

Divisible Statistics and Their Partial  
Sum Processes:  
*Asymptotic Properties and  
Applications*

by

Haizhen Wu

A thesis  
submitted to the Victoria University of Wellington  
in fulfilment of the  
requirements for the degree of  
Doctor of Philosophy  
in Statistics.

Victoria University of Wellington

2010



# Abstract

Divisible statistics have been widely used in many areas of statistical analysis. For example, Pearson's Chi-square statistic and the log-likelihood ratio statistic are frequently used in goodness of fit (GOF) and categorical analysis; the maximum likelihood (ML) estimators of the Shannon's and Simpson's diversity indices are often used as measure of diversity; and the spectral statistic plays a key role in the theory of large number of rare events.

In the classical multinomial model, where the number of disjoint events  $N$  and their probabilities  $\{p_i\}_{1 \leq i \leq N}$  are all fixed, limit distributions of many divisible statistics have gradually been established. However, most of the results are based on the asymptotic equivalence of these statistics to Pearson's Chi-square statistic and the known limit distribution of the latter. In fact, with deeper analysis, one can conclude that the key point is not the asymptotic behavior of the Chi-square statistic, but that of the normalized frequencies. Based on the asymptotic normality of the normalized frequencies in the classical model, a unified approach to the limit theorems of more general divisible statistics can be established, of which the case of the Chi-square statistic is simply a natural corollary.

In many applications, however, the classical multinomial model is not appropriate, and an extension to new models becomes necessary. This new type of model, called "non-classical" multinomial models, considers the case when  $N$  increases and the  $\{p_{ni}\}$  change as sample size  $n$  increases. As we will see, in

these non-classical models, both the asymptotic normality of the normalized frequencies and the asymptotic equivalence of many divisible statistics to the Chi-square statistic are lost, and the limit theorems established in classical model are no longer valid in non-classical models.

The extension to non-classical models not only met the demands of many real world applications, but also opened a new research area in statistical analysis, which has not been thoroughly investigated so far. Although some results on the limit distributions of the divisible statistics in non-classical models have been acquired, e.g., Holst (1972); Morris (1975); Ivchenko and Levin (1976); Ivchenko and Medvedev (1979), they are far from complete.

Though not yet attracting much attention by many applied statisticians, another advanced approach, introduced by Khmaladze (1984), makes use of modern martingale theory to establish functional limit theorems of the partial sum processes of divisible statistics successfully. In the main part of this thesis, we show that this martingale approach can be extended to more general situations where both Gaussian and Poissonian frequencies exist, and further discuss the properties and applications of the limiting processes, especially in constructing distribution-free statistics.

The last part of the thesis is about the statistical analysis of large number of rare events (LNRE), which is an important class of non-classical multinomial models and presented in numerous applications. In LNRE models, most of the frequencies are very small and it is not immediately clear how consistent and reliable inference can be achieved. Based on the definitions and key concepts firstly introduced by Khmaladze (1988), we discuss a particular model with the context of diversity of questionnaires. The advanced statistical techniques such as large deviation, contiguity and Edgeworth expansion used in establishing limit theorems underpin the potential of LNRE theory to become a fruitful research area in future.

# Acknowledgments

Firstly, I am very grateful to my supervisors, Prof. Estate V. Khmaladze and Dr. Leigh Roberts, for their patient guidance, their encouragement, and their invaluable comments and assistance in the preparation of this thesis. In particular, I would like to thank Estate for introducing me to this interesting and challenging topic and spending lots of time on training me rigorously in mathematics and also for his patience when I progressed slowly at some stage. To Leigh, I very much appreciate his encouragement in spirit and tolerance of my deviating from the finance area.

I am also indebted to staff members in the School of Mathematics, Statistics and Operations research including A/Prof. Megan Clark, Dr. Dong Wang, Dr. Richard Arnold, Dr. Ivy Liu and many others, for their encouragement and support during my study.

My deepest thanks goes to my family including my wife, Ying Xu, my son Jack Wu, my parents, Shenghe Hu and Chaoren Wu and my parents in law Xianqing Jiang and Linyuan Xu, for their unreserved love and unconditional support.

Finally, I would like to acknowledge the financial assistance from the Victoria University of Wellington, who provided Victoria PhD scholarships.



# Contents

List of Figures	vii
List of Tables	viii
Abbreviations	ix
<b>1 Introduction</b>	<b>1</b>
1.1 Multinomial Model . . . . .	2
1.1.1 Classical Model . . . . .	2
1.1.2 Non-Classical Models . . . . .	3
1.2 Divisible Statistics . . . . .	3
1.2.1 Definitions . . . . .	3
1.2.2 Traditional Goodness of Fit Statistics . . . . .	5
1.2.3 Power-divergence Test Statistics . . . . .	6
1.2.4 The $\phi$ -divergence Test Statistics . . . . .	7
1.2.5 $L_1$ -distance . . . . .	7
1.2.6 Estimates of Diversity Indices . . . . .	8
1.2.7 Spectral Statistics . . . . .	8
1.3 Aim and Outline of the Thesis . . . . .	9
<b>2 Divisible Statistics in Classical Models</b>	<b>11</b>
2.1 Applications . . . . .	12
2.1.1 GOF Test with Specified Distributions . . . . .	12
2.1.2 GOF Test with Parameters Estimated . . . . .	13

2.1.3	Measures of Diversity . . . . .	17
2.2	Exact Distributions of Divisible Statistics . . . . .	18
2.3	Limit Distributions in Classical Models . . . . .	21
2.3.1	Limit Distributions of the Normalized Frequencies . . .	21
2.3.2	Limit Distributions of Divisible Statistics . . . . .	24
2.3.3	Asymptotic Equivalence to Chi-square Statistic . . . .	26
2.3.4	Limit Distributions of Chi-square Statistic . . . . .	27
<b>3</b>	<b>Divisible Statistics in Non-classical Models</b>	<b>29</b>
3.1	Some Objects in Non-classical Models . . . . .	30
3.1.1	Objects . . . . .	30
3.1.2	Limit behavior . . . . .	31
3.2	Typical Examples of Non-classical Multinomial Models . . . .	33
3.2.1	GOF Test for Continuous Distributions . . . . .	33
3.2.2	Statistical Analysis to Linguistic Data . . . . .	35
3.2.3	Diversity Analysis of Responses in Questionnaires . . .	36
3.3	Difficulties in Non-Classical Models . . . . .	37
3.3.1	Loss of Asymptotic Normality . . . . .	37
3.3.2	Loss of Asymptotic Equivalence . . . . .	39
<b>4</b>	<b>FLTs for Partial Sum Processes</b>	<b>40</b>
4.1	Preliminaries . . . . .	42
4.1.1	Probability spaces and martingales . . . . .	42
4.1.2	Triangular Array and Lindeberg Theorems . . . . .	46
4.1.3	Martingale Central Limit Theorem . . . . .	47
4.2	FLT for Models with Poissonian Frequencies . . . . .	48
4.2.1	FLT for Martingale Component . . . . .	50
4.2.2	FLT for Compensator . . . . .	51
4.3	FLT in Models with Mixed Frequencies . . . . .	53
4.3.1	Auxiliary Lemmas . . . . .	55
4.3.2	FLT for Martingale Component . . . . .	63
4.3.3	FLT for the Compensator Component . . . . .	65



4.3.4	Alternative Proof to FLT of Compensator of Chi-square Process . . . . .	68
4.4	FLT for Partial Sum Processes . . . . .	70
4.5	FLTs under Contiguous Alternatives . . . . .	71
4.6	Comparison of FLTs in two Schemes . . . . .	77
<b>5</b>	<b>Properties of Limiting Processes and Applications</b>	<b>79</b>
5.1	Characteristics of Limiting Processes . . . . .	80
5.1.1	Chi-square Process . . . . .	80
5.1.2	Log-likelihood Processes . . . . .	81
5.1.3	Spectral Statistics Processes . . . . .	82
5.2	Some Properties of Limiting Processes . . . . .	83
5.2.1	Properties in Models with all Gaussian Frequencies . .	83
5.2.2	Transformation of Partial Sum Processes . . . . .	84
5.3	GOF tests based on Limiting Processes . . . . .	85
5.3.1	Evaluation of convergence of $\tilde{W}_n$ and $W_n$ . . . . .	85
5.3.2	Properties of Statistics based on $\tilde{W}_n$ and $W_n$ . . . . .	94
<b>6</b>	<b>Asymptotics in the LNRE models</b>	<b>98</b>
6.1	Introduction to LNRE theory . . . . .	99
6.1.1	Definition of LNRE . . . . .	100
6.1.2	Law of Large Numbers of Spectral Statistics . . . . .	101
6.1.3	Models of LNRE . . . . .	102
6.2	LNRE in multiple-choice questionnaires . . . . .	104
6.2.1	The Model . . . . .	105
6.2.2	Preliminary Discussion . . . . .	106
6.2.3	Limit theorem for neutral questionnaires . . . . .	109
6.2.4	Limit theorem for contiguous neighborhood of neutral questionnaires . . . . .	110
6.2.5	Limit theorem for general cases . . . . .	115
<b>7</b>	<b>Conclusions</b>	<b>121</b>

<i>CONTENTS</i>	vii
A Some discussion on Large Deviations	124
Bibliography	126

# List of Figures

2.1	Comparison of exact and limit distribution of divisible statistics	20
2.2	Limit distributions of general divisible statistics . . . . .	25
3.1	Loss of Power of GOF Test to Continuous Distribution . . . .	33
3.2	Loss of Asymptotic Normality of Normalized Frequencies . . .	38
3.3	Loss of Asymptotic Equivalence to Chisquare Statistic . . . .	39
5.1	Density function of Beta(2,4) and Beta(0.5,0.5) . . . . .	88
5.2	Evaluation of Convergence of $W_n$ in Scheme A . . . . .	89
5.3	Evaluation of Convergence of $\tilde{W}_n$ in Scheme A . . . . .	90
5.4	Evaluation of Convergence of $W_n$ in Scheme B . . . . .	91
5.5	Evaluation of Convergence of $\tilde{W}_n$ in Scheme B . . . . .	92
5.6	Convergence Rate of $\tilde{W}_n, W_n$ for Different Divisible Statistic .	93
5.7	Distribution-free Property of Statistics Based on $\tilde{W}_n$ and $W_n$ .	95
5.8	Power in Detecting some Contiguous Alternatives . . . . .	97

# List of Tables

3.1	Word Frequencies of BNC BROWN Corpus . . . . .	36
6.1	Connectivity of Global Internet Routing System . . . . .	100

# Abbreviations

BNC	British National Corpus.
CDF	cumulative distribution function.
CLT	central limit theorem.
GOF	goodness of fit.
LNRE	large number of rare events.
ML	maximum likelihood.



# Chapter 1

## Introduction

In many applications of statistical analysis, such as the GOF test of finite discrete distributions (see e.g. Read and Cressie, 1988), inference in categorical analysis (see e.g. Agresti, 2002) and statistical analysis of diversity of finite populations (see e.g. Magurran, 2004; Jost and Chao, 2010), the multinomial model plays the central role.

The multinomial model can also be employed in many other applications, but often it provides only an approximation to the “real” model. For example, in Pearson’s Chi-square test applied to a continuous distribution, we have to partition the support of the distribution into a finite number of intervals; in the diversity analysis of a population with unseen species, the multinomial model can only include those species which we have seen. In both cases, when sample size increases, we would wish to adopt a “closer” approximation to the real model. In other words, we intend to use a sequence of multinomial models with an increasing number of disjoint events to approximate the real model, as the sample size increases.

The introduction of this “non-classical” multinomial model was felt necessary for applications in GOF tests, categorical analysis and the statistical analysis of diversity. This extension of the applicability, as we will show later,

is closely associated with the theory of large number of rare events (LNRE) (see Khmaladze, 1988).

Most statistics involved in these applications, such as various GOF test statistics, the ML estimates of diversity indices and the spectral statistics, can be classified as belonging to a class of statistics called divisible statistics, and their asymptotic behavior can be discussed under a unified framework.

In this chapter, we will introduce the definition of classical and non-classical multinomial models and divisible statistics, give examples of some important divisible statistics and outline the overall structure of the thesis.

## 1.1 Multinomial Model

Despite the importance of the multinomial model in applications, the definition is simple. Disregarding the context in different applications, the classical multinomial model can be defined as following.

### 1.1.1 Classical Model

Given a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ , let  $A_1, \dots, A_N$  be mutually disjoint measurable sets of  $\Omega$  (disjoint events) with

$$\bigcup_{i=1}^N A_i = \Omega,$$

so that  $\mathbf{A} = \{A_i\}_{1 \leq i \leq N}$  forms a partition of  $\Omega$ , and let the probability of each event be

$$p_i = \mathbf{P}(A_i).$$



Considering a set of independent samples  $\{\omega_j\}_{1 \leq j \leq n}$ , the frequencies of the events can be defined by

$$\nu_i = \sum_{j=1}^n \mathbf{I}\{\omega_j \in A_i\}.$$

In this model, the number of partitions  $N$  and the probabilities  $\{p_i\}_{1 \leq i \leq N}$  are fixed - they do not change as the sample size  $n$  changes. However, in many applications (see Chapter 3), the number of partitions  $N$  may increase as the sample size increases. This requires introducing the non-classical models, which allow  $N$  to increase and  $\{p_i\}_{1 \leq i \leq N}$  to change as  $n$  increases.

### 1.1.2 Non-Classical Models

In non-classical models, we consider instead a sequence  $\{\mathbf{A}_n\}$  of partitions with  $\mathbf{A}_n = \{A_{ni}\}_{1 \leq i \leq N}$ , where the number of partitions  $N$  increases as sample size  $n$  increases. The corresponding probabilities are

$$p_{ni} = \mathbf{P}(A_{ni}),$$

and the frequencies of the events become

$$\nu_{ni} = \sum_{j=1}^n \mathbf{I}\{\omega_j \in A_{ni}\}.$$

It is worth noting that the classical models can be regarded as a special case of the non-classical models.

## 1.2 Divisible Statistics

### 1.2.1 Definitions

According to Kudlaev (1990), the term divisible statistics can be traced back to Medvedev (1970). In some publications, two synonyms - separable statis-

tics and decomposable statistics - are also often used.

**Definition 1.1.** *The sum*

$$D_n = \sum_{i=1}^N g_{ni}(\nu_{ni}, p_{ni}, n, N).$$

*with  $g_{ni}$  being functions of  $\nu_{ni}$ ,  $p_{ni}$ ,  $n$  and  $N$  is called a divisible statistic.*

The subscript  $i$  in  $g_{ni}$  implies that the functions  $g_{ni}$  may change over  $i$ . Since  $np_{ni}$  is the expectation of  $\nu_{ni}$  and such quantities play crucial role in the asymptotic behavior of divisible statistics in non-classical models, it is more convenient that we adopt the expression

$$D_n = \sum_{i=1}^N g_{ni}(\nu_{ni}, np_{ni}). \quad (1.1)$$

In some models, when the expectations  $np_{ni}$  of the frequencies  $\nu_{ni}$  may tend to infinity, we consider the normalized frequencies

$$Y_{ni} = \frac{\nu_{ni} - np_{ni}}{\sqrt{np_{ni}}}$$

instead of the frequencies  $\nu_{ni}$ , and denote divisible statistics as sum of functions  $h_{ni}$ , i.e.,

$$D_n = \sum_{i=1}^N h_{ni}(Y_{ni}, np_{ni}). \quad (1.2)$$

Sometime the quantities  $g_{ni}$  do not involve  $np_{ni}$ , (1.1) can be reduced to

$$D_n = \sum_{i=1}^N g_{ni}(\nu_{ni}). \quad (1.3)$$

Likewise, in some models, one can reduce (1.2) to

$$D_n = \sum_{i=1}^N h_{ni}(Y_{ni}). \quad (1.4)$$

### Symmetric Divisible Statistics

An important class of divisible statistics is symmetric divisible statistics.

**Definition 1.2.** *The statistic*

$$D_n = \sum_{i=1}^N g_n(\nu_{ni}).$$

*with  $g_n$  independent of  $i$  is called a symmetric divisible statistic.*

Note that in the equiprobable multinomial model with  $p_{ni} = 1/N$ , the divisible statistics with  $g_n(\nu_{ni}, np_{ni})$  are symmetric divisible statistics.

### m-Divisible Statistics

A generalization of divisible statistics which is also of interest in statistical analysis is  $m$ -divisible statistics.

**Definition 1.3.** *The sum of functions*

$$D_n = \sum_{i=1}^{N-m} g_{ni}(\nu_{ni}, \nu_{n,i+1}, \dots, \nu_{n,i+m})$$

*for  $m \geq 1$  is called an  $m$ -divisible statistic.*

Although  $m$ -divisible statistics covers a broader range of statistics, we will limit our discussion to divisible statistics in this thesis.

## 1.2.2 Traditional Goodness of Fit Statistics

### Pearson's Chi-square Statistic

$$\chi^2 = \sum_{i=1}^N \frac{(\nu_{ni} - np_{ni})^2}{np_{ni}} = \sum_{i=1}^N Y_{ni}^2 \quad (1.5)$$

One of the most frequently used statistics in goodness-of-fit and categorical data analysis is Pearson's Chi-square statistic which was introduced by Pearson (1900). The reason for the popularity of this statistic is that its limit distribution in the classical model is known and many other GOF statistics are asymptotically equivalent to it.

### Log-likelihood Ratio Statistic

$$G^2 = \sum_{i=1}^N 2\nu_{ni} \log \frac{\nu_{ni}}{np_{ni}} \quad (1.6)$$

Another popular statistic in goodness-of-fit theory is the log-likelihood ratio statistic, since the likelihood ratio test is the uniformly most powerful test or uniformly most powerful unbiased test (see, e.g., p429, Shao, 2003).

### 1.2.3 Power-divergence Test Statistics

Both Pearson's Chi-square statistic and the log-likelihood ratio statistic can be included in a larger class of statistics, which is the Power-divergence test statistics  $I_n^\lambda$  introduced by Cressie and Read (1984), with definition:

$$2nI_n^\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^N \nu_{ni} \left( \left( \frac{\nu_{ni}}{np_{ni}} \right)^\lambda - 1 \right) \quad \text{for } \lambda \in \mathbb{R}. \quad (1.7)$$

Apart from Pearson's Chi-square statistic ( $\lambda = 1$ ) and log-likelihood ratio statistic ( $\lambda = 0$ ), many other well-known statistics such as Freeman-Tukey statistic ( $\lambda = -1/2$ ), Modified likelihood ratio test statistic ( $\lambda = -1$ ) and the Neyman modified Chi-square statistic ( $\lambda = -2$ ), all belong to this class.

The power-divergence test statistics “linked the traditional test statistics through a single-valued parameter, and provides a way to consolidate and extend the current fragmented literature” (Pardo, 2006). Cressie and Read (1984) also proposed a new GOF statistic with  $\lambda = 2/3$ , and claimed that it is

a promising alternative to Pearson's Chi-square statistic and Log-likelihood Ratio statistic for some sparse data.

### 1.2.4 The $\phi$ -divergence Test Statistics

Another class of statistics, which is highly connected to information theory, is  $\phi$ -divergence statistics. Since the ML estimates of the probabilities  $\mathbf{p}_n = (p_{n1}, \dots, p_{nN})^T$  in multinomial models is

$$\hat{\mathbf{p}}_n = (\hat{p}_{n1}, \dots, \hat{p}_{nN})^T = \left( \frac{\nu_{n1}}{n}, \dots, \frac{\nu_{nN}}{n} \right)^T,$$

the  $\phi$ -divergence goodness-of-test statistic is just the  $\phi$ -divergence measure between  $\hat{\mathbf{p}}_n$  and null hypothesis  $\mathbf{p}_n$ .

$$T_\phi = \frac{2n}{\phi''(1)} \sum_{i=1}^N \phi \left( \frac{\hat{p}_{ni}}{p_{ni}} \right) p_{ni} = \frac{2n}{\phi''(1)} \sum_{i=1}^N \phi \left( \frac{\nu_{ni}}{np_{ni}} \right) p_{ni}, \quad \phi \in \Phi^* \quad (1.8)$$

where  $\phi : (0, \infty) \rightarrow \mathbb{R}$  is twice continuously differentiable function with  $\phi''(1) \neq 0$ . This was introduced simultaneously by Csiszár (1963) and Ali and Silvey (1966).

The  $\phi$ -divergence statistics include most traditional GOF statistics and power divergence statistics with  $\lambda \in \mathbb{R} \setminus (0, 1)$ .

### 1.2.5 $L_1$ -distance

Some other divergence measures such as  $L_p$ -distance between  $\hat{\mathbf{p}}_n$  and  $\mathbf{p}_n$  also belong to the class of divisible statistics. For example,  $L_1$ -distance

$$\sum_{i=1}^N \left| \frac{\nu_{ni}}{n} - p_{ni} \right| \quad (1.9)$$

played an important role in investigating the consistency of the estimates  $\hat{\mathbf{p}}_n$  to  $\mathbf{p}_n$  in LNRE models (Khmaladze, 1988).

### 1.2.6 Estimates of Diversity Indices

In addition to GOF statistics, the ML estimates of diversity indices can also be classified as a divisible statistic. For example, the most widely used diversity measures include Simpson's diversity index (Simpson, 1949)

$$H_{GS} = 1 - D = 1 - \sum_{i=1}^N p_{ni}^2$$

and Shannon's entropy (Shannon, 1948)

$$H = - \sum_{i=1}^N p_{ni} \ln p_{ni},$$

which is also called Shannon's diversity index. It is easy to see that the ML estimates

$$\hat{D} = \sum_{i=1}^N \left( \frac{\nu_{ni}}{n} \right)^2$$

and

$$\hat{H} = - \sum_{i=1}^N \frac{\nu_{ni}}{n} \ln \left( \frac{\nu_{ni}}{n} \right)$$

are both divisible statistics.

### 1.2.7 Spectral Statistics

Another class of divisible statistics, which plays an important role in LNRE theory (Khmaladze, 1988), is that of the so-called spectral statistics

$$\mu_n(m) = \sum_{i=1}^N \mathbf{I}\{\nu_{ni} = m\}$$

and vocabulary

$$\mu_n = \sum_{i=1}^N \mathbf{I}\{\nu_{ni} > 0\}.$$

It is necessary to point out that any symmetric divisible statistic can be expressed as the linear combination of the spectral statistics:

$$\sum_{i=1}^N g_n(\nu_{ni}) = \sum_{m=1}^n g_n(m) \mu_n(m).$$

Therefore, the asymptotic properties of symmetric divisible statistic may be investigated through the spectral statistics.

### 1.3 Aim and Outline of the Thesis

Divisible statistics have been used in statistical analysis for long time, and some classes of divisible statistics have been intensively studied, whereas little attention has been paid to establishing a unified framework to analyze divisible statistics. One purpose of this thesis is to attempt to fill this blank, to construct a general framework for considering divisible statistics, especially their asymptotic properties and applications.

Although some attention will be devoted to the classical multinomial model in showing a unified approach to the asymptotics of divisible statistics, our special emphasis is on asymptotic properties of divisible statistics in non-classical models. Unlike the popular treatments, which consider the divisible statistics only as a whole, we will discuss an advanced approach proposed by Khmaladze (1984), which focuses on the partial sum processes constructed by the divisible statistics, and the martingale and compensator components arising from them. Taking use of modern martingale theory, functional limit theorems (FLT) can be established. Apart from describing this martingale method in detail, we also show that it is possible to extend it to a more general situation and we show some desirable properties and applications of the limiting processes.

Another theme of this thesis is LNRE theory, which was developed by Khmaladze in the 1980s. As an important class of non-classical multinomial models, the LNRE models have many real world applications (see, e.g. Baayen, 2001). In addition to reviewing the general framework of LNRE theory, we discuss an interesting model within the context of diversity of questionnaires and establish limit theorems.

More specifically, the following is the structure of this thesis. In chapter 2, we summarize the limit theorems and applications of divisible statistics in the classical model. In chapter 3, we will illustrate the wide applications of divisible statistics in non-classical multinomial models; introduce some important objects and discuss their asymptotic probabilities, and show the loss of some good properties in non-classical models. A functional limit theorem for partial sum processes will be established in chapter 4. And in chapter 5, we shall discuss the properties and the applications of the limiting processes. Finally, in chapter 6, we will discuss LNRE theory and the analysis of multiple-choice questionnaires.



## Chapter 2

# Divisible Statistics in Classical Models

Since the non-classical multinomial models can be regarded as a sequence of classical models, and many applications of the non-classical model are closely related to applications of the classical model, a thorough review of the application and asymptotic properties of divisible statistics in classical models becomes a good starting point.

Due to a long history and the wide application of divisible statistics in classical multinomial model in statistical analysis, the asymptotic properties of many particular statistics or particular classes of divisible statistics have been thoroughly studied, and many results appear in classical and contemporary books e.g., Rao (1973), Kendall et al. (1987), Read and Cressie (1988), Agresti (2002), and Pardo (2006) etc.

Most of these results focus on those statistics which are asymptotically chi-squared distributed. These results essentially take advantage of the fact that such statistics are asymptotically equivalent to Pearson's Chi-square statistic, of which the limit distribution is known.

However, once we realize that the key to those results is the asymptotic normality of the normalized frequencies, we do not need to restrict ourselves to those asymptotically chi-squared distributed statistics. Based on this point, a unified approach to the limit theorems of more general divisible statistics can be established, of which the result for the Chi-square statistic is simply a natural corollary.

In section 2.1, we will first list some typical applications of divisible statistics in classical models. Then, in section 2.2, we will discuss the exact distributions of divisible statistics and the motivation for investigating limit theorems. Finally, in section 2.3, we will discuss the limit theorems of divisible statistics based on a unified approach.

## 2.1 Applications

### 2.1.1 GOF Test with Specified Distributions

The GOF test is one of the major areas of applications of divisible statistics, and tests how well the model fits a set of observations. If the hypothetical distribution is a finite discrete distribution, i.e.,

$$H_0 : \mathbf{p} = \mathbf{p}^0$$

with  $\mathbf{p}^0 = (p_1, \dots, p_i, \dots, p_N)^T$ , then this is exactly the classical multinomial model. A typical example is the so-called “test of discrete uniformity” with all  $p_i = 1/N$ .

If the hypothetical distribution is a continuous distribution or infinite discrete distribution, then the null hypothesis may be expressed as

$$H_0 : F = F^0.$$

Although the classical multinomial model is not applicable directly to this problem, by partitioning or grouping the sample space into  $N$  subsets  $\{A_i\}_{1 \leq i \leq N}$  (in the case of real line, for example, partitioning the real line into  $N$  disjoint intervals), we can approximate the hypothesis by

$$H_0 : \mathbf{p} = \mathbf{p}^0$$

with  $\mathbf{p}^0 = (p_1, \dots, p_i, \dots, p_N)^T$  and all  $p_i = \int_{A_i} dF^0$ .

Various divisible statistics such as described in (1.5-1.8), which measure the distance between the ML estimates  $\hat{\mathbf{p}}$  and the null hypothetical probabilities  $\mathbf{p}^0$ , are often used here to implement the goodness-of-fit test. These divisible statistics are usually of the form

$$D_n = \sum_{i=1}^N g_{ni}(\nu_i, np_i) = \sum_{i=1}^N h_{ni}(Y_i, np_i).$$

### 2.1.2 GOF Test with Parameters Estimated

In some applications, the probabilities are not fully specified. These probabilities are functions of some vector of parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)^T$ , but the values of these parameters are not specified. In this case, the null hypothesis becomes

$$H_0 : \mathbf{p} = \mathbf{p}^0(\boldsymbol{\theta})$$

with  $\mathbf{p}^0(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_N(\boldsymbol{\theta}))^T$ .

Since the parameters  $\boldsymbol{\theta}$  are not specified, the probabilities  $\mathbf{p}^0(\boldsymbol{\theta})$  are also unspecified. To test the hypothesis, we usually construct test statistics with  $\boldsymbol{\theta}$  replaced by some (usually ML) estimates  $\hat{\boldsymbol{\theta}}_n$ , i.e.  $\hat{\mathbf{p}}_n^0(\boldsymbol{\theta}) = (p_1(\hat{\boldsymbol{\theta}}_n), \dots, p_N(\hat{\boldsymbol{\theta}}_n))^T$ . The goodness-of-fit test with estimated parameters have broad applications in goodness-of-fit and categorical analysis. Below we will show some examples.

**Pearson's Chi-square Test for Normality**

Although there are many other competing alternative approaches to test normality, Pearson's Chi-square test for normality is a frequently used one. This method can be extended to test other family of distribution, such as test of exponentiality.

Considering a set of independent random variables  $X_1, \dots, X_j, \dots, X_n$ , and we want to test if the sample comes from the normal distribution. We can first estimate  $\hat{\mu}$  and  $\hat{\sigma}$  based on the sample,<sup>1</sup> then partition the whole real line into  $N$  intervals  $A_i$  and define

$$\hat{p}_i = \int_{A_i} \phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right) dx$$

with  $\phi$  is standard normal density function. We can also find the frequencies,

$$\hat{\nu}_i = \sum_{j=1}^n \mathbf{I}\{X_j \in A_i\}$$

and construct the test statistic

$$\hat{\chi}^2 = \sum_{i=1}^N \frac{(\nu_i - n\hat{p}_i)^2}{n\hat{p}_i}.$$

The Chi-square statistic is not the only one for conducting this type of test; other statistics such as log-likelihood ratio, power-divergence (see, e.g. Read and Cressie, 1988) and  $\phi$ -divergence (see, e.g. Pardo, 2006) test statistics can be used as well.

---

<sup>1</sup>The procedure described here estimate the parameter based on the observations, rather than multinomial frequencies. This will lead to different asymptotic distribution, which was discussed in detail in 30.15-30.16 of Kendall et al. (1987). Although we may carry out the ML estimation based on the multinomial frequencies, "it is almost always difficult" (p723, Devore and Berk, 2006).

### Test of Independence

The structure of dependence between two random variables is often of great interest in statistical analysis. If two variables  $X$  and  $Y$  have both finite discrete distributions (with sample space  $(1, \dots, I)$  and  $(1, \dots, J)$  respectively), the structure of dependence can be described by a  $I \times J$  table with each entry being the joint probability,

$$p_{ij} = \mathbf{P}(X = i, Y = j).$$

After introducing the frequencies

$$\nu_{ij} = \sum_{k=1}^n \mathbf{I}\{X_k = i, Y_k = j\},^2$$

based on  $n$  observations, a test statistic

$$D_n = \sum_{i=1}^I \sum_{j=1}^J g_{ij}(\nu_{ij}, np_{ij})$$

can be constructed to test if the hypothetical structure of dependence is valid.

In this case of testing independence, the null hypothesis is,

$$H_0 : p_{ij} = p_{i.}p_{.j} \quad \forall \quad i, j$$

with  $p_{i.} = \mathbf{P}(X = i)$  and  $p_{.j} = \mathbf{P}(Y = j)$  remaining unspecified. This null hypothesis is composite and the ML estimator  $\hat{p}_{i.} = \sum_{j=1}^J \nu_{ij}/n$  and  $\hat{p}_{.j} = \sum_{i=1}^I \nu_{ij}/n$  are used to construct test statistic

$$D_n = \sum_{i=1}^I \sum_{j=1}^J g_{ij}(\nu_{ij}, n\hat{p}_{ij}).$$

---

<sup>2</sup>The  $I \times J$  table with entries  $\nu_{ij}$  is a two-way contingency table.

### Test of Homogeneity

The test of homogeneity is designed to check if a set of random variables  $X_1, \dots, X_I$  with common sample space  $(1, \dots, J)$  are all of the same distribution but with unknown probabilities. Setting the sample sizes of  $X_1, \dots, X_I$  to be  $n_1, \dots, n_I$  respectively, the test statistic is based on the frequencies

$$\nu_{ij} = \sum_{k=1}^{n_i} \mathbf{I}\{X_{ik} = j\}.$$

The null hypothesis is

$$H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_I = \mathbf{p}^0$$

with  $\mathbf{p}^0 = (p_1, \dots, p_J)^T$ . This is equivalent to

$$H_0 : p_{ij} = p_j, \quad j \in (1, \dots, J)$$

with  $p_j$ 's being unspecified.

The ML estimates of  $p_j$  in this case is  $\hat{p}_j = \left( \sum_{i=1}^I \nu_{ij} \right) / n$  with  $n = \sum_{i=1}^I n_i$  and the test statistic is of the form,

$$D_n = \sum_{i=1}^I \sum_{j=1}^J g_{ij}(\nu_{ij}, n\hat{p}_{ij}).$$

### Test of Symmetry

In the two-way contingency table, when the number of rows  $I$  is same as the number of columns  $J$ , it is often interesting to investigate whether there are symmetric patterns in the data. In terms of null hypothesis, it is

$$H_0 : p_{ij} = p_{ji}.$$

This is also a problem of goodness-of-fit with composite null hypothesis and a similar approach can be used to construct test statistics.

Strictly speaking, the test statistics discussed above are not divisible statistics in themselves, but the divisible statistics with estimated parameters, and usually of the form

$$D_n = \sum_{i=1}^N g_{ni}(\nu_{ni}, n\hat{p}_i) = \sum_{i=1}^N h_{ni}(\hat{Y}_{ni}, n\hat{p}_i),$$

with

$$\hat{Y}_{ni} = \frac{\nu_{ni} - n\hat{p}_{ni}}{\sqrt{n\hat{p}_{ni}}}.$$

### 2.1.3 Measures of Diversity

The concept of diversity appears in many different research areas and is a popular topic in some areas such as biology (Magurran, 2004), ecology (Williams, 1964), genetics, economics, linguistics, etc. Hence a mathematical model of measure of diversity is often of great interest. Based on Rao (1982), a general definition of the measure of diversity in multinomial model could be,

**Definition 2.1.** *Let  $\mathcal{P}$  be the set of all possible probability vectors  $\mathbf{p} = (p_1, \dots, p_N)^T$ . A function  $H$  mapping  $\mathcal{P}$  into the real line is said to be a measure of diversity if it satisfies the following conditions:*

*i  $H(\mathbf{p}) \geq 0 \quad \forall \mathbf{p} \in \mathcal{P}$  and  $H(\mathbf{p}) = 0$  if and only if  $\mathbf{p}$  is degenerate.*

*ii  $H$  is a concave function on  $\mathcal{P}$ .*

This definition is far more general than what we are going to discuss. In this thesis, we consider only a subclass with an extra constraint

$$H(\mathbf{p}) = \sum_{i=1}^N H_i(p_i),$$

such that it can be a class of divisible statistics.

This class include Simpson's and Shannon's diversity indices and many others, such as the diversity index of degree- $\beta$

$$H^\beta(\mathbf{p}) = \begin{cases} \left(1 - \sum_{i=1}^N p_i^{\beta+1}\right) / \beta & \beta > 0 \\ \lim_{\beta \rightarrow 0} \left(1 - \sum_{i=1}^N p_i^{\beta+1}\right) / \beta & \beta = 0 \end{cases}$$

described by Patil and Taillie (1982) and  $\phi$ -entropy measure (Burbea and Rao, 1982b,a)

$$H^\phi(\mathbf{p}) = \sum_{i=1}^N \phi(p_i),$$

where  $\phi : (0, \infty) \rightarrow \mathbb{R}$  is a continuous concave function.

Although the definition of measures of diversity is based on the probabilities  $p_i$ , this is of less interest in practice, since these probabilities are usually unknown. Instead, the estimates of the measures of diversity are often used. Since the natural estimates of  $p_i$ 's are ML estimates  $\hat{p}_i = \nu_i/n$ , the estimates of the measure of diversity

$$\hat{H}(\mathbf{p}) = H\left(\frac{\boldsymbol{\nu}}{n}\right) = \sum_{i=1}^N H_i\left(\frac{\nu_i}{n}\right)$$

are divisible statistics.

In classic model, it is easy to see, by the law of large numbers, that

$$H\left(\frac{\boldsymbol{\nu}}{n}\right) \xrightarrow{a.s.} H(\mathbf{p}) = \sum_{i=1}^N H_i(p_i). \quad (2.1)$$

## 2.2 Exact Distributions of Divisible Statistics

In principle, since the joint distribution of the frequencies is simply the multinomial,

$$\mathbf{P}\{\boldsymbol{\nu} = \mathbf{k}\} = \frac{n!}{\prod_{i=1}^N k_i!} \prod_{i=1}^N p_i^{k_i}, \quad (2.2)$$



the exact distribution of the divisible statistics under the null hypothesis can be calculated.

The comparison between exact distributions and limiting distributions of Chi-square and exact log-likelihood test, especially in small sample cases, was a popular topic and many corrections and adjustments were proposed to improve the accuracy of limiting distributions to exact distributions (see e.g. Read and Cressie, 1988 for reviews).

However, the procedures of calculating exact distribution are usually computationally intensive, and hence it is only feasible in some small-sample cases, where both  $n$  and  $N$  are not large. In fact, as we can see in figure 2.1, the approximation of the limiting distribution to the exact distribution is good enough even when the sample size is not very large. Therefore, the limit distributions are of more interest.

Another reason of favoring limit distributions is that, in the case of estimated parameters, the exact distribution is usually not easy to obtain. Moreover, all exact distributions depend on the hypothetical distributions (probabilities) while the limit distributions are distribution-free.

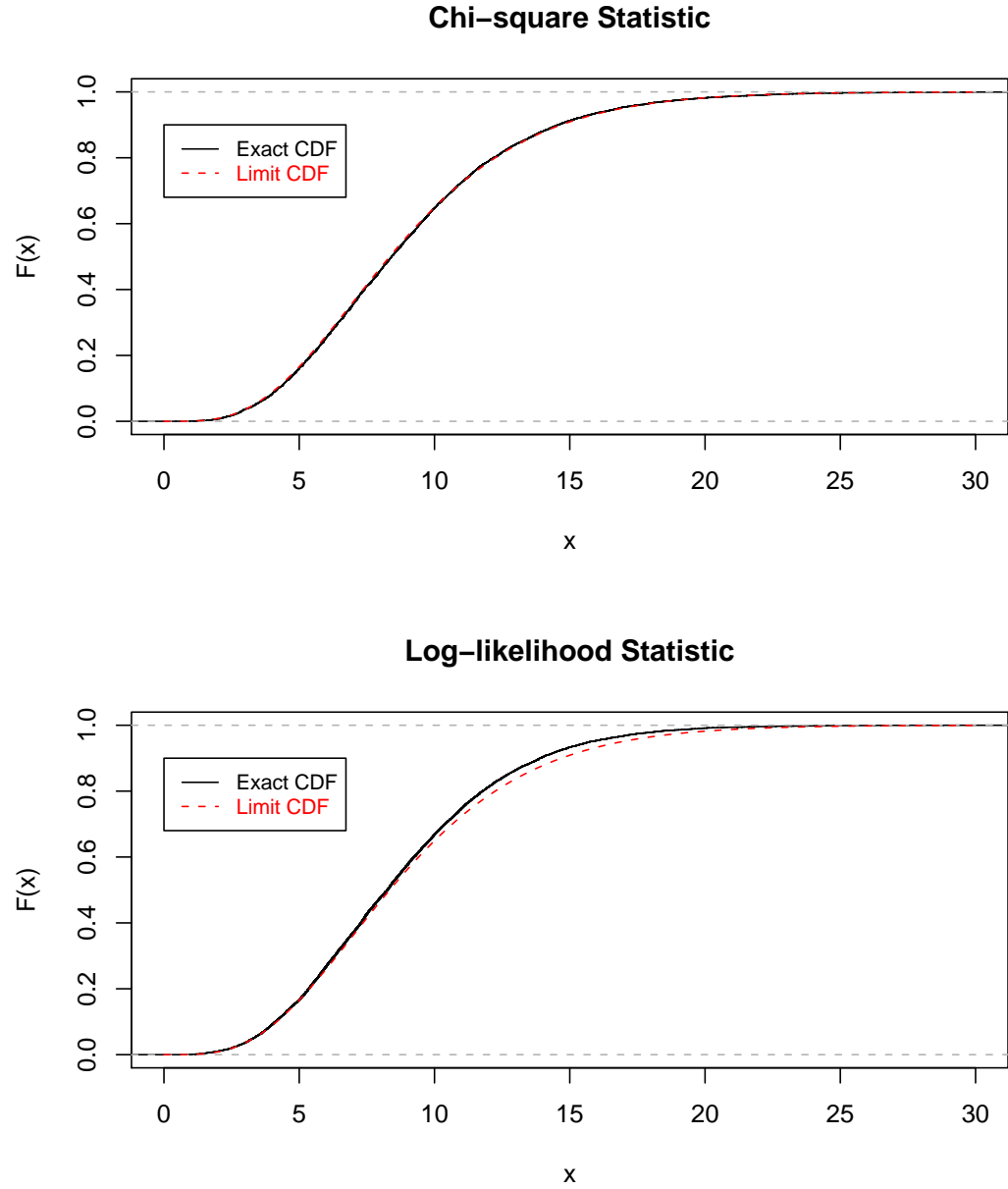


Figure 2.1: Comparison of exact and limit distribution of divisible statistics:  $n=50$ ,  $N=10$ ,  $\mathbf{p}=(0.10, 0.06, 0.10, 0.13, 0.12, 0.13, 0.06, 0.10, 0.10, 0.10)$

## 2.3 Limit Distributions in Classical Models

The limit theorems for some, or some classes of, divisible statistics such as Chi-square and log-likelihood statistics have been established for long time. In this section, we discuss the asymptotic behaviour of general divisible statistics from a different angle.

### 2.3.1 Limit Distributions of the Normalized Frequencies

There are several types of limit distributions of divisible statistics which are of interest. Apart from the limit distribution under the null hypothesis and with estimated parameters, the limit distributions under the contiguous alternatives are also interesting. Correspondingly, we need to establish the limit theorems of the vector of normalized frequencies under these three situations. It is revealed that the limit distributions are all multivariate normal distribution, but with different mean vector and covariance matrix. The proofs in this section basically follow the approach of Kendall et al. (1987) and Rao (1973).

#### Limit Distributions Under Null Hypothesis

We firstly establish the limit distribution of the normalized frequencies under null hypothesis.

**Theorem 2.1.** *Define the vector  $\mathbf{q} = (\sqrt{p_1^0}, \dots, \sqrt{p_i^0}, \dots, \sqrt{p_N^0})^T$ . Under the null hypothesis*

$$H_0 : \mathbf{p} = \mathbf{p}^0 = (p_1^0, \dots, p_i^0, \dots, p_N^0)^T$$

*we have*

$$\mathbf{Y}_n \xrightarrow{d} \mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I} - \mathbf{q}\mathbf{q}^T).$$

*Proof.* It is easy to see that  $\mathbb{E}\mathbf{Y}_n = \mathbf{0}$  and the variance-covariance matrix of  $\mathbf{Y}_n$  is

$$\mathbb{E} \left[ (\mathbf{Y}_n - \mathbb{E}\mathbf{Y}_n) (\mathbf{Y}_n - \mathbb{E}\mathbf{Y}_n)^T \right] = \mathbf{C} = \mathbf{I} - \mathbf{q}\mathbf{q}^T$$

for any  $n$ . By the central limit theorem (CLT) for multivariate random variables,  $\mathbf{Y}_n$  converges in distribution to a normal vector. Hence, the theorem is proved.  $\square$

### Limit Distributions Under Contiguous Alternatives

Apart from the limit distribution under the null hypothesis, the limit distributions of divisible statistics under contiguous alternatives are often of great interest.

**Theorem 2.2.** *Define the vector  $\mathbf{q} = (\sqrt{p_1^0}, \dots, \sqrt{p_i^0}, \dots, \sqrt{p_N^0})^T$  and  $\mathbf{Q} = \text{diag}(\mathbf{q})$ . Consider a sequence of contiguous alternatives to the null hypothesis as  $n$  increases, i.e.*

$$H_1^{(n)} : \mathbf{p} = \mathbf{p}_n = (p_{n1}, \dots, p_{ni}, \dots, p_{nN})^T = \mathbf{p}^0 + \frac{\mathbf{d}}{\sqrt{n}}$$

with  $\mathbf{d} = (d_1, \dots, d_N)^T$  such that  $\sum_{i=1}^N d_i = 0$ . Under any such alternatives, we have

$$\mathbf{Y}_n \xrightarrow{d} \mathbf{Y} + \mathbf{Q}^{-1}\mathbf{d} \sim \mathcal{N}(\mathbf{Q}^{-1}\mathbf{d}, \mathbf{I} - \mathbf{q}\mathbf{q}^T).$$

*Proof.* Let  $\mathbf{q}_n = (\sqrt{p_{n1}^0}, \dots, \sqrt{p_{ni}^0}, \dots, \sqrt{p_{nN}^0})^T$  and  $\mathbf{Q}_n = \text{diag}(\mathbf{q}_n)$ . Then  $\mathbf{q}_n \rightarrow \mathbf{q}$  implies, under  $H_1^{(n)}$ ,

$$\tilde{\mathbf{Y}}_n = \mathbf{Q}_n^{-1}(\boldsymbol{\nu}_n - n\mathbf{p}_n) / \sqrt{n} \xrightarrow{d} \mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I} - \mathbf{q}\mathbf{q}^T).$$

Since

$$\mathbf{Q}_n \tilde{\mathbf{Y}}_n = (\boldsymbol{\nu}_n - n\mathbf{p}_n) / \sqrt{n} = (\boldsymbol{\nu}_n - n\mathbf{p}^0) / \sqrt{n} - \mathbf{d} = \mathbf{Q}\mathbf{Y}_n - \mathbf{d}$$

By Slutsky's theorem, the theorem is proved.  $\square$

### Limit Distributions With Estimated Parameter

The limit behaviour when parameters are estimated is more complicated. However, under suitable assumptions, the limit distribution of the normalized frequencies can be derived.

**Theorem 2.3.** *Consider the composite null hypothesis*

$$H_0 : \mathbf{p} = \mathbf{p}^0(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_N(\boldsymbol{\theta}))^T$$

with  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)^T$  and  $\mathbf{p}^0$  is differentiable at  $\boldsymbol{\theta}$ . Define the vector  $\mathbf{q} = \left( \sqrt{p_1(\boldsymbol{\theta})}, \dots, \sqrt{p_N(\boldsymbol{\theta})} \right)^T$  and matrix  $\mathbf{Q} = \text{diag}(\mathbf{q})$ . If the ML estimates  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  is regular<sup>3</sup>, i.e.,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = (\mathbf{A}\mathbf{Y}_n + o(1)) / \sqrt{n}, \quad (2.3)$$

with  $\mathbf{Y}_n = \mathbf{Q}^{-1}(\boldsymbol{\nu}_n - n\mathbf{p}^0(\boldsymbol{\theta})) / \sqrt{n}$  and some  $(s \times N)$  matrix  $\mathbf{A}$ , then

$$\hat{\mathbf{Y}}_n = \hat{\mathbf{Q}}^{-1}(\boldsymbol{\nu}_n - n\mathbf{p}^0(\hat{\boldsymbol{\theta}})) / \sqrt{n} \xrightarrow{d} (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{B}\mathbf{A}) \mathbf{Y}.$$

Where  $\hat{\mathbf{Q}} = \text{diag}(\hat{\mathbf{q}})$  with  $\hat{\mathbf{q}} = \left( \sqrt{p_1(\hat{\boldsymbol{\theta}})}, \dots, \sqrt{p_N(\hat{\boldsymbol{\theta}})} \right)^T$  and  $\mathbf{B}$  is a  $(N \times s)$  matrix with elements

$$b_{ij} = \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j} \frac{1}{p_i(\boldsymbol{\theta})^{\frac{1}{2}}}$$

*Proof.* We can express

$$\hat{\mathbf{Q}}\hat{\mathbf{Y}}_n = (\boldsymbol{\nu}_n - n\mathbf{p}^0(\hat{\boldsymbol{\theta}})) / \sqrt{n} = \mathbf{Q}\mathbf{Y}_n - \sqrt{n}(\mathbf{p}^0(\hat{\boldsymbol{\theta}}) - \mathbf{p}^0(\boldsymbol{\theta})).$$

By a Taylor expansion, and considering (2.3), we have

$$\sqrt{n}(\mathbf{p}^0(\hat{\boldsymbol{\theta}}) - \mathbf{p}^0(\boldsymbol{\theta})) = \mathbf{Q} \left[ \sqrt{n}\mathbf{B}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1) \right] = \mathbf{Q}[\mathbf{B}\mathbf{A}\mathbf{Y}_n + o_p(1)].^4$$

By Slutsky's theorem,

$$\hat{\mathbf{Y}}_n = \hat{\mathbf{Q}}^{-1}\mathbf{Q}[(\mathbf{I} - \mathbf{B}\mathbf{A})\mathbf{Y}_n + o_p(1)] \xrightarrow{d} (\mathbf{I} - \mathbf{B}\mathbf{A})\mathbf{Y}.$$

□

<sup>3</sup>See 30.10 of Kendall et al. (1987) for detail.

<sup>4</sup>We say a sequence of random variables  $X_n = o_p(1)$  if  $X_n \xrightarrow{p} 0$  and  $X_n = O_p(1)$  if  $X_n$  is bounded in probability.

According to (30.14, Kendall et al., 1987), if  $\hat{\boldsymbol{\theta}}$  is the multinomial ML estimator, we have

$$\mathbf{B}^T \mathbf{q} = \mathbf{0} \quad (2.4)$$

and

$$\mathbf{A} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T.$$

Therefore,

$$\hat{\mathbf{Y}}_n \xrightarrow{d} (\mathbf{I} - \mathbf{B}\mathbf{A}) \mathbf{Y} = (\mathbf{I} - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) \mathbf{Y}.$$

### 2.3.2 Limit Distributions of Divisible Statistics

Based on the limit distributions of the normalized frequency vector, the limit distributions of divisible statistics can be obtained, which can be expressed as the distribution of a function of the limiting normal vector.

**Theorem 2.4.** *Consider a divisible statistic*

$$D_n = \sum_{i=1}^N h_i(Y_{ni}, np_i) = \sum_{i=1}^N h_i^*(Y_{ni}, \epsilon_{ni})$$

with  $Y_{ni} = (\nu_{ni} - np_i) / \sqrt{np_i}$  and  $\epsilon_{ni} = 1/np_i$ . If  $h_i^*$  are continuous at all points  $(y, 0)$ , and

$$\mathbf{Y}_n = (Y_{n1}, \dots, Y_{ni}, \dots, Y_{nN})^T \xrightarrow{d} \mathbf{Y}^* = (Y_1^*, \dots, Y_i^*, \dots, Y_N^*)^T$$

then

$$D_n \xrightarrow{d} \sum_{i=1}^N h_i^*(Y_i^*, 0) = h^*(\mathbf{Y}^*).$$

*Proof.* The proof is easily based on the Mann-Wald theorem (Continuous mapping theorem) and the fact that all  $np_i \rightarrow \infty$ .  $\square$

**Example:** For log-likelihood Ratio Statistic

$$G^2 = 2 \sum_{i=1}^N \nu_{ni} \log \frac{\nu_{ni}}{np_{ni}} = 2 \sum_{i=1}^N \left[ np_i \left( 1 + \frac{Y_{ni}}{\sqrt{np_i}} \right) \log \left( 1 + \frac{Y_{ni}}{\sqrt{np_i}} \right) - \sqrt{np_i} Y_{ni} \right],$$

we have

$$h_i^*(y, 0) = 2 \lim_{\epsilon \rightarrow 0} \frac{(1 + y\sqrt{\epsilon}) \log(1 + y\sqrt{\epsilon}) - y\sqrt{\epsilon}}{\epsilon} = y^2$$

and hence

$$G^2 \xrightarrow{d} \sum_{i=1}^N Y_i^{*2} = \mathbf{Y}^{*T} \mathbf{Y}^*.$$

Note that this theorem applies for general divisible statistics, which are not necessarily asymptotically chi-square distributed, as shown in Figure 2.2.

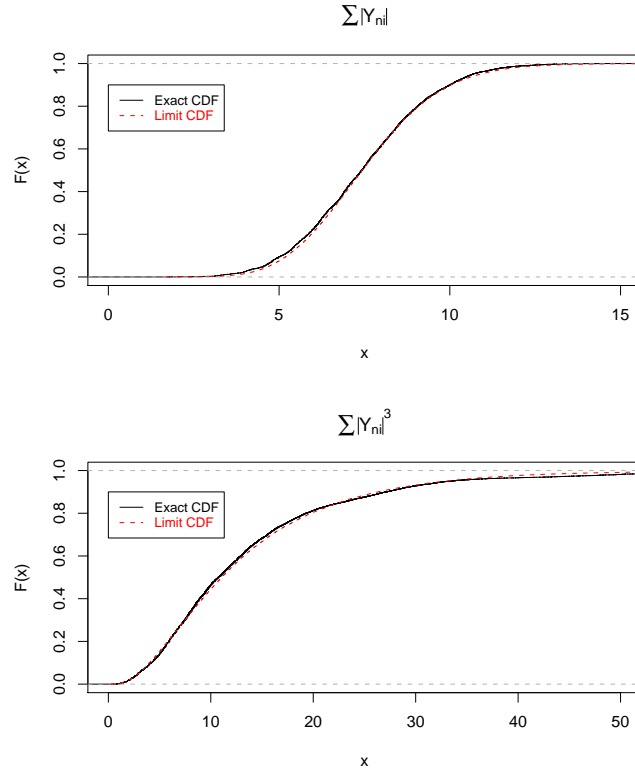


Figure 2.2: Limit distributions of general divisible statistics:  $\sum_{i=1}^N |Y_{ni}|$  and  $\sum_{i=1}^N |Y_{ni}|^3$  with  $n=20$ ,  $N=10$ ,  $\mathbf{p}=(0.10, 0.06, 0.10, 0.13, 0.12, 0.13, 0.06, 0.10, 0.10, 0.10)$ . The limit distribution is calculated by Monte Carlo simulation based on Theorem 2.4. The generation of the multivariate normal vector uses R MVTNORM package (Genz et al., 2010) .

Based on this theorem, the limit distributions of divisible statistics under the null hypothesis, contiguous alternatives or estimated parameters can be easily acquired, with  $\mathbf{Y}^*$  replaced by  $\mathbf{Y}$ ,  $\mathbf{Y} + \mathbf{Q}^{-1}\mathbf{d}$  and  $(\mathbf{I} - \mathbf{Q}^{-1}\mathbf{BA})\mathbf{Y}$  respectively.

### 2.3.3 Asymptotic Equivalence to Chi-square Statistic

Recall that, in general, the limit distributions  $\sum_{i=1}^N h_i^*(Y_i, 0)$  of divisible statistics are not necessarily chi-squared. However, many widely used goodness-of-fit test statistics are asymptotically equivalent to Chi-square statistics in the classical multinomial model. Theorem (2.5) shows the conditions of this asymptotic equivalence.

**Theorem 2.5.** *If the divisible statistic is of the form*

$$D_n = \sum_{i=1}^N h_i \left( \frac{\nu_{ni}}{np_i} \right) np_i$$

*with  $h_i$ 's being continuously twice differentiable with  $h_i(1) = 0$  and  $h_i''(1) = 2$  then  $D_n$  is asymptotically equivalent to Pearson's Chi-square statistics, i.e.,*

$$D_n \xrightarrow{d} \mathbf{Y}^{*T} \mathbf{Y}^*$$

*as  $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}^*$ .*

*Proof.* Since

$$D_n = \sum_{i=1}^N h_i \left( \frac{\nu_{ni}}{np_i} \right) np_i = \sum_{i=1}^N np_i \left[ h_i \left( 1 + \frac{Y_i}{\sqrt{np_i}} \right) - h_i'(1) \frac{Y_i}{\sqrt{np_i}} \right]$$

and

$$\lim_{\epsilon \rightarrow 0} \frac{h_i(1 + y\sqrt{\epsilon}) - h_i(1) - h_i'(1)y\sqrt{\epsilon}}{\epsilon} = h_i''(1) \frac{y^2}{2} = y^2.$$

Invoking theorem 2.4, we prove the theorem. □



It is easy to verify that most widely used goodness-of-fit test statistics, such as the power-divergence statistics and  $\phi$ -divergence statistics, belong to this class.

### 2.3.4 Limit Distributions of Chi-square Statistic

Unlike general functions  $h^*(\mathbf{Y}^*)$ , the quadratic form  $\mathbf{Y}^{*T}\mathbf{Y}^*$  has a closed-form distribution. Based on the following Lemma, the limit distributions of divisible statistics, which are asymptotically equivalent to Chi-square Statistics, can be established.

**Lemma 2.1.** (*p63, Ferguson, 1996*) Suppose  $\mathbf{Y}$  is  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . If  $\boldsymbol{\Sigma}$  is idempotent of rank  $r$  and  $\boldsymbol{\Sigma}\boldsymbol{\mu} = \boldsymbol{\mu}$ , the distribution of  $\mathbf{Y}^T\mathbf{Y}$  is noncentral chi-square with  $r$  degrees of freedom and noncentrality parameter  $\boldsymbol{\mu}^T\boldsymbol{\mu}$ .

**Corollary 2.1.** When  $\mathbf{Y}^* = \mathbf{Y}$ ,

$$\mathbf{Y}^{*T}\mathbf{Y}^* \sim \chi_{N-1}^2(0)$$

*Proof.* By theorem 2.1,  $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T$ . Since  $\mathbf{q}^T\mathbf{q} = 1$ ,

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{q}\mathbf{q}^T)(\mathbf{I} - \mathbf{q}\mathbf{q}^T) = \mathbf{I} - 2\mathbf{q}\mathbf{q}^T + \mathbf{q}\mathbf{q}^T\mathbf{q}\mathbf{q}^T = \mathbf{I} - \mathbf{q}\mathbf{q}^T = \boldsymbol{\Sigma},$$

$\boldsymbol{\Sigma}$  is idempotent. Then since  $\text{rank}(\boldsymbol{\Sigma}) = \text{trace}(\boldsymbol{\Sigma}) = N - 1$  and  $\boldsymbol{\mu} = \mathbf{0}$ , the corollary is proved.  $\square$

**Corollary 2.2.** When  $\mathbf{Y}^* = \mathbf{Y} + \mathbf{Q}^{-1}\mathbf{d}$ ,

$$\mathbf{Y}^{*T}\mathbf{Y}^* \sim \chi_{N-1}^2\left(\sum_{i=1}^N \frac{d_i^2}{p_i^0}\right)$$

*Proof.* By theorem 2.2,  $\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q}\mathbf{q}^T$  is idempotent and of rank  $N - 1$ . Since  $\boldsymbol{\mu} = \mathbf{Q}^{-1}\mathbf{d}$ ,

$$\boldsymbol{\mu}^T\boldsymbol{\mu} = \mathbf{d}^T (\mathbf{Q}^{-1})^T (\mathbf{Q}^{-1}) \mathbf{d} = \sum_{i=1}^N \frac{d_i^2}{p_i^0}$$

the corollary is proved.  $\square$

**Corollary 2.3.** *When  $\mathbf{Y}^* = (\mathbf{I} - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) \mathbf{Y}$ ,*

$$\mathbf{Y}^{*T} \mathbf{Y}^* \sim \chi_{N-1-s}^2(0)$$

*Proof.* By theorem 2.1 and theorem 2.3,

$$\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) (\mathbf{I} - \mathbf{q} \mathbf{q}^T) (\mathbf{I} - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T)^T.$$

Since  $(\mathbf{I} - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T)$  is idempotent and  $\mathbf{B}^T \mathbf{q} = \mathbf{0}$  by (2.4),

$$\boldsymbol{\Sigma} \boldsymbol{\Sigma} = \mathbf{I} - \mathbf{q} \mathbf{q}^T - \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T.$$

Since  $\text{rank}(\boldsymbol{\Sigma}) = \text{trace}(\boldsymbol{\Sigma}) = \text{trace}(\mathbf{I} - \mathbf{q} \mathbf{q}^T) - \text{trace}(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) = N - 1 - s$  and  $\boldsymbol{\mu} = \mathbf{0}$ , the corollary is proved.  $\square$

This alternative establishment of the limit theorems emphasizes the important role played by normalized frequencies in asymptotic behaviour of divisible statistics in classical multinomial models.

## Chapter 3

# Divisible Statistics in Non-classical Models

As we have shown in chapter 2, the asymptotic properties of divisible statistics essentially rely on the asymptotic normality of the normalized frequencies, which depend on the fact that all  $np_{ni}$  diverge to infinity and all the frequencies are asymptotically Gaussian. In particular, the marginal distribution of the normalized frequencies converges to the Gaussian.

However, in non-classical multinomial models, as the sample size increases, the number of disjoint events also increases. Many probabilities tend to 0 such that some frequencies are asymptotically Poissonian in the sense that the marginal distribution of the frequencies converge to a Poisson distribution. Therefore, the asymptotics of divisible statistics in non-classical models are much more complicated.

In this chapter, we will firstly discuss the asymptotic properties of some objects which play an important role in non-classical multinomial models. Then, in section 3.2, we will illustrate with some typical examples. Finally, we will discuss difficulties in non-classical asymptotics of divisible statistics in section 3.3.

### 3.1 Some Objects in Non-classical Models

Although it looks simple, it is necessary to discuss some objects in non-classical models and their relations before we proceed to further discussion.

#### 3.1.1 Objects

The most important objects in non-classical multinomial models are the vectors of probabilities  $\mathbf{p}_n = \{p_{ni}\}_{1 \leq i \leq N}$ , since they specify the model. The parameter  $N$  (more properly expressed as  $N(n)$ ) is naturally associated to  $\mathbf{p}_n$ , and when we are interested in the asymptotic behaviour of individual event,

$$f_{ni} = Np_{ni}$$

is often of interest. However, when we study the overall properties of the probability vectors, the distribution function, defined as a partial sum of the probabilities

$$F_n(t) = \sum_{i=1}^{Nt} p_{ni} = \int_0^t f_n(s) ds, \quad t \in [0, 1]$$

with  $f_n(t) = Np_{n[Nt]}$ , will become one of the key object.

Another interesting parameter is  $n$ . In fact,  $n$  plays two roles in this scheme. One is the sample size and another is index of the sequence. As sample size,  $np_{ni}$  is simply the expectation of the frequencies  $\nu_{ni}$ , and can be expressed as

$$\lambda_{ni} = np_{ni}.$$

When we focus on the overall properties, the continuous version

$$\lambda_n(t) = np_{n[Nt]}$$

are important. Similar to  $f_n(t)$  and  $F_n(t)$ , we may define

$$\Lambda_n(t) = \sum_{i=1}^{Nt} np_{ni} \frac{1}{N} = \int_0^t \lambda_n(s) ds, \quad t \in [0, 1]$$

Since both  $n$  and  $N$  tend to infinity in non-classical models, the “rate per cell”  $d_n = n/N$  is often crucial as a measure of the relative rates of increasing. Typically, it is assumed that  $d_n \rightarrow d \in (0, \infty)$ , and sometimes, one simply assumes  $n = N$ . The models with  $d_n \rightarrow 0$  are called the LNRE models with very rare events, and the functional limit theorems in these models have been studied by Mnatsakanov (1986, 1987). In this thesis, we only consider the cases of  $d_n \rightarrow \infty$  and  $d_n \rightarrow d \in (0, \infty)$ .

### 3.1.2 Limit behavior

The objects introduced above specified the non-classical multinomial models. In the discussion of asymptotic behaviour of divisible statistics, the limit behaviours of these objects are of interest.

For instance, when we discuss the asymptotic behaviour of the partial sum processes in chapter 4 and 5, it is necessary to require that the limits  $\lambda(t) = \lim \lambda_n(t)$ ,  $\Lambda(t) = \lim \Lambda_n(t)$ ,  $f(t) = \lim f_n(t)$ , and  $F(t) = \lim F_n(t)$  exist for all  $t \in [0, 1]$ .

For different limit of  $d_n$ , the relations between  $\lambda(t)$ ,  $\Lambda(t)$  and  $f(t)$ ,  $F(t)$  would be:

- $d_n \rightarrow 0$ ,

$$\begin{aligned} \Lambda(t) &= 0 \quad \forall \quad t \in [0, 1] \\ \lambda(t) &\begin{cases} = 0 & f(t) < \infty \\ > 0 & f(t) = \infty \end{cases} \quad \& \quad \liminf(d_n f_{n[Nt]}) > 0 \end{aligned}$$

- $d_n \rightarrow d \in (0, \infty)$ ,

$$\Lambda(t) = d \cdot F(t)$$

$$\lambda(t) = d \cdot f(t)$$

- $d_n \rightarrow \infty$ ,

$$\Lambda(t) = \infty \quad \text{if} \quad F(t) > 0$$

$$\lambda(t) \begin{cases} = \infty & f(t) > 0 \\ < \infty & f(t) = 0 \end{cases} \quad \& \quad \limsup(d_n f_{n[Nt]}) < \infty$$

Based on the limit of  $\lambda_n(t)$  or equivalently  $\lambda_{n[Nt]}$ , the limiting distribution of the frequencies and the normalized frequencies can be easily seen as the following:

- If  $\lambda_{n[Nt]} \rightarrow 0$ ,

$$\nu_{n[Nt]} \xrightarrow{p} 0$$

$$Y_{n[Nt]} \xrightarrow{p} 0$$

- if  $\lambda_{n[Nt]} \rightarrow \lambda(t)$ ,

$$\begin{aligned} \nu_{n[Nt]} &\xrightarrow{d} Z(t) \\ Y_{n[Nt]} &\xrightarrow{d} Y(t) = \frac{Z(t) - \lambda(t)}{\sqrt{\lambda(t)}} \end{aligned}$$

with  $Z(t) \sim Poi(\lambda(t))^1$ .

- If  $\lambda_{n[Nt]} \rightarrow \infty$ ,

$$\nu_{n[Nt]} \xrightarrow{p} \infty$$

$$Y_{n[Nt]} \xrightarrow{d} Y(t) = Y$$

with  $Y \sim \mathcal{N}(0, 1)$ .

These results imply that the asymptotic normality of the vector of the normalized frequencies has in general been broken in non-classical multinomial models.

---

<sup>1</sup>Note that we do not consider  $Z(t)$  and  $Y(t)$  as stochastic processes. We simply consider they are random variables with corresponding parameters.

## 3.2 Typical Examples of Non-classical Multinomial Models

### 3.2.1 GOF Test for Continuous Distributions

One important application of non-classical multinomial models is that of the goodness of fit test to continuous distributions. The motivation is that for a continuous hypothetical distribution, many GOF tests are not consistent against all alternatives. For example, if  $F^1$  is such that all  $\int_{A_i} dF^1 = \int_{A_i} dF^0$ , then the Chi-square test will have no power to test  $F^0$  against  $F^1$  (see Figure 3.1).

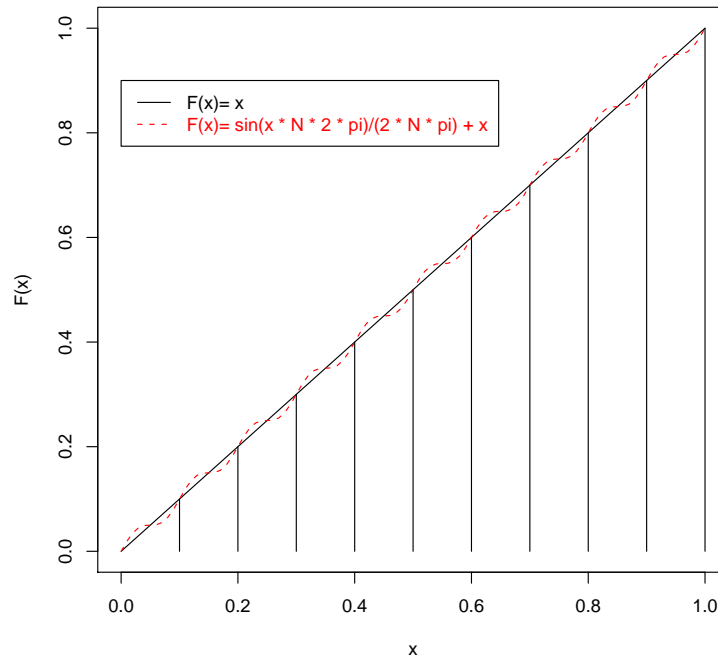


Figure 3.1: Loss of Power of GOF Test to Continuous Distribution:  $N=10$ ,  $A_i = [\frac{i-1}{N}, \frac{i}{N})$ ,  $F^0(x) = x$ ,  $F^1(x) = \frac{\sin(2*N*\pi*x)}{2*N*\pi} + x$ .

“Therefore, it is tempting to allow  $N$  to increase with  $n$  in order to obtain power against an even broader range of alternatives” (Lehmann and Romano, 2005). And this leads to non-classical multinomial models with  $p_{ni} = \int_{A_{ni}} dF^0$ .

The probabilities  $p_{ni}$  not only depend on the null hypothetic distribution  $F^0$ , but also rely on the sequence of partitions  $\mathbf{A}_n = \{A_{ni}\}_{1 \leq i \leq N}$ . Hence the limit of  $F_n$  depends on both the underlying distribution  $F^0$  and the sequence of partitions  $\mathbf{A}_n$ .

In particular, when  $F^0$  is concentrated on  $[0, 1]$ ,<sup>2</sup> the equipartitions with  $A_{ni} = [\frac{i-1}{N}, \frac{i}{N})$  are often employed. Under this setting,

$$F_n(t) = \sum_{i=1}^{Nt} p_{ni} \rightarrow F^0(t)$$

and

$$f_n(t) = Np_{n[Nt]} \rightarrow f^0(t)$$

if the density  $f^0$  of  $F^0$  exists.

It is worth noting that for general partitions the limits  $F = \lim F_n$  and  $f = \lim f_n$  are not necessarily  $F^0$  and  $f^0$  respectively.

Another important parameter which specifies the asymptotic behaviour of the non-classical model is  $d_n = n/N$ . Based on the different limit of  $d_n$ , the following typical non-classical multinomial models can be obtained.

---

<sup>2</sup>In cases the original distribution  $G^0$  is concentrated on a large set, say the whole real line, we can always use some fixed (absolutely continuous) distribution function  $\Phi$  to construct  $F^0(t) = G^0[\Phi^{-1}(t)]$ , such that  $F^0$  is concentrated on  $[0, 1]$ .



### Non-classical Model with Asymptotically Gaussian Frequencies

If  $d_n \rightarrow \infty$ , with the further assumption that  $\inf_t f(t) > \delta > 0$ , we get a non-classical model with all frequencies being asymptotically Gaussian with  $\inf_t np_{n[Nt]} \rightarrow \infty$ .

### Non-classical Model with Asymptotically Poissonian Frequencies

On the other hand, if  $d_n \rightarrow d \in (0, \infty)$  and  $\sup_t f(t) < \infty$ , all the frequencies are asymptotically Poissonian, with all  $np_{n[Nt]} \rightarrow d \cdot f(t) < \sup_t d \cdot f(t) < \infty$ .

### Non-classical Model with Equiprobable Frequencies

If we further choose partitions  $\mathbf{A}_n$  such that all  $p_{ni} = \int_{A_{ni}} dF^0 = 1/N$ , then this is a non-classical model with all frequencies equiprobable.

## 3.2.2 Statistical Analysis to Linguistic Data

Statistical linguistics is one of major application areas for the non-classical multinomial models, in which the word frequencies in corpora (large and structured set of texts) are the major objects (see, e.g. Baayen, 2001). One of the notable feature of linguistic data is that, apart from some words with high frequencies, most of the words are very rare in the sense that the frequencies are very small. The traditional, but improper treatment, was to group these words. However, these rare words are usually very important - they constitute the major proportion of the vocabulary. For example, in table 3.1, 9573 words out of 14715, which is over 65%, have frequencies less than 10 in a sample of over 1 million words.

This feature of some events being of high frequencies but most being very rare is the common feature of many statistical data in practice. An advanced formal approach to analyzing this sort of data is the LNRE theory which will be discussed in chapter 6.

No.	Word	Frequency	Frequency Range	Number of words
1	THE	70,002	>10000	11
2	BE	44,293	1000-10000	82
.....	.....	.....	100-1000	1152
721	SON	182	10-100	3897
725	FLOOR	182	1-10	9573
740	SIZE	179		
.....	.....	.....		
5,755	BOUNDS	8		
9,154	KINETIC	3		
9,512	OSLO	3		
11,084	QUARTZ	2		
14,715	MAYA	1		
Total		1,003,116		
Average		68		

Table 3.1: Word Frequencies of BNC BROWN Corpus.

### Non-classical Model with Mixed Frequencies

In table 3.1, many frequencies are very high and should not be treated as Poissonian frequencies. Hence we consider this as a typical example of non-classical model with both asymptotically Gaussian and Poissonian frequencies existing. It worth noting that since the frequencies on all the words observed in the corpus are at least 1, it is reasonable to consider  $\inf_i np_{ni} > \delta > 0$ , i.e., the expectation of the frequencies are bounded from below.

### 3.2.3 Diversity Analysis of Responses in Questionnaires

Another interesting example came from the study of the diversity of responses in questionnaires (Khmaladze, 2009).

Consider a questionnaire with  $q$  questions with binary (0-1) answers, which

is completed by  $n$  individuals. Obviously, there are  $N = 2^q$  possible ways in which to fill out such a questionnaire. Each possible way stands for an event in multinomial models and can be denoted by a  $q$ -dimensional vector  $\vec{x}_q = (x_1, \dots, x_i, \dots, x_q)$  with coordinates 0 or 1. Let  $a_i$  be the probability of the answer "1" to  $i$ -th question, then the probabilities  $p_{ni}$  become  $p(\vec{x}_q)$  with

$$p(\vec{x}_q) = \prod_{i=1}^q a_i^{x_i} (1 - a_i)^{1-x_i} , \quad (3.1)$$

in this context.

If the number  $n$  of people interviewed is much larger than the number of possible answers  $N$ , then it can be treated as a classical multinomial problem. However, as is often the case, if many case,  $n/N$  is not very large, the non-classical model is more suitable.

If most of the  $a_i$  deviate from  $1/2$ , then most of the  $np_{ni}$  tend to 0, and the asymptotic behaviour of divisible statistics on this model becomes very complicated. This was the main theme of Khmaladze (2009), and is highly relevant to the theory of LNRE.

### 3.3 Difficulties in Non-Classical Models

#### 3.3.1 Loss of Asymptotic Normality

Figure 3.2 shows the cumulative distribution function (CDF) of two divisible statistics,  $\sum_{i=1}^N |Y_{ni}|$  and  $\sum_{i=1}^N |Y_{ni}|^3$ , in a non-classical multinomial model with  $\mathbf{p} = \{1/N\}$ . The fact that the simulated CDF deviates from the limit CDF under the assumption that  $\{Y_{ni}\}$  is asymptotically a normal random vector, implies the loss of asymptotic normality of normalized frequencies in non-classical models.

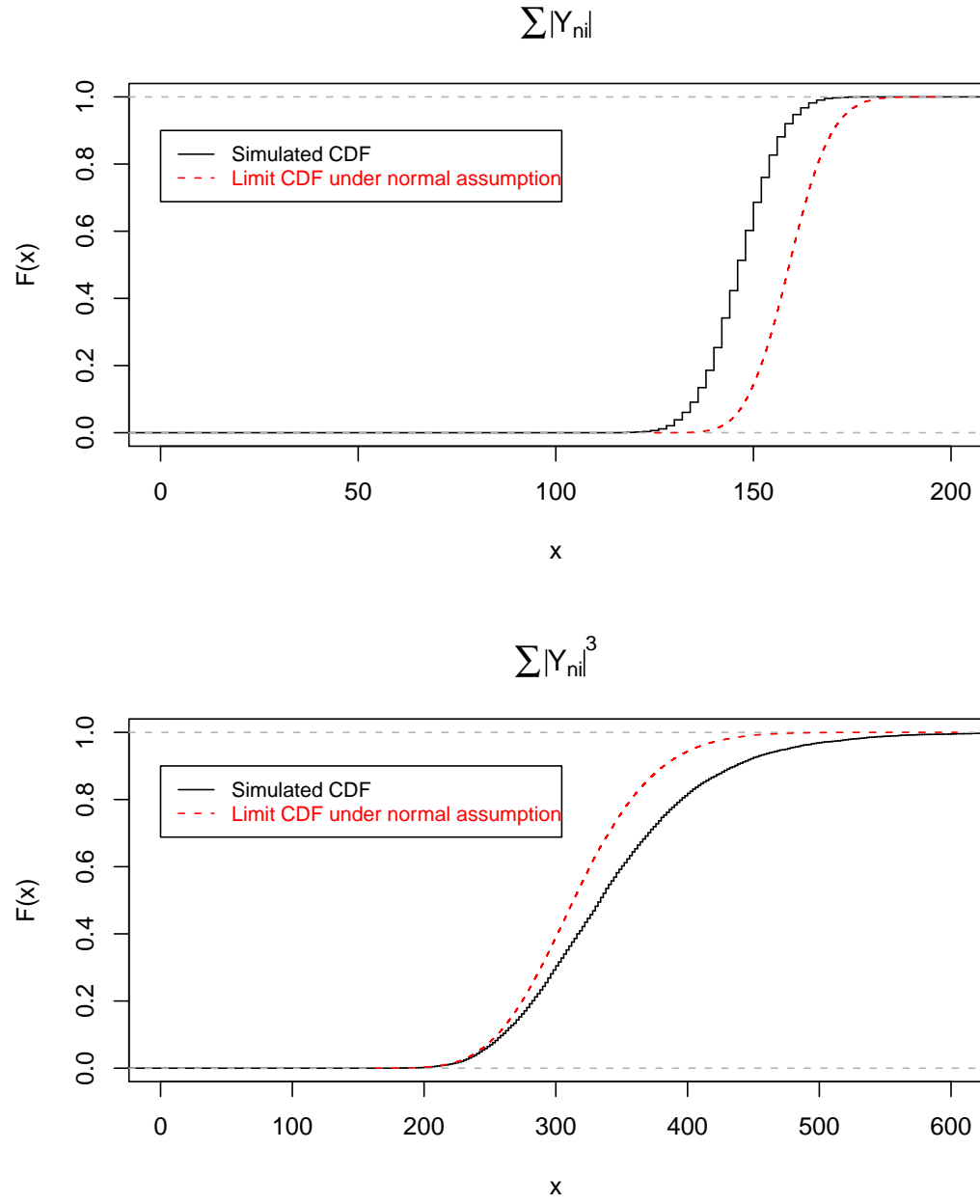


Figure 3.2: Loss of Asymptotic Normality of Normalized Frequencies:  $n = N = 200$  and  $\mathbf{p} = \{1/N\}$ .

### 3.3.2 Loss of Asymptotic Equivalence

Figure 3.3 shows the comparison of the CDF of the log-likelihood statistic and the Freeman-Tukey statistic to that of the Chi-square statistic. It shows that when  $n = 20N$  and the model is close to the classical multinomial model, the three CDFs are very close. But when  $n = N$ , for which the model is non-classical, the CDF of both the log-likelihood statistic and the Freeman-Tukey statistic deviate significantly from the CDF of the Chi-square statistic.

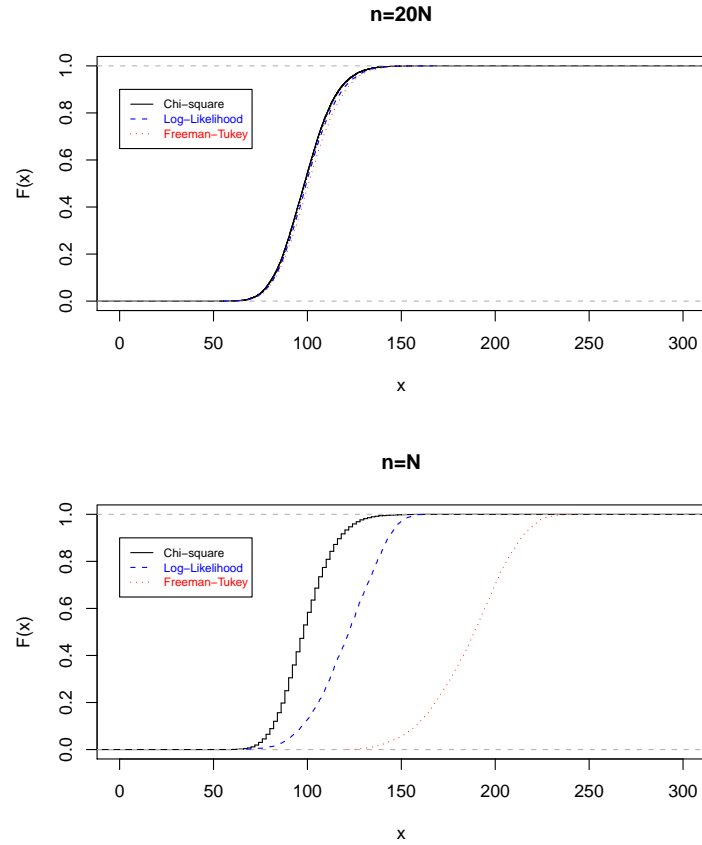


Figure 3.3: Loss of Asymptotic Equivalence to Chi-square Statistic:  $N = 100$  and  $\mathbf{p}=\{1/N\}$ .



## Chapter 4

# FLTs for Partial Sum Processes

As we have seen in previous sections, two reasons which make the analysis of non-classical asymptotics of divisible statistics more complicated are: the asymptotic normality of the normalized frequencies has been lost; and for those statistics which are asymptotically equivalent to Pearson's Chi-square statistics, this asymptotic equivalence is also lost. There is actually a third reason which we will show in this chapter: the dependence among the frequencies (or the summands of the divisible statistics) is not asymptotically negligible.

Some efforts have been dedicated to non-classical asymptotics (see e.g., Holst (1972); Morris (1975); Ivchenko and Levin (1976); Medvedev (1977, 1978); Ivchenko and Medvedev (1981); Györfi and Vajda (2002), etc.), but they are neither complete nor systematic. Most of them focused on the divisible statistic as a whole. However, since the divisible statistics are sums of a sequence of random variables, it is natural for us to consider the partial sum process

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (g_{ni}(\nu_{ni}, np_{ni}) - \mathbb{E}[g_{ni}(\nu_{ni}, np_{ni})]), \quad t \in [0, 1] \quad (4.1)$$

as the main object. With the knowledge of martingale theory, we can decompose this partial sum process  $X_n$  into a martingale component  $W_n$  and

a compensator component  $K_n$ .

One advantage of this approach is that the limit distributions of these two components have a convenient form - both of them are Gaussian processes. Another advantage is that both components have sensible meanings. Since the distribution of the limiting process of the martingale component is exactly the same as that of partial sum process as if all summands were mutually independent, the compensator component can be considered as the one which reflects the “dependence” of the summands, while the martingale component account to the “size” of the summands.

It was in Khmaladze (1984) that the first successful treatment had been proposed, where Khmaladze introduced a special filtration  $\{\mathcal{F}_i^{(n)}\}_{0 \leq i \leq N}$  with  $\sigma$ -algebras  $\mathcal{F}_i^{(n)} = \sigma\{\nu_{nk} : k \leq i\}$ , generated by the frequencies  $\nu_{ni}$  instead of that generated by the summands  $g_{ni}(\nu_{ni}, np_{ni})$ . This construction makes it possible to derive the limiting distributions of the partial sum process and their components.

Although the approach of Khmaladze (1984) was successful, it covers only the models where  $n \sim N$  and all  $np_{ni} = \mathbb{E}\nu_{ni}$  are bounded, which implies that all the frequencies are asymptotically Poissonian. In many practical situations, such as the linguistic data we showed in Table 3.1, it is necessary to consider models with a mixture of asymptotically Gaussian and asymptotically Poissonian frequencies. In many other non-classical models with all frequencies being asymptotically Gaussian, the case of  $n \gg N$  needs to be considered as well.

In this chapter, we will describe in detail the martingale approach to the FLTs of the partial sum processes of divisible statistics and show that the martingale approach can be extended to the non-classical models with a mixture of asymptotically Gaussian and asymptotically Poissonian frequencies.



The FLTs for this extended class of models will be established, and a comparison is made between this class and the one with all Poissonian frequencies.

## 4.1 Preliminaries

### 4.1.1 Probability spaces and martingales

The analysis in this chapter relies heavily on martingale theory. Therefore, it is necessary for us to review some concepts in martingale theory. The approach here basically follows from Williams (1991); Shiryaev (1995); Brzezniak and Zastawniak (2000).

**Definition 4.1.** *Let  $\Omega$  be the set of all possible outcomes. Then  $\Omega$  is the sample space.*

**Definition 4.2.** *A family  $\mathcal{F}$  of subsets of a space  $\Omega$  is called a  $\sigma$ -algebra if it contains  $\Omega$  itself and is closed under complements and countable unions, i.e.,*

1.  $\Omega \in \mathcal{F}$
2.  $A \in \mathcal{F}$  implies  $A^c \in \mathcal{F}$
3.  $A_1, A_2, \dots \in \mathcal{F}$  implies  $A_1 \cup A_2 \cup \dots \in \mathcal{F}$ .

The pair  $(\Omega, \mathcal{F})$  is called a *measurable space*.

**Definition 4.3.** *Let  $\mathcal{A}$  be a collection of subsets of  $\Omega$ . The  $\sigma$ -algebra generated by  $\mathcal{A}$ , denoted by  $\sigma(\mathcal{A})$ , is the smallest  $\sigma$ -algebra of  $\Omega$  which contains  $\mathcal{A}$ .*

**Definition 4.4.** *The Borel  $\sigma$ -algebra of  $\mathbb{R}$ , denoted by  $\mathcal{B}(\mathbb{R})$ , is the  $\sigma$ -algebra generated by all open subsets of  $\mathbb{R}$ .*

**Definition 4.5.** A real-valued function  $\mathbf{P}$  defined on a  $\sigma$ -algebra  $\mathcal{F}$  is a probability measure if it satisfies the following conditions:

1.  $0 \leq \mathbf{P}(A) \leq 1$  for  $A \in \mathcal{F}$ .
2.  $\mathbf{P}(\emptyset) = 0, \mathbf{P}(\Omega) = 1$ .
3. If  $A_i \in \mathcal{F}$  is a countable union of disjoint sets, then

$$\mathbf{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbf{P}(A_i).$$

The triple  $(\Omega, \mathcal{F}, \mathbf{P})$  is called a *probability space*.

**Definition 4.6.** If  $\mathcal{F}$  is a  $\sigma$ -algebra of  $\Omega$ , then a function  $X : \Omega \rightarrow \mathbb{R}$  is said to be  $\mathcal{F}$ -measurable if

$$\{X \in B\} = \{\omega : X(\omega) \in B\} \in \mathcal{F} \text{ for all } B \in \mathcal{B}(\mathbb{R})$$

If we consider a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ , then  $X : \Omega \rightarrow \mathbb{R}$  is a random variable.

**Definition 4.7.** The  $\sigma$ -algebra  $\sigma(X)$  generated by a random variable  $X$  consists of all  $\{X \in B\}$  with  $B \in \mathcal{B}(\mathbb{R})$ .

**Definition 4.8.** The  $\sigma$ -algebra  $\sigma(X_i : i \in I)$  generated by a class of random variables  $\{X_i : i \in I\}$  is defined to be the smallest  $\sigma$ -algebra containing all events  $\{X_i \in B\}$  with  $B \in \mathcal{B}(\mathbb{R})$ .

**Definition 4.9.** A filtration is a sequence  $\{\mathcal{F}_i\}_{i \geq 0}$  of  $\sigma$ -algebras on  $\Omega$  with

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$$

**Definition 4.10.** A sequence  $\{X_i\}_{i \geq 0}$  of random variables is adapted to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$  if  $X_i$  is  $\mathcal{F}_i$ -measurable for all  $i = 0, 1, 2, \dots$

For a sequence  $\{X_i\}_{i \geq 0}$  of random variables, let  $\mathcal{F}_i$  be the  $\sigma$ -algebra generated by  $X_0, X_1, \dots, X_i$ , i.e.,

$$\mathcal{F}_i = \sigma(X_0, X_1, \dots, X_i);$$

then  $\{\mathcal{F}_i\}_{i \geq 0}$  forms a filtration and  $\{X_i\}_{i \geq 0}$  is adapted to  $\mathcal{F}$ .

**Definition 4.11.** A sequence  $\{X_i\}_{i \geq 0}$  of random variables is a martingale with respect to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$  if

- $X_0, X_1, \dots$  is adapted to  $\{\mathcal{F}_i\}_{i \geq 0}$ ,
- $\mathbb{E}|X_i| < \infty \forall i$ , and
- $\mathbb{E}(X_{i+1}|\mathcal{F}_i) = X_i$  a.s. for all  $i = 0, 1, \dots$

$\{X_i\}_{i \geq 0}$  is a submartingale if

- $\mathbb{E}(X_{i+1}|\mathcal{F}_i) \geq X_i$  a.s. for all  $i = 0, 1, \dots$

**Definition 4.12.** A sequence  $\{\xi_i\}_{i \geq 0}$  of random variables is a martingale-difference with respect to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$  if  $\{\xi_i\}_{i \geq 0}$  is adapted to  $\{\mathcal{F}_i\}_{i \geq 0}$ ,  $\mathbb{E}|\xi_i| < \infty \forall i$  and

$$\mathbb{E}[\xi_{i+1}|\mathcal{F}_i] = 0 \quad \text{a.s.}$$

For a martingale  $\{X_i\}_{i \geq 0}$ , then the sequence  $\{\xi_i\}_{i \geq 0}$  with  $\xi_0 = X_0$  and  $\xi_i = \Delta X_i = X_i - X_{i-1}, i \geq 1$ , is a martingale-difference. On the other hand, for a martingale-difference  $\{\xi_i\}_{i \geq 0}$ , the sequence  $\{X_i\}_{i \geq 0}$  with  $X_i = \sum_{j=0}^i \xi_j$  is a martingale.

**Lemma 4.1.** If  $\{X_i\}_{i \geq 0}$  is a martingale with respect to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$ , and  $g(x)$  is convex with  $\mathbb{E}|g(X_i)| < \infty, \forall i$ , then  $\{g(X_i)\}_{i \geq 0}$  is a submartingale.

**Definition 4.13.** A sequence  $\{X_i\}_{i \geq 0}$  of random variables is a predictable sequence with respect to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$  if for all  $i$ ,  $X_i$  is  $\mathcal{F}_{i-1}$ -measurable and  $\mathcal{F}_{-1} = \mathcal{F}_0$ .

**Theorem 4.1. (Doob)** If  $\{X_i\}_{i \geq 0}$  is a submartingale with respect to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$ , then there is a martingale  $\{m_i\}_{i \geq 0}$  and a predictable increasing sequence  $\{A_i\}_{i \geq 0}$  (compensator) with respect to  $\{\mathcal{F}_i\}_{i \geq 0}$ , such that,

$$X_i = m_i + A_i \quad \text{a.s.},$$

and the decomposition of this kind is unique. Further

$$m_i = m_0 + \sum_{j=1}^i (X_j - \mathbb{E}[X_j | \mathcal{F}_{j-1}]), \quad (4.2)$$

and

$$A_i = \sum_{j=1}^i (\mathbb{E}[X_j | \mathcal{F}_{j-1}] - X_{j-1}) \quad (4.3)$$

with  $m_0 = X_0$  and  $A_0 = 0$ .

*Proof.* See Theorem 2, p.482, Shiryaev (1995). □

**Theorem 4.2.** *If  $\{X_i\}_{i \geq 0}$  is adapted to a filtration  $\{\mathcal{F}_i\}_{i \geq 0}$  and all  $X_i$ 's are integrable, i.e., all  $\mathbb{E}|X_i| < \infty$ , then the decomposition similar to Theorem 4.1 can be established, except that  $\{A_i\}_{i \geq 0}$  is not necessarily increasing. Such a decomposition is called a semi-martingale decomposition and  $\{X_i\}_{i \geq 0}$  is a semi-martingale with respect to  $\{\mathcal{F}_i\}_{i \geq 0}$ .*

*Proof.* See p.121, Williams (1991). □

Let  $M = \{M_i\}_{i \geq 0}$  be a square integrable martingale with respect to  $\{\mathcal{F}_i\}_{i \geq 0}$  with  $\mathbb{E}M_i^2 < \infty \quad \forall i$ . By lemma 4.1,  $M^2 = \{M_i^2\}_{i \geq 0}$  is a submartingale. By theorem 4.1,  $M^2$  can be decomposed into a martingale and a compensator  $\langle M \rangle = \{\langle M \rangle_i\}_{i \geq 0}$  with

$$\langle M \rangle_i = \sum_{j=1}^i (\mathbb{E}[M_j^2 | \mathcal{F}_{j-1}] - M_{j-1}^2) = \sum_{j=1}^i \mathbb{E}[(\Delta M_j)^2 | \mathcal{F}_{j-1}]$$

**Definition 4.14.** *The sequence  $\langle M \rangle$  is called the predictable quadratic variation or the quadratic characteristic of  $M$ .*

**Remark 4.1.** *Since  $\{\Delta M_i\}_{i \geq 0}$  is a martingale-difference,  $\mathbb{E}[\Delta M_j | \mathcal{F}_{j-1}] = 0$  and  $\langle M \rangle_i$  is the sum of the conditional variances of  $\{\Delta M_j\}_{0 \leq j \leq i}$ .*

In the theory of martingales, another important concept which is often connected to quadratic characteristic is *quadratic variation*.

**Definition 4.15.** *The quadratic variation of a process  $X$  is  $[X] = \{[X]_i\}_{i \geq 0}$  with*

$$[X]_i = \sum_{j=0}^i (\Delta X_j)^2.$$

The quadratic characteristic is actually the compensator part of the quadratic variation.

### 4.1.2 Triangular Array and Lindeberg Theorems

In many situations, a sum of many small random variables may be approximately normally distributed and hence a sequence of sums of an increasing number of small random variables may converge in distribution to a normal random variable.

**Definition 4.16.** *A double sequence  $\{\xi_{ni}\}_{n \geq 1, 1 \leq i \leq n}$  of random variables, i.e.,*

$$\begin{array}{c} \xi_{11} \\ \xi_{21}, \xi_{22} \\ \dots\dots\dots \\ \xi_{n1}, \xi_{n2}, \dots\dots\dots, \xi_{nn} \end{array}$$

*is called a triangular array.*

Often we are interested in the row sums of the triangular array  $S_n = \sum_{i=1}^n \xi_{ni}$ . In the classical central limit theorem, if for any  $n$ ,  $\{\xi_{ni}\}_{1 \leq i \leq n}$  is a sequence of i.i.d. random variables with zero mean and variance  $\sigma^2/n$ , then the sum

$$\sum_{i=1}^n \xi_{ni} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

converges in distribution to a normal random variable.

However, the “identical” condition is actually not a necessary condition for normality.

**Theorem 4.3** (Lindeberg’s Theorem). *Suppose that the elements of each row of a triangular array are mutually independent, zero-mean random variables and the condition*

$$\sum_{i=1}^n \mathbb{E} \xi_{ni}^2 \rightarrow \sigma^2$$

*is satisfied. Then Lindeberg’s condition:*

$$\sum_{i=1}^n \mathbb{E} [\xi_{ni}^2 \mathbf{I}\{\xi_{ni}^2 > \epsilon\}] \rightarrow 0 \quad \forall \epsilon > 0, \quad \text{as } n \rightarrow \infty \quad (L)$$

*implies  $S_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ .*

*Proof.* See Theorem 2, p334, in Shiryaev (1995); or Feller (1970).  $\square$

### 4.1.3 Martingale Central Limit Theorem

A further step to generalize the central limit theorem is to consider the triangular array with dependent row elements. In particular, it is of great interest to consider the elements of each row forming a martingale difference with respect to a sequence of filtrations.

**Definition 4.17.** *In a triangular array, if for each  $n$ ,  $\{\xi_{ni}\}_{1 \leq i \leq n}$  is a martingale difference with respect to the filtration  $\mathcal{F}^{(n)} = \{\mathcal{F}_i^{(n)}\}_{1 \leq i \leq n}$ , then this triangular array is called a martingale difference array with respect to  $\mathcal{F}^{(n)}$ .*

It is easy to see that the partial sum of each row

$$X_{ni} = \sum_{j=1}^i \xi_{nj}$$

is a martingale with respect to  $\mathcal{F}^{(n)}$ . Under certain conditions the FLT for the partial sum process

$$X_n(t) = \sum_{i=1}^{[nt]} \xi_{ni}$$

can be established. This is formally called *the martingale central limit theorem*, which was established in the late 1970s (Shiryaev, 1995, Durrett, 2004 etc.).

**Theorem 4.4.** *Consider a martingale difference array  $\{\xi_{ni}\}_{n \geq 1, 1 \leq i \leq n}$  with respect to  $\mathcal{F}^{(n)} = \{\mathcal{F}_i^{(n)}\}_{1 \leq i \leq n}$ , for which all elements  $\xi_{ni}$  are square-integrable. If the quadratic characteristic converges in probability, i.e.,*

$$\sum_{i=1}^{[nt]} \mathbb{E} [\xi_{ni}^2 | \mathcal{F}_{i-1}] \xrightarrow{p} \sigma^2(t)$$

*or the quadratic variation converges in probability, i.e.,*

$$\sum_{i=1}^{[nt]} \xi_{ni}^2 \xrightarrow{p} \sigma^2(t);$$

*and the conditional Lindeberg condition*

$$\sum_{i=1}^{[nt]} \mathbb{E} [\xi_{ni}^2 \mathbf{I}\{\xi_{ni}^2 > \epsilon\} | \mathcal{F}_{i-1}] \xrightarrow{p} 0 \quad \forall \epsilon > 0 \quad (CL)$$

*is satisfied, then the partial sum process*

$$X_n \xrightarrow{d} W,$$

*where  $W$  is a Brownian motion with respect to time  $\sigma^2(t)$ .*

*Proof.* See Liptser and Shiryaev (1981). □

## 4.2 FLT for Models with Poissonian Frequencies

As we have seen in equation (4.1), the object of interest here is the partial sum process,

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (g_{ni}(\nu_{ni}, np_{ni}) - \mathbb{E}[g_{ni}(\nu_{ni}, np_{ni})]).$$

The process itself is hard to analyze directly due to the dependence of summands. However, based on a semi-martingale decomposition, with respect to a suitable filtration, the partial sum process can be decomposed into a martingale component

$$W_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( g_{ni}(\nu_{ni}, np_{ni}) - \mathbb{E} \left[ g_{ni}(\nu_{ni}, np_{ni}) | \mathcal{F}_{i-1}^{(n)} \right] \right),$$

and a compensator component

$$K_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \mathbb{E} \left[ g_{ni}(\nu_{ni}, np_{ni}) | \mathcal{F}_{i-1}^{(n)} \right] - \mathbb{E} [g_{ni}(\nu_{ni}, np_{ni})] \right).$$

The filtration used for this purpose could be the natural filtration  $\{\mathcal{F}_i^{(n)}\}_{0 \leq i \leq N}$ , with  $\mathcal{F}_i^{(n)} = \sigma\{g_{nk}(\nu_{nk}, np_{nk}) : k \leq i\}$ , but it is not fine enough to establish limit theorems. To cope with this deficiency, a special filtration with  $\mathcal{F}_i^{(n)} = \sigma\{\nu_{nk} : k \leq i\}$  was introduced by Khmaladze (1984) and we shall see later that this finer filtration is sufficient to investigate the asymptotic behaviour of the partial sum processes.

As  $N \rightarrow \infty$ , the number of the summands increases to infinity. It is easy to see that the effect of an increasing number of summands can be “offset” by the normalizing coefficient  $1/\sqrt{N}$ . Therefore, in discussion of the asymptotic behaviour of the process, the focus turned to the asymptotic properties of  $g_{ni}(\nu_{ni}, np_{ni})$ , which is controlled by the frequencies  $\nu_{ni}$  and their expectation  $np_{ni}$ .

In those models where  $np_{ni}$  satisfies

$$\limsup_n \sup_i (np_{ni}) = \Lambda < \infty \quad (\text{C1})$$

such that all  $\nu_{ni}$  are asymptotically Poissonian frequencies, the summands of the partial sum processes are uniformly square-integrable, for a large class of divisible statistics with  $g_{ni}$  satisfying

$$|g_{ni}(x, \lambda)| < be^{ax}, \quad |g_t(x, \lambda)| < be^{ax} \quad (\text{C2})$$



and

$$g_{n[Nt]}(x, \lambda) \rightarrow g_t(x, \lambda) \quad \text{as} \quad n \rightarrow \infty$$

for some positive  $a$  and  $b$  and for all  $x$ ,  $\lambda$  and  $t$ . According to Khmaladze (1984), the FLTs of the partial sum processes and its components can be established for this scheme. The proofs were technical and here we only illustrate them conceptually.

### 4.2.1 FLT for Martingale Component

Under Conditions C1 and C2, it can be shown that the conditional Lindeberg condition  $CL$  is satisfied. At the same time, under the additional assumption

$$f(t) = \lim_{n \rightarrow \infty} Np_{n[Nt]} \quad \text{for all } t \in [0, 1] \quad (C3)$$

the quadratic characteristic of  $W_n$  converges in probability to some deterministic function  $\tau(t)$ . The martingale limit theorem then implies that the martingale component converges in distribution to a Brownian motion with respect to time  $\tau(t)$ .

The original proof of Khmaladze was based on  $n = N$ , but it can be easily generalized to  $n = d \times N$  for fixed  $d$  with  $0 < d < \infty$ . Under this setting,

$$\lambda(t) = \lim_{n \rightarrow \infty} np_{n[Nt]} = d \times f(t)$$

and

$$\nu_{n[Nt]} \xrightarrow{d} Z(t) \sim Poi(\lambda(t)).$$

And the theorem can be formulated as following.

**Theorem 4.5** (Khmaladze 1984). *Assume the Conditions C1-C3 are satisfied. Let  $w$  denote a standard Brownian motion on  $\mathbb{R}^+$ ; and let  $W(t) = w(\tau(t))$  with*

$$\tau(t) = \int_0^t \sigma^2(s) ds, \quad (4.4)$$

and  $\sigma^2(s)$  being the variance of  $g_s(Z(s), \lambda(s))$ . Then

$$W_n \xrightarrow{d} W.$$

Note that if the summands of the partial sum process  $X_n$  were treated as being mutually independent, the distribution of the limiting process would be the same as  $W$ .

### 4.2.2 FLT for Compensator

In order to investigate the limiting behaviour of the compensator component, one must start from the properties of the frequencies  $\nu_{ni}$ .

Marginally, the frequency  $\nu_{ni}$  is binomially distributed with parameters  $n$  and  $p_{ni}$ . Under condition C1, for  $n$  large enough,  $\nu_{ni}$  is approximately Poisson distributed with parameter  $\lambda_{ni} = np_{ni}$ , and hence

$$\mathbb{E}g_{ni}(\nu_{ni}, np_{ni}) \approx \mathbb{E}g_{ni}(Z_{ni}, \lambda_{ni}) = \sum_{k=0}^{\infty} g_{ni}(k, \lambda_{ni}) \pi(k, \lambda_{ni})$$

with  $Z_{ni} \sim Poi(\lambda_{ni})$ .

Conditional on  $\mathcal{F}_{i-1}^{(n)}$ , the frequency  $\nu_{ni}$  is again binomially distributed but with different parameters

$$\tilde{n}_{ni} = n - \sum_{j=1}^{i-1} \nu_{nj} \quad \text{and} \quad \tilde{p}_{ni} = p_{ni} / (1 - \sum_{j=1}^{i-1} p_{nj}).$$

If for any  $\epsilon > 0$  and  $T_n = 1 - \epsilon/\sqrt{n}$ ,

$$\lim_{n \rightarrow \infty} \sqrt{n} [1 - F(T_n)] \rightarrow \delta(\epsilon) > 0, \tag{C4}$$

then for any  $i \leq NT$ , as  $n \rightarrow \infty$

$$\tilde{n}_{ni} \tilde{p}_{ni} \sim np_{ni}$$

and  $\tilde{n}_{ni}$  is large enough so that the conditional distribution of  $\nu_{ni}$  under  $\mathcal{F}_{i-1}^{(n)}$  is approximately Poisson with parameter  $\tilde{\lambda}_{ni} = \tilde{n}_{ni}\tilde{p}_{ni}$  and

$$\mathbb{E} \left[ g_{ni}(\nu_{ni}, np_{ni}) | \mathcal{F}_{i-1}^{(n)} \right] \approx \mathbb{E} g_{ni}(\tilde{Z}_{ni}, \lambda_{ni}) = \sum_{k=0}^{\infty} g_{ni}(k, \lambda_{ni}) \pi(k, \tilde{\lambda}_{ni})$$

with  $\tilde{Z}_{ni} \sim Poi(\tilde{\lambda}_{ni})$ .

Based on Lemma 2 of Khmaladze (1984), it can be shown that for  $t < T_n$ , the compensator component  $K_n(t)$  can be approximated by

$$\begin{aligned} K_n^\pi(t) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \mathbb{E} g_{ni}(\tilde{Z}_{ni}, \lambda_{ni}) - \mathbb{E} g_{ni}(Z_{ni}, \lambda_{ni}) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \sum_{k=0}^{\infty} g_{ni}(k, \lambda_{ni}) \left( \pi(k, \tilde{\lambda}_{ni}) - \pi(k, \lambda_{ni}) \right). \end{aligned} \quad (4.5)$$

Then since both  $\tilde{\lambda}_{n[Nt]}$  and  $\lambda_{n[Nt]}$  converge to  $\lambda(t)$  as  $n \rightarrow \infty$ , Taylor's expansion shows that

$$\begin{aligned} \left( \pi(k, \tilde{\lambda}_{ni}) - \pi(k, \lambda_{ni}) \right) &\approx \frac{\partial \pi(k, \lambda)}{\partial \lambda} \Big|_{\lambda=\lambda_{ni}} \cdot (\tilde{\lambda}_{ni} - \lambda_{ni}) \\ &= \frac{k - \lambda_{ni}}{\lambda_{ni}} \pi(k, \lambda_{ni}) (\tilde{\lambda}_{ni} - \lambda_{ni}). \end{aligned}$$

If we let  $F_n(t) = \sum_{i=1}^{Nt} p_{ni}$  be the distribution function defined by  $\{p_{ni}\}$ ,  $\hat{F}_n(t) = \frac{1}{N} \sum_{i=1}^{Nt} \nu_{ni}$  be the corresponding empirical distribution function, and  $v_n(t) = \sqrt{n}(\hat{F}_n(t) - F_n(t))$ , then

$$\tilde{\lambda}_{ni} - \lambda_{ni} = \lambda_{ni} \frac{F_n\left(\frac{i-1}{N}\right) - \hat{F}_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)} = -\frac{\lambda_{ni}}{\sqrt{n}} \frac{v_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)},$$

and hence the right hand side of equation (4.5) can be approximated by

$$\frac{1}{\sqrt{nN}} \sum_{i=1}^{Nt} \mathbb{E} [g_{ni}(Z_{ni}, \lambda_{ni}) (Z_{ni} - \lambda_{ni})] \left( \frac{-v_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)} \right). \quad (4.6)$$

Since

$$\frac{v_n(t)}{1 - F_n(t)} \xrightarrow{d} \frac{v(t)}{1 - F(t)},$$

where  $v$  is a Brownian bridge with respect to time  $F = \lim_{n \rightarrow \infty} F_n$ , (4.6) can be approximated by

$$K(t) = -\frac{1}{\sqrt{d}} \int_0^t \mathbb{E}[g_s(Z(s), \lambda(s))(Z(s) - \lambda(s))] \frac{v(s)}{1 - F(s)} ds.$$

Finally, the limit theorem for  $K_n(t)$  can be formulated as follows.

**Theorem 4.6** (Khmaladze 1984). *Define the operator*

$$\mathcal{K}x(t) = -\int_0^t k(s)x(s)ds$$

with

$$k(s) = \frac{1}{\sqrt{d}} \mathbb{E}[g_s(Z(s), \lambda(s))(Z(s) - \lambda(s))].$$

If Conditions C1-C4 are satisfied,

$$K_n \xrightarrow{d} K = \mathcal{K} \frac{v}{1 - F}.$$

### 4.3 FLT in Models with Mixed Frequencies

The proofs of FLTs in non-classical models with all asymptotically Poissonian frequencies essentially made use of the fact that both conditional and unconditional distribution of the frequencies are nearly Poissonian and they converge to the same limiting distribution if the expectation  $np_{ni}$  of the frequencies are bounded. However, this fact is not valid any more as soon as

some of the  $np_{ni}$  diverge to infinity.

In non-classical models with mixed frequencies, where some frequencies are asymptotically Gaussian, it is more appropriate to focus on the asymptotic behaviour of the the normalized frequencies  $Y_{ni}$ , and express the divisible statistics by an equivalent expression

$$D_n = \sum_{i=1}^N h_{ni}(Y_{ni}, np_{ni}).$$

Correspondingly, the partial sum process becomes

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (h_{ni}(Y_{ni}, np_{ni}) - \mathbb{E}[h_{ni}(Y_{ni}, np_{ni})]).$$

Consequently, the martingale component is

$$W_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( h_{ni}(Y_{ni}, np_{ni}) - \mathbb{E} \left[ h_{ni}(Y_{ni}, np_{ni}) \mid \mathcal{F}_{i-1}^{(n)} \right] \right),$$

and the compensator component becomes,

$$K_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \mathbb{E} \left[ h_{ni}(Y_{ni}, np_{ni}) \mid \mathcal{F}_{i-1}^{(n)} \right] - \mathbb{E}[h_{ni}(Y_{ni}, np_{ni})] \right).$$

Since some  $np_{ni}$  tend to infinity, the frequency  $\nu_{ni}$  is not bounded in probability, and the summands of the partial sum processes are not necessarily uniformly square-integrable under condition C2. However, if we replace the Conditions C1,C2 by

$$\liminf_n \inf_i (np_{ni}) = \delta > 0, \quad (\text{C1}^*)$$

and

$$|h_{ni}(y, \lambda)| < be^{a|y|}, \quad |h_t(y, \lambda)| < be^{a|y|}, \quad (\text{C2}^*)$$

$$h_{n[Nt]}(y, \lambda) \rightarrow h_t(y, \lambda) \quad \text{as} \quad n \rightarrow \infty$$

for some positive  $a$  and  $b$  and for all  $y$ ,  $\lambda$  and  $t$ , we can again establish FLT's. The class describe in C2\* is large enough to include most widely used divisible statistics.

For the theorems to be established, it is necessary to assume further that

$$f(t) = \lim_{n \rightarrow \infty} Np_{n[Nt]} \quad \text{and} \quad \lambda(t) = \lim_{n \rightarrow \infty} np_{n[Nt]} \quad (\text{C3}^*)$$

exist for  $t \in [0, 1]$ .

### 4.3.1 Auxiliary Lemmas

Before we formally establish the theorems, we shall review some properties of the frequencies  $\{\nu_{ni}\}$  and the normalized frequencies  $\{Y_{ni}\}$ , and establish some auxiliary lemmas.

Recall that for  $\lambda(t) < \infty$

$$Y_{n[Nt]} \xrightarrow{d} Y(t) = \frac{Z(t) - \lambda(t)}{\sqrt{\lambda(t)}} \quad \text{with} \quad Z(t) \sim \text{Poi}(\lambda(t)) \quad (4.7)$$

and for  $\lambda(t) = \infty$  and  $p_{n[Nt]} \rightarrow 0$ ,

$$Y_{n[Nt]} \xrightarrow{d} Y(t) \sim \mathcal{N}(0, 1). \quad (4.8)$$

The following lemma concerns the tail probability of the normalized frequencies.

**Lemma 4.2** (Exponential Inequality). *Consider a binomial random variable  $\nu$  with parameter  $n$  and  $p$ . The tail probabilities of the normalized binomial random variable*

$$Y = \frac{\nu - np}{\sqrt{np}}$$

satisfy

$$\mathbf{P}(Y > y) \leq \exp \left( -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np}} \right) \right) \quad \text{for } y > 0, \quad (4.9)$$

and

$$\mathbf{P}(Y < y) \leq \exp \left( -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np}} \right) \right) \quad \text{for } -\sqrt{np} \leq y \leq 0, \quad (4.10)$$

with  $\psi(\lambda) = (2/\lambda^2)[(1+\lambda)\ln(1+\lambda) - \lambda] = (2/\lambda^2) \int_0^\lambda \ln(1+x)dx$ .

*Proof.* For  $y > 0$  and  $t > 0$ ,

$$\begin{aligned} \mathbf{P}(Y > y) &= \mathbf{P}(e^{tY-ty} > 1) \leq \inf_{t>0} Ee^{tY-ty} \\ &= \inf_{t>0} e^{-ty-t\sqrt{np}} Ee^{\frac{t}{\sqrt{np}}\nu} = \inf_{t>0} e^{-ty-t\sqrt{np}} \left( 1 - p + pe^{\frac{t}{\sqrt{np}}} \right)^n \\ &= \inf_{t>0} \exp \left( n \ln \left( 1 + p(e^{\frac{t}{\sqrt{np}}} - 1) \right) - ty - t\sqrt{np} \right) \\ &\leq \inf_{t>0} \exp \left( np(e^{\frac{t}{\sqrt{np}}} - 1) - ty - t\sqrt{np} \right). \end{aligned}$$

Letting  $f(t) = np(e^{\frac{t}{\sqrt{np}}} - 1) - ty - t\sqrt{np}$ . Then

$$f'(t^*) = \sqrt{np} e^{\frac{t^*}{\sqrt{np}}} - y - \sqrt{np} = 0$$

implies

$$\exp \left( \frac{t^*}{\sqrt{np}} \right) = 1 + \frac{y}{\sqrt{np}}$$

and

$$t^* = \sqrt{np} \ln \left( 1 + \frac{y}{\sqrt{np}} \right).$$

Since

$$f''(t^*) = \exp \left( \frac{t^*}{\sqrt{np}} \right) = 1 + \frac{y}{\sqrt{np}} \geq 0,$$

we have

$$\inf_{t>0} f(t) = f(t^*) = np \left[ \frac{y}{\sqrt{np}} - \left( 1 + \frac{y}{\sqrt{np}} \right) \ln \left( 1 + \frac{y}{\sqrt{np}} \right) \right] = -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np}} \right)$$

and hence,

$$\mathbf{P}(Y > y) \leq \exp \left( -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np}} \right) \right).$$

On the other hand, for  $-\sqrt{np} \leq y \leq 0$  and  $t > 0$ ,

$$\begin{aligned} \mathbf{P}(Y < y) &= \mathbf{P}(e^{-tY+ty} > 1) \leq \inf_{t>0} E e^{-tY+ty} \\ &= \inf_{t>0} e^{ty+t\sqrt{np}} E e^{-\frac{t}{\sqrt{np}}\nu} = \inf_{t>0} e^{ty+t\sqrt{np}} \left( 1 - p + p e^{-\frac{t}{\sqrt{np}}} \right)^n \\ &= \inf_{t>0} \exp \left( n \ln \left( 1 + p(e^{-\frac{t}{\sqrt{np}}} - 1) \right) + ty + t\sqrt{np} \right) \\ &\leq \inf_{t>0} \exp \left( np(e^{-\frac{t}{\sqrt{np}}} - 1) + ty + t\sqrt{np} \right) \end{aligned}$$

let  $f(t) = np(e^{-\frac{t}{\sqrt{np}}} - 1) + ty + t\sqrt{np}$ , then

$$f'(t^*) = -\sqrt{np} e^{-\frac{t^*}{\sqrt{np}}} + y + \sqrt{np} = 0$$

implies

$$\exp \left( -\frac{t^*}{\sqrt{np}} \right) = 1 + \frac{y}{\sqrt{np}}$$

and

$$t^* = -\sqrt{np} \ln \left( 1 + \frac{y}{\sqrt{np}} \right).$$

Since

$$f''(t^*) = \exp \left( -\frac{t^*}{\sqrt{np}} \right) = 1 + \frac{y}{\sqrt{np}} \geq 0,$$

we have

$$\inf_{t>0} f(t) = f(t^*) = np \left[ \frac{y}{\sqrt{np}} - \left( 1 + \frac{y}{\sqrt{np}} \right) \ln \left( 1 + \frac{y}{\sqrt{np}} \right) \right] = -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np}} \right)$$

and hence,

$$\mathbf{P}(Y < y) \leq \exp \left( -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np}} \right) \right).$$

□

Based on this exponential inequality, a crucial but not so obvious fact can be revealed in the following lemma.



**Lemma 4.3.** *Under condition C1\*, for any  $a > 0$ , there is  $c_0(a, \delta) > 0$  such that for  $c > c_0(a, \delta)$ ,*

$$\sup_{n,i} \mathbb{E} \left[ e^{a|Y_{ni}|} \mathbf{I} \{|Y_{ni}| > c\} \right] \leq 2(a+1)e^{-c}. \quad (4.11)$$

*Proof.* By lemma 4.2, for  $y > 0$ ,

$$1 - F_{Y_{ni}}(y) = \mathbf{P}(Y_{ni} > y) \leq \exp \left( -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np_{ni}}} \right) \right)$$

with  $\psi(\lambda) = (2/\lambda^2)[(1+\lambda)\ln(1+\lambda) - \lambda]$ .

Since  $\psi$  is decreasing and  $np_{ni} > \delta$ ,

$$\psi \left( \frac{y}{\sqrt{np_{ni}}} \right) \geq \psi \left( \frac{y}{\sqrt{\delta}} \right) \quad \text{and} \quad \mathbf{P}(Y_{ni} > y) \leq e^{-\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{\delta}} \right)}.$$

Then since  $\lambda\psi(\lambda)$  is increasing and  $\lim_{x \rightarrow 0} \psi(x) = 1$ , if we let  $c_1(a, \delta)$  be the solution of  $y\psi \left( y/\sqrt{\delta} \right) / 2 = a + 1$ ,

$$\mathbf{P}(Y_{ni} > y) \leq e^{-(a+1)y}$$

for all  $y > c_1(a, \delta)$ . Hence for  $c > c_1(a, \delta)$ ,

$$\int_{y>c} e^{ay} dF_{Y_{ni}}(y) = e^{ac}[1 - F_{Y_{ni}}(c)] + a \int_c^\infty [1 - F_{Y_{ni}}(y)] e^{ay} dy \leq (a+1)e^{-c}. \quad (4.12)$$

On the other hand, for  $-\sqrt{np} \leq y < 0$

$$F_{Y_{ni}}(y) = \mathbf{P}(Y_{ni} < y) \leq \exp \left( -\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np_{ni}}} \right) \right),$$

then, since  $\psi(\lambda) \geq 1$  for  $-1 \leq \lambda < 0$ ,

$$-\frac{y^2}{2} \psi \left( \frac{y}{\sqrt{np_{ni}}} \right) \leq -\frac{y^2}{2}.$$

Bearing in mind that  $\mathbf{P}(Y_{ni} < -\sqrt{np}) = 0$ , we have

$$\mathbf{P}(Y_{ni} < y) \leq e^{-\frac{y^2}{2}}.$$

If we let  $c_2(a) = 2(a + 1)$ , then

$$\mathbf{P}(Y_{ni} < y) \leq e^{(a+1)y}$$

for all  $y < -c_2(a)$ . And for  $c > c_2(a)$ ,

$$\int_{y < -c} e^{-ay} dF_{Y_{ni}}(y) = e^{ac} F_{Y_{ni}}(-c) + a \int_{-\infty}^{-c} F_{Y_{ni}}(y) e^{-ay} dy \leq (a+1)e^{-c}. \quad (4.13)$$

Combining (4.12) and (4.13), for  $c > c_0(a, \delta) = \max(c_1(a, \delta), c_2(a))$ , we have (4.11). □

The immediate consequence of Lemma 4.3 is that  $\{be^{a|Y_{ni}|}\}_{n \geq 1, 1 \leq i \leq N}$  is uniformly integrable and

$$\sup_{n,i} \mathbb{E} [be^{a|Y_{ni}|}] \leq M$$

for some  $M$ , which depends only on  $a$ ,  $b$  and  $\delta$ .

Now consider the conditional distribution of  $\nu_{ni}$  given  $\mathcal{F}_{i-1}^{(n)}$ . It is again binomial, but with sample size  $\tilde{n}_{ni} = n - \sum_{j=1}^{i-1} \nu_{nj}$  and probability  $\tilde{p}_{ni} = p_{ni} / (1 - \sum_{j=1}^{i-1} p_{nj})$ .

Recall that if we let  $F_n(t) = \sum_{i=1}^{Nt} p_{ni}$  be the distribution function defined by  $\{p_{ni}\}$ , which converges to a continuous distribution function  $F(t)$ , and  $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{Nt} \nu_{ni}$  be the corresponding empirical distribution function, then  $v_n(t) = \sqrt{n}(\hat{F}_n(t) - F_n(t))$  converges in distribution to the Brownian bridge  $v(t)$  with respect to time  $F(t)$ .

Consider the ratio

$$r_{ni} = \frac{\tilde{n}_{ni}\tilde{p}_{ni}}{np_{ni}} = \frac{1 - \hat{F}_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)}.$$

Then for any  $T < 1$  such that

$$\liminf_n (1 - F_n(T)) > 0, \quad (4.14)$$

by the Kolmogorov-Smirnov theorem, we can easily see that

$$\begin{aligned} \sup_{i \leq NT} |\sqrt{n}(r_{ni} - 1)| &= \sup_i \left| \frac{-v_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)} \right| \\ &\leq \frac{\sup_i |v_n\left(\frac{i-1}{N}\right)|}{\liminf_n (1 - F_n(T))} = O_p(1). \end{aligned} \quad (4.15)$$

Based on these properties, we can establish the uniform integrability of  $\{be^{a|Y_{ni}|}\}_{1 \leq i \leq N}$  in probability under conditional measures.

**Lemma 4.4.** *If Conditions C1\* and*

$$\sup_i p_{ni} \rightarrow 0, \quad (C4^*)$$

*are satisfied, then for any  $0 < \lambda < 1$  and  $\epsilon > 0$ , there is an  $n_\epsilon$  such that for all  $n > n_\epsilon$  and sufficiently large  $c$ ,*

$$\mathbf{P} \left( \sup_i \mathbb{E} \left[ e^{a|Y_{ni}|} \mathbf{I} \{|Y_{ni}| > c\} | \mathcal{F}_{i-1}^{(n)} \right] \leq 2(\tilde{a} + 1)e^{a\lambda - \tilde{c}} \right) > 1 - \epsilon \quad (4.16)$$

*with  $\tilde{a} = a\sqrt{1 + \lambda}$ ,  $\tilde{\delta} = (1 - \lambda)\delta$  and  $\tilde{c} = (c - \lambda)/\sqrt{1 + \lambda}$ . Then,*

$$\mathbf{P} \left( \sup_i \mathbb{E} \left[ be^{a|Y_{ni}|} | \mathcal{F}_{i-1}^{(n)} \right] \leq \tilde{M} \right) > 1 - \epsilon \quad (4.17)$$

*for some  $\tilde{M}$  which depends only on  $a$ ,  $b$ ,  $\delta$  and  $\lambda$ .*

*Proof.* Let  $\tilde{Y}_{ni} = (\nu_{ni} - \tilde{n}_{ni}\tilde{p}_{ni})/\sqrt{\tilde{n}_{ni}\tilde{p}_{ni}}$  then

$$Y_{ni} = \sqrt{r_{ni}}\tilde{Y}_{ni} + \sqrt{np_{ni}}(r_{ni} - 1).$$

Under condition (4.14), the Glivenko-Cantelli theorem implies that

$$\sup_{i \leq NT} |r_{ni} - 1| \xrightarrow{a.s.} 0. \quad (4.18)$$

Under C4\* and by (4.15), we have

$$\sup_{i \leq NT} |\sqrt{np_{ni}}(r_{ni} - 1)| \leq \sup_{i \leq NT} \sqrt{p_{ni}} \sup_{i \leq NT} |\sqrt{n}(r_{ni} - 1)| = o_p(1) \quad (4.19)$$

For any  $0 < \lambda < 1$  and  $\epsilon > 0$ , (4.18) and (4.19) imply that there exists  $n_\epsilon$  such that for all  $n > n_\epsilon$ ,

$$\mathbf{P} \left( \sup_{i \leq NT} |r_{ni} - 1| < \lambda, \sup_{i \leq NT} |\sqrt{np_{ni}}(r_{ni} - 1)| < \lambda \right) > 1 - \epsilon. \quad (4.20)$$

$\sup_{i \leq NT} |r_{ni} - 1| < \lambda$  and  $\sup_{i \leq NT} |\sqrt{np_{ni}}(r_{ni} - 1)| < \lambda$  imply,

$$e^{a|Y_{ni}|} \leq e^{a(\sqrt{1+\lambda}|\tilde{Y}_{ni}|+\lambda)} = e^{a\lambda} e^{a\sqrt{1+\lambda}|\tilde{Y}_{ni}|},$$

$$\inf_i (\tilde{n}_{ni}\tilde{p}_{ni}) = \inf_i (r_{ni}np_{ni}) \geq (1 - \lambda)\delta = \tilde{\delta} > 0$$

and

$$\mathbf{I}\{|Y_{ni}| > c\} \leq \mathbf{I}\left\{\sqrt{1+\lambda}|\tilde{Y}_{ni}| + \lambda > c\right\} = \mathbf{I}\left\{|\tilde{Y}_{ni}| > \tilde{c}\right\}.$$

Therefore,

$$\sup_i \mathbb{E} \left[ e^{a|Y_{ni}|} \mathbf{I}\{|Y_{ni}| > c\} | \mathcal{F}_{i-1}^{(n)} \right] \leq e^{a\lambda} \sup_i \mathbb{E} \left[ e^{a\sqrt{1+\lambda}|\tilde{Y}_{ni}|} \mathbf{I}\left\{|\tilde{Y}_{ni}| > \tilde{c}\right\} | \mathcal{F}_{i-1}^{(n)} \right].$$

By applying lemma 4.3, we get (4.16) and hence (4.17). □

Another crucial but not well-known fact is presented in the following lemma.

**Lemma 4.5.** *If Conditions  $C1^*$  and  $C4^*$  are satisfied, and if*

$$|y_{ni}| = \left| \frac{k - np_{ni}}{\sqrt{np_{ni}}} \right| \leq c_{ni} = \frac{1}{\sqrt{p_{ni}}} o(1)$$

*with  $o(1)$  independent of  $i$ , then for all  $i \leq NT$ ,*

$$\begin{aligned} D_{ni}(y_{ni}) &= \frac{B(k, \tilde{n}_{ni}, \tilde{p}_{ni})}{B(k, n, p_{ni})} - 1 \\ &= y_{ni} \sqrt{np_{ni}}(r_{ni} - 1) + p_{ni} O_p(1) + y_{ni}^2 p_{ni} O_p(1) + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \quad (4.21)$$

*with  $O_p(1)$  and  $o_p\left(\frac{1}{\sqrt{n}}\right)$  both independent of  $i$ .*

*Proof.* For the sake of simplicity, we use  $p, \tilde{p}, \tilde{n}, y, c$  and  $r$  in place of  $p_{ni}, \tilde{p}_{ni}, \tilde{n}_{ni}, y_{ni}, c_{ni}$  and  $r_{ni}$  respectively. It is easy to see that  $|y\sqrt{p/n}| = o(1/\sqrt{n})$ . Let  $\tilde{y} = (k - \tilde{n}\tilde{p})/\sqrt{\tilde{n}\tilde{p}}$ . Under Conditions C1\* and C4\*, (4.15) implies that

$$|\tilde{y} - y| = \left| \left( \frac{1}{\sqrt{r}} - 1 \right) y - \frac{1}{\sqrt{r}} \sqrt{np}(r - 1) \right| \leq \frac{1}{\delta} o_p(1) + \sqrt{p} O_p(1) = o_p(1).$$

Since  $\sqrt{\tilde{p}/\tilde{n}} = \sqrt{p/n} O_p(1)$ ,

$$\left| \tilde{y} \sqrt{\frac{\tilde{p}}{\tilde{n}}} \right| \leq (|y| + |\tilde{y} - y|) \sqrt{\frac{\tilde{p}}{\tilde{n}}} = o_p \left( \frac{1}{\sqrt{n}} \right).$$

Therefore, we have

$$\frac{k}{n} = p + y\sqrt{\frac{p}{n}} = p + o \left( \frac{1}{\sqrt{n}} \right) \quad \text{and} \quad \frac{k}{\tilde{n}} = \tilde{p} + o_p \left( \frac{1}{\sqrt{n}} \right).$$

Applying Stirling's formula, we have,

$$\begin{aligned} \frac{B(k, \tilde{n}, \tilde{p})}{B(k, n, p)} &= \left( \frac{\tilde{n}\tilde{p}}{np} \right)^k \left( \frac{1 - \frac{k}{n}}{1 - p} \right)^{n-k} \left( \frac{1 - \tilde{p}}{1 - \frac{k}{\tilde{n}}} \right)^{\tilde{n}-k} \sqrt{\frac{1 - \frac{k}{n}}{1 - \frac{k}{\tilde{n}}}} \left( 1 + o \left( \frac{1}{\sqrt{n}} \right) \right) \\ &= e^A \left( 1 + o \left( \frac{1}{\sqrt{n}} \right) \right) \end{aligned}$$

with

$$A = k \ln \left( \frac{\tilde{n}\tilde{p}}{np} \right) + (n - k) \ln \left( \frac{1 - \frac{k}{n}}{1 - p} \right) - (\tilde{n} - k) \ln \left( \frac{1 - \frac{k}{\tilde{n}}}{1 - \tilde{p}} \right) + \frac{1}{2} \ln \left( \frac{1 - \frac{k}{n}}{1 - \frac{k}{\tilde{n}}} \right).$$

Let's consider each term. We have

$$k \ln \left( \frac{\tilde{n}\tilde{p}}{np} \right) = np(r - 1) + y\sqrt{np}(r - 1) + \frac{np(r - 1)^2}{2} + o_p \left( \frac{1}{\sqrt{n}} \right), \quad (4.22)$$

$$(n - k) \ln \left( \frac{1 - \frac{k}{n}}{1 - p} \right) = -y\sqrt{np} + \frac{y^2 p}{2(1 - p)^2} + o \left( \frac{1}{\sqrt{n}} \right), \quad (4.23)$$

$$(\tilde{n} - k) \ln \left( \frac{1 - \frac{k}{\tilde{n}}}{1 - \tilde{p}} \right) = -\tilde{y}\sqrt{\tilde{n}\tilde{p}} + \frac{\tilde{y}^2 \tilde{p}}{2(1 - \tilde{p})^2} + o_p \left( \frac{1}{\sqrt{n}} \right), \quad (4.24)$$

and

$$\ln \left( \frac{1 - \frac{k}{n}}{1 - \frac{k}{\tilde{n}}} \right) = \ln \left( 1 + \frac{\frac{k}{\tilde{n}} - \frac{k}{n}}{1 - \frac{k}{\tilde{n}}} \right) = \tilde{p} - p + p o_p(1) + o_p \left( \frac{1}{\sqrt{n}} \right). \quad (4.25)$$

Since

$$\tilde{y}\sqrt{\tilde{n}\tilde{p}} - y\sqrt{np} = np - \tilde{n}\tilde{p} = -np(r-1)$$

(4.22)-(4.25) imply that

$$\begin{aligned} A &= y\sqrt{np}(r-1) + \frac{np(r-1)^2}{2} + \frac{y^2p}{2(1-p)^2} - \frac{\tilde{y}^2\tilde{p}}{2(1-\tilde{p})^2} + \frac{\tilde{p}-p}{2} + p \cdot o_p(1) + o_p \left( \frac{1}{\sqrt{n}} \right) \\ &= y\sqrt{np}(r-1) + p \cdot O_p(1) + y^2p \cdot O_p(1) + o_p \left( \frac{1}{\sqrt{n}} \right) \end{aligned}$$

with  $O_p(1)$  and  $o_p \left( \frac{1}{\sqrt{n}} \right)$  independent of  $i$ . Since  $A \rightarrow 0$ ,

$$e^A - 1 = A + A^2/2 + o(A)$$

which implies (4.21). □

### 4.3.2 FLT for Martingale Component

Based on the auxiliary lemmas proved in previous section, we can establish the FLTs. We first establish the limit theorem for the martingale component.

**Theorem 4.7.** *Assume the Conditions  $C1^*$ - $C4^*$  are satisfied and hence*

$$h_{n[Nt]} \left( Y_{n[Nt]}, np_{n[Nt]} \right) \xrightarrow{d} \xi(t) = h_t(Y(t), \lambda(t)). \quad (4.26)$$

Let us denote by  $w$  a standard Brownian motion on  $\mathbb{R}^+$ , and let  $W(t) = w(\tau(t))$  with

$$\tau(t) = \int_0^t \sigma^2(s) ds, \quad (4.27)$$

with  $\sigma^2(s)$  being the variance of  $\xi(s)$ . Then for  $t < T$ ,

$$W_n \xrightarrow{d} W.$$

*Proof.* Letting  $\xi_{ni} = h_{ni}(Y_{ni}, np_{ni})$  and  $\eta_{ni} = \mathbb{E}[\xi_{ni} | \mathcal{F}_{i-1}]$ , so that  $W_n$  is a martingale with differences  $(\xi_{ni} - \eta_{ni}) / \sqrt{N}$ . According to Theorem 4.4, to prove the theorem, it is necessary and sufficient to prove that

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ (\xi_{ni} - \eta_{ni})^2 \mathbf{I}\{(\xi_{ni} - \eta_{ni})^2 > \epsilon N\} | \mathcal{F}_{i-1}^{(n)} \right] \xrightarrow{P} 0 \quad \forall \epsilon > 0 \quad (\alpha)$$

and

$$\frac{1}{N} \sum_{i=1}^{Nt} \mathbb{E} \left[ (\xi_{ni} - \eta_{ni})^2 | \mathcal{F}_{i-1}^{(n)} \right] \xrightarrow{P} \tau(t) \quad (\beta)$$

are satisfied.

To verify  $(\alpha)$ , it is sufficient to show that

$$\sup_i \mathbb{E} \left[ (\xi_{ni} - \eta_{ni})^2 \mathbf{I}\{(\xi_{ni} - \eta_{ni})^2 > \epsilon N\} | \mathcal{F}_{i-1}^{(n)} \right] \xrightarrow{P} 0. \quad (4.28)$$

Recall that under condition C2\*,  $|\xi_{ni}| \leq be^{a|Y_{ni}|}$ . By Lemma 4.4, consider a fixed  $0 < \lambda < 1$ ; then for any  $\epsilon > 0$ , there exists  $n_\epsilon$  such that for all  $n > n_\epsilon$ , (4.20) holds and

$$\mathbf{P} \left( \sup_i |\eta_{ni}| \leq \sup_i \mathbb{E} \left[ |\xi_{ni}| | \mathcal{F}_{i-1}^{(n)} \right] \leq \tilde{M} \right) > 1 - \epsilon. \quad (4.29)$$

with some  $\tilde{M} < \infty$  only depending on  $a, b, \delta, \lambda$ .

Since  $|\eta_{ni}| \leq \tilde{M}$  implies that

$$(\xi_{ni} - \eta_{ni})^2 \leq 2\xi_{ni}^2 + 2\eta_{ni}^2 \leq 2\xi_{ni}^2 + 2\tilde{M}^2 \leq 2(b^2 + \tilde{M}^2) e^{2a|Y_{ni}|}$$

and hence

$$\mathbf{I}\{(\xi_{ni} - \eta_{ni})^2 > \epsilon N\} \leq \mathbf{I}\left\{|Y_{ni}| > \frac{1}{2a} \ln \frac{\epsilon N}{2(b^2 + \tilde{M}^2)} = c\right\},$$

(4.16) implies that the left side of (4.28) is bounded by

$$\sup_i \mathbb{E} \left[ 2(b^2 + \tilde{M}^2) e^{2a|Y_{ni}|} \mathbf{I}\{|Y_{ni}| > c\} | \mathcal{F}_{i-1}^{(n)} \right] \leq 4(b^2 + \tilde{M}^2) (2\tilde{a} + 1) e^{2a\lambda - \tilde{c}},$$

with  $\tilde{a}$  and  $\tilde{c}$  defined in Lemma 4.4.

For sufficiently large  $n$ ,  $N$  and hence  $c$  are large enough and therefore the right side can be arbitrarily small. Since  $\epsilon$  is arbitrary, (4.29) holds and  $(\alpha)$  is satisfied.

For  $(\beta)$ , consider the step functions

$$\sigma_n^2(t) = \mathbb{E} \left[ \left( \xi_{n[Nt]} - \eta_{n[Nt]} \right)^2 \middle| \mathcal{F}_{[Nt]-1}^{(n)} \right].$$

Then Lemma 4.4 and (4.26) imply  $\sigma_n^2(t) \xrightarrow{P} \sigma^2(t)$  and

$$\sigma_n^2(t) \leq \mathbb{E} \left[ \xi_{n[Nt]}^2 \middle| \mathcal{F}_{[Nt]-1}^{(n)} \right] \leq \sup_{t \leq T} \mathbb{E} \left[ \xi_{n[Nt]}^2 \middle| \mathcal{F}_{[Nt]-1}^{(n)} \right] < \infty$$

in probability. Obviously,  $\sup_{t \leq T} \mathbb{E} \left[ \xi_{n[Nt]}^2 \middle| \mathcal{F}_{[Nt]-1}^{(n)} \right]$  is integrable with respect to  $t \in [0, T]$ . By the dominated convergence theorem, for  $t \leq T$ ,

$$\int_0^t \sigma_n^2(s) ds \xrightarrow{P} \tau(t).$$

Hence  $(\beta)$  holds and therefore the theorem is proved. □

### 4.3.3 FLT for the Compensator Component

The limit theorem for the compensator component is established below.

**Theorem 4.8.** *Define the operator*

$$\mathcal{K}x(t) = - \int_0^t k(s)x(s)ds$$

with

$$k(s) = \mathbb{E} [\xi(s)Y(s)]\sqrt{f(s)} = \mathbb{E} [h_s(Y(s), \lambda(s))Y(s)]\sqrt{f(s)}$$

If conditions  $C1^*$ - $C4^*$  are satisfied,

$$\limsup_n (Np_{n[Nt]}) < Cf(t) \quad \forall \quad t < T \tag{C5*}$$



for some  $C < \infty$  and there is some  $\alpha > 0$  such that

$$\left( \sup_i p_{ni} \right)^{-1/2+\alpha} - \frac{1}{2} \ln N \rightarrow \infty. \quad (\text{C6}^*)$$

Then for  $t < T$

$$K_n \xrightarrow{d} K = \mathcal{K} \frac{v}{1-F}.$$

*Proof.* Instead of  $K_n(t)$ , we consider the truncated process

$$K_n^c(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \mathbb{E} \left[ \xi_{ni} \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \} \mid \mathcal{F}_{i-1}^{(n)} \right] - \mathbb{E} [\xi_{ni} \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \}] \right)$$

with  $c_{ni} = p_{ni}^{-1/2+\alpha/2} = p_{ni}^{-1/2} o(1)$ .

It can be shown by Lemma 4.3 and 4.4 that the difference

$$K_n(t) - K_n^c(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \mathbb{E} \left[ \xi_{ni} \mathbf{I} \{ |Y_{ni}| > c_{ni} \} \mid \mathcal{F}_{i-1}^{(n)} \right] - \mathbb{E} [\xi_{ni} \mathbf{I} \{ |Y_{ni}| > c_{ni} \}] \right)$$

is asymptotically negligible.

C6\* implies  $\sqrt{N} \exp(-\inf_i c_{ni}) \rightarrow 0$ . By Lemma 4.3, we have

$$\begin{aligned} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \mathbb{E} [\xi_{ni} \mathbf{I} \{ |Y_{ni}| > c_{ni} \}] \right| &\leq \sqrt{N} \sup_i \mathbb{E} \left[ |\xi_{ni}| \mathbf{I} \left\{ |Y_{ni}| > \inf_i c_{ni} \right\} \right] \\ &\leq b\sqrt{N} \sup_i \mathbb{E} \left[ e^{a|Y_{ni}|} \mathbf{I} \left\{ |Y_{ni}| > \inf_i c_{ni} \right\} \right] \leq 2b(a+1)\sqrt{N} e^{-\inf_i c_{ni}} \rightarrow 0. \end{aligned}$$

Similarly, by Lemma 4.4, for any fixed  $\lambda$  with  $0 < \lambda < 1$ , and setting  $\tilde{c}_{ni} = (c_{ni} - \lambda)/\sqrt{1+\lambda}$ ,  $\sqrt{N} \exp(-\inf_i \tilde{c}_{ni}) \rightarrow 0$  as well. Then we have

$$\left| \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \mathbb{E} \left[ \xi_{ni} \mathbf{I} \{ |Y_{ni}| > c_{ni} \} \mid \mathcal{F}_{i-1}^{(n)} \right] \right| \leq 2b(\tilde{a}+1)\sqrt{N} e^{a\lambda - \inf_i \tilde{c}_{ni}} \rightarrow 0$$

in probability. Since  $\inf_i c_{ni} = (\sup_i p_{ni})^{-1/2+\alpha/2} \rightarrow \infty$ ,

$$\sup_{t \leq NT} |K_n(t) - K_n^c(t)| \xrightarrow{p} 0. \quad (4.30)$$

The proof of weak convergence of  $K_n^c(t)$  is based on the following equality,

$$\mathbb{E} \left[ \xi_{ni} \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \} | \mathcal{F}_{i-1}^{(n)} \right] - \mathbb{E} [\xi_{ni} \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \}] = \mathbb{E} [\xi_{ni} D_{ni}(Y_{ni}) \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \}]$$

with  $D_{ni}(y)$  defined in (4.21).

Since

$$\sup_{t \leq NT} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \mathbb{E} [\xi_{ni} Y_{ni}^2 \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \}] p_{ni} \right| \xrightarrow{p} 0$$

and

$$\sup_{t \leq NT} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} p_{ni} \right| \rightarrow 0,$$

Lemma 4.5 implies

$$\sup_{t \leq NT} \left| K_n^c(t) - \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \mathbb{E} [\xi_{ni} Y_{ni} \mathbf{I} \{ |Y_{ni}| \leq c_{ni} \}] \sqrt{np_{ni}} (r_{ni} - 1) \right| = o_p(1). \quad (4.31)$$

Let

$$k_n^c(t) = E [\xi_{n[Nt]} Y_{n[Nt]} \mathbf{I} (|Y_{n[Nt]}| \leq c_{n[Nt]})] \sqrt{Np_{n[Nt]}}$$

and

$$\sqrt{n}(r_{n[Nt]} - 1) = -\frac{v_n(t)}{(1 - F_n(t))}.$$

Then (4.30) and (4.31) imply

$$\sup_{t \leq NT} \left| K_n(t) + \int_0^t k_n^c(s) \frac{v_n(s)}{1 - F_n(s)} ds \right| = o_p(1).$$

Since  $\sup_t |E [\xi_{n[Nt]} Y_{n[Nt]} \mathbf{I} (|Y_{n[Nt]}| \leq c_{n[Nt]})]|$  is bounded in probability, under condition C5\*,  $k_n^c(t)$  is integrable. Further since  $k_n^c(t) \rightarrow k(t)$  for all  $t < T$ ,

$$\sup_{t \leq NT} \left| \int_0^t k_n^c(s) \frac{v_n(s)}{1 - F_n(s)} ds - \int_0^t k(s) \frac{v_n(s)}{1 - F_n(s)} ds \right| = o_p(1).$$

Since the operator  $\mathcal{K}^*$  is continuous and for  $t < T$ ,

$$\frac{v_n(t)}{1 - F_n(t)} \xrightarrow{d} \frac{v(t)}{1 - F(t)},$$

the theorem is proved.  $\square$

#### 4.3.4 Alternative Proof to FLT of Compensator of Chi-square Process

The FLTs established in Sections 4.2.2 and 4.3.3 are for partial sum processes of general divisible statistic and the proof was very technical. However, for the Chi-square process

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (Y_{ni}^2 - \mathbb{E}[Y_{ni}^2]) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (Y_{ni}^2 - (1 - p_{ni}))$$

a simpler alternative proof can be acquired. The consistency of the results from both methods illustrate the power of the theorems we have established.

Let

$$\tilde{Y}_{ni} = \frac{\nu_{ni} - \tilde{n}\tilde{p}_{ni}}{\sqrt{\tilde{n}\tilde{p}_{ni}}},$$

then

$$\begin{aligned} Y_{ni}^2 &= \frac{(\nu_{ni} - np_{ni})^2}{np_{ni}} = \frac{(\sqrt{\tilde{n}\tilde{p}_{ni}}\tilde{Y}_{ni} + \tilde{n}\tilde{p}_{ni} - np_{ni})^2}{np_{ni}} \\ &= \tilde{Y}_{ni}^2 \frac{\tilde{n}\tilde{p}_{ni}}{np_{ni}} + 2\tilde{Y}_{ni} \frac{\sqrt{\tilde{n}\tilde{p}_{ni}}(\tilde{n}\tilde{p}_{ni} - np_{ni})}{np_{ni}} + \frac{(\tilde{n}\tilde{p}_{ni} - np_{ni})^2}{np_{ni}}. \end{aligned}$$

Since  $\mathbb{E}(\tilde{Y}_{ni}^2 \mid \mathcal{F}_{i-1}) = 1 - \tilde{p}_{ni}$ ,  $\mathbb{E}Y_{ni} = 0 = \mathbb{E}(\tilde{Y}_{ni} \mid \mathcal{F}_{i-1}^{(n)})$  and

$$\begin{aligned} \tilde{n}\tilde{p}_{ni} - np_{ni} &= np_{ni} \left[ \frac{1 - \hat{F}_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)} - 1 \right] \\ &= -np_{ni} \frac{\hat{F}_n\left(\frac{i-1}{N}\right) - F_n\left(\frac{i-1}{N}\right)}{1 - F_n\left(\frac{i-1}{N}\right)} \\ &= -\sqrt{n} \frac{p_{ni}}{1 - F_n\left(\frac{i-1}{N}\right)} v_n \left( \frac{i-1}{N} \right), \end{aligned}$$

we have

$$\mathbb{E}(Y_{ni}^2 \mid \mathcal{F}_{i-1}^{(n)}) = (1 - \tilde{p}_{ni}) \frac{\tilde{n}\tilde{p}_{ni}}{np_{ni}} + p_{ni} \left( \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right)^2.$$

Then since  $\mathbb{E}Y_{ni}^2 = 1 - p_{ni}$ ,

$$\begin{aligned} & \mathbb{E}(Y_{ni}^2 \mid \mathcal{F}_{i-1}^{(n)}) - \mathbb{E}Y_{ni}^2 \\ &= (1 - \tilde{p}_{ni}) \frac{\tilde{n}\tilde{p}_{ni}}{np_{ni}} + p_{ni} \left( \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right)^2 - (1 - p_{ni}) \\ &= (1 - \tilde{p}_{ni}) \left( \frac{\tilde{n}\tilde{p}_{ni}}{np_{ni}} - 1 \right) + (p_{ni} - \tilde{p}_{ni}) + p_{ni} \left( \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right)^2 \\ &= -\frac{1}{\sqrt{n}}(1 - \tilde{p}_{ni}) \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} - \frac{F_n(\frac{i-1}{N})p_{ni}}{1 - F_n(\frac{i-1}{N})} + p_{ni} \left( \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right)^2. \end{aligned}$$

Next, for  $1 - F_n(T) > \delta > 0$  and  $i \leq NT$ ,

$$\left| -\frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \frac{F_n(\frac{i-1}{N})p_{ni}}{1 - F_n(\frac{i-1}{N})} \right| \leq \frac{F_n(T)}{\sqrt{N}(1 - F_n(T))} \rightarrow 0.$$

Since  $\mathbf{P}(\sup_i |v_n(\frac{i-1}{N})| > \lambda) < 58 \exp(-2\lambda^2)$  (Dvoretzky, Kiefer, Wolfowitz; see p354, Shorack and Wellner, 1986),

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( p_{ni} \left( \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right)^2 \right) \leq \frac{1}{(1 - F_n(T))^2} \sup \frac{v_n(\frac{i-1}{N})^2}{\sqrt{N}} \xrightarrow{p} 0.$$

Similarly, we have

$$\left| \frac{1}{\sqrt{nN}} \sum_{i=1}^{Nt} \left( \tilde{p}_{ni} \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right) \right| \leq \frac{1}{(1 - F_n(T))^2} \sup \frac{|v_n(\frac{i-1}{N})|}{\sqrt{nN}} \xrightarrow{p} 0.$$

Therefore, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (\mathbb{E}(Y_{ni}^2 \mid \mathcal{F}_{i-1}) - \mathbb{E}Y_{ni}^2) = -\frac{1}{\sqrt{nN}} \sum_{i=1}^{Nt} \left( \frac{v_n(\frac{i-1}{N})}{1 - F_n(\frac{i-1}{N})} \right) + o_p(1).$$

And finally,

$$K_n(t) \xrightarrow{d} -\frac{1}{\sqrt{d}} \int_0^t \frac{v(s)}{1-F(s)} ds. \quad (4.32)$$

We will show in chapter 5 that for the Chi-square process,  $k(s) = 1/\sqrt{d}$ , and the consistency of (4.32) on the one hand, and the results of Theorems 4.6 and 4.8 on the other hand, is obvious.

## 4.4 FLT for Partial Sum Processes

The separate limit processes  $W$  and  $K$  of the martingale and compensator components are not sufficient to describe the limiting behaviour of the partial sum process  $X_n$ , unless we find the correlation structure of the two limit processes. Khmaladze (1984) has revealed this correlation structure for models with asymptotically Poissonian frequencies in the following theorem.

Let  $(v, w)$  be a two-dimensional Gaussian process adapted to the flow of  $\sigma$ -algebra generated by the process  $v$ , where the correlation function of  $v$  is  $F(t \wedge s) - F(t)F(s)$ , the correlation function of  $w$  is  $\tau(t \wedge s)$  and the mutual correlation function of  $v$  and  $w$  is

$$\mathbb{E}v(t)w(s) = [1 - F(t)] \int_0^{t \wedge s} \frac{k(y)}{1 - F(y)} dy, \quad (4.33)$$

where  $k$  is as defined in theorem 4.6.

Based on this correlation structure between  $v$  and  $w$ , the convergence of  $X_n$  is then described in the following theorem.

**Theorem 4.9** (Khmaladze (1984)). *Suppose the conditions (C1-C4) hold. Then*

$$X_n \xrightarrow{d} X \quad \text{as } n \rightarrow \infty$$

where

$$X(t) = - \int_0^t k(s) \frac{v(s)}{1 - F(s)} ds + w(t).$$

*Proof.* See Khmaladze (1984). □

**Remark 4.2.** The FLTs established in Theorems 4.5 and 4.6 for models with Poissonian frequencies and Theorems 4.7 and 4.8 for models with mixed frequencies are of the same structure, where  $\sigma^2(s)$  and  $k(s)$  play the same roles the two schemes. Since the establishment of the correlation function (4.33) and the proof of Theorem 4.9 are irrelevant to the specific form of  $\sigma^2(s)$  and  $k(s)$ , both (4.33) and Theorem 4.9 apply to the new scheme with the models for mixed frequencies.

## 4.5 FLTs under Contiguous Alternatives

The limit theorems established in previous sections were under the null hypothesis. However, it is also important to consider the asymptotic behaviour of the processes under the alternative hypotheses, i.e., the “true” probabilities are not  $\mathbf{p}_n = \{p_{ni}\}_{1 \leq i \leq N}$  but  $\bar{\mathbf{p}}_n = \{\bar{p}_{ni}\}_{1 \leq i \leq N}$ . In particular, it is often interesting to consider  $\bar{\mathbf{p}}_n$  as a sequence of probability measures contiguous to  $\mathbf{p}_n$  in the following form.

**Condition 4.1.** Let  $\bar{\mathbf{p}}_n = \{\bar{p}_{ni}\}_{1 \leq i \leq N}$  be a sequence of probabilities with  $\bar{p}_{ni} = p_{ni}(1 + c_{ni}/\sqrt{n})$ ; and suppose there exists a function  $c(t)$  such that

$$\int_0^1 c^2(t) f(t) dt < \infty \quad \text{and} \quad \int_0^1 (c_{n[Nt]} - c(t))^2 f(t) dt \rightarrow 0,$$

where  $f(t) = \lim f_n(t) = \lim(Np_{n[Nt]})$ .

Under this sequence of contiguous alternatives, (C1\*-C6\*) hold if (C1\*-C6\*)

hold under the null hypothesis. Therefore, if we define

$$\bar{X}_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) - \bar{\mathbb{E}}[h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni})]),$$

$$\bar{W}_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) - \bar{\mathbb{E}}[h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) | \mathcal{F}_{i-1}^{(n)}] \right),$$

and

$$\bar{K}_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \bar{\mathbb{E}}[h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) | \mathcal{F}_{i-1}^{(n)}] - \bar{\mathbb{E}}[h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni})] \right),$$

with  $\bar{Y}_{ni} = (\nu_{ni} - n\bar{p}_{ni})/\sqrt{n\bar{p}_{ni}}$  and  $\bar{E}$  being expectation under the alternative, then  $\bar{W}_n \xrightarrow{d} W$ ,  $\bar{K}_n \xrightarrow{d} K$  and  $\bar{X}_n \xrightarrow{d} X$  under the alternative.

Now we introduce the definitions of some shift processes:

$$m_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (\bar{\mathbb{E}}[h_{ni}(Y_{ni}, np_{ni})] - \mathbb{E}[h_{ni}(Y_{ni}, np_{ni})]),$$

$$l_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \bar{\mathbb{E}}[h_{ni}(Y_{ni}, np_{ni}) | \mathcal{F}_{i-1}^{(n)}] - \mathbb{E}[h_{ni}(Y_{ni}, np_{ni}) | \mathcal{F}_{i-1}^{(n)}] \right),$$

and

$$q_n = m_n - l_n.$$

Then for all  $t < 1$ ,

$$\begin{aligned} W_n(t) - \bar{W}_n(t) - l_n(t) = & \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( (h_{ni}(Y_{ni}, np_{ni}) - h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni})) \right. \\ & \left. - \bar{\mathbb{E}}[h_{ni}(Y_{ni}, np_{ni}) - h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) | \mathcal{F}_{i-1}^{(n)}] \right), \quad (4.34) \end{aligned}$$

$$\begin{aligned}
X_n(t) - \bar{X}_n(t) - m_n(t) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} ((h_{ni}(Y_{ni}, np_{ni}) - h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni})) \\
&\quad - \bar{\mathbb{E}}[h_{ni}(Y_{ni}, np_{ni}) - h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni})]), \quad (4.35)
\end{aligned}$$

and

$$\begin{aligned}
K_n(t) - \bar{K}_n(t) - q_n(t) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (\bar{\mathbb{E}}[h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) - h_{ni}(Y_{ni}, np_{ni})] \\
&\quad - \bar{\mathbb{E}}[h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) - h_{ni}(Y_{ni}, np_{ni}) \mid \mathcal{F}_{i-1}^{(n)}]) \quad (4.36)
\end{aligned}$$

Next, considering functions  $g_{ni}^*$  such that

$$g_{ni}^*(\nu_{ni}, \sqrt{np_{ni}}) = h_{ni}(Y_{ni}, np_{ni}) = g_{ni}(\nu_{ni}, np_{ni}),$$

we can introduce

$$g_{ni}^\#(\nu_{ni}, np_{ni}) = \frac{\partial g_{ni}^*(\nu_{ni}, \sqrt{np_{ni}})}{\partial \sqrt{np_{ni}}}$$

and

$$h_{ni}^\#(Y_{ni}, np_{ni}) = g_{ni}^\#(\nu_{ni}, np_{ni}),$$

from which the following lemma can be established.

**Lemma 4.6.** *If conditions (C1\*-C6\*) hold and*

$$h_{ni}^\# \text{ satisfies } C2^*, \quad (C7^*)$$

*then for all  $t < T$  with  $T$  satisfying (4.14),*

$$\|W_n - \bar{W}_n - l_n\| \xrightarrow{p} 0,^1$$

$$\|X_n - \bar{X}_n - m_n\| \xrightarrow{p} 0,$$

*and*

$$\|K_n - \bar{K}_n - q_n\| \xrightarrow{p} 0.$$

---

<sup>1</sup>Since  $\bar{\mathbf{p}}_n$  are contiguous to  $\mathbf{p}_n$ , the convergence in probability to 0 under  $\bar{\mathbf{p}}_n$  is equivalent to that under  $\mathbf{p}_n$ , and we do not distinguish them here.



*Proof.* Since

$$\sqrt{n\bar{p}_{ni}} - \sqrt{np_{ni}} = \sqrt{np_{ni}} \left( \sqrt{1 + \frac{c_{ni}}{\sqrt{n}}} - 1 \right) = O(\sqrt{p_{ni}}) \rightarrow 0,$$

we have

$$h_{ni}(Y_{ni}, np_{ni}) - h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni}) = h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) O(\sqrt{\bar{p}_{ni}}),$$

and

$$\begin{aligned} & \frac{1}{\sqrt{N}} \sum_{i=1}^N (h_{ni}(Y_{ni}, np_{ni}) - h_{ni}(\bar{Y}_{ni}, n\bar{p}_{ni})) \\ &= O\left(\sqrt{\sup_i \bar{p}_{ni}}\right) \frac{1}{\sqrt{N}} \sum_{i=1}^N h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}). \end{aligned}$$

Since  $h_{ni}^{\#}$  satisfies C2\*, then theorems 4.7 and 4.8 imply that

$$\bar{X}_n^{\#}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) - \bar{\mathbb{E}} \left[ h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) \right] \right),$$

$$\bar{W}_n^{\#}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) - \bar{\mathbb{E}} \left[ h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) \mid \mathcal{F}_{i-1}^{(n)} \right] \right),$$

and

$$\bar{K}_n^{\#}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \bar{\mathbb{E}} \left[ h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) \mid \mathcal{F}_{i-1}^{(n)} \right] - \bar{\mathbb{E}} \left[ h_{ni}^{\#}(\bar{Y}_{ni}, n\bar{p}_{ni}) \right] \right),$$

converge in distribution to  $X^{\#}$ ,  $W^{\#}$  and  $K^{\#}$  respectively. Hence  $\|\bar{X}_n^{\#}\| = O_p(1)$ ,  $\|\bar{W}_n^{\#}\| = O_p(1)$  and  $\|\bar{K}_n^{\#}\| = O_p(1)$  for  $t < T$  and  $\|W_n - \bar{W}_n - l_n\|$ ,  $\|X_n - \bar{X}_n - m_n\|$  and  $\|K_n - \bar{K}_n - q_n\|$  are all  $O_p(\sqrt{\sup \bar{p}_{ni}})$  for  $t < T$ . Since  $\sup \bar{p}_{ni} \rightarrow 0$  by C4\*, the lemma is established.  $\square$

Based on this lemma, the focus can turn to the limit theorems of the shifts  $m_n$ ,  $l_n$  and  $q_n$ .

**Theorem 4.10.** *If the conditions (C1\*-C6\*) hold, then for  $t < T$*

$$m_n \xrightarrow{d} m = -\mathcal{K}c.$$

*Proof.* The proof in this theorem is analogous to that of theorem 4.8, except that the role of

$$\sqrt{n}(r_{n[Nt]} - 1) = -\frac{v_n(t)}{(1 - F_n(t))}$$

has been played by

$$\sqrt{n}(\bar{r}_{n[Nt]} - 1) = \sqrt{n} \left( \frac{n\bar{p}_{n[Nt]}}{np_{n[Nt]}} - 1 \right) = c_{n[Nt]},$$

and  $c_{n[Nt]} \rightarrow c(t)$ . □

**Theorem 4.11.** *If the conditions (C1\*-C6\*) hold, then for  $t < T$*

$$q_n \xrightarrow{d} q = -\mathcal{K} \frac{C}{1 - F}$$

with  $C(t) = \int_0^t c(s)f(s)ds$ .

*Proof.* If let

$$\bar{K}_n^*(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \bar{\mathbb{E}} \left[ h_{ni}(Y_{ni}, np_{ni}) \mid \mathcal{F}_{i-1}^{(n)} \right] - \bar{\mathbb{E}}[h_{ni}(Y_{ni}, np_{ni})] \right),$$

then we can get another expression for  $q_n$ :

$$q_n = K_n - \bar{K}_n^*.$$

The limit of  $\bar{K}_n^*(t)$  can be derived in an analogous fashion to that in the proof of theorem 4.8, except that the role of

$$\frac{v_n(t)}{(1 - F_n(t))} = \frac{\sqrt{n}(\hat{F}_n(t) - F_n(t))}{1 - F_n(t)}$$

is now played by

$$\frac{\bar{v}_n(t)}{(1 - \bar{F}_n(t))} = \frac{\sqrt{n}(\hat{F}_n(t) - \bar{F}_n(t))}{1 - \bar{F}_n(t)}.$$

Since

$$\bar{F}_n(t) = \sum_{i=1}^{Nt} \bar{p}_{ni} = \sum_{i=1}^{Nt} \bar{p}_{ni} \left(1 + \frac{c_{ni}}{\sqrt{n}}\right) = F_n(t) + \frac{1}{\sqrt{n}} \sum_{i=1}^{Nt} c_{ni} p_{ni},$$

we have

$$\frac{\bar{v}_n(t)}{(1 - \bar{F}_n(t))} = \frac{\sqrt{n} \left( \hat{F}_n(t) - F_n(t) \right)}{1 - \bar{F}_n(t)} + \frac{\sum_{i=1}^{Nt} c_{ni} p_{ni}}{1 - \bar{F}_n(t)} \xrightarrow{d} \frac{v(t)}{1 - F(t)} + \frac{C(t)}{1 - F(t)},$$

and hence

$$\bar{K}_n^* \xrightarrow{d} \mathcal{K} \left( \frac{v}{1 - F} + \frac{C}{1 - F} \right).$$

Since

$$K_n \xrightarrow{d} \mathcal{K} \left( \frac{v}{1 - F} \right),$$

the theorem is proved.  $\square$

Since both  $m$  and  $q$  are deterministic functions, the following assertions can be established,

$$l_n \xrightarrow{d} l = m - q = \mathcal{K} \left( c - \frac{C}{1 - F} \right),$$

$$W_n \xrightarrow{d} W + l,$$

$$K_n \xrightarrow{d} K + q,$$

and

$$X_n \xrightarrow{d} X + m.$$

In the models with all Poissonian frequencies, replacing C7\* by

$$g_{ni}^{\#} \text{ satisfies C2,} \tag{C5}$$

will achieve similar results. In particular, if  $g_{ni}(\nu_{ni}, np_{ni})$  is independent of  $np_{ni}$ , i.e.,  $g_{ni}(\nu_{ni}, np_{ni}) = g_{ni}(\nu_{ni})$  then simply  $g_{ni}^{\#}(\nu_{ni}, np_{ni}) = 0$ . This was the case considered by Khmaladze (1984).

## 4.6 Comparison of FLTs in two Schemes

The FLTs discussed in section 4.2 and 4.3 covered two different schemes: Scheme A dealt with divisible statistics such that

$$|g_{ni}(\nu_{ni}, np_{ni})| < be^{a\nu_{ni}} \quad (4.37)$$

in nonclassical multinomial models with all asymptotically Poissonian frequencies (all  $np_{ni}$  bounded from above); and scheme B considered divisible statistics with

$$|h_{ni}(Y_{ni}, np_{ni})| < be^{a|Y_{ni}|} \quad (4.38)$$

in nonclassical multinomial models with mixed asymptotically Poissonian and Gaussian frequencies (but all  $np_{ni}$  bounded from below).

On the one hand, these two schemes complement each other. For models in which  $np_{ni}$  can diverge to infinity but is bounded from below<sup>2</sup>, some divisible statistics like  $\sum_{i=1}^N (\nu_{ni} - np_{ni})^2$  may diverge. Scheme B is appropriate and we should choose divisible statistics satisfying (4.38), say the Chi-square statistic. While for models in which  $np_{ni}$  is bounded from above but can tends to 0, the Chi-square statistic is not applicable<sup>3</sup>, and we have to use scheme A and choose a divisible statistic satisfying (4.37) such as  $\sum_{i=1}^N (\nu_{ni} - np_{ni})^2$  or  $\sum_{i=1}^N |\nu_{ni} - np_{ni}|$ .

On the other hand, there is some overlap between these two schemes and they coincide to each other in such models. Consider models with  $d_n \rightarrow d$  and all frequencies bounded from both above and below:

$$0 < \delta = \liminf_n \inf_i (np_{ni}) < \limsup_n \sup_i (np_{ni}) = \Lambda < \infty. \quad (4.39)$$

---

<sup>2</sup>Linguistic data is an obvious example.

<sup>3</sup>For instance, if we do a GOF test for a beta distribution with both parameters greater than 1, and  $N$  increases proportionally as  $n$ , then some of the  $np_{ni}$  tend to 0, when  $i$  is close to 1 or  $N$ .

It is not hard to see that (4.37) and (4.38) are equivalent under this condition.

Further,

$$g_s(Z(s), \lambda(s)) = h_s(Y(s), \lambda(s))$$

implies that  $\sigma^2(s)$  defined in the theorems 4.5 and 4.7 are equivalent. And the following equality

$$k(s) = \frac{1}{\sqrt{d}} \mathbb{E} [g_s(Z(s), \lambda(s)) (Z(s) - \lambda(s))] = \mathbb{E} [h_s(Y(s), \lambda(s)) Y(s)] \sqrt{f(s)}$$

shows the equivalence of the theorems 4.6 and 4.8.



## Chapter 5

# Properties of Limiting Processes and Applications

Apart from  $\lambda(s)$  and  $f(s)$ , which represent the nature of the non-classical models, the characteristics of the limiting processes can be fully described by two functions  $\sigma^2(s)$  and  $k(s)$ , which depend on the structure of the divisible statistics. Based on these characteristics for several examples of limiting processes of divisible statistics, we can discuss some properties of the limiting processes.

Another concern in this chapter is the application of the limiting processes. We will show that two processes, a martingale component  $W_n$  and a 1-to-1 transformation of the partial sum processes  $\tilde{W}_n$ , can both be used to construct distribution-free statistics for GOF tests. The convergence of these two processes and the statistics based on them will be evaluated and the properties of these statistics will be discussed.

## 5.1 Characteristics of Limiting Processes

The limiting processes for  $W_n$ ,  $K_n$  and  $X_n$  have the same structure for all divisible statistics, except that two functions  $\sigma^2(s)$  and  $k(s)$  depend on the form of the divisible statistics. Therefore,  $\sigma^2(s)$  and  $k(s)$  can be regarded as the characteristics of the limiting processes, and to describe the limit processes, it is sufficient to specify these two functions.

We will show in this section, that the characteristics for some divisible statistics are simple and explicit. Even when the characteristics do not have a convenient form, they can still easily be calculated.

### 5.1.1 Chi-square Process

Like its counterpart (Pearson's Chi-square statistic) in the classical model, the limiting distribution of the Chi-square process

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (Y_{ni}^2 - \mathbb{E}[Y_{ni}^2]) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (Y_{ni}^2 - (1 - p_{ni}))$$

has a simple explicit expression.

#### Characteristic of the martingale component

Recall that for  $\lambda(s) < \infty$

$$Y(s) = \frac{Z(s) - \lambda(s)}{\sqrt{\lambda(s)}} \quad \text{with} \quad Z(s) \sim \text{Poi}(\lambda(s)),$$

so that we have

$$\mathbb{E}Y^2(s) = \mathbb{E} \left[ \frac{(Z(s) - \lambda(s))^2}{\lambda(s)} \right] = 1$$

and

$$\mathbb{E}Y^4(s) = \mathbb{E} \left[ \frac{(Z(s) - \lambda(s))^4}{\lambda^2(s)} \right] = 3 + \frac{1}{\lambda(s)}.$$



Hence

$$\sigma^2(s) = \text{var} [Y^2(s)] = 2 + \frac{1}{\lambda(s)}.$$

On the other hand,  $Y(s) \sim \mathcal{N}(0, 1)$  for  $\lambda(s) = \infty$ , so that

$$\sigma^2(s) = \text{var} [Y^2(s)] = 2.$$

### Characteristic of the compensator component

Since

$$k_{ni} = \sqrt{Np_{ni}} \mathbb{E} Y_{ni}^3 = \frac{1}{\sqrt{d_n}} \mathbb{E} \frac{(\nu_{ni} - np_{ni})^3}{np_{ni}} = \frac{1}{\sqrt{d_n}} (1 - p_{ni})(1 - 2p_{ni}),$$

we have

$$k(s) = \lim_{n \rightarrow \infty} k_{n[Ns]} = \frac{1}{\sqrt{d}}.$$

**Remark 5.1.** Note that as  $d_n \rightarrow \infty$ , for the Chi-square process,  $k(s) = 0$  implies  $K(t) = 0$  and

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (Y_{ni}^2 - (1 - p_{ni})) \xrightarrow{d} w(\tau(t))$$

*This is an example of when the dependence of summands in the partial sum process is asymptotically negligible.*

### 5.1.2 Log-likelihood Processes

For partial sum process of log-likelihood statistic,

$$X_n(t) = \frac{2}{\sqrt{N}} \sum_{i=1}^{Nt} \left( \nu_{ni} \log \frac{\nu_{ni}}{np_{ni}} - \mathbb{E} \left[ \nu_{ni} \log \frac{\nu_{ni}}{np_{ni}} \right] \right)$$

the characteristics of the limiting processes can be expressed as follows.

**the martingale component**

For  $\lambda(s) < \infty$ ,

$$\sigma^2(s) = \text{var} \left[ 2Z(s) \log \frac{Z(s)}{\lambda(s)} \right];$$

while for  $\lambda(s) = \infty$ ,  $h_s(Y(s)) = Y(s)^2$  with  $Y(s) \sim \mathcal{N}(0, 1)$  and

$$\sigma^2(s) = 2.$$

**the compensator component**

For  $\lambda(s) < \infty$ ,

$$k(s) = \frac{2}{\sqrt{d}} \mathbb{E} \left[ Z(s)(Z(s) - \lambda(s)) \log \frac{Z(s)}{\lambda(s)} \right];$$

while for  $\lambda(s) = \infty$ ,  $h_s(Y(s)) = Y(s)^2$  with  $Y(s) \sim \mathcal{N}(0, 1)$  and

$$k(s) = \sqrt{f(s)} \mathbb{E} [h_s(Y(s))Y(s)] = 0.$$

**5.1.3 Spectral Statistics Processes**

For partial sums of spectral statistic

$$X_n(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} (\mathbf{I}\{\nu_{ni} = m\} - \mathbf{P}(\nu_{ni} = m)),$$

the limiting processes are also simple.

**Characteristic of the martingale component**

For  $\lambda(s) < \infty$ ,

$$\sigma^2(s) = \frac{(\lambda(s))^m}{m!} e^{-\lambda(s)} \left[ 1 - \frac{(\lambda(s))^m}{m!} e^{-\lambda(s)} \right].$$

For  $\lambda(s) = \infty$ ,  $h_s(Y(s)) = \mathbf{I}\{Y(s) = -\infty\}$  with  $Y(s) \sim \mathcal{N}(0, 1)$  and

$$\sigma^2(s) = \mathbf{P}\{Y(s) = -\infty\} - (\mathbf{P}\{Y(s) = -\infty\})^2 = 0.$$

**Characteristic of the compensator component**

For  $\lambda(s) < \infty$ ,

$$k(s) = \frac{1}{\sqrt{d}}(m - \lambda(s)) \frac{(\lambda(s))^m}{m!} e^{-\lambda(s)};$$

while for  $\lambda(s) = \infty$ ,

$$k_{ni} = \frac{1}{\sqrt{d_n}}(m - np_{ni}) \binom{n}{m} p_{ni}^m (1 - p_{ni})^{n-m}$$

which implies

$$k(s) = \lim_{n \rightarrow \infty} k_{n[Ns]} = 0.$$

**5.2 Some Properties of Limiting Processes****5.2.1 Properties in Models with all Gaussian Frequencies**

Theorems 4.7 and 4.8 do not require  $n \sim N$ . Therefore they can apply to the model where  $n \gg N$  such that all frequencies are asymptotically Gaussian. The limit theorems for this model exhibit interesting properties for some divisible statistics.

**Asymptotic Equivalence to Chi-square Process**

One remarkable property is that, in this model, the partial sum processes of those divisible statistics which are asymptotically equivalent to the Chi-square statistics, are asymptotically equivalent to the Chi-square process. As we have seen in Log-likelihood process, when all  $\lambda_{ni} = np_{ni} \rightarrow \infty$ , the characteristics are the same as those of the Chi-square process.

In fact, for those statistics which satisfy the conditions stated in Theorem 2.5, p. 26, it is not difficult to verify the asymptotic equivalence of the corresponding partial sum process to the Chi-square process.

### Asymptotic Negligibility of the compensator component

As we said, the limit of the compensator component of Chi-square process vanishes in this model. However, the Chi-square process is not the only one which possesses this property. For divisible statistics with  $h_{ni}(y) = |y|^m$  ( $m$  being positive integers), the compensator components are also asymptotically negligible.

### 5.2.2 Transformation of Partial Sum Processes

Khmaladze (1984) has showed that based on the following lemma, one can construct an 1-to-1 mapping of the processes  $X_n$  into a process converging in distribution to a Wiener process.

**Lemma 5.1** (Khmaladze (1984)). *The process  $X$  is a diffusion type process with stochastic differential*

$$dX(t) = -\frac{k(t)}{1 - \rho(t)} \int_0^t \frac{k(s)}{\sigma^2(s)} dX(s) dt + d\tilde{W}(t),$$

where  $\tilde{W}$  is a Wiener process with respect to time  $\tau$  and

$$r(t) = \frac{k^2(s)}{\sigma^2(s)}, \quad \rho(t) = \int_0^t r(s) ds.$$

The correspondence between the trajectories of  $X$  and  $\tilde{W}$  is 1-to-1.

The transformed process  $\tilde{W}_n$  with increment

$$\Delta \tilde{W}_n \left( \frac{i}{N} \right) = \Delta X_n \left( \frac{i}{N} \right) + \frac{k(i/N)}{N - \sum_{j \leq i} r(i/N)} \sum_{j \leq i} \left( \frac{k(j/N)}{\sigma^2(j/N)} \Delta X_n \left( \frac{j}{N} \right) \right), \quad (5.1)$$

converges in distribution to  $\tilde{W}$ .

Due to the 1-to-1 property, distinguishing between null and alternative hypotheses based on the partial sum process  $X_n$  and the transformed processes  $\tilde{W}_n$  are equivalent, while the limiting distribution of  $\tilde{W}_n$  is much simpler and easy to compute.

By the same logic as that in Remark 4.2, we can see that this transformation also applies to the scheme of models with mixed frequencies.

### 5.3 GOF tests based on Limiting Processes

Apart from the transformed process  $\tilde{W}_n$ , the martingale component  $W_n$  also converges in distribution to a Brownian motion with respect to time  $\tau$ . The computation of  $\tilde{W}_n$  is a bit simpler than that of  $W_n$ , while the conditions needed for FLT of the martingale component are less onerous than that of the partial sum process. Since the extra computational effort needed for  $W_n$  and the extra conditions needed for  $\tilde{W}_n$  are both insignificant, we are free to choose either one as the basis for constructing testing statistics.

#### 5.3.1 Evaluation of convergence of $\tilde{W}_n$ and $W_n$

As remarked in Haywood and Khmaladze (2008), “it is not entirely clear how to best evaluate the convergence of distribution of a sequence of processes to limiting process.” Therefore, we need to consider several statistics which describe different feature of the processes to access the difference between  $\tilde{W}_n$ ,  $W_n$  and Brownian Motion.

Instead of using  $\tilde{W}_n$  and  $W_n$ , we consider standardized versions of the processes,

$$\tilde{w}_n(s) = \frac{\tilde{W}_n(t)}{\sqrt{\tau_n(1)}} \quad \text{and} \quad w_n(s) = \frac{W_n(t)}{\sqrt{\tau_n(1)}},$$

with  $s = \tau_n(t)/\tau_n(1)$ . Then both  $\tilde{w}_n$  and  $w_n$  converge in distribution to standard Brownian motion  $w$ . And the following statistics can be constructed: one-sided and two-sided Kolmogorov-Smirnov statistics,

$${}^w d_n^+ = \sup_{0 < s < 1} w_n(s), \quad {}^w d_n^- = - \inf_{0 < s < 1} w_n(s), \quad {}^w d_n = \sup_{0 < s < 1} |w_n(s)|;$$

the Cramér-von Mises statistic,

$${}^w \omega_n^2 = \int_0^1 w_n^2(s) ds;$$

and the Anderson-Darling statistic,

$${}^w A_n^2 = \int_0^1 \frac{w_n^2(s)}{s} ds.$$

We can also consider a further transform from  $\tilde{w}_n$  and  $w_n$  to

$$\tilde{v}_n(s) = \tilde{w}_n(s) - s\tilde{w}_n(1) \quad \text{and} \quad v_n(s) = w_n(s) - sw_n(1),$$

then  $\tilde{v}_n$  and  $v_n$  converge in distribution to standard Brownian bridge  $v$ . And a set of statistics “with respect to Brownian bridge” can be constructed: Kolmogorov-Smirnov statistics,

$${}^v d_n^+ = \sup_{0 < s < 1} v_n(s), \quad {}^v d_n^- = - \inf_{0 < s < 1} v_n(s), \quad {}^v d_n = \sup_{0 < s < 1} |v_n(s)|;$$

the Cramér-von Mises statistic,

$${}^v \omega_n^2 = \int_0^1 v_n^2(s) ds;$$

and the Anderson-Darling statistic,

$${}^v A_n^2 = \int_0^1 \frac{v_n^2(s)}{s(1-s)} ds.$$

**Remark 5.2.** *The transformation from Brownian motion to Brownian bridge bears some risk of losing testing power. Suppose when the alternative hypothesis is true, the limiting distribution of  $w_n(s)$  is  $w(s) + l(S)$  with  $l(s) = s$ ,*

instead of  $w(s)$  when null hypothesis is true. Then the statistics based on Brownian motion, such as  ${}^w d_n$  can distinguish the alternative properly. However, the limiting distribution of  $v_n(s)$  is  $(w(s) + l(s)) - s(w(1) + l(1)) = v(s)$  under both null and alternative. The shift  $l(t) = t$  has been offset by the transformation, and the statistics based on  $v_n(s)$  has no power against this alternative. However, it is worth pointing out that this is a special situation.

The limiting distribution functions of the above statistics are all available. For  ${}^w d_n^+$  and  ${}^w d_n^-$ , the limit distribution follows from the reflection principle and can be seen from e.g., Feller (1970). The limiting distribution functions of  ${}^w d_n$ ,  ${}^v d_n^+$ ,  ${}^v d_n^-$  and  ${}^v d_n$  can be found from Shorack and Wellner (1986). The table of the limit distribution of  ${}^w \omega_n^2$  was presented in Orlov (1973), and analytical expression of  ${}^w \omega_n^2$ ,  ${}^w A_n^2$  and  ${}^v \omega_n^2$ ,  ${}^v A_n^2$  can be obtained from Deheuvels and Martynov (2003). Shorack and Wellner (1986) also provided expressions of distributions  ${}^v \omega_n^2$  and  ${}^v A_n^2$  in terms of the distribution of weighted sums of i.i.d. Chi-square random variables.

Figure 5.2-5.5 compared  ${}^w d_n^+$ ,  ${}^w d_n^-$ ,  ${}^w d_n$ ,  ${}^v d_n^+$ ,  ${}^v d_n^-$ ,  ${}^v d_n$ ,  ${}^w \omega_n^2$  and  ${}^v A_n^2$ , based on  $W_n$  and  $\tilde{W}_n$ , to their limiting distributions. Figure 5.2 and 5.3 are based on scheme A with divisible statistic being  $\sum_{i=1}^N (\nu_{ni} - np_{ni})^2$  and probabilities generated from Beta(2,4) (bounded from above, see Figure 5.1); Figure 5.4 and 5.5 are based on scheme B with divisible statistic being Chi-square and probabilities generated from Beta(0.5,0.5) (bounded from below).

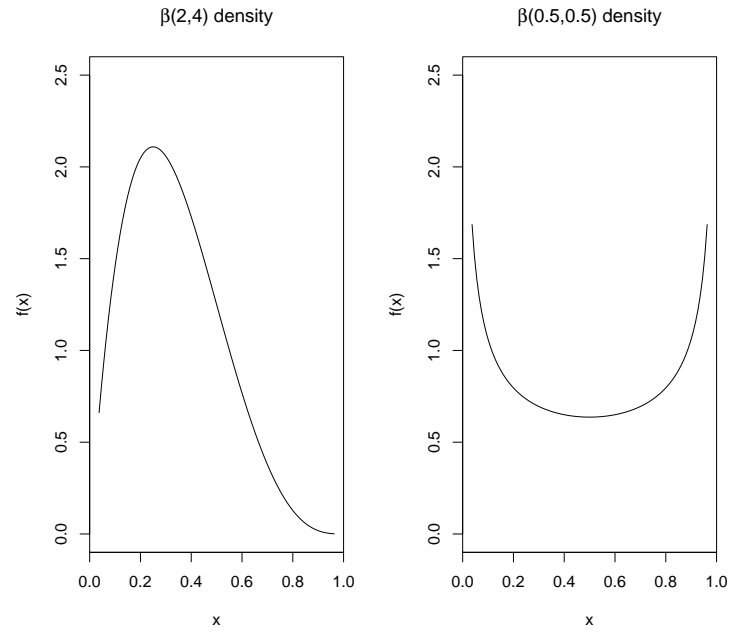


Figure 5.1: Density function of Beta(2,4) and Beta(0.5,0.5).

The results of Figure 5.2-5.5 shows that, for both scheme, the empirical distributions of statistics are very close to their limiting distributions, and the difference between statistics base on  $W_n$  and that based on  $\tilde{W}_n$  are very small. Therefore, the convergence of both  $W_n$  and  $\tilde{W}_n$  are quick enough and they can be good candidates for constructing GOF test statistics.



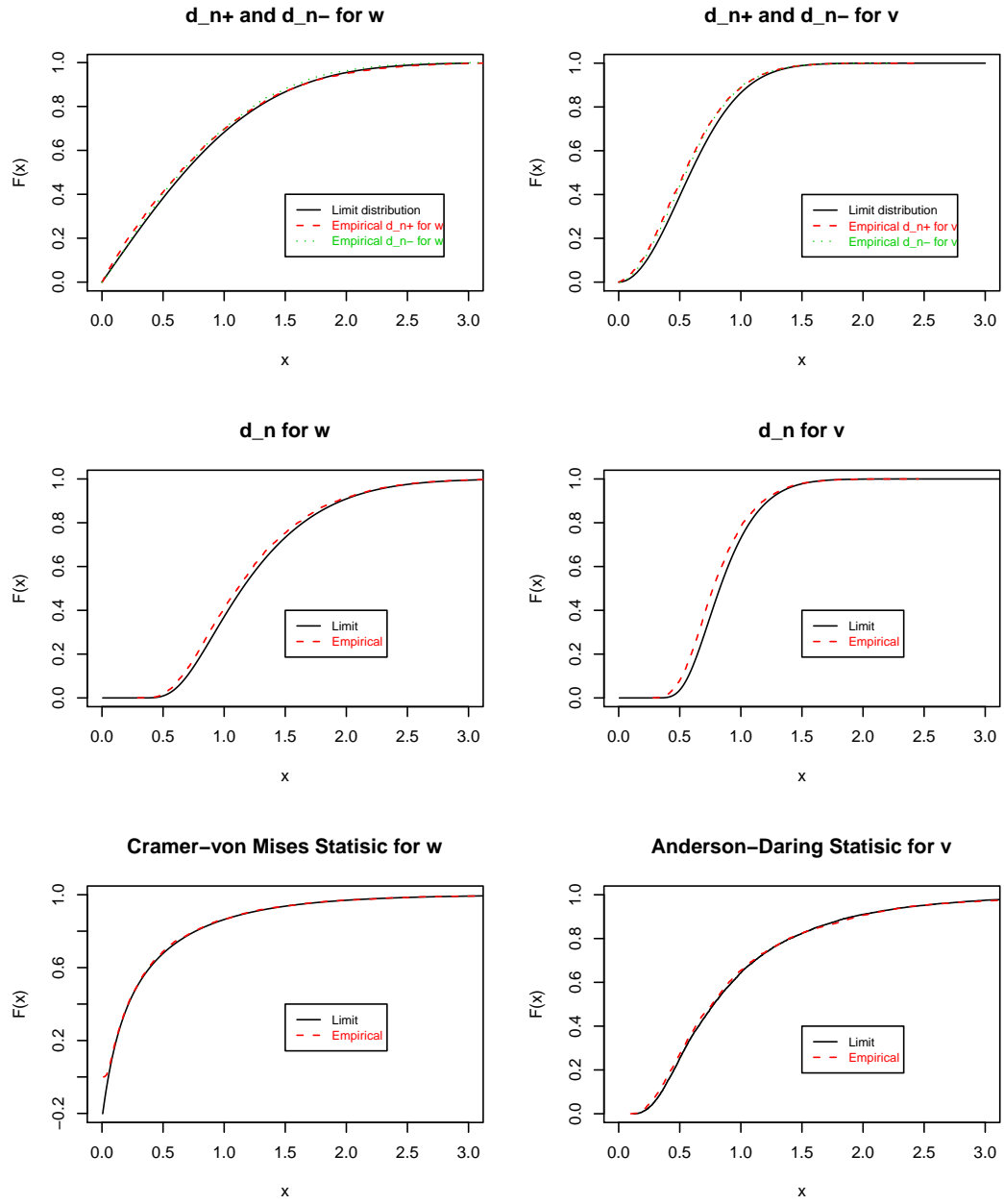


Figure 5.2: Evaluation of Convergence of  $W_n$  in Scheme A:  $n=N=1000$ ; Probabilities generated from Beta(2,4) by equally partitioning;  $W_n$  being the martingale component of partial sum process of  $\sum_{i=1}^N (\nu_{ni} - np_{ni})^2$ ; 5000 replicates.

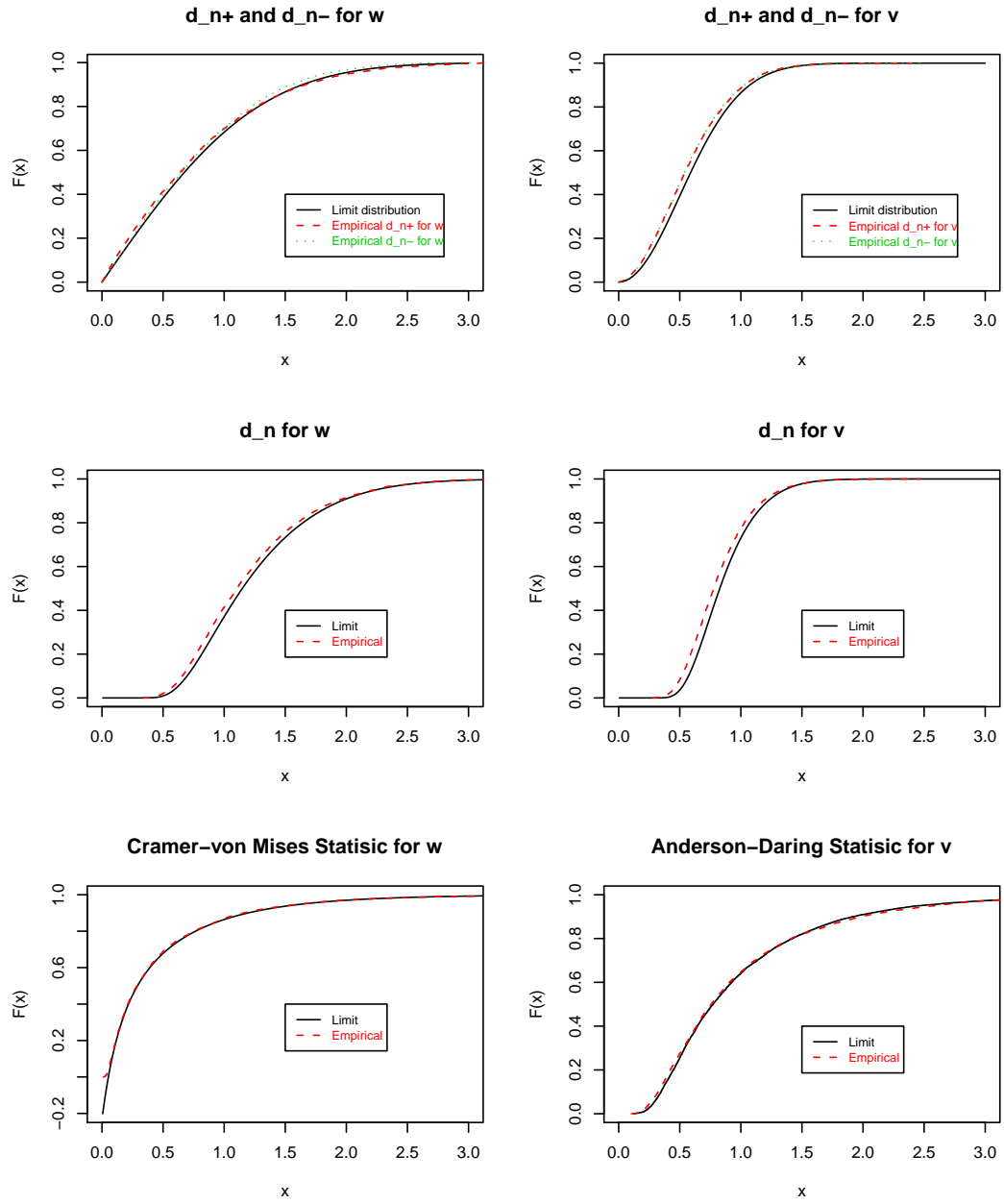


Figure 5.3: Evaluation of Convergence of  $\tilde{W}_n$  in Scheme A:  $n=N=1000$ ; Probabilities generated from Beta(2,4) by equally partitioning;  $\tilde{W}_n$  is the transformed process of partial sum process of  $\sum_{i=1}^N (\nu_{ni} - np_{ni})^2$ ; 5000 replicates.

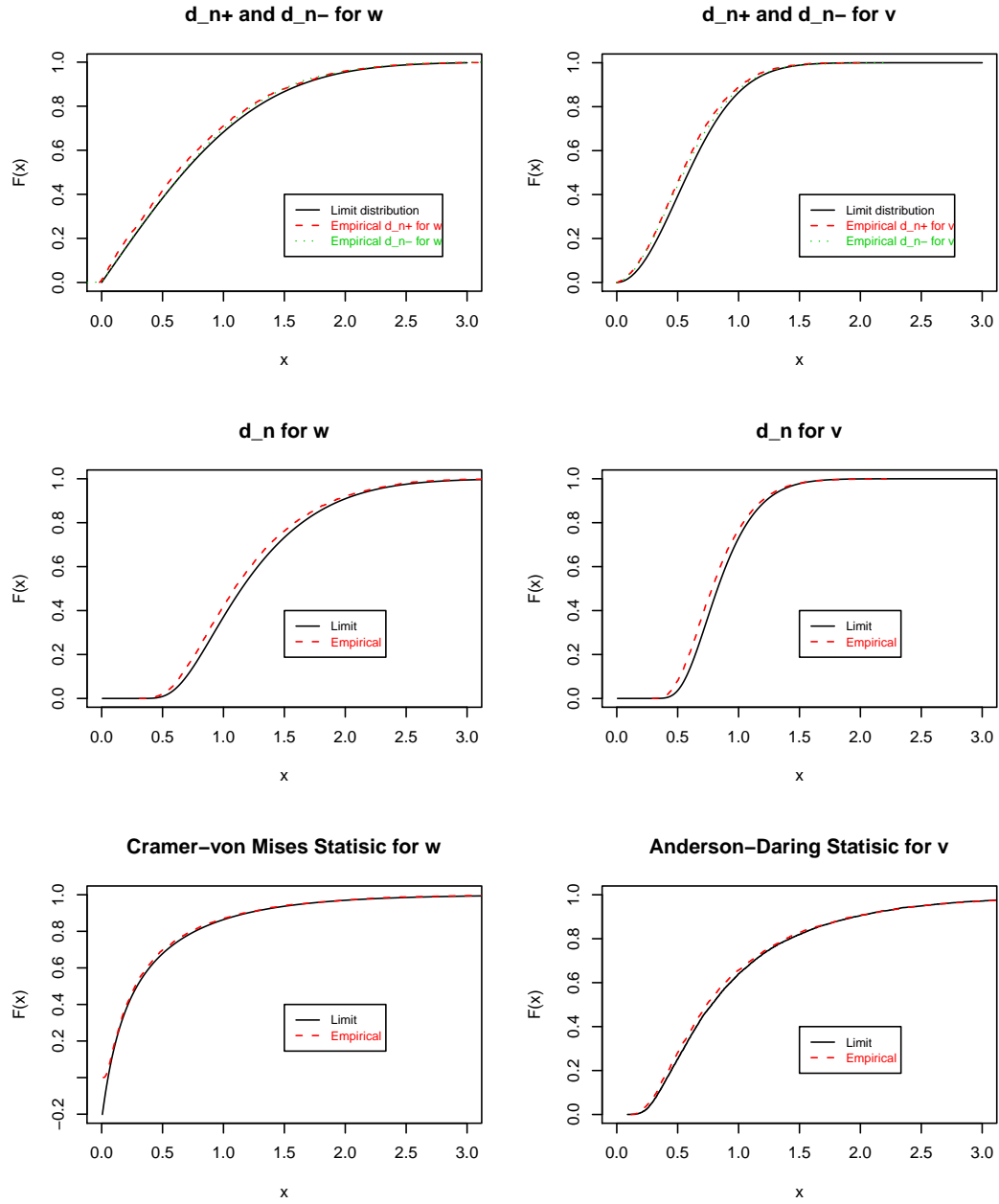


Figure 5.4: Evaluation of Convergence of  $W_n$  in Scheme B:  $n=N=1000$ ; Probabilities generated from Beta(0.5,0.5) by equally partitioning;  $W_n$  being the martingale component of Chi-square process; 5000 replicates.

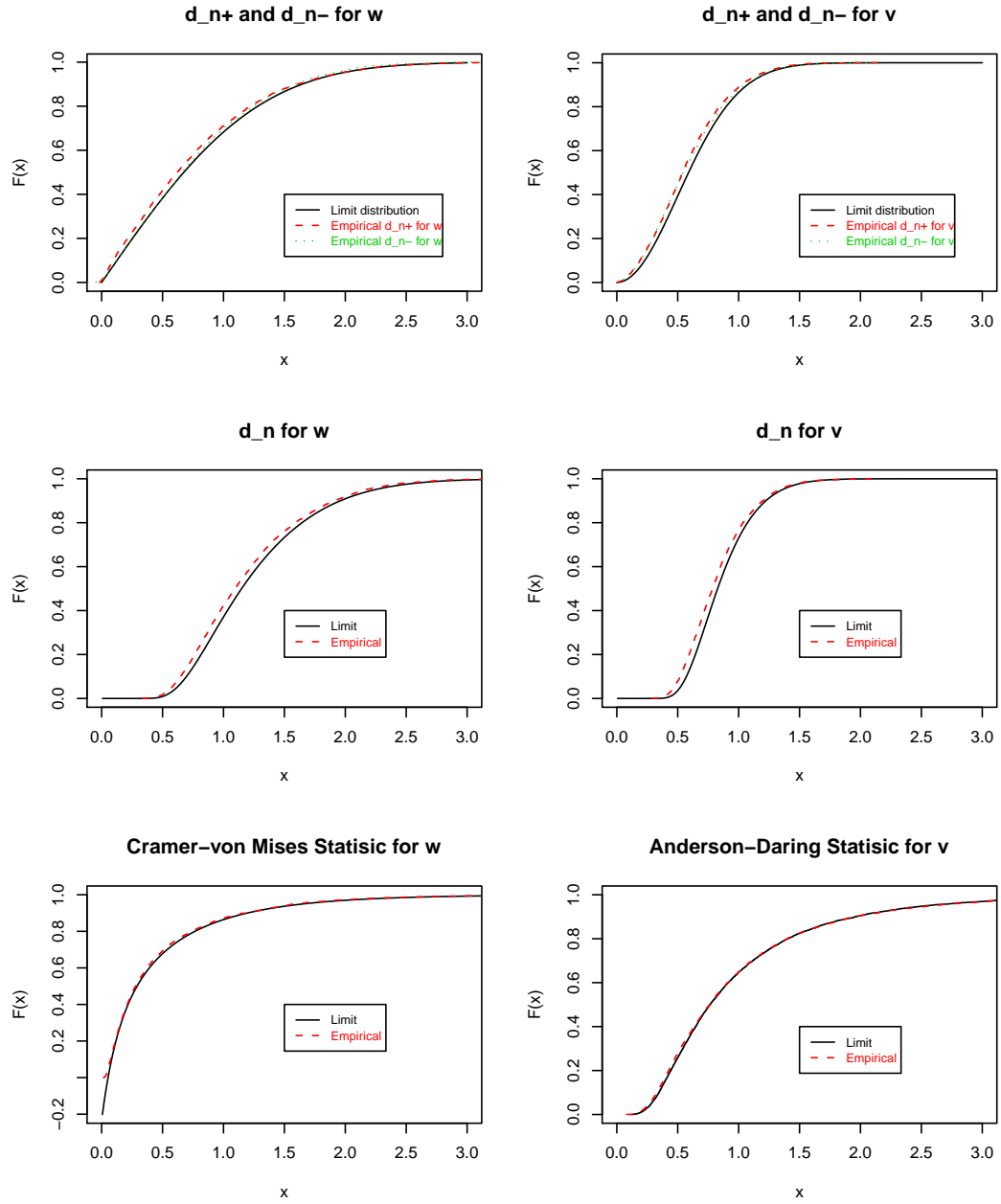


Figure 5.5: Evaluation of Convergence of  $\tilde{W}_n$  in Scheme B:  $n=N=1000$ ; Probabilities generated from Beta(0.5,0.5) by equally partitioning;  $\tilde{W}_n$  is the transformed process of Chi-square process; 5000 replicates.

We can see further in Figure 5.6 that, for divisible statistics with  $h_{ni} = |Y_{ni}|$ ,  $|Y_{ni}|^2$  and  $|Y_{ni}|^3$  respectively, the empirical distributions of Cramér-von Mises statistic  ${}^w\omega_n^2$  based on  $\tilde{W}_n$  or  $W_n$  are very close. In other words, the rate of convergence of  $\tilde{W}_n$  or  $W_n$  to Brownian motion is good for many different  $h_{ni}$ .

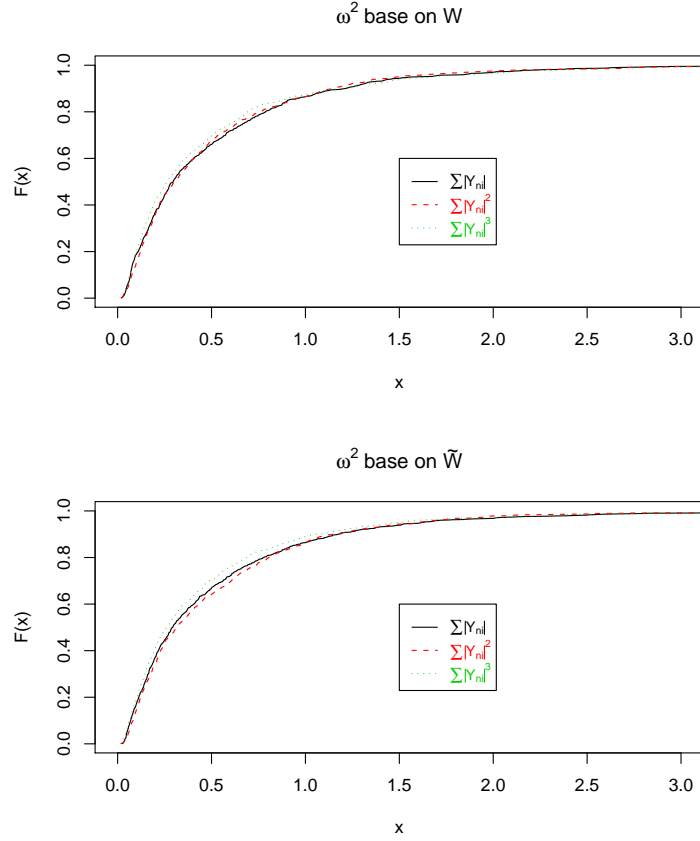


Figure 5.6: Convergence Rate of  $\tilde{W}_n$ ,  $W_n$  for Different Divisible Statistic:  $n = N = 1000$ ; Probabilities generated from beta (0.1,0.9) by equally partitioning; Empirical distribution of  ${}^w\omega_n^2$  based on  $\tilde{W}_n$ ,  $W_n$  of divisible statistics with  $h_{ni} = |Y_{ni}|$ ,  $|Y_{ni}|^2$  and  $|Y_{ni}|^3$  respectively; 1000 replicates.

### 5.3.2 Properties of Statistics based on $\tilde{W}_n$ and $W_n$

One could think that, in the move from the classical multinomial models to non-classical models, the distribution-free will be lost and the distribution of the divisible statistics do depends on the hypothetic probabilities. However, the statistics based on the transformed partial sum process  $\tilde{W}_n$  or the martingale component  $W_n$  are again distribution-free. Furthermore, for those divisible statistic with non-degenerate limiting processes, the original forms of the divisible statistic are even irrelevant. These are also highly desirable properties, since we can carry out statistical tests without calculating many different limiting distributions.

#### Distribution-Free

Figure 5.4 shows that the empirical distributions of Kolmogorov-Smirnov statistics  $^w d_n$  based on Chi-square processes are independent of the probabilities, which are generated from Beta(0.1,0.9), Beta(0.5,0.5) and Beta(0.9,0.1) respectively, by equally partitioning. We can see that in both  $\tilde{W}_n$  and  $W_n$  cases, the empirical distributions are almost indistinguishable.

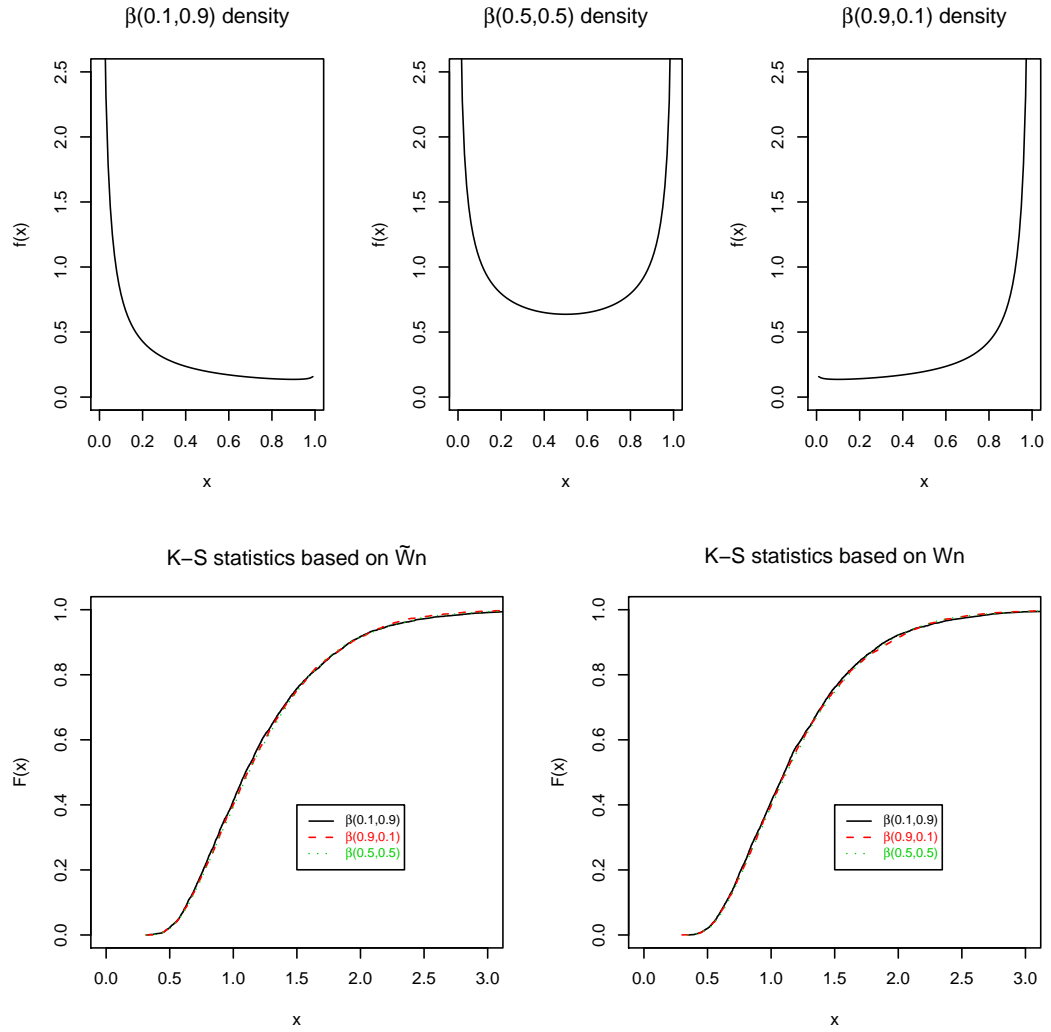


Figure 5.7: Distribution-free Property of Statistics Based on  $\tilde{W}_n$  and  $W_n$ :  $n = N = 1000$ ; Probabilities generated from Beta(0.1,0.9), Beta(0.5,0.5), Beta(0.9,0.1) by equally partitioning;  $\tilde{W}_n$ ,  $W_n$  are the transformed process and the martingale component of Chi-square process respectively; 5000 replicates.

### Power in Detecting some Contiguous Alternatives

The move from the divisible statistics to corresponding partial sum processes, not only provide us an alternative approach to limit theorems of the divisible statistics, but also reveal much richer information. For example, we cannot use divisible statistics to distinguish between null hypothesis to some contiguous alternative, while the statistics based on  $\tilde{W}_n$  or  $W_n$  can fulfil this mission. This is also justify the necessity of this moving.

The tests based on divisible statistic are equivalent to tests based on  $X_n(1)$ . However, for those contiguous alternatives with  $m(1) = 0$ , the limiting distributions of  $X_n(1)$  are the same under both null and alternative. Recall that

$$m(t) = -\mathcal{K}c = \int_0^1 k(s)c(s)ds,$$

any alternative satisfying  $\int_0^1 k(s)c(s)ds = 0$  can not be distinguished from null by test based on  $X_n(1)$ .

We will show such an example in Figure 5.8. For simplicity, we choose  $n = N$  and Chi-square process such that  $k(s) = 1$ . Therefore, if let  $c(s) = 3 \cos(\pi s)$ , then

$$m(t) = \int_0^t k(s)c(s) = \frac{3}{\pi} \sin(\pi t)$$

implies  $m(1) = 0$ . We then compare the empirical distributions of  $X_n(1)$  under null and alternative. It shows that they are indistinguishable. While for the empirical distributions of statistics based on  $\tilde{W}_n$  or  $W_n$ , such as  ${}^w d_n$  and  ${}^w \omega_n$ , the deviation of the empirical distributions under alternative from that under null and from limiting distributions are both significant.



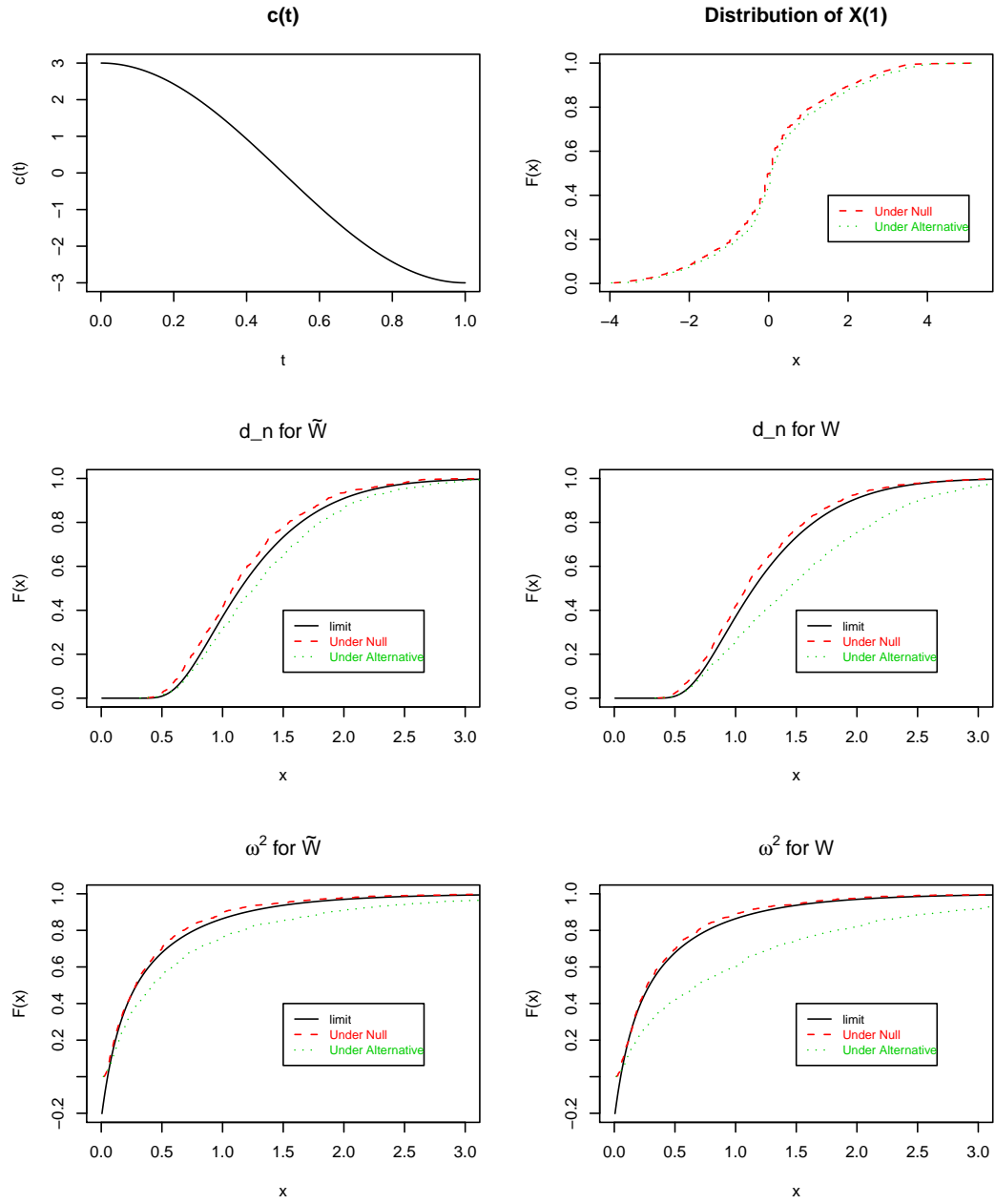


Figure 5.8: Power of statistics based on  $\tilde{W}_n$  and  $W_n$  in Detecting some Contiguous Alternatives:  $p_{ni} = 1/N$  under Null;  $\bar{p}_{ni} = p_{ni}(1 + c(i/N)/\sqrt{n})$  under alternative;  $n = N = 1000$ ; Empirical distribution of  $^w d_n$  and  $^w \omega_n^2$  based on  $\tilde{W}_n$ ,  $W_n$  of Chi-square process; 1000 replicates.



## Chapter 6

# Asymptotics in the LNRE models

Most non-classical multinomial models belong to the so-called LNRE model, where many frequencies are relatively small. Apart from the model with all Gaussian frequencies, other models discussed in chapters 4 and 5 are all LNRE models.

The spectral statistics  $\mu_n(m)$  and empirical vocabulary  $\mu_n$ , which were introduced in section (1.2.7) play a vital role in the discussion of asymptotics in the LNRE models. One reason is that the ratios

$$\frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n} \quad \text{for } m = 1, 2, \dots$$

usually specify different LNRE models. Besides, the symmetric divisible statistics, such as the ML estimates of diversity measures, can be expressed as a linear combination of the spectral statistics

$$\sum_{i=1}^N g_n(\nu_{ni}) = \sum_{m=1}^n g_n(m)\mu_n(m).$$

The asymptotics of these symmetric divisible statistics is therefore naturally related to the asymptotic behaviour of spectral statistics.

In the models discussed in chapters 4 and 5, we assume the density  $\lambda(t) = d \times f(t)$  is a continuous function. Hence when  $n \sim N$ ,

$$\frac{\mathbb{E}\mu_n(m)}{n} = \frac{1}{d_n N} \sum_{i=1}^N \mathbf{P}(\nu_{ni} = m) \rightarrow \frac{1}{d} \int_0^1 \frac{\lambda^m(s)}{m!} e^{-\lambda(s)} ds > 0,$$

such that both (d1) and (d2) conditions of LNRE model are satisfied. While for models with density concentrated at countable many points in  $[0,1]$ ,

$$\frac{\mathbb{E}\mu_n(1)}{n} \rightarrow 0,$$

and the tools developed in chapters 4 and 5 are not applicable. This is the case where (d2) are satisfied but not (d1).

In this chapter, apart from introducing the general framework and models of LNRE in section 6.1, we focus on a particular model within the context of multiple-choice questionnaires in section 6.2. We will establish limit theorems and show that this is the case that (d2) are satisfied but not (d1).

## 6.1 Introduction to LNRE theory

Apart from linguistic data (see Chapter 3), there are plenty of statistical data having the LNRE properties. For example, if we are interested in the connectivity of autonomous systems (AS) which compose the global Internet routing system, then most of them are only connected to a few of the other systems, and the number of connections of these systems are well described by an LNRE model.

number of connections	1	2	3-10	>10
number of ASs	12184	12929	6679	1534

Table 6.1: The Connectivity of Global Internet Routing System(CAIDA AS Relationships Dataset, 2009/12/15).

The LNRE theory can be traced back to the 1980s, when the formal statistical analysis of LNRE was established by Khmaladze (1988), in which Khmaladze introduced the notions, studied its different forms and found various necessary and sufficient conditions for so-called (d1) and (d2) zone.

### 6.1.1 Definition of LNRE

Recall the definition of spectral statistics

$$\mu_n(m) = \sum_{i=1}^N \mathbf{I}\{\nu_{ni} = m\}, \quad m = 1, 2, \dots$$

and empirical vocabulary

$$\mu_n = \sum_{i=1}^N \mathbf{I}\{\nu_{ni} > 0\},$$

introduced in Section 1.2.7.

**Definition 6.1.** *A sequence of random vectors  $\{\boldsymbol{\nu}_n\}_{n \geq 1}$  forms a sequence of large number of rare events if*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}\mu_n(1)}{n} > 0. \quad (\text{d1})$$

**Definition 6.2.** *A sequence of random vectors  $\{\boldsymbol{\nu}_n\}_{n \geq 1}$  forms a sequence of large number of rare events if*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}\mu_n(1)}{\mathbb{E}\mu_n} > 0, \quad \lim_{n \rightarrow \infty} \mathbb{E}\mu_n = \infty. \quad (\text{d2})$$

For convenience, we will say that we are in  $(d1)$  or  $(d2)$  zone of LNRE if  $(d1)$  or  $(d2)$  is satisfied respectively. These two definitions are not equivalent: namely  $(d1) \Rightarrow (d2)$ , but not vice versa.

It is easy to observe that, in classical multinomial models with fixed  $N$  and vector of probabilities  $\mathbf{p}$  each frequency  $\nu_{ni} \rightarrow \infty$  a.s. as  $n \rightarrow \infty$ , and therefore  $\mu_n(1) \rightarrow 0$  and  $\mu_n \rightarrow N$  a.s. Consequently  $(d1)$  and  $(d2)$  both cannot be satisfied.

**Remark 6.1.** *As Khmaladze (1988) indicates, some other definitions also may correspond to the intuitive understanding of the expression “large number of rare events”. For example  $\{\nu_n\}$  could be called an LNRE sequence if*

$$\lim_{n \rightarrow \infty} \mathbb{E}\mu_n(1) = \infty,$$

*or if*

$$\lim_{n \rightarrow \infty} \mathbb{E}\mu_n(1) > 0,$$

*or if*

$$\lim_{n \rightarrow \infty} \mathbb{E}\mu_n = \infty.$$

However, these definitions are less interesting than Definition 6.1 and 6.2, since the ratios<sup>1</sup>

$$\alpha_n(m) = \frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n}$$

play key roles in classifying the LNRE models.

### 6.1.2 Law of Large Numbers of Spectral Statistics

The definition of the LNRE models involves only  $\mathbb{E}\mu_n(m)$  and  $\mathbb{E}\mu_n$ , while in empirical data, only  $\mu_n(m)$  and  $\mu_n$  are available. Therefore, we need to firstly establish the connection between  $\mu_n(m)$  and  $\mu_n$  and their expectations. The

---

<sup>1</sup>Called relative spectrum elements in Baayen (2001).

following theorem achieves this target.

**Theorem 6.1.** *If for every  $m = 1, 2, \dots$ ,  $\mathbb{E}\mu_n(m)$  and  $\mathbb{E}\mu_n \geq a(\ln n)^2$  for sufficiently large  $n$ , then*

$$\frac{\mu_n}{\mathbb{E}\mu_n} \rightarrow 1 \text{ a.s.}$$

and

$$\frac{\mu_n(m)}{\mathbb{E}\mu_n(m)} \rightarrow 1 \text{ a.s.}$$

*Proof.* See Theorem 6 of Kvizhinadze (2010). □

**Remark 6.2.** *It is necessary to point out that this condition is not difficult to satisfy. Kvizhinadze (2010) has shown, in the LNRE model which will be discussed in Section 6.2, that  $\mathbb{E}\mu_n(m)$  and  $\mathbb{E}\mu_n$  are actually of the same order of  $N^u$  with  $0 < u < 1$ , which is much stronger than what is required in the theorem.*

In this thesis, we assume this condition is satisfied and hence we are free to exchange  $\mu_n(m)$ ,  $\mu_n$  and  $\mathbb{E}\mu_n(m)$ ,  $\mathbb{E}\mu_n$ .

### 6.1.3 Models of LNRE

The classification of the LNRE models were based on the limits of the ratio of the spectral statistics to the empirical vocabulary

$$\frac{\mu_n(m)}{\mu_n},$$

or to the relative spectrum elements

$$\frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n}.$$

There is no significant difference between these two as long as the condition required in Theorem 6.1 is satisfied. However, we prefer the latter, since they exclude the randomness and only reflect the structure of the LNRE models.

The most profound the LNRE model is Zip's Law observed by French stenographer J.B. Estoup and which was systematically studied by the American linguist and philologist G.K. Zipf (1935), which states that the ratio of spectral statistics to empirical vocabulary is approximately

$$\frac{\mu_n(m)}{\mu_n} \approx \frac{1}{m(m+1)}.$$

Other popular LNRE laws include Zipf-Mandelbrot law (Mandelbrot, 1952), Yule law (Yule, 1924; Simon, 1955), Yule-Simon law (Simon, 1960), Karlin-Rouault law (Rouault, 1978).

It was revealed that the above laws can subsumed in the Zipfian family of the LNRE models (Baayen, 2001)<sup>2</sup> and have a general expression (introduced by Orlov and Chitashvili, 1983a,b) in terms of the limit of the relative spectrum elements,

$$\alpha(m, \alpha, \beta, \gamma) = \lim_{n \rightarrow \infty} \alpha_n(m) = \frac{\int_0^\infty \frac{(\log(1+x))^{\gamma-1} x^\alpha}{(1+x)^{m+1} (1+x)^\beta} dx}{\int_0^\infty \frac{(\log(1+x))^{\gamma-1} x^{\alpha-1}}{(1+x)^{\beta+1}} dx}$$

This expression defines a family of models, which reduce to specific laws for specific choices of the parameters  $\alpha$ ,  $\beta$  and  $\gamma$ .

### Zipf

$$\alpha(m, 1, 1, 1) = \frac{1}{m(m+1)}$$

### Zipf-Mandelbrot

$$\alpha(m, 1, 1, \gamma) = \frac{1}{m^\gamma} - \frac{1}{(m+1)^\gamma}, \quad (\gamma > 0)$$

---

<sup>2</sup>Baayen (2001) consider  $\alpha(m, N, \alpha, \beta, \gamma)$  instead of  $\alpha(m, \alpha, \beta, \gamma)$ . But this is incorrect, since the right side of the expression does not depend on  $N$ .



**Yule**

$$\alpha(m, \beta, \beta, 1) = \frac{\Gamma(\beta + 1)\Gamma(m)\beta}{\Gamma(m + \beta + 1)}$$

**Yule-Simon**

$$\alpha(m, 1, \beta, 1) = \frac{\beta}{(m + \beta - 1)(m + \beta)}, \quad (\beta > 0)$$

**Waring-Herdan-Muller(Herdan,1960,1964;Muller,1979)**

$$\alpha(m, \alpha, \beta, 1) = \frac{\Gamma(\beta + 1)\alpha}{\Gamma(\beta + 1 - \alpha)} \cdot \frac{\Gamma(m + \beta - \alpha)}{\Gamma(m + \beta + 1)}, \quad (\alpha > 1, \beta \geq \alpha)$$

**Karlin-Rouault**

$$\alpha(m, \alpha, 0, 1) = \frac{\alpha\Gamma(m - \alpha)}{\Gamma(1 - \alpha)\Gamma(m + 1)}, \quad (0 < \alpha < 1)$$

**6.2 LNRE in multiple-choice questionnaires**

In practice there are plenty of examples when the random variable is a  $q$  dimensional vector with coordinates taking different number of values. For example, imagine a system which contains a large number (say, 20) of components and these components can be in various states in any given second, so that the state of a system could be characterized as 20 dimensional vector. Biologists use a sequence of procedures to identify bacteria. During each procedure bacteria is placed in a certain chemical substance where it changes its colour. Looking on these sequence of colour biologists can say which group of bacteria it belongs to. It is obvious that the result of this tests is also an example of multi dimensional random variable. And finally as

a basic example we will take a questionnaire with  $q$  multiple choice answers.

### 6.2.1 The Model

Consider a set of the disjoint events which can be indexed by  $q$ -dimensional vectors  $\vec{x}_q = (x_1, \dots, x_i, \dots, x_q)$  with coordinates  $x_i$  ranging from 1 to  $k_i$  respectively. Then  $p_{ni}$  and  $\nu_{ni}$ ,  $i = 1, \dots, N$ , in previous setting become  $p(\vec{x}_q)$  and  $\nu_n(\vec{x}_q)$ , respectively, with  $\vec{x}_q \in \Xi_q = \times_{i=1}^q \{1, \dots, k_i\}$  and  $N = \prod_{i=1}^q k_i$  being the cardinality of  $\Xi_q$ . Hence  $\mu_n(m)$  becomes,

$$\mu_n(m) = \sum_{\vec{x}_q \in \Xi_q} I\{\nu_n(\vec{x}_q) = m\}.$$

Therefore,

$$\mathbb{E}\mu_n(m) = \sum_{\vec{x}_q \in \Xi_q} \mathbb{P}\{\nu_n(\vec{x}_q) = m\} = \sum_{\vec{x}_q \in \Xi_q} \binom{n}{m} p(\vec{x}_q)^m (1 - p(\vec{x}_q))^{n-m}$$

and

$$\mathbb{E}\mu_n = \sum_{\vec{x}_q \in \Xi_q} (1 - (1 - p(\vec{x}_q))^n).$$

In the context of multiple-choice questionnaires,  $\vec{x}_q$  can be interpreted as an opinion in a questionnaire with  $q$  multi-option or multi-choice questions (the  $i$ -th question has  $k_i$  options). And the ratios

$$\frac{\mathbb{E}\mu_n(m)}{n} \quad \text{and} \quad \frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n} \tag{6.1}$$

can be interpreted as: the proportion of the number of opinions with  $m$  supporters in all  $n$  responses; and the total number of opinions with at least 1 supporter, respectively.

The main setting of the framework was given in Khmaladze (2009). In that paper all  $x_i$  were binary. However, in this section, we want to take advantage

of the fact that the proofs given in Khmaladze (2009) are of a more general nature. We demonstrate this by extending settings for questionnaires with a general structure.

### 6.2.2 Preliminary Discussion

#### The relation among $n$ , $q$ and $N$

There are three variables  $n$ ,  $q$  and  $N$ , which control the asymptotic behavior of the ratios. Among them,  $q$  and  $N$  are directly associated with each other by the definition of  $N$ . Therefore, it is sufficient that we discuss the relation between  $n$  and  $N$ .

Since  $N$  is the number of disjoint events (opinions) and  $n$  is sample size (the number of responses), when  $n = o(N)$ , most frequencies tend to zero and those nonzero frequencies will mostly be 1. On the other hand, in the situation of  $N = o(n)$ , most frequencies are nonzero and eventually tend to infinity. However, it is more interesting to investigate the situation where  $N$  and  $n$  are comparable, particularly,  $n = \lambda N$  for some constant  $0 < \lambda < \infty$ . In this section, we focus on the last case.

#### Introduction of likelihood ratio $M_q$

Let  $\mathbf{P}_q$  denote the probability measure on  $\Xi_q$  which is defined by probabilities  $p(\vec{x}_q)$ :

$$\mathbf{P}_q(\vec{X}_q = \vec{x}_q) = p(\vec{x}_q)$$

and let  $\mathbf{P}_{0,q}$  denote the uniform measure on  $\Xi_q$ :

$$\mathbf{P}_{0,q}(\vec{X}_q = \vec{x}_q) = p_0(\vec{x}_q) = \frac{1}{N}.$$

Then,

$$\begin{aligned}\mathbb{E}\mu_n(m) &= \sum_{\vec{x}_q \in \Xi_q} \binom{n}{m} p(\vec{x}_q)^m (1 - p(\vec{x}_q))^{n-m} \\ &= N \mathbb{E}_{\mathbf{P}_{0,q}} \left[ \binom{n}{m} p(\vec{X}_q)^m (1 - p(\vec{X}_q))^{n-m} \right]\end{aligned}$$

and

$$\mathbb{E}\mu_n = N \mathbb{E}_{\mathbf{P}_{0,q}} \left[ 1 - (1 - p(\vec{X}_q))^n \right].$$

Define  $M_q$  as likelihood ratio of the alternative measure  $\mathbf{P}_q$  to null measure  $\mathbf{P}_{0,q}$ , i.e.

$$M_q(\vec{x}_q) = \frac{d\mathbf{P}_q}{d\mathbf{P}_{0,q}}(\vec{x}_q) = \frac{p(\vec{x}_q)}{p_0(\vec{x}_q)} = Np(\vec{x}_q). \quad (6.2)$$

Then we have

$$\mathbb{E}\mu_n(m) = N \binom{n}{m} \frac{1}{n^m} \mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m \left( 1 - \frac{\lambda M_q(\vec{X}_q)}{n} \right)^{n-m} \right] \quad (6.3)$$

and

$$\mathbb{E}\mu_n = N \left( 1 - \mathbb{E}_{\mathbf{P}_{0,q}} \left[ \left( 1 - \frac{\lambda M_q(\vec{X}_q)}{n} \right)^n \right] \right). \quad (6.4)$$

At first sight, it looks artificial that we introduce such a likelihood ratio  $M_q$ . However, as it was suggested in Khmaladze (2009) the benefit of this introduction is significant. Although the “physical” measure of  $\vec{X}_q$  is  $\mathbf{P}_q$ , using  $M_q$  we can exploit its asymptotic properties as if  $\vec{X}_q$  had the uniform distribution  $\mathbf{P}_{0,q}$ . As a likelihood ratio and a martingale in  $\mathbf{q}$ ,  $M_q(\vec{X}_q)$  possesses some good and well-known asymptotic properties, which is very convenient.

Further, according to the Lemma 6.1 below, expressions in the right hand side of (6.3) and (6.4) can be replaced by Poissonian limits. This suggests that we can lay aside the role of sample size  $n$  in the asymptotic behaviour of the ratios, and focus on the limiting behaviour of distribution of  $M_q(\vec{X}_q)$ , or

equivalently,  $Np(\vec{X}_q)$ , under the measure  $\mathbf{P}_{0,q}$  (We will use the same notation later on).

**Lemma 6.1.** *For  $M_q(\vec{X}_q)$  defined by (6.2),*

$$\mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m \left( 1 - \frac{\lambda M_q(\vec{X}_q)}{n} \right)^{n-m} \right] = \mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m e^{-\lambda M_q(\vec{X}_q)} \right] + O\left(\frac{1}{n}\right).$$

**Proof:** Since

$$\sup_{0 \leq x \leq n} \left| x^m \left( 1 - \frac{x}{n} \right)^{n-m} - x^m e^{-x} \right| = O\left(\frac{1}{n}\right)$$

and  $0 \leq \lambda M_q(\vec{X}_q) \leq n$ ,

$$\begin{aligned} & \left| \mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m \left( 1 - \frac{\lambda M_q(\vec{X}_q)}{n} \right)^{n-m} \right] - \mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m e^{-\lambda M_q(\vec{X}_q)} \right] \right| \\ & \leq \int_{-\infty}^{\infty} \left| x^m \left( 1 - \frac{x}{n} \right)^{n-m} - x^m e^{-x} \right| dF_{\lambda M_q(\vec{X}_q)}(x) \leq O\left(\frac{1}{n}\right). \end{aligned}$$

□

**The structure of  $p(\vec{x}_q)$**

By definition,  $p(\vec{x}_q)$  is the probability of  $\{\vec{X}_q = \vec{x}_q\}$ , and we can define

$$a_i(j) = \mathbf{P}_q(X_i = j)$$

to be the probability of answering "j" to the  $i$ -th question. In the case that  $X_1, \dots, X_q$  are independent,

$$p(\vec{x}_q) = \prod_{i=1}^q a_i(x_i)$$

and

$$M_q(\vec{x}_q) = Np(\vec{x}_q) = \prod_{i=1}^q k_i a_i(x_i).$$

If we consider

$$\xi_i = \ln(k_i a_i(X_i))$$

then we can define

$$L_q(\vec{X}_q) = \ln M_q(\vec{X}_q) = \sum_{i=1}^q \ln(k_i a_i(X_i)) = \sum_{i=1}^q \xi_i.$$

In principle, discussions based on  $M_q$  and  $L_q$  are equivalent. Since  $L_q$  can be expressed as a sum of  $q$  random variables, it is more convenient to discuss the limiting distribution of  $L_q$ .

### 6.2.3 Limit theorem for neutral questionnaires

Let us call a questionnaire “neutral” if the distribution of each  $X_i$  is uniform on its possible values. In this case  $a_i(x_i) = \frac{1}{k_i}$  and there is no need to study  $M_q$ , as it is simply 1. Then we have

$$\mathbb{E}\mu_n(m) = N \binom{n}{m} \frac{1}{n^m} \lambda^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \sim N \frac{\lambda^m e^{-\lambda}}{m!},$$

and

$$\mathbb{E}\mu_n = N \left(1 - \left(1 - \frac{\lambda}{n}\right)^n\right) \sim N(1 - e^{-\lambda}).$$

The limits of the ratios are therefore:

$$\frac{\mathbb{E}\mu_n(1)}{n} \rightarrow e^{-\lambda}$$

and

$$\frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n} = \frac{N \binom{n}{m} \frac{1}{n^m} \lambda^m \left(1 - \frac{\lambda}{n}\right)^{n-m}}{N \left(1 - \left(1 - \frac{\lambda}{n}\right)^n\right)} \rightarrow \frac{\lambda^m e^{-\lambda}}{m!(1 - e^{-\lambda})}.$$

Note that  $\mathbb{E}\mu_n(1) \sim n$  in this case, and hence the frequencies defined here form a sequence of large number of rare events in sense of both (d1) and (d2).

In practice, the questionnaires can be neither absolutely neutral nor too “far” from the neutral case. In other words, they are “nearly neutral”. In this case, we assume the sequence of measures  $\mathbf{P}_q$  is contiguous to the sequence of measure  $\mathbf{P}_{0,q}$ .

In more general situations, where  $\{a_i(j)\}$  were assumed to be an arbitrary sequence, the asymptotic behaviour of ratios in (6.1) is more complicated. We will show that, under certain conditions, the limit theorems can still be established.

#### 6.2.4 Limit theorem for contiguous neighborhood of neutral questionnaires

As mentioned, one reason of introducing the likelihood ratio  $M_q$  is its possession of good asymptotic properties. The asymptotic normality of log-likelihood ratio (see e.g., Oosterhoff and Zwet (1979) and Greenwood and Shirayev (1985)) shows that if  $\{\mathbf{P}_q\}$  is contiguous to  $\{\mathbf{P}_{0,q}\}$ , and satisfies some additional conditions, the distribution of  $L_q$  converges to the normal distribution  $\mathcal{N}(-\frac{1}{2}\sigma^2, \sigma^2)$ ; i.e. the distribution of  $M_q$  converges to a log-normal distribution. The limit theorem under this condition can therefore be formulated.

**Definition 6.3.** *A sequence  $\mathbf{P}_q$  of probability measure is called contiguous with respect to another sequence  $\mathbf{P}_{0,q}$  of probability measure if  $\lim_{q \rightarrow \infty} \mathbf{P}_{0,q}(A_q) = 0$  implies  $\lim_{q \rightarrow \infty} \mathbf{P}_q(A_q) = 0$  for any sequence of measurable sets  $A_q$ . This is called one-sided contiguity and can be denoted by  $\mathbf{P}_q \triangleleft \mathbf{P}_{0,q}$ .*

The sequences are said to be contiguous with respect to each other if both  $\mathbf{P}_q \triangleleft \mathbf{P}_{0,q}$  and  $\mathbf{P}_{0,q} \triangleleft \mathbf{P}_q$ . This two-sided contiguity concept is denoted by

$\mathbf{P}_q \triangleleft \mathbf{P}_{0q}$  (Oosterhoff and Zwet, 1979).

**Definition 6.4.** The Hellinger distance  $H(\mathbf{P}, \mathbf{P}_0)$  between two probability measures  $\mathbf{P}$  and  $\mathbf{P}_0$  is defined as follows:

$$H(\mathbf{P}, \mathbf{P}_0) = \left( \int (\sqrt{p} - \sqrt{p_0})^2 d\mu \right)^{\frac{1}{2}} = \left( 2 - 2 \int \sqrt{p} \sqrt{p_0} d\mu \right)^{\frac{1}{2}}$$

where  $p = \frac{d\mathbf{P}}{d\mu}$  and  $p_0 = \frac{d\mathbf{P}_0}{d\mu}$  are corresponding Radon-Nikodym derivatives with respect  $\sigma$ -finite measure  $\mu$  dominating  $\mathbf{P} + \mathbf{P}_0$ .

Suppose  $\mathbf{P}_q = \prod_{i=1}^q \mathbf{P}_{qi}$  and  $\mathbf{P}_{0q} = \prod_{i=1}^q \mathbf{P}_{0,qi}$ , so  $\mathbf{P}_q$  and  $\mathbf{P}_{0q}$  are product measures. Then, the Hellinger distance between product measures and that of their marginals are connected by the relationship

$$H^2(\mathbf{P}_q, \mathbf{P}_{0q}) = 2 - 2 \prod_{i=1}^q \left( 1 - \frac{1}{2} H^2(\mathbf{P}_{qi}, \mathbf{P}_{0,qi}) \right).$$

The following theorem from Oosterhoff and Zwet (1979) established the connection between the Hellinger distance and contiguity.

**Theorem 6.2.** (Oosterhoff and Zwet, 1979)  $\mathbf{P}_q \triangleleft \mathbf{P}_{0q}$  iff

$$\limsup_{q \rightarrow \infty} \sum_{i=1}^q H^2(\mathbf{P}_{qi}, \mathbf{P}_{0,qi}) < \infty$$

and

$$\limsup_{q \rightarrow \infty} \sum_{i=1}^q \mathbf{P}_{qi} \left( \frac{p_{qi}}{p_{0,qi}}(X_{qi}) \geq c_q \right) = 0$$

whenever  $c_q \rightarrow \infty$ .

Consider log-likelihood  $\mathbf{P}_q$  with respect to  $\mathbf{P}_{0q}$

$$L_q = \sum_{i=1}^q \ln \frac{p_{qi}}{p_{0,qi}}(X_{qi}).$$



**Theorem 6.3** (Oosterhoff and Zwet (1979)). *For a given  $\sigma \geq 0$*

$$L_q \xrightarrow{d} \mathcal{N}\left(-\frac{1}{2}\sigma^2; \sigma^2\right)$$

*under measure  $\mathbf{P}_{0q}$  and*

$$\lim_{q \rightarrow \infty} \max_{1 \leq i \leq q} \mathbf{P}_{0,qi} \left( \left| \ln \frac{p_{qi}}{p_{0,qi}}(X_{qi}) \right| \geq \epsilon \right) = 0$$

*for every  $\epsilon > 0$ , iff for every  $\epsilon > 0$*

$$\lim_{q \rightarrow \infty} \sum_{i=1}^q H^2(\mathbf{P}_{qi}, \mathbf{P}_{0,qi}) = \frac{1}{4}\sigma^2$$

*and*

$$\lim_{q \rightarrow \infty} \sum_{i=1}^q \int_{|p_{0,qi} - p_{qi}| \geq \epsilon p_{qi}} (\sqrt{p_{0,qi}} - \sqrt{p_{qi}})^2 d\mu_{qi} = 0.$$

In our context,

$$\mathbf{P}_{qi}(x_i) = a_i(x_i)$$

and

$$\mathbf{P}_{0,qi}(x_i) = 1/k_i,$$

We can define the Hellinger distance between  $\mathbf{P}_{qi}$  and  $\mathbf{P}_{0,qi}$  as follows:

$$H(\mathbf{P}_{qi}, \mathbf{P}_{0,qi}) = \left( 2 - 2 \int \left( \frac{d\mathbf{P}_{qi}}{d\mathbf{P}_{0,qi}} \right)^{\frac{1}{2}} d\mathbf{P}_{0,qi} \right)^{\frac{1}{2}} = \left( 2 - 2 \int \sqrt{k_i a_i(x_i)} d\mathbf{P}_{0,qi} \right)^{\frac{1}{2}},$$

and establish the limit theorem for contiguous neighborhood of neutral questionnaires.

**Theorem 6.4.** *If*

$$\lim_{q \rightarrow \infty} \sum_{i=1}^q H(\mathbf{P}_{qi}, \mathbf{P}_{0,qi})^2 = \frac{1}{4}\sigma^2 < \infty \quad (6.5)$$

*and for every  $\epsilon > 0$ ,*

$$\lim_{q \rightarrow \infty} \sum_{i=1}^q \int_{|k_i a_i(x_i) - 1| \geq \epsilon} \left( \left( \sqrt{k_i a_i(x_i)} - 1 \right) \right)^2 d\mathbf{P}_{0,qi} = 0, \quad (6.6)$$

then

$$\mathbb{E}\mu_n(m) \sim N \frac{1}{m!} \mathbb{E} \left[ \lambda^m e^{mL} e^{-\lambda e^L} \right]$$

and

$$\mathbb{E}\mu_n \sim N \left( 1 - \mathbb{E} \left[ e^{-\lambda e^L} \right] \right).$$

Further,

$$\frac{\mathbb{E}\mu_n(m)}{n} \rightarrow \frac{1}{\lambda m!} \mathbb{E} \left[ \lambda^m e^{mL} e^{-\lambda e^L} \right] \quad (6.7)$$

and

$$\frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n} \rightarrow \frac{\mathbb{E} \left[ \lambda^m e^{mL} e^{-\lambda e^L} \right]}{m! \left( 1 - \mathbb{E} \left[ e^{-\lambda e^L} \right] \right)}, \quad (6.8)$$

with  $L \sim \mathcal{N} \left( -\frac{\sigma^2}{2}, \sigma^2 \right)$ .

**Proof:** Conditions 6.5 and 6.6 imply  $\{\mathbf{P}_q\} \triangleleft \{\mathbf{P}_{0,q}\}$ , and guarantee the asymptotic normality of  $L_q$  (Oosterhoff and Zwet (1979), theorem 2),

$$L_q = \ln M_q \xrightarrow{d(\mathbf{P}_{0,q})} \mathcal{N} \left( -\frac{\sigma^2}{2}, \sigma^2 \right).$$

By Lemma 6.1,

$$\begin{aligned} \mathbb{E}\mu_n(m) &= N \binom{n}{m} \frac{1}{n^m} \mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m \left( 1 - \frac{\lambda M_q(\vec{X}_q)}{n} \right)^{n-m} \right] \\ &\sim N \frac{1}{m!} \mathbb{E} \left[ \lambda^m e^{mL} e^{-\lambda e^L} \right] \end{aligned} \quad (6.9)$$

and

$$\begin{aligned} \mathbb{E}\mu_n &= N \left( 1 - \mathbb{E}_{\mathbf{P}_{0,q}} \left[ \left( 1 - \frac{\lambda M_q(\vec{X}_q)}{n} \right)^n \right] \right) \\ &\sim N \left( 1 - \mathbb{E} \left[ e^{-\lambda e^L} \right] \right), \end{aligned} \quad (6.10)$$

which yields (6.7) and (6.8).  $\square$

In this case, both ratios are strictly greater than 0, and  $\mathbb{E}\mu_n \rightarrow \infty$ . Again, both (d1) and (d2) conditions of LNRE are satisfied.

**Example:** Suppose we have

$$\mathbf{P}_{qi}(j) = a_i(j) = \frac{1 + \frac{e_{ij}}{\sqrt{q}}}{k_i},$$

where  $\{e_{ij}\}$  satisfies  $-1 \leq e_{ij} \leq 1$  and

$$\lim_{q \rightarrow \infty} \frac{1}{q} \sum_{i=1}^q \frac{1}{k_i} \sum_{j=1}^{k_i} e_{ij}^2 \rightarrow \sigma^2 < \infty$$

with constraint  $\sum_{j=1}^{k_i} e_{ij} = 0$ . Then the square of the Hellinger distance between  $\mathbf{P}_{qi}$  and  $\mathbf{P}_{0,qi}$  becomes

$$H(\mathbf{P}_{qi}, \mathbf{P}_{0,qi})^2 = 2 - 2 \int \sqrt{k_i a_i(x_i)} d\mathbf{P}_{0,qi} = 2 - 2 \frac{1}{k_i} \sum_{j=1}^{k_i} \sqrt{1 + \frac{e_{ij}}{\sqrt{q}}}.$$

Using Taylor's expansion we get

$$\sqrt{1 + \frac{e_{ij}}{\sqrt{q}}} = 1 + \frac{1}{2} \frac{e_{ij}}{\sqrt{q}} - \frac{1}{8} \frac{e_{ij}^2}{q} + \frac{1}{16} \frac{e_{ij}^3}{q\sqrt{q}} \dots,$$

and hence

$$\sum_{i=1}^q H(\mathbf{P}_{qi}, \mathbf{P}_{0,qi})^2 = \frac{1}{4} \sigma^2 + O\left(\frac{1}{\sqrt{q}}\right) \rightarrow \frac{1}{4} \sigma^2.$$

Since when  $q > \frac{1}{\epsilon^2}$ ,  $|k_i a_i(x_i) - 1| < \epsilon$  for all  $i$ , it is easy to see that (6.6) is satisfied. This implies the asymptotic normality of  $L_q$ .

**Remark 1.** In our treatment in this section, we assumed that the components of  $\vec{X}_q$  are independent. However, this is not a necessary condition. In the case that components of  $\vec{X}_q$  are dependent, we can simply replace  $k_i a(x_i)$  by conditional probabilities  $k_i a(x_i | \vec{x}_{i-1})$ , to achieve the same result (see Greenwood and Shirayev (1985)).

### 6.2.5 Limit theorem for general cases

In general, if  $\{a_i(j)\}$  is an arbitrary sequence of distributions, then unlike the contiguity case in previous section where  $L_q(\vec{X}_q)$  converge in distribution to normal random variable, the expectation of  $L_q(\vec{X}_q)$  usually tends to  $-\infty$  while the variance goes to  $\infty$ .

In this situation we can use similar technique which typically are used in the theory of large deviations (see e.g. Feller (1970) and Kallenberg (2002)). After applying Esscher's transform (see, e.g. Feller (1970)),  $Y_q = \frac{L_q(\vec{X}_q)}{\sqrt{q}}$  will converge under the adjoint measure, and the distribution of  $Y_q$  can be approximated by Edgeworth series (see e.g. Feller (1970) and Kolassa (2006)).

Under necessary conditions, we shall see that, in this case, the limit theorem can be established and result agrees with Karlin-Rouault's law (see, e.g. Khmaladze and Chitashvili (1989), Baayen (2001)).

For any fixed sequence  $\{a_i(j)\}$ , the cumulant generating function of  $\xi_i$  under  $\mathbf{P}_{0,qi}$  can be defined by

$$\psi_i(u) = \ln \mathbb{E}_{\mathbf{P}_{0,qi}} e^{u\xi_i} = \ln \left( \sum_{j=1}^{k_i} [k_i a_i(j)]^u \right) - \ln(k_i).$$

Then the cumulant generating function of  $L_q(\vec{X}_q)$  is

$$\ln \mathbb{E}_{\mathbf{P}_{0,q}} e^{uL_q(\vec{X}_q)} = \sum_{i=1}^q \psi_i(u).$$

By Esscher's transform, the distribution  $\mathbf{Q}_{u,q}$  of  $L_q(\vec{X}_q)$  adjoint to  $\mathbf{P}_{0,q}$  can be defined as follows,

$$\frac{d\mathbf{Q}_{u,q,L_q(\vec{X}_q)}}{d\mathbf{P}_{0,q,L_q(\vec{X}_q)}}(z) = e^{uz - \sum_{i=1}^q \psi_i(u)}.$$

Consequently, the logarithm of moment generating function of  $Y_q = \frac{L_q(\vec{X}_q)}{\sqrt{q}}$  under  $\mathbf{Q}_{u,q}$  is,

$$\ln \mathbb{E}_{\mathbf{Q}_{u,q}} e^{rY_q} = \sum_{i=1}^q \psi_i\left(u + \frac{r}{\sqrt{q}}\right) - \sum_{i=1}^q \psi_i(u).$$

We can choose  $u = u_q$  such that  $\mathbb{E}_{\mathbf{Q}_{u_q,q}} L_q(\vec{X}_q) = \sum_{i=1}^q \psi'_i(u_q) = 0$ . The variance of  $Y_q$  under  $\mathbf{Q}_{u_q,q}$  is

$$\sigma_q^2 = \frac{1}{q} \sum_{i=1}^q \psi''_i(u_q),$$

and therefore  $Y_q = \frac{L_q(\vec{X}_q)}{\sqrt{q}}$  becomes a random variable with mean 0 and variance  $\sigma_q^2$  under  $\mathbf{Q}_{u_q,q}$ .

**Theorem 6.5.** Assume  $u_q$  is the solution of  $\sum_{i=1}^q \psi'_i(u) = 0$ . If  $\{a_i(j)\}$  is such that

$$c < \frac{1}{q} \sum_{i=1}^q \psi''_i(u_q) < C, \quad (6.11)$$

and if there exists  $\delta > 0$  such that

$$\left| e^{\sum_{i=1}^q [\psi_i(u_q+r) - \psi_i(u_q)]} \right| = o\left(\frac{1}{\sqrt{q}}\right) \quad \text{uniformly in } r > \delta > 0, \quad (6.12)$$

then

$$\mathbb{E}\mu_n(m) \sim N e^{\sum_{i=1}^q \psi_i(u_q)} \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \frac{\Gamma(m - u_q)}{m!}, \quad (6.13)$$

$$\mathbb{E}\mu_n \sim N e^{\sum_{i=1}^q \psi_i(u_q)} \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \frac{\Gamma(1 - u_q)}{u_q} \quad (6.14)$$

and

$$\frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n} \rightarrow \frac{u^* \Gamma(m - u^*)}{\Gamma(m + 1) \Gamma(1 - u^*)}, \quad (6.15)$$

where  $u^* = \lim_{q \rightarrow \infty} u_q$ .

**Proof:** Applying Esscher's transform we get

$$\begin{aligned} & \mathbb{E}_{\mathbf{P}_{0,q}} \left[ (\lambda M_q(\vec{X}_q))^m e^{-\lambda M_q(\vec{X}_q)} \right] \\ &= e^{\sum_{i=1}^q \psi_i(u_q)} \int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)x} e^{-\lambda e^x} d\mathbf{Q}_{u_q,q,L_q(\vec{X}_q)}(x), \end{aligned} \quad (6.16)$$

then replace  $L_q(\vec{X}_q)$  by  $Y_q$ ,

$$\begin{aligned} & \int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)x} e^{-\lambda e^x} d\mathbf{Q}_{u_q,q,L_q(\vec{X}_q)}(x) \\ &= \int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\mathbf{Q}_{u_q,q,Y_q}(y). \end{aligned} \quad (6.17)$$

In Lemma 6.2 , we will prove that under condition (6.11) and (6.12),

$$\begin{aligned} & \int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\mathbf{Q}_{u_q,q,Y_q}(y) \\ &= \int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) + o\left(\frac{1}{\sqrt{q}}\right), \end{aligned} \quad (6.18)$$

where  $\Phi_{0,\sigma_q^2}$  is normal distribution function with mean 0 and variance  $\sigma_q^2$ . Then by Lemma 6.3,

$$\begin{aligned} & \int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) \\ & \sim \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \Gamma(m - u_q) = O\left(\frac{1}{\sqrt{q}}\right). \end{aligned} \quad (6.19)$$

Combine (6.9), Lemma 6.1, (6.16), (6.17), (6.18), (6.19), and note that  $\frac{1}{n} = o\left(\frac{1}{\sqrt{q}}\right)$ , we conclude that for any  $m \geq 1$ ,

$$\mathbb{E}\mu_n(m) \sim N e^{\sum_{i=1}^q \psi_i(u_q)} \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \frac{\Gamma(m - u_q)}{m!}.$$

Combining the last equation with

$$\sum_{m=1}^{\infty} \frac{\Gamma(m - u_q)}{m!} = \frac{\Gamma(1 - u_q)}{u_q}$$

and considering  $\mathbb{E}\mu_n(m) \geq 0$ , we get

$$\mathbb{E}\mu_n \sim Ne^{\sum_{i=1}^q \psi_i(u_q)} \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0, \sigma_q^2}(0) \frac{\Gamma(1 - u_q)}{u_q}$$

and consequently (6.15). □

**Lemma 6.2.** *If conditions (6.11) and (6.12) are satisfied, then (6.18) holds.*

**Proof:** Denote  $g(y, q) = \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}}$  and  $g'(y, q) = \frac{\partial g(y, q)}{\partial y}$ , then  $\lim_{y \rightarrow \infty} g(y, q) = 0$  and  $\lim_{y \rightarrow -\infty} g(y, q) = 0$  when  $u_q < m$ ,

$$\begin{aligned} & \int_{-\infty}^{\infty} g(y, q) d\mathbf{Q}_{u_q, q, Y_q}(y) - \int_{-\infty}^{\infty} g(y, q) d\Phi_{0, \sigma_q^2}(y) \\ &= \int_{-\infty}^{\infty} \left( \mathbf{Q}_{u_q, q, Y_q}(y) - \Phi_{0, \sigma_q^2}(y) \right) g'(y, q) dy. \end{aligned}$$

Under the condition (6.11) and (6.12), the Edgeworth expansion (see Feller (1970)) shows,

$$\mathbf{Q}_{u_q, q, Y_q}(y) = \Phi_{0, \sigma_q^2}(y) - \frac{\sum_{i=1}^q \psi_i^{(3)}(u_q)}{6\sigma_q^3 q^{\frac{3}{2}}} H_2(\sigma_q y) \phi(\sigma_q y) + o\left(\frac{1}{\sqrt{q}}\right).$$

Here  $H_2(y) = y^2 - 1$  is the second Hermite polynomial. Since  $\int_{-\infty}^{\infty} |g'(y, q)| dy$  is bounded,

$$\begin{aligned} & \int_{-\infty}^{\infty} \left( \mathbf{Q}_{u_q, q, Y_q}(y) - \Phi_{0, \sigma_q^2}(y) \right) g'(y, q) dy \\ &= - \int_{-\infty}^{\infty} \frac{\sum_{i=1}^q \psi_i^{(3)}(u_q)}{6\sigma_q^3 q^{\frac{3}{2}}} H_2(\sigma_q y) \phi(\sigma_q y) g'(y, q) dy + o\left(\frac{1}{\sqrt{q}}\right) \\ &= - \frac{\frac{1}{q} \sum_{i=1}^q \psi_i^{(3)}(u_q)}{6\sigma_q^2 \sqrt{q}} \int_{-\infty}^{\infty} H_3(\sigma_q y) \phi(\sigma_q y) g(y, q) dy + o\left(\frac{1}{\sqrt{q}}\right). \end{aligned} \quad (6.20)$$

Here  $H_3(y) = y^3 - 3y$  is the third Hermite polynomial. Now, since

$$\begin{aligned} & \int_{-\infty}^{\infty} H_3(\sigma_q y) \phi(\sigma_q y) g(y, q) dy \\ &= \int_{-\infty}^{\infty} ((\sigma_q y)^3 - 3\sigma_q y) \phi(\sigma_q y) \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} dy \rightarrow 0 \end{aligned}$$

and  $\lim_{q \rightarrow \infty} \frac{1}{q} \sum_{i=1}^q \psi_i^{(3)}(u_q) < \infty$ , the right side of (6.20) is  $o(\frac{1}{\sqrt{q}})$  and (6.18) holds.  $\square$

**Lemma 6.3.** *Suppose  $u_q$  is solution of  $\sum_{i=1}^q \psi_i'(u) = 0$ , then*

$$\int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) \sim \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \Gamma(m - u_q).$$

**Proof:** Since for any  $\beta > 0$  and  $m \geq 1 > u_q$ , we have for large values of  $q$ ,

$$\begin{aligned} & \int_{-\infty}^{-\beta q^{-\frac{1}{4}}} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) \\ & \leq \lambda^m e^{-(m-u_q)\beta q^{\frac{1}{4}}} e^{-\lambda e^{-\beta q^{\frac{1}{4}}}} \int_{-\infty}^{-\beta q^{-\frac{1}{4}}} d\Phi_{0,\sigma_q^2}(y) < o\left(\frac{1}{\sqrt{q}}\right) \end{aligned}$$

and

$$\begin{aligned} & \int_{\beta q^{-\frac{1}{4}}}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) \\ & \leq \lambda^m e^{(m-u_q)\beta q^{\frac{1}{4}}} e^{-\lambda e^{\beta q^{\frac{1}{4}}}} \int_{\beta q^{-\frac{1}{4}}}^{\infty} d\Phi_{0,\sigma_q^2}(y) < o\left(\frac{1}{\sqrt{q}}\right) \end{aligned}$$

while,

$$\begin{aligned} & \int_{-\beta q^{-\frac{1}{4}}}^{\beta q^{-\frac{1}{4}}} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) \\ & = \lambda^{u_q} \int_{-\beta q^{\frac{1}{4}}}^{\beta q^{\frac{1}{4}}} (\lambda e^z)^{m-u_q} e^{-\lambda e^z} d\Phi_{0,q\sigma_q^2}(z) \\ & = \lambda^{u_q} \int_{-\beta q^{\frac{1}{4}}}^{\beta q^{\frac{1}{4}}} (\lambda e^z)^{m-u_q} e^{-\lambda e^z} \frac{1}{\sigma_q \sqrt{2\pi q}} e^{-\frac{z^2}{2q\sigma_q^2}} dz \\ & \sim \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \int_{-\infty}^{\infty} (\lambda e^z)^{m-u_q} e^{-\lambda e^z} dz \\ & = \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \Gamma(m - u_q). \end{aligned}$$

Hence, we conclude,

$$\int_{-\infty}^{\infty} \lambda^m e^{(m-u_q)\sqrt{q}y} e^{-\lambda e^{\sqrt{q}y}} d\Phi_{0,\sigma_q^2}(y) \sim \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0,\sigma_q^2}(0) \Gamma(m - u_q).$$

$\square$



Note that in this case,

$$\mathbb{E}\mu_n(m) \sim N e^{\sum_{i=1}^q \psi_i(u_q)} \frac{\lambda^{u_q}}{\sqrt{q}} \phi_{0, \sigma_q^2}(0) \frac{\Gamma(m - u_q)}{m!},$$

implies

$$\frac{\mathbb{E}\mu_n(1)}{n} \sim O\left(\frac{1}{\sqrt{q}}\right) \rightarrow 0,$$

and therefore (d2) are satisfied but not (d1).



# Chapter 7

## Conclusions

In spite of having wide application and being of various forms, the essential nature of divisible statistics is simple: they are the sum of functions of a sequence of frequencies, which follow a joint multinomial distribution. Therefore, the asymptotic properties of divisible statistics can be discussed in a unified framework. In this thesis, we established such a framework for divisible statistics in both classical multinomial models and non-classical models.

In classical multinomial models, the probabilities are fixed and the expectation of all frequencies tends to infinity, hence the limit distribution of the vector of normalized frequencies is multivariate normal. As we have shown in chapter 2, based on this limiting distribution of normalized frequencies, the limit theorems of general divisible statistics can be established; and some useful properties such as the asymptotic equivalence to Chi-square for some classes of divisible statistics, and the distribution-free nature of this class, can be obtained.

However, in the move from classical to non-classical models, these good properties are lost. The asymptotic behaviour of the divisible statistics becomes very complicated and it is difficult to establish limit theorems. We illustrate these difficulties in chapter 3 and carried out preliminary analysis to the

asymptotic behaviour of some objects, which are important in specifying the models.

The most important results of this thesis come from chapter 4, in which we discussed a powerful method for establishing limit theorems of divisible statistics. This advanced approach, first established by Khmaladze (1984), has not yet attracted much attention in applied statistics. In addition to describing this martingale approach in detail, we show that this approach can be extended to a more general situation, where both asymptotically Poissonian and Gaussian frequencies can be allowed. The functional limit theorems have been established in this new scheme and the comparison between these two schemes has been discussed.

Chapter 5 was devoted to discussion of the properties and applications of the limiting processes, arising from the FLTs of chapter 4. We showed that the limiting processes can be determined by two characteristics, and gave examples of them for different divisible statistics. Some properties of these characteristics have been discussed, in particular for the non-classical models with all frequencies asymptotically Gaussian. Furthermore, we have shown that a new class of statistics can be constructed based on the martingale component or a 1-to-1 transformation of the partial sum process. We have shown that this new class of statistics possesses some very desirable properties such as being distribution-free and having power to detect some contiguous alternatives which cannot be detected by divisible statistics.

At the end of this thesis, we discussed selected topics in LNRE theory. The LNRE models covered most non-classical multinomial models, including those discussed in chapters 4 and 5. While in chapter 6, we focused on the LNRE models which satisfy (d2) but not (d1) condition. In particular, we established limit theorems for the model with context of diversity of questionnaires. The asymptotics of spectral statistics and vocabulary was revealed

and the ratios

$$\alpha_n(m) = \frac{\mathbb{E}\mu_n(m)}{\mathbb{E}\mu_n}$$

can be seen satisfying the so-called Karlin-Rouault law.

Due to the complex nature of the problem, there are still some aspects which need to be investigated in the future, such as the limit theorems of partial sum processes when parameters are estimated and the convergence rates of the partial sum processes. There is also room in further development in LNRE theory.



# Appendix A

## Some discussion on Large Deviations

### Large Deviation

Consider a sequence of i.i.d. random variables  $\xi_1, \dots, \xi_i, \dots, \xi_n$ . If  $\mathbb{E}\xi_i = \mu < \infty$  and  $\sigma^2 = \text{var}(\xi_i) < \infty$ , then the CLT implies as  $n \rightarrow \infty$ ,

$$Y_n = \frac{S_n - n\mu}{\sqrt{n}} \rightarrow Y \sim \mathcal{N}(0, \sigma^2)$$

so that  $Y_n$  can be approximated by a normal random variable  $Y$ .

However, the approximation of  $Y_n$  by  $Y$  is good only in the central part, but not in the tail. For example if all  $\xi_i$  are iid Bernoulli r.v. with parameter  $p$ , then  $\mathbf{P}(Y_n > \sqrt{n}(1-p)) = 0$  which is obviously different from  $\mathbf{P}(Y > \sqrt{n}(1-p))$ .

There are many situation where the tail is of interest. For example, sometimes we want to investigate the relative error of  $1 - F_{Y_n}(x)$  approximated by  $1 - F_Y(x)$ , when  $y \sim n^\alpha$ .

### Esscher's transforms

To cope with the difficulty resulted from the “large deviation”, we can employ the technique of Esscher's transforms.

**Definition A.1.** *Given a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  and a random variable  $Y$ , the adjoint probability measures  $Q_u$  (sometime called conjugate measure) with Radon-Nikodým density*

$$\frac{dQ_u}{dP}(y) = e^{uy - \Lambda_{Y,P}(u)}$$

for some real number  $u$ , with  $\Lambda_{Y,P}(u) = \ln E_P e^{uY}$  being cumulant-generating function of  $Y$  under  $P$ , are called Esscher transforms (Esscher 1932).

The cumulant-generating function of  $Y_n$  under  $Q_u$  is

$$\Lambda_{Y_n, Q_u}(r) = \ln \int e^{ry} \frac{dQ_u}{dP} dP = \Lambda_{Y_n, P}(u + r) - \Lambda_{Y_n, P}(u)$$

then expectation of  $Y_n$  under  $Q_u$  is therefore a function of  $u$ ,

$$\mathbb{E}_{Q_u}(Y_n) = \left. \frac{d\Lambda_{Y_n, Q_u}(r)}{dr} \right|_{r=0} = \Lambda'_{Y_n, P}(u).$$

Hence we can **choose**  $u$  such that,

$$\mathbb{E}_{Q_u}(Y_n) = \Lambda'_{Y_n, P}(u) = cn^\alpha$$

such that the deviations are NOT large any more. The tail in origin measure is now moved to the central of the adjoint measure.

### Example: Asymptotic Expansion of the Distribution of a Sum

Consider the following problem: For  $x > 0$ , what is the asymptotic expansion of

$$\mathbf{P}(S_n > x) = \mathbf{P}\left(Y_n > \frac{x}{\sqrt{n}} - \mu\sqrt{n}\right)$$



Choose  $u$  such that  $\mathbb{E}_{Q_u} Y_n = \Lambda'_{Y_n, P}(u) = -\mu\sqrt{n}$ , then  $\text{var}_{Q_u} Y_n = \Lambda''_{Y_n, P}(u) = \sigma^2$  and

$$\begin{aligned}
\mathbf{P}(S_n > x) &= e^{\Lambda_{Y_n, P}(u)} \int_{\frac{x}{\sqrt{n}} + \Lambda'_{Y_n, P}(u)}^{\infty} e^{-uy} dQ_{u, Y_n}(y) \\
&= e^{\Lambda_{Y_n, P}(u) - u\Lambda'_{Y_n, P}(u)} \int_{\frac{x}{\sigma\sqrt{n}}}^{\infty} e^{-u\sigma z} dQ_{u, Z_n}(z) \\
&\sim e^{\Lambda_{Y_n, P}(u) - u\Lambda'_{Y_n, P}(u)} \int_{\frac{x}{\sigma\sqrt{n}}}^{\infty} e^{-u\sigma z} d\Phi(z) \\
&= e^{\Lambda_{Y_n, P}(u) - u\Lambda'_{Y_n, P}(u) + \frac{u^2 \Lambda''_{Y_n, P}(u)}{2}} \left[ 1 - \Phi\left(\frac{x}{\sigma\sqrt{n}} + u\sigma\right) \right]
\end{aligned}$$

Further,

$$1 - \Phi\left(\frac{x}{\sqrt{n}} + u\sigma\right) \sim \frac{\varphi\left(\frac{x}{\sqrt{n}} + u\sigma\right)}{\frac{x}{\sqrt{n}} + u\sigma} \sim \frac{1}{\sqrt{2\pi}u\sigma} e^{-\frac{\left(\frac{x}{\sqrt{n}} + u\sigma\right)^2}{2}} \sim \frac{e^{\frac{u^2\sigma^2}{2}}}{\sqrt{2\pi}u\sigma} e^{-\frac{xu}{\sqrt{n}}}$$

Set  $t = \frac{u}{\sqrt{n}}$ , then since

$$\Lambda'_{Y_n, P}(u) = n\lambda'\left(\frac{u}{\sqrt{n}}\right) - \mu\sqrt{n} = -\mu\sqrt{n},$$

$\lambda'(t) = 0$  and eventually,

$$\mathbf{P}(S_n > x) \sim e^{\Lambda_{Y_n, P}(u) - u\Lambda'_{Y_n, P}(u) + \frac{u^2 \Lambda''_{Y_n, P}(u)}{2}} \frac{e^{\frac{u^2\sigma^2}{2}}}{\sqrt{2\pi}u\sigma} e^{-\frac{xu}{\sqrt{n}}} = A(t)e^{-xt}$$

with  $A(t)$  being some term irrelevant to  $x$ .



# Bibliography

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience.
- Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):131–142.
- Baayen, R. H. (2001). *Word Frequency Distributions*. Springer.
- Brzezniak, Z. and Zastawniak, T. (2000). *Basic Stochastic Processes*. Springer.
- Burbea, J. and Rao, C. R. (1982a). On the convexity of higher order jensen differences based on entropy functions. *IEEE Transactions on Information Theory*, 28(6):961–963.
- Burbea, J. and Rao, C. R. (1982b). On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28(3):489 – 495.
- Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):440–464.
- Deheuvels, P. and Martynov, G. (2003). Karhunenloev expansions for weighted wiener processes and brownian bridges via bessel functions. *Progress in Probability*, 55:57–93.

- Devore, J. L. and Berk, K. N. (2006). *Modern Mathematical Statistics with Applications*. Duxbury Press.
- Durrett, R. (2004). *Probability: Theory and Examples (Probability: Theory & Examples)*. Duxbury Press.
- Feller, W. (1970). *An Introduction to Probability Theory and Its Applications, Vol 2*. Wiley.
- Ferguson, T. S. (1996). *A Course in Large Sample Theory*. CRC Press.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2010). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-92.
- Greenwood, P. E. and Shiryayev, A. N. (1985). *Contiguity and the statistical invariance principle*. Gordon and Breach.
- Gyorfi, L. and Vajda, I. (2002). Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models. *Statistics & Probability Letters*, 56(1):57 – 67.
- Haywood, J. and Khmaladze, E. (2008). On distribution-free goodness-of-fit testing of exponentiality. *Journal of Econometrics*, 143(1):5–18.
- Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika*, 59(1):137–145.
- Ivchenko, G. I. and Levin, V. V. (1976). Asymptotic normality of a class of statistics in the multinomial scheme. *Theory of Probability and its Applications*, 21(1):188–192.
- Ivchenko, G. I. and Medvedev, Y. I. (1979). Separable statistics and hypothesis testing. the case of small samples. *Theory of Probability and its Applications*, 23(4):764–775.

- Ivchenko, G. I. and Medvedev, Y. I. (1981). Decomposable statistics and hypothesis testing for grouped data. *Theory of Probability and its Applications*, 25(3):540–551.
- Jost, L. and Chao, A. (2010). *Diversity Analysis*. Chapman and Hall/CRC Applied Environmental Statistics Series. Taylor and Francis.
- Kallenberg, O. (2002). *Foundations of modern probability*. Springer, 2 edition.
- Kendall, M. G., Stuart, A., and Ord, J. K. (1987). *Kendall's advanced theory of statistics*, volume 2A. Griffin, London.
- Khmaladze, E. V. (1984). Martingale limit theorems for divisible statistics. *Theory of Probability and its Applications*, 28(3):530–548.
- Khmaladze, E. V. (1988). The statistical analysis of a large number of rare events. *CWI Report MS-R8804*.
- Khmaladze, E. V. (2009). Diversity of responses in questionnaires and similar objects. Technical Report 09-3, MSOR, Victoria University of Wellington.
- Khmaladze, E. V. and Chitashvili, R. J. (1989). The statistical analysis of a large number of rare events and the related problem. *Proc. Tbilisi Mathematical Institute*, 92:196–245.
- Kolassa, J. E. (2006). *Series Approximation Methods in Statistics*, volume 88 of *Lecture Notes in Statistics*. Springer-Verlag New York.
- Kudlaev, E. M. (1990). Divisible statistics. *Journal of Mathematical Sciences*, 50(3):1601–1642.
- Kvizhinadze, G. (2010). *Large number of rare events: Diversity analysis in multiple choice questionnaires and related topics*. PhD thesis, Victoria University of Wellington.
- Lehmann, E. and Romano, J. P. (2005). *Testing Statistical Hypotheses (Springer Texts in Statistics)*. Springer.

- Liptser, R. S. and Shirayayev, A. N. (1981). A functional central limit theorem for semimartingales. *Theory of Probability and its Applications*, 25(4):667–688.
- Magurran, A. E. (2004). *Measuring Biological Diversity*. Blackwell Science Ltd.
- Mandelbrot, B. (1952). An information theory of the statistical structure of language. *Proceedings of the Symposium on Applications of communication Theory*, pages 486–500.
- Medvedev, Y. I. (1970). Theorems on asymptotic distribution of the statistic  $X^2$ . *Dokl. Akad. Nauk SSSR*, 192(5):987–989.
- Medvedev, Y. I. (1977). Separable statistics in a polynomial scheme, I. *Theory of Probability and its Applications*, 22(1):1–15.
- Medvedev, Y. I. (1978). Separable statistics in a polynomial scheme, II. *Theory of Probability and its Applications*, 22(3):607–615.
- Mnatsakanov, R. M. (1986). A functional limit theorem for additively separable statistics in the case of very rare events. *Theory of Probability and its Applications*, 30(3):622–626.
- Mnatsakanov, R. M. (1987). On the convergence of separable statistics to a wiener process. *Theory of Probability and its Applications*, 32(1):152–157.
- Morris, C. (1975). Central limit theorems for multinomial sums. *The Annals of Statistics*, 3(1):165–188.
- Oosterhoff, J. and Zwet, W. R. v. (1979). A note on contiguity and hellinger distance. *Contributions to Statistics, Reidel, Dordrecht*, pages 157–166.
- Orlov, A. I. (1973). On testing the symmetry of distributions. *Theory of Probability and its Applications*, 17(2):357–361.

- Orlov, J. K. and Chitashvili, R. Y. (1983a). Generalized z-distribution generating the well-known "rank-distributions.". *Bulletin of the Academy of Sciences, Georgia*, 110:269–272.
- Orlov, J. K. and Chitashvili, R. Y. (1983b). On the statistical interpretation of zipf's law. *Bulletin of the Academy of Sciences, Georgia*, 109:505–508.
- Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*, volume 185 of *Statistics, textbooks and monographs*. Taylor & Francis Group, LLC.
- Patil, G. P. and Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American Statistical Association*, 77(379):548–561.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine*, 50:157–172.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley & Sons.
- Rao, C. R. (1982). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhya: The Indian Journal of Statistics, Series A*, 44(1):1–22.
- Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer series in statistics. Springer-Verlag New York.
- Rouault, A. (1978). Loi de zipf et sources markoviennes. *Annales de l'Institute H. Poincare*, 14:169–188.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656.
- Shao, J. (2003). *Mathematical Statistics*. Springer, 2 edition.

- Shiryaev, A. N. (1995). *Probability (Graduate Texts in Mathematics)*. Springer.
- Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes With Applications to Statistics*. John Wiley & Sons Inc.
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.
- Simon, H. A. (1960). Some further notes on a class of skew distribution functions. *Information and Control*, 3(1):80 – 88.
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163:688.
- Williams, C. B. (1964). *Patterns in the balance of nature and related problems in quantitative ecology*. London, New York, Academic Press.
- Williams, D. (1991). *Probability with martingales*. Cambridge University Press.
- Yule, G. U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London.*, 212:21–87.
- Zipf, G. (1935). *The Psycho-Biology of Language*. Houghton Mifflin, Boston.