_i \le \max_{1\le i\le k} x_i \le \sqrt{\frac{1}{k^2}+\frac{\delta}{k}}$. Then

$$\begin{aligned}
d\left(x,\left(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots\right)\right) &= \sum_{i=1}^{k}\frac{|x_i-1/k|}{2^i}\\
&\le \sum_{i=1}^{k}\frac{1}{2^i}\max_{1\le i\le k}|x_i-\frac{1}{k}| = \left(1-\frac{1}{2^k}\right)\max_{1\le i\le k}|x_i-\frac{1}{k}|\\
&\le \max\left\{\left|\sqrt{\frac{1}{k^2}-\frac{\delta}{k}}-\frac{1}{k}\right|,\left|\sqrt{\frac{1}{k^2}+\frac{\delta}{k}}-\frac{1}{k}\right|\right\}\\
&= \max\left\{\frac{\frac{\delta}{k}}{\sqrt{\frac{1}{k^2}-\frac{\delta}{k}}+\frac{1}{k}},\frac{\frac{\delta}{k}}{\sqrt{\frac{1}{k^2}+\frac{\delta}{k}}+\frac{1}{k}}\right\}\\
&< \delta.
\end{aligned}$$

Therefore $(S_\lambda < \delta)\cap(|\varphi_2-\frac{1}{k}|<\delta)\subset B_\delta(\frac{1}{k},\cdots,\frac{1}{k},0,\cdots)$, the claim (5.1.5) is thus proved. □

Now we will embark on a long journey to prove the previous lemmas.
[PROOF OF LEMMA 5.1]:

Proof. Define $V_1 = U_1, V_i = (1-U_1)\cdots(1-U_{i-1})U_i, i\ge 2$, where $\{U_i, i\ge 1\}$ are i.i.d. Beta(1,θ). Then $(V_1,V_2,\cdots)$ follows the GEM distribution. Denote $(V_{(1)},V_{(2)},\cdots)$ the order statistics of $(V_1,V_2,\cdots)$ by decreasing order; then $(V_{(1)},V_{(2)},\cdots)$ follows the Poisson-Dirichlet distribution. Since $H_2=\sum_{i=1}^\infty V_{(i)}^2=\sum_{i=1}^\infty V_i^2$, one can observe that $1-H_2=(1-U_1^2)-(1-U_1)^2+(1-U_1)^2(1-\tilde{H}_2)=2U_1(1-U_1)+(1-U_1)^2(1-\tilde{H}_2)$, where $\tilde{H}_2=\sum_{i=1}^\infty \tilde{V}_i^2$, and $\tilde{V}_1=U_2,\tilde{V}_i=(1-U_2)\cdots(1-U_i)U_{i+1}$. We can see that $(\tilde{V}_1,\tilde{V}_2,\cdots)$ follows the GEM distribution as well and is independent of U_1. Thus, $E(1-H_2)^k=E(1-\tilde{H}_2)^k$, and

$$\begin{aligned}
m_k =& E(1-H_2)^k = \left(2U_1(1-U_1)+(1-U_1)^2(1-\tilde{H}_2)\right)^k\\
=& E\sum_{l=0}^{k}\binom{k}{l}\left(2U_1(1-U_1)\right)^{k-l}(1-\tilde{H}_2)^l\\
=& \sum_{l=0}^{k}\binom{k}{l}E\left(2U_1(1-U_1)\right)^{k-l}E(1-\tilde{H}_2)^l\\
=& \sum_{l=0}^{k}\binom{k}{l}\frac{2^{k-l}\Gamma(k-l+1)\Gamma(k+l+\theta)\theta}{\Gamma(2k+1+\theta)}m_l.
\end{aligned}$$

If we isolate m_k, we have

$$m_k = \theta \sum_{l=1}^{k-1} \frac{2k+\theta}{2k} \frac{2^k k!}{2^l l!} \frac{\Gamma(k+l+\theta)}{\Gamma(2k+1+\theta)} m_l + \frac{2k+\theta}{2k} \frac{2^k k! \Gamma(k+\theta)}{\Gamma(2k+1+\theta)} \theta, \tag{5.1.6}$$

where $k \geq 2$, and $m_1 = \frac{\theta}{1+\theta}$. We claim that m_k has the following expansion

$$m_k = \sum_{l=1}^{k} A_{k,l}(\theta)\theta^l,$$

where

$$A_{k,l}(\theta) = \sum_{u=l-1}^{k-1} \frac{2k+\theta}{2k} \frac{2^k k!}{2^u u!} \frac{\Gamma(k+u+\theta)}{\Gamma(2k+1+\theta)} A_{u,l-1}(\theta), l \geq 2; A_{k,1}(\theta) = \frac{2k+\theta}{2k} \frac{2^k k! \Gamma(k+\theta)\theta}{\Gamma(2k+\theta+1)}.$$

Indeed, for $k = 1$, this is obvious. Assume that m_{k-1} has the above expression, then for m_k, by (5.1.6), we have

$$\begin{aligned} &\theta \sum_{l=1}^{k-1} \frac{2k+\theta}{2k} \frac{2^k k!}{2^l l!} \frac{\Gamma(k+l+\theta)}{\Gamma(2k+1+\theta)} \sum_{u=1}^{l} A_{l,u}(\theta)\theta^u + \frac{2k+\theta}{2k} \frac{2^k k! \Gamma(k+\theta)}{\Gamma(2k+1+\theta)} \theta \\ =& \sum_{u=1}^{k-1} \left[\sum_{l=u}^{k-1} \frac{2k+\theta}{2k} \frac{2^k k!}{2^l l!} \frac{\Gamma(k+l+\theta)}{\Gamma(2k+1+\theta)} A_{l,u}(\theta) \right] \theta^{u+1} + \frac{2k+\theta}{2k} \frac{2^k k! \Gamma(k+\theta)}{\Gamma(2k+1+\theta)} \theta. \end{aligned}$$

Let us denote $p = u+1$; then we have $m_k = \sum_{p=1}^{k} A_{k,p}(\theta)\theta^p$, where

$$\begin{aligned} A_{k,1}(\theta) =& \frac{2k+\theta}{2k} \frac{2^k k! \Gamma(k+\theta)}{\Gamma(2k+1+\theta)}, \\ A_{k,p}(\theta) =& \sum_{l=p-1}^{k-1} \frac{2k+\theta}{2k} \frac{2^k k!}{2^l l!} \frac{\Gamma(k+l+\theta)}{\Gamma(2k+1+\theta)} A_{l,p-1}(\theta), \quad p \geq 2. \end{aligned}$$

□

[PROOF OF LEMMA 5.2]:

Proof. We use mathematical induction with respect to p to show these conclusions.
Step1: We are going to show $\frac{1}{2^{p+1}} A_{k,p+1} \leq A_{k,p+1}(\theta) \leq A_{k,p+1}$.
When $p = 1$, we have, $\forall \theta \in [0,1]$,

$$A_{k,1}(\theta) = \frac{2k+\theta}{2k} \frac{2^k k!}{(2k+\theta)\cdots(k+\theta)} = \frac{2^{k-1}(k-1)!}{(2k-1+\theta)\cdots(k+\theta)}$$

$$\leq \frac{2^{k-1}(k-1)!}{(2k-1)\cdots k} = \frac{2^k(k-1)!k!}{(2k)!} = A_{k,1},$$

and

$$A_{k,1}(\theta) = \frac{2^{k-1}(k-1)!}{(2k-1+\theta)\cdots(k+\theta)} \geq \frac{2^{k-1}(k-1)!}{2k\cdots(k+1)} = \frac{1}{2}A_{k,1}.$$

Moreover, $A_{k,1} = \frac{2^k k!(k-1)!}{(2k)!} \leq 1 < \frac{1}{2^{1-2}} = 2$; therefore,

$$\frac{1}{2}A_{k,1} \leq A_{k,1}(\theta) \leq A_{k,1} \leq \frac{1}{2^{1-2}}.$$

Now we assume that $A_{k,p}(\theta)$ satisfies the inequality

$$\frac{1}{2^p}A_{k,p} \leq A_{k,p}(\theta) \leq A_{k,p} \leq \frac{1}{2^{p-2}}, k \geq p \geq 1. \tag{5.1.7}$$

Since $A_{k,p+1}(\theta) = \sum_{l=p}^{k-1} \frac{1}{2k}\frac{2^k k!}{2^l l!}\frac{A_{l,p}(\theta)}{(2k-1+\theta)\cdots(k+l+\theta)}$, by assumption (5.1.7), we have

$$\frac{1}{2^p}\sum_{l=p}^{k-1}\frac{1}{2k}\frac{2^k k!}{2^l l!}\frac{A_{l,p}}{(2k-1+1)\cdots(k+l+1)} \leq A_{k,p+1}(\theta)$$
$$\leq \sum_{l=p}^{k-1}\frac{1}{2k}\frac{2^k k!}{2^l l!}\frac{A_{l,p}}{(2k-1)\cdots(k+l)}.$$

Thus,

$$\frac{1}{2^p}\sum_{l=p}^{k-1}\frac{k+l}{2k}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l)}{\Gamma(2k+1)}A_{l,p} \leq A_{k,p+1}(\theta) \leq \sum_{l=p}^{k-1}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l)}{\Gamma(2k+1)}A_{l,p} = A_{k,p+1}.$$

But $\frac{k+l}{2k} > \frac{k}{2k} = \frac{1}{2}$, then

$$\frac{1}{2^p}\sum_{l=p}^{k-1}\frac{k+l}{2k}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l)}{\Gamma(2k+1)}A_{l,p} \geq \frac{1}{2^{p+1}}\sum_{l=p}^{k-1}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l)}{\Gamma(2k+1)}A_{l,p} = \frac{1}{2^{p+1}}A_{k,p+1}.$$

Hence,

$$\frac{1}{2^{p+1}}A_{k,p+1} \leq A_{k,p+1}(\theta) \leq A_{k,p+1}.$$

Step 2: We are going to show

$$A_{k,p} \leq \frac{1}{2^{p-2}}, \forall k \geq p \geq 1. \tag{5.1.8}$$

Since $A_{k,1} \leq \frac{1}{2^{1-2}}$, we can assume that $A_{k,p} \leq \frac{1}{2^{p-2}}, k \geq p$. So

$$A_{k,p+1} = \sum_{l=p}^{k-1} \frac{2^k k!}{2^l l!} \frac{\Gamma(k+l)}{\Gamma(2k+1)} A_{l,p} \leq \frac{1}{2^{p-2}} \sum_{l=p}^{k-1} \frac{2^k k! \Gamma(k+l)}{2^l l! \Gamma(2k+1)},$$

where we will show that $\sum_{l=p}^{k-1} \frac{2^k k! \Gamma(k+l)}{2^l l! \Gamma(2k+1)} < \frac{1}{2}$. Therefore, $A_{k,p+1} \leq \frac{1}{2^{p-1}}$, and (5.1.8) is thus proved. Next, to show $\sum_{l=p}^{k-1} \frac{2^k k! \Gamma(k+l)}{2^l l! \Gamma(2k+1)} < \frac{1}{2}$, let us define $B_{k,l} = \frac{2^k k! \Gamma(k+l)}{2^l l! \Gamma(2k+1)}, \quad k-1 \geq l \geq p \geq 2$, where $B_{k,l}$ is increasing in l only if $l < k-2$. Because when $l < k-2$

$$\frac{B_{k,l+1}}{B_{k,l}} = \frac{2^l l! \Gamma(k+l+1)}{2^{l+1}(l+1)! \Gamma(k+l)} = \frac{k+l}{2(l+1)} > 1.$$

Therefore, $B_{k,k-2}$ or $B_{k,k-1}$ should be the maximum term. Since

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

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

we obtain that

$$\sum_{l=p}^{k-1} \frac{2^k k! \Gamma(k+l)}{2^l l! \Gamma(2k+1)} \leq \frac{k-p}{2k-1} = \frac{k-p}{2(k-p)+2p-1} < \frac{1}{2}.$$

Step 3: We are going to show

$$\Big| A_{k,p}(\theta) - A_{k,p} \Big| \leq \theta p A_{k,p}.$$

When $p = 1$, $\Big| A_{k,1}(\theta) - A_{k,1} \Big| \leq \theta A_{k,1}$, for

$$\begin{aligned}
\Big| A_{k,1}(\theta) - A_{k,1} \Big| =& \Big| \frac{2^k k! \Gamma(k+\theta)}{2k\Gamma(2k+\theta)} - \frac{2^k k! \Gamma(k)}{\Gamma(2k+1)} \Big| \\
=& \Big| \frac{2^k k! \Gamma(k)}{\Gamma(2k+1)} \Big(\frac{\Gamma(2k)\Gamma(k+\theta)}{\Gamma(2k+\theta)\Gamma(k)} - 1 \Big) \Big| \\
\leq& A_{k,1} \Big| \frac{(2k-1)\cdots k}{(2k+\theta-1)\cdots(k+\theta)} - 1 \Big| \\
=& A_{k,1} \Big| (1 - \frac{\theta}{2k+\theta-1}) \cdots (1 - \frac{\theta}{k+\theta}) - 1 \Big|
\end{aligned}$$

$$
\begin{aligned}
=&\theta A_{k,1}\Big|\sum_{u=1}^{k}(-1)^u \sum_{k\le l_1<\cdots<l_u\le 2k-1}\frac{\theta^{u-1}}{(l_1+\theta)\cdots(l_u+\theta)}\Big|\\
\le&\theta A_{k,1}\sum_{u=1}^{k}\sum_{k\le l_1<\cdots<l_u\le 2k-1}\frac{\theta^{u-1}}{(l_1+\theta)\cdots(l_u+\theta)}\\
\le&\theta A_{k,1}\sum_{u=1}^{k}\sum_{k\le l_1<\cdots<l_u\le 2k-1}\frac{1}{l_1\cdots l_u}\\
=&\theta A_{k,1}\Big|(1+\frac{1}{2k-1})\cdots(1+\frac{1}{k})-1\Big|\\
=&\theta A_{k,1}\Big|\frac{2k}{2k-1}\frac{2k-1}{2k-2}\cdots\frac{k+1}{k}-1\Big| = \theta A_{k,1}|2-1| = \theta A_{k,1}.
\end{aligned}
$$

Therefore we assume that

$$
\Big|A_{k,p}(\theta)-A_{k,p}\Big| \le \theta p A_{k,p}, \tag{5.1.9}
$$

then, for $\Big|A_{k,p+1}(\theta)-A_{k,p+1}\Big|$, we have

$$
\begin{aligned}
\Big|A_{k,p+1}(\theta)-A_{k,p+1}\Big| =&\Big|\sum_{l=p}^{k-1}\frac{2k+\theta}{2k}\frac{2^k k!\Gamma(k+l+\theta)}{2^l l!\Gamma(2k+1+\theta)}A_{l,p}(\theta)-\sum_{l=p}^{k-1}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p}\Big|\\
\le&\Big|\sum_{l=p}^{k-1}\frac{2k+\theta}{2k}\frac{2^k k!\Gamma(k+l+\theta)}{2^l l!\Gamma(2k+1+\theta)}(A_{l,p}(\theta)-A_{l,p})\Big|\\
&+\Big|\sum_{l=p}^{k-1}\frac{2k+\theta}{2k}\frac{2^k k!\Gamma(k+l+\theta)}{2^l l!\Gamma(2k+1+\theta)}A_{l,p}-\sum_{l=p}^{k-1}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p}\Big|\\
\le&\sum_{l=p}^{k-1}\frac{2k+\theta}{2k}\frac{2^k k!\Gamma(k+l+\theta)}{2^l l!\Gamma(2k+1+\theta)}\Big|A_{l,p}(\theta)-A_{l,p}\Big|\\
&+\sum_{l=p}^{k-1}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p}\Big|\frac{\Gamma(k+l+\theta)\Gamma(2k)}{\Gamma(k+l)\Gamma(2k+\theta)}-1\Big|.
\end{aligned}
$$

By the assumption (5.1.9), we have

$$
\begin{aligned}
\Big|A_{k,p+1}(\theta)-A_{k,p+1}\Big| \le&\theta p\sum_{l=p}^{k-1}\frac{2^k k!\Gamma(k+l+\theta)}{2^l l! 2k\Gamma(2k+\theta)}A_{l,p}\\
&+\sum_{l=p}^{k-1}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p}\Big|\frac{\Gamma(k+l+\theta)\Gamma(2k)}{\Gamma(2k+\theta)\Gamma(k+l)}-1\Big|,
\end{aligned}
$$

where

$$\begin{aligned}
\Big|\frac{\Gamma(k+l+\theta)\Gamma(2k)}{\Gamma(2k+\theta)\Gamma(k+l)}-1\Big| =&\Big|\frac{(2k-1)\cdots(k+l)}{(2k+\theta-1)\cdots(k+l+\theta)}-1\Big| \\
=&\Big|(1-\frac{\theta}{2k-1+\theta})\cdots(1-\frac{\theta}{k+l+\theta})-1\Big| \\
=&\Big|\sum_{u=1}^{k-l}(-1)^u\sum_{k+l\le l_1<\cdots<l_u\le 2k-1}\frac{\theta^u}{(l_1+\theta)\cdots(l_u+\theta)}\Big| \\
\le&\sum_{u=1}^{k-l}\sum_{k+l\le l_1<\cdots<l_u\le 2k-1}\frac{\theta^u}{(l_1+\theta)\cdots(l_u+\theta)} \\
\le&\theta\sum_{u=1}^{k-l}\sum_{k+l\le l_1<\cdots<l_u\le 2k-1}\frac{1}{l_1\cdots l_u} \\
=&\theta\Big|(1+\frac{1}{2k-1})\cdots(1+\frac{1}{k+l})-1\Big| \\
=&\theta\frac{k-l}{k+l}<\theta,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\Gamma(k+l+\theta)}{2k\Gamma(2k+\theta)}=&\frac{1}{2k(2k+\theta-1)\cdots(k+l+\theta)} \\
\le&\frac{1}{2k(2k-1)\cdots(k+l)}=\frac{\Gamma(k+l)}{\Gamma(2k)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Big|A_{k,p+1}(\theta)-A_{k,p+1}\Big| \le&\theta p\sum_{l=p}^{k-1}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l)}{\Gamma(2k+1)}A_{l,p}+\theta\sum_{l=p}^{k-1}\frac{2^k k!}{2^l l!}\frac{\Gamma(k+l)}{\Gamma(2k+1)}A_{l,p} \\
=&\theta(p+1)A_{k,p+1}.
\end{aligned}$$

Thus, we have proved the lemma.

□

[PROOF OF LEMMA 5.3]:

Proof. By Lemma 5.2, we have

$$\sum_{l=[\lambda]+1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k,l}(\theta)\le\sum_{l=[\lambda]+1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\frac{1}{2^{l-2}}$$

$$
\begin{aligned}
&=4\sum_{l=[\lambda]+1}^{\infty}\frac{\theta^l}{2^l}\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}\\
&\leq 4\sum_{l=[\lambda]+1}^{\infty}(\frac{\theta}{2})^l e^{\lambda\log\frac{1}{\theta}} = 4\sum_{l=[\lambda]+1}^{\infty}\frac{\theta^{l-\lambda}}{2^l} = 4\frac{\theta^{[\lambda]-\lambda+1}}{2^{[\lambda]+1}}\frac{2}{2-\theta}\\
&\to 0,\ \text{as } \theta\to 0,\ \text{due to } [\lambda]+1>\lambda.
\end{aligned}
$$

Similarly, we can also show

$$
\lim_{\theta\to 0}\sum_{l=[\lambda]+1}^{\infty}\theta^l\sum_{k=l}^{\infty}\frac{(\lambda\log\frac{1}{\theta})^k}{k!}A_{k+n,l}(\theta)=0.
$$

$\square$

[PROOF OF LEMMA 5.5]:

Proof. By mathematical induction with respect on p, we can prove this lemma. For $p=1$, by Stirling's formula,

$$
\Gamma(z)\sim\sqrt{\frac{2\pi}{z}}(\frac{z}{e})^z, \tag{5.1.10}
$$

we have $A_{k,1}\sim\sqrt{\pi}\frac{1}{\sqrt{k}}(\frac{1}{2})^k$, as $k\to+\infty$. We can therefore assume that, as $k\to+\infty$,

$$
A_{k,p}\sim C_p\frac{1}{k^{\frac{p}{2}}}(\frac{p}{p+1})^k,\quad C_p=C_{p-1}\sqrt{\pi}(\frac{p+1}{p-1})^{\frac{p-1}{2}}. \tag{5.1.11}
$$

For $A_{k,p+1}$, we have

$$
A_{k,p+1}=\sum_{l=p}^{k-1}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p}.
$$

By the assumption (5.1.11), $\forall\epsilon>0,\exists M>0$, such that $\forall k>M$,

$$
1-\epsilon<\frac{A_{k,p}}{C_p\frac{1}{k^{\frac{p}{2}}}(\frac{p}{p+1})^k}<1+\epsilon;
$$

then we rewrite $A_{k,p+1}$ as $X+Y$, where

$$
X=\sum_{l=p}^{M}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p},\qquad Y=\sum_{l=M+1}^{k-1}\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}A_{l,p}.
$$

Define $a_k(l)=\frac{2^k k!\Gamma(k+l)}{2^l l!\Gamma(2k+1)}(\frac{p}{p+1})^l$, and $\Sigma_1=\sum_{l=p}^{k-1}\frac{C_p}{l^{\frac{p}{2}}}a_k(l)$. Now we are going to show $\lim_{k\to+\infty}\frac{X}{\Sigma_1}=0$, and $\lim_{k\to+\infty}\frac{Y}{\Sigma_1}=1$.

Indeed, since

$$0 \leq X \leq \frac{\max_{p\leq l\leq M}\{A_{l,p}\}(M-p+1)}{2^p p!}\frac{2^k k!\Gamma(k+M)}{\Gamma(2k+1)}$$

and

$$(1-\epsilon)\sum_{l=M+1}^{k-1}\frac{C_p}{l^{\frac{p}{2}}}a_k(l) \leq Y \leq (1+\epsilon)\sum_{l=M+1}^{k-1}\frac{C_p}{l^{\frac{p}{2}}}a_k(l),$$

we have

$$\begin{aligned}0 \leq & \frac{X}{\Sigma_1} \leq \frac{X}{\frac{C_p}{(k-1)^{\frac{p}{2}}}a_k(k-1)} \leq \frac{\max_{p\leq l\leq M}\{A_{l,p}\}(M-p+1)}{2^p p!}(k-1)^{\frac{p}{2}}\frac{2^k k!\Gamma(k+M)}{\Gamma(2k+1)a_k(k-1)} \\ \leq & \frac{\max_{p\leq l\leq M}\{A_{l,p}\}(M-p+1)}{2^p p!}(k-1)^{\frac{p}{2}}(\frac{2(p+1)}{p})^{k-1}\frac{\Gamma(k+M)\Gamma(k)}{\Gamma(2k-1)}.\end{aligned}$$

By Stirling's formula (5.1.10), we have, as $k \to +\infty$,

$$(k-1)^{\frac{p}{2}}(\frac{2(p+1)}{p})^{k-1}\frac{\Gamma(k+M)\Gamma(k)}{\Gamma(2k-1)} \sim k^{M+\frac{p+1}{2}}(\frac{p+1}{2p})^{k-1} \to 0. \tag{5.1.12}$$

Thus, $\lim_{k\to+\infty}\frac{X}{\Sigma_1} = 0$. Similarly,

$$0 \leq \sum_{l=p}^{M}\frac{C_p}{l^{\frac{p}{2}}}a_k(l) \leq \frac{C_p(M-p+1)}{p^{\frac{p}{2}}2^p p!}(\frac{p}{p+1})^p\frac{2^k k!\Gamma(k+M)}{\Gamma(2k+1)},$$

then

$$0 \leq \frac{\sum_{l=p}^{M}\frac{C_p}{l^{\frac{p}{2}}}a_k(l)}{\Sigma_1} \leq \frac{(M-p+1)(\frac{p}{p+1})^p}{p^{\frac{p}{2}}2^p p!}(k-1)^{\frac{p}{2}}(\frac{2(p+1)}{p})^{k-1}\frac{\Gamma(k+M)\Gamma(k)}{\Gamma(k-1)}.$$

Thus, $\frac{\sum_{l=p}^{M}\frac{C_p}{l^{\frac{p}{2}}}a_k(l)}{\Sigma_1} \to 0$ due to (5.1.12) . Therefore, for the following inequality,

$$(1-\epsilon)\left(1-\frac{\sum_{l=p}^{M}\frac{C_p}{l^{\frac{p}{2}}}a_k(l)}{\Sigma_1}\right) \leq \frac{Y}{\Sigma_1} \leq (1+\epsilon)\left(1-\frac{\sum_{l=p}^{M}\frac{C_p}{l^{\frac{p}{2}}}a_k(l)}{\Sigma_1}\right),$$

if we let $k \to +\infty$, and $\epsilon \to 0$, then we have

$$\lim_{k\to\infty}\frac{Y}{\Sigma_1} = 1;$$

hence $A_{k,p+1} \sim \Sigma_1$. Moreover,

$$a_k(l) = \frac{2^k k!(k-1)!}{\Gamma(2k+1)} \left(\frac{2(p+1)}{p+2}\right)^k \frac{\Gamma(k+l)}{l!\Gamma(k)} \left(1 - \frac{p+2}{2(p+1)}\right)^l \left(\frac{p+2}{2(p+1)}\right)^k.$$

Let X_α^k be negative binomial, $NB(k, \alpha)$, where $\alpha = \frac{p+2}{2(p+1)}$, then

$$a_k(l) = \frac{2^k k!(k-1)!}{\Gamma(2k+1)} \left(\frac{2(p+1)}{p+2}\right)^k P(X_\alpha^k = l),$$

and

$$\Sigma_1 = C_p \frac{2^k k!(k-1)!}{\Gamma(2k+1)} \left(\frac{2(p+1)}{p+2}\right)^k \sum_{l=p}^{k-1} \frac{1}{l^{\frac{p}{2}}} P(X_\alpha^k = l).$$

We claim that, as $k \to +\infty$,

$$\sum_{l=p}^{k-1} \frac{1}{l^{\frac{p}{2}}} P(X_\alpha^k = l) \sim \frac{1}{l_0^{\frac{p}{2}}} \sum_{l=p}^{k-1} P(X_\alpha^k = l) \sim \frac{1}{l_0^{\frac{p}{2}}}, \tag{5.1.13}$$

where $l_0 = \frac{(1-\alpha)k}{\alpha}$. Therefore,

$$\Sigma_1 \sim \frac{1}{l_0^{\frac{p}{2}}} C_p \frac{4^k k!(k-1)!}{\Gamma(2k+1)} \left(\frac{p+1}{p+2}\right)^k.$$

Then, by Stirling's formula (5.1.10), we know

$$A_{k,p+1} \sim \Sigma_1 \sim C_p \sqrt{\pi} \left(\frac{p+2}{p}\right)^{p/2} \frac{1}{k^{\frac{p+1}{2}}} \left(\frac{p+1}{p+2}\right)^k.$$

This lemma is thus proved. Now we only need to show claim (5.1.13). Indeed,

$$\frac{\sum_{l=p}^{k-1} \frac{1}{l^{\frac{p}{2}}} P(X_\alpha^k = l)}{\frac{1}{l_0^{\frac{p}{2}}}} = \sum_{l=p}^{k-1} \left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k = l).$$

$\forall \epsilon > 0$, we have

$$\begin{aligned} &\sum_{l=p}^{k-1} \left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k = l) \\ = &\sum_{p \le l \le l_0(1-\epsilon)} \left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k = l) + \sum_{l_0(1-\epsilon) \le l \le l_0(1+\epsilon)} \left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k = l) \end{aligned}$$

$$+\sum_{l_0(1+\epsilon)\le l\le k-1}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l),$$

where

$$\begin{aligned}0\le\sum_{p\le l\le l_0(1-\epsilon)}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l) &\le \left(\sqrt{\frac{l_0}{p}}\right)^p\sum_{p\le l\le l_0(1-\epsilon)}P(X_\alpha^k=l)\\ &=\left(\sqrt{\frac{l_0}{p}}\right)^p P(X_\alpha^k\le l_0(1-\epsilon)),\end{aligned}$$

$$\begin{aligned}0\le\sum_{l_0(1+\epsilon)\le l\le k-1}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l) &\le \left(\sqrt{\frac{l_0}{l_0(1+\epsilon)}}\right)^p\sum_{l_0(1+\epsilon)\le l\le k-1}P(X_\alpha^k=l)\\ &\le\left(\sqrt{\frac{1}{1+\epsilon}}\right)^p P(X_\alpha^k\ge l_0(1+\epsilon)),\end{aligned}$$

and

$$\begin{aligned}&\left(\sqrt{\frac{l_0}{l_0(1+\epsilon)}}\right)^p P(l_0(1-\epsilon)\le X_\alpha^k\le l_0(1+\epsilon))\\ \le&\sum_{l_0(1-\epsilon)\le l\le l_0(1+\epsilon)}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l)\\ \le&\left(\sqrt{\frac{l_0}{l_0(1-\epsilon)}}\right)^p P(l_0(1-\epsilon)\le X_\alpha^k\le l_0(1+\epsilon)).\end{aligned}$$

By LDP for $NB(k,\alpha)$ in example 2.4, we have

$$P(X_\alpha^k\le l_0(1-\epsilon))\sim e^{-k\left[\inf_{x<(-\epsilon+1)\frac{1-\alpha}{\alpha}} I_1(x)\right]},$$

and

$$P(X_\alpha^k\ge l_0(1+\epsilon))\sim e^{-k\left[\inf_{x>(\epsilon+1)\frac{1-\alpha}{\alpha}} I_1(x)\right]}.$$

Therefore, as $k\to+\infty$,

$$\sum_{p\le l\le l_0(1-\epsilon)}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l)\to 0,$$

and

$$\sum_{l_0(1+\epsilon)\le l\le k-1}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l)\to 0.$$

By the central limit theorem of X_α^k, if we let $k\to+\infty$, then $\epsilon\to 0$, we have

$$\sum_{l_0(1-\epsilon)\le l\le l_0(1+\epsilon)}\left(\sqrt{\frac{l_0}{l}}\right)^p P(X_\alpha^k=l)\to 1.$$

The claim (5.1.13) is thus proved. □

[PROOF OF LEMMA 5.6]:

Proof. Define

$$\Sigma_2=\sum_{s=0}^{k-l}\binom{k}{s}\left(\frac{\lambda-l}{\lambda}\right)^s C_l\frac{1}{(k-s)^{\frac{l}{2}}}\left(\frac{l}{l+1}\right)^{k-s}.$$

Then

$$\Sigma_2=C_l\left(\frac{\lambda-l}{\lambda}+\frac{l}{l+1}\right)^k\sum_{s=0}^{k-l}\frac{1}{(k-s)^{\frac{l}{2}}}P(X_\beta^k=s),$$

where X_β^k follows binomial distribution $B(k,\beta)$, with $\beta=\frac{\frac{\lambda-l}{\lambda}}{\frac{\lambda-l}{\lambda}+\frac{l}{l+1}}$.

Next, we show, as $k\to+\infty$,

$$\sum_{s=0}^{k-l}\frac{1}{(k-s)^{\frac{l}{2}}}P(X_\beta^k=s)\sim\left(\frac{\beta}{(1-\beta)s_0}\right)^{\frac{l}{2}}, \tag{5.1.14}$$

where $s_0=\beta k$. To this end, $\forall\epsilon>0$, let us consider

$$s_0^{\frac{l}{2}}\sum_{s=0}^{k-l}\frac{1}{(k-s)^{\frac{l}{2}}}P(X_\beta^k=s)=\sum_{s=0}^{k-l}\left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k=s).$$

$\forall\epsilon>0$, we have,

$$\begin{aligned}&\sum_{s=0}^{k-l}\left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k=s)\\ =&\sum_{0\le s\le s_0(1-\epsilon)}\left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k=s)+\sum_{(1-\epsilon)s_0\le s\le(1+\epsilon)s_0}\left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k=s)\end{aligned}$$

$$+ \sum_{(1+\epsilon)s_0 \le s \le k-1} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s),$$

where

$$0 \le \sum_{0 \le s \le s_0(1-\epsilon)} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s) \le \left(\sqrt{\frac{s_0}{k-s_0(1-\epsilon)}}\right)^l P(X_\beta^k \le s_0(1-\epsilon))$$

and

$$0 \le \sum_{(1+\epsilon)s_0 \le s \le k-1} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s) \le \left(\sqrt{\frac{s_0}{1}}\right)^l P(X_\beta^k \ge s_0(1+\epsilon)).$$

Then by the LDP for $B(k, \beta)$ in example 2.4, we have

$$\lim_{k \to +\infty} \sum_{0 \le s \le s_0(1-\epsilon)} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s) = 0,$$

and

$$\lim_{k \to +\infty} \sum_{k-l \ge s \ge s_0(1+\epsilon)} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s) = 0.$$

Moreover,

$$\begin{aligned} & \left(\sqrt{\frac{s_0}{k-(1+\epsilon)s_0}}\right)^l P((1-\epsilon)s_0 \le X_\beta^k \le (1+\epsilon)s_0) \\ \le & \sum_{(1-\epsilon)s_0 \le s \le (1+\epsilon)s_0} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s) \\ \le & \left(\sqrt{\frac{s_0}{k-(1-\epsilon)s_0}}\right)^l P((1-\epsilon)s_0 \le X_\beta^k \le (1+\epsilon)s_0). \end{aligned}$$

Analogously, if we let $k \to \infty$ and $\epsilon \to 0$, then we have

$$\sum_{(1-\epsilon)s_0 \le s \le (1+\epsilon)s_0} \left(\sqrt{\frac{s_0}{k-s}}\right)^l P(X_\beta^k = s) \to \left(\frac{\beta}{1-\beta}\right)^{\frac{l}{2}},$$

due to the central limit theorem of binomial distributions. Therefore,

$$\Sigma_2 \sim C_l \left(\frac{1}{1-\beta}\right)^{\frac{l}{2}} \frac{1}{k^{\frac{l}{2}}} \left(\frac{\lambda - l}{\lambda} + \frac{l}{l+1}\right)^k.$$

Next we are going to show $C_{k,l} \sim \Sigma_2$. Because of Lemma 5.5 ,$\forall \epsilon > 0, \exists M > l$, such that $\forall k - s > M$,

$$1 - \epsilon \leq \frac{A_{k-s,l}}{C_l \frac{1}{(k-s)^{\frac{l}{2}}} (\frac{l}{l+1})^{k-s}} \leq 1 + \epsilon.$$

Then we rewrite $C_{k,l}$ as $A + B$, where

$$A = \sum_{s=0}^{k-M-1} \binom{k}{s} \left(\frac{\lambda - l}{\lambda} \right)^s A_{k-s,l}$$

and

$$B = \sum_{k-M \leq s \leq k-l} \binom{k}{s} \left(\frac{\lambda - l}{\lambda} \right)^s A_{k-s,l}.$$

Since

$$0 \leq B \leq \max_{k-M \leq s \leq k-l} \{A_{k-s,l}\} \binom{k}{k-M} \left(\frac{\lambda - l}{\lambda} \right)^{k-M},$$

and

$$\begin{aligned} 0 \leq \frac{B}{\Sigma_2} \leq & \frac{\max_{k-M \leq s \leq k-l} \{A_{k-s,l}\} \binom{k}{k-M} (\frac{\lambda-l}{\lambda})^{k-M}}{\Sigma_2} \\ \sim & \frac{\max_{k-M \leq s \leq k-l} \{A_{k-s,l}\}}{C_l M!} \left(\frac{\lambda}{\lambda - l} \right)^M (1-\beta)^{\frac{l}{2}} k^{\frac{l}{2}+M} \left(\frac{\frac{\lambda-l}{\lambda}}{\frac{\lambda-l}{\lambda} + \frac{l}{l+1}} \right)^k \\ \to & 0. \end{aligned}$$

we can easily see $\lim_{k \to +\infty} \frac{B}{\Sigma_2} = 0$. Moreover,

$$\begin{aligned} & (1-\epsilon) \sum_{0 \leq s \leq k-M-1} \binom{k}{s} \left(\frac{\lambda - l}{\lambda} \right)^s C_l \frac{1}{(k-s)^{\frac{l}{2}}} \left(\frac{l}{l+1} \right)^{k-s} \\ \leq & A \\ \leq & (1+\epsilon) \sum_{0 \leq s \leq k-M-1} \binom{k}{s} \left(\frac{\lambda - l}{\lambda} \right)^s C_l \frac{1}{(k-s)^{\frac{l}{2}}} \left(\frac{l}{l+1} \right)^{k-s}; \end{aligned}$$

and also

$$\begin{aligned} & (1-\epsilon) \left(\Sigma_2 - \sum_{k-M \leq s \leq k-l} \binom{k}{s} \left(\frac{\lambda - l}{\lambda} \right)^s C_l \frac{1}{(k-s)^{\frac{l}{2}}} \left(\frac{l}{l+1} \right)^{k-s} \right) \\ \leq & A \leq (1+\epsilon) \left(\Sigma_2 - \sum_{k-M \leq s \leq k-l} \binom{k}{s} \left(\frac{\lambda - l}{\lambda} \right)^s C_l \frac{1}{(k-s)^{\frac{l}{2}}} \left(\frac{l}{l+1} \right)^{k-s} \right). \end{aligned}$$

Since

$$\begin{aligned}0 \leq & \sum_{k-M\leq s\leq k-l} \binom{k}{s} \left(\frac{\lambda - l}{\lambda}\right)^s C_l \frac{1}{(k-s)^{\frac{l}{2}}} \left(\frac{l}{l+1}\right)^{k-s} \\ \leq & C_l \frac{1}{l^{\frac{l}{2}}} (\frac{l}{l+1})^l \binom{k}{k-M} \left(\frac{\lambda - l}{\lambda}\right)^{k-M},\end{aligned}$$

we have

$$\begin{aligned}0 \leq & \frac{\sum_{k-M\leq s\leq k-l} \binom{k}{s} (\frac{\lambda-l}{\lambda})^s C_l \frac{1}{(k-s)^{\frac{l}{2}}} (\frac{l}{l+1})^{k-s}}{\Sigma_2} \\ \leq & \frac{C_l \frac{1}{l^{\frac{l}{2}}} (\frac{l}{l+1})^l \binom{k}{k-M} (\frac{\lambda-l}{\lambda})^{k-M}}{\Sigma_2} \to 0 \text{ as } k \to \infty.\end{aligned}$$

Hence, letting $k \to \infty$, $\epsilon \to 0$, we have $A \sim \Sigma_2$. Thus, $C_{k,l} \sim \Sigma_2$. Then,

$$\begin{aligned}C_{k,l} \sim \Sigma_2 \sim & C_l \left(\frac{1}{1-\beta}\right)^{\frac{l}{2}} \frac{1}{k^{\frac{l}{2}}} \left(\frac{\lambda - l}{\lambda} + \frac{l}{l+1}\right)^k \\ = & C_l \left(1 + \frac{(\lambda - l)(l+1)}{\lambda l}\right)^{\frac{l}{2}} \frac{1}{k^{\frac{l}{2}}} \left(\frac{\lambda - l}{\lambda} + \frac{l}{l+1}\right)^k.\end{aligned}$$

□

[PROOF OF LEMMA 5.7]:

Proof. Let us assume that

$$\lim_{\theta\to 0} \tilde{K}_n^\lambda(\theta) = (\frac{u-1}{u})^n \text{ for } u(u-1) < \lambda \leq u(u+1), u > 2; \tag{5.1.15}$$

then we are going to show $\lim_{\theta\to 0} K_n^\lambda(\theta) = \lim_{\theta\to 0} \tilde{K}_n^\lambda(\theta)$. Note that $K_n^\lambda(\theta)$ can be rewritten as

$$K_n^\lambda(\theta) = \frac{\tilde{K}_n^\lambda(\theta) + F_n^\lambda(\theta)}{1 + G^\lambda(\theta)},$$

where

$$F_n^\lambda(\theta) = \frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} (A_{k+n,l}(\theta) - A_{k+n,l})}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}},$$

and

$$G^\lambda(\theta) = \frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} (A_{k,l}(\theta) - A_{k,l})}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}}.$$

We claim that $\lim_{\theta\to 0} F_n^\lambda(\theta) = \lim_{\theta\to 0} G^\lambda(\theta) = 0$. Then

$$\lim_{\theta\to 0} K_n^\lambda(\theta) = \lim_{\theta\to 0} \tilde{K}_n^\lambda(\theta)$$

follows. Indeed, by Lemma 5.2, we have

$$\begin{aligned} 0 \le |F_n^\lambda(\theta)| \le & \frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} |A_{k+n,l}(\theta) - A_{k+n,l}|}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} \\ \le & \frac{\sum_{l=1}^{[\lambda]} \theta^l l\theta \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} \\ \le & [\lambda]\theta \frac{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,l}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} = [\lambda]\theta \tilde{K}_n^\lambda(\theta) \to 0. \end{aligned}$$

Similarly, by Lemma 5.2, we also have

$$0 \le |G^\lambda(\theta)| \le [\lambda]\theta \to 0, \text{ as } \theta \to 0.$$

Now we are going to show the assumption (5.1.15). We can rewrite $\tilde{K}_n(\theta)$ as

$$\sum_{v=1}^{[\lambda]} \frac{\theta^v \sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} \frac{\sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,v}}{\sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}.$$

Then,

$$\lim_{\theta\to 0} \tilde{K}_n^\lambda(\theta) = \sum_{v=1}^{[\lambda]} \lim_{\theta\to 0} \frac{\theta^v \sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} \lim_{\theta\to 0} \frac{\sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,v}}{\sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}.$$

By Lemma 5.4 and Lemma 5.5, we know

$$\lim_{\theta\to 0} \frac{\sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k+n,v}}{\sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}} = (\frac{v}{v+1})^n.$$

Then we need to show

$$\lim_{\theta\to 0} \frac{\theta^v \sum_{k=v}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^\infty \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} = \delta_{(u-1)}(v). \tag{5.1.16}$$

Once we have obtained this, then $\lim_{\theta\to 0} \tilde{K}_n^\lambda(\theta) = \sum_{v=1}^{[\lambda]} \delta_{(u-1)}(v)(\frac{v}{v+1})^n = (\frac{u-1}{u})^n$. To this end, both the numerator and the denumerator of

$$\frac{\theta^v \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}},$$

are divided by θ^λ. Thus, we need consider

$$\frac{\theta^{v-\lambda} \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^{l-\lambda} \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}}.$$

Since $1 \leq v \leq [\lambda] \leq \lambda$, it is not difficult to see that

$$\begin{aligned}
\theta^{v-\lambda} \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v} =& (\frac{1}{\theta})^{\lambda-v} \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v} \\
=& \left(\sum_{s=0}^{\infty} \frac{(\log \frac{1}{\theta})^s}{s!} (\lambda - v)^s \right) \left(\sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v} \right) \\
=& \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} C_{k,v},
\end{aligned}$$

where

$$C_{k,v} = \sum_{s=0}^{k-v} \binom{k}{s} \left(\frac{\lambda - v}{\lambda} \right)^s A_{k-s,v}.$$

Then

$$\frac{\theta^{v-\lambda} \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^{l-\lambda} \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} = \frac{\sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} C_{k,v}}{\sum_{l=1}^{[\lambda]} \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} C_{k,l}}.$$

Thus, to figure out the limit (5.1.16), we must find the leading term among

$$\sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} C_{k,l}, 1 \leq l \leq [\lambda].$$

By Lemma 5.4 and Lemma 5.6, we have

$$\begin{aligned}
&\lim_{\theta \to 0} \frac{\sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} C_{k,v}}{\sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} C_{k,l}} = \lim_{k \to +\infty} \frac{C_{k,v}}{C_{k,l}} \\
=& \lim_{k \to +\infty} \frac{(1 + \frac{(\lambda - v)(v+1)}{\lambda v})^{\frac{v}{2}}}{(1 + \frac{(\lambda - l)(l+1)}{\lambda l})^{\frac{l}{2}}} \frac{1}{k^{\frac{v-l}{2}}} \left(\frac{\frac{\lambda - v}{\lambda} + \frac{v}{v+1}}{\frac{\lambda - l}{\lambda} + \frac{l}{l+1}} \right)^k.
\end{aligned}$$

To find the leading term, we need to figure our the maximum term among $\frac{\lambda-l}{\lambda}+\frac{l}{l+1}, 1 \leq l \leq [\lambda]$. Consider $f(x)=\frac{\lambda-x}{\lambda}+\frac{x}{x+1}=2-(\frac{x}{\lambda}+\frac{1}{x+1})$; then

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

We know $f^{'}(x)$ is

$$\begin{cases} \geq 0, & \text{if } x \leq \sqrt{\lambda}-1 \\ < 0, & \text{if } x > \sqrt{\lambda}-1. \end{cases}$$

Therefore, $\frac{\lambda-l}{\lambda}+\frac{l}{l+1}$ attains its maximum at $[\sqrt{\lambda}]-1$ or $[\sqrt{\lambda}]$.

Case 1: For $(u-1)u<\lambda<u^2, [\sqrt{\lambda}]=u-1$, since

$$f(u-2)=2-(\frac{u-2}{\lambda}+\frac{1}{u-1})$$

and

$$f(u-1)=2-(\frac{u-1}{\lambda}+\frac{1}{u}),$$

$f(u-2)-f(u-1)=\frac{1}{\lambda}-\frac{1}{u(u-1)}<0$. So $f(u-1)$ is the maximum term.

$$\frac{\frac{\lambda-l}{\lambda}+\frac{l}{l+1}}{\frac{\lambda-u+1}{\lambda}+\frac{u-1}{u}}<1, \forall 1 \leq l \leq [\lambda], l \neq u-1.$$

Thus

$$\lim_{k \to +\infty} \frac{C_{k,l}}{C_{k,u-1}}=\delta_{u-1}(l), \forall 1 \leq l \leq [\lambda].$$

Case 2: For $u^2 \leq \lambda < u(u+1), [\sqrt{\lambda}]=u$; then the maximum term of $\frac{\lambda-l}{\lambda}+\frac{l}{l+1}$ should be $f(u-1)$ or $f(u)$. Since

$$f(u-1)-f(u)=\frac{1}{\lambda}-\frac{1}{u(u+1)}>0,$$

we can see $f(u-1)$ is the maximum term; and

$$\frac{\frac{\lambda-l}{\lambda}+\frac{l}{l+1}}{\frac{\lambda-u+1}{\lambda}+\frac{u-1}{u}}<1, \forall 1 \leq l \leq [\lambda], l \neq u-1.$$

Thus,

$$\lim_{k \to +\infty} \frac{C_{k,l}}{C_{k,u-1}}=\delta_{u-1}(l), \forall 1 \leq l \leq [\lambda].$$

Case 3: For $\lambda=u(u+1), [\sqrt{\lambda}]=u$, then the maximum term of $\frac{\lambda-l}{\lambda}+\frac{l}{l+1}$ is

$f(u-1)$ or $f(u)$. Since

$$f(u-1) - f(u) = \frac{1}{\lambda} - \frac{1}{u(u+1)} = 0,$$

one can have

$$\frac{\frac{\lambda-l}{\lambda} + \frac{l}{l+1}}{\frac{\lambda-u+1}{\lambda} + \frac{u-1}{u}} < 1, \forall 1 \leq l \leq [\lambda], l \neq u, u-1,$$

and

$$\frac{\frac{\lambda-u+1}{\lambda} + \frac{u-1}{u}}{\frac{\lambda-u}{\lambda} + \frac{u}{u+1}} = 1.$$

But

$$\lim_{k\to+\infty} \frac{C_{k,u}}{C_{k,u-1}} = \lim_{k\to+\infty} \frac{(1+\frac{(\lambda-u)(u+1)}{\lambda u})^{\frac{u}{2}}}{(1+\frac{(\lambda-u+1)u}{\lambda(u-1)})^{\frac{u-1}{2}}} \frac{1}{k^{\frac{1}{2}}} = 0;$$

therefore, $C_{k,u-1}$ is the leading term among $C_{k,l}, 1 \leq l \leq [\lambda]$. We have

$$\lim_{k\to+\infty} \frac{C_{k,l}}{C_{k,u-1}} = \delta_{u-1}(l), 1 \leq l \leq [\lambda].$$

Thus,

$$\lim_{\theta\to 0} \frac{\theta^v \sum_{k=v}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,v}}{\sum_{l=1}^{[\lambda]} \theta^l \sum_{k=l}^{\infty} \frac{(\lambda \log \frac{1}{\theta})^k}{k!} A_{k,l}} = \delta_{(u-1)}(v), 1 \leq v \leq [\lambda].$$

□

5.2 Some Partial Results on Asymptotic behaviours of Transient Distributions

In this section, we can show that the limits of the transient distributions of the one-parameter selective model with symmetric overdominance coincide with that of the one-parameter neutral model. In this sense, Gillespie's conjecture is verified. For the one-parameter selective model with symmetric underdominance, however, the limiting transient distributions can only be obtained when $\sigma = \lambda\alpha(\theta), \lambda < 2$.

5.2.1 One-parameter Selective Model with Symmetric Overdominance

Let us present several lemmas before we show the main result.

Lemma 5.8. *Suppose that $f(t)$ and $g(t)$ are differentiable and locally integrable functions in $[0,+\infty)$ respectively. If*

$$\frac{df(t)}{dt} \le -\theta f(t) + g(t),$$

where $\theta > 0$,

$$f(t) \le f(0)e^{-\theta t} + \int_0^t e^{-\theta(t-s)} g(s)ds, \quad \forall t \ge 0.$$

Proof. Notice that

$$e^{\theta t} f(t) - f(0) = \int_0^t (e^{\theta s} f(s))' ds.$$

Since $(e^{\theta t} f(t))' = \theta e^{\theta t} f(t) + e^{\theta t} f'(t) \le e^{\theta} g(t)$, integrating on both sides gives rise to

$$e^{\theta t} f(t) - f(0) \le \int_0^t e^{\theta s} g(s)ds.$$

The lemma is thus proved! $\square$

Theorem 5.6. *Let $X_t^{\sigma(\theta)}$ be the one-parameter selective model with symmetric over-dominance, where $\sigma(\theta) = \lambda\alpha(\theta)$, and $\alpha(\theta) > 0, \lambda < 0$. Then, for any given $t > 0$, we have*

$$X_t^{\sigma(\theta)} \to \delta_{(0,0,\cdots)} \text{ weakly , as } \theta \to +\infty.$$

Proof. Since $\{X_t^{\sigma(\theta)}, \theta > 0\}$ is a family of $\bar{\nabla}_\infty$-valued random variables, and $\bar{\nabla}_\infty$ is compact, by the Prokhorov theorem we can conclude that $\{X_t^{\sigma(\theta)}, \theta > 0\}$ is tight, and hence relative compact. For any sequence $\{\theta_k, k \ge 1\}$, where $\lim_{k\to+\infty} \theta_k = +\infty$, there must be a subsequence $\theta_{k_l}, l \ge 1$ such that $X_t^{\sigma(\theta_{k_l})}$ weakly converges to a measure $\mu_0(dx)$. If we can confirm that $\mu_0(dx) = \delta_{(0,0,\cdots)}$, then the proof is completed. To this end, we consider $E\varphi_2(X_t^{\sigma(\theta)})$. Then

$$\begin{aligned} E\varphi_2(X_t^{\sigma(\theta)}) - \varphi_2(0) = \int_0^t \Bigg[& \big(1 - (\theta+1)E\varphi_2(X_s^{\sigma(\theta)})\big) \\ & + 2\sigma(\theta)E\Big(\varphi_3(X_s^{\sigma(\theta)}) - \varphi_2^2(X_s^{\sigma(\theta)})\Big)\Bigg] ds. \end{aligned} \tag{5.2.1}$$

By the Cauchy-Schwarz inequality, we know

$$\varphi_2(x) = \sum_{i=1}^{\infty} x_i^{\frac{1}{2}} x_i^{\frac{3}{2}} \le \sqrt{(\sum_{i=1}^{\infty} x_i)}\sqrt{(\sum_{i=1}^{\infty} x_i^3)} \le \sqrt{(\sum_{i=1}^{\infty} x_i^3)};$$

thus, $\varphi_3(x) \geq \varphi_2^2(x)$. Besides, $\sigma(\theta) < 0$; therefore

$$\frac{dE\varphi_2(X_t^{\sigma(\theta)})}{dt} \leq 1 - (\theta+1)E\varphi_2(X_t^{\sigma(\theta)}).$$

By Lemma 5.8, we know

$$E\varphi_2(X_t^{\sigma(\theta)}) \leq e^{-(\theta+1)t}\varphi_2(x) + \int_0^t e^{-(t-s)(\theta+1)}ds = e^{-(\theta+1)t}\varphi_2(x) + \frac{1-e^{-(\theta+1)t}}{\theta+1}.$$

Letting $\theta \to +\infty$, we have

$$\int_{\bar{\nabla}_\infty} \varphi_2(x)\mu_0(dx) = \lim_{l\to+\infty} EX_t^{\sigma(\theta_{k_l})} = 0.$$

Then $\mu_0 = \delta_{(0,0,\cdots)}$. □

Remark 5.4. *As can be seen, for any fixed $t > 0$, the transient distribution of the one-parameter selective model with symmetric dominance converges weakly to $\delta_{(0,0,\cdots)}$ like the one-parameter neutral model. In this sense, Gillespie's conjecture, therefore, is right. Moreover, since*

$$E\varphi_2(X_t^{\sigma(\theta)}) \leq e^{-(\theta+1)t}\varphi_2(x) + \frac{1-e^{-(\theta+1)t}}{\theta+1},$$

we conclude, as long as $\lim_{\theta\to+\infty}\theta t(\theta) = +\infty$, $X_{t(\theta)}^{\sigma(\theta)} \to \delta_{(0,0,\cdots)}$ weakly as $\theta \to +\infty$.

5.2.2 One-parameter Selective Model with Symmetric Underdominance

Let $X_t^{\sigma(\theta)}$ be the one-parameter selective model with symmetric underdominance, in which $\sigma(\theta) = \lambda\alpha(\theta)(\lambda > 0, \alpha(\theta) > 0)$ and $\lim_{\theta\to+\infty}\frac{\alpha(\theta)}{\theta} = 0$ or 1. We are going to show that, for any fixed $t > 0$, $X_t^{\sigma(\theta)}$ converges weakly to $\delta_{(0,0,\cdots)}$ only if $\lambda < 2$. To this end, we need the following lemma.

Lemma 5.9. *For $0 < \lambda < 2$, we have*

$$\inf_{x\in\bar{\nabla}_\infty}\left\{1 - 2\lambda\frac{\varphi_3(x)-\varphi_2^2(x)}{\varphi_2(x)}\right\} \geq 1 - \frac{\lambda}{2} > 0.$$

Proof.

$$\inf_{x\in\bar{\nabla}_\infty}\left\{1 - 2\lambda\frac{\varphi_3(x)-\varphi_2^2(x)}{\varphi_2(x)}\right\} = \inf_{r\in[0,1]}\inf_{\varphi_2(x)=r}\left\{1 - 2\lambda\frac{\varphi_3(x)-r^2}{r}\right\}$$

$$= \inf_{r\in[0,1]} \left\{ 1 - 2\lambda \sup_{\varphi_2(x)=r} (\frac{\varphi_3(x)}{r} - r) \right\}.$$

Moreover, we claim that $\sup_{\varphi_2(x)=r} \varphi_3(x) = r^{\frac{3}{2}}$. Indeed, set $y_i = \frac{x_i^2}{r}, i \geq 1$; then

$$\sum_{i=1}^{\infty} x_i^3 = r^{\frac{3}{2}} \sum_{i=1}^{\infty} y_i^{\frac{3}{2}} \leq r^{\frac{3}{2}} \sum_{i=1}^{\infty} y_i \leq r^{\frac{3}{2}}.$$

Taking $x = (\sqrt{r}, 0, 0, \cdots)$, we have $\sup_{\varphi_2(x)=r} \varphi_3(x) = r^{\frac{3}{2}}$. Therefore,

$$\inf_{x\in\bar{\nabla}_\infty} \left\{ 1 - 2\lambda \frac{\varphi_3(x) - \varphi_2^2(x)}{\varphi_2(x)} \right\} = \inf_{r\in[0,1]} \left\{ 1 - 2\lambda(\sqrt{r} - r) \right\} = 1 - \frac{\lambda}{2} > 0.$$

□

Theorem 5.7. *Let $X_t^{\sigma(\theta)}$ be the one-parameter selective model with symmetric underdominance. When $0 < \lambda < 2$, then, for any given $t > 0$, we have, as $\theta \to +\infty$,*

$$X_t^{\sigma(\theta)} \to \delta_{(0,0,\cdots)} \text{ weakly }.$$

Proof. Since $\{X_t^{\sigma(\theta)}, \theta > 0\}$ is a family of $\bar{\nabla}_\infty$-valued random variables, and $\bar{\nabla}_\infty$ is compact, by the Prokhorov theorem $\{X_t^{\sigma(\theta)}, \theta > 0\}$ is tight, and hence relative compact. For any sequence $\{\theta_k, k \geq 1\}$, where $\lim_{k\to+\infty} \theta_k = +\infty$, there must exist a subsequence $\theta_{k_l}, l \geq 1$ such that $X_t^{\sigma(\theta_{k_l})}$ weakly converges to a measure ν_0. If we can show $\nu_0 = \delta_{(0,0,\cdots)}$, then the proof is completed. To this end, we consider $E\varphi_2(X_t^{\sigma(\theta)})$; similarly, we have

$$\begin{aligned}
E\varphi_2(X_t^{\sigma(\theta)}) - \varphi_2(0) &= \int_0^t \Big[(1 - (\theta+1)E\varphi_2(X_s^{\sigma(\theta)})) \\
&\quad + 2\sigma(\theta)E\Big(\varphi_3(X_s^{\sigma(\theta)}) - \varphi_2^2(X_s^{\sigma(\theta)})\Big)\Big] ds. \\
&= \int_0^t \Big[(1 - (\theta+1)E\varphi_2(X_s^{\sigma(\theta)})) + 2\lambda\theta E\Big(\varphi_3(X_s^{\sigma(\theta)}) - \varphi_2^2(X_s^{\sigma(\theta)})\Big) \\
&\quad + 2\lambda(\alpha(\theta) - \theta)E\Big(\varphi_3(X_s^{\sigma(\theta)}) - \varphi_2^2(X_s^{\sigma(\theta)})\Big)\Big] ds \\
&= \int_0^t (1 - E\varphi_2(X_s^{\sigma(\theta)}))ds - \theta \int_0^t E\varphi_2(X_s^{\sigma(\theta)}) \\
&\quad \left[1 - 2\lambda \frac{\varphi_3 - \varphi_2^2}{\varphi_2} - 2\lambda \left(\frac{\alpha(\theta)}{\theta} - 1 \right) \frac{\varphi_3 - \varphi_2^2}{\varphi_2} \right] (X_s^{\sigma(\theta)}) ds.
\end{aligned}$$

Therefore,

$$\frac{dE\varphi_2(X_t^{\sigma(\theta)})}{dt} \leq 1 - \theta E\varphi_2(X_t^{\sigma(\theta)}) \Bigg[\inf_{x\in\bar{\nabla}_\infty} \left(1 - 2\lambda \frac{\varphi_3 - \varphi_2^2}{\varphi_2}\right) - 2\lambda\left|\frac{\alpha(\theta)}{\theta} - 1\right| \sup_{x\in\bar{\nabla}_\infty} \frac{\varphi_3 - \varphi_2^2}{\varphi_2}\Bigg].$$

Due to Lemma 5.9, we have

$$\frac{dE\varphi_2(X_t^{\sigma(\theta)})}{dt} \leq 1 - \theta E\varphi_2(X_t^{\sigma(\theta)}) \left(1 - \frac{\lambda}{2} - \frac{\lambda}{2}\left|\frac{\alpha(\theta)}{\theta} - 1\right|\right).$$

Since $\lim_{\theta\to+\infty} \frac{\alpha(\theta)}{\theta} = 1$, then $\exists\theta_0 > 0$, such that $\forall\theta > \theta_0$, we have

$$1 - \frac{\lambda}{2} - \frac{\lambda}{2}\left|\frac{\alpha(\theta)}{\theta} - 1\right| > \frac{1}{2}(1 - \frac{\lambda}{2}) > 0.$$

Thus, $\forall\theta > \theta_0$, we have

$$\frac{dE\varphi_2(X_t^{\sigma(\theta)})}{dt} \leq 1 - \theta E\varphi_2(X_t^{\sigma(\theta)})\frac{1}{2}\left(1 - \frac{\lambda}{2}\right).$$

By Lemma 5.8, we have

$$E\varphi_2(X_t^{\sigma(\theta)}) \leq e^{-\theta\frac{1}{2}(1-\frac{\lambda}{2})t}\varphi_2(x) + \frac{1}{\theta\frac{1}{2}(1 - \frac{\lambda}{2})}.$$

Letting $\theta \to +\infty$, we have $\lim_{\theta\to+\infty} E\varphi_2(X_t^{\sigma(\theta)}) = 0$. The theorem is thus proved. $\square$

References

[Arratia et al., 2003] Arratia, R., Barbour, A. D., and Tavaré, S. (2003). *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich.

[Barbour et al., 2000] Barbour, A. D., Ethier, S. N., and Griffiths, R. C. (2000). A transition function expansion for a diffusion model with selection. *Ann. Appl. Probab.*, 10(1):123–162.

[Chen, 2005] Chen, M.-F. (2005). *Eigenvalues, inequalities, and ergodic theory*. Probability and its Applications (New York). Springer-Verlag London Ltd., London.

[Dawson and Feng, 1998] Dawson, D. A. and Feng, S. (1998). Large deviations for the Fleming-Viot process with neutral mutation and selection. *Stochastic Process. Appl.*, 77(2):207–232.

[Dawson and Feng, 2001] Dawson, D. A. and Feng, S. (2001). Large deviations for the Fleming-Viot process with neutral mutation and selection. II. *Stochastic Process. Appl.*, 92(1):131–162.

[Dawson and Feng, 2006] Dawson, D. A. and Feng, S. (2006). Asymptotic behavior of the Poisson-Dirichlet distribution for large mutation rate. *Ann. Appl. Probab.*, 16(2):562–582.

[Dembo and Zeitouni, 2010] Dembo, A. and Zeitouni, O. (2010). *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin. Corrected reprint of the second (1998) edition.

[Ethier, 1992] Ethier, S. N. (1992). Eigenstructure of the infinitely-many-neutral-alleles diffusion model. *J. Appl. Probab.*, 29(3):487–498.

[Ethier and Griffiths, 1993] Ethier, S. N. and Griffiths, R. C. (1993). The transition function of a Fleming-Viot process. *Ann. Probab.*, 21(3):1571–1590.

[Ethier and Kurtz, 1981] Ethier, S. N. and Kurtz, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.*, 13(3):429–452.

[Ethier and Kurtz, 1986] Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.

[Ethier and Kurtz, 1993] Ethier, S. N. and Kurtz, T. G. (1993). Fleming-Viot processes in population genetics. *SIAM J. Control Optim.*, 31(2):345–386.

[Ethier and Kurtz, 1994] Ethier, S. N. and Kurtz, T. G. (1994). Convergence to Fleming-Viot processes in the weak atomic topology. *Stochastic Process. Appl.*, 54(1):1–27.

[Ethier and Kurtz, 1998] Ethier, S. N. and Kurtz, T. G. (1998). Coupling and ergodic theorems for Fleming-Viot processes. *Ann. Probab.*, 26(2):533–561.

[Feng, 2007a] Feng, S. (2007a). Large deviations associated with Poisson-Dirichlet distribution and Ewens sampling formula. *Ann. Appl. Probab.*, 17(5-6):1570–1595.

[Feng, 2007b] Feng, S. (2007b). Large deviations for Dirichlet processes and Poisson-Dirichlet distribution with two parameters. *Electron. J. Probab.*, 12:no. 27, 787–807.

[Feng, 2009] Feng, S. (2009). Poisson-Dirichlet distribution with small mutation rate. *Stochastic Process. Appl.*, 119(6):2082–2094.

[Feng, 2010] Feng, S. (2010). *The Poisson-Dirichlet distribution and related topics: Models and asymptotic behaviors*. Probability and its Applications (New York). Springer, Heidelberg.

[Feng and Gao, 2010] Feng, S. and Gao, F. (2010). Asymptotic results for the two-parameter Poisson-Dirichlet distribution. *Stochastic Process. Appl.*, 120(7):1159–1177.

[Feng and Sun, 2010] Feng, S. and Sun, W. (2010). Some diffusion processes associated with two parameter Poisson-Dirichlet distribution and Dirichlet process. *Probab. Theory Related Fields*, 148(3-4):501–525.

[Feng et al., 2011] Feng, S., Sun, W., Wang, F.-Y., and Xu, F. (2011). Functional inequalities for the two-parameter extension of the infinitely-many-neutral-alleles diffusion. *J. Funct. Anal.*, 260(2):399–413.

[Feng and Xiong, 2002] Feng, S. and Xiong, J. (2002). Large deviations and quasi-potential of a Fleming-Viot process. *Electron. Comm. Probab.*, 7:13–25 (electronic).

[Gillespie, 1999] Gillespie, J. H. (1999). The role of population size in molecular evolution. *Theoretical Population Biology*, 55(2):145–156.

[Gillespie, 2004] Gillespie, J. H. (2004). *Population Genetics: A Concise Guide*. The Johns Hopkins University Press, second edition.

[Griffiths, 1979] Griffiths, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. in Appl. Probab.*, 11(2):310–325.

[Grote and Speed, 2002] Grote, M. N. and Speed, T. P. (2002). Approximate Ewens formulae for symmetric overdominance selection. *Ann. Appl. Probab.*, 12(2):637–663.

[Handa, 2005] Handa, K. (2005). Sampling formulae for symmetric selection. *Electron. Comm. Probab.*, 10:223–234.

[Hogg and Craig, 1978] Hogg, R. V. and Craig, A. T. (1978). *Introduction to mathematical statistics*. Macmillan Publishing Co., Inc., New York, fourth edition.

[Huillet, 2007] Huillet, T. (2007). Ewens sampling formulae with and without selection. *J. Comput. Appl. Math.*, 206(2):755–773.

[Joyce et al., 2003] Joyce, P., Krone, S. M., and Kurtz, T. G. (2003). When can one detect overdominant selection in the infinite-alleles model? *Ann. Appl. Probab.*, 13(1):181–212.

[Karlin and Taylor, 1981] Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.

[Kingman, 1977] Kingman, J. F. C. (1977). The population structure associated with the Ewens sampling formula. *Theoret. Population Biology*, 11(2):274–283.

[Kingman, 1978a] Kingman, J. F. C. (1978a). Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A*, 361(1704):1–20.

[Kingman, 1978b] Kingman, J. F. C. (1978b). The representation of partition structures. *J. London Math. Soc. (2)*, 18(2):374–380.

[Kingman, 1993] Kingman, J. F. C. (1993). *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York. Oxford Science Publications.

[Kingman et al., 1975] Kingman, J. F. C., Taylor, S. J., Hawkes, A. G., Walker, A. M., Cox, D. R., Smith, A. F. M., Hill, B. M., Burville, P. J., and Leonard, T. (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B*, 37:1–22. With a discussion by S. J. Taylor, A. G. Hawkes, A. M. Walker, D. R. Cox, A. F. M. Smith, B. M. Hill, P. J. Burville, T. Leonard and a reply by the author.

[Li et al., 1999] Li, Z., Shiga, T., and Yao, L. (1999). A reversibility problem for Fleming-Viot processes. *Electron. Comm. Probab.*, 4:65–76 (electronic).

[Macdonald, 1995] Macdonald, I. G. (1995). *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition. With contributions by A. Zelevinsky, Oxford Science Publications.

[Overbeck and Röckner, 1997] Overbeck, L. and Röckner, M. (1997). Geometric aspects of finite- and infinite-dimensional Fleming-Viot processes. *Random Oper. Stochastic Equations*, 5(1):35–58.

[Petrov, 2009] Petrov, L. A. (2009). A two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funktsional. Anal. i Prilozhen.*, 43(4):45–66.

[Pitman, 2006] Pitman, J. (2006). *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.

[Pitman and Yor, 1997] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25(2):855–900.

[Puhalskii, 1991] Puhalskii, A. (1991). On functional principle of large deviations. In *New trends in probability and statistics, Vol. 1 (Bakuriani, 1990)*, pages 198–218. VSP, Utrecht.

[Qian and Zheng, 2004] Qian, Z. and Zheng, W. (2004). A representation formula for transition probability densities of diffusions and applications. *Stochastic Process. Appl.*, 111(1):57–76.

[Revuz and Yor, 1999] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition.

[Stannat, 2000] Stannat, W. (2000). On the validity of the log-Sobolev inequality for symmetric Fleming-Viot operators. *Ann. Probab.*, 28(2):667–684.

[Steinröcken and Song, 2013] Steinröcken, M. and Song, S. Y. (2013). A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*, 190(3):1117–1129.

[Steinröcken et al., 2013] Steinröcken, M., Wang, Y. R., and Song, Y. (2013). An explicit transition density expansion for a multi-allelic wright-fisher diffusion with general diploid selection. *Theoretical Population Biology*, 83(2):1–14.

[Stroock and Varadhan, 2006] Stroock, D. W. and Varadhan, S. R. S. (2006). *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin. Reprint of the 1997 edition.

[Tavaré, 1984] Tavaré, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoret. Population Biol.*, 26(2):119–164.

[Wang, 2000] Wang, F.-Y. (2000). Functional inequalities, semigroup properties and spectrum estimates. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 3(2):263–295.

[Watterson, 1977a] Watterson, G. A. (1977a). Correction to: "The stationary distribution of the infinitely-many neutral alleles diffusion model". *J. Appl. Probability*, 14(4):897.

[Watterson, 1977b] Watterson, G. A. (1977b). Heterosis or neutrality? *Genetics*, 85(4):789–814.