جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل/ كلية التربية

حول نظرية التوزيعات التقاربية

رسالة
مقدمة الى قسم الرياضيات- كلية التربية- جامعة بابل وهو جزء من متطلبات
نيل درجة الماجستير في علوم الرياضيات

من قبل

جنان حمزة فرهود الخناني

بإشراف

الدكتورة كريمة عبد الكاظم الخفاجي

1426هـ 2005 م

REPUBLIC OF IRAQ,
MINISTRY OF HIGHER EDUCATION,
AND SCIENTIFIC RESEARCH,
BABYLON UNIVERSITY,
COLLEGE OF EDUCATION

ON ASYMPTOTIC DISTRIBUTIONS THEORY

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS- COLLEGE OF EDUCATION- UNIVERSITY OF BABYLON IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN MATHEMATICS

BY
JINAN HAMZA FARHOOD AL-KHINANI

SUPERVISED
DR. KAREEMA ABDUL-KADDUM ALKAFAGI

2005

ABSTRACT

The essence of asymptotic methods is approximation. The main object of this thesis is to give a unified derivation of some results and theorems.

Also, this thesis deals with asymptotic distributions that is the distributions we obtained by letting the time horizon (sample size) tends to infinity. The research methodology is theoretical. The present study consists of three chapters.

In chapter one we give general introduction, review of literature and out line of this thesis.

Chapter two contains the key tools of asymptotic analysis and presents limit theorems for a univariate case (for example about sequence of random variables).

Chapter three contains the key tools of asymptotic analysis and presents limit theorems for a multivariate case (for example about sequence of random vectors).

Furthermore, we obtain some results about the asymptotic theory.

حول نظرية التوزيعات التقاربية

الخلاصة

ان جوهر الطرق التقاربية هو التقريب وان الهدف الرئيسي من هذه الرسالة هو اعطاء اشتقاق موحد لبعض النتائج والنظريات.

ان هذه الرسالة تتعامل ايضاً مع التوزيعات التقاربية التي نحصل عليها عندما يقترب حجم العينة من المالانهاية (يتزايد بصورة غير متناهية) وتتضمن الدراسة الحالية ثلاثة فصول.

نعطي في الفصل الاول مقدمة عامة وخلفيات البحث واطار عام لهذه الرسالة.

اما الفصل الثاني فانه يحتوي على الادوات الاساسية والرئيسية للتحليل التقاربي ويقدم نظريات الغاية لحالة المتغير الواحد (على سبيل المثال متتابعة المتغيرات العشوائية).

والفصل الثالث ضم الادوات الاساسية والرئيسية للتحليل التقاربي وقدم نظريات الغاية لحالة اكثر من متغير (على سبيل المثال متتابعة المتجهات العشوائية).

وبالاضافة الى ذلك، اوجدنا بعض النتائج حول نظرية التقارب.

SUPERVISOR CERTIFICATION

I certify that this thesis was prepared under my supervision at the Department of Mathematics/ College of Education/ University of Babylon, as a partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

Signature:

Name: ***Dr. Kareema Abdul-Kaddum AlKhafaji***

Title: Lecture

Date: / / 2005

RECOMMENDATION OF THE HEAD OF THE DEPARTMENT

In view of the available recommendations, I forward this research for debate by the Examining Committee.

Signature:

Name: Dr. Lua'y Abdul Hani Al-Siweedy

Title: Assistant Professor

Date: / / 2005

EXAMINING COMMITTEE CERTIFICATION

We certify that we have read this thesis entitled "On Asymptotic Distributions Theory" and, as an Examining Committee, we examined the student in it's content, and what is related to it, and that in our opinion it is adequate with standing as a thesis for the degree of Master of Science in Mathematics.

Chairman

Signature:

Name: Dr. Sa'ad Abid Madhi

Title: Assistant Professor

Date: / / 2005

Member

Signature:

Name: Dr. Abdul- Ameer Abdul- Hussein Al-Ali

Title: Assistant Professor

Date: / /2005

Member

Signature:

Name: Dr. Udei Sabri Abdul-Razzak

Title: Assistant Professor

Date: / / 2005

Member (Supervisor)

Signature:

Name: Dr. Kareema Abdul- Kaddum

Title: Lecture

Date: / / 2005

Approved by the Dean of the College of Education

Signature:

Name: Dr. Abdul- Ilah Rizzoki Karbal

Title: Professor

Date: / / 2005

ON ASYMPTOTIC DISTRIBUTIONS THEORY

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS- COLLEGE OF EDUCATION- UNIVERSITY OF BABYLON IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN MATHEMATICS

BY
JINAN HAMZA FARHOOD AL-KHINANI

SUPERVISED
DR. KAREEMA ABDUL-KADDUM ALKAFAGI

٢٠٠٥

{

Zفَأَمَّا الزَّبَدُ فَيَذْهَبُ جُفَاءً وَأَمَّا مَا
يَنْفَعُ النَّاسَ فَيَمْكُثُ فِي الأَرْضِ
كَذَلِكَ يَضْرِبُ اللَّهُ الأَمْثَالَE

صدق الله العظيم

سورة الرعد-١٧

الاهداء

الى روح والدتي الطاهرة...

الى القلب الذي وهبني الامان....

وانار لي الدرب...

وتحدى من اجلي الكرب...

واعطى فأجزل العطاء...

والدي

الى نوارس المحبة والبسمة الدائمة...

اخوتي

واخواتي وبالاخص وجدان

الى الذين لمست منهم التشجيع والعون في طلب العلم...

الى من يسعده نجاحي

اهدي هذا الجهد المتواضع

جنان

SUPERVISOR CERTIFICATION

I certify that this thesis was prepared under my supervision at the Department of Mathematics/ College of Education/ University of Babylon, as a partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

Signature:

Name: ***Dr. Kareema Abdul-Kaddum AlKhafaji***

Title: Lecture

Date: / / ٢٠٠٥

RECOMMENDATION OF THE HEAD OF THE DEPARTMENT

In view of the available recommendations, I forward this research for debate by the Examining Committee.

Signature:

Name: Dr. Lua'y Abdul Hani Al-Siweedy

Title: Assistant Professor

Date: / / ٢٠٠٥

EXAMINING COMMITTEE CERTIFICATION

We certify that we have read this thesis entitled "On Asymptotic Distributions Theory" and, as an Examining Committee, we examined the student in it's content, and what is related to it, and that in our opinion it is adequate with standing as a thesis for the degree of Master of Science in Mathematics.

Chairman

Signature:

Name: Dr. Sa'ad Abid Madhi

Title: Assistant Professor

Date: / /٢٠٠٥

Member

Signature:

Name: Dr. Abdul- Ameer Abdul- Hussein Al-Ali

Title: Assistant Professor

Date: / /٢٠٠٥

Member

Signature:

Name: Dr. Udei Sabri Abdul-Razzak

Title: Assistant Professor

Date: / /٢٠٠٥

Member (Supervisor)

Signature:

Name: Dr. Kareema Abdul- Kaddum

Title: Lecture

Date: / /٢٠٠٥

Approved by the Dean of the College of Education

Signature:

Name: Dr. Abdul- Ilah Rizzoki Karbal

Title: Professor

Date: /   /٢٠٠٥

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor Dr. Kareema Abdul-Kaddum for her advice, guidance and encouragement that I received from here during the course of this investigation. I'm also acknowledged a great deal of gratitude for Assist. Prof. Dr.Lua'y Abdul-Hani Al-Suwidi for his continual encouragement and support.

Thanks are due to the staff of the Department of Mathematics/ College of Education/ University of Babylon for their cooperation.

Also I thank all those helped me very much in the fulfillment of this work.

My thanks are due to my family for their spiritual support during the period of writing this thesis.

ABSTRACT

The essence of asymptotic methods is approximation. The main object of this thesis is to give a unified derivation of some results and theorems.

Also, this thesis deals with asymptotic distributions that is the distributions we obtained by letting the time horizon (sample size) tends to infinity. The research methodology is theoretical. The present study consists of three chapters.

In chapter one we give general introduction, review of literature and out line of this thesis.

Chapter two contains the key tools of asymptotic analysis and presents limit theorems for a univariate case (for example about sequence of random variables).

Chapter three contains the key tools of asymptotic analysis and presents limit theorems for a multivariate case (for example about sequence of random vectors).

Furthermore, we obtain some results about the asymptotic theory.

CHAPTER ONE

INTRODUCTION

In this chapter, we give three sections. Section one includes a general introduction, section two includes the review of literature and section three includes the out line of this thesis.

١,١- GENERAL INTRODUCTION

Exact distribution theory is limited to very special cases (normal independent identically distributed (i.i.d.) errors linear estimators), or involves very difficult calculations. This is too restrictive for applications. By making approximations based on large sample sizes, we can obtain distribution theory that is applicable in a much wider range of circumstances. These approximations are sometimes quite accurate and can often be constructed without a complete specification of the population distribution for the data. Suppose $F_n(x)$ is the (unknown) cumulative distribution function for some statistic based on a sample of size *n*. If it can be shown that the sequence of functions $F_1(x), F_2(x), \ldots$ converges rapidly to a known limit $F(x)$ as *n* tends to infinity, then we might use $F(x)$ as an approximation to $F_n(x)$ even for moderate values of *n*. The quality of the approximation depends on the speed of convergence, but can be checked by computer simulation.

The simplest example of this approach `is the average of independent draws from a distribution possessing a finite variance. Let $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$, where the X's are i.i.d. with population mean $= E(X_i) = \mu$ and population variance $= \mathrm{Var}(X_i) = \sigma^2$. By an easy calculation, we find that $\overline{X}_n$ has mean μ and variance σ^2 / n. Although the exact distribution of $\overline{X}_n$ depends on the distribution of X's, a simple asymptotic approximation is always available. The cumulative distribution function F_n for $\overline{X}_n$ is quite sensitive to the value of *n* so we would not expect the limit of the sequence F_1, F_2,… to yield a good approximation to F_n unless *n* is very large. But the standardized random variable $S_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ has mean zero and variance one for every *n*; its cumulative distribution function, say F_n^*, is much less sensitive to the value of *n*. Thus, if we could find the limit F^* of the sequence F_1^*, F_2^*, …, we might be willing to use it as an approximation to the distribution of S_n. The sequence F_1^*, F_2^*, … necessarily converges to the standard normal cumulative distribution function. This leads us to approximate F_n by the cumulative distribution function of a $N(\mu, \sigma^2 / n)$ distribution.

People may ask; since asymptotic distribution is only an approximation, why we are not using the exact distribution instead? Unfortunately, the exact finite sample distribution in many cases are too complicated to derive, even for Gaussian processes. Therefore, we use asymptotic distribution as alternatives.

Many estimators and test statistics used in econometrics are complicated nonlinear functions of the data and are not covered by the standard theorems which deal with sums of random variables. However, these statistics can often be approximated by sums and are still amenable to asymptotic analysis.

Most econometric methods used in applied economics are asymptotic in the sense that they are likely to hold only when the sample size is "large enough".

Asymptotic theory involves generalizing the usual notions of convergence for real sequences to allow for random variables. It is important to emphasis that limiting distributions obtained by central limit theorem (CLT) all involve unknown parameters which we seek to estimate.

We offer some of theorems, propositions and corollaries without proof.

This thesis develops the basic asymptotic results that will be used in subsequent chapters. We summarizes the key tools of the general introduction, review of literature and out line of this thesis in first chapter and the key tools of asymptotic analysis and presents limit theorems for a univariate case (for example about sequence of random variables) in second chapter and for a multivariate case (for example about sequence of random vectors) in third chapter and included some results in these chapters about the asymptotic theory.

١,٢- REVIEW OF LITERATURE

The joint asymptotic distribution of the sample mean and the sample median was found by Laplace almost ٢٠٠ years ago. See Stigler [١٣] for an interesting historical discussion of this achievement. After this, Ferguson, T.S. in [٢٣] derived the asymptotic joint distribution of the sample mean and an arbitrary quantile and he hoped that the proof may be new and of interest.

In (١٩٨٦) Irvine, J.M. [٢٥] derived the asymptotic distribution of the likelihood ratio test that is one technique for detecting a shift in the mean of a sequence of independent normal random variables. In (١٩٩٨) Rust ,J. [٣٠] studied the empirical process proof of the asymptotic distribution of sample quantiles.

In (١٩٩٩) Alves, M.I.F. [١٤] derived the asymptotic distribution of gumbel statistic in a semi-parametric approach (*) where this note is an answer to some open problems connected with recent developments for appropriate methodologies for making inferences on the tail of a distribution function (d.f.). Namely, in Fraga Alves and Gomes (١٩٩٦), the Gumbel statistic, based on the top part of a sample, is used in a semi-parametric approach, in order to fit an appropriate tail to the underlying model to a data set. The problem of statistical inference about external observations is handled there according to a test for choosing the most appropriate domain of attraction for the tail distribution, which gives preference to the Gumbel domain for the null hypothesis. The asymptotic behaviour of the referred statistic is derived therein under that null hypothesis and Alves, M.I.F. presented similar extended results under the alternative conditions, i.e., for d.f. that belongs to the other Generalized Extreme value domains, as an accomplishment to the promise made in last chapters of Fraga Alves and Gomes (١٩٩٥; ١٩٩٦).

In the same year Flinn, C. [٢٤] studied the asymptotic results for the linear regression model and Kelejian, H.H and Prucha, I. R. [٢٦] first gave a general result concerning the large sample distribution of Moran I type test statistic and applied this result to derive the large sample distribution of the Moran I test statistic for a variety of important models for which general spatial correlation testing procedures are not available.

In the same year Pötscher, B.M, and Prucha, I. R. [٢٩] provided a review of basic elements of asymptotic theory. Topics included modes of convergence, laws of large numbers and central limit theorems.

In (٢٠٠١) Sarno, E. [٣١] studied the asymptotic distribution of the Euclidean distance between time series models proposed by Piccolo (١٩٨٤, ١٩٩٠) in the case of Moving Averages models comparisons when least squares estimates are used for the unknown parameters. Piccolo showed that, when purely Autoregressive models are compared, the distribution of the distance estimator is a linear combination of independent Chi-squared random variables and invertible Moving Average models are considered to show that under the same assumptions stated by Berks (١٩٧٤) and Bhansali (١٩٧٨) for autoregressive model fitting of stationary processes, a similar result holds. Some descriptive statistics are then used to evaluate the effectiveness of such asymptotic result on finite samples sizes under specific model comparisons.

In (٢٠٠٤) Dufour, J.M. and Jouini, T.[٢١] studied the asymptotic distribution of a simple two-stage (Hannan-Rissanen-type) linear estimator for stationary invertible vector autoregressive moving average (VARMA) models in the echelon form representation. General conditions for consistency and asymptotic normality are given. A consistent estimator of the asymptotic covariance matrix of the estimator is also provided, so that tests and confidence-intervals can easily be constructed.

In the same year Bertola, M. [١٦] presented the formal properties of corrrelators of eigenvalues in the so-called planar limit (semiclassical) of various matrix models in terms of certain algebra-geometric data and Anderson, T.W. [١٥] studied the asymptotic distribution of a set of linear restrictions on regression coefficients where reduced rank regression analysis provided maximum likelihood estimators of a matrix of regression coefficients of a specified rank and of corresponding linear restrictions on such a matrix. These estimators depended on the eigenvectors of an "effect" matrix in the metric of an error covariance matrix and shown that the maximum likelihood estimator of the restrictions can be approximated by a function of the effect matrix alone. The procedures are applied to a block of simultaneous equations. The block may be over-identified in the entire model and the individual equations just-identified within the block.

١,٣- OUT LINE OF THIS THESIS

This thesis is divided in three chapters

Chapter one contains the general introduction, review of literature and out line of this thesis.

Chapter two contains the key tools of asymptotic analysis and presents limit theorems for a univariate case.

Chapter three contains the key tools of asymptotic analysis and presents limit theorems for a multivariate case.

Furthermore, we obtain some results about the asymptotic theory.

CHAPTER TWO
UNIVARIATE CONVERGENCE

The purpose of this chapter is to introduce many basic concepts, definitions and results which find use in this chapter and in the later chapter.

٢,١- Basic Concepts and Definitions

Definition (*Limit of Sequence*) (٢,١,١) [١]

Let $\{A_n\}$ be a sequence of sets. The set of all points *w* in the set Ω that belong to A_n for infinitely many values of *n* is called the limit superior of the sequence $\{A_n\}$ and is written as $\limsup\limits_{n\to\infty} A_n$.

The set of all points that belong to A_n for all but a finite number of values of *n* is called the limit inferior of the sequence $\{A_n\}$ and is written as $\liminf\limits_{n\to\infty} A_n$.

If $\limsup\limits_{n\to\infty} A_n = \liminf\limits_{n\to\infty} A_n$,

we say that limit exists of the sequence $\{A_n\}$ and we denote it by $\lim\limits_{n\to\infty} A_n$

For an increasing sequence $A_{١} \subset A_{٢} \subset \ldots$, $\lim\limits_{n\to\infty} A_n = \bigcup\limits_n A_n$.

For a decreasing sequence $A_{١} \supset A_{٢} \supset \ldots$, $\lim\limits_{n\to\infty} A_n = \bigcap\limits_n A_n$

In the case of an arbitrary sequence of sets $A_{١}, A_{٢}, \ldots$, we have

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k, \qquad \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k.$$

Definition (*Probability Space*) (٢,١,٢) [٣٦]

A probability space consists of a triple (Ω, F, P) where

(i) Ω is a space of points *w*, called the sample space and sample points. It is a nonempty set that represents the collection of all possible outcomes of an experiment.

(ii) *F* is a σ-field of subsets of Ω. It includes the empty set as well as the set Ω and is closed under the set operations of complements and finite or countable unions and intersections. The elements of *F* are called *measurable events*, or simply *events*.

(iii) $P(.)$ is a probability measure on *F*; henceforth refer to *P* as simply a probability.

Definition (*Convex function*) (٢,١,٣) [٣٧]

A continuous function *f* with domain and counterdomain the real line is called convex if for every $x_\circ$ on the real line, there exists a line which goes through the point $(x_\circ, f(x_\circ))$ and lies on or under the graph of the function *f*.

i.e., A real valued function f is *convex* if

(i) $f(\alpha\, x + (1-\alpha)\, y) \le \alpha\, f(x) + (1-\alpha) f(y)$.

for all x, y and all $0 \le \alpha \le 1$. This holds if and only if

(ii) $f(\frac{1}{2}(x+y)) \le \frac{1}{2}[f(x) + f(y)]$

for all x, y. Note that (ii) holds if and only if

(iii) $f(s) \le \frac{1}{2}[f(s-r) + f(s+r)]$ for all r, s

Also

(iv) $f''(x) \ge 0$ for all x implies f is convex .

Definition (*Convergence of a Sequence*) (٢,١,٤) [٣٢], [٣]

Let $\{a_n\}_{n=1}^{\infty}$ be denote a sequence of constants. The sequence a_n converges to a limiting value a if every neighborhood of a contains all but a finite number of the full sequence; that is, for each $\varepsilon > 0$, there exists some n_ε such that $|a_n - a| \le \varepsilon$ for all $n \ge n_\varepsilon$.

Usual notation is $\lim_{n\to\infty} a_n = a$ or $a_n \to a$, as $n \to \infty$.

Example (٢,١,٥) [٣٤]

(١) A very important example of a convergent sequence is as follows.

For any real number c, $\left(1+\frac{c}{n}\right)^n \to e^c$, as $n \to \infty$ (*Eulers limit*).

(٢) Let $a_n = 1 - \frac{1}{n}$; $a_n \to$ ١, as $n \to \infty$

Definition (٢,١,٦) [٢٧]

The sequence of real numbers $\{a_n\}$ is asymptotically equivalent to the sequence $\{b_n\}$, written $a_n \sim b_n$, if $(a_n / b_n) \to$ ١.

Equivalently, $a_n \sim b_n$ if and only if.

$$\left|\frac{a_n - b_n}{a_n}\right| \to 0$$

The left–hand expression above is called the relative error in approximating a_n by b_n.

We introduce some concepts about order

Order

Conventions (٢,١,٧) [١٧]

We examine the relative behavior of two sequences a_n and b_n as $n \to \infty$. The following list illustrates how to verbalize the notation

(١) $a_n = o(b_n) \Rightarrow a_n$ is little o of b_n.

(٢) $a_n = O(b_n) \Rightarrow a_n$ is big O of b_n.

(٣) $a_n = o_p(b_n) \Rightarrow a_n$ is little o p of b_n.

(٤) $a_n = O_p(b_n) \Rightarrow a_n$ is big O p of b_n.

The equal sign is read "is" rather than "equals". It signifies a relation between the left –and right– hand sides.

The *O*, *o* Notation (٢,١,٨) [٣٦]

Before the discussion of the concept of convergence for random variable, we will give a quick review of ways of comparing the magnitude of two sequences. A notation that is especially useful for keeping track of the order of an approximation is the "big *O*, little *o* ".

Let $\{a_n\}$ and $\{b_n\}$ be two sequences of real numbers. We have the following three concept of comparison:

(١) $a_n = O(b_n)$ if the ratio a_n / b_n is bounded for large n, if there exists a number K and an integer $n(K)$ such that if $n \geq n(K)$, then $|a_n| < K|b_n|$.

(٢) $a_n = o(b_n)$ if the ratio a_n / b_n converges to ٠, as $n \to \infty$.

(٣) $a_n \sim b_n$ iff $a_n / b_n = c + o(١)$.

Example (٢,١,٩) [٣٦]

Taylor expansion of a function $f(.)$ about the value c can be stated as

$$f(x) = f(c) + (x-c)f'(c) + o(|x-c|) \quad as \quad x \to c.$$

Theorem (٢,١,١٠) (*Taylor*) [٣٦]

Let the function f have a finite nth derivatives $f^{(n)}$ everywhere in the open interval (a, b) and (n –١)th derivative $f^{(n-1)}$ continuous in the closed interval [a, b]. Let $x \in$ [a, b]. For each point $y \in [a, b]$, $y \neq x$, there exists a point z interior to the interval joining x and y such that:

$$f(y) = f(x) + \sum_{k=1}^{n-1} \frac{f^{(k)}(x)}{k!}(y-x)^k + \frac{f^{(n)}(z)}{n!}(y-x)^n.$$

Or

$$f(y) = f(x) + \sum_{k=1}^{n-1} \frac{f^{(k)}(x)}{k!}(y-x)^k + o\left(|y-x|^{n-1}\right) \text{as } y \to x.$$

٢,٢- Concepts of Convergence of a Sequence of Random Variables. [٣٦]

Consider a sequence of random variables X١(*w*), X٢(*w*), …defined on the probability space (Ω, *F*, *P*). The random variable X(*w*) is a transformation of Ω on to the real line. The sequence of random variables X١(*w*), X٢(*w*), … is usually denoted by X١, X٢, …or simple by $\{X_n\}$, *n* = ١, ٢, … .

A sequence of random variables, assuming that it converges, can either converge to a constant or to a random variable. In the case where $\{X_n\}$ converges to a random variable, say X, the distribution function of X is said to asymptotically approximate that of X_n.

There are several different modes of convergence of a sequence of random variables. These are "convergence in probability", "convergence with probability one", "convergence in *r*th mean" and "convergence in distribution".

Definition (*Convergence in Probability*) (٢,٢,١) [٣٢], [٣٤], [١٩]

A sequence of random variables X١, X٢, ..., is said to *converge in probability* to a random variable X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

or, equivalently

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1, \qquad \text{………..(٢,١)}$$

It is common to write $X_n \xrightarrow{P} X$ or $\underset{n\to\infty}{\operatorname{plim}} X_n = X$.

Remark (٢,٢,٢) [٣٢]

For all $\varepsilon > 0$, the quantities $P(|X_n - X| \geq \varepsilon)$ and $P(|X_n - X| < \varepsilon)$ are just constants Thus, when we discuss convergence in probability, we are basically dealing with sequences of numbers. Informally, one might say that $X_n \xrightarrow{P} X$ if the probability that X_n "stays away from" X gets sufficiently small as *n* gets large.

Example (٢,٢,٣)

Suppose X_n is normally distributed with mean $\mu_n = \mu + \frac{k}{n}$ and the variance $\sigma_n^2 = \frac{\sigma^2}{n}$. To show that $\{X_n\}$ converges in probability to μ_n, a fixed constant. Here we show the convergence of X_n to μ_n by obtaining $P(|X_n - \mu_n| < \varepsilon)$ directly. But as it becomes clear later, the result can be established much more easily using general results on convergence in probability. Since $X_n - \mu_n \sim N\left(\frac{k}{n}, \frac{\sigma^2}{n}\right)$ it is easily seen that

$$P\left(\left|X_n - \mu_n\right| < \varepsilon\right) = \phi\left(\frac{\varepsilon - \frac{k}{n}}{\frac{\sigma}{\sqrt{n}}}\right) - \phi\left(\frac{-\varepsilon - \frac{k}{n}}{\frac{\sigma}{\sqrt{n}}}\right),$$

where $\phi(.)$ represents the cumulative distribution function of a standard normal variate. But

$$\lim_{n\to\infty} \phi\left(\frac{\varepsilon - \frac{k}{n}}{\frac{\sigma}{\sqrt{n}}}\right) = \lim_{n\to\infty} \phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1, \quad \text{for any} \quad \varepsilon > 0$$

and

$$\lim_{n\to\infty} \phi\left(\frac{-\varepsilon - \frac{k}{n}}{\frac{\sigma}{\sqrt{n}}}\right) = \lim_{n\to\infty} \phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right) = 0, \quad \text{for any} \quad \varepsilon > 0$$

Therefore

$$\lim_{n\to\infty} P\left(\left|X_n - \mu_n\right| < \varepsilon\right) = 1,$$

as required

For a given value of ε, the rate of convergence of X_n to μ_n clear depends on k, σ and the shape of the distribution function $\phi(.)$. The larger the value of σ, the slower will be the rate of convergence of X_n to μ_n.

Theorem (٢،٢،٤) (*Continuity*) [٣٢]

Suppose that $X_n \xrightarrow{p} X$ and let $g : R \to R$ be a real continuous function. Then, $g(X_n) \xrightarrow{p} g(X)$

Proof

Suppose that $X_n \xrightarrow{p} X$. Since g is continuous, we know that for all $\varepsilon > 0$, there exists some $\delta_\varepsilon > 0$ such that $\left|x_n - x\right| < \delta_\varepsilon \Rightarrow \left|g(x_n) - g(x)\right| < \varepsilon$ (definition of Continuity of a function). Define the events $A = \{x : |x_n - x| < \delta_\varepsilon\}$ and $B = \{x : \left|g(x_n) - g(x)\right| < \varepsilon\}$. Clearly, $A \subseteq B$ Furthermore, we have

$P(|X_n - X| < \delta_\varepsilon) = P(X_n \in A) \le P(X_n \in B) = P\{|g(X_n) - g(X)| < \varepsilon\}$. But, since $X_n \xrightarrow{p} X$, it must be that

$$P(|X_n - X| < \delta_\varepsilon) \to 1.$$

Hence, the result follows. ■

Definition (*Convergence With Probability One*) (٢,٢,٥) [٣٦]

Consider random variables X١, X٢, … and X, we say that X_n *Converges with Probability One* (or strongly, almost surely, almost every where, … etc.) to X if

$$P\left(w : \lim_{n\to\infty} X_n(w) = X(w)\right) = 1.$$

This is written $X_n \xrightarrow{\omega.p.1} X$, $n \to \infty$. To be better understanding this convergence, we give the following equivalent condition:

$$\lim_{n\to\infty} P(|X_m - X| < \varepsilon, \ \text{for all} \ \ m \ge n) = 1, \ \text{for every} \ \varepsilon > 0 \ldots\ldots\ldots\ldots (٢,٢)$$

The concept of convergence in probability defined by (٢,١) is a special case of (٢,٢) (setting *m* = *n* in (٢,٢) delivers (٢,١). But as we shall see below the reverse is not necessarily true. The concept of convergence with probability ١ is stronger than convergence in probability and is often referred to as the “strong convergence” as compared to convergence in probability which is referred to as “weak convergence”.

Definition (*Indicator Function*) (٢,٢,٦) [٣٣]

I(*A*) = ١ if *A* is true.

I(*A*) = ٠ if *A* is not true.

Proposition (٢,٢,٧) [٣٣]

$E[I(X \in A)] = P(X \in A)$.

Proposition (٢,٢,٨) [٣٥]

If $X_n \xrightarrow{\omega.p.1} X$, then $X_n \xrightarrow{p} X$.

However, the converse is not true.

Theorem (٢,٢,٩) [٣٢]

Suppose that $X_n \xrightarrow{\omega.p.1} X$ and let $g : R \to R$ be a real continuous function. Then, $g(X_n) \xrightarrow{\omega.p.1} g(X)$.

Definition (*Convergence in rth Mean*) (٢,٢,١٠) [٣٦], [٥], [١٩]

We say that X_n *converges in rth mean* to X if

$$\lim_{n\to\infty} E|X_n - X|^r = 0.$$

This is written $X_n \xrightarrow{rth} X, \quad n \to \infty.$

Remark (٢,٢,١١) [٣٦]

The method we just use can be viewed as an application of the fact of *Convergence in rth Mean Implies Convergence in Probability.*

Theorem (٢,٢,١٢)

(i) $X_n \xrightarrow{rth} X \Rightarrow X_n \xrightarrow{p} X, \forall r > 0.$

(ii) $X_n \xrightarrow{sth} X \Rightarrow X_n \xrightarrow{rth} X, \forall s > r.$ [٣٣]

Proof

(i) For any $\varepsilon > 0$, we write

$$E|X_n - X|^r = \{E|X_n - X|^r I(|X_n - X| \le \varepsilon)\} + E\{|X_n - X|^r I(|X_n - X| > \varepsilon)\}$$

$$E|X_n - X|^r \ge E\{|X_n - X|^r I(|X_n - X| > \varepsilon)\} \ge \varepsilon^r P(|X_n - X| > \varepsilon)$$

and thus

$$P(|X_n - X| > \varepsilon) \le \varepsilon^{-r} E|X_n - X|^r \to 0, \quad n \to \infty.$$ [٣٦] ■

Remark (٢,٢,١٣) [٢٧]

The case r =٢ is referred to as "*convergence in mean square*" or "*convergence in quadratic mean*" and denoted as $X_n \xrightarrow{m.s.} X$ or $X_n \xrightarrow{q.m} X.$

Notice, however, that as the following example demonstrates convergence in probability does not necessarily imply convergence in quadratic mean.

Example (٢,٢,١٤) [٢٧]

Consider the following sequence of random variables:

$$X_n = \begin{cases} n \text{ with probability } ١/n \\ ٠ \text{ with probability } ١-١/n. \end{cases}$$

Then $X_n \xrightarrow{p} 0$. But $E(X_n) = 1$ and $\text{Var}(X_n) = n - 1$.

Definition (*Characteristic Function*) (٢,٢,١٥) [٧], [١١]

The characteristic function of a random variable X is defined by

$$\phi_X(\theta) = E(e^{i\theta X}), \quad -\infty < \theta < \infty$$

where $i=\sqrt{-1}$. The essential property of a characteristic function is that it uniquely determined by the distribution function. In particular, if X has a probability density function $f(x)$ so that

$$\phi_X(\theta)=\int_{-\infty}^{\infty} e^{i\theta X} f(x)dx = \int_{-\infty}^{\infty} e^{i\theta X} dF(x).$$

***Properties of Characteristic Functions (٢,٢,١٦)* [٧]**

For all real θ, we have

(i) $\phi_X(\theta)$ is continuous everywhere, i.e., $\phi_X(\theta)$ is a continuous function of θ in $[-\infty,\infty]$. Rather $\phi_X(\theta)$ is uniformly continuous in θ.

(ii) $\phi_X(\theta)$ is defined in every finite θ interval.

(iii) $\phi_X(0)=\int_{-\infty}^{\infty} dF(x)=1$

(iv) $\phi_X(\theta)$ and $\phi_X(-\theta)$ are conjugate quantities.

(v) $|\phi_X(\theta)| = \left|\int_{-\infty}^{\infty} e^{i\theta X} f(x)dx\right| \le \int_{-\infty}^{\infty} \left|e^{i\theta X}\right| f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1 = \phi_X(0)$.

Since $|\phi_X(\theta)|\le 1$, characteristic function $\phi_X(\theta)$ always exists.

Definition (*Convergence in Distribution*) (٢,٢,١٧)

Consider distribution functions F١(.), F٢(.), … and F(.). Let X١, X٢, … and X denote random variables (not necessarily on a common probability space) having these distributions, respectively. We say that X_n *converges in distribution* (or in law) to X if

$\lim_{n\to\infty} F_n(v) = F(v)$, for all v which are continuity points of F.

This is written $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{L} X$ or $F_n \xrightarrow{w} F$. [٣٦], [٣٧], [٥]

The limiting distribution function, F, is referred to as the asymptotic distribution of X_n, and provides the basis for approximating the distribution of X_n, as n increases without bounds.

In practice when the mean or variance of X_n increase with n, in deriving the asymptotic distribution of X_n it is necessary to consider the limiting distribution of normalized or rescaled random variable, $Z_n = \dfrac{X_n - \mu_n}{\sigma_n}$, where μ_n and σ_n are appropriate constants.

In general, we would like to say that the distribution of the random variables X_n converges to the distribution of X if

$F_n(x) = P(X_n < x) \to F(x) = F(X < x)$ for every $x \in R$. [٣٦]

Example (٢,٢,١٨) [٣٦]

Consider random variables X_n which take values ١ − $n^{-١}$ or ١ + $n^{-١}$

with probabilities ١/٢. We would want the values of X_n to be more and more concentrated about ١. Note that the distribution function of X_n is

$$F_n(x) = \begin{cases} 0, & x < 1 - n^{-1} \\ 1/2, & 1 - n^{-1} \le x < 1 + n^{-1} \\ 1, & x \ge 1 + n^{-1}. \end{cases}$$

By calculation, we have $F_n(x) \to F^*(x)$ as $n \to \infty$ where

$$F^*(x) = \begin{cases} 0, & x < 1 \\ 1/2, & x = 1 \\ 1, & 1 < x. \end{cases}$$

On the other hand, for the random variable X taking value ١ with probability ١. The distribution of X is

$$F(x) = \begin{cases} 0, & x < 1 \\ 1, & x \ge 1. \end{cases}$$

Apparently, not much should be assumed about what happens for x at a discontinuity point of $F(x)$. Therefore, we can only consider convergence in distribution at continuity points of F. Read Example ١٤,٣,٢ (pp ٤٦٧) of [٢]. Another important tool for establishing convergence in distribution is to use moment-generating function or characteristic function. The characteristic function is used most often. Read example ١٤,٣,٣ (pp ٤٦٧) of [٢].

Theorem (٢,٢,١٩) [١٢]

Let the distribution functions F, F_1, F_2, … possess respective characteristic functions ϕ, ϕ_1, ϕ_2, … . The following statements are equivalent:

(i) $F_n \xrightarrow{w} F$ (or $X_n \xrightarrow{d} X$);

(ii) $\lim_{n\to\infty} \phi_n(\theta) = \phi(\theta)$, for each real θ;

(iii) $\lim_{n\to\infty} \int g \, dF_n = \int g \, dF$, for each bounded continuous function g.

Theorem (٢,٢,٢٠) [٣٦], [٢٨]

If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$.

Proof:

For $\varepsilon > 0$ and x is a point of continuity of F, we have

$$P(X_n \le x) \ge P(X_n \le x, X \le x - \varepsilon) \qquad \text{……..(٢,٣)}$$

because the first even is larger. Then we use the basic relation

$$P(A) = P(A \cap B) + P(A \cap B^c),$$

which holds for any events *A, B*. Take $A=\{X \le x-\varepsilon\}$ and $B=\{X_n \le x\}$. It follows that $A \cap B^c \subseteq \{|X_n - X| \ge \varepsilon\}$ and so

$$P(X \le x-\varepsilon) \le P(X_n \le x, X \le x-\varepsilon) + P(|X_n - X| \ge \varepsilon) \quad .... (٢,٤)$$

Combining (٢,٣) and (٢,٤) establishes

$$P(X \le x-\varepsilon) \le P(X_n \le x) + P(|X_n - X| \ge \varepsilon),$$

$$F(x-\varepsilon) \le F_n(x) + P(|X_n - X| \ge \varepsilon).$$

Since $X_n \xrightarrow{p} X$, we have $P(|X_n - X| \ge \varepsilon) \to 0$.

Thus $F(x-\varepsilon) \le \liminf_{n\to\infty} F_n(x)$. By a similar argument, we have

$\limsup_{n\to\infty} F_n(x) \le F(x+\varepsilon)$. Since x is a point of continuity of F, we have

$$\liminf_{n\to\infty} F_n(x) = \limsup_{n\to\infty} F_n(x) = F(x). \blacksquare$$

Remark (٢,٢,٢١) [٣٤]

Generally, $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$; however, if X is not a degenerate random variable (i.e., not constant), then

$$X_n \xrightarrow{d} X \text{ does not implies } X_n \xrightarrow{p} X.$$

Example (٢,٢,٢٢) [٣٤]

$X_n = Z$ with probability $(n-١)/n$

$X_n = n$ with probability $١/n$.

Therefore $\operatorname{plim}_{n\to\infty} X_n = Z$; $X_n \xrightarrow{p} Z$; $X_n \xrightarrow{d} Z$,

convergence in probability implies that convergence in distribution.

Proposition (٢,٢,٢٣) [٣٥]

$X_n \xrightarrow{p} c \Leftrightarrow X_n \xrightarrow{d} c$ where c is a constant

Example (٢,٢,٢٤) [٣٤]

$X_n = ٠$ with probability $١ - (١/n)$

$X_n = n$ with probability $١/n$

$X_n \xrightarrow{p} 0; X_n \xrightarrow{d} 0$

X_n has non stochastic $\operatorname{plim}_{n\to\infty}(0)$

We introduce some definitions about order in probability.

Definition "*Big*" O_p (٢,٢,٢٥) [٢٠]

Let $\{X_n\}$ denote a sequence of random variables. X_n is at *most of order in probability* n^k if, for all $\varepsilon > 0$, there exist constants $n_\varepsilon, M_\varepsilon > 0$ such that

$$P(n^{-k}|X_n| < M_\varepsilon) > 1-\varepsilon$$

for all $n > n_\varepsilon$. The notation is $X_n = O_p(n^k)$.

Example (٢,٢,٢٦) [٢٠]

If k = -١/٢, then $n^{1/2}X_n$ stays bounded as n gets large with high probability. In particular, with high probability, X_n is bounded by $M_\varepsilon n^{-1/2}$. This must mean that X_n itself is getting "small" as n gets large.

If k = ٠, then $X_n = O_p(1)$. From the above definition, this says that X_n remains bounded by the constant M_ε for n large with high probability. In this case, X_n is said to be *bounded in probability*.

Definition (٢,٢,٢٧) [١٧], [٣٣]

Let $\{a_n\}$ be a sequence of real numbers and let $\{X_n\}$ be a sequence of random variables. $X_n = O_p(a_n)$ if $X_n/\,a_n = O_p$ (١).

The above definition can be generalized for two sequences of random variables $\{X_n\}$ and $\{Y_n\}$. The notation $X_n = O_p(Y_n)$ denotes that the sequence $\left\{\frac{X_n}{Y_n}\right\}$ is $O_p(1)$.

Definition "*Little*" o_p (٢,٢,٢٨) [٢٠]

Let $\{X_n\}$ denote a sequence of random variables. X_n is said to be of *smaller order in probability* than n^k if

$$n^{-k}X_n \xrightarrow{p} 0, \text{ as } n \to \infty.$$

The notation is $X_n = o_p(n^k)$.

Example (۲,۲,۲۹) [۲۰]

(۱) The practical case of most interest to us is $k = ۰$. It is immediate that this is the same as $X_n \xrightarrow{p} 0$. Thus, writing $X_n = o_p(۱)$ is a shorthand way of saying that X_n converges in probability to zero. More generally, the case $k \leq ۰$ is the most interesting.

(۲) if $k = -1/2, X_n = o_p(n^{-1/2})$, then $n^{1/2} X_n \xrightarrow{p} 0$.

Definition (۲,۲,۳۰) [۱۷]

Let $\{X_n\}$ be a sequence of random variables. $X_n = o_p(1)$ if $X_n \xrightarrow{p} 0$. That is,

for every $\varepsilon > 0$, $\lim_{n \to \infty} P(|X_n| < \varepsilon) = 1$

or, equivalently, for every $\varepsilon > 0$ and for every $\eta > 0, \exists$ an integer $n(\varepsilon, \eta)$

Such that if $n >$ $n(\varepsilon, \eta)$ then

$P(|X_n| < \varepsilon) \geq 1 - \eta$.

One can say, informally, that $X_n = o_p(1)$ if $X_n = o(۱)$ with arbitrarily high probability.

Definition (۲,۲,۳۱) [۱۷], [۳۳]

Let $\{a_n\}$ be a sequence of real numbers and let $\{X_n\}$ be a sequence of random variables. $X_n = o_p(a_n)$ if $X_n / a_n = o_p$ (۱).

The above definition can be generalized for two sequences of random variables $\{X_n\}$ and $\{Y_n\}$. The notation $X_n = o_p(Y_n)$ means that $\frac{X_n}{Y_n} \xrightarrow{p} 0$.

Remark (۲,۲,۳۲) [۱۷]

(۱) $a_n \ o_p(۱) = o_p(a_n)$.

(۲) $a_n \ O_p(۱) = O_p(a_n)$.

(۳) $o_p(o_p(۱)) = o_p(۱)$.

Theorem (٢,٢,٣٣) [٣٣]

For a finite constant c,

$$X_n \xrightarrow{p} c \Rightarrow X_n = O_p(1).$$

For c = ٠,

$$X_n \xrightarrow{p} 0 \Leftrightarrow X_n = o_p(1).$$

Proposition (٢,٢,٣٤) [٣٥]

If X_n and Y_n are random variables defined in the same probability space and $a_n >$ ٠, $b_n >$ ٠, then

(i) If $X_n = o_p(a_n)$ and $Y_n = o_p(b_n)$, we have

$$X_n Y_n = o_p(a_n b_n).$$

$$X_n + Y_n = o_p(\max(a_n, b_n)).$$

$$|X_n|^r = o_p(a_n^r) \text{ for } r > 0.$$

(ii) If $X_n = o_p(a_n)$ and $Y_n = O_p(b_n)$, we have $X_n Y_n = o_p(a_n b_n)$.

Example (٢,٢,٣٥) [٣٥]

(١) If $X_n = o_p(n^k)$ and $Y_n = o_p(n^j)$, then $X_n Y_n = o_p(n^{k+j})$.

(٢) If $X_n = o_p(n^{-١})$ and $Y_n = o_p(n^{-٢})$, then $X_n + Y_n = o_p(n^{-١})$.

Remark (٢,٢,٣٦) [٣٥]

The notations above can be naturally extended from sequence of scaler to sequence of vector or matrix. In particular, $\underline{X}_n = o_p(n^k)$ if and only if all elements in $\underline{X}$ converges to zero at order n^{-k}. Using Euclidean distance $\|\underline{X}_n - \underline{X}\| = \left(\sum_{i=1}^{m}(X_{in} - X_i)^2\right)^{1/2}$, where m is the dimension of X_n.

Convergence in L_p Norm (٢،٢،٣٧) [٣٥]

When $E(|X_n|^p) < \infty$ with $p >$ ٠, X_n is said to be L_p*-bounded.* Define that the L_p norm of X is $\|X\|_p = (E|X|^p)^{1/p}$.

Before we define L_p convergence, we first review some useful inequalities.

Proposition (Cauchy- Schwarz Inequality) (٢،٢،٣٨) [٢٨], [٣٧]

$|E(XY)| \le E|XY| \le (E|X|^٢)^{١/٢} (E(|Y|^٢)^{١/٢}$.

Proposition (*Jensen's Inequality*) (٢،٢،٣٩) [٥]

Let X be random variable with finite mean E(X), and let f be a convex function. Then

$$E[f(X)] \ge f[E(X)]$$

Proposition (*Markov's Inequality*) (٢،٢،٤٠) [٣٥], [٣٧]

If $E|X|^p < \infty$, $p \ge$ ٠ and $\varepsilon > 0$, then

$$P(|X| \ge \varepsilon) \le \varepsilon^{-p} E|X|^p.$$

Remark (٢،٢،٤١) [٣٥]

In the Markov's inequality, we can also replace |X| with |X – c|, where c can be any real number.

Proposition (*Chebyshev's Inequality*) (٢،٢،٤٢) [٧], [١١]

If X is a random variable with mean μ and variance $\sigma^٢$, then for any positive number λ, we have

$$P(|X-\mu| \ge \lambda\sigma) \le \frac{1}{\lambda^2}.$$

Proposition (*Holder's Inequality*) (٢،٢،٤٣) [٣٥], [٢٨]

$$E|XY| \le \|X\|_p \|Y\|_q,$$

where $q = p/(p -$ ١$)$ if $p >$١ and $q = \infty$ if $p =$ ١.

Proposition (*Liapunov's Inequality*) (٢,٢,٤٤) [٣٥]

If $p > q > 0$, then $\|X\|_p \geq \|X\|_q$.

Proof

Let $Z = |X|^q$, Y = ١, $s = p/q$, then by Holder's inequality,

$E|ZY| \leq \|Z\|_s \|Y\|_{s/(s-1)}$, or

$E\left(|X|^q\right) \leq \left(E|X|^{qs}\right)^{1/s} = \left(E|X|^p\right)^{q/p}$. ■

Definition (L_p *Convergence*) (٢,٢,٤٥) [٣٥]

If $\|X_n\|_p < \infty$ for all n with $p >$ ٠ and $\lim_{n\to\infty}\|X_n - X\|_p = 0$, then X_n is said to *converge in L_p norm* to X, written $X_n \xrightarrow{L_p} X$. When p = ٢, we say it *converges in mean square*, written as $X_n \xrightarrow{m.s} X$.

For any $p > q > 0$, L_p convergences implies L_q convergence by Liaponov's inequality. We can take convergence in probability as an L٠ convergence, therefore, L_p convergence implies convergence in probability:

Proposition (L_p*convergence implies convergence in probability*)(٢,٢,٤٦)[٣٥]

If $X_n \xrightarrow{L_p} X$ then $X_n \xrightarrow{p} X$.

Proof

$P(|X_n - X| > \varepsilon)$

$\leq \varepsilon^{-p} E|X_n - X|^p$ by Markov's inequality

$\to$ ٠. ■

we introduce the relationships between the convergence in probability, convergence with probability one, convergence in distribution and L_p convergence.

$$
\begin{tikzcd}[column sep=small]
\text{convergence with probability ١} \arrow[dr] & & \text{L}_p \text{ convergence} \arrow[dl] \\
& \text{convergence in probability} \arrow[d] & \\
& \text{convergence in distribution} &
\end{tikzcd}
$$

The Law of Large Numbers (٢,٢,٤٧) [١٨]

Consider a sequence $\{X_n\}$ of random variables. The *strong law of large numbers* (SLLN) states conditions under which the average $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges with probability ١ to some constant μ. The corresponding convergence in probability result is called the *weak law of large numbers* (WLLN).

We shall first review the versions of the law for sequences of independently identically distributed (i.i.d.) random variables. Subsequently we consider the case where the random variables are independently distributed.

Theorem (*Weak Law of Large Numbers*) (٢,٢,٤٨) [٣٢], [٧]

Suppose that X١,X٢, …, are i.i.d. random variables with $E(X_i) = \mu$ and $\text{Var}\,(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, $\overline{X}_n \xrightarrow{p} \mu$. That is, the sample mean $\overline{X}_n$ converges in probability to μ.

Theorem (٢..٢,٤٩) [١٨]

Let $\{X_n\}$ be a sequence of independent random variables, with $E(X_i) = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$. Then a sufficient condition for

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow{p} 0$$

is that

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = 0.$$

Proof

Define $C_n = \sum_{i=1}^{n}(X_i - \mu_i)$ and $c_n^2 = \text{Var}(C_n) = \sum_{i=1}^{n}\sigma_i^2$, where the final equality follows from $E(C_n) = 0$ and independence of the X_i's. Chebyshev's inequality implies that for all $\varepsilon > 0$,

$$P\left(\left|\frac{1}{n}C_n\right| \geq \varepsilon\right) \leq \frac{c_n^2}{n^2\varepsilon^2},$$

and because the right-hand side converges to zero by assumption, it follows that $n^{-1}C_n \xrightarrow{p} 0$. ■

Theorem (٢,٢,٤٨) and Theorem (٢,٢,٤٩) give different conditions for the weak convergence of the sums of random variables.

Example (٢,٢,٥٠) [٣٢]

Suppose that X١, X٢, …, X_n~i.i.d. $f_X(x|p) = p^x(1\text{-}p)^{1\text{-}x}$, I (x = ٠, ١), where ٠ < p < ١. Define the statistic

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

for each n. From the WLLN, we know that $\hat{p}_n \xrightarrow{p} p$. It follows that

$$\log\left(\frac{\hat{p}_n}{1-\hat{p}_n}\right) \xrightarrow{p} \log\left(\frac{p}{1-p}\right).$$

The quantity log {p / (١− p)} is called the *log-odds* of p and is often abbreviated as logit (p).

Lemma (٢,٢,٥١) [٣٢]

Suppose that the non-random sequence $a_n \to a$ and that the stochastic sequence $X_n \xrightarrow{p} X$. Then, $a_n X_n \xrightarrow{p} aX$.

Example (٢,٢,٥٢) [٣٢]

Let $X_1, X_2, \ldots$ be n independent random variables with mean μ, common variance σ^2, and common third and fourth moments about their mean, μ_3 and μ_4, respectively (that is $\mu_r = E(X_i - \mu_i)^r$).
To show $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ converges to σ^2 in probability.

Solution

Straightforward algebra shows that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2)$$

by the WLLN, we know that $\frac{1}{n}\sum_{i=1}^{n}X_i^2 \xrightarrow{p} E(X_i^2) = \sigma^2 + \mu^2$.

We have that $\overline{X}^2 \xrightarrow{p} \mu^2$. Now just note that

$$\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right) = \underbrace{\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}X_i^2}_{\xrightarrow{p}\sigma^2+\mu^2} - \underbrace{\frac{n}{n-1}\overline{X}^2}_{\xrightarrow{p}\mu^2}.$$

Using the above Lemma with $a_n = n/(n-1)$, for $n > 1$, the result follows, i.e., we have $s^2 \xrightarrow{p} \sigma^2$ since $X_n + Y_n \xrightarrow{p} c + d$ when $X_n \xrightarrow{p} c$ and $Y_n \xrightarrow{p} d$.

Theorem (*Strong Law of Large Numbers*) (2,2,53) [32], [5]

Suppose that $X_1, X_2, \ldots$, are i.i.d. random variables with $E(X_i) = \mu$ and Var $(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$. Then, $\overline{X}_n \xrightarrow{w.p.1} \mu$. That is, the sample mean $\overline{X}_n$ converges with probability 1 to μ.

The Central Limit Theorem (2,2,54) [18]

Let $\{X_n\}$ be denote a sequence of random variables, with $E(X_i) = \mu_i$. and let $\left\{C_n = \sum_{i=1}^{n}(X_i - \mu_i)\right\}$ denote the sequence of partial sums. A fundamental weak convergence result is the *central limit theorem* (CLT), which states that under suitable conditions, the standardized sum $c_n^{-1}C_n$, with, $c_n^2 = \text{Var}(C_n)$, converges in distribution to a standard normal random variables Z, denoted by $c_n^{-1}C_n \xrightarrow{d} Z$ or $c_n^{-1}C_n \xrightarrow{d} N(0,1)$.

Before we proceed, we illustrate the theorem with two examples in Figures (٢,١) and (٢,٢). In the first of these, the density of $c_n^{-1}C_n$ is depicted for the case where $\{X_n\}$ is a sequence of independent random variables, uniformly distributed on the [٠, ١] interval (i.e, their probability density function is $f(x) = I_{[0,1]}(x)$). In Figure (٢,٢), the random variables X_i are

independent exponentially distributed with mean ١ (with probability density function $g(x) = e^{-x} I_{[0,\infty]}(x)$).

The figures illustrate that however for the density of X_i may be from the normal density, the process of summation yields a distribution which tends to the normal.

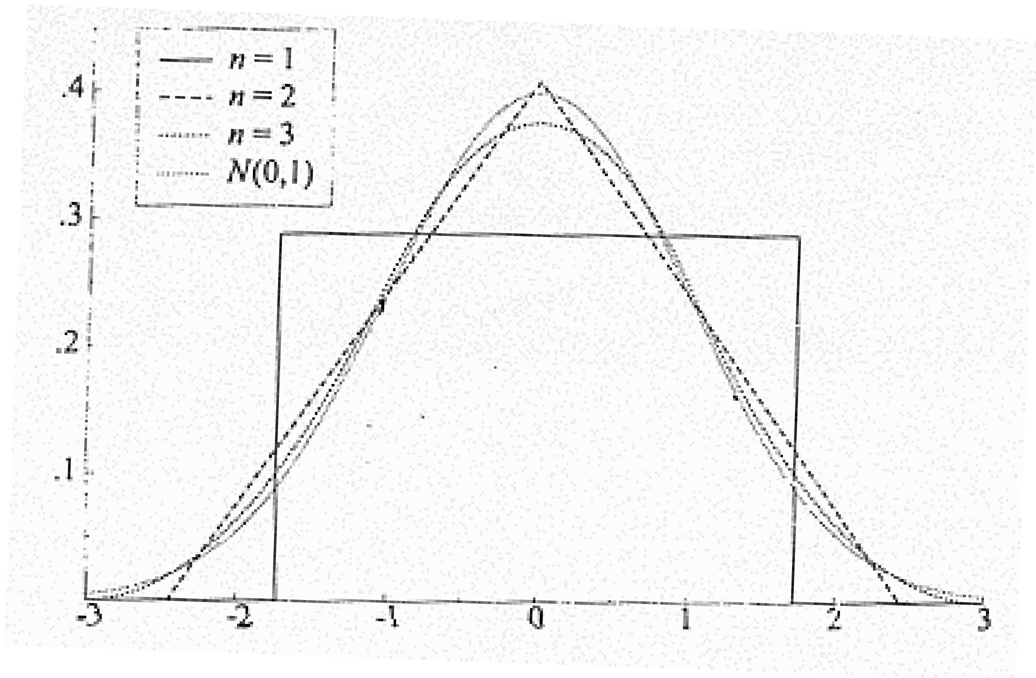

FIG. (٢,١): Probability density functions of $c_n^{-1}C_n$, uniform random variables

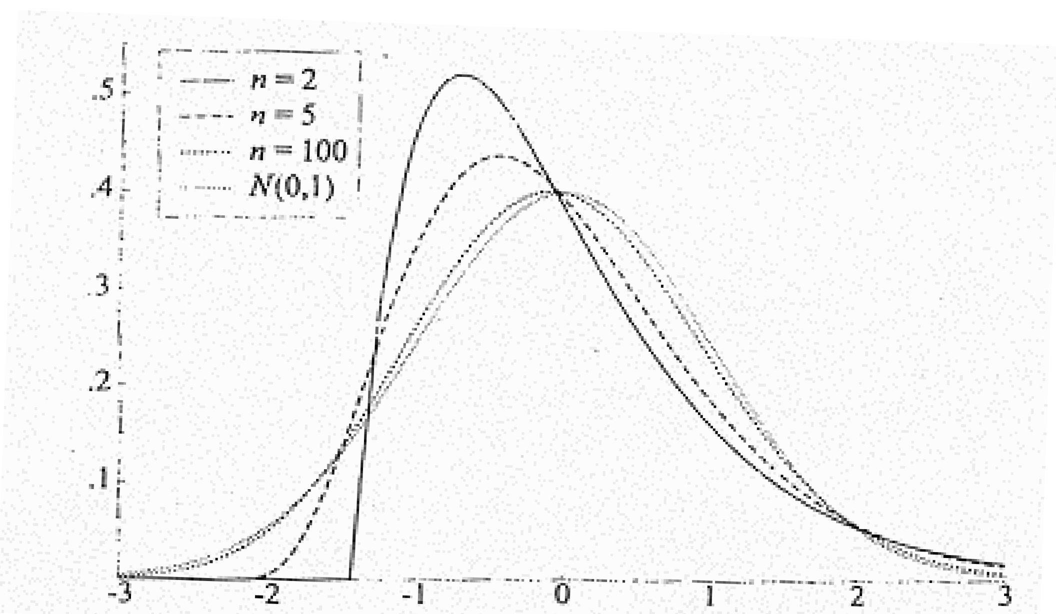

FIG. (٢,٢): Probability density function of $c_n^{-1}C_n$, exponential random variables

Definition (*Asymptotic Normality*) (٢,٢,٥٥) [٣٥]

A sequence of random variables $\{X_n\}$ is said to be asymptotic normal with mean μ_n and standard deviation σ_n if $\sigma_n > 0$ for n sufficiently large and

$$(X_n - \mu_n)/\sigma_n \xrightarrow{d} N(0,1).$$

Remark (٢,٢,٥٦) [٢٠]

In particular, μ_n is called the "*asymptotic mean*" and σ_n^2 is called the "*asymptotic variance*" such that $X_n \sim N(\mu_n, \sigma_n^2)$.

The basic CLT for i.i.d. random variables is as follows:

Theorem (*Lindeberg- Levy* Theorem) (٢,٢,٥٧) [١٨]

Let $\{X_n\}$ be a sequence of i.i.d. random variables, with $E(X_i)$ = ٠ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$, if $C_n = \sum_{i=1}^{n} X_i$, then

$$\frac{C_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Proof:

Let $\phi(\theta)$ be the characteristic function of X_i, and let $\phi_n(\theta)$ be the characteristic function of $C_n/(\sigma\sqrt{n})$. Then using the independence of X_i's, we have

$$\phi_n(\theta) = \left[\phi\left(\frac{\theta}{\sigma\sqrt{n}}\right)\right]^n.$$

Since the mean and variance of X_i exist, then we can write

$$\phi\left(\frac{\theta}{\sigma\sqrt{n}}\right) = 1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{\sigma^2 n}\right).$$

Therefore,

$$\log \phi_n(\theta) = n \log\left\{1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{\sigma^2 n}\right)\right\}$$

and

$$\lim_{n\to\infty} \log\{\phi_n(\theta)\} = \frac{-\theta^2}{2}$$

and hence $\phi_n(\theta) \to e^{-\frac{1}{2}\theta^2}$, as $n \to \infty$, which is the characteristic function of a standard normal variate, namely $\frac{C_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$. ■

A comparison of Theorem (٢,٢,٤٨) and Theorem (٢,٢,٥٧) clearly shows the additional assumption needed when moving from the WLLN to the CLT, namely that the CLT requires the existence of the second moments, while the WLLN for i.id. random variables only needs the existence of first moments.

Remark (٢,٢,٥٨) [٣٢]

It is common to write things like $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ and $\overline{X}_n \sim N(\mu, \sigma^2/n)$. Personally, we enjoy writing

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Example (٢,٢,٥٩) [٣٢]

Suppose that $X_n \sim$ binomial (n, p). Because $X_n = \sum_{i=1}^{n} Y_i$, where $Y_i \sim$ i.i.d. Bernoulli, $E(Y_i) = p$ and Var $(Y_i) = p(1-p) < \infty$, we have $\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1)$, or, equivalently $\sqrt{n}(\hat{p}_n - p) \xrightarrow{d} N\{0, p(1-p)\}$, where $\hat{p}_n = X_n / n = \overline{Y}_n$.

Proposition (٢,٢,٦٠) [٢٢]

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables such that $X_n - Y_n \xrightarrow{p} 0$ and $Y_n \xrightarrow{d} Y$, and let $g : R \to R$ be a continuous function. Then

(a) $X_n \xrightarrow{d} Y$;

(b) $g(X_n) - g(Y_n) \xrightarrow{p} 0$;

(c) $g(\mathrm{X}_n)\xrightarrow{d} g(\mathrm{Y})$.

Mann-Wald Approximation Theorem (٢,٢,٦١) [٣٨]

Let $\{\mathrm{X}_n, \mathrm{Y}_n\}$ be a sequence of pairs of random variables such that $\mathrm{X}_n - \mathrm{Y}_n \xrightarrow{p} 0$.

Then if $\mathrm{X}_n \xrightarrow{d} \mathrm{X} \Rightarrow \mathrm{Y}_n \xrightarrow{d} \mathrm{X}$.

Theorem (*Slutsky*) (٢,٢,٦٢) [٣٦], [٢٢]

Let $\mathrm{X}_n \xrightarrow{d} \mathrm{X}$ and $\mathrm{Y}_n \xrightarrow{p} \mathrm{c}$, where c is a constant not infinity. Then

(a) $\mathrm{X}_n + \mathrm{Y}_n \xrightarrow{d} \mathrm{X} + \mathrm{c}$;

(b) $\mathrm{X}_n \mathrm{Y}_n \xrightarrow{d} \mathrm{cX}$;

(c) $\dfrac{\mathrm{X}_n}{\mathrm{Y}_n} \xrightarrow{d} \dfrac{\mathrm{X}}{\mathrm{c}}$ if $\mathrm{c} \neq 0$.

Example (٢,٢,٦٣) [٣٢]

Suppose that X١, X٢, …, X_n are i.i.d. random variables with mean μ and variance σ٢. From CLT, we know that

$$\frac{\sum_{i=1}^{n} \mathrm{X}_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\overline{\mathrm{X}}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Furthermore, we know that $s^2 \xrightarrow{p} \sigma^2$, and hence, we have $\sigma/s \xrightarrow{p} 1$. Thus

$$\frac{\overline{\mathrm{X}}_n - \mu}{s/\sqrt{n}} = \underbrace{\frac{\sigma}{s}}_{\xrightarrow{p} 1} \times \underbrace{\frac{\overline{\mathrm{X}}_n - \mu}{\sigma/\sqrt{n}}}_{\xrightarrow{d} N(0,1)} \xrightarrow{d} N(0,1)$$

by Theorem (٢,٢,٦٢)

Example (٢,٢,٦٤) [٣٢]

We know that $\hat{p}_n \xrightarrow{p} p$ (WLLN) and that $\hat{p}_n(1-\hat{p}_n) \xrightarrow{p} p(1-p)$.

$$\sqrt{\frac{p(1-p)}{\hat{p}_n(1-\hat{p}_n)}} \xrightarrow{p} 1.$$

Therefore,

$$\frac{\hat{p}_n - p}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} = \underbrace{\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}}}_{\xrightarrow{d} N(0,1)} \times \underbrace{\sqrt{\frac{p(1-p)}{\hat{p}_n(1-\hat{p}_n)}}}_{\xrightarrow{p} 1} \xrightarrow{d} N(0,1)$$

by Theorem (٢,٢,٦٢).

We obtain some results about the asymptotic theory that is:

Lemma (٢,٢,٦٥)

Let $\{X_n\}$ be a sequence of independent random variables such that $X_{ni} \xrightarrow{d} X$, i = ١, …, m, where X is a random variable. Then $\sum_{i=1}^{m} X_{ni} \xrightarrow{d} mX$.

Proof:

Let $\phi_n(\theta)$ be the characteristic function of $\sum_{i=1}^{m} X_{ni}$, for any real θ. Then

$$\phi_n(\theta) = E\left(e^{i\theta \sum_{i=1}^{m} X_{ni}} \right)$$

and

$$\lim_{n\to\infty} \phi_n(\theta) = E\left(e^{i\theta m X}\right)$$

which is the characteristic function of mX and this is implies that

$\sum_{i=1}^{m} X_{ni} \xrightarrow{d} mX$. (by Theorem (٢,٢,١٩)(i),(ii)).

Theorem (٢,٢,٦٦)

Let $\{X_n\}$ be a sequence of independent random variables and $\{Y_n\}$ be a sequences of random variables such that $X_{ni} \xrightarrow{d} X$, i = ١, …, m and $Y_n \xrightarrow{p} c$, where X is a random variable and c is a constant not infinity. Then

(a) $\sum_{i=1}^{m} X_{ni} \pm Y_n \xrightarrow{d} mX \pm c$;

(b) $\sum_{i=1}^{m} X_{ni} Y_n \xrightarrow{d} mcX$;

(c) $\dfrac{\sum_{i=1}^{m} \mathrm{X}_{ni}}{\mathrm{Y}_n} \xrightarrow{d} \dfrac{m\mathrm{X}}{c}$ if $\mathrm{c} \neq 0$.

Proof

(a) Choose and fix v such that v–c is a continuity point of $\mathrm{F}_{m\mathrm{X}}(v)$. Let $\varepsilon > 0$ be such that $v - c + \varepsilon$ and $v - c - \varepsilon$ are also continuity points of $\mathrm{F}_{m\mathrm{X}}(v)$. Then

$$\mathrm{F}_{\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n}(v) = P\left(\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n \leq v \right)$$

$$\leq P\left(\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n \leq v, |\,\mathrm{Y}_n - c\,| < \varepsilon \right) + P(|\,\mathrm{Y}_n - c\,| \geq \varepsilon)$$

$$\leq P\left(\sum_{i=1}^{m} \mathrm{X}_{ni} \leq v - c + \varepsilon \right) + P(|\,\mathrm{Y}_n - c\,| \geq \varepsilon).$$

Hence by the hypotheses of the theorem, and by the choice of $v - c + \varepsilon$,

$$\limsup_{n\to\infty} \mathrm{F}_{\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n}(v) \leq \limsup_{n\to\infty} P\left(\sum_{i=1}^{m} \mathrm{X}_{ni} \leq v - \mathrm{c} + \varepsilon \right) + \limsup_{n\to\infty} P(|\,\mathrm{Y}_n - \mathrm{c}\,| \geq \varepsilon)$$

$$= \mathrm{F}_{m\mathrm{X}}\ (v - c + \varepsilon).$$

Since $\mathrm{Y}_n \xrightarrow{p} \mathrm{c}$, we have $\limsup_{n\to\infty} P(|\,\mathrm{Y}_n - \mathrm{c}\,| \geq \varepsilon) = 0$.

Similarly,

$$P\left(\sum_{i=1}^{m} \mathrm{X}_{ni} \leq v - \mathrm{c} - \varepsilon \right) \leq P\left(\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n \leq v \right) + P(|\,\mathrm{Y}_n - \mathrm{c}\,| \geq \varepsilon)$$

and thus

$$\mathrm{F}_{m\mathrm{X}}(v - c - \varepsilon) \leq \liminf_{n\to\infty} \mathrm{F}_{\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n}(v).$$

Since v – c is a continuity point of $\mathrm{F}_{m\mathrm{X}}(v)$, and since ε may be taken arbitrarily small, we have

$$\lim_{n\to\infty} \mathrm{F}_{\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n}(v) = \mathrm{F}_{m\mathrm{X}}(v - c) = \mathrm{F}_{m\mathrm{X+c}}(v).$$

This is follows that

$$\sum_{i=1}^{m} \mathrm{X}_{ni} + \mathrm{Y}_n \xrightarrow{d} m\mathrm{X} + c \text{ (by Definition (٢٫٢٫١٧))}.$$

And to proof the other direction, we have

Let $\phi_n(\theta)$ be the characteristic function of $\sum_{i=1}^{m} X_{ni} - Y_n$, for any real θ. Then

$$\phi_n(\theta) = E\left(e^{i\theta\left(\sum_{i=1}^{m} X_{ni} - Y_n\right)}\right) = e^{-i\theta Y_n} E\left(e^{i\theta \sum_{i=1}^{m} X_{ni}}\right)$$

and

$$\lim_{n\to\infty} \phi_n(\theta) = e^{-i\theta c} E(e^{i\theta m X})$$

which is the characteristic function of mX–c and this is implies that

$$\sum_{i=1}^{m} X_{ni} - Y_n \xrightarrow{d} mX - c \text{ (by theorem (٢،٢،١٩)(i),(ii)).}$$

(b) Let $Z_n = \sum_{i=1}^{m} X_{ni}(Y_n - c)$, and for arbitrary positive constants ε and δ, consider

$$P(|Z_n| > \varepsilon) = P\left(\left|\sum_{i=1}^{m} X_{ni}\right| |Y_n - c| > \varepsilon, |Y_n - c| < \frac{\varepsilon}{\delta}\right) +$$

$$P\left(\left|\sum_{i=1}^{m} X_{ni}\right| |Y_n - c| > \varepsilon, |Y_n - c| \geq \frac{\varepsilon}{\delta}\right)$$

$$\leq P\left(\left|\sum_{i=1}^{m} X_{ni}\right| \geq \delta\right) + P\left(|Y_n - c| \geq \frac{\varepsilon}{\delta}\right).$$

For any fixed δ, taking limits of both sides of the above inequality, and noting that by assumption $Y_n \xrightarrow{p} c$ and $\sum_{i=1}^{m} X_{ni} \xrightarrow{d} mX$, we have

$$\lim_{n\to\infty} P(|Z_n| > \varepsilon) \leq \lim_{n\to\infty} P\left(\left|\sum_{i=1}^{m} X_{ni}\right| > \delta\right) = P\left(\left|\sum_{i=1}^{m} X_{ni}\right| > \delta\right).$$

But δ is arbitrary and hence $P\left(\left|\sum_{i=1}^{m} X_{ni}\right| > \delta\right)$ can be made as small as desired by choosing a large enough value for δ. Therefore $\lim_{n\to\infty} P(|Z_n| > \varepsilon) = 0$ and $Z_n \xrightarrow{P} 0$. Hence by proposition (٢،٢،٦٠)(a), $\sum_{i=1}^{m} X_{ni} Y_n$ and $c\sum_{i=1}^{m} X_{ni}$ will have the same asymptotic distribution given by the distribution of mcX.

(c) Let $Z_n = \sum_{i=1}^{m} X_{ni}\left(\frac{1}{Y_n} - \frac{1}{c}\right)$, and for arbitrary positive constants ε and δ, consider

$$P\left(|Z_n| > \varepsilon\right) = P\left(\left|\sum_{i=1}^{m} X_{ni}\right|\left|\frac{1}{Y_n} - \frac{1}{c}\right| > \varepsilon, \left|Y_n - c\right| < \frac{\varepsilon}{\delta}\right) +$$

$$P\left(\left|\sum_{i=1}^{m} X_{ni}\right|\left|\frac{1}{Y_n} - \frac{1}{c}\right| > \varepsilon, \left|Y_n - c\right| \geq \frac{\varepsilon}{\delta}\right)$$

$$\leq P\left(\frac{\left|\sum_{i=1}^{m} X_{ni}\right|}{\left|cY_n\right|} \geq \delta\right) + P\left(\left|Y_n - c\right| \geq \frac{\varepsilon}{\delta}\right).$$

For any fixed δ, taking limits of both sides of the above inequality, and noting that by assumption $Y_n \xrightarrow{p} c$ and $\sum_{i=1}^{m} X_{ni} \xrightarrow{d} mX$, we have

$$\lim_{n\to\infty} P(|Z_n| > \varepsilon) \leq \lim_{n\to\infty} P\left(\frac{\left|\sum_{i=1}^{m} X_{ni}\right|}{\left|cY_n\right|} > \delta\right) = P\left(\frac{\left|\sum_{i=1}^{m} X_{ni}\right|}{\left|cY_n\right|} > \delta\right).$$

But δ is arbitrary and hence $P\left(\frac{\left|\sum_{i=1}^{m} X_{ni}\right|}{\left|cY_n\right|} > \delta\right)$ can be made as small as desired by choosing a large enough value for δ. Therefore $\lim_{n\to\infty} P(|Z_n| > \varepsilon) = 0$ and $Z_n \xrightarrow{P} 0$. Hence by proposition (٢,٢,٦٠)(a), $\frac{\sum_{i=1}^{m} X_{ni}}{Y_n}$ and $\frac{\sum_{i=1}^{m} X_{ni}}{c}$ will have the same asymptotic distribution given by the distribution of $\frac{mX}{c}$. ■

Example (٢,٢,٦٧)

Suppose that $\{X_n\}$ be a sequence of independent random variables and $\{Y_n\}$ be a sequences of random variables such that $X_{ni} \xrightarrow{d} X$, $i =$ ١, …,

m and $Y_n \xrightarrow{p} c$, where X~ N (μ,σ٢) and c is a constant not infinity and suppose that there are two constants such as a and b. Then

(١) $a\left(\sum_{i=1}^{m} X_{ni}\right) \pm bY_n \xrightarrow{d} N(ma\mu \pm bc, ma^2\sigma^2)$.

(٢) $ab\left(\sum_{i=1}^{m} X_{ni}Y_n\right) \xrightarrow{d} N(mabc\mu, ma^2b^2c^2\sigma^2)$.

(٣) $\dfrac{a\left(\sum_{i=1}^{m} X_{ni}\right)}{bY_n} \xrightarrow{d} N(\dfrac{ma\mu}{bc}, \dfrac{ma^2\sigma^2}{b^2c^2})$.

Theorem (٢,٢,٦٨)

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of independent random variables such that $X_{ni} \xrightarrow{d} X$, i = ١, …, m and $Y_n \xrightarrow{d} Y$, where X and Y are two random variables. Suppose X_n and Y_n are independent for $n \geq 1$. Then X and Y are independent, and

(a) $\sum_{i=1}^{m} X_{ni} \pm Y_n \xrightarrow{d} mX \pm Y$;

(b) $\sum_{i=1}^{m} X_{ni}Y_n \xrightarrow{d} mXY$;

(c) $\dfrac{\sum_{i=1}^{m} X_{ni}}{Y_n} \xrightarrow{d} \dfrac{mX}{Y}$.

Proof:

(a) Let $\phi_n(\theta)$ be the characteristic function of $\sum_{i=1}^{m} X_{ni} + Y_n$, for every real θ. Then

$$\phi_n(\theta) = E\left(e^{i\theta\left(\sum_{i=1}^{m} X_{ni} + Y_n\right)}\right) = E\left(e^{i\theta\sum_{i=1}^{m} X_{ni}} e^{i\theta Y_n}\right)$$

and taking limits we have:

$$\lim_{n\to\infty} \phi_n(\theta) = E\left(e^{i\theta mX} e^{i\theta Y}\right)$$

which is the characteristic function of mX+Y, and consequently

$$\sum_{i=1}^{m} X_{ni} + Y_n \xrightarrow{d} mX + Y \text{ (by Theorem (٢,٢,١٩)(i),(ii)).}$$

And to proof the other direction, we have

Let $\phi_n(\theta)$ be the characteristic function of $\sum_{i=1}^{m} X_{ni} - Y_n$, for every real θ. Then

$$\phi_n(\theta) = E\left(e^{i\theta\left(\sum_{i=1}^{m} X_{ni} - Y_n\right)} \right) = E\left(e^{i\theta \sum_{i=1}^{m} X_{ni}} e^{-i\theta Y_n} \right)$$

and

$$\lim_{n\to\infty} \phi_n(\theta) = E\left(e^{i\theta m X} e^{-i\theta Y}\right)$$

which is the characteristic function of mX–Y, and this is implies that

$$\sum_{i=1}^{m} X_{ni} - Y_n \xrightarrow{d} mX - Y \text{ (by Theorem (٢,٢,١٩)(i),(ii)).}$$

(b) Let $\phi_n(\theta)$ be denote the characteristic function of $\sum_{i=1}^{m} X_{ni} Y_n$, for every real θ. Then

$$\phi_n(\theta) = E\left(e^{i\theta\left(\sum_{i=1}^{m} X_{ni} Y_n\right)} \right).$$

Taking limits we have:

$$\lim_{n\to\infty} \phi_n(\theta) = E\left(e^{i\theta(mXY)}\right)$$

which is the characteristic function of mXY, and consequently

$$\sum_{i=1}^{m} X_{ni} Y_n \xrightarrow{d} mXY \text{ (by Theorem (٢,٢,١٩)(i),(ii)).}$$

(c) The proof is similar to that given above for (a), (b). ■

Example (٢,٢,٦٩)

Suppose that $\{X_n\}$ and $\{Y_n\}$ two sequences of independent random variables such that $X_{ni} \xrightarrow{d} X$, i = ١, …, m and $Y_n \xrightarrow{d} Y$, where X and Y are independent such that X~$N(\mu_١, \sigma_1^2)$ and Y~$N(\mu_٢, \sigma_2^2)$ and suppose that there are two constants such as a and b. Then

(١) $a\left(\sum_{i=1}^{m} X_{ni}\right) \pm bY_n \xrightarrow{d} N(ma\mu_1 \pm b\mu_2, ma^2\sigma_1^2 + b^2\sigma_2^2)$.

$(٢)\ ab\left(\sum_{i=1}^{m} X_{ni} Y_n\right) \xrightarrow{d} N(mab\mu_1\mu_2, ma^2b^2\sigma_1^2\sigma_2^2).$

The Delta Method

Remark (٢,٢,٧٠) [٣٢]

Suppose the real function $g: R \to R$ has derivatives of all orders at c. Then, the *Taylor series* for g at c is given by

$$g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(\mathrm{c})}{n!}(x-\mathrm{c})^n$$

If g is differentiable n times at x = c, then the polynomial

$$g_{r,c}(x) = \sum_{k=0}^{n} \frac{g^{(k)}(\mathrm{c})}{k!}(x-\mathrm{c})^k$$

is called the *nth Taylor polynomial* of g at c. The quantity $g(x) - g_{r,c}(x)$ is sometimes called the "remainder term".

Application (٢,٢,٧١) [٣٢]

Suppose that X is a random variable with mean μ and variance $\sigma^٢$. Let $g{:}R \to R$ be differentiable at μ, and consider the first order stochastic Taylor series expansion

$$g(\mathrm{X}) \approx g(\mu) + g'(\mu)\,(\mathrm{X}-\mu).$$

Taking expectations and variances, we have that

$$E\{\, g(\mathrm{X})\} \approx g(\mu)$$

and

$$\mathrm{Var}\ \{\, g(\mathrm{X})\} \approx \{\, g'(\mu)\}^٢\ \sigma^٢$$

Example (٢,٢,٧٢) [٣٢]

In Example (٢,٢,٥٠), we looked at the log-odds of $\hat{p}_n$. Here, we examine just the odds of $\hat{p}_n$; i.e., $g(\hat{p}_n) = \hat{p}_n/(1-\hat{p}_n)$. We are going to use the above result to find the asymptotic mean and variance of $g(\hat{p}_n)$. First, not that $E(\hat{p}_n) = p$ and Var $(\hat{p}_n) = p(١-p)/\,n$. Let $g(p) = p/(١-p)$. This is a differentiable function for $٠ < p < ١$, and it is easy to compute $g'(p) = ١/(١-p)^٢$. Thus,

$$\mathrm{E}\{g(\hat{p}_n)\} \approx g(p) = \frac{p}{1-p}$$

and

$$\text{Var}\,\{g(\hat{p}_n)\} \approx \{g'(p)\}^2\ \text{Var}\,(\hat{p}_n) = \left\{\frac{1}{(1-p)^2}\right\}^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}\,.$$

Remark (٢,٢,٧٣) [٣٢]

Taylor series expansion for the basis for the following useful generalization of the CLT-know as the *Delta Method*.

The Delta Method (٢,٢,٧٤) [٣٢], [٢٧]

Suppose that $\sqrt{n}(\mathrm{Y}_n - \theta) \xrightarrow{d} N(0,\sigma^2)$ and that $g{:}R \to R$ is differentiable at θ (and is not zero). Then,

$$\sqrt{n}\{g(\mathrm{Y}_n) - g(\theta)\} \xrightarrow{d} N[0, \{g'(\theta)\}^2 \sigma^2];$$

i.e., $g(\mathrm{Y}_n) \sim N[g(\theta), \{g'(\theta)\}^2 \sigma^2 / n]$.

Proof

Consider the stochastic expansion

$$g(\mathrm{Y}_n) = g(\theta) + g'(\theta)(\mathrm{Y}_n - \theta) + R_n(\theta)$$

Now, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\,\mathrm{Y}_n - \theta\,| \ge \varepsilon) = \lim_{n\to\infty} P(\underbrace{\sqrt{n}\,|\,\mathrm{Y}_n - \theta\,|}_{\xrightarrow{d}|Z|} \ge \sqrt{n}\varepsilon) = \lim_{n\to\infty} P(|\,\mathrm{Z}\,| \ge \sqrt{n}\varepsilon) = 0,$$

where $\mathrm{Z} \sim N(٠, \sigma^٢)$. Thus, it follows that $\mathrm{Y}_n \xrightarrow{p} \theta$, and hence, $\mathrm{Y}_n - \theta \xrightarrow{p} 0$. Also, all the terms in the remainder involving $(\mathrm{Y}_n - \theta)^k$, $k >$ ١, converges in probability to zero; hence $R_n(\theta) \xrightarrow{p} 0$. Dropping the remainder above becomes

$$\sqrt{n}\{g(\mathrm{Y}_n) - g(\theta)\} = g'(\theta)\underbrace{\sqrt{n}(\mathrm{Y}_n - \theta)}_{\xrightarrow{d} N(0,\sigma^2)} \to N[0, \{g'(\theta)^2\}\sigma^2]$$

Applying Theorem (٢,٢,٦٢) again to the right-hand side, the result follows.■

Example (٢,٢,٧٥) [٣٢]

Applying the Delta Method to $g(\hat{p}_n) = \hat{p}_n / (1 - \hat{p}_n)$ in Example (٢,٢,٧٢), we have that

$$\sqrt{n}\left\{\frac{\hat{p}_n}{1-\hat{p}_n} - \frac{p}{1-p}\right\} \xrightarrow{d} N\left\{0, \frac{p}{(1-p)^3}\right\},$$

or, in other words,

$$\frac{\hat{p}_n}{1-\hat{p}_n} \sim N\left\{\frac{p}{1-p}, \frac{p}{n(1-p)^3}\right\}.$$

Example (٢,٢,٧٦) [٣٢]

Suppose that $X_1, X_2, \ldots, X_n$ i.i.d. Poisson (θ). To find the asymptotic distribution of a properly centered and scaled version of exp $(-\overline{X}_n)$.

Solution

We know from the CLT that $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{d} N(0,\theta)$. Furthermore, $g(x) = e^{-x}$ is differentiable at θ (and is not zero). Thus, we can apply the Delta Method. Noting that $g'(\theta) = -e^{-\theta}$, we have

$$\sqrt{n}\left\{\exp(-\overline{X}_n) - \exp(-\theta)\right\} \xrightarrow{d} N\left\{0, \theta\exp(-2\theta)\right\};$$

i.e., $e^{-\overline{X}_n} \sim N(e^{-\theta}, \theta e^{-2\theta}/n)$.

Example (٢,٢,٧٧) [٨]

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. with mean $\mu \neq 0$ and variance σ^2. To find the asymptotic distribution of $1/\overline{X}_n$.

Solution

We know the CLT that $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0,\sigma^2)$.

Also, $g(x) = 1/x$ is continuous at $x = \mu$. Thus, we can apply the Delta Method. Noting that $g'(\mu) = -1/\mu^2$, we have

$$\sqrt{n}\{1/\overline{X}_n - 1/\mu\} \xrightarrow{d} N\{0, \sigma^2/\mu^4\};$$

i.e., $1/\overline{X}_n \sim N(1/\mu, \sigma^2/n\mu^4)$.

CHAPTER THREE
MULTIVARIATE CONVERGENCE

In this chapter, we cover thoroughly many concepts, definitions and results which find use in this chapter.

٣,١- Multivariate Notions of Convergence [٢٧]

We now consider vectors $\underline{x} = (x_1, x_2, ..., x_k)' \in R^k$ where R^k is the k-dimensional Euclidean space. We must define a norm on R^k. We are interested primarily in the norm of a vector tent to zero, a concept for which any norm will suffice, so we may as well as take the Euclidean norm:

$$\|\underline{x}\| = \left(\sum_{i=1}^{k} x_i^2\right)^{1/2}.$$

Definition (٣,١,١) [١٨], [٢٧]

A function $g : R^k \to R^\ell$ is said to be continuous at $\underline{c} \in R^k$ if for all $\varepsilon > 0$, there is a $\delta > 0$ such that $\|\underline{x} - \underline{c}\| < \delta$ implies $\|g(\underline{x}) - g(\underline{c})\| < \varepsilon$. If $\underline{x}_n$ is a sequence of vectors in R^k converging to $\underline{x}$, then continuity at $\underline{x}$ immediately implies $g(\underline{x}_n) \to g(\underline{x})$.

Definition (٣,١,٢) [٣٢]

A k-vector (non-random) sequence $\underline{x}_n = (x_{1n}, x_{2n}, ..., x_{kn})'$ converges to $\underline{c} = (c_1, c_2, ..., c_k)'$ if the Euclidean distance between $\underline{x}_n$ and $\underline{c}$; (i.e., $\|\underline{x}_n - \underline{c}\|$, where the notation $\|\underline{u}\| = \left(\sum_{i=1}^{k} u_i^2\right)^{1/2}$), tends to zero, as $n \to \infty$. Of course, this is equivalent to the convergence of the coordinate sequences x_{in} to c_i for i= ١, ٢, …, k.

٣,٢- Concepts of Convergence of a Sequence of Random Vectors.

The concepts in previous chapter are readily extended to multivariate cases where $\{\underline{X}_n\}$ denote a sequence of k-dimensional random vectors, i.e., the concepts extend immediately to vectors and matrices of finite dimension. [٣٨]

For a random vector $\underline{X} = (X_1, X_2, ..., X_k)' \in R^k$, the distribution function of $\underline{X}$, defined for $\underline{x} = (x_1, x_2, ..., x_k)' \in R^k$, is denoted by:

$F_{\underline{X}}(\underline{x}) = P(X_1 \le x_1, ..., X_k \le x_k) \in R^k$. [٤], [٦]

We want to talk about convergence of random vectors $\underline{X}_n$ to a limiting random vector $\underline{X}$ (a gain, usually a constant vector $\underline{c}$). [٣٢]

There are several different modes of convergence of a sequence of random vectors. These are "convergence in probability", "convergence almost surely", "convergence in *r*th mean" and "convergence in distribution".

Definition (*Convergence in Probability*) (٣,٢,١) [٣٢]

A sequence of random vectors $\underline{X}_1$, $\underline{X}_2$, …, is said to *converge in probability* to a random vector $\underline{X}$ if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\|\underline{X}_n - \underline{X}\| \geq \varepsilon\right) = 0,$$

or, equivalently,

$$\lim_{n\to\infty} P\left(\|\underline{X}_n - \underline{X}\| < \varepsilon\right) = 1.$$

We write $\underline{X}_n \xrightarrow{p} \underline{X}$ or, equivalently, $\operatorname*{plim}_{n\to\infty} \underline{X}_n = \underline{X}$.

Remark (٣,٢,٢) [٣٨]

Convergence in probability is said to hold for a random vector if it holds for each its components

$$\underline{X}_n = (X_{1n}, \ldots, X_{kn})', \quad \underline{X} = (X_1, \ldots, X_k)'$$

$$\underline{X}_n \xrightarrow{P} \underline{X} \text{ if } X_{in} \xrightarrow{P} X_i \ (i = 1, \ldots, k).$$

Proposition (٣,٢,٣) [٣٥]

If $\{\underline{X}_n\}$ is a sequence of *k*–dimensional random vectors such that $\underline{X}_n \xrightarrow{P} \underline{X}$ and if $g : R^k \to R^\ell$ is a continuous mapping, then

$$g(\underline{X}_n) \xrightarrow{p} g(\underline{X}).$$

Example (٣,٢,٤)

Suppose that $X_{1n} \xrightarrow{p} X_1$ and $X_{2n} \xrightarrow{p} X_2$. Then,

(١) $X_{1n} + X_{2n} \xrightarrow{P} X_1 + X_2$.

(٢) $X_{1n} X_{2n} \xrightarrow{P} X_1 X_2$.

(٣) $\dfrac{X_{1n}}{X_{2n}} \xrightarrow{P} \dfrac{X_1}{X_2}$.

These are straightforward applications of continuity.

Proposition (٣,٢,٥)

Let $\{\underline{X}_n\}$ denote a sequence of (k × ١) random vectors with plim $\underline{c}$, and let g ($\underline{c}$) be a vector–valued function, $g: R^k \to R^\ell$, where $g(.)$ is continuous at $\underline{c}$ and does not depend on n. Then $g(\underline{X}_n) \xrightarrow{p} g(\underline{c})$. [٨]

Proof

Since g is continuous at $\underline{c}$, then for every $\varepsilon > 0$, there exists a constant $\delta > 0$ such that

$$\|\underline{x} - \underline{c}\| < \delta \Rightarrow \|g(\underline{x}) - g(\underline{c})\| < \varepsilon$$

so that

$$P(\|\underline{X}_n - \underline{c}\| < \delta) \le P(\|g(\underline{X}_n) - g(\underline{c})\| < \varepsilon).$$

But $\underline{X}_n \xrightarrow{p} \underline{c}$, so $P(\|\underline{X}_n - \underline{c}\| < \delta) \to 1$. This is implies that

$$P(\|g(\underline{X}_n) - g(\underline{c})\| < \varepsilon) \to 1.$$

Or, $g(\underline{X}_n) \xrightarrow{P} g(\underline{c})$. [٣٦] ■

Example (٣,٢,٦)

If $X_{1n} \xrightarrow{p} c_1$ and $X_{2n} \xrightarrow{p} c_2$, then

(١) $X_{1n} + X_{2n} \xrightarrow{p} c_1 + c_2$.

(٢) $X_{1n}X_{2n} \xrightarrow{p} c_1c_2$.

(٣) $\dfrac{X_{1n}}{X_{2n}} \xrightarrow{p} \dfrac{c_1}{c_2}$, if $c_2 \ne 0$.

These are also straightforward applications of continuity.

Definition (*Convergence Almost Surely*) (٣,٢,٧)[٤], [٦]

The sequence $\{\underline{X}_n\}$ *converges almost surely to* $\underline{X}, \underline{X}_n \xrightarrow{a.s} \underline{X}$, if

$$P\{\lim \underline{X}_n = \underline{X}\} = 1.$$

Almost sure converges is sometimes called convergence with probability ١ (*w.p.*١) or strong convergence.

Equivalently, $\underline{X}_n \xrightarrow{a.s} \underline{X}$ iff for every $\varepsilon > 0$,

$$P\{\| \underline{X}_m - \underline{X} \| < \varepsilon, \text{for all } m \ge n\} \to 1 \text{ as } n \to \infty.$$

Theorem (٣،٢،٨) [٣٢]

Suppose that $\underline{X}_n \xrightarrow{a.s} \underline{X}$ and let $g: R^k \to R^\ell$ (often ℓ=١) be a real vectors-valued continuous function. Then, $g(\underline{X}_n) \xrightarrow{a.s} g(\underline{X})$.

Definition (*Convergence in rth Mean*) (٣،٢،٩) [٤], [٦]

For a real number $r >$ ٠, the sequence $\{\underline{X}_n\}$ *converges in the rth mean* to $\underline{X}, \underline{X}_n \xrightarrow{rth} \underline{X}$, if

$$E\|\underline{X}_n - \underline{X}\|^r \to 0, \text{ as } n \to \infty.$$

Remark (٣،٢،١٠) [٢٢]

$\underline{X}_n \xrightarrow{q.m} \underline{X}$ means that $E\{(\underline{X}_n - \underline{X})'(\underline{X}_n - \underline{X})\} \to 0$.

Definition (*Convergence in Distribution*) (٣،٢،١١) [٤], [٦]

The sequence $\{\underline{X}_n\}$ *converges in law* to $\underline{X}, \underline{X}_n \xrightarrow{L} \underline{X}$, if $F_{\underline{X}_n}(\underline{x}) \to F_{\underline{X}}(\underline{x})$ as $n \to \infty$, for all points $\underline{x}$ at which $F_{\underline{X}}(\underline{x})$ is continuous. It is sometimes called convergence in distribution or weak convergence.

Remark (٣،٢،١٢)

Convergence in distribution is usually denoted by $\underline{X}_n \xrightarrow{d} \underline{X}$.

Theorem (٣،٢،١٣) [٣٣]

$$\underline{X}_n \xrightarrow{d} \underline{X} \Leftrightarrow \phi_{\underline{X}_n}(\underline{u}) \to \phi_{\underline{X}}(\underline{u}), \forall \text{ vector } \underline{u}$$

$$\Leftrightarrow E\{g(\underline{X}_n) \to E\{g(\underline{X})\}, \forall \text{ bounded continuous } g.$$

Remark (٣،٢،١٤) [٣٥]

A simple way to verify convergence in distribution of a $k \times$ ١ vector is the following. If the scalar $(\lambda_1 X_{1n} + \lambda_2 X_{2n} + ... + \lambda_k X_{kn})$ converges in distribution to $(\lambda_1 X_1 + \lambda_2 X_2 + ... + \lambda_k X_k)$ for any real values of $(\lambda_1, \lambda_2, ..., \lambda_k)$, then the vector $\underline{X}_n = (X_{1n}, X_{2n}, ..., X_{kn})'$ converges in distribution to the vector $(X_1, X_2, ..., X_k)'$.

Theorem (*Cramer-Wold*) (٣،٢،١٥) [٣٣]

$$\underline{X}_n \xrightarrow{d} \underline{X} \Leftrightarrow \underline{\lambda}'\underline{X}_n \xrightarrow{d} \underline{\lambda}'\underline{X} \quad \forall \text{ vector } \underline{\lambda}. \qquad \text{.............(٣،١)}$$

Proof

$$\underline{\lambda}'\underline{X}_n \xrightarrow{d} \underline{\lambda}'\underline{X} \Rightarrow E(e^{i\theta\underline{\lambda}'\underline{X}_n}) \to E(e^{i\theta\underline{\lambda}'\underline{X}})$$
$$\Rightarrow \phi_{\underline{X}_n}(\underline{u}) \to \phi_{\underline{X}}(\underline{u}), \ \ \forall \ \text{vector}\ \underline{u}$$
$$\Rightarrow \underline{X}_n \xrightarrow{d} \underline{X}.$$

The converse is proved similarly. ■

Remark (٣,٢,١٦) [٣٣]

This result allow us to show first the usually much simpler rhs of (٣,١) in order to get a proof that the lhs holds. This is known as the *Cramer-Wold device*.

Theorem (٣,٢,١٧) [٣٣]

For vectors $\underline{X}_n, \underline{X}, \underline{c}$,

(i) $\underline{X}_n \xrightarrow{p} \underline{X} \Rightarrow \underline{X}_n \xrightarrow{d} \underline{X}$.

(ii) $\underline{X}_n \xrightarrow{p} \underline{c} \Leftrightarrow \underline{X}_n \xrightarrow{d} \underline{c}$.

Proposition (٣,٢,١٨) [٣٥]

If $\{\underline{X}_n\}$ is a sequence of random k-vectors with $\underline{X}_n \xrightarrow{d} \underline{X}$ and if $g: R^k \to R^\ell$ is a continuous function. Then $g(\underline{X}_n) \xrightarrow{d} g(\underline{X})$.

Theorem (٣,٢,١٩) [٢٧]

If $\underline{X}_n \xrightarrow{d} \underline{X} \in R^k$ and $\underline{Y}_n \xrightarrow{p} \underline{c} \in R^\ell$, then

$$g\begin{pmatrix}\underline{X}_n \\ \underline{Y}_n\end{pmatrix} \xrightarrow{d} g\begin{pmatrix}\underline{X} \\ \underline{c}\end{pmatrix} \qquad \text{..........(٣,٢)}$$

for any continuous function $g: R^{k+\ell} \to R^p$. Expression (٣,٢) will sometimes be written $g(\underline{X}_n, \underline{Y}_n) \xrightarrow{d} g(\underline{X}, \underline{c})$.

Corollary (٣,٢,٢٠) [٢٧]

If $\underline{X}_n \xrightarrow{d} \underline{X}$, then $X_{in} \xrightarrow{d} X_i$ for all i, $١ \le i \le k$.

Remark (٣,٢,٢١) [٢٧]

Theorem (٣,٢,١٩) may be strengthened if $\underline{X}_n$ is known to be independent of $\underline{Y}_n$, as follows

Theorem (٣،٢،٢٢) [٢٧]

If $\underline{X}_n \xrightarrow{d} \underline{X}$ and $\underline{Y}_n \xrightarrow{d} \underline{Y}$, where $\underline{X}_n$ is independent of $\underline{Y}_n$, then for a continuous function g,

$$g(\underline{X}_n, \underline{Y}_n) \xrightarrow{d} g(\underline{X}, \underline{Y}),$$

where $\underline{X}$ and $\underline{Y}$ are taken to be independent.

The proof is easy, since the independence of random vectors implies that $(\underline{a}, \underline{b})$ is a continuity point of the joint distribution if and only if $\underline{a}$ is a continuity point of $\underline{X}$ and $\underline{b}$ is a continuity point of $\underline{Y}$. Thus, for continuity points $\underline{a}$ of $\underline{X}$ and $\underline{b}$ of $\underline{Y}$,

$F_{(\underline{X}_n, \underline{Y}_n)}(\underline{a}, \underline{b}) = F_{\underline{X}_n}(\underline{a})\, F_{\underline{Y}_n}(\underline{b}) \to F_{\underline{X}}(\underline{a})\, F_{\underline{Y}}(\underline{b}) = F_{(\underline{X}, \underline{Y})}(\underline{a}, \underline{b})$.

Therefore, $(\underline{X}_n, \underline{Y}_n) \xrightarrow{d} (\underline{X}, \underline{Y})$, and so $g(\underline{X}_n, \underline{Y}_n) \xrightarrow{d} g(\underline{X}, \underline{Y})$ follows by Theorem (٣،٢،١٩). ■

The Multivariate Normal Distribution (٣،٢،٢٣) [٢٧]

Given a mean vector $\underline{\mu} \in R^k$ and a positive definite ($k \times k$) covariance matrix V, $\underline{X}$ has a multivariate normal distribution with mean $\underline{\mu}$ and covariance V, written $\underline{X} \sim N(\underline{\mu}, V)$, if its density on R^k is

$$f(\underline{x}) = \mathrm{C} \exp\left\{ -\frac{1}{2} (\underline{x} - \underline{\mu})' V^{-1} (\underline{x} - \underline{\mu}) \right\}. \qquad \ldots\ldots.(٣،٣)$$

In expression (٣،٣), the constant $\mathrm{C} = (2^k \pi^k |V|)^{-1/2}$, where $|V|$ denotes the determinant of V. Because of the assumption that V is positive definite.

As a special case, consider the bivariate normal distribution, where for some $\sigma_1^2 > 0, \sigma_2^2 > 0$ and $-1 < \rho_{12} < 1$ we have

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

thus

$$|V| = \sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2) \text{ and } V^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2)} \begin{pmatrix} \sigma_2^2 & -\rho_{12}\sigma_1\sigma_2 \\ -\rho_{12}\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

In this case, of course, X١ and X٢ have correlation ρ١٢ and marginal variances σ_1^2 and σ_2^2.

Theorem (٣،٢،٢٤) [٣٣], [٩]

$$\underline{X} \sim N(\underline{\mu}, V) \Leftrightarrow \underline{\lambda}'\underline{X} \sim N(\underline{\lambda}'\underline{\mu}, \underline{\lambda}'V\underline{\lambda}), \forall \text{ vector } \underline{\lambda}.$$

Proof

Note that

$$\underline{X} \sim N(\underline{\mu}, V) \Leftrightarrow \phi_{\underline{X}}(\underline{u}) = e^{i\underline{u}'\underline{\mu} - \frac{1}{2}\underline{u}'V\underline{u}}, \forall \text{ vector } \underline{u}$$

Now

$$\underline{\lambda}'\underline{X} \sim N(\underline{\lambda}'\underline{\mu}, \underline{\lambda}'V\underline{\lambda}) \Rightarrow \phi_{\underline{\lambda}'\underline{X}}(\theta) = e^{i\theta\underline{\lambda}'\underline{\mu} - \frac{1}{2}\theta^2\underline{\lambda}'V\underline{\lambda}}, \forall \text{ real } \theta, \text{ vector } \underline{\lambda}$$

$$= \phi_{\underline{X}}(\theta\underline{\lambda}) \Rightarrow \underline{X} \sim N(\underline{\mu}, V)$$

because $\underline{\lambda}$ is arbitrary.

Converse follows from the fact that linear combinations of $\underline{X}$ are normal.■

Remark (٣،٢،٢٥) [٣٣]

Theorem (٣،٢،٢٤) implies that to show a vector variable $\underline{X}_n$ is multi-normal, it is necessary and sufficient to show that all linear combinations are univariate normal. Incorporating Theorem (٣،٢،١٥), we see that to show that $\underline{X}_n$ is asymptotically multi-normal we show that all the linear combinations are asymptotically univariate normal,

$$\underline{\lambda}'\underline{X}_n \xrightarrow{d} \underline{\lambda}'\underline{X} \sim N(\underline{\lambda}'\underline{\mu}, \underline{\lambda}'V\underline{\lambda}) \Leftrightarrow \underline{X}_n \xrightarrow{d} \underline{X} \sim N(\underline{\mu}, V).$$

The Law of Large Numbers

Multivariate WLLN (٣،٢،٢٦) [٣٢]

Suppose that $\underline{X}_1, \underline{X}_2, \ldots,$ are i.i.d. random vectors with $E(\underline{X}_i) = \underline{\mu}$ and $\mathrm{Var}(\underline{X}) = V$. Define $\overline{\underline{X}}_n = (\overline{X}_{1n}, \overline{X}_{2n}, \ldots, \overline{X}_{kn})'$. Then $\overline{\underline{X}}_n \xrightarrow{p} \underline{\mu}$.

Example (٣،٢،٢٧)[٣٢]

Suppose that $\underline{X}_1, \underline{X}_2, \ldots,$ are i.i.d. bivariate normal random vectors with mean $\underline{\mu} = (\mu_1, \mu_2)'$ and variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}.$$

By the WLLN, the sequence $\overline{\underline{X}}_n = (\overline{X}_{1n}, \overline{X}_{2n})'$ converges in probability to $\underline{\mu} = (\mu_1, \mu_2)'$.

Multivariate SLLN (٣،٢،٢٨) [٣٢]

Suppose that $\underline{X}_1, \underline{X}_2, \ldots,$ are i.i.d. random vectors with $E(\underline{X}_i) = \underline{\mu}$ and $\mathrm{Var}(\underline{X}) = V$. Define $\overline{\underline{X}}_n = (\overline{X}_{1n}, \overline{X}_{2n}, \ldots, \overline{X}_{kn})'$. Then $\overline{\underline{X}}_n \xrightarrow{a.s} \underline{\mu}$.

Multivariate CLT (٣,٢,٢٩) [٣٢], [١٠]

Let $\underline{X}_1, \underline{X}_2, ...,$ denote a sequence of i.i.d. random k-vectors with mean $\underline{\mu} = (\mu_1, \mu_2, ..., \mu_k)'$ and variance-covariance matrix V (with finite determinant) .Define

$$\overline{\underline{X}}_n = \frac{1}{n}\sum_{j=1}^{n} \underline{X}_j \,.$$

Then, $\sqrt{n}(\overline{\underline{X}}_n - \underline{\mu}) \xrightarrow{d} N(\underline{0}, V)$.

Example (٣,٢,٣٠) [٣٢]

Suppose that $\underline{X}_1, \underline{X}_2, ...,$ are i.i.d. bivariate random vectors with mean $\underline{\mu} = (\mu_1, \mu_2)'$ and variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix},$$

where V is symetric matrix.

The multivariate CLT says that $\sqrt{n}(\overline{\underline{X}}_n - \underline{\mu}) \xrightarrow{d} N(\underline{0}, V)$;

i.e., $\overline{\underline{X}}_n \sim N(\underline{\mu}, V/n)$.

Writing $\underline{X}_j = (X_{1j}, X_{2j})'$, we have that $\overline{\underline{X}}_n = (\overline{X}_{1n}, \overline{X}_{2n})'$, where

$$\overline{X}_{in} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$

for i = ١, ٢, . Marginally, it follows that

$\overline{X}_{1n} \sim N(\mu_1, \sigma_1^2/n)$ and $\overline{X}_{2n} \sim N(\mu_2, \sigma_2^2/n)$.

Proposition (٣,٢,٣١) [٨]

Let $\{\underline{Y}_n\}$ be a sequence of (k × ١) random vectors with $\underline{Y}_n \xrightarrow{d} \underline{Y}$. Suppose that $\{\underline{X}_n\}$ is a sequence of (k × ١) random vectors such that $(\underline{X}_n - \underline{Y}_n) \xrightarrow{p} \underline{0}$. Then $\underline{X}_n \xrightarrow{d} \underline{Y}$; that is, $\underline{X}_n$ and $\underline{Y}_n$ have the same limiting distribution.

Slutsky's Theorem (٣,٢,٣٢) [٣٨], [٨]

Let $\{\underline{X}_n\}$ and $\{\underline{Y}_n\}$ be sequences of random vectors such that $\underline{X}_n \xrightarrow{d} \underline{X}$ and $\underline{Y}_n \xrightarrow{p} \underline{c}$ (constant). Then

(a) $\underline{X}_n + \underline{Y}_n \xrightarrow{d} \underline{X} + \underline{c}$.

(b) $\underline{Y}'_n \underline{X}_n \xrightarrow{d} \underline{c}'\underline{X}$.

Lemma (٣,٢,٣٣) [١٩]

Let $\{\underline{X}_n\}$ and $\{\underline{Y}_n\}$ be sequences of random ($k \times$ ١) vectors. Then:

(a) If $(\underline{X}_n - \underline{Y}_n) \xrightarrow{p} \underline{0}$ and $\underline{X}_n \xrightarrow{d} \underline{X} \Rightarrow \underline{Y}_n \xrightarrow{d} \underline{X}$.

(b) If $\underline{X}_n \xrightarrow{d} \underline{X}$ and $\underline{Y}_n \xrightarrow{p} \underline{0} \Rightarrow \underline{Y}'_n \underline{X}_n \xrightarrow{p} \underline{0}$.

Corollary (٣,٢,٣٤) [٣٣]

(i) $\underline{X}_n \xrightarrow{d} \underline{X},\ \mathrm{Y}_n \xrightarrow{p} \mathrm{C} \Rightarrow \mathrm{Y}_n \underline{X}_n \xrightarrow{d} \mathrm{C}\underline{X}, \mathrm{Y}_n$ matrix, $\underline{X}_n$ vector.

(ii) $\underline{X}_n \xrightarrow{d} \underline{X},\ \mathrm{Y}_n \xrightarrow{p} \mathrm{C} \Rightarrow \mathrm{Y}_n^{-1} \underline{X}_n \xrightarrow{d} \mathrm{C}^{-1}\underline{X}, |\mathrm{C}| \neq \underline{0}$.

We obtain some results about the asymptotic theory that is:

Lemma (٣,٢,٣٥)

Let $\{\underline{X}_n\}$ and $\{\underline{Z}_n\}$ be two sequences of ($k \times$ ١) independent random vectors such that $\mathrm{Z}_{1n} = \mathrm{X}_{1n1} + \mathrm{X}_{1n2} + \dots + \mathrm{X}_{1nm}$,
$\mathrm{Z}_{2n} = \mathrm{X}_{2n1} + \mathrm{X}_{2n2} + \dots + \mathrm{X}_{2nm}$ and $\mathrm{Z}_{kn} = \mathrm{X}_{kn1} + \mathrm{X}_{kn2} + \dots + \mathrm{X}_{knm}$. Then

$$\sum_{t=1}^{m} \underline{X}_{nt} = \underline{Z}_n$$

Proof

$$\sum_{t=1}^{m} \underline{X}_{nt} = \underline{X}_{n1} + \underline{X}_{n2} + \dots + \underline{X}_{nm}$$

$$= \begin{pmatrix} \mathrm{X}_{1n1} \\ \mathrm{X}_{2n1} \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{X}_{kn1} \end{pmatrix} + \begin{pmatrix} \mathrm{X}_{1n2} \\ \mathrm{X}_{2n2} \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{X}_{kn2} \end{pmatrix} + \dots + \begin{pmatrix} \mathrm{X}_{1nm} \\ \mathrm{X}_{2nm} \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{X}_{knm} \end{pmatrix}$$

$$= \begin{pmatrix} \mathrm{X}_{1n1} + \mathrm{X}_{1n2} + \ . \ . \ . + \mathrm{X}_{1nm} \\ \mathrm{X}_{2n1} + \mathrm{X}_{2n2} + \ . \ . \ . + \mathrm{X}_{2nm} \\ \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \\ \mathrm{X}_{kn1} + \mathrm{X}_{kn2} + \ . \ . \ . + \mathrm{X}_{knm} \end{pmatrix} = \begin{pmatrix} \mathrm{Z}_{1n} \\ \mathrm{Z}_{2n} \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{Z}_{kn} \end{pmatrix} = \underline{Z}_n. \ \blacksquare$$

Lemma (٣،٢،٣٦)

Let $\{\underline{X}_n\}$ be a sequence of (k × ١) independent random vectors. If $\underline{X}_{nt} \xrightarrow{d} \underline{X}$; t = ١,…, m, then $\sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\underline{X}$.

Proof

Denote the characteristic functions of $\underline{\lambda}' \sum_{t=1}^{m} \underline{X}_{nt}$ and $m\underline{\lambda}'\underline{X}$ by $\phi_{\underline{\lambda}' \sum_{t=1}^{m} \underline{X}_{nt}}(\theta)$ and $\phi_{m\underline{\lambda}'\underline{X}}(\theta)$, respectively,

when $\underline{\lambda} = (\underline{\lambda}_1, ..., \underline{\lambda}_k)'$ is an arbitrary vector of fixed constants and θ is any real.

By using Theorem (٢،٢،١٩)(i),(ii) and Theorem (٣،٢،١٥), we have

$$\phi_{\underline{\lambda}' \sum_{t=1}^{m} \underline{X}_{nt}}(\theta) = \phi_{\underline{\lambda}'\underline{Z}_n}(\theta) \qquad \text{(by Lemma (٣،٢،٣٥))}$$

$$= E\left(\mathrm{e}^{i\theta\underline{\lambda}' \underline{Z}_n}\right) = E\left(\mathrm{e}^{i\theta \sum_{i=1}^{k} \lambda_i Z_{in}}\right),$$

and

$$\lim_{n\to\infty} \phi_{\underline{\lambda}' \sum_{t=1}^{m} \underline{X}_{nt}}(\theta) = E\left(\mathrm{e}^{i\theta \sum_{i=1}^{k} \lambda_i Z_i}\right) = E\left(\mathrm{e}^{i\theta m \sum_{i=1}^{k} \lambda_i X_i}\right)$$

$$= E\left(\mathrm{e}^{i\theta m \underline{\lambda}'\underline{X}}\right) = \phi_{m\underline{\lambda}'\underline{X}}(\theta)$$

$$\lim_{n\to\infty} \phi_{\underline{\lambda}' \sum_{t=1}^{m} \underline{X}_{nt}}(\theta) = \phi_{m\underline{\lambda}'\underline{X}}(\theta), \forall \underline{\lambda} \in R^k, \underline{\lambda} \neq \underline{0}, \forall \theta \in R$$

$$\Rightarrow \underline{\lambda}' \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\underline{\lambda}'\underline{X}, \forall \underline{\lambda} \in R^k, \underline{\lambda} \neq \underline{0} \text{ (by Theorem (٢،٢،١٩)(i),(ii))}$$

$$\Rightarrow \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\underline{X} \text{ (by Theorem (٣،٢،١٥)).} \blacksquare$$

Theorem (٣،٢،٣٧)

Let $\{\underline{X}_n\}$ be a sequence of (k × ١) independent random vectors with $\underline{X}_{nt} \xrightarrow{d} \underline{X}$, t = ١,…, m, and let $\{\underline{Y}_n\}$ be a sequence of (k × ١) random vectors with $\underline{Y}_n \xrightarrow{p} \underline{c}$, where $\underline{c}$ be a vector of constants not infinity. Then

(a) $\sum_{t=1}^{m} \underline{X}_{nt} \pm \underline{Y}_n \xrightarrow{d} m\underline{X} \pm \underline{c}$;

(b) $\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\underline{c}'\underline{X}$.

Proof

(a) Choose and fix v such that $v - \sum_{i=1}^{k} \lambda_i c_i$ is a continuity point of $F_{m\sum_{i=1}^{k}\lambda_i X_i}(v)$. Let $\varepsilon > 0$ be such that $v - \sum_{t=1}^{k} \lambda_i c_i + \varepsilon$ and $v - \sum_{i=1}^{k} \lambda_i c_i - \varepsilon$ are also continuity points of $F_{m\sum_{i=1}^{k}\lambda_i X_i}(v)$.

Denote the distribution functions of $\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)$ and $\underline{\lambda}'(m\underline{X} + \underline{c})$ by $F_{\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)}(v)$ and $F_{\underline{\lambda}'(m\underline{X}+\underline{c})}(v)$, respectively,

when $\underline{\lambda} = (\lambda_1, ..., \lambda_k)'$ is an arbitrary vector of fixed constants. By using Definition (٢،٢،١٧) and Theorem (٣،٢،١٥), we have

$$F_{\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)}(v) = F_{\underline{\lambda}'(\underline{Z}_n + \underline{Y}_n)}(v)$$

$$= F_{\sum_{i=1}^{k}\lambda_i Z_{in} + \sum_{i=1}^{k}\lambda_i Y_{in}}(v)$$

$$= P(\sum_{i=1}^{k}\lambda_i Z_{in} + \sum_{i=1}^{k}\lambda_i Y_{in} \le v)$$

$$\le P\left(\sum_{i=1}^{k}\lambda_i Z_{in} + \sum_{i=1}^{k}\lambda_i Y_{in} \le v, \left|\sum_{i=1}^{k}\lambda_i Y_{in} - \sum_{i=1}^{k}\lambda_i c_i\right| < \varepsilon\right)$$
$$+ P\left(\left|\sum_{i=1}^{k}\lambda_i Y_{in} - \sum_{i=1}^{k}\lambda_i c_i\right| \ge \varepsilon\right)$$

$$\le P\left(\sum_{i=1}^{k}\lambda_i Z_{in} \le v - \sum_{i=1}^{k}\lambda_i c_i + \varepsilon\right) + P\left(\left|\sum_{i=1}^{k}\lambda_i Y_{in} - \sum_{i=1}^{k}\lambda_i c_i\right| \ge \varepsilon\right).$$

Hence, by the hypotheses of the theorem, and by the choice of $v - \sum_{i=1}^{k}\lambda_i c_i + \varepsilon$,

$$\limsup_{n\to\infty} F_{\sum_{i=1}^{k}\lambda_i Z_{in}+\sum_{i=1}^{k}\lambda_i Y_{in}}(v)$$

$$\le \limsup_{n\to\infty} P\left(\sum_{i=1}^{k}\lambda_i Z_{in} \le v-\sum_{i=1}^{k}\lambda_i c_i+\varepsilon\right)+\limsup_{n\to\infty} P\left(\left|\sum_{i=1}^{k}\lambda_i Y_{in}-\sum_{i=1}^{k}\lambda_i c_i\right|\ge\varepsilon\right)$$
$$= F_{\sum_{i=1}^{k}\lambda_i Z_i}\left(v-\sum_{i=1}^{k}\lambda_i c_i+\varepsilon\right)=F_{m\sum_{i=1}^{k}\lambda_i X_i}\left(v-\sum_{i=1}^{k}\lambda_i c_i+\varepsilon\right).$$

Similarly,

$$P\left(\sum_{i=1}^{k}\lambda_i Z_{in}\le v-\sum_{i=1}^{k}\lambda_i c_i-\varepsilon\right)\le P\left(\sum_{i=1}^{k}\lambda_i Z_{in}+\sum_{i=1}^{k}\lambda_i Y_{in}\le v\right)+P\left(\left|\sum_{i=1}^{k}\lambda_i Y_{in}-\sum_{i=1}^{k}\lambda_i c_i\right|\ge\varepsilon\right)$$

and thus

$$F_{\sum_{i=1}^{k}\lambda_i Z_i}\left(v-\sum_{i=1}^{k}\lambda_i c_i-\varepsilon\right)=F_{m\sum_{i=1}^{k}\lambda_i X_i}\left(v-\sum_{i=1}^{k}\lambda_i c_i-\varepsilon\right)\le\liminf_{n\to\infty}F_{\sum_{i=1}^{k}\lambda_i Z_{in}+\sum_{i=1}^{k}\lambda_i Y_{in}}(v).$$

Since $v-\sum_{i=1}^{k}\lambda_i c_i$ is a continuity point of $F_{m\sum_{i=1}^{k}\lambda_i X_i}(v)$, and since ε may be taken arbitrarily small, we have

$$\lim_{n\to\infty}F_{\sum_{i=1}^{k}\lambda_i Z_{in}+\sum_{i=1}^{k}\lambda_i Y_{in}}(v)=F_{\sum_{i=1}^{k}\lambda_i Z_i}\left(v-\sum_{i=1}^{k}\lambda_i c_i\right)=F_{m\sum_{i=1}^{k}\lambda_i X_i}\left(v-\sum_{i=1}^{k}\lambda_i c_i\right)$$
$$=F_{m\sum_{i=1}^{k}\lambda_i X_i+\sum_{i=1}^{k}\lambda_i c_i}(v)$$

i.e.,
$$\lim_{n\to\infty}F_{\underline{\lambda}'\left(\sum_{t=1}^{m}\underline{X}_{nt}+\underline{Y}_n\right)}(v)=F_{\underline{\lambda}'(m\underline{X}+\underline{c})}(v)$$

$$\Rightarrow \underline{\lambda}'\left(\sum_{t=1}^{m}\underline{X}_{nt}+\underline{Y}_n\right)\xrightarrow{d}\underline{\lambda}'(m\underline{X}+\underline{c}),\forall\underline{\lambda}\in R^k,\underline{\lambda}\ne\underline{0}$$

(by Definition (٢,٢,١٧))

$$\Rightarrow \sum_{t=1}^{m}\underline{X}_{nt}+\underline{Y}_n\xrightarrow{d}m\underline{X}+\underline{c}\ \text{(by Theorem (٣,٢,١٥)).}$$

The proof of the other direction is similar to that given above.

(b) Denote the characteristic functions of $\underline{Y}'_n\sum_{t=1}^{m}\underline{X}_{nt}$ and $m\underline{c}'\underline{X}$ by $\phi_{\underline{Y}'_n\sum_{t=1}^{m}\underline{X}_{nt}}(\theta)$ and $\phi_{m\underline{c}'\underline{X}}(\theta)$, respectively, when θ is any real.

By using Theorem (٢,٢,١٩)(i),(ii), we have

$$\phi_{\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt}}(\theta) = \phi_{\underline{Y}'_n \underline{Z}_n}(\theta) \text{ (by Lemma (٣,٢,٣٥))}$$

$$= E\left(e^{i\theta \underline{Y}'_n \underline{Z}_n}\right) = E\left(e^{i\theta \sum_{i=1}^{k} Y_{in} Z_{in}}\right),$$

and hence

$$\lim_{n\to\infty} \phi_{\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt}}(\theta) = E\left(e^{i\theta \sum_{i=1}^{k} c_i Z_i}\right) = E\left(e^{i\theta m \sum_{i=1}^{k} c_i X_i}\right)$$

$$= E\left(e^{i\theta m \underline{c}' \underline{X}}\right) = \phi_{m\underline{c}'\underline{X}}(\theta)$$

$\lim_{n\to\infty} \phi_{\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt}}(\theta) = \phi_{m\underline{c}'\underline{X}}(\theta), \forall \theta \in R$, it then follows that

$$\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\underline{c}'\underline{X} \text{ (by Theorem (٢,٢,١٩)(i),(ii)).} \blacksquare$$

Example (٣,٢,٣٨)

Let $\{\underline{X}_n\}$ be a sequence of (k × ١) independent random vectors with $\underline{X}_{nt} \xrightarrow{d} \underline{X}$, t = ١,…, m, where $\underline{X} \sim N(\underline{\mu}, V)$ and let $\{\underline{Y}_n\}$ be a sequence of (k×١) random vectors with $\underline{Y}_n \xrightarrow{p} \underline{c}$, where $\underline{c}$ be a vector of constants not infinity. Then

(١) The limiting distribution of $\sum_{t=1}^{m} \underline{X}_{nt} \pm \underline{Y}_n$ is the same as that of $m\underline{X} \pm \underline{c}$; that is,

using Theorem (٣,٢,٢٤), we obtain that

$$\forall \underline{\lambda} = (\lambda_1, \ldots, \lambda_k)' \in R^k, \underline{\lambda} \neq \underline{0}$$

$$\underline{\lambda}'(m\underline{X}) \sim N(m\underline{\lambda}'\underline{\mu}, m\underline{\lambda}'V\underline{\lambda}).$$

Using the other direction of the Theorem (٣,٢,٢٤), we now find

$$m\underline{X} \sim N(m\underline{\mu}, mV)$$

where mV is positive definite because $0 < m\underline{\lambda}'V\underline{\lambda} < \infty$ for all $\underline{\lambda} \neq \underline{0}$. Finally, it follows that

$$m\underline{X} \pm \underline{c} \sim N\left(m\underline{\mu} \pm \underline{c}, mV\right) \text{ such that}$$

$$\sum_{t=1}^{m} \underline{X}_{nt} \pm \underline{Y}_n \xrightarrow{d} N\left(m\underline{\mu} \pm \underline{c}, mV\right).$$

(٢) The limiting distribution of $\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt}$ is the same as that of $m\underline{c}'\underline{X}$. Note that

$$m\underline{c}'\underline{X} \sim N\left(m\underline{c}'\underline{\mu}, m\underline{c}'V\underline{c}\right), \text{ and this is follows that}$$

$$\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} N(m\underline{c}'\underline{\mu}, m\underline{c}'V\underline{c}).$$

Theorem (٣,٢,٣٩)

Let $\{\underline{X}_n\}$ be a sequence of (k × ١) independent random vectors with $\underline{X}_{nt} \xrightarrow{d} \underline{X}$, t = ١,…, m, and let $\{\underline{Y}_n\}$ be a sequence of (k × ١) independent random vectors with $\underline{Y}_n \xrightarrow{d} \underline{Y}$, where $\underline{X}$ and $\underline{Y}$ are two random vectors. Suppose $\underline{X}_n$ and $\underline{Y}_n$ are independent for $n \geq 1$. Then $\underline{X}$ and $\underline{Y}$ are independent, and

(a) $\sum_{t=1}^{m} \underline{X}_{nt} \pm \underline{Y}_n \xrightarrow{d} m\underline{X} \pm \underline{Y}$;

(b) $\underline{Y}'_n \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\underline{Y}'\underline{X}$.

Proof:

(a) Let $\phi_{\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)}(\theta)$ be the characteristic function of $\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)$ and let $\phi_{\underline{\lambda}'(m\underline{X}+\underline{Y})}(\theta)$ be the characteristic function of $\underline{\lambda}'(m\underline{X} + \underline{Y})$, for any vector $\underline{\lambda} = (\lambda_1, \ldots, \lambda_k)' \in R^k$ and any real θ.

Using for proof Theorem (٢,٢,١٩)(i),(ii) and Theorem (٣,٢,١٥), we have

$$\phi_{\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)}(\theta) = \phi_{\underline{\lambda}'(\underline{Z}_n + \underline{Y}_n)}(\theta) \qquad \text{(by Lemma (٣,٢,٣٥))}$$

$$= E\left(e^{i\theta\underline{\lambda}'(\underline{Z}_n + \underline{Y}_n)}\right)$$

$$= E\left(e^{i\theta\left(\sum_{i=1}^{k} \lambda_i Z_{in} + \sum_{i=1}^{k} \lambda_i Y_{in}\right)} \right)$$

$$= E\left(e^{i\theta\sum_{i=1}^{k} \lambda_i Z_{in}} e^{i\theta\sum_{i=1}^{k} \lambda_i Y_{in}} \right).$$

Now taking limits of the above, we have

$$\lim_{n\to\infty} \phi_{\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n\right)}(\theta) = E\left(e^{i\theta\sum_{i=1}^{k} \lambda_i Z_i} e^{i\theta\sum_{i=1}^{k} \lambda_i Y_i} \right) = E\left(e^{i\theta m\sum_{i=1}^{k} \lambda_i X_i} e^{i\theta\sum_{i=1}^{k} \lambda_i Y_i} \right)$$

$$= E\left(e^{i\theta\left(m\sum_{i=1}^{k} \lambda_i X_i + \sum_{i=1}^{k} \lambda_i Y_i\right)} \right)$$

$$= E\left(e^{i\theta\underline{\lambda}'(m\underline{X}+\underline{Y})} \right) = \phi_{\underline{\lambda}'(m\underline{X}+\underline{Y})}(\theta).$$

Therefore, we have

$\underline{\lambda}'\left(\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n \right) \xrightarrow{d} \underline{\lambda}'(m\underline{X} + \underline{Y}), \forall \underline{\lambda} \in R^k, \underline{\lambda} \neq \underline{0}$ (by Theorem (٢,٢,١٩)(i),(ii)). It then follows that

$\sum_{t=1}^{m} \underline{X}_{nt} + \underline{Y}_n \xrightarrow{d} m\underline{X} + \underline{Y}$ (by Theorem (٣,٢,١٥)).

The proof of the other direction is similar to that given above. ■
(b) the proof is similar to that given above for (*a*).

Example (٣,٢,٤٠)

Let $\{\underline{X}_n\}$ be denote a sequence of (k × ١) independent random vectors with $\underline{X}_{nt} \xrightarrow{d} \underline{X}$, t = ١,…, m, and let $\{\underline{Y}_n\}$ be denote a sequence of (k × ١) independent random vectors with $\underline{Y}_n \xrightarrow{d} \underline{Y}$. Suppose that $\underline{X}_n$ and $\underline{Y}_n$ are independent for $n \geq 1$. Then $\underline{X}$ and $\underline{Y}$ are independent such that $\underline{X} \sim N(\underline{\mu}_1, V_1)$ and $\underline{Y} \sim N(\underline{\mu}_2, V_2)$, and the limiting distribution of $\sum_{t=1}^{m} \underline{X}_{nt} \pm \underline{Y}_n$ is the same as that of $m\underline{X} \pm \underline{Y}$; that is, note that by taking Theorem (٣,٢,٢٤), we obtain that

$$\forall \underline{\lambda} = (\lambda_1,\ldots,\lambda_k)' \in R^{k}, \underline{\lambda} \neq \underline{0}$$
$$\underline{\lambda}'(m\underline{X} \pm \underline{Y}) \sim N(m\underline{\lambda}'\underline{\mu}_1 \pm \underline{\lambda}'\underline{\mu}_2, m\underline{\lambda}'V_1\underline{\lambda} + \underline{\lambda}'V_2\underline{\lambda}).$$
Using the other direction of the Theorem (٣,٢,٢٤), we now find
$$m\underline{X} \pm \underline{Y} \sim N(m\underline{\mu}_1 \pm \underline{\mu}_2, mV_1 + V_2)$$
where $mV_1 + V_2$ is positive definite since $\forall \underline{\lambda} \in R^k, \underline{\lambda} \neq \underline{0}$: $m\underline{\lambda}'V_1\underline{\lambda} + \underline{\lambda}'V_2\underline{\lambda} > 0$, it then follows that
$$\sum_{t=1}^{m} \underline{X}_{nt} \pm \underline{Y}_n \xrightarrow{d} N(m\underline{\mu}_1 \pm \underline{\mu}_2, mV_1 + V_2).$$

Theorem (٣,٢,٤١)

Let $\{\underline{X}_n\}$ be a sequence of (k × ١) independent random vectors with $\underline{X}_{nt} \xrightarrow{d} \underline{X}$, t = ١,…, m, and let $\{\mathrm{Y}_n\}$ be a sequence of (ω × k) random matrices with $\mathrm{Y}_n \xrightarrow{p} \mathrm{C}$. Then $\mathrm{Y}_n \sum_{t=1}^{m} \underline{X}_{nt} \xrightarrow{d} m\mathrm{C}\underline{X}$.

Proof:

Let $\phi_{\underline{\lambda}'\left(\mathrm{Y}_n \sum_{t=1}^{m} \underline{X}_{nt}\right)}(\theta)$ and $\phi_{\underline{\lambda}'(m\mathrm{C}\underline{X})}(\theta)$ be the characteristic functions of $\underline{\lambda}'\left(\mathrm{Y}_n \sum_{t=1}^{m} \underline{X}_{nt}\right)$ and $\underline{\lambda}'(m\mathrm{C}\underline{X})$, respectively, when $\underline{\lambda} = (\lambda_1,\ldots,\lambda_\omega)'$ is an arbitrary vector of fixed constants and θ is any real.

$$\phi_{\underline{\lambda}'\left(\mathrm{Y}_n \sum_{t=1}^{m} \underline{X}_{nt}\right)}(\theta) = \phi_{\underline{\lambda}'(\mathrm{Y}_n \underline{Z}_n)}(\theta) \qquad \text{(by Lemma (٣,٢,٣٥))}$$
$$= E\left(\mathrm{e}^{i\theta\underline{\lambda}'(\mathrm{Y}_n \underline{Z}_n)}\right) = E\left(\mathrm{e}^{i\theta \sum_{j=1}^{\omega} \sum_{i=1}^{k} \lambda_j \mathrm{y}_{jin} \mathrm{Z}_{in}}\right),$$

and hence

$$\lim_{n\to\infty} \phi_{\underline{\lambda}'\left(\mathrm{Y}_n \sum_{t=1}^{m} \underline{X}_{nt}\right)}(\theta) = E\left(\mathrm{e}^{i\theta \sum_{j=1}^{\omega} \sum_{i=1}^{k} \lambda_j \mathrm{c}_{ji} Z_i}\right) = E\left(\mathrm{e}^{i\theta m \sum_{j=1}^{\omega} \sum_{i=1}^{k} \lambda_j \mathrm{c}_{ji} \mathrm{X}_i}\right)$$
$$= E\left(\mathrm{e}^{i\theta\underline{\lambda}'(m\mathrm{C}\underline{X})}\right) = \phi_{\underline{\lambda}'(m\mathrm{C}\underline{X})}(\theta).$$
$$\Rightarrow \underline{\lambda}'\left(\mathrm{Y}_n \sum_{t=1}^{m} \underline{X}_{nt}\right) \xrightarrow{d} \underline{\lambda}'(m\mathrm{C}\underline{X}), \forall \underline{\lambda} \in R^{\omega}, \underline{\lambda} \neq \underline{0} \text{ (by}$$
Theorem(٢,٢,١٩)(i),(ii)).

$$\Rightarrow \mathrm{Y}_n \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt} \xrightarrow{d} m\mathrm{C}\underline{\mathrm{X}} \text{ (by Theorem (٣,٢,١٥)).} \blacksquare$$

Example (٣,٢,٤٢)

Let $\{\underline{\mathrm{X}}_n\}$ be denote a sequence of $(k \times ١)$ independent random vectors with $\underline{\mathrm{X}}_{nt} \xrightarrow{d} \underline{\mathrm{X}}$, t = ١,…, m, where $\underline{\mathrm{X}} \sim N(\underline{\mu}, V)$ and let $\{\mathrm{Y}_n\}$ be denote a sequence of $(\omega \times k)$ random matrices with $\mathrm{Y}_n \xrightarrow{p} \mathrm{C}$. Then the limiting distribution of $\mathrm{Y}_n \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt}$ is the same as that of $m\mathrm{C}\underline{\mathrm{X}}$, in other words,

$$\mathrm{Y}_n \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt} \xrightarrow{d} N(m\mathrm{C}\underline{\mu}, m\mathrm{C}V\mathrm{C}').$$

Theorem (٣,٢,٤٣)

Let $\{\underline{\mathrm{X}}_n\}$ be a sequence of $(k \times ١)$ independent random vectors with $\underline{\mathrm{X}}_{nt} \xrightarrow{d} \underline{\mathrm{X}}$, t = ١,…, m and let $\{\mathrm{Y}_n\}$ be a sequence of $(k \times k)$ random matrices with $\mathrm{Y}_n \xrightarrow{p} \mathrm{C}$, a nonsingular matrix. Then $\mathrm{Y}_n^{-1} \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt} \xrightarrow{d} m\mathrm{C}^{-1}\underline{\mathrm{X}}$.

Proof

Suppose that

$$\mathrm{Y}_n^{-1} = \begin{bmatrix} a_{11n} & . & . & . & a_{1kn} \\ . & & & & \\ . & & & & \\ . & & & & \\ a_{k1n} & . & . & . & a_{kkn} \end{bmatrix} \text{ and } \mathrm{C}^{-1} = \begin{bmatrix} d_{11} & . & . & . & d_{1k} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ d_{k1} & . & . & . & d_{kk} \end{bmatrix}.$$

Now using for proof Theorem (٢,٢,١٩)(i),(ii) and Theorem (٣,٢,١٥) by denoting the characteristic functions of $\underline{\lambda}'\left(\mathrm{Y}_n^{-1} \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt}\right)$ and $\underline{\lambda}'(m\mathrm{C}^{-1}\underline{\mathrm{X}})$ by

$\phi_{\underline{\lambda}'\left(\mathrm{Y}_n^{-1} \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt}\right)}(\theta)$ and $\phi_{\underline{\lambda}'(m\mathrm{C}^{-1}\underline{\mathrm{X}})}(\theta)$, respectively, $\forall \underline{\lambda} = (\lambda_1, \ldots, \lambda_k)' \in R^k$, $\underline{\lambda} \neq \underline{0}, \forall \theta \in R$:

$$\phi_{\underline{\lambda}'\left(\mathrm{Y}_n^{-1} \sum_{t=1}^{m} \underline{\mathrm{X}}_{nt}\right)}(\theta) = \phi_{\underline{\lambda}'(\mathrm{Y}_n^{-1}\underline{\mathrm{Z}}_n)}(\theta) \qquad \text{(by Lemma (٣,٢,٣٥))}$$

$$= E\left(\mathrm{e}^{i\theta\underline{\lambda}'(\mathrm{Y}_n^{-1}\underline{\mathrm{Z}}_n)}\right) = E\left(\mathrm{e}^{\mathrm{i}\theta \sum_{j=1}^{k} \sum_{i=1}^{k} \lambda_j a_{jin} \mathrm{Z}_{in}}\right),$$

by taking limits, we obtain that

$$\lim_{n\to\infty} \phi_{\underline{\lambda}'\left(Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt}\right)}(\theta) = E\left(e^{i\theta\sum_{j=1}^{k}\sum_{i=1}^{k}\lambda_j d_{ji} Z_i}\right) = E\left(e^{i\theta m\sum_{j=1}^{k}\sum_{i=1}^{k}\lambda_j d_{ji} X_i}\right)$$

$$= E\left(e^{i\theta\underline{\lambda}'(mC^{-1}\underline{X})}\right) = \phi_{\underline{\lambda}'(mC^{-1}\underline{X})}(\theta).$$

Hence, by above, we have

$$\underline{\lambda}'\left(Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt}\right) \xrightarrow{d} \underline{\lambda}'(mC^{-1}\underline{X}), \forall \underline{\lambda}\in R^k, \underline{\lambda}\neq \underline{0}$$ (by Theorem (٢,٢,١٩)(i),(ii)), and this is follows that

$$Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt} \xrightarrow{d} mC^{-1}\underline{X}$$ (by Theorem (٣,٢,١٥)). ■

Example (٣,٢,٤٤)

Let $\{\underline{X}_n\}$ be denote a sequence of (k × ١) independent random vectors with $\underline{X}_{nt} \xrightarrow{d} \underline{X}$, t = ١,…, m where $\underline{X} \sim N(\underline{\mu}, V)$ and let $\{Y_n\}$ be denote a sequence of (k × k) random matrices with $Y_n \xrightarrow{p} C$. Then the limiting distribution of $Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt}$ is the same as that of $mC^{-1}\underline{X}$, in other words

$$Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt} \xrightarrow{d} N\left(mC^{-1}\underline{\mu}, mC^{-1}V[C^{-1}]'\right).$$

Lemma (٣,٢,٤٥)

let $\{\underline{X}_n\}$ be a sequence of (k ×١) random vectors. If $\underline{X}_{nt} \xrightarrow{p} \underline{c}$, t = ١, …, m. Then $\sum_{t=1}^{m}\underline{X}_{nt} \xrightarrow{p} m\underline{c}$.

Theorem (٣,٢,٤٦)

Let $\{\underline{X}_n\}$ be a sequence of (k ×١) random vectors with $\underline{X}_{nt} \xrightarrow{p} \underline{c}_1$, t = ١,…, m and let $\{Y_n\}$ be a sequence of (k × k) random matrices with $Y_n \xrightarrow{p} C_2$, a nonsingular matrix. Then $Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt} \xrightarrow{p} mC_2^{-1}\underline{c}_1$.

Proof

To proof this, note that the elements of the matrix Y_n^{-1} are continuous functions of the elements of Y_n at $Y_n = C_2$, since C_2^{-1} exists. Thus, $Y_n^{-1} \to C_2^{-1}$.

Similarly, the elements of $Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt}$ are sums of products of elements of Y_n^{-1} with those of $\sum_{t=1}^{m}\underline{X}_{nt}$. Since each sum is again a continuous function of Y_n^{-1} and $\sum_{t=1}^{m}\underline{X}_{nt}$,

$$\operatorname*{plim}_{n\to\infty}\left(Y_n^{-1}\sum_{t=1}^{m}\underline{X}_{nt}\right) = \left(\operatorname*{plim}_{n\to\infty} Y_n\right)^{-1}\operatorname*{plim}_{n\to\infty}\sum_{t=1}^{m}\underline{X}_{nt} = mC_2^{-1}\underline{c}_1 . \blacksquare$$

Example (٣,٢,٤٧)

Let $\{\underline{X}_n\}$ be denote a sequence of (k × ١) random vectors with $\underline{X}_{nt} \xrightarrow{p} \underline{c}_1$, t = ١,…, m and let $\{\underline{Y}_n\}$ be denote a sequence of (k × ١) random vectors with $\underline{Y}_n \xrightarrow{p} \underline{c}_2$ where $\underline{c}$١ and $\underline{c}$٢ are two vectors of constants. Then

(١) $\operatorname*{plim}_{n\to\infty}\left(\sum_{t=1}^{m}\underline{X}_{nt} \pm \underline{Y}_n\right) = \operatorname*{plim}_{n\to\infty}\sum_{t=1}^{m}\underline{X}_{nt} \pm \operatorname*{plim}_{n\to\infty}\underline{Y}_n = m\underline{c}_1 \pm \underline{c}_2$.

(٢) $\operatorname*{plim}_{n\to\infty}\left(\underline{Y}_n'\sum_{t=1}^{m}\underline{X}_{nt}\right) = \left(\operatorname*{plim}_{n\to\infty} Y_n\right)'\operatorname*{plim}_{n\to\infty}\sum_{t=1}^{m}\underline{X}_{nt} = m\underline{c}'_2\underline{c}_1$.

Multivariate Delta Method (٣,٢,٤٨) [٣٢], [١٧]

Suppose that $\{\underline{X}_n\}$ is a sequence of random k-vectors such that $\sqrt{n}(\underline{X}_n - \underline{\mu}) \xrightarrow{d} N(\underline{0}, V)$ and let $g : R^k \to R$ is differentiable at $\underline{\mu}$. Then,

$$\sqrt{n}\{g(\underline{X}_n) - g(\underline{\mu})\} \xrightarrow{d} N\left\{0, \frac{\partial g(\underline{\mu})}{\partial \underline{x}'} V \frac{\partial g(\underline{\mu})}{\partial \underline{x}}\right\},$$

where

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'} = \left.\left(\frac{\partial g(\underline{x})}{\partial x_1}, \frac{\partial g(\underline{x})}{\partial x_2}, \ldots, \frac{\partial g(\underline{x})}{\partial x_k}\right)\right|_{\underline{x}=\underline{\mu}}$$

Definition (٣,٢,٤٩) [٢٧]

Write $g : R^k \to R^\ell$ as $g(\underline{x}) = [g_1(\underline{x}),\ldots,g_\ell(\underline{x})]'$ and suppose that for $j = ١,\ldots,\ell$, $g_j(\underline{x})$ has partial derivatives with respect to $x_١$, …, x_k. Then the derivative of $g(\underline{x})$ at $\underline{\mu}$, denote G, is the $\ell \times k$ matrix.

$$\mathrm{G} = \left. \begin{bmatrix} \dfrac{\partial g_1(\underline{x})}{\partial x_1} & . \; . \; . & \dfrac{\partial g_1(\underline{x})}{\partial x_k} \\ . & & . \\ . & & . \\ . & & . \\ \dfrac{\partial g_\ell(\underline{x})}{\partial x_1} & . \; . \; . & \dfrac{\partial g_\ell(\underline{x})}{\partial x_k} \end{bmatrix} \right|_{\underline{x}=\underline{\mu}}$$

Theorem (*Multivariate Delta Method*) (٣,٢,٥٠) [٢٧], [٣٧]

If $g : R^k \to R^\ell$ has a derivative at $\underline{\mu} \in R^k$ and

$$n^b(\underline{\mathrm{X}}_n - \underline{\mu}) \xrightarrow{d} \underline{\mathrm{Y}}$$

from some k-vector $\underline{\mathrm{Y}}$ and some sequence $\underline{\mathrm{X}}_١$, $\underline{\mathrm{X}}_٢$, … of k vectors, where $b >$ ٠, then

$$n^b\left\{ g(\underline{\mathrm{X}}_n) - g(\underline{\mu}) \right\} \xrightarrow{d} \mathrm{G}\,\underline{\mathrm{Y}}.$$

Example (٣,٢,٥١)

Continuing with Example (٣,٢,٣٠). To find the asymptotic distribution of

(١) $g(\underline{\overline{\mathrm{X}}}_n) = g(\overline{\mathrm{X}}_{1n}, \overline{\mathrm{X}}_{2n}) = \overline{\mathrm{X}}_{1n} - \overline{\mathrm{X}}_{2n}$.

(٢) $g(\underline{\overline{\mathrm{X}}}_n) = g(\overline{\mathrm{X}}_{1n}, \overline{\mathrm{X}}_{2n}) = \dfrac{\overline{\mathrm{X}}_{1n}}{\overline{\mathrm{X}}_{2n}}$.

Solution

(١) Form Example (٣,٢,٣٠), we know that $\sqrt{n}\left(\underline{\overline{\mathrm{X}}}_n - \underline{\mu}\right) \xrightarrow{d} N\left(\underline{0}, V\right)$.

Since $g(\underline{x}) = g(x_1, x_2) = x_1 - x_2$ is differentiable over all $R^٢$, we can apply the Delta Method. Note that

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'} = \left. \left(\frac{\partial g(\underline{x})}{\partial x_1}, \frac{\partial g(\underline{x})}{\partial x_2} \right) \right|_{\underline{x}=\underline{\mu}} = (1,-1),$$

and that

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'} V \frac{\partial g(\underline{\mu})}{\partial \underline{x}} = (1,-1) \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

Applying the Delta Method, we have that

$$\sqrt{n}\left\{ g\left(\overline{\mathrm{X}}_n\right) - g\left(\underline{\mu}\right)\right\} \xrightarrow{d} N\left(0, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}\right);$$

i.e., $\overline{\mathrm{X}}_{1n} - \overline{\mathrm{X}}_{2n} \sim N\left\{\mu_1 - \mu_2, \left(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}\right)\big/ n\right\}.$

(٢) From Example (٣,٢,٣٠), we know that $\sqrt{n}\left(\overline{\mathrm{X}}_n - \underline{\mu}\right) \xrightarrow{d} N\left(\underline{0}, V\right)$.

Since $g(\underline{x}) = g(x_1, x_2) = \dfrac{x_1}{x_2}$ is differentiable over all $R^{٢}$, we can apply the Delta Method. Note that

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'} = \left(\frac{\partial g(\underline{x})}{\partial x_1}, \frac{\partial g(\underline{x})}{\partial x_2}\right)\Bigg|_{\underline{x}=\underline{\mu}} = \left(\frac{1}{\mu_2}, \frac{-\mu_1}{\mu_2^2}\right),$$

and

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'} V \frac{\partial g(\underline{\mu})}{\partial \underline{x}} = \left(\frac{1}{\mu_2}, \frac{-\mu_1}{\mu_2^2}\right)\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}\begin{pmatrix} \dfrac{1}{\mu_2} \\ \dfrac{-\mu_1}{\mu_2^2} \end{pmatrix}$$

$$= \frac{1}{\mu_2^2}\left(\sigma_1^2 + \sigma_2^2 \frac{\mu_1^2}{\mu_2^2} - 2\sigma_{12}\frac{\mu_1}{\mu_2}\right)$$

Applying the Delta Method, we obtain that

$$\sqrt{n}\left\{ g\left(\overline{\mathrm{X}}_n\right) - g\left(\underline{\mu}\right)\right\} \xrightarrow{d} N\left(0, \frac{1}{\mu_2^2}\left(\sigma_1^2 + \sigma_2^2 \frac{\mu_1^2}{\mu_2^2} - 2\sigma_{12}\frac{\mu_1}{\mu_2}\right)\right);$$

i.e., $\dfrac{\overline{\mathrm{X}}_{1n}}{\overline{\mathrm{X}}_{2n}} \sim N\left(\dfrac{\mu_1}{\mu_2}, \left(\dfrac{1}{\mu_2^2}\left(\sigma_1^2 + \sigma_2^2 \dfrac{\mu_1^2}{\mu_2^2} - 2\sigma_{12}\dfrac{\mu_1}{\mu_2}\right)\right)\Big/ n\right).$

Example (٣,٢,٥٢)

Suppose that $\underline{\mathrm{X}}_{١}$, $\underline{\mathrm{X}}_{٢}$, …, are i.i.d. random vectors with mean $\underline{\mu} = (\mu_1, \mu_2, \mu_3)'$ and variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix},$$

where V is symetric matrix.

The multivariate CLT says that $\sqrt{n}\left(\overline{\mathrm{X}}_n - \underline{\mu}\right) \xrightarrow{d} N\left(\underline{0}, V\right)$;

i.e., $\overline{\mathrm{X}}_n \sim N\left(\underline{\mu}, V/n\right)$. To find the asymptotic distribution of

$$g(\underline{\overline{X}}_n)=g(\overline{X}_{1n},\overline{X}_{2n},\overline{X}_{3n})=\overline{X}_{1n}+\overline{X}_{2n}+\overline{X}_{3n}.$$

Solution

Since $g(\underline{x})=g(x_1,x_2,x_3)=x_1+x_2+x_3$ is differentiable over all $R^{٣}$, we have

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'}=\left(\frac{\partial g(\underline{x})}{\partial x_1},\frac{\partial g(\underline{x})}{\partial x_2},\frac{\partial g(\underline{x})}{\partial x_3}\right)\Bigg|_{\underline{x}=\underline{\mu}}=(1,1,1),$$

and that

$$\frac{\partial g(\underline{\mu})}{\partial \underline{x}'}V\frac{\partial g(\underline{\mu})}{\partial \underline{x}}=(1,1,1)\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_2^2 & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_3^2\end{bmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$=\sigma_1^2+\sigma_2^2+\sigma_3^2+2\sigma_{12}+2\sigma_{13}+2\sigma_{23}.$$

Therefore, the Delta Method gives

$$\sqrt{n}\left\{g(\underline{\overline{X}}_n)-g(\underline{\mu})\right\}\xrightarrow{d}N\left(0,\sigma_1^2+\sigma_2^2+\sigma_3^2+2\sigma_{12}+2\sigma_{13}+2\sigma_{23}\right);$$

i.e., $\overline{X}_{1n}+\overline{X}_{2n}+\overline{X}_{3n}\sim N\left(\mu_1+\mu_2+\mu_3,\left(\sigma_1^2+\sigma_2^2+\sigma_3^2+2\sigma_{12}+2\sigma_{13}+2\sigma_{23}\right)\big/n\right).$

Example (٣،٢،٥٣)

Continuing with Example (٣،٢،٣٠). Find the asymptotic distribution of $g(\underline{\overline{X}}_n)=[g_1(\underline{\overline{X}}_n),g_2(\underline{\overline{X}}_n),g_3(\underline{\overline{X}}_n)]'$ such that

$$g_1(\underline{\overline{X}}_n)=g_1(\overline{X}_{1n},\overline{X}_{2n})=\overline{X}_{1n}+\overline{X}_{2n}$$

$$g_2(\underline{\overline{X}}_n)=g_2(\overline{X}_{1n},\overline{X}_{2n})=\overline{X}_{1n}-\overline{X}_{2n}$$

$$g_3(\underline{\overline{X}}_n)=g_3(\overline{X}_{1n},\overline{X}_{2n})=\frac{\overline{X}_{1n}}{\overline{X}_{2n}}$$

Solution

From Example (٣،٢،٣٠), we know that $\sqrt{n}(\underline{\overline{X}}_n-\underline{\mu})\xrightarrow{d}N(\underline{0},V)$. Since $g(\underline{x})=[g_1(\underline{x}),g_2(\underline{x}),g_3(\underline{x})]'$ has a derivative over all $R^{٢}$, such that

$$g_1(\underline{x})=g_1(x_1,x_2)=x_1+x_2$$

$$g_2(\underline{x})=g_2(x_1,x_2)=x_1-x_2$$

$$g_3(\underline{x})=g_3(x_1,x_2)=\frac{x_1}{x_2}$$

We have

$$G = \begin{bmatrix} \dfrac{\partial g_1(\underline{x})}{\partial x_1} & \dfrac{\partial g_1(\underline{x})}{\partial x_2} \\ \dfrac{\partial g_2(\underline{x})}{\partial x_1} & \dfrac{\partial g_2(\underline{x})}{\partial x_2} \\ \dfrac{\partial g_3(\underline{x})}{\partial x_1} & \dfrac{\partial g_3(\underline{x})}{\partial x_2} \end{bmatrix}\Bigg|_{\underline{x}=\underline{\mu}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ \dfrac{1}{\mu_2} & \dfrac{-\mu_1}{\mu_2^2} \end{bmatrix}$$

such that

$$G\,V\,G' = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ \dfrac{1}{\mu_2} & \dfrac{-\mu_1}{\mu_2^2} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & \dfrac{1}{\mu_2} \\ 1 & -1 & \dfrac{-\mu_1}{\mu_2^2} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

where

$$a_{11} = \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}$$

$$a_{22} = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

$$a_{33} = \sigma_1^2 \frac{1}{\mu_2^2} + \sigma_2^2 \frac{\mu_1^2}{\mu_2^4} - 2\sigma_{12} \frac{\mu_1}{\mu_2^3}$$

$$a_{12} = \sigma_1^2 - \sigma_2^2$$

$$a_{13} = \sigma_1^2 \frac{1}{\mu_2} - \sigma_2^2 \frac{\mu_1}{\mu_2^2} - \sigma_{12} \frac{\mu_1}{\mu_2^2} + \sigma_{21} \frac{1}{\mu_2}$$

$$a_{23} = \sigma_1^2 \frac{1}{\mu_2} + \sigma_2^2 \frac{\mu_1}{\mu_2^2} - \sigma_{12} \frac{\mu_1}{\mu_2^2} - \sigma_{21} \frac{1}{\mu_2}$$

Therefore, the Delta Method gives

$$\sqrt{n}\left\{ g(\underline{\overline{X}}_n) - g(\underline{\mu}) \right\} \xrightarrow{d} N(0, GVG');$$

i.e., $g(\underline{\mathrm{X}}_n) \sim N\{g(\underline{\mu}), (\mathrm{G}V\mathrm{G}')/n\}$.

Theorem (٣,٢,٥٤)

Suppose that $\underline{\mathrm{X}}_n = (\mathrm{X}_{1n}, ..., \mathrm{X}_{kn})'$ is asymptotically distributed as $N(\underline{\mu}_1, V_1/n)$ and $\underline{\mathrm{Y}}_n = (\mathrm{Y}_{1n}, ..., \mathrm{Y}_{kn})'$ is asymptotically distributed as $N(\underline{\mu}_2, V_2/n)$, where V١ and V٢ are two fixed matrices, $\underline{\mathrm{X}}_n$ and $\underline{\mathrm{Y}}_n$ are independent. Consider the two random vectors $\underline{\mathrm{X}} = (\mathrm{X}_1, ..., \mathrm{X}_k)'$ and $\underline{\mathrm{Y}} = (\mathrm{Y}_1, ..., \mathrm{Y}_k)'$, let $g(\underline{\mathrm{X}}) = [g_1(\underline{\mathrm{X}}), ..., g_\ell(\underline{\mathrm{X}})]'$ be a vector-valued function with non-zero differentials

$$\mathrm{G} = \left[\frac{\partial g_{\mathrm{j}}(\underline{\mathrm{X}})}{\partial \mathrm{X}_i}\right]\Bigg|_{\underline{\mathrm{X}}=\underline{\mu}_1}$$

which is an $\ell \times k$ matrix, let $f(\underline{\mathrm{Y}}) = [f_1(\underline{\mathrm{Y}}), ..., f_\ell(\underline{\mathrm{Y}})]'$ be a vector-valued function with non-zero differentials

$$\mathrm{F} = \left[\frac{\partial f_{\mathrm{j}}(\underline{\mathrm{Y}})}{\partial \mathrm{Y}_i}\right]\Bigg|_{\underline{\mathrm{Y}}=\underline{\mu}_2}$$

which is an $\ell \times k$ matrix. Suppose $g(\underline{\mathrm{X}}_n)$ and $f(\underline{\mathrm{Y}}_n)$ are independent. Then the asymptotic distribution of $g(\underline{\mathrm{X}}_n) + f(\underline{\mathrm{Y}}_n)$ also normal with mean $g(\underline{\mu}_1) + f(\underline{\mu}_2)$ and covariance matrix $(\mathrm{G}V_1\mathrm{G}' + \mathrm{F}V_2\mathrm{F}')/n$.

Proof

Since $\mathrm{Var}(\underline{\mathrm{X}}_n) = \mathrm{V}_1/n \to 0$ as $n \to \infty$, it then follows that $\underline{\mathrm{X}}_n \xrightarrow{p} \underline{\mu}_1$, and

$$\| \underline{\mathrm{X}}_n - \underline{\mu}_1 \| = o_p(1),$$

and since $\mathrm{Var}(\underline{\mathrm{Y}}_n) = V_2/n \to 0$, as $n \to \infty$, it then follows that $\underline{\mathrm{Y}}_n \xrightarrow{p} \underline{\mu}_2$, and

$$\| \underline{\mathrm{Y}}_n - \underline{\mu}_2 \| = o_p(1).$$

Now using the Taylor series, approximation result for stochastic processes we have:

$$g(\underline{\mathrm{X}}_n) = g(\underline{\mu}_1) + \mathrm{G}(\underline{\mathrm{X}}_n - \underline{\mu}_1) + \mathrm{Z}_n$$

where $\mathrm{Z}_n = o_p(\| \underline{\mathrm{X}}_n - \underline{\mu}_1 \|$, and

$$f(\underline{\mathrm{Y}}_n) = f(\underline{\mu}_2) + \mathrm{F}(\underline{\mathrm{Y}}_n - \underline{\mu}_2) + \mathrm{H}_n$$

where $\mathrm{H}_n = o_p(\| \underline{\mathrm{Y}}_n - \underline{\mu}_2 \|$, and this is follows that

$$g(\underline{\mathrm{X}}_n) + f(\underline{\mathrm{Y}}_n) = g(\underline{\mu}_1) + f(\underline{\mu}_2) + \mathrm{G}(\underline{\mathrm{X}}_n - \underline{\mu}_1) + \mathrm{F}(\underline{\mathrm{Y}}_n - \underline{\mu}_2) + \mathrm{Z}_n + \mathrm{H}_n.$$

Now using Lemma (٣,٢,٣٣) (a), $\sqrt{n}\left\{g(\underline{X}_n)+f(\underline{Y}_n)-g(\underline{\mu}_1)-f(\underline{\mu}_2)\right\}$ and $\sqrt{n}\left\{G(\underline{X}_n-\underline{\mu}_1)+F(\underline{Y}_n-\underline{\mu}_2)\right\}$, will have the same limiting distribution if we show that

$$\text{plim}_{n\to\infty}\left\{\sqrt{n}(Z_n+H_n)\right\}=\underline{0}.$$

But since $Z_n=o_p(\|\underline{X}_n-\underline{\mu}_1\|=o_p(1)$ and $H_n=o_p(\|\underline{Y}_n-\underline{\mu}_2\|=o_p(1)$, then by Definition (٢,٢,٣٠)

$$\underset{n\to\infty}{\text{plim}}\left\{\frac{Z_n}{\|\underline{X}_n-\underline{\mu}_1\|}\right\}=\underline{0} \qquad \text{or} \qquad \underset{n\to\infty}{\text{plim}}\left\{\frac{\sqrt{n}Z_n}{\sqrt{n}\|\underline{X}_n-\underline{\mu}_1\|}\right\}=\underline{0}.$$

and

$$\underset{n\to\infty}{\text{plim}}\left\{\frac{H_n}{\|\underline{Y}_n-\underline{\mu}_2\|}\right\}=\underline{0} \qquad \text{or} \qquad \underset{n\to\infty}{\text{plim}}\left\{\frac{\sqrt{n}H_n}{\sqrt{n}\|\underline{Y}_n-\underline{\mu}_2\|}\right\}=\underline{0}.$$

However, by assumption $\sqrt{n}(\underline{X}_n-\underline{\mu})$ has a finite limiting normal distribution and is bounded stochastically,
that is

$$\sqrt{n}\|\underline{X}_n-\underline{\mu}_1\|=o_p(1),$$

and also $\sqrt{n}(\underline{Y}_n-\underline{\mu}_2)$ has a finite limiting normal distribution and bounded stochastically,
that is

$$\sqrt{n}\|\underline{Y}_n-\underline{\mu}_2\|=o_p(1)$$

and therefore we have
$\text{plim}_{n\to\infty}\sqrt{n}(Z_n)=\underline{0}$ and $\text{plim}_{n\to\infty}\sqrt{n}(H_n)=\underline{0}$, and this is follows that

$$\text{plim}_{n\to\infty}\left\{\sqrt{n}(Z_n+H_n)\right\}=\underline{0}.$$

Hence

$$\begin{aligned}&\sqrt{n}\left\{g(\underline{X}_n)+f(\underline{Y}_n)-g(\underline{\mu}_1)-f(\underline{\mu}_2)\right\}\\&=\sqrt{n}\left\{g(\underline{X}_n)-g(\underline{\mu}_1)\right\}+\sqrt{n}\left\{f(\underline{Y}_n)-f(\underline{\mu}_2)\right\}\\&\xrightarrow{d}G\left\{\sqrt{n}(\underline{X}_n-\underline{\mu}_1)\right\}+F\left\{\sqrt{n}(\underline{Y}_n-\underline{\mu}_2)\right\}\xrightarrow{d}N(\underline{0},GV_1G'+FV_2F').\blacksquare\end{aligned}$$

Theorem (٣,٢,٥٥)

Suppose that $\underline{X}_n=(X_{1n},\ldots,X_{kn})'$ is asymptotically distributed as $N(\underline{\mu},V/n)$, where V is a fixed matrix. Consider the random vector $\underline{X}=(X_1,\ldots,X_k)'$, let $g(\underline{X})=[g_1(\underline{X}),\ldots,g_\ell(\underline{X})]'$ be a vector-valued function with non-zero differentials

$$G = \left[\frac{\partial g_{\mathrm{j}}(\underline{X})}{\partial X_i} \right] \Bigg|_{\underline{X}=\underline{\mu}}$$

which is an $\ell \times k$ matrix, let $f(\underline{X}) = [f_1(\underline{X}),...,f_\ell(\underline{X})]'$ be a vector-valued function with non-zero differentials

$$F = \left[\frac{\partial f_{\mathrm{j}}(\underline{X})}{\partial X_i} \right] \Bigg|_{\underline{X}=\underline{\mu}}$$

which is an $\ell \times k$ matrix. Suppose $g(\underline{X}_n)$ and $f(\underline{X}_n)$ are independent. Then the asymptotic distribution of $g(\underline{X}_n) + f(\underline{X}_n)$ also normal with mean $g(\underline{\mu}) + f(\underline{\mu})$ and covariance matrix $(GVG' + FVF')/n$.

Proof

The proof is the same way for proof theorem (٣,٢,٥٤).

REFERENCES

[١] Ash, R. (١٩٧٢). "Real Analysis and Probability", Academic Press, New York.

[٢] Bishop, Y. M. M., Fienberg, S.E., and Holland, P.W. (١٩٧٥). "Discrete Multivariate Analysis", MIT Press, Cambridge, Mass.

[٣] Burrill, C.W. and Knudsen, J.R. (١٩٦٩). "Real Variables", Holt, Rinehart and Winston, Inc., New York.

[٤] Dedwics, E. J. (١٩٧٦). "Introduction to Statistics and Probability", Holt, Rinehart and Winston, U.S.A.

[٥] Fahady, K.S. and Shamoon, P.J. (١٩٩٠). "Probability", Ministry of Higher Education and Scientific Research.

[٦] Ferguson, T.S. (١٩٩٦). "A Course in Large Sample Theory", Chapman and Hall, Great British, New York.

[٧] Gupta, S.C. and Kapoor, V.K. (١٩٧٧). "Elements of Mathematical Statistics", ٢d ed., Sultan Chand &Sons, New Delhi.

[٨] Hamilton, J.D. (١٩٩٤). "Time Series Analysis", Princeton University Press, New Jersey.

[٩] Johnson, R. A. and Wichern, D. W. (١٩٩٨). "Applied Multivariate Statistical Analysis", ٤th ed., Prentice Hall, New Jersey.

[١٠] Klimov, G. (١٩٨٦). "Probability Theory and Mathematical Statistics", Mir, Publishers, Moscow.

[١١] Parzen, E. (١٩٦٠). "Modern Probability Theory and Its Applications", John Wiley & Sons, Inc., New York, London.

[١٢] Serfling, R.J. (١٩٨٠). "Approximation Theorems of Mathematical Statistics", John Wiley & Sons, New York.

[١٣] Stephem M. Stigler (١٩٧٣). “Laplace, Fisher, and the Discovery of the Concept of Sufficiency”, Biometrika ٦٠, ٤٣٩- ٤٤٥.

And the references from the Internet is:

[١٤] Alves, M.I.F. (١٩٩٩). “Asymptotic Distribution of Gumbel Statistic in a Semi-Parametric Approach**(*)**”, Internet. www.emis.de/journals/PM/ ٥٦f٣/ pm٥٦f٣٠٤.pdf.

[١٥] Andersn, T.W. (٢٠٠٤). “Asymptotic Distribution of a set of Linear Restrictions on Regression Coefficients”, Internet. matrix٠٤. amu. edu. pl/pdf/anderson.pdf.

[١٦] Bertola, M. (٢٠٠٤). “Formal Properties of the Asymptotic Distribution of Eigenvalues of Random Matrix Models”, Internet.www.crm.umotreal.ca/ semi/pdf/bertola.pdf.

[١٧] Boik, R.J. (٢٠٠٤). “Lecture Notes: Statistics ٥٥٠ Spring ٢٠٠٤”, Internet. www.math.montana.edu/~rjboik/classes/٥٥٠/notes_٥٥٠.pdf.

[١٨] Boswijk, H.P. “Asymptotic Theory for Integrated Processes”, Internet. www١.fee.uva.nl/ke/boswijk/book.pdf.

[١٩] Bunzel, H. “Asymptotic Theory- Handout #١”, Internet. www. econ. iastate.edu/classes/econ٦٧١/Bunzel/Asymptotic.pdf.

[٢٠] Davidian, M. “٨ Large Sample Theory: A Casual Approach”, Internet. www.state.ncsu.edu/~st٧٦٢_info/davidian/notes/chap٨.pdf.

[٢١] Dufour, J. and Jouini, T. (٢٠٠٤). “Asymptotic Distribution of a Simple Linear Estimator for VARMA Models in Echelon Form”, Internet.www. fas.umontreal.ca/SCECO/dufour/Dufour_Jouini_٢٠٠٣_LinearEstVARMAEchelon_W.pdf.

[٢٢] Dufour, J. (٢٠٠٣). "Notions of Asymptotic Theory", Internet.www.fas.umontreal.ca/SCECO/dufour/ResE/Dufour_١٩٩٩_C_TS_Asymptotic Theory.pdf.

[٢٣] Ferguson, T.S. "Asymptotic Joint Distribution of Sample Mean and a Sample Quantile",Internet.www.math.ucla.edu/~tom/papers/unpublished/meanmed.pdf.

[٢٤] Flinn, C. (١٩٩٩). "Asymptotic Results for the Linear Regression Model", Internet.www.econ.nyu.edu/user/flinnc/notes١.pdf.

[٢٥] Irvine, J.M. (١٩٨٦). "The Asymptotic Distribution of the Likelihood Ratio Test for a Change in the Mean", Internet.www.census.gov/srd/papers/pdf/rr٨٦-١٠.pdf.

[٢٦] Kelejian, H.H. and Prucha, I.R. (١٩٩٩). "On the Asymptotic Distribution of the Moran I Test Statistic with Applications", Internet.econpapers.hhs.se/paper/undumdeco/Prucha٢.htm-١٠k-٢٤.Dez.٢٠٠٤.

[٢٧] Lehmann, E. (٢٠٠٣). "Statistics ٥٩٧ A Asymptotic Tools Fall ٢٠٠٣. Lecture Notes", Internet.www.stat.psu.edu/~dhunter/asymp/fall٢٠٠٣/lectures/-٤k.

[٢٨] Linton, O. (٢٠٠٤). "AC٤٣٧: Lectures", Internet.econ.lse.ac.٤k/staff/olinton/ac٤٣٧/lecac٤٣٧٢٠٠٣.pdf.

[٢٩] Pötscher, B.M. and Prucha, I.R. (١٩٩٩). "Basic Elements of Asymptotic Theory", Internet.ideas.repec.org/p/umd/umdeco/prucha١.html-٨k.

[٣٠] Rust, J. (١٩٩٨). "Empirical Process Proof of the Asymptotic Distribution of Sample Quantiles", Internet.gemini.econ.umd.edu/jrust/econ٥٥١/lectures/quantile.html-١٢k.

[٣١] Sarno, E. (٢٠٠١). "Further Results on the Asymptotic Distribution of the Euclidean Distance Between MA Models", Internet.www.dipstat.unina.it/Quaderni٪٢٠di٪٢٠statistica/sarno.pdf.

[٣٢] Tebbs, J.M. “Stat ٧٧١ Statistical Theory II Spring, ٢٠٠٥ Lectures Notes”, Internet.www-personal.ksu.edu/~tebbs/stat٧٧١/s٠٥notes.pdf.

[٣٣] Velasco, C. (٢٠٠١). “Asymptotic Theory”, Internet.halweb.uc٣m.es/esp/personal/personas/carelos/econometrics١/ps_pdf١/.pdf.

[٣٤] Weeks, D.M.(٢٠٠٣). “Introduction to Asymptotic Concepts in Statistics”, Internet.www.econ.cam.ac.uk/faculty/weeks/pre_٦sts.htm-٧k- .

[٣٥] Internet. Swan.econ.ohio-state.edu/econ٨٤٠/note٤.pdf., “Lecture ٤: Asymptotic Distribution Theory”.

[٣٦] Internet.www.math.ntu.edu.tw/~hchen/teaching/Large Sample/notes/noteprobability.pdf., “Chapter ٣. Asymptotic Methods”.

[٣٧]Internet.www.stat.washington.edu/jaw/COURSES/٥٨٠s/٥٨١/LectNotes/ch٢.pdf.,“Chapter ٢ Some Basic Large Sample Theory”.

[٣٨] Internet. www.utdt.edu/~economia/econometria/asymptotic.pdf., “Asymptotic Theory”.

حول نظرية التوزيعات التقاربية

الخلاصة

ان جوهر الطرق التقاربية هو التقريب وان الهدف الرئيسي من هذه الرسالة هو اعطـاء اشتقاق موحد لبعض النتائج والنظريات.

ان هذه الرسالة تتعامل ايضاً مع التوزيعات التقاربية التي نحصل عليها عندما يقترب حجم العينة من المالانهاية (يتزايد بصورة غير متناهية) وتتضمن الدراسة الحالية ثلاثة فصول.

نعطي في الفصل الاول مقدمة عامة وخلفيات البحث واطار عام لهذه الرسالة.

اما الفصل الثاني فانه يحتوي على الادوات الاساسية والرئيسية للتحليل التقـاربي ويقـدم نظريات الغاية لحالة المتغير الواحد (على سبيل المثال متتابعة المتغيرات العشوائية).

والفصل الثالث ضم الادوات الاساسية والرئيسية للتحليل التقاربي وقدم نظريـات الغايـة لحالة اكثر من متغير (على سبيل المثال متتابعة المتجهات العشوائية).

وبالاضافة الى ذلك، اوجدنا بعض النتائج حول نظرية التقارب.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل/ كلية التربية

حول نظرية التوزيعات التقاربية

رسالة

مقدمة الى قسم الرياضيات- كلية التربية- جامعة بابل وهو جزء من
متطلبات نيل درجة الماجستير في علوم الرياضيات

من قبل

جنان حمزة فرهود الخناني

بإشراف

الدكتورة كريمة عبد الكاظم الخفاجي

١٤٢٦هـ ٢٠٠٥ م

REPUBLIC OF IRAQ,
MINISTRY OF HIGHER EDUCATION,
AND SCIENTIFIC RESEARCH,
BABYLON UNIVERSITY,
COLLEGE OF EDUCATION