1

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

0-315-19944-X

National Library
of CanadaBibliothèque nationale
du Canada

Canadian Theses Division

Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

67851

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

SUSANA GRACIELA RUBIN BLEUER

Date of Birth — Date de naissance

JUNE 27, 1948

Country of Birth — Lieu de naissance

ARGENTINA

Permanent Address — Résidence fixe

62 GLENDENNING DRIVE
NEPEAN, ONTARIO
CANADA, K2H-7Y9

Title of Thesis — Titre de la thèse

CONFIDENCE BANDS FOR QUANTILE-QUANTILE PLOTS AND TESTING
FOR NORMALITY.

University — Université

CARLETON UNIVERSITY

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

DOCTOR OF PHILOSOPHY

Year this degree conferred — Année d'obtention de ce grade

1985

Name of Supervisor — Nom du directeur de thèse

MIKLÓS CSÖRGÖ

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies of
the film.L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et de
prêter ou de vendre des exemplaires du film.The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.L'auteur se réserve les autres droits de publication; ni la thèse
ni de longs extraits de celle-ci ne doivent être imprimés ou
autrement reproduits sans l'autorisation écrite de l'auteur.

Date

MAY 13, 1985

Signature

CONFIDENCE BANDS FOR QUANTILE-QUANTILE PLOTS
AND TESTING FOR NORMALITY

by

SUSANA RUBIN BLEUER, Licenciada en Matemáticas, M.Sc.

A thesis submitted to the Faculty of
Graduate Studies and Research in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University

Ottawa, Canada

May, 1985

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis

"Confidence Bands for Quantile-Quantile Plots and Testing for Normality"

submitted by Susana Rubin Bleuer, Licenciada en Matemáticas, M.Sc.,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

Miklós Csörgő

Dr. M. Csörgő, Thesis Supervisor

B.M. Puttaswamaiah

Dr. B.M. Puttaswamaiah, Chairman
Department of Mathematics & Statistics

Lajos Horváth

Dr. L. Horváth, External Examiner

Carleton University

May 1985

ABSTRACT

We study several statistical problems from the quantile theory point of view. First we derive a probability inequality for general quantile processes and give an example of its application. Then, we rigorously obtain confidence bands for Q-Q plots via the strong approximation results of Aly (1983) and M. Csörgő and Révész (1978). Our confidence bands are modified versions of M. Csörgő and Révész (1981a)'s bands. Also, we derive the asymptotic distribution of the modified Kolmogorov-Smirnov test for univariate normality proposed by Kumar and Pathak (1977). This test is based on the notion of sufficiency and it deals with the case of having unknown mean and variance.

Finally, we describe the distribution theory corresponding to a class of invariant procedures for assessing multivariate normality when the parameters are unknown. These procedures were studied by Koziol (1982). We investigate and correct some of Koziol (1982)'s conclusions. Our approach on this subject is based on the "estimated empirical process"-point of view. We define a new Cramér-von Mises type invariant statistic, L_n , and under the null hypothesis of joint normality, we obtain an asymptotic representation of L_n , from which we shall obtain tables of critical values in the near future. We describe another statistic M_n , which under the null hypothesis converges in distribution to a completely specified normal random

variable. The statistic M_n provides a two-tailed test. We prove that the L_n and M_n -tests are consistent when we test the null hypothesis against a certain type of alternatives. We also examine a set of quadrivariate data and compare the M_n -test with other tests on the same data.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
FOREWORD	ii
I. PRELIMINARIES AND A KOMLÓS-MAJOR-TUSNÁDY-TYPE INEQUALITY FOR GENERAL QUANTILES.	1
1.1. Introduction	1
1.2. A probability Inequality for General Quantiles and Applications	11
II. CONFIDENCE BANDS FOR QUANTILE-QUANTILE PLOTS DERIVED BY STRONG APPROXIMATION.	18
2.1. Introduction	18
2.2. Confidence Bands for Q-Q Plots via Density Estimation	25
2.3. Direct Confidence Bands for the Quantile-Quantile Plot	35
III. ON THE KUMAR-PATHAK TEST FOR NORMALITY.	47
3.1. Introduction	47
3.2. On the Strong Approximation of $t_n(\cdot)$	49
IV. TESTING FOR MULTIVARIATE NORMALITY	65
4.1. Introduction	65
4.2. Calculation of the Limit of $Z_n(\cdot)$	80
4.3. Weak Convergence of Cramér-von Mises-type Functionals of the process $W_n(\cdot)$	98
4.4. On Assessing Multivariate Normality	110
4.5. An Example	115

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Miklós Csörgő, my thesis supervisor, for his patience, invaluable advice and unfailing encouragement.

I wish to thank Dr. Emad-Eldin Aly for his helpful suggestions and collaboration.

I also wish to express my appreciation to Mrs. Gill Murray, who typed this thesis, for her fast and accurate work.

Finally, I wish to thank my husband for his loving support and encouragement.

FOREWORD

In recent years the strong approximation methodologies have played a role in tackling some classical statistical problems.

M. Csörgő and Révész (1975, 1978) obtained almost sure representations for the general sample quantile process $\rho_n(y) := n^{1/2} f(F_n^{-1}(y)) (F_n^{-1}(y) - F^{-1}(y))$, $0 < y < 1$, where $f = F'$ and F_n^{-1} is the sample quantile function. This enabled the above mentioned authors to develop a new methodology by means of which several important problems could be solved. Some of the numerous applications can be found in M. Csörgő and Révész (1981b), and in M. Csörgő (1983).

In this dissertation our original purpose was to tackle different problems from the quantile theory point of view. We have succeeded in finding confidence bands for the quantile-quantile plot (Chapter II) and derived the asymptotic distribution of a test for univariate normality (Chapter III) as applications of the said quantile theory. However the main subject of our work is the problem of composite goodness of fit for multivariate normality (Chapter IV), where our solution has developed from the point of view of an "estimated empirical process" by strong approximation methodology.

In Chapter I we introduce some notation and we review the main results in quantile theory which will be used throughout this dissertation. We also prove a Komlós, Major and Tusnády-type inequality for the general quantile process and give an example of its application. The results given in Theorems 1.2.1 and 1.2.2 are joint with M. Csörgő.

Chapter II deals with constructing confidence bands for the quantile-quantile plot. This is a problem which we believe had not been previously solved despite attempts by other authors, e.g., Doksum (1974), Switzer (1976) and Doksum and Sievers (1976). In Section 2.1 we explain why their bands, obtained by the frequently misused method of inverting the empirical distribution function, are not necessarily correct. The need for these bands arises when comparing two populations on the basis of two independent random samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, each with distribution function F and G respectively. We denote by $QF = G^{-1}(F(\cdot))$, the quantile-quantile plot of G against F . The main result of this chapter is the construction of confidence bands like

$$1-\alpha = \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c_\alpha) \leq QF(x) \leq Q_n(F_n(x) + n^{-1/2}c_\alpha) \text{ for all } x \text{ such that } F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)\},$$

where $Q_n = G_n^{-1}$ is the empirical quantile function based on the Y -sample and F_n is the empirical distribution function based on the X -sample, ϵ_n is an appropriate sequence of real numbers and c_α is defined by $P(\sup_{0 \leq y \leq 1} |B(y)| \leq c_\alpha/\sqrt{2}) = 1-\alpha$, with $B(\cdot)$ a Brownian Bridge. We follow the methodology of M. Csörgő and Révész (1981a), where they obtain confidence bands for the quantile function Q . In our case, the argument of the empirical quantile function Q_n is random, a complication solved by the use of some results of Wellner (1978) on the ratio of the empirical distribution function to the true

distribution function. We obtain also that the confidence bands for the quantile-quantile plot have a random range of definition (i.e. $F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)$) when the distribution function F is unknown. The results in this chapter are joint with E. Aly and they can also be found in a paper by Aly and Bleuer (1983), also to appear in *Statistics and Decisions* in 1985.

In Chapter III we are concerned with composite goodness of fit when testing for univariate normality. This subject has received a great deal of attention in the past two decades and several solutions were proposed. Concerning the classical Cramér-von Mises and Kolmogorov-Smirnov statistics the work of Darling (1955), Kac, Kiefer and Wolfowitz (1955), Shapiro and Wilk (1965), Shapiro and Francia (1972), de Wet and Venter (1972, 1973), Durbin (1973, 1976), Neuhaus (1974, 1976), M. Csörgö *et al.* (1974), Burke *et al.* (1979), M. Csörgö and Révész (1980 I-II, 1981c) and Aly (1983c) has been discussed by M. Csörgö and Révész (1981c) and Aly and M. Csörgö (1985). More recently La Riccia and Mason (1985) also offered a solution. We derive the asymptotic distribution of the modified Kolmogorov-Smirnov test for normality proposed by Kumar and Pathak (1977). Their statistic R_n is based on the notion of sufficiency. We obtain an almost sure representation of the process involved by a two-parameter Gaussian Process, and show that the test based on R_n can be performed with the use of Lilliefors' (1967) tables. These results are joint with Aly and M. Csörgö and the chapter is based on a paper by Aly, Bleuer and M. Csörgö (1984).

Finally, Chapter IV addresses the problem of assessing multivariate normality when the parameters are unknown. It is in general

difficult to extend univariate methods to the multivariate problem. Different approaches were proposed by Weiss (1958), Anderson (1966), Cox (1968), Healy (1968), Wagle (1968), Wilk and Gnanadesikan (1968), Day (1969), Mardia (1970, 1974, 1975), Andrews, Gnanadesikan and Warner (1971, 1972, 1973), Malkovich (1971), Aitkin (1972), Gnanadesikan and Kettenring (1972), Kessel and Fukunaga (1972), Dahiya and Gurland (1973), Malkovich and Afifi (1973), Sarkadi and Tusnády (1974), Mardia and Zemroch (1975), Giorgi and Fattorini (1976), Hensler, Mehrota and Michalek (1977), Small (1978, 1980), de Wet, Venter and van Wyk (1979), Pettit (1979), Rincón-Gallardo, Quesenberry and O'Reilly (1979), Hawkins (1981), Moore and Stubblebine (1981), Yang (1981), Koziol (1982, 1983) and S. Csörgő (1984). Most of these works have been reviewed and classified by Gnanadesikan (1977), Cox and Small (1978) and Mardia (1980). Some of the proposals are graphical procedures, or univariate techniques for evaluating marginal normality. Marginal normality does not imply multivariate normality but the presence of non-normality is often reflected in the marginal distributions. As Gnanadesikan (1977) stresses, there is a need for tests which involve the multivariate nature of the data in order to obtain more sensitivity. There are few formal methods available and for those which do exist, the computational efforts tend to be prohibitive.

The composite goodness of fit approach by the estimated multivariate empirical process was discussed in Burke et al. (1979) and Koziol (1983). In terms of this approach, the resulting asymptotic Gaussian Process depends on the unknown parameters and hence little

can be said about the classical Kolmogorov-Smirnov or the Cramér-von Mises statistics of the process. The problem of an asymptotic process depending on the unknown parameters will not exist if the statistic considered is invariant under affine transformations of the observed values (vectors of the sample): in such a case, we may assume, without loss of generality, that the X -sample is from a multivariate normal population with zero vector mean and unit covariance matrix. In this chapter we investigate the univariate process $W_n(\cdot)$ proposed by Koziol (1982), which is invariant. $W_n(\cdot)$ is based on some earlier graphical techniques: given that X is distributed with a k -variate normal law $N_{(k)}(\mu, \Sigma)$, the random variable $Y = (X-\mu)^T \Sigma^{-1} (X-\mu)$ is distributed as a chi-squared with k degrees of freedom (with distribution function denoted by $F_{(k)}(\cdot)$). The process $W_n(\cdot)$ is sensitive to this characteristic of the multivariate normal distribution. We propose a new statistic L_n , which is a Cramér-von Mises type functional of $W_n(\cdot)$ and prove that it converges in distribution to a completely specified random variable that can be expressed as an infinite sum of weighted independent chi-squared with one degree of freedom random variables. This result provides a consistent test against a certain type of alternatives. We propose also another test: we prove that the statistic $M_n = \int W_n(\cdot) dF_{(k)}(\cdot)$ converges in distribution to a normal random variable with zero mean and variance σ_k^2 , depending on the number k of variables of the vector X . We also prove that the corresponding two-tailed test is consistent against a group of alternatives, slightly more restrictive than those for the Cramér-von Mises test. The results in this chapter are believed to be new.

We have tried to be as consistent as possible to preserve the notation established in Chapter I. However some symbols represent different statistics from chapter to chapter. To avoid confusion, we define most of the variables, functions and processes involved at the beginning of each chapter.

May, 1985.

I. PRELIMINARIES AND A KOMLÓS-MAJOR-TSNÁDY INEQUALITY FOR GENERAL QUANTILES

§1.1. Introduction. Let $X_1, X_2, \dots, X_n, \dots$ be independent identically distributed random variables (i.i.d.r.v.) with distribution function F and let F_n be the empirical distribution function of the first $n \geq 1$ of these random variables. Define the empirical process β_n by

$$\beta_n(x) := n^{1/2} \{F_n(x) - F(x)\}, \quad x \in \mathbb{R}$$

where $\mathbb{R}$ denotes the set of real numbers.

Two Gaussian Processes play an important role in this exposition. They are defined as follows:

- (a) a Brownian Bridge $\{B(y) : 0 \leq y \leq 1\}$ is a separable Gaussian Process with $E(B(y)) = 0$ and covariance function $E(B(y)B(x)) = y \wedge x - yx$;
- (b) a Kiefer Process $\{K(y, t) : 0 \leq y \leq 1, 0 \leq t < +\infty\}$ is a separable Gaussian Process with $E[K(y, t)] = 0$ and covariance function $E[K(y_1, t)K(y_2, s)] = (s \wedge t) \cdot (y_1 \wedge y_2 - y_1 \cdot y_2)$.

Both of these processes can be represented in terms of Wiener Processes:

$$B(y) \stackrel{D}{=} W(y) - yW(1), \quad 0 \leq y \leq 1;$$

$$K(y, t) \stackrel{D}{=} W(y, t) - yW(1, t), \quad 0 \leq y \leq 1, t \geq 0$$

where $\stackrel{D}{=}$ stands for equality of all finite dimensional distributions, and

where $\{W(y) : 0 \leq y \leq 1\}$ is a standard Wiener Process with $E(W(y)) = 0$ and $E[W(y)W(x)] = x \wedge y$ and $\{W(y,t) : 0 \leq y, t \leq \infty\}$ is a standard two-dimensional Wiener Process with $E(W(y,t)) = 0$ and $E[W(y_1,t)W(y_2,s)] = (y_1 \wedge y_2) \cdot (s \wedge t)$.

Then we have

THEOREM 1.1.1 (Komlós, Major and Tusnády (1975)). For $X_1, X_2, \dots$ there exists a probability space on which we can define a Brownian Bridge $\{B_n(y) : 0 \leq y \leq 1\}$ for each n , and a Kiefer Process $\{K(y,t) : 0 \leq y \leq 1, t \geq 0\}$ such that

$$(1.1.1) \quad P\left\{\sup_{x \in \mathbb{R}} |\beta_n(x) - B_n(F(x))| > n^{-1/2}(C \log n + z)\right\} \leq Le^{-\lambda z}$$

and

$$(1.1.2) \quad P\left\{\sup_{x \in \mathbb{R}} |n^{1/2}\beta_n(x) - K(F(x), n)| > (C \log n + z) \log n\right\} \leq Le^{-\lambda z}$$

for all $z \in \mathbb{R}$ and n , where C, L, λ are positive absolute constants.

Consequently,

$$(1.1.3) \quad \sup_{x \in \mathbb{R}} |\beta_n(x) - B_n(F(x))| = O(n^{-1/2} \log n) \text{ a.s.}$$

(almost surely), and

$$(1.1.4) \quad \max_{1 \leq k \leq n} \sup_{x \in \mathbb{R}} k^{-1/2} |k^{1/2}\beta_k(x) - K(F(x), k)| = O(n^{-1/2} \log n) \text{ a.s.}$$

Let $U_1, U_2, \dots$ be i.i.d. uniform $(0,1)$ rv. Let E_n be the empirical distribution function of the first n of these rv. and let $U_n(\cdot)$ be the uniform sample quantile function of them, i.e.,

with the order statistics $U_{1:n} \leq U_{2:n} \leq \dots \leq U_{n:n}$ of the first n of these rv. we have

$$(1.1.5) \quad U_n(y) := E_n^{-1}(y) := \inf \{u : E_n(u) \geq y\}$$

$$= \begin{cases} 0 & \text{if } y = 0 \\ U_{k:n} & \text{if } (k-1)/n < y \leq k/n \quad k=1, \dots, n \end{cases}$$

Define the uniform quantile process u_n by

$$(1.1.6) \quad u_n(y) = n^{1/2} \{U_n(y) - y\} \quad 0 \leq y \leq 1.$$

THEOREM 1.1.2 (M. Csörgö and Révész (1975, 1978)). For $U_1, U_2, \dots$ there exists a probability space on which one can define a Brownian Bridge $\{B_n(y); 0 \leq y \leq 1\}$ for each n , and a Kiefer Process $\{K(y, t); 0 \leq y \leq 1, t \geq 0\}$ such that

$$(1.1.7) \quad \max_{1 \leq k \leq n} \sup_{0 \leq y \leq 1} k^{-1/2} |k^{1/2} u_k(y) - K(y, k)| = O(n^{-1/2} (\log \log n)^{1/2} (\log n)^{1/2}) \text{ a.s.}$$

and

$$(1.1.8) \quad P\left\{ \sup_{0 \leq y \leq 1} |u_n(y) - B_n(y)| > n^{-1/2} (C \log n + z) \right\} \leq L e^{-\lambda z}$$

for all $|z| \leq cn, c > 0$ and n , where C, L, λ are positive absolute constants. Consequently,

$$(1.1.9) \quad \sup_{0 \leq y \leq 1} |u_n(y) - B_n(y)| = O(n^{-1/2} \log n) \text{ a.s.}$$

Assume that the distribution function F of the i.i.d.rv $X_1, X_2, \dots$ is continuous and let $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be the

order statistics of the first of these random variables. Define the general sample quantile function Q_n by

$$(1.1.10) \quad Q_n(y) = F_n^{-1}(y) := \inf \{x : F_n(x) \geq y\} \\ = X_{k:n} \quad \text{if } (k-1)/n < y \leq k/n$$

$$(k = 1, 2, \dots, n), \quad \text{and } F_n^{-1}(0) = F_n^{-1}(0+).$$

Let $Q = F^{-1}$, the quantile function of the distribution function F and assume that $f := F'$ exists. Define the general quantile process ρ_n by

$$(1.1.11) \quad \rho_n(y) := n^{1/2} f(Q(y)) \{Q_n(y) - Q(y)\} \quad 0 \leq y \leq 1.$$

Then we have

THEOREM 1.1.3 (M. Csörgő and Révész (1978), M. Csörgő et al.

(1982). Assume that the following conditions hold:

$$(1.1.12) \quad \left\{ \begin{array}{l} \text{(i)} \quad F \text{ is twice differentiable on } (a,b), \text{ the support of } F, \\ \text{where } a := \sup \{x : F(x) = 0\} \quad b := \inf \{x : F(x) = 1\}, \\ \quad \quad -\infty \leq a < b \leq +\infty. \\ \text{(ii)} \quad F'(x) = f(x) > 0, \quad x \in (a,b), \\ \text{(iii)} \quad \text{for some } \gamma > 0 \text{ we have} \\ \quad \quad \sup_{a < x < b} F(x)(1-F(x)) \frac{|f'(x)|}{f^2(x)} = \sup_{0 < y < 1} y(1-y) \frac{|f'Q(y)|}{f^2Q(y)} \leq \gamma. \end{array} \right.$$

Then, on the probability space of Theorem 1.1.2, we have, almost surely,

$$(1.1.13) \quad \sup_{1/(n+1) \leq y \leq n/(n+1)} |\rho_n(y) - B_n(y)| = \begin{cases} O(n^{-1/2} \log n) & \text{if } \gamma < 2 \\ O(n^{-1/2} (\log \log n)^\gamma (\log n)^{(1+\varepsilon)(\gamma-1)}) & \text{if } \gamma \geq 2, \end{cases}$$

where $\varepsilon > 0$ is arbitrary and γ is as in condition (1.1.12) (iii), and

$$(1.1.14) \quad \max_{1 \leq k \leq n} \sup_{1/(n+1) \leq y \leq n/(n+1)} k^{-1/2} |k^{1/2} \rho_k(y) - K(y, k)| = O(n^{-1/2} (\log \log n)^{1/2} (\log n)^{1/2})$$

almost surely.

If, in addition to conditions (1.1.12), we also assume

$$(1.1.15) \quad \left\{ \begin{array}{l} \text{(iv) } A := \limsup_{x \downarrow a} f(x) < \infty, \quad B := \limsup_{x \uparrow b} f(x) < \infty, \\ \text{(v) one of} \\ \quad \text{(v, } \alpha) \text{ } \min(A, B) > 0 \\ \quad \text{(v, } \beta) \text{ if } A = 0 \text{ (resp. } B = 0) \text{ the } f \text{ is nondecreasing (resp.} \\ \quad \text{nonincreasing) on an interval to the right of } a \\ \quad \text{(to the left of } b) \end{array} \right.$$

then, if (v, α) obtains,

$$(1.1.16) \quad \sup_{0 \leq y \leq 1} |\rho_n(y) - B_n(y)| = O(n^{-1/2} \log n) \quad \text{a.s.}$$

and

$$(1.1.17) \quad \max_{1 \leq k \leq n} \sup_{0 \leq y \leq 1} k^{-1/2} |k^{1/2} \rho_k(y) - K(y, k)| = O(n^{-1/2} (\log \log n)^{1/2} (\log n)^{1/2}),$$

a.s., and if (v, β) obtains,

$$(1.1.18) \sup_{0 < y < 1} |\rho_n(y) - B_n(y)| = \begin{cases} O(n^{-1/2} \log n) \text{ a.s.} & \text{if } \gamma < 2 \\ O(n^{-1/2} (\log \log n)^\gamma (\log n)^{(1+\varepsilon)(\gamma-1)}) \text{ a.s.} & \text{if } \gamma \geq 2 \end{cases}$$

where $\varepsilon > 0$ is arbitrary and

$$(1.1.19) \max_{1 \leq k \leq n} \sup_{0 < y < 1} k^{1/2} |\rho_k(y) - K(y, k)| = O(n^{-1/2} (\log \log n)^{1/2} (\log n)^{1/2}) \text{ a.s.}$$

Now let

$$\alpha_n(y) := n^{1/2} (E_n(y) - y), \quad 0 \leq y \leq 1,$$

denote the uniform empirical process, where $E_n(\cdot)$ is the uniform empirical distribution function based on the sample $U_1, \dots, U_n$ previously defined. The following properties shall be used throughout this dissertation.

LEMMA 1.1.1. (M. Csörgő and Révész, 1978). Let F satisfy the conditions (1.1.12). Then we have, with $f = F'$,

$$\frac{f(F^{-1}(y_1))}{f(F^{-1}(y_2))} \leq \left(\frac{y_1 \vee y_2}{y_1 \wedge y_2} \cdot \frac{1 - y_1 \wedge y_2}{1 - y_1 \vee y_2} \right)^\gamma$$

for every pair $y_1, y_2 \in (0, 1)$.

LEMMA 1.1.2 (Wellner, 1978). For all $\lambda > 1$ and $0 \leq b \leq 1$ we have

$$(i) \quad P\left(\sup_{b < y < 1} \frac{Y}{U_n(y)} \geq \lambda\right) \leq \exp\{-nbh(1/\lambda)\}$$

and

$$(ii) \quad P\left(\sup_{b \leq y \leq 1} \frac{U_n(y)}{y} \geq \lambda\right) \leq \exp\{-nbh(\lambda)\},$$

where

$$h(\lambda) := \lambda + \log(1/\lambda) - 1.$$

LEMMA 1.1.3 (Wellner, 1978, Lemma 1, p. 75). For all $\lambda \geq 1$ and $0 \leq a \leq 1$, we have

$$(i) \quad P\left(\sup_{a \leq y \leq 1} \frac{E_n(y)}{y} \geq \lambda\right) \leq \exp(-na\lambda h(1/\lambda)),$$

and

$$(ii) \quad P\left(\sup_{a \leq y \leq 1} \frac{Y}{E_n(y)} \geq \lambda\right) \leq \exp(-na\lambda^{-1}h(\lambda)),$$

where

$$h(\lambda) := \lambda + \log(1/\lambda) - 1.$$

LEMMA 1.1.4 (Csaki's Functional Law of the Iterated Logarithm for the Uniform Empirical Process, cf. M. Csörgő and Révész, 1981b, Theorem 4.5.4). With $\delta_n = dn^{-1} \log \log n$ and $d \geq 0.236$, we have

$$\lim_{n \rightarrow \infty} \sup_{\delta_n \leq y \leq 1 - \delta_n} (y(1-y) \log \log n)^{-1/2} |\alpha_n(y)| = 2 \text{ a.s.}$$

LEMMA 1.1.5. (Functional Law of the Iterated Logarithm for the Uniform Quantile Process, M. Csörgő and Révész, 1978, Theorem 2).

With $\delta_n = 25n^{-1} \log \log n$ we have

$$\lim_{n \rightarrow \infty} \sup_{\delta_n \leq y \leq 1 - \delta_n} (y(1-y) \log \log n)^{-1/2} |u_n(y)| \leq 4 \text{ a.s.,}$$

where $u_n(\cdot)$ is as in (1.1.6).

LEMMA 1.1.6. (P. Lévy's Modulus of Continuity for Brownian Bridges, cf. M. Csörgő and Révész, 1981b, Theorem 1.4.1). For $B(\cdot)$ a Brownian Bridge, we have

$$\lim_{h \rightarrow 0} \sup_{0 \leq s \leq 1-h} \frac{|B(s+h) - B(s)|}{\sqrt{2h \log 1/h}} = 1 \text{ a.s.}$$

LEMMA 1.1.7. (Modulus of Continuity for the Kiefer Process, M. Csörgő and Révész, 1981b, Theorem 1.15.2). Let $\{h_n\}$ be a sequence of positive numbers for which

$$\lim_{n \rightarrow \infty} \frac{\log h_n^{-1}}{\log \log n} = \infty.$$

Then

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq 1-h_n} \gamma_n |K(t+h_n, n) - K(t, n)| = 1 \text{ a.s.}$$

where $\gamma_n = (2nh_n \log 1/h_n)^{-1/2}$, and $K(\cdot, \cdot)$ is any Kiefer Process.

LEMMA 1.1.8 (Functional Law of the Iterated Logarithm for the Kiefer Process, M. Csörgő and Révész, 1981b, Corollary 1.15.2). If $K(\cdot, \cdot)$ is a Kiefer Process, then

$$\lim_{t \rightarrow \infty} \sup_{0 < y < 1} \frac{|K(y, t)|}{\sqrt{4y(1-y)t} \log \log (t/y(1-y))} = 1 \text{ a.s.}$$

LEMMA 1.1.9 (Integral representation of a sequence of estimators, Burke et al. 1979, Theorem 3.1 and Lemma 3.2). Let $X_1, X_2, \dots$ be a sequence of i.i.d.r.v. from a family of distribution functions $\{F(x, \theta); x \in \mathbb{R}, \theta \in \mathbb{R}^p\}$. Let $\{\hat{\theta}_n\} = \{(\hat{\theta}_{n1}, \dots, \hat{\theta}_{np})\}$ be a sequence of estimators of the row vector θ based on the random sample $X_1, \dots, X_n$. Assume that

$$(i) \quad n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{j=1}^n \ell(X_j, \theta_0) + \epsilon_{1n},$$

where θ_0 is the true value of θ , $\ell(\cdot, \theta_0)$ is a measurable p -dimensional row vector valued function, and ϵ_{1n} converges to zero almost surely.

$$(ii) \quad E(\ell(X_j, \theta_0)) = 0$$

$$(iii) \quad M(\theta_0) = E\{\ell(X_j, \theta_0)^T \ell(X_j, \theta_0)\} \text{ is a finite nonnegative definite matrix, where we denote by } \ell(X_j, \theta_0)^T \text{ the transpose vector of } \ell(X_j, \theta_0).$$

$$(iv) \quad \text{Each component of the vector function } \ell(x, \theta_0) \text{ is of bounded variation on each finite interval.}$$

$$(v) \quad \lim_{s \downarrow 0} (s \log \log 1/s)^{1/2} \|\ell(\text{inv } F(s; \theta_0), \theta_0)\| = 0$$

and

$$\lim_{s \uparrow 1} ((1-s) \log \log 1/(1-s))^{1/2} \|\ell(\text{inv } F(s, \theta_0), \theta_0)\| = 0,$$

where $\text{inv } F(s; \theta_0) = \inf\{x : F(x; \theta_0) \geq s\}$ and $\|\cdot\|$ stands for any Euclidean norm.

$$(vi) \quad s \left\| (\partial/\partial s) \ell(\text{inv } F(s; \theta_0), \theta_0) \right\| \leq C, \quad 0 < s < 1/2$$

and

$$(1-s) \left\| (\partial/\partial s) \ell(\text{inv } F(x; \theta_0), \theta_0) \right\| \leq C, \quad 1/2 < s < 1$$

for some positive constant C , where the vector of partial derivatives of the components of $\ell(\text{inv } F(x; \theta_0), \theta_0)$

with respect to s , exists for all $s \in (0,1)$. Then there

exists a Kiefer Process $K(\cdot, \cdot)$ such that

$$\sup_x \left| n^{1/2} \{F_n(x) - F(x; \theta_0)\} - n^{-1/2} K(F(x; \theta_0), n) \right| = O(n^{-1/2} (\log n)^2) \text{ a.s.}$$

and

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \int \ell(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) + L_n + \epsilon_{1n}$$

where

$$L_n = O(n^{-\epsilon}) \quad \text{a.s.}$$

for some $\epsilon > 0$.

On comparing Theorems 1.1.1 to 1.1.3 we notice that the analogue of (1.1.2) is missing in Theorem 1.1.2. Also, the analogues of (1.1.1) (or that of (1.1.8)) and (1.1.2) are missing in Theorem 1.1.3. Our aim here is to find an analogue of (1.1.8) for the general quantile process. To this purpose M. Csörgö and Révész (1984, unpublished notes) proved a probability inequality for the rv.

$$\sup_y \frac{|u_n(y)|}{(y(1-y))^{1/2}}$$

from which an inequality of the type (1.1.8) follows as a corollary.

LEMMA 1.1.10 (M. Csörgő and Révész, 1984, unpublished notes).

Let a_n be a sequence of positive numbers such that

$$(1.1.20) \quad a_n / \log n \rightarrow \infty \quad \text{and} \quad a_n \leq dn$$

where $d > 0$. Then, there exist constants $C > 0$ and $L > 0$ such that

$$(1.1.21) \quad P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} \frac{|u_n(y)|}{(y(1-y))^{1/2}} > a_n \right\} \leq L \exp\{-C(k^4 a_n \wedge n)\}$$

for any $1 \leq k \leq a_n^2 \wedge n/2$ and $n = 1, 2, \dots$.

§1.2. A Probability Inequality for General Quantiles and Applications.

THEOREM 1.2.1. Let a_n be a sequence of numbers such that a_n satisfies the conditions (1.1.20). Assume the conditions (1.1.12) of Theorem 1.1.3. Then; for all $1 \leq k \leq a_n^4 \wedge n/2$ and $n \geq 1$, $\alpha > 0$, we have

$$(1.2.1) \quad P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} |\rho_n(y) - u_n(y)| > n^{-1/2} a_n^{1+\alpha} \right\} \leq L \exp\{-k \log(1+\gamma)\}$$

where $L > 0$ is an absolute constant and γ is the constant of condition (1.1.12) (iii).

REMARK 1.2.1. As a consequence of (1.2.1), when $k = a_n$

we have with $\alpha^* = \frac{\alpha}{2(1+\gamma)}$,

$$(1.2.2) \quad P\left\{ \sup_{\frac{a_n}{n+1} \leq y \leq 1 - \frac{a_n}{n+1}} |\rho_n(y) - u_n(y)| > n^{-\frac{1}{2}} a_n^{1+\alpha} \right\} \leq L(e^{-a_n \log a_n^{\alpha^*}}).$$

Proof of Theorem 1.2.1. By a two-term Taylor expansion we get

$$\begin{aligned} \rho_n(y) &= n^{\frac{1}{2}} f_Q(y) \{Q_n(y) - Q(y)\} = n^{\frac{1}{2}} f_Q(y) \{Q(U_n(y)) - Q(y)\} \\ &= n^{\frac{1}{2}} f_Q(y) \left\{ \frac{u_n(y)}{n^{\frac{1}{2}} f_Q(y)} - \frac{u_n^2(y)}{2n} \frac{f'_Q(\xi)}{f^3_Q(\xi)} \right\} \end{aligned}$$

where $y \wedge U_n(y) < \xi < y \vee U_n(y)$ and $U_n(y)$ is the uniform quantile function. Thus

$$|\rho_n(y) - u_n(y)| \leq \frac{n^{-\frac{1}{2}}}{2} u_n(y) f_Q(y) \frac{|f'_Q(\xi)|}{f^3_Q(\xi)}$$

and

$$(1.2.3) \quad n^{\frac{1}{2}} |\rho_n(y) - u_n(y)| \leq \left\{ \frac{1}{2} \frac{u_n^2(y)}{y(1-y)} \right\} \cdot \left\{ \frac{y(1-y)}{\xi(1-\xi)} \right\} \cdot \left\{ \xi(1-\xi) \frac{|f'_Q(\xi)|}{f^2_Q(\xi)} \right\} \cdot \left\{ \frac{f_Q(y)}{f_Q(\xi)} \right\}.$$

The third factor in (1.2.3) is bounded by γ (cf. condition 1.1.12 (iii)). Applying Lemma 1.1.1 to the fourth factor we get

$$\frac{f_Q(y)}{f_Q(\xi)} \leq \left(\frac{y \vee \xi}{y \wedge \xi} \cdot \frac{1 - y \wedge \xi}{1 - y \vee \xi} \right)^\gamma \quad \text{for every } y, \xi \in (0,1),$$

hence, when $y \leq \xi \leq U_n(y)$,

$$\frac{y(1-y)}{\xi(1-\xi)} \cdot \frac{f_Q(y)}{f_Q(\xi)} \leq \frac{y(1-y)}{y(1-U_n(y))} \cdot \left(\frac{U_n(y)}{y} \cdot \frac{1-y}{1-U_n(y)} \right)^\gamma$$

$$= \left(\frac{U_n(y)}{y} \right)^\gamma \left(\frac{1-y}{1-U_n(y)} \right)^{1+\gamma},$$

and when $U_n(y) \leq \xi \leq y$,

$$\frac{y(1-y)}{\xi(1-\xi)} \frac{fQ(y)}{fQ(\xi)} \leq \left(\frac{y}{U_n(y)} \right)^{1+\gamma} \left(\frac{1-U_n(y)}{1-y} \right)^\gamma.$$

Then with $\alpha > 0$,

$$\begin{aligned} (1.2.4) \quad & P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} |\rho_n(y) - u_n(y)| > n^{-1/2} a_n^{1+\alpha} \right\} \leq \\ & \leq P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} \frac{|u_n(y)|}{[y(1-y)]^{1/2}} > (2a_n/\gamma)^{1/2} \right\} + \\ & + P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} \left(\frac{U_n(y)}{y} \right)^\gamma > a_n^{\alpha/2} \right\} + \\ & + P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} \left(\frac{y}{U_n(y)} \right)^{1+\gamma} > a_n^{\alpha/2} \right\}. \end{aligned}$$

Here we applied the fact that

$$(1.2.5) \quad \left\{ 1 - U_n \left(\frac{k}{n+1} \right); 1 \leq k \leq n \right\} \stackrel{D}{=} \left\{ U_n \left(1 - \frac{k}{n+1} \right); 1 \leq k \leq n \right\}.$$

By Lemma 1.1.2. we have

$$P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} \frac{U_n(y)}{y} > a_n^{\alpha/2\gamma} \right\} \leq \exp\left(-n \left(\frac{k}{n+1} \right) h(a_n^{\alpha/2\gamma})\right)$$

and

$$P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} \frac{y}{U_n(y)} > a_n^{\alpha/2(1+\gamma)} \right\} \leq \exp\left(-n \left(\frac{k}{n+1} \right) h(1/a_n^{\alpha/2(1+\gamma)})\right)$$

where $h(\lambda) = \lambda + \log(1/\lambda) - 1$. Hence, from (1.2.4) and Lemma 1.1.10 we obtain, with absolute positive constants L_1, C_1 and C_2 ,

$$\begin{aligned} (1.2.6) \quad & P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} |\rho_n(y) - u_n(y)| \geq n^{-1/2} a_n^{1+\alpha} \right\} \leq \\ & \leq L_1 \exp(-C_1 k^{1/2} a_n^{1/2}) + \exp(-C_2 k a_n^{\alpha/2\gamma}) + \exp(-k \log a_n^{\frac{\alpha}{2(1+\gamma)}}) \\ & \leq L \exp(-k \log a_n^{\frac{\alpha}{2(1+\gamma)}}), \end{aligned}$$

where $L > 0$. Thus the theorem is proved.

THEOREM 1.2.2. Let a_n be a sequence of positive numbers such that $a_n^{1/2}$ satisfies the conditions (1.1.20). Assume all the conditions of Theorem 1.2.1. Then we have

$$\begin{aligned} (1.2.7) \quad & P\left\{ \sup_{k/(n+1) \leq y \leq 1-k/(n+1)} |\rho_n(y) - B_n(y)| > n^{-1/2} a_n^{1+\alpha} \right\} \\ & \leq L \exp \left\{ -k \log a_n^{\frac{\alpha}{2(1+\gamma)}} \right\} \end{aligned}$$

where $L > 0$ is an absolute constant and γ is the constant of conditions (1.1.12) (iii).

The proof follows from Theorem 1.2.1 and Theorem 1.1.3.

REMARK 1.2.2. The statement (1.2.7) does not imply any almost sure rate of convergence but it could be useful in trying to estimate the rate of convergence of the distribution function of functionals of interest of $\rho_n(\cdot)$ to those of a Brownian Bridge.

For example, let

$$(1.2.8) \quad T_n = n^{-1} \sum_{i=1}^n c_{in} X_{i:n}$$

be a linear combination of order statistics where $X_{i:n}$ are as in Section 1, the order statistics of a sample from a population with distribution function F , and the c_{in} are real numbers. We say that T_n admits a "Cramér type large deviation result" if, for an appropriate standardization by constants τ_n and σ_n it holds that

$$P(n^{1/2} \frac{(T_n - \tau_n)}{\sigma_n} > x) = (1 - \Phi(x))(1 + o(1))$$

as $n \rightarrow \infty$, uniformly, in a range of the form $-A \leq x \leq o(n^\alpha)$, $A \geq 0$, $\alpha > 0$, where $\Phi(\cdot)$ denotes the standard normal distribution function.

Jureckova and Veraverbeke argued that with the help of a probability inequality for the quantile process, this can be obtained under conditions on F like those in (1.1.12). Indeed, they prove the following

THEOREM 1.2.3 (Jureckova and Veraverbeke private communication, 1984). Let T_n be given by (1.2.8), define

$$(1.2.9) \quad \tau_n := n^{-1} \sum_{i=1}^n c_{in} F^{-1}(i/(n+1))$$

$$(1.2.10) \quad \sigma_n^2 = \text{var}(n^{1/2}(T_n - \tau_n))$$

and assume

i) F satisfies the conditions (1.1.12)

$$ii) \quad \frac{1}{n\sigma_n} \sum_{i=1}^n \frac{|c_{in}|}{fQ(i/(n+1))} = O(1) .$$

Then as $n \rightarrow \infty$,

$$P(n^{1/2} \frac{(T_n - \tau_n)}{\sigma_n} > x) = [1 - \Phi(x)] [1 + o(1)]$$

uniformly in the range $-A \leq x \leq O[(\log n)^{1/2}]$.

REMARK 1.2.3. We are interested in showing how the "in probability" inequality of (1.2.7) can be applied, therefore we shall omit parts of the proof where it is not involved.

Proof: They prove first that

$$n^{1/2} \frac{(T_n - \tau_n)}{\sigma_n} = N(0,1) + R_n$$

where $N(0,1)$ denotes a normal random variable with mean zero and variance 1, and

$$R_n = \sum_{i=1}^n \frac{c_{in}}{n \sigma_n^2 f_Q(i/(n+1))} \left| \rho_n\left(\frac{i}{n+1}\right) - B_n\left(\frac{i}{n+1}\right) \right|$$

Here $\rho_n(\cdot)$ denotes the general quantile process for the sample $x_1, \dots, x_n$ and $B_n(\cdot)$ the sequence of Brownian Bridges of Theorem 1.1.3. Hence

$$\begin{aligned} 1 - \Phi(x + \varepsilon_n) - P(|R_n| > \varepsilon_n) &\leq P\left\{n^{1/2} \frac{(T_n - \tau_n)}{\sigma_n} > x\right\} \leq \\ &\leq 1 - \Phi(x - \varepsilon_n) + P(|R_n| > \varepsilon_n) \end{aligned}$$

Applying now (1.2.7),

$$P(|R_n| > \varepsilon_n) \leq P\left(\frac{1}{n+1} \sup_{\frac{1}{n+1} \leq y \leq \frac{n}{n+1}} n^{1/2} |\rho_n(y) - B_n(y)| \leq K \varepsilon_n n^{1/2}\right)$$

where K is some constant depending on condition (ii). Taking

$$\epsilon_n = \frac{1}{K} n^{-\frac{1}{2}} a_n^{1+\alpha}, \quad \alpha > 0,$$

we have

$$P(|R_n| > \epsilon_n) \leq L \exp(-\log a_n^{\frac{\alpha}{2(1+\gamma)}}).$$

Let $\delta := \alpha/2(1+\gamma)$.

Now for $a_n \rightarrow \infty$, we have that

$$\frac{\exp(-\log a_n^\delta)}{1 - \phi(x)} \leq \exp(-x^2 - \log a_n^\delta) \rightarrow 0$$

uniformly in x if

$$x < (\log a_n^\delta)^{1/2}.$$

On the other hand

$$\frac{1 - \phi(x + \frac{1}{K} n^{-\frac{1}{2}} a_n^{1+\alpha})}{1 - \phi(x)} \rightarrow 1$$

uniformly in $-A \leq x \leq 0(n^{\frac{1}{2}} a_n^{-(1+\alpha)})$. A choice for a_n could be

$$a_n = n^{\frac{1}{3+2\alpha}}.$$

Hence the range of convergence is

$$-A \leq x \leq \left(\frac{\delta}{2} \log n^{\frac{1}{3+2\alpha}}\right)^{1/2} = O(\log n)^{1/2},$$

which proves the theorem.

II. CONFIDENCE BANDS FOR QUANTILE-QUANTILE PLOTS DERIVED BY STRONG APPROXIMATION

§2.1. Introduction. We consider the problem of comparing two populations with distribution functions F and G when we have available two independent random samples $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ respectively. One of the useful tools for comparing two samples is the empirical quantile-quantile plot (EQ-Q plot) technique. This is essentially a plot of the order statistics of one sample against those from another and it was formally introduced and first discussed by Wilk and Gnanadesikan (1968). The quantile-quantile plot (Q-Q plot) of G against F is defined by

$$QF(x) := G^{-1}[F(x)], \quad x \in \text{support}(F),$$

where Q denotes the inverse of the distribution function G , i.e.,

$$Q(y) = G^{-1}(y) := \inf\{u : G(u) \geq y\}.$$

Another way to look at comparing two populations is to assume a general model where the difference between two generic random variables X and Y is given in terms of a function $\Delta(\cdot)$, such that

$$Y \stackrel{D}{=} X + \Delta(X).$$

Doksum (1974) studied this model. Under minor restrictions the function $\Delta(\cdot)$ can be uniquely defined by

$$G(\Delta(x) + x) = F(x), \quad x \in \text{support}(F)$$

or by

$$\Delta(x) = G^{-1}[F(x)] - x.$$

If, for example, the X 's are the responses on untreated subjects (control responses) and the Y 's are the responses on the treated subjects, then $\Delta(X)$ gives the effect of the treatment as a function of the control response and is called the response function (Doksum and Sievers, 1976). A natural estimator of the response function is

$$\Delta_n(x) = Q_n^*(F_n(x)) - x,$$

where $F_n(\cdot)$ denotes the empirical distribution function (EDF) based on the X -sample and $Q_n(\cdot)$ denotes the empirical quantile function (EQF) based on the Y -sample. The function $Q_n^*F_n(\cdot)$ is called EQ-Q plot.

To obtain more information about the response function we consider next the empirical quantile-quantile process

$$\begin{aligned} \tau_n(x) &= n^{1/2} g_{QF}(x) (\Delta_n(x) - \Delta(x)) \\ &= n^{1/2} g[Q(F(x))] [Q_n(F_n(x)) - Q(F(x))], \quad x \in S(F) \end{aligned}$$

where $g(x) = G'(x)$ and $S(F) = \text{support}(F) = \{x : 0 < F(x) < 1\}$.

Doksum (1974) derived weak convergence results for the process $\{\tau_n(x)/g_{QF}(x), x \in S(F)\}$ under the condition of F and G having finite support. However, the limiting process is not distribution free.

Aly (1983) obtained strong approximation theorems for the empirical $Q-Q$ process $\{\tau_n(x), x \in S(F)\}$. We have:

THEOREM 2.1.1 (Aly, 1983). Let $Y_1, Y_2, \dots$ be i.i.d.r.v. with a continuous distribution function G satisfying conditions (1.1.12) and let $X_1, X_2, \dots$ be i.i.d.r.v. independent from the sequence of the Y 's with a continuous distribution function F . Then there exists a sequence of two independent Kiefer processes $K_1(\cdot, \cdot)$ and $K_2(\cdot, \cdot)$ such that

$$\begin{aligned} & \sup_{F^{-1}(1/n+1) \leq x \leq F^{-1}(n/n+1)} |\tau_n(x) - n^{-1/2}\{K_1(F(x), n) + K_2(F(x), n)\}| = \\ &= \sup_{1/(n+1) \leq y \leq n/(n+1)} |\tau_n(F^{-1}(y)) - n^{-1/2}\{K_1(y, n) + K_2(y, n)\}| \\ &= O(n^{-1/4}(\log \log n)^{1/2}(\log n)^{1/2}) \quad \text{a.s.} \end{aligned}$$

This result is an analogue of M. Csörgő and Révész (1978)'s strong approximation theorem for the general quantile process. As an application of this methodology M. Csörgő and Révész (1981a) obtained simultaneous confidence bands for the quantile function through two different approaches. The first approach depends on constructing a sequence of consistent estimators $\psi_n^0(\cdot)$ and $\psi_n(\cdot)$ for $1/gQ(\cdot)$. We have

THEOREM 2.1.2 (M. Csörgő and Révész 1981a). Let $Y_1, Y_2, \dots$ be i.i.d.r.v. with a continuous distribution function $G(\cdot)$ satisfying conditions (1.1.12), (1.1.15) and

$$(2.1.2) \quad \sup_{a < x < b} \{G(x)(1-G(x))\}^2 / g(x) \leq C \text{ with some } C > 0.$$

where (a,b) = support of $G(\cdot)$ and $g(\cdot) = G'(\cdot)$. Then

$$(2.13) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{\varepsilon_n \leq y \leq 1 - \varepsilon_n} \left| \frac{Q_n(y+a_n) - Q_n(y-a_n)}{2a_n} - \frac{1}{gQ(y)} \right| = 0 \text{ a.s.}$$

where $a_n = n^{-\alpha}$, $\varepsilon_n = n^{-\beta}$, $\omega(n) = n^\delta$ and $3\beta + \delta < \alpha < 1/2$, $2\delta + 4\beta + \alpha < 1$.

We shall write

$$(2.1.4) \quad \psi_n^0(y) := (2a_n)^{-1} \{Q_n(y+a_n) - Q_n(y-a_n)\}.$$

THEOREM 2.1.3. (M. Csörgő and Révész, 1981a) Let $Y_1, Y_2, \dots$ be i.i.d.r.v. with a continuous distribution function $G(\cdot)$ satisfying conditions (1.1.12), (1.1.15) and

$$(2.1.5) \quad \sup_{a < x < b} |g''(x)| \leq C \text{ with some } C > 0,$$

where (a,b) is the support of G and $g(\cdot) = G'(\cdot)$. Let $\lambda(\cdot)$ be a density function which is absolutely continuous on $(-\infty, +\infty)$ and vanishing outside of the interval $-1/2 \leq x \leq 1/2$ and for which

$$(2.1.6) \quad \begin{cases} \int_{-1/2}^{1/2} x \lambda(x) dx = 0 \\ \int_{-1/2}^{1/2} x^2 \lambda(x) dx < +\infty \\ \sup_x |\lambda'(x)| < +\infty. \end{cases}$$

then we have

$$(2.1.7) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{\varepsilon_n \leq y \leq 1 - \varepsilon_n} \left| \frac{1}{a_n} \int_0^1 \lambda\left(\frac{y-u}{a_n}\right) dQ_n - \frac{1}{fQ(y)} \right| \stackrel{a.s.}{=} 0$$

where $a_n = n^{-\alpha}$, $\omega(n) = n^\delta$, $\varepsilon_n = n^{-\beta}$ and $\delta + 10\beta < 2\alpha$, $\alpha + 2\beta + \delta < 1/2$.

We shall write

$$(2.1.8) \quad \psi_n(y) := a_n^{-1} \int_0^1 \lambda\left(\frac{y-u}{a_n}\right) dQ_n(u).$$

Combining theorems (1.1.3), (2.1.2) and (2.1.3) M. Csörgö and Révész (1981a) obtained confidence bands of the form

$$Q_n(y) - n^{-1/2} c \psi_n(y) \leq Q(y) \leq Q_n(y) + n^{-1/2} c \psi_n(y)$$

where the constant c depends on the significance level of the distribution of the rv. $\sup_{0 \leq y \leq 1} |B(y)|$. We apply these results and Theorem 2.1.1 to obtain first a sequence of consistent estimators $\phi_n(\cdot)$ for $1/gQF(\cdot)$, and then simultaneous confidence bands for the Q - Q plot of the form

$$Q_n(F_n(x)) - n^{-1/2} c \phi_n(x) \leq QF(x) \leq Q_n(F_n(x)) + n^{-1/2} c \phi_n(x)$$

where the constant c depends now on the significance level of the distribution of the rv. $\sup_{0 \leq y \leq 1} \sqrt{2} |B(y)|$.

The second approach is motivated by the discussion in M. Csörgö and Révész (1981a) about the validity of the "usual" argument based on inverting the empirical distribution function. Doksum (1974), Switzer (1976) and Doksum and Sievers (1976) derived confidence bands

for the Q-Q plot by taking the following heuristic steps:

Assume $F(x) = G(x + \Delta(x))$. Then $G_n(x + \Delta(x))$ is equal in distribution and independent from the empirical distribution function of the random sample from F . Consequently, the limiting distribution of the Kolmogorov-Smirnov statistic

$$\sup_x n^{1/2} |F_n(x) - G_n(x + \Delta(x))|$$

is

$$\sup_{0 \leq y \leq 1} 2^{1/2} |B(y)|$$

where $\{B(y), 0 \leq y \leq 1\}$ is a Brownian Bridge. Let $c = c(\alpha)$ be defined by

$$P\left\{ \sup_{0 \leq y \leq 1} 2^{1/2} |B(y)| \leq c \right\} = 1 - \alpha.$$

Then

$$\begin{aligned} 1 - \alpha &= \lim_{n \rightarrow \infty} P\left\{ \sup_{-\infty < x < +\infty} n^{1/2} |F_n(x) - G_n(x + \Delta(x))| \leq c \right\} \\ &= \lim_{n \rightarrow \infty} P\{F_n(x) - n^{-1/2}c \leq G_n(x + \Delta(x)) \leq F_n(x) + n^{-1/2}c, \quad -\infty < x < +\infty\}. \end{aligned}$$

Hence we should get

$$(2.1.9) \quad 1 - \alpha = \lim_{n \rightarrow \infty} P\{G_n^{-1}(F_n(x) - n^{-1/2}c) \leq x + \Delta(x) \leq G_n^{-1}(F_n(x) + n^{-1/2}c), \quad -\infty < x < +\infty\}$$

where

$$G_n^{-1}(y) := \inf \{x : G_n(x) \geq y\}$$

and

$$G_n^{-I}(y) := \sup \{x : G_n(x) \leq y\}.$$

The last step is not correct. First, $G_n^{-1}(\cdot)$ and $G_n^{-I}(\cdot)$ are defined only for those x such that $F_n(x) - n^{-1/2}c \geq 0$ and $F_n(x) + n^{-1/2}c \leq 1$.

Now the event

$$\{F_n(x) - n^{-1/2}c \leq G_n(x + \Delta x) \leq F_n(x) + n^{-1/2}c, \quad -\infty < x < +\infty\}$$

is equal to

$$\begin{aligned} \{F_n(x) - n^{-1/2}c \leq G_n(x + \Delta x) \leq F_n(x) + n^{-1/2}c \text{ for all } x \text{ such that} \\ n^{-1/2}c \leq F_n(x) \leq 1 - n^{-1/2}c\} \end{aligned}$$

since $G_n(\cdot)$ varies between 0 and 1. For the new set of x 's, we have

$$\begin{aligned} \{G_n^{-1}(F_n(x) - n^{-1/2}c) \leq x + \Delta x, \text{ for all } x \text{ such that } n^{-1/2}c \leq F_n(x)\} \\ = \{F_n(x) - n^{-1/2}c \leq G_n(x + \Delta x), \text{ for all } x \text{ such that } n^{-1/2}c \leq F_n(x)\}. \end{aligned}$$

However, even for the smaller set $\{x : n^{-1/2}c \leq F_n(x) \leq 1 - n^{-1/2}c\}$, the statement

$$x + \Delta x \leq G_n^{-I}(F_n(x) + n^{-1/2}c)$$

does not necessarily imply

$$G_n(x + \Delta x) \leq F_n(x) + n^{-1/2}c.$$

Therefore with this technique one can at most obtain simultaneous confidence lower bands and in a smaller range of x 's than is often

proposed. For example we obtain

$$1-\alpha = P\left(\sup_{0 \leq y \leq 1} B(y) \leq c/\sqrt{2}\right) = \lim_{n \rightarrow \infty} P(F_n(x) - n^{-1/2}c \leq G_n(x+\Delta x); x \text{ such}$$

$$\text{that } F_n(x) \geq n^{-1/2}c)$$

$$= \lim_{n \rightarrow \infty} P(G_n^{-1}(F_n(x) - n^{-1/2}c) \leq x + \Delta x; x \text{ such that } F_n(x) \geq n^{-1/2}c).$$

Here again, we use M. Csörgő and Révész' (1981a) method to rigorously obtain a statement in the fashion of (2.1.9). Our confidence bands will be derived directly from the classical strong approximation theorem for the quantile process.

§2.2. Confidence Bands for Q-Q plots via density estimation.

We propose to estimate $1/gQ(F(x))$, $x \in S(F)$ by the sequence of estimators

$$\phi_n^O(x) := \psi_n^O(F_n(x))$$

and

$$\phi_n(x) := \psi_n(F_n(x)),$$

where $\psi_n^O(\cdot)$ and $\psi_n(\cdot)$ are defined by (2.1.4) and (2.1.8) respectively.

We shall see that $\phi_n^O(\cdot)$ and $\phi_n(\cdot)$ are consistent estimators of $1/gQ(F(\cdot))$. Throughout this chapter let

$$(2.2.1) \quad \alpha_n(y) := n^{1/2}(E_n(y) - y), \quad 0 \leq y \leq 1,$$

denote the uniform empirical process based on the random sample

$F(X_1), \dots, F(X_n)$ defined in Section 2.1. We first prove the following

THEOREM 2.2.1. Let $X_1, X_2, \dots, X_n, \dots$ be i.i.d.r.v. with continuous distribution function F and assume that F satisfies the conditions (1.1.12). Assume that ϵ_n, ϵ'_n and ϵ''_n are sequences of positive numbers converging to zero such that

$$(a) \quad \epsilon_n = k\epsilon'_n \text{ and } \epsilon'_n = k\epsilon''_n \text{ with } 0 < k < 1, \text{ i.e. } \epsilon_n < \epsilon'_n < \epsilon''_n,$$

and

$$(b) \quad n^{\frac{1}{2}}\epsilon_n \rightarrow \infty.$$

We write

$$I(\epsilon_n) := \{F^{-1}(\epsilon_n) \leq y \leq F^{-1}(1-\epsilon_n)\},$$

$$I(\epsilon'_n) := \{F^{-1}(\epsilon'_n) \leq y \leq F^{-1}(1-\epsilon'_n)\},$$

$$I(\epsilon''_n) := \{F^{-1}(\epsilon''_n) \leq y \leq F^{-1}(1-\epsilon''_n)\}.$$

Then we have that for almost all ω there exists n_0 such that if $n \geq n_0$ then

$$(2.2.2) \quad I(\epsilon''_n) \subset I(\epsilon'_n) \subset I(\epsilon_n)$$

and we also have

$$(2.2.3) \quad \begin{cases} \lim_{n \rightarrow \infty} P(I(\epsilon'_n) \not\subset I(\epsilon_n)) = \lim_{n \rightarrow \infty} P(F^{-1}(\epsilon'_n) < F^{-1}(\epsilon_n) \text{ or } F^{-1}(1-\epsilon'_n) > F^{-1}(1-\epsilon_n)) \\ \quad \quad \quad = 0 \\ \lim_{n \rightarrow \infty} P(I(\epsilon''_n) \not\subset I(\epsilon'_n)) = \lim_{n \rightarrow \infty} P(F^{-1}(\epsilon''_n) < F^{-1}(\epsilon'_n) \text{ or } F^{-1}(1-\epsilon''_n) > F^{-1}(1-\epsilon'_n)) \\ \quad \quad \quad = 0 \end{cases}$$

REMARK 2.2.1. The statements above can be proven with any three sequences satisfying $\epsilon_n < \epsilon'_n < \epsilon''_n$ but not too far from each other (the meaning of this will be clear from the proof) and such that

$$n \epsilon_n \rightarrow +\infty.$$

Proof. We are going to show that

$$(2.2.4) \quad Z_n := F_n^{-1}(\epsilon'_n) - F^{-1}(\epsilon_n) \geq 0^* \quad \text{a.s. for } n \text{ large,}$$

and

$$(2.2.5) \quad \lim_{n \rightarrow \infty} P(Z_n < 0) = 0.$$

Also

$$(2.2.6) \quad W_n := F_n^{-1}(\epsilon''_n) - F_n^{-1}(\epsilon'_n) \geq 0 \quad \text{a.s. for } n \text{ large,}$$

and

$$(2.2.7) \quad \lim_{n \rightarrow \infty} P(W_n < 0) = 0.$$

The other two inequalities needed to prove (2.2.2), namely

$$F^{-1}(1-\epsilon''_n) \leq F_n^{-1}(1-\epsilon'_n) \leq F^{-1}(1-\epsilon_n) \quad \text{a.s. for } n \text{ large}$$

and the following results needed to prove (2.2.3),

$$\lim_{n \rightarrow \infty} P(F_n^{-1}(1-\epsilon'_n) > F^{-1}(1-\epsilon_n)) = 0$$

and

$$\lim_{n \rightarrow \infty} P(F^{-1}(1-\epsilon''_n) > F_n^{-1}(1-\epsilon'_n)) = 0,$$

can be obtained similarly.

Let $\rho_n^F(\cdot)$ be the general quantile process for the sample $X_1, X_2, \dots, X_n$ from F , i.e.,

$$\rho_n^F(y) := n f F^{-1}(y) (F_n^{-1}(y) - F^{-1}(y)), \quad 0 \leq y \leq 1,$$

where $f(x) := F'(x)$. Then by one term Taylor expansion with

$\epsilon_n' < \epsilon < \epsilon_n''$, we have

$$\begin{aligned}
 (2.2.8) \quad Z_n &= F_n^{-1}(\epsilon'_n) - F^{-1}(\epsilon_n) = F^{-1}(\epsilon'_n) - F^{-1}(\epsilon'_n) + n^{-1/2} \rho_n^F(\epsilon'_n) / fF^{-1}(\epsilon'_n) \\
 &= \left\{ (\epsilon'_n - \epsilon_n) + n^{-1/2} \rho_n^F(\epsilon'_n) \frac{fF^{-1}(\xi)}{fF^{-1}(\epsilon'_n)} \right\} / fF^{-1}(\xi) .
 \end{aligned}$$

Now by Lemma 1.1.1 we have

$$(2.2.9) \quad \frac{fF^{-1}(\xi)}{fF^{-1}(\epsilon'_n)} \leq \left(\frac{\xi \vee \epsilon'_n}{\xi \wedge \epsilon'_n} \cdot \frac{1 - \xi \wedge \epsilon'_n}{1 - \xi \vee \epsilon'_n} \right)^\gamma$$

where $\gamma > 0$ is the constant of condition (1.1.12) (iii). Since

$$\epsilon_n < \xi < \epsilon'_n ,$$

$$\frac{fF^{-1}(\xi)}{fF^{-1}(\epsilon'_n)} \leq \left(2 \frac{\epsilon'_n}{\epsilon_n} \right)^\gamma = (2/k)^\gamma ,$$

we get

$$\begin{aligned}
 P(Z_n < 0) &= P(n^{1/2} fF^{-1}(\xi) Z_n < 0) \\
 &= P(n^{1/2} (\epsilon'_n - \epsilon_n) + \rho_n^F(\epsilon'_n) \frac{fF^{-1}(\xi)}{fF^{-1}(\epsilon'_n)} < 0) \\
 &\leq P(|\rho_n^F(\epsilon'_n)| (2/k)^\gamma > n^{1/2} (\epsilon'_n - \epsilon_n)) \\
 &= P(|\rho_n^F(\epsilon'_n)| > C n^{1/2} \epsilon'_n) ,
 \end{aligned}$$

where $C = (1-k)(k/2)^\gamma$. Now by Theorem 1.1.3 there exists a Kiefer Process $K(.,.)$ such that

$$\sup_{1/(n+1) \leq y \leq n/(n+1)} |\rho_n^F(y) - n^{-1/2} K(y, n)| = o(1) \quad \text{a.s.} .$$

Hence

$$\begin{aligned} P(Z_n < 0) &= P(|\rho_n^F(\varepsilon'_n)| > Cn^{\frac{1}{2}}\varepsilon'_n) \\ &= P(|K(\varepsilon'_n, n)| > Cn\varepsilon'_n) + o(1) \\ &= P(|N(0,1)| > C_1[n\varepsilon'_n]^{\frac{1}{2}}) + o(1) \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ since $n^{\frac{1}{2}}\varepsilon'_n \rightarrow \infty$ (here $N(0,1)$ represents a standard normal rv. and $C_1 > 0$ a fixed constant), which proves (2.2.5). Now by the Functional Law of the Iterated Logarithm for the Kiefer Process (Lemma 1.1.8) we have

$$(2.2.10) \quad n^{-\frac{1}{2}}K(\varepsilon'_n, n) = O(\varepsilon'_n \log \log n)^{\frac{1}{2}} \text{ a.s.}$$

Combining (2.2.8) and (2.2.10) we obtain

$$\begin{aligned} n^{\frac{1}{2}}fF^{-1}(\xi)Z_n &\geq \{(1-k)n^{\frac{1}{2}}\varepsilon'_n - O(\varepsilon'_n \log \log n)^{\frac{1}{2}}\} \\ &= (n^{\frac{1}{2}}\varepsilon'_n)^{\frac{1}{2}} \left\{ (1-k)(n^{\frac{1}{2}}\varepsilon'_n)^{\frac{1}{2}} - O\left(\frac{\log \log n}{n^{\frac{1}{2}}}\right)^{\frac{1}{2}} \right\} \geq 0 \text{ a.s. for } n \text{ large,} \end{aligned}$$

i.e. we obtained (2.2.4). Consider now W_n and apply a one term Taylor expansion. With $\varepsilon'_n < \eta < \varepsilon''_n$, we have

$$W_n := F^{-1}(\varepsilon''_n) - F^{-1}(\varepsilon'_n) = \{(\varepsilon''_n - \varepsilon'_n) - n^{-\frac{1}{2}}\rho_n^F(\varepsilon'_n) \frac{fF^{-1}(\eta)}{fF^{-1}(\varepsilon'_n)}\} / fF^{-1}(\eta).$$

We apply (2.2.9) to obtain

$$\frac{fF^{-1}(\eta)}{fF^{-1}(\varepsilon'_n)} \leq 2\left(\frac{\varepsilon''_n}{\varepsilon'_n}\right)^{\gamma} = 2\left(\frac{2}{k}\right)^{\gamma};$$

Hence

$$n^{\frac{1}{2}}fF^{-1}(\eta)W_n \geq kn^{\frac{1}{2}}\varepsilon'_n - 2\left(\frac{2}{k}\right)^{\gamma} |\rho_n^F(\varepsilon'_n)|.$$

Following the same argument as for the proof of (2.2.5) we have

$$\begin{aligned} P(W_n < 0) &\leq P(|\rho_n^F(\varepsilon'_n)| > Cn^{\frac{1}{2}}\varepsilon'_n) \\ &= P(|N(0,1)| > C_1 n^{\frac{1}{2}}\varepsilon'_n) + o(1) \\ &= o(1), \quad \text{as } n \rightarrow \infty, \end{aligned}$$

and

$$n^{\frac{1}{2}} f F^{-1}(n) W_n \geq kn^{\frac{1}{2}}\varepsilon'_n - (n^{\frac{1}{2}}\varepsilon'_n)^{\frac{1}{2}} O\left(\frac{\log \log n}{n^{\frac{1}{2}}}\right)^{\frac{1}{2}} \geq 0, \quad \text{a.s. for } n \text{ large,}$$

which proves (2.2.7) and (2.2.6) respectively.

THEOREM 2.2.2. Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ be two independent random samples with continuous distribution functions F and G , respectively. Assume that G satisfies the conditions of Theorem 2.1.2. Then,

$$(2.2.11) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{F^{-1}(\varepsilon_n) \leq x \leq F^{-1}(1-\varepsilon_n)} \left| \phi_n^O(x) - \frac{1}{g_{QF}(x)} \right| = 0 \quad \text{a.s.}$$

where $\phi_n^O(x) = \psi_n^O(F_n(x))$, that is

$$\phi_n^O(x) = (2a_n)^{-1} \{ Q_n(F_n(x) + a_n) - Q_n(F_n(x) - a_n) \}$$

and $\omega(n) = n^\delta$, $a_n = n^{-\alpha}$, $\varepsilon_n = n^{-\beta}$ with $3\beta + \delta < \alpha < 1/2$,
 $2\delta + 4\beta + \alpha < 1$.

If, in addition, F satisfies the conditions (1.1.12), then

$$(2.1.12) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{F_n^{-1}(\varepsilon_n) \leq x \leq F_n^{-1}(1-\varepsilon_n)} \left| \phi_n^O(x) - \frac{1}{g_{QF}(x)} \right| = 0 \quad \text{a.s.}$$

Proof. Let $E_n(\cdot)$ be the empirical distribution function of the uniform $(0,1)$ random sample $F(X_1), \dots, F(X_n)$ and let $\{\alpha_n(y); 0 \leq y \leq 1\}$ be the corresponding uniform empirical process. Let $\lambda_n = n^{-\sigma}$ where $\lambda_n < \epsilon_n$, such that $3\sigma + \delta < \alpha < 1/2$ and $2\delta + 4\sigma + \alpha < 1$. If we show that for

$$(2.2.13) \quad \epsilon_n \leq y \leq 1 - \epsilon_n \Rightarrow \lambda_n \leq E_n(y) \leq 1 - \lambda_n \text{ a.s. for } n \text{ large,}$$

then we can apply Theorem 2.1.2 with $y = E_n(u)$, $\epsilon_n \leq y \leq 1 - \epsilon_n$, that is,

$$(2.2.14) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} \left| \frac{Q_n(E_n(y) + a_n) - Q_n(E_n(y) - a_n)}{2a_n} - \frac{1}{gQE_n(y)} \right| = 0 \text{ a.s.}$$

Now to show (2.2.13), by Lemma 1.1.4, i.e., the Functional Law of the Iterated Logarithm for the Uniform Empirical Process, we have

$$(2.2.15) \quad \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} |\alpha_n(y)| = O(\epsilon_n \log \log n)^{1/2} \text{ a.s.}$$

hence for $y \geq \epsilon_n$,

$$E_n(y) \geq E_n(\epsilon_n) = \epsilon_n + \alpha_n(\epsilon_n) n^{-1/2} \geq \epsilon_n - O\left(\frac{\epsilon_n \log \log n}{n}\right)^{1/2} \geq \lambda_n \text{ a.s.}$$

for n large. We can prove that $E_n(y) \leq 1 - \lambda_n$ a.s. in a similar way.

The statement (2.2.11) will follow if we observe that for $y = F(x)$,

$$F_n(x) = E_n(F(x)) = E_n(y)$$

and we prove

$$(2.2.16) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{\epsilon_n \leq y \leq 1-\epsilon_n} \left| \frac{1}{gQ(E_n(y))} - \frac{1}{gQ(y)} \right| = 0.$$

By the one-term Taylor expansion we have, with ξ between y and $E_n(y)$,

$$\left| \frac{1}{gQ(E_n(y))} - \frac{1}{gQ(y)} \right| \leq \sup_{\epsilon_n \leq y \leq 1-\epsilon_n} |a_n(y)| n^{-1/2} \frac{|g'Q(\xi)|}{g^3Q(\xi)}.$$

Then $\lambda_n \leq \xi \leq 1-\lambda_n$. We apply condition (1.1.12) (iii) with $\gamma > 0$ and condition (2.1.2) of Theorem 2.1.2 with $C > 0$ to the derivative of the quantile density function to obtain

$$\left| \frac{g'Q(\xi)}{g^3Q(\xi)} \right| \leq \frac{\gamma \cdot C}{[\xi(1-\xi)]^3} \leq \frac{C_1}{(\lambda_n)^3} \text{ a.s.}$$

This together with (2.2.15) imply

$$\begin{aligned} n^{\delta} \sup_{\epsilon_n \leq y \leq 1-\lambda_n} \left| \frac{1}{gQ(E_n(y))} - \frac{1}{gQ(y)} \right| &= O(n^{\delta+3\sigma} n^{-\beta/2} (\log \log n)^{1/2} n^{-1/2}) \\ &= o(1) \text{ a.s.} \end{aligned}$$

as $n \rightarrow \infty$, which proves (2.2.16).

For the second statement of the theorem, let $\epsilon'_n = k\epsilon_n$, with $k > 1$, as in Theorem 2.1.1. Then we have, for n large

$$[F_n^{-1}(\epsilon'_n), F_n^{-1}(1-\epsilon'_n)] \subset [F_n^{-1}(\epsilon_n), F_n^{-1}(1-\epsilon_n)] \text{ a.s.,}$$

and let $\beta' < \beta$ such that β' satisfies the same conditions for β and recall that $F_n^{-1}(\cdot)$ is non-decreasing. Then, for n sufficiently large, (depending on k and β)

$$[F_n^{-1}(n^{-\beta'}), F_n^{-1}(1-n^{-\beta'})] \subset [F_n^{-1}(\epsilon_n'), F_n^{-1}(1-\epsilon_n')]$$

and hence

$$\begin{aligned} & \lim_{n \rightarrow \infty} \omega(n) \sup_{F_n^{-1}(n^{-\beta'}) \leq x \leq F_n^{-1}(1-n^{-\beta'})} \left| \phi_n^O(x) - \frac{1}{gQF(x)} \right| \\ & \leq \lim_{n \rightarrow \infty} \omega(n) \sup_{F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)} \left| \phi_n^O(x) - \frac{1}{gQF(x)} \right| = 0 \quad \text{a.s.} \end{aligned}$$

COROLLARY 2.2.1. Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ be two independent random samples with continuous distribution functions $F(\cdot)$ and $G(\cdot)$ respectively. Assume that $F(\cdot)$ and $G(\cdot)$ satisfy all the conditions of Theorem 2.2.2. Then one can define a sequence of Brownian Bridges $B_n(\cdot)$ such that

$$\begin{aligned} (2.2.17) \quad & \sup_{F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)} \left| \frac{n^{1/2}}{\phi_n^O(x)} \{Q_n F_n(x) - QF(x)\} + \sqrt{2} B_n(F(x)) \right| \\ & = O\left(\frac{1}{\omega(n)}\right) \quad \text{a.s.,} \end{aligned}$$

$$(2.2.18) \quad \lim_{n \rightarrow \infty} P\{Q_n F_n(x) - cn^{-1/2} \phi_n^O(x) \leq QF(x) \leq Q_n F_n(x) + cn^{-1/2} \phi_n^O(x);$$

$$\text{for } F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)\} = P\left\{ \sup_{0 \leq y \leq 1} |B(y)| \leq c/\sqrt{2} \right\} =$$

$$= 1 - \sum_{k \neq 0} (-1)^{k+1} e^{-k^2 c^2}.$$

The proof follows from Theorems 2.1.1, 2.2.2, and from the fact that the two independent Kiefer Processes $K_1(\cdot, \cdot)$ and $K_2(\cdot, \cdot)$ of

Theorem 2.1.1 can be expressed as

$$n^{1/2} \{K_1(\cdot, n) + K_2(\cdot, n)\} \stackrel{D}{=} \{\sqrt{2} B(\cdot)\}$$

where $\{B(\cdot)\}$ is a Brownian Bridge and $\stackrel{D}{=}$ stands for equality of all the finite dimensional distributions. The details of the proof are similar to those of Consequence 1 of Csörgő and Révész (1981a).

THEOREM 2.2.3. Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ be two independent random samples with continuous distribution functions F and G respectively. Assume that G satisfies the conditions of Theorem 2.1.3. Then,

$$(2.2.29) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)} \left| \phi_n(x) - \frac{1}{g_{QF}(x)} \right| = 0 \text{ a.s.,}$$

where $\phi_n(x) = \psi_n(F_n(x))$, that is

$$\phi_n(x) = a_n^{-1} \int_0^1 \lambda\left(\frac{F_n(x) - u}{a_n}\right) dQ_n(u),$$

where $\omega(n) = n^\delta$, $a_n = n^{-\alpha}$, $\epsilon_n = n^{-\beta}$ with $\delta + 10\beta < 2\alpha$, $\alpha + 2\beta + \delta < 1/2$ and $\lambda(\cdot)$ is a density function which is absolutely continuous on $(-\infty, +\infty)$ and vanishing outside the interval $(-1/2, 1/2)$ and for which the conditions (2.1.6) are satisfied. If, in addition, we assume that F satisfies the conditions (1.1.12) then we have

$$(2.2.20) \quad \lim_{n \rightarrow \infty} \omega(n) \sup_{F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)} \left| \phi_n(x) - \frac{1}{g_{QF}(x)} \right| = 0 \text{ a.s.}$$

The proof is similar to that of Theorem 2.2.2.

COROLLARY 2.2.2. Assume that the conditions of Theorem 2.2.3 are satisfied. Then the statements of Corollary 2.2.1 will be true with the corresponding sequences of numbers ε_n , $\omega(n)$ and a_n of Theorem 2.2.3 when we replace $\phi_n^0(x)$ by $\phi_n(x)$ in them, where

$$\phi_n(x) = a_n^{-1} \int_0^1 \lambda\left(\frac{F_n(x) - u}{a_n}\right) dQ_n(u).$$

§2.3. Direct Confidence Bands for the Quantile-Quantile Plot.

In the present section we are going to construct confidence bands for $Q F(x)$ by a direct route similar to that taken by Csorgö and Revesz (1981a). We cannot apply their result directly as we did in the previous section but we do follow their methodology.

THEOREM 2.3.1. Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ be two independent random samples with continuous distributions functions F and G respectively. Assume that G satisfies (1.1.12). Then

$$\begin{aligned} (2.3.1) \quad & \lim_{n \rightarrow \infty} P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c); F^{-1}(\varepsilon_n) \leq x \leq F^{-1}(1-\varepsilon_n)\} \\ &= \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c) \leq QF(x); F^{-1}(\varepsilon_n) \leq x \leq F^{-1}(1-\varepsilon_n)\} \\ &= P\left\{ \sup_{0 \leq y \leq 1} B(y) \leq c/\sqrt{2} \right\} = 1 - e^{-c^2}, \end{aligned}$$

where $\{B(\cdot)\}$ is a Brownian Bridge, $c > 0$ and $\varepsilon_n > 0$ is a sequence converging to zero such that

$$n^{1/2}\varepsilon_n \rightarrow \infty.$$

If, in addition, F satisfies conditions (1.1.12), then

$$\begin{aligned}
 (2.3.2) \quad & \lim_{n \rightarrow \infty} P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c); F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)\} \\
 &= \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c) \leq QF(x); F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)\} \\
 &= 1 - e^{-c^2}.
 \end{aligned}$$

Proof. Recall that for $y = F(x)$, $E_n(y) = F_n[F^{-1}(y)]$, $0 \leq y \leq 1$, is the EDF obtained from the sample $F(X_1), \dots, F(X_n)$, and $\{\alpha_n(\cdot)\}$ is the corresponding uniform empirical process. Let $\rho_n^G(\cdot)$ be the normed quantile process of the Y -sample and let

$$\begin{aligned}
 I_n &:= \{y : \epsilon_n \leq y \leq 1-\epsilon_n\}, \\
 J_n &:= \{\xi : (E_n(y) + n^{-1/2}c) \wedge y \leq \xi \leq (E_n(y) + n^{-1/2}c) \vee y\}, \\
 y_n &:= E_n(y) + n^{-1/2}c.
 \end{aligned}$$

Consider now

$$(2.3.3) \quad \{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c); F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-\epsilon_n)\}.$$

When $y = F(x)$ and $\epsilon_n \leq y \leq 1-\epsilon_n$, then

$$y_n - y = n^{-1/2}(\alpha_n(y) + c),$$

and

$$0 \leq Q_n(y_n) - Q(y) = Q_n(y_n) - Q(y_n) + Q(y_n) - Q(y),$$

$$\begin{aligned}
 0 &\leq \rho_n^G(y_n) + n^{-1/2} gQ(y_n) [Q(y_n) - Q(y)] \\
 &= \rho_n^G(y_n) + \frac{gQ(y_n)}{gQ(\xi_n)} [\alpha_n(y) + c] \\
 &= \rho_n^G(y_n) + \alpha_n(y) + c + \left[\frac{gQ(y_n)}{gQ(\xi_n)} - 1 \right] (\alpha_n(y) + c),
 \end{aligned}$$

where $y \in I_n$ and $\xi_n \in J_n$. Therefore the set in (2.3.3) is equal to

$$(2.3.4) \quad \{H_n(y, \xi_n, c) \leq c, \quad y \in I_n, \quad \xi_n \in J_n\},$$

where

$$H_n(y, \xi_n, c) := -\rho_n^G(y_n) - \alpha_n(y) - (\alpha_n(y) + c) \left[\frac{gQ(y_n)}{gQ(\xi_n)} - 1 \right].$$

We are going to show that

$$(2.3.5) \quad H_n(y, \xi_n, c) = n^{-1/2} \{K_1(y, n) + K_2(y, n)\} + o_p(1), \quad y \in I_n, \quad \xi_n \in J_n,$$

where $\{K_1(\cdot, \cdot)\}$ and $\{K_2(\cdot, \cdot)\}$ are independent Kiefer processes.

Given that $\{\alpha_n(\cdot) + c\}$ converges weakly to a Brownian Bridge $B(\cdot) + c$, if we prove that

$$(2.3.6) \quad T_n := \sup_{y \in I_n, \xi_n \in J_n} \left| \frac{gQ(y_n)}{gQ(\xi_n)} - 1 \right| \stackrel{P}{=} o(1), \quad y_n = y_n(y),$$

that is, T_n converges to zero in probability, then we obtain

$$\sup_{y \in I_n, \xi_n \in J_n} \left| \frac{gQ(y_n)}{gQ(\xi_n)} - 1 \right| |\alpha_n(y) + c| \stackrel{P}{=} o(1).$$

To prove (2.3.6), let $\gamma > 0$ be the constant of the condition

(1.1.12) (iii) for $Q(\cdot) := G^{-1}(\cdot)$. We have, by Lemma 1.1.1,

$$\begin{aligned} T_n &\leq \sup_{I_n, J_n} \left\{ \left(\frac{gQ(y_n)}{gQ(\xi_n)} - 1 \right) \vee \left(\frac{gQ(\xi_n)}{gQ(y_n)} - 1 \right) \right\} \\ &\leq \sup_{I_n, J_n} \left\{ \left(\frac{y_n \vee \xi_n}{y_n \wedge \xi_n} \cdot \frac{1 - y_n \wedge \xi_n}{1 - y_n \vee \xi_n} \right)^\gamma - 1 \right\}. \end{aligned}$$

Hence, for an arbitrary $\varepsilon > 0$,

$$\begin{aligned} P(T_n > \varepsilon) &\leq P \left(\sup_{I_n, J_n} \left\{ \left(\frac{y_n \vee \xi_n}{y_n \wedge \xi_n} \cdot \frac{1 - y_n \wedge \xi_n}{1 - y_n \vee \xi_n} \right)^\gamma - 1 \right\} > \varepsilon \right) \\ &\leq P \left(\sup_{I_n, J_n} \frac{y_n(1-\xi_n)}{\xi_n(1-y_n)} > (1+\varepsilon)^{1/\gamma} \text{ and } y \leq E_n(y) \right) + \\ &\quad + P \left(\sup_{I_n, J_n} \frac{\xi_n(1-y_n)}{y_n(1-\xi_n)} > (1+\varepsilon)^{1/\gamma} \text{ and } E_n(y) \leq y \right) = p_{1n} + p_{2n}. \end{aligned}$$

Since

$$\frac{y_n(1-\xi_n)}{\xi_n(1-y_n)} \leq \frac{y_n(1-y)}{y(1-y_n)} \text{ when } y \leq \xi_n \leq E_n(y), y \in I_n$$

it is clear that

$$\begin{aligned} (2.3.7) \quad p_{1n} &\leq P \left(\sup_{I_n, J_n} \frac{y_n(1-\xi_n)}{\xi_n(1-y_n)} > (1+\varepsilon)^{1/\gamma} \text{ and } y \leq \xi_n \leq E_n(y) \right) \\ &\leq P \left(\sup_{I_n} \frac{y_n}{y} > (1+\varepsilon)^{1/2\gamma} \right) + P \left(\sup_{I_n} \frac{1-y}{1-y_n} > (1+\varepsilon)^{1/2\gamma} \right). \end{aligned}$$

Now we apply Lemma 1.1.3.

$$P \left(\sup_{y \in I_n} \frac{y_n}{y} > (1+\varepsilon)^{1/2\gamma} \right) = P \left(\sup_{\varepsilon_n \leq y \leq 1-\varepsilon_n} \frac{E_n(y) + n^{-1/2}c}{y} > (1+\varepsilon)^{1/2\gamma} \right)$$

$$\begin{aligned} &\leq P\left(\sup_{\epsilon_n < y \leq 1-\epsilon_n} \frac{E_n(y)}{y} > (1+\epsilon)^{1/2\gamma} - n^{-1/2} c/\epsilon_n\right) \\ &\leq \exp \{-n\epsilon_n(1+\delta)h[1/(1+\delta)]\}, \end{aligned}$$

where $h(\cdot)$ is as in Lemma 1.1.3 and for n large we have

$$(1+\epsilon)^{1/2\gamma} - n^{-1/2} c/\epsilon_n \geq 1+\delta > 1.$$

This is due to the fact that $n^{1/2}\epsilon_n \rightarrow +\infty$. Hence,

$$(2.3.8) \quad P\left(\sup_{\epsilon_n < y \leq 1-\epsilon_n} \frac{y}{y} > (1+\epsilon)^{1/2\gamma}\right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Next, we observe that

$$\begin{aligned} P\left(\sup_{I_n} \frac{1-y}{1-y_n} > (1+\epsilon)^{1/2\gamma}\right) &= P\left(\sup_{I_n} \frac{1-y}{1-E_n(y) - n^{-1/2} c} > (1+\epsilon)^{1/2\gamma}\right) \\ &\leq P\left(\sup_{I_n} \frac{1-y}{1-E_n(y)} > 1/[(1+\epsilon)^{-1/2\gamma} + c/n^{1/2}\epsilon_n]\right) \\ &= P\left(\sup_{I_n} \frac{1-y}{1-E_n(y)} > 1+\delta^*\right), \end{aligned}$$

where $1+\delta^* \leq [1/(1+\epsilon)^{1/2\gamma} + c/n^{1/2}\epsilon_n]^{-1}$, $\delta^* > 0$ for n large.

By symmetry we have

$$\{1 - E_n(1-y); 0 \leq y \leq 1\} \stackrel{D}{=} \{E_n(y); 0 \leq y \leq 1\}$$

hence, with $h(\cdot)$ as in Lemma 1.1.3, we have

$$\begin{aligned} P\left(\sup_{I_n} \frac{1-y}{1-E_n(y)} > 1+\delta^*\right) &= P\left(\sup_{I_n} \frac{y}{E_n(y)} \geq 1+\delta^*\right) \\ &\leq \exp \{-n\epsilon_n(1+\delta^*)^{-1}h(1+\delta^*)\}. \end{aligned}$$

This last inequality is due to Lemma 1.1.3. Hence

$$(2.3.9) \quad P(\sup_{I_n} \frac{1-y}{1-y_n} > (1+\epsilon)^{1/2\gamma}) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Combining (2.3.7), (2.3.8) and (2.3.9) we obtain

$$(2.3.10) \quad p_{1n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Next we estimate p_{2n} . We observe that in this case

$$\frac{\xi_n(1-y_n)}{y_n(1-y)} \leq \frac{y(1-y_n)}{y_n(1-y)}, \text{ when } E_n(y) \leq \xi_n \leq y,$$

and

$$\frac{y(1-y_n)}{y_n(1-y)} \leq \frac{y}{E_n(y)} \cdot \frac{1-E_n(y)}{1-y}, \quad y \in I_n.$$

Hence

$$\begin{aligned} p_{2n} &\leq P(\sup_{I_n} \frac{y}{E_n(y)} \cdot \frac{1-E_n(y)}{1-y} > (1+\epsilon)^{1/\gamma}) \\ &\leq P(\sup_{I_n} \frac{y}{E_n(y)} > (1+\epsilon)^{1/2\gamma}) + P(\sup_{I_n} \frac{1-E_n(y)}{1-y} > (1+\epsilon)^{1/2\gamma}) \\ &\leq \exp[-n\epsilon_n(1+\epsilon)^{-1/2\gamma} \cdot h((1+\epsilon)^{1/2\gamma})] + \exp[-n\epsilon_n(1+\epsilon)^{1/2\gamma} \cdot h(1/(1+\epsilon)^{1/2\gamma})] \end{aligned}$$

by symmetry and Lemma 1.1.3. Then

$$(2.3.11) \quad p_{2n} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

and given $\epsilon > 0$,

$$P(T_n > \epsilon) \leq p_{1n} + p_{2n} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

which proves (2.3.6).

Consider now the first two terms of $H(y, \xi_n, c)$, that is,

$$(2.3.12) \quad -\rho_n^G(E_n(y) + n^{-1/2}c) - \alpha_n(y), \quad \epsilon_n \leq y \leq 1 - \epsilon_n.$$

Since the distribution function G satisfies the conditions (1.1.12)

there exists a Kiefer Process $\{K_1(\cdot, \cdot)\}$ such that

$$(2.3.13) \quad \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} |\rho_n^G(E_n(y) + n^{-1/2}c) - n^{-1/2}K_1(E_n(y) + n^{-1/2}c, n)| = o(1) \text{ a.s.}$$

by (1.1.14) and the fact that when $\epsilon_n \leq y \leq 1 - \epsilon_n$, the quantity

$E_n(y) + n^{-1/2}c$ is within the proper limits, indeed we have

$$E_n(y) + n^{-1/2}c \geq E_n(\epsilon_n) + n^{-1/2}c = \epsilon_n + n^{-1/2}(\alpha_n(\epsilon_n) + c),$$

and by Lemma 2.2.3

$$\alpha_n(\epsilon_n) = O(\epsilon_n \log \log n)^{1/2} \text{ a.s.}$$

hence

$$n^{1/2}\{\epsilon_n + n^{-1/2}(\alpha_n(\epsilon_n) + c)\} = n^{1/2}\epsilon_n + O((\epsilon_n \log \log n)^{1/2}) + c \text{ a.s.}$$

From $n^{1/2}\epsilon_n \rightarrow \infty$ we see that for n large the quantity above is larger

than 1, a.s., hence

$$(n+1)\{E_n(y) + n^{-1/2}c\} \geq (n+1)\{\epsilon_n + O((\epsilon_n \log \log n)^{1/2}) + n^{-1/2}c\} \geq 1 \text{ a.s.}$$

for n large. The same is valid for the other end of the interval I_n . Now by Lemma 1.1.7, i.e., by the Modulus of Continuity for the Kiefer Process, we have

$$(2.3.14) \quad \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} n^{-1/2} |K_1(E_n(y) + n^{-1/2}c, n) - K_1(y, n)| = o(1) \quad \text{a.s.}$$

Given that the X - and Y -samples are independent, the $F(X)$ - and Y -samples are independent. Since the uniform empirical process $\{\alpha_n(\cdot)\}$ is defined from the $F(X)$ sample, there exists another Kiefer Process $\{K_2(\cdot, \cdot)\}$, independent of $\{K_1(\cdot, \cdot)\}$, such that

$$(2.3.15) \quad \sup_{0 \leq y \leq 1} |\alpha_n(y) - n^{-1/2} K_2(y, n)| = o(1) \quad \text{a.s.}$$

Also we have

$$\{n^{-1/2}(K_1(y, n) + K_2(y, n)); 0 \leq y \leq 1\} \stackrel{D}{=} \{\sqrt{2} B(y); 0 \leq y \leq 1\},$$

where $B(\cdot)$ is a Brownian Bridge, and so

$$\begin{aligned} (2.3.16) \quad & P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F^{-1}(\epsilon_n) \leq x \leq F^{-1}(1 - \epsilon_n)\} \\ &= P\{H(Y, \xi_n, c) \leq c, Y \in I_n, \xi_n \in J_n\} \\ &= P\left\{ \sup_{\epsilon_n \leq y \leq 1 - \epsilon_n} \sqrt{2} B(y) + o_p(1) \leq c \right\}. \end{aligned}$$

Finally, by the P. Lévy Modulus of Continuity for a Brownian Bridge (Lemma 1.1.6) we have

$$\sup_{1 - \epsilon_n \leq y \leq 1} B(y) = o(1) \quad \text{a.s.}$$

and

$$\sup_{0 \leq y \leq \epsilon_n} B(y) = o(1) \quad \text{a.s.}$$

From this and (2.3.15) we obtain the first statement of the theorem. To prove (2.3.2), assume that F satisfies conditions (1.1.12) and define $\epsilon_n < \epsilon'_n < \epsilon''_n$ as in Theorem 2.1.1. Now let

$$A_n := \{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F^{-1}(\epsilon_n) \leq x \leq F^{-1}(1-\epsilon_n)\}$$

$$B_n := \{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F^{-1}(\epsilon'_n) \leq x \leq F^{-1}(1-\epsilon'_n)\}$$

$$C_n := \{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F^{-1}(\epsilon''_n) \leq x \leq F^{-1}(1-\epsilon''_n)\}$$

and let $I(\epsilon_n)$, $I(\epsilon'_n)$, $I(\epsilon''_n)$ be the intervals defined in Theorem 2.2.1.

We have by (2.2.3)

$$P(A_n) = P(A_n \text{ and } I(\epsilon'_n) \subset I(\epsilon_n)) + P(A_n \text{ and } I(\epsilon'_n) \not\subset I(\epsilon_n))$$

$$\leq P(B_n) + o(1),$$

and

$$P(B_n) = P(B_n \text{ and } I(\epsilon''_n) \subset I(\epsilon'_n)) + P(B_n \text{ and } I(\epsilon''_n) \not\subset I(\epsilon'_n))$$

$$\leq P(C_n) + o(1).$$

Hence

$$\begin{aligned} 1 - e^{-c^2} &= \lim_{n \rightarrow \infty} P(A_n) \leq \lim_{n \rightarrow \infty} P(B_n) \leq \lim_{n \rightarrow \infty} P(C_n) \\ &= 1 - e^{-c^2}. \end{aligned}$$

This proves the first part of (2.3.2). Similarly, we can prove the second part of the statements (2.3.1) and (2.3.2) which complete

the statements of Theorem 2.3.1.

A combination of lines like those in the proof of Theorem 2.3.1 yields

THEOREM 2.3.2. Let $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$ be two independent random samples with continuous distribution functions F and G respectively. Assume that F and G both satisfy the conditions (1.1.12). Then

$$(2.3.17) \quad \lim_{n \rightarrow \infty} P\{Q_n[F_n(x) - n^{-1/2}c] \leq QF(x) \leq Q_n[F_n(x) + n^{-1/2}c] \text{ for all } x \text{ such that} \\ F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1 - \epsilon_n)\} \\ = P\left\{ \sup_{0 \leq y \leq 1} |B(y)| \leq c/\sqrt{2} \right\} = K(c/\sqrt{2}),$$

where $c > 0$, $B(\cdot)$ is a Brownian Bridge and ϵ_n a sequence of positive numbers converging to zero such that $n^{1/2}\epsilon_n \rightarrow 0$ and

$$K(c/\sqrt{2}) = 1 - \sum_{k=0}^{\infty} (-1)^{k+1} e^{-k^2 c^2 / 2}.$$

REMARK 2.3.1. All our results here are dealing with the case of having equal sample sizes. The generalization to the case of having unequal sizes is straightforward. The details are omitted.

The next theorems can be proven as Theorem 2.3.1. The results give narrower bands than (2.3.2) and (2.3.17) when we are interested in calculating simultaneous confidence bands in a smaller interval contained in $[0,1]$. The formulas for the limiting distributions are given in Csáki (1974) and in M. Csörgő and Révész (1981a), and the

calculation of these distributions, (as well as those given in the previous theorems) can be obtained using computer programs by Chung (1984).

THEOREM 2.3.3. Assume the conditions of Theorems 2.3.1 and

2.3.2. Let $0 < a_1 < a_2 < 1$ be fixed. Then

$$\begin{aligned}
 (2.3.1) \quad & \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c) \leq QF(x) \text{ for all } x \text{ such that } F_n^{-1}(a_1+) \leq x \leq F_n^{-1}(a_2)\} \\
 & = \lim_{n \rightarrow \infty} P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F_n^{-1}((1-a_2)+) \\
 & \quad \leq x \leq F_n^{-1}(1-a_1)\} \\
 & = P\left\{ \sup_{a_n \leq y \leq a_2} B(y) \leq c/\sqrt{2} \right\}, \quad c > 0.
 \end{aligned}$$

THEOREM 2.3.4. With $a_1 = \epsilon_n$ and a_2 fixed as before, under the conditions of Theorem 2.3.2, we have

$$\begin{aligned}
 (2.3.19) \quad & \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c) \leq QF(x) \text{ for all } x \text{ such that } F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(a_2)\} \\
 & = \lim_{n \rightarrow \infty} P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F_n^{-1}[(1-a_2)+] \\
 & \quad \leq x \leq F_n^{-1}(1-\epsilon_n)\} \\
 & = P\left\{ \sup_{0 \leq y \leq a_2} B(y) \leq c/\sqrt{2} \right\}, \quad c > 0,
 \end{aligned}$$

and if $a_2 = 1 - \epsilon_n$ and a_1 is fixed as before,

$$\begin{aligned}
 (2.3.20) \quad & \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c) \leq QF(x) \text{ for all } x \text{ such that } F_n^{-1}(a_1+) \leq x \leq F_n^{-1}(1-\epsilon_n)\} \\
 & = \lim_{n \rightarrow \infty} P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that } F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1-a_1)\}
 \end{aligned}$$

$$= P\left\{ \sup_{a_1 \leq y \leq 1} B(y) \leq c/\sqrt{2} \right\}, \quad c > 0$$

and if $0 < a_1 \leq a_2 < 1$ are fixed, ϵ_n as before,

$$(2.3.21) \quad \lim_{n \rightarrow \infty} P\{Q_n(F_n(x) - n^{-1/2}c) \leq QF(x) \text{ for all } x \text{ such that}$$

$$F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(a_1) \text{ and } F_n^{-1}(a_2) \leq x \leq F_n^{-1}(1 - \epsilon_n)\}$$

$$= \lim_{n \rightarrow \infty} P\{QF(x) \leq Q_n(F_n(x) + n^{-1/2}c) \text{ for all } x \text{ such that}$$

$$F_n^{-1}(\epsilon_n) \leq x \leq F_n^{-1}(1 - a_2) \text{ and } F_n^{-1}(1 - a_1) \leq x \leq F_n^{-1}(1 - \epsilon_n)\}$$

$$= P\left\{ \sup_{0 \leq y \leq a_1, a_2 \leq y \leq 1} B(y) \leq c/\sqrt{2} \right\}.$$

III. ON THE KUMAR-PATHAK TEST FOR NORMALITY

§3.1. Introduction. The standard Kolmogorov-Smirnov tests are among the most popular tests in dealing with simple goodness-of-fit hypotheses. In this chapter we are interested in an appropriate statistic for composite goodness-of-fit when testing for univariate normality. Kumar and Pathak (1977) proposed an approach based on the notion of sufficiency to obtain a modified version of the Kolmogorov-Smirnov test when dealing with nuisance parameters. However they did not address the problem of obtaining asymptotic distributions of their proposed statistic. In this section we are interested in exploring the asymptotic theory of Kumar and Pathak (1977)'s modified Kolmogorov-Smirnov statistic in the case of testing for normality. We begin with a brief description of the problem at hand.

Suppose $X_1, X_2, \dots, X_n$ is a univariate random sample drawn from a normally distributed population with distribution function $\Phi(\cdot, \mu, \sigma)$ where $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$. Let F_n be the empirical distribution function based on $X_1, \dots, X_n$ and let

$$(3.1.1) \quad \bar{X}_n := n^{-1} \sum_{j=1}^n X_j$$

and

$$(3.1.2) \quad S_n^2 := (n-1)^{-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

Then $(\bar{X}_n, S_n^2)$ is a complete sufficient statistic for (μ, σ) .

Kumar and Pathak (1977) considered the statistic

$$(3.1.3) \quad R_n := \sup_{x \in R} |K_n(x) - F_n(x)|,$$

where

$$K_n(x) := P(X_1 \leq x | \bar{X}_n, S_n^2),$$

and proved that $K_n(\cdot)$ can be replaced by $K_n^*(\cdot)$ when the sample size n is large enough, where $K_n^*(\cdot)$ is defined by

$$(3.1.4) \quad K_n^*(x) = \begin{cases} 0 & \text{if } X_{1:n} > x \\ \Phi(y_k) & \text{if } X_{k:n} \leq x < X_{k+1:n} \\ 1 & \text{if } X_{n:n} \leq x \end{cases}$$

where $\Phi(\cdot)$ is the standard normal distribution function,

$$(3.1.5) \quad y_k := (X_{k:n} - \bar{X}_n) / \left\{ S_n^2 \left(1 + \frac{1}{n(n-2)} \right) - \frac{1}{(n-2)} (X_{k:n} - \bar{X}_n)^2 \right\}^{1/2},$$

$$1 \leq k \leq n$$

and $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ are the order statistics of the given sample.

Now define for $1 \leq k \leq n$,

$$(3.1.6) \quad t_n(y) := n^{1/2} \{ \Phi(y_k) - y \} \quad \text{if } \frac{k-1}{n+1} < y \leq \frac{k}{n+1}$$

and

$$t_n(y) := n^{1/2} \{ \Phi(y_n) - y \} \quad \text{if } \frac{n}{n+1} < y \leq 1$$

and observe that

$$R_n = \max_{1 \leq k \leq n} n^{-1/2} |t_n(k/(n+1))|.$$

Hence, in order to study the asymptotic theory of the Kumar and Pathak (1977)'s test for normality, one may study the asymptotic theory of the process $t_n(\cdot)$ of (3.1.6). We note that the distribution of the statistic R_n is parameter free under the assumption of normality. Hence in the sequel we will assume that $\mu = 0$ and $\sigma = 1$.

Let us denote by $\phi = \Phi'$, i.e. the derivative of the standard normal distribution function, by $Q := \Phi^{-1}$ the inverse of the standard normal distribution function and by Q_n , the sample quantile function. The corresponding general quantile process is defined in 1.1.11 and the parameter estimated quantile process (cf. M. Csörgö and Révész (1981b), p. 205) is defined as

$$(3.1.7) \quad \hat{\rho}_n(y) := n^{1/2} \phi Q(y) \left\{ \frac{Q_n(y) - \bar{X}_n}{S_n} - Q(y) \right\}, \quad 0 < y < 1.$$

In the sequel the following notation will be used.

$$(3.1.8) \quad r_1(n) = \{n^{-1}(\log n) \log \log n\}^{1/2}$$

$$(3.1.9) \quad r_2(n) = \{n^{-1}(\log n)^{3/2}\}$$

and

$$(3.1.10) \quad r_3 = (n, \epsilon) = \max(n^{-\epsilon}, n^{-1/2}(\log \log n)^{1/2}(\log n)^{1/2}), \quad \epsilon > 0.$$

§3.2. On the strong approximation of $t_n(\cdot)$. We first observe that

$$t_n\left(\frac{k}{n+1}\right) = n^{1/2} \left\{ \Phi(y_k) - \Phi Q\left(\frac{k}{n+1}\right) \right\}, \quad 1 \leq k \leq n.$$

Hence, by a two-term Taylor expansion we have

$$(3.2.1) \quad t_n\left(\frac{k}{n+1}\right) = b_n\left(\frac{k}{n+1}\right) + a_n\left(\frac{k}{n+1}\right), \quad 1 \leq k \leq n,$$

where

$$(3.2.2) \quad \begin{cases} b_n\left(\frac{k}{n+1}\right) = n^{\frac{1}{2}} \phi Q\left(\frac{k}{n+1}\right) (y_k - Q\left(\frac{k}{n+1}\right)) \\ a_n\left(\frac{k}{n+1}\right) = b_n^2\left(\frac{k}{n+1}\right) \phi' Q(\theta_{k,n}) / [2\phi^2 Q\left(\frac{k}{n+1}\right) n^{\frac{1}{2}}] \end{cases},$$

with

$$|\theta_{k,n} - \frac{k}{n+1}| \leq |\phi(y_k) - \frac{k}{n+1}|.$$

As to $b_n(\cdot)$ of (3.2.2) we have the following theorem

THEOREM 3.2.1. Let $\hat{\rho}_n(\cdot)$ be as in (3.1.7). Then, for $1 \leq k \leq n$ we have

$$(3.2.3) \quad b_n\left(\frac{k}{n+1}\right) = \hat{\rho}_n\left(\frac{k}{n+1}\right) + \phi Q\left(\frac{k}{n+1}\right) O(n^{\frac{1}{2}} r_2(n)) \quad \text{a.s. .}$$

Theorem 3.2.1 is a direct consequence of definition (3.1.7) and the following two lemmas.

LEMMA 3.2.1. For $\frac{1}{n+1} \leq y \leq \frac{n}{n+1}$ we have

$$(3.2.4) \quad Q_n(y) = O(2 \log n)^{\frac{1}{2}} \quad \text{a.s. .}$$

Proof. By definition of the normed quantile process,

$$(3.2.5) \quad Q_n(y) = \rho_n(y) / n^{\frac{1}{2}} \phi Q(y) + Q(y).$$

We divide the proof in two steps.

Step 1: $\delta_n \leq y \leq 1-\delta_n$ and $\delta_n = 25n^{-1} \log \log n$. In this case

$$(3.2.6) \quad Q_n(y) = Q(y) + \{ \rho_n(y) / (y(1-y) \log \log n)^{1/2} \} \left\{ \frac{y(1-y) \log \log n}{n \phi^2 Q(y)} \right\}^{1/2}$$

$$\leq Q(y) + c_1 \left\{ \frac{y(1-y) \log \log n}{n \phi^2 Q(y)} \right\}^{1/2} \text{ a.s.,}$$

where the last line above follows by Remark 4 of M. Csörgő and Révész (1978) which states that

$$\limsup_{n \rightarrow \infty} \sup_{\delta_n \leq y \leq 1-\delta_n} \frac{|\rho_n(y)|}{(y(1-y) \log \log n)^{1/2}} \leq c_1 \text{ a.s.}$$

where $c_1 > 0$ is an absolute constant.

Now, in order to estimate the right hand side of (3.2.6) we use the following representation for $\phi Q(y)$ as $y \rightarrow 0$ (cf., e.g., Parzen, 1979)

$$(3.2.7) \quad \phi Q(y) \sim y(2 \log \frac{1}{y})^{1/2} \text{ as } y \rightarrow 0.$$

Consequently

$$(3.2.8) \quad Q(y) \sim \left(\log \frac{1}{y^2 \log \frac{1}{y^2}} \right)^{1/2} \text{ as } y \rightarrow 0.$$

By symmetry of the normal p.d.f. $\phi(\cdot)$ there is a similar representation of $\phi Q(\cdot)$ and $Q(\cdot)$ as $(1-y) \rightarrow 0$. Hence

$$(3.2.9) \quad Q(y) = O(2 \log n)^{1/2}$$

for $1/(n+1) \leq y \leq n/(n+1)$. Then the second term in (3.2.6) behaves like

$$\left(\frac{y(1-y) \log \log n}{n \phi^2 Q(y)} \right)^{1/2} \sim \left(\frac{\log \log n}{\log n} \right)^{1/2} \rightarrow 0$$

in $\delta_n \leq y \leq 1 - \delta_n$, as $n \rightarrow \infty$. This proves (3.2.4) in $\delta_n \leq y \leq 1 - \delta_n$.

Step 2: $1/(n+1) \leq y \leq \delta_n$ or $1 - \delta_n \leq y \leq n/(n+1)$. Given (3.2.9), we only need to prove that the first term of (3.2.5) converges almost surely to zero uniformly in the corresponding intervals. Now M. Csörgő et al. (1982, Theorem 3.1) proved that

$$(3.2.10) \quad \sup_{1/(n+1) \leq y \leq \delta_n} |\rho_n(y)| = \sup_{1 - \delta_n \leq y \leq n/(n+1)} |\rho_n(y)| = O(n^{-1/2} (\log \log n)^{1+\gamma}) \quad \text{a.s.}$$

if $\gamma \leq 1$, where γ is the constant of condition (1.1.12) (iii). For the standard normal distribution function we have $\gamma \leq 1$. Combining (3.2.10) and (3.2.7) we have

$$\sup_{1/(n+1) \leq y \leq \delta_n} |\rho_n(y)| / n^{1/2} \phi Q(y) = O\left(\frac{(\log \log n)^2}{(\log n)^{1/2}} \right) \quad \text{a.s.}$$

and by symmetry we obtain the same on the interval $1 - \delta_n \leq y \leq n/(n+1)$.

This completes the proof of Lemma 3.2.1.

LEMMA 3.2.2. Let y_k be as in (3.1.5). Then

$$(3.2.11) \quad \begin{aligned} (i) \quad y_k &= x_{k:n} + O(r_1(n)) \quad \text{a.s., } 1 \leq k \leq n, \\ (ii) \quad y_k &= \frac{x_{k:n} - \bar{x}_n}{s_n} + O(r_2(n)) \quad \text{a.s., } 1 \leq k \leq n. \end{aligned}$$

Proof. By the law of the iterated logarithm, since $\mu = 0$ and $\sigma = 1$, we have

$$\bar{X}_n = n^{-1} \sum_{j=1}^n X_j = O((n^{-1} \log \log n)^{1/2}) \text{ a.s.} \quad (3.2.12)$$

$$\begin{aligned} S_n^2 - 1 &= (n-1)^{-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2 - 1 \\ &= O((n^{-1} \log \log n)^{1/2}) \text{ a.s.} \end{aligned}$$

This combined with (3.2.4) implies

$$(X_{k:n} - \bar{X}_n)^2 / (n-2) = O(n^{-1} \log n) \text{ a.s., } 1 \leq k \leq n,$$

which yields

$$(3.2.13) \quad D_n := \left\{ S_n^2 \left(1 + \frac{1}{n(n-2)} \right) - (X_{k:n} - \bar{X}_n)^2 / (n-2) \right\}^{1/2} = S_n + O(n^{-1} \log n) \text{ a.s.}$$

Since $S_n^2 = 1 + O((n^{-1} \log \log n)^{1/2})$ a.s. then $S_n = 1 + O((n^{-1} \log \log n)^{1/2})$ a.s. hence D_n , the denominator of Y_k , can be expressed as

$$D_n = 1 + O((n^{-1} \log \log n)^{1/2}) \text{ a.s.}$$

Hence

$$\begin{aligned} Y_k &= (X_{k:n} - \bar{X}_n) \{ 1 + O((n^{-1} \log \log n)^{1/2}) \} \\ &= X_{k:n} + O(r_1(n)) \text{ a.s.,} \end{aligned}$$

This proves part (i) of the lemma. Now consider the denominator of Y_k and express it as

$$D_n = S_n + O(n^{-1} \log n) = S_n (1 + O(n^{-1} \log n)) \text{ a.s.}$$

hence

$$y_k = \frac{x_{k:n} - \bar{x}_n}{s_n} (1 + O(n^{-1} \log n)) \quad \text{a.s.}$$

$$= \frac{x_{k:n} - \bar{x}_n}{s_n} + O(r_2(n)) \quad \text{a.s.}$$

The proof of Lemma 3.2.2 is now complete.

REMARK 3.2.1. Let $\rho_n(\cdot)$ be the general quantile process based on the sample $x_1, \dots, x_n$. Then, for some $\varepsilon > 0$ and $1 \leq k \leq n$, we have

$$(3.2.14) \quad \hat{\rho}_n\left(\frac{k}{n+1}\right) = \rho_n\left(\frac{k}{n+1}\right) - \phi Q\left(\frac{k}{n+1}\right) I_n^0 - \phi Q\left(\frac{k}{n+1}\right) Q\left(\frac{k}{n+1}\right) I_n^1 + O(n^{-\varepsilon}) \quad \text{a.s.},$$

where

$$(3.2.15) \quad I_n^0 = \int_0^1 n^{-1/2} K(y, n) dQ(y),$$

and

$$(3.2.16) \quad I_n^1 = \frac{1}{2} \int_0^1 n^{-1/2} K(y, n) dQ^2(y)$$

with $K(\cdot, \cdot)$ the Kiefer Process of Lemma 1.1.9.

The proof of Remark 3.2.1 results from a combination of ideas used in Burke et al. (1979) and M. Csörgő and Révész (1981b). Indeed, M. Csörgő and Révész (1981b, Corollary 5.8.2) expressed the estimated quantile process $\hat{\rho}_n(\cdot)$ in the form

$$\begin{aligned} \hat{\rho}_n\left(\frac{k}{n+1}\right) &= n^{1/2} \phi Q\left(\frac{k}{n+1}\right) \left\{ \frac{x_{k:n} - \bar{x}_n}{s_n} - Q\left(\frac{k}{n+1}\right) \right\} \\ &= n^{1/2} \phi Q\left(\frac{k}{n+1}\right) \frac{1}{s_n} \left\{ x_{k:n} - Q\left(\frac{k}{n+1}\right) - \bar{x}_n - (s_n - 1) Q\left(\frac{k}{n+1}\right) \right\}. \end{aligned}$$

By the law of the iterated logarithm

$$1/S_n = 1 + O((n^{-1} \log \log n)^{1/2}) \quad \text{a.s.},$$

hence

$$(3.2.17) \quad \hat{\rho}_n\left(\frac{k}{n+1}\right) = \{1 + O((n^{-1} \log \log n)^{1/2})\} \left\{ \rho_n\left(\frac{k}{n+1}\right) - \phi Q\left(\frac{k}{n+1}\right) n^{1/2} \bar{X}_n \right. \\ \left. - \phi Q\left(\frac{k}{n+1}\right) Q\left(\frac{k}{n+1}\right) n^{1/2} (S_n - 1) \right\} \quad \text{a.s.}$$

Also we have

$$S_n^2 - 1 = n^{-1} \sum_{j=1}^n \{ (n/(n-1)) X_j^2 - 1 \} + \varepsilon_n$$

where $\varepsilon_n = \frac{n}{n-1} \bar{X}_n^2 = O(n^{-1} \log \log n)$ a.s.. Applying a two-term Taylor expansion to the function $f(x) := (1+x)^{1/2}$ near $x = 0$, we obtain the following representation for S_n :

$$S_n - 1 = (2n)^{-1} \sum_{j=1}^n \{ X_j^2 - 1 \} + \varepsilon'_n$$

where $\varepsilon'_n = O(n^{-1} \log \log n)$ a.s..

Now with $(\mu, \sigma) = (0, 1)$, let

$$\ell_1(X_j, \mu, \sigma) := X_j$$

$$\ell_2(X_j, \mu, \sigma) := (X_j^2 - 1)/2$$

therefore the estimators $(\bar{X}_n, S_n)$ of (μ, σ) can be written as

$$(3.2.18) \quad \begin{cases} n^{\frac{1}{2}} \bar{X}_n = n^{-\frac{1}{2}} \sum_{j=1}^n \ell_1(X_j, \mu, \sigma) \\ n^{\frac{1}{2}} (S_n - 1) = n^{-\frac{1}{2}} \sum_{j=1}^n \ell_2(X_j, \mu, \sigma) + n^{\frac{1}{2}} \varepsilon'_n \end{cases}$$

where $\ell_i(\cdot, \mu, \sigma)$ are measurable functions such that $E(\ell_i(X_j, \mu, \sigma)) = 0$
 $i = 1, 2$, and $n^{\frac{1}{2}} \varepsilon'_n = O(n^{-\frac{1}{2}} \log \log n)$ a.s.

Since the standard normal distribution function Φ satisfies all the conditions of Lemma 1.1.9 (the tail conditions are verified by the property (3.2.7)) we can apply this lemma to obtain, for $K(\cdot, \cdot)$ of (3.2.15),

$$n^{\frac{1}{2}} \bar{X}_n = \int_{-\infty}^{+\infty} x d_x [n^{-\frac{1}{2}} K(\Phi(x), n)] + O(n^{-\varepsilon}) = \int_0^1 n^{-\frac{1}{2}} K(y, n) dQ(y) + O(n^{-\varepsilon})$$

a.s.

and

$$\begin{aligned} n^{\frac{1}{2}} (S_n - 1) &= \int_{-\infty}^{+\infty} (x^2 - 1)/2 d_x [n^{-\frac{1}{2}} K(\Phi(x), n)] + O(n^{-\varepsilon}) \\ &= \frac{1}{2} \int_0^1 n^{-\frac{1}{2}} K(y, n) dQ^2(y) + O(n^{-\varepsilon}) \quad \text{a.s.} \end{aligned}$$

which, together with (3.2.17) imply (3.2.14).

COROLLARY 3.2.1. Using the notation of Remark 3.2.1, we have

for $1 \leq k \leq n$,

$$(3.2.19) \quad b_n\left(\frac{k}{n+1}\right) \stackrel{\text{a.s.}}{=} n^{-\frac{1}{2}} K\left(\frac{k}{n+1}, n\right) - \Phi Q\left(\frac{k}{n+1}\right) I_n^0 - \Phi Q\left(\frac{k}{n+1}\right) Q\left(\frac{k}{n+1}\right) I_n^1 + O(r_3(n, \varepsilon))$$

where $\varepsilon > 0$. (Here, equality with probability one is denoted by $\stackrel{\text{a.s.}}{=}$.)

This is a consequence of (3.2.14) and Theorem 1.1.3.

LEMMA 3.2.3. Let y_k be as in (3.1.5). Then

$$(3.2.20) \quad \phi(y_k) \stackrel{a.s.}{=} \phi(x_{k:n}) + \phi Q\left(\frac{k}{n+1}\right) O(r_1(n)), \quad \delta_n \leq \frac{k}{n+1} \leq 1 - \delta_n$$

$$(3.2.21) \quad \phi(y_k) \stackrel{a.s.}{=} \phi(x_{k:n}) + \phi Q\left(\frac{k}{n+1}\right) o(n^\delta r_1(n)), \quad 1 \leq k \leq n,$$

where $\delta > 0$ is arbitrary.

Proof. By (3.2.11) we have for $1 \leq k \leq n$,

$$(3.2.22) \quad \phi(y_k) = \phi(x_{k:n} + O(r_1(n))) \quad a.s.$$

$$= \phi(x_{k,n}) + O(r_1(n)) \phi(\eta_{k,n}) \quad a.s.,$$

where $\eta_{k,n} = x_{k:n} + \epsilon_{k,n}$ and $\epsilon_{k,n} \stackrel{a.s.}{=} O(r_1(n))$. Hence

$$\eta_{k,n} = Q\left(\frac{k}{n+1}\right) + \omega_k(n),$$

where

$$\omega_k(n) = \epsilon_{k,n} + \rho_n(k/(n+1)) / n^{1/2} \phi Q(k/(n+1)).$$

Now the same argument of Lemma 3.2.1 can be used to prove

$$(3.2.23) \quad \sup_{1 \leq k \leq n} |\omega_k(n)| = O((\log \log n)^2 / (\log n)^{1/2}) \quad a.s.$$

this implies for $1 \leq k \leq n$,

$$\phi(\eta_{k,n}) = \phi Q(k/(n+1)) \exp\{-(1/2) \omega_k^2(n) - \omega_k(n) Q(k/(n+1))\}$$

$$\leq \phi Q(k/(n+1)) \exp\{-C_1 (\log \log n)^4 / \log n\} \exp\{(\log \log n)\}$$

a.s.

$$\leq \phi Q(k/(n+1)) o(n^\delta) \quad a.s.,$$

for any $\delta > 0$ and n sufficiently large. This proves (3.2.21).

From (3.2.22) we can prove (3.2.20) in a similar fashion.

The following theorem estimates $a_n(\cdot)$ of (3.2.1) on $\delta_n \leq \frac{k}{n+1} \leq 1 - \delta_n$.

THEOREM 3.2.2. Let $a_n(\cdot)$ and $\theta_{k,n}$ be as in (3.2.1) and let $\delta_n = 25 n^{-1} \log \log n$. Then

$$(3.2.24) \quad a_n(k/(n+1)) = O(n^{-1/2} \log \log n) \text{ a.s., } \delta_n \leq k/(n+1) \leq 1 - \delta_n.$$

Proof of Theorem 3.2.2.: We observe that

$$a_n\left(\frac{k}{n+1}\right) = (1/2 n^{1/2}) \prod_{i=1}^4 R_{ikn},$$

where

$$(3.2.25) \quad \begin{cases} R_{1kn} = b_n^2\left(\frac{k}{n+1}\right) / \left\{ \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \right\} \\ R_{2kn} = \left\{ \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \right\} / \left\{ \theta_{k,n} (1 - \theta_{k,n}) \right\} \\ R_{3kn} = \theta_{k,n} (1 - \theta_{k,n}) \phi' Q(\theta_{k,n}) / \phi^2 Q(\theta_{k,n}) \\ R_{4kn} = \left\{ \phi' Q(\theta_{k,n}) / \phi Q(k/(n+1)) \right\}^2. \end{cases}$$

By Lemma 3.2.2 (i) we have

$$b_n\left(\frac{k}{n+1}\right) = \rho_n\left(\frac{k}{n+1}\right) + n^{1/2} \phi Q\left(\frac{k}{n+1}\right) O(r_1(n)) \text{ a.s.}$$

Hence,

$$(3.2.26) \quad R_{1kn} \stackrel{\text{a.s.}}{=} \rho_n\left(\frac{k}{n+1}\right) / \left\{ \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \right\} + O(r_1(n)) n^{1/2} \phi Q\left(\frac{k}{n+1}\right) / \left\{ \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \right\}^{1/2}$$

By Remark 4 of M. Csörgő and Révész (1978) which we stated in the proof of Lemma 3.2.1, we have

$$\sup_{\delta_n \leq \frac{k}{n+1} \leq 1-\delta_n} \left| \rho_n\left(\frac{k}{n+1}\right) \right| / \left\{ \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \right\}^{\frac{1}{2}} = O(\log \log n)^{\frac{1}{2}} \text{ a.s..}$$

On the other hand, the second term of the right hand side can be estimated by using (3.2.7). Hence we obtain

$$(3.2.27) \quad \sup_{\delta_n \leq \frac{k}{n+1} \leq 1-\delta_n} |R_{1kn}| = O(\log \log n) \text{ a.s. .}$$

As to R_{2kn} of (3.2.25) we recall that $\theta_{k,n}$ is defined in (3.2.2) such that

$$\left| \frac{k}{n+1} - \theta_{k,n} \right| \leq \left| \phi(y_k) - \frac{k}{n+1} \right|.$$

Hence, by (3.2.20) we have, for $\delta_n \leq \frac{k}{n+1} \leq 1-\delta_n$, that

$$(3.2.28) \quad \left| \frac{k}{n+1} - \theta_{k,n} \right| \leq \left| \phi(x_{k:n}) - \frac{k}{n+1} \right| + O(r_1(n)) \phi Q\left(\frac{k}{n+1}\right) \text{ a.s.}$$

$$\leq p_n := 4 \left\{ n^{-1} \frac{k}{n+1} \left(1 - \frac{k}{n+1}\right) \log \log n \right\}^{\frac{1}{2}} + O(r_1(n)) Q\left(\frac{k}{n+1}\right) \text{ a.s.}$$

where the last line above follows by Lemma 1.1.5. Now for $\delta_n \leq \frac{k}{n+1} \leq 1-\delta_n$ we have

$$\frac{k}{n+1} / \theta_{k,n} \leq 1 + \frac{p_n / \left(\frac{k}{n+1}\right)}{1 - p_n / \left(\frac{k}{n+1}\right)}.$$

On the other hand by (3.2.7) we have

$$\begin{aligned} \frac{p_n}{k/(n+1)} &= 4 \left\{ \frac{1 - \frac{k}{n+1}}{k/(n+1)} \cdot n^{-1} \log \log n \right\}^{\frac{1}{2}} + O(r_1(n)) \frac{\phi_Q(\frac{k}{n+1})}{k/(n+1)} \\ &\leq \frac{4}{5} + O((\log n)^{\frac{1}{2}} r_1(n)) . \end{aligned}$$

Hence

$$(3.2.29) \quad \frac{k}{n+1} / \theta_{k,n} \leq 6 \quad \text{a.s. .}$$

Similarly we can prove that

$$(3.2.30) \quad (1 - \frac{k}{n+1}) / (1 - \theta_{k,n}) \leq 6 \quad \text{a.s. .}$$

Hence

$$(3.2.31) \quad \sup_{\delta_n \leq \frac{k}{n+1} \leq 1 - \delta_n} |R_{2kn}| \leq 36 \quad \text{a.s. .}$$

Now the normal distribution $\Phi(\cdot)$ satisfy the conditions 1.1.12 (iii) with $\gamma > 0$, hence

$$(3.2.32) \quad \sup_{\frac{1}{n+1} \leq \frac{k}{n+1} \leq \frac{n}{n+1}} |R_{3kn}| \leq \gamma .$$

By Lemma 1.1.1, since $\Phi(\cdot)$ satisfies 1.1.12 (iii),

$$R_{4kn}^{\frac{1}{2}} := \frac{\phi_Q(\theta_{k,n})}{\phi_Q(\frac{k}{n+1})} \leq \left\{ \frac{\theta_{k,n} \vee (\frac{k}{n+1})}{\theta_{k,n} \wedge (\frac{k}{n+1})} \cdot \frac{1 - \theta_{k,n} \wedge (\frac{k}{n+1})}{1 - \theta_{k,n} \vee (\frac{k}{n+1})} \right\}^{\gamma} ,$$

and by the argument used to obtain (3.2.29) and (3.2.30) we have

$$(3.2.33) \quad R_{4kn} \leq (2 \times 36)^2 \quad \text{a.s.}$$

The required result follows by combining (3.2.27), (3.2.31),

(3.2.32) and (3.2.33).

The main result of this section is the following

THEOREM 3.2.3. Let $t_n(\cdot)$ be as in (3.1.6). Let $G(\cdot, \cdot)$ be a two parameter Gaussian Process defined by

$$(3.2.34) \quad G(y, n) = n^{-1/2} K(y, n) - \phi Q(y) \int_0^1 n^{-1/2} K(x, n) dQ(x) - \phi(Q(y)) Q(y) \cdot \int_0^1 \frac{1}{2} n^{-1/2} K(x, n) dQ^2(x)$$

where $K(\cdot, \cdot)$ is the Kiefer Process defined by (1.1.14). Then, for some $\varepsilon > 0$,

$$(3.2.35) \quad \sup_{\frac{1}{n+1} \leq y \leq \frac{n}{n+1}} |t_n(y) - G(y, n)| = O(r_3(n, \varepsilon)) \text{ a.s.}$$

where $r_3(\cdot, \cdot)$ is as in (3.1.10).

Proof. We recall that

$$t_n(y) = b_n(y) + a_n(y),$$

where by Theorem 3.2.2 the last term is like $O(n^{-1/2} \log \log n)$ in $\delta_n \leq y \leq 1 - \delta_n$ and the first term can be expressed as

$$b_n(y) = G(y, n) + O(r_3(n, \varepsilon)) \text{ a.s., } \frac{1}{n+1} \leq y \leq \frac{n}{n+1}, \quad \varepsilon > 0$$

by Corollary 3.2.1. Hence, with $\varepsilon > 0$

$$(3.2.36) \quad \sup_{\delta_n \leq y \leq 1 - \delta_n} |t_n(y) - G(y, n)| = O(r_3(n, \varepsilon)) \text{ a.s..}$$

Now if we show

$$(3.2.37) \quad \sup_{\frac{1}{n+1} \leq y \leq \delta_n} |t_n(y)| = O(n^{-\frac{1}{2}} \log \log n) \text{ a.s.}$$

and

$$(3.2.38) \quad \sup_{\frac{1}{n+1} \leq y \leq \delta_n} |G(y, n)| = O(n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} (\log \log n)^{\frac{1}{2}}) \text{ a.s.}$$

we shall obtain the required result (due to symmetry the same rates are obtained on the tails near 1). To prove (3.2.37), we have for any $\delta > 0$ by (3.2.21),

$$\begin{aligned} \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq y \leq \delta_n} |t_n(y)| &= \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq y \leq \delta_n} n^{\frac{1}{2}} \left| \phi(y_k) - \frac{k}{n+1} \right| \\ &= \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq y \leq \delta_n} n^{\frac{1}{2}} \left| \phi(X_{k:n}) - \frac{k}{n+1} + \phi Q(y) \circ (n^{\delta} r_1(n)) \right| \\ &\leq \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq y \leq \delta_n} n^{\frac{1}{2}} \left| \phi(X_{k:n}) - \frac{k}{n+1} \right| + \\ &\quad + o(n^{\frac{1}{2}+\delta} r_1(n)) \cdot \phi Q(\delta_n) \text{ a.s.} \end{aligned}$$

Now M. Csörgő and Révész (1981b, Theorem 4.5.6) proved that the uniform quantile process $u_n(\cdot)$ is uniformly bounded in the interval

$0 \leq y \leq \delta_n$ by

$$\sup_{0 \leq y \leq \delta_n} |u_n(y)| = O(n^{-\frac{1}{2}} \log \log n) \text{ a.s.,}$$

hence

$$(3.2.39) \quad \sup_{1/(n+1) \leq y \leq \delta_n} n^{1/2} |\phi(X_{k:n}) - \frac{k}{n+1}| = O(n^{-1/2} \log \log n) \text{ a.s. .}$$

Also the behaviour of $\phi Q(\cdot)$ as $y \rightarrow 0$ implies (cf. 3.2.7),

$$(3.2.40) \quad o(n^{1/2+\delta} r_1(n)) \phi Q(\delta_n) = o(n^{-1+\delta} (\log n) (\log \log n)^{3/2}) .$$

Hence (3.2.37) is obtained.

To prove (3.2.38) we observe that by the law of the iterated logarithm and (3.2.7), (3.2.8),

$$(3.2.41) \quad \frac{1}{n+1} \sup_{1/(n+1) \leq y \leq \delta_n} |G(y, n)| \leq \frac{1}{n+1} \sup_{1/(n+1) \leq y \leq \delta_n} \left| \frac{K(y, n)}{n^{1/2}} \right| + O(\log \log n)^{1/2} .$$

$$\{ \phi Q(\delta_n) + \phi Q(\delta_n) \cdot Q(\delta_n) \} \text{ a.s.}$$

$$\leq \frac{1}{n+1} \sup_{1/(n+1) \leq y \leq \delta_n} \frac{|K(y, n)|}{n^{1/2}} + O(n^{-1} (\log \log n)^{3/2} \log n) .$$

Next, by the Functional Law of the Iterated Logarithm for the Kiefer Process (Lemma 1.1.8) we have

$$(3.2.42) \quad \sup_{\frac{1}{n+1} \leq y \leq \delta_n} |K(y, n)| = O((\log n)^{1/2} (\log \log n)^{1/2}) , \text{ a.s. .}$$

This together with (3.2.41) imply (3.2.38). The proof of the theorem is now complete.

COROLLARY 3.2.2. Let R_n be as in (3.1.2), let $B(\cdot)$ be a Brownian Bridge and define

$$(3.2.43) \quad G(y) := B(y) - \phi Q(y) \int_0^1 B(t) dQ(t) - \phi(Q(y)) Q(y) \int_0^1 \frac{1}{2} B(t) dQ^2(t)$$

then

$$(3.2.44) \quad R_n \xrightarrow{D} \sup_{0 < y < 1} |G(y)|$$

where $\xrightarrow{D}$ stands for convergence in distribution.

REMARK 3.2.2. We observe that Aly (1983b) proved that Lilliefors' statistic for testing normality has the same limit in distribution as R_n above. Hence when using the Kumar-Pathak test we may use Lilliefors' (1967) tables when the sample size is large.

IV TESTING FOR MULTIVARIATE NORMALITY

§4.1. Introduction. We are concerned with assessing multivariate normality when the parameters are unknown. A formal classical approach is based on the following.

Suppose $X_1, X_2, \dots, X_n$ is a sample of k -dimensional random vectors drawn from a normally distributed population with cumulative distribution function $\Phi(\cdot, \theta_0)$, with $\theta_0 = (\mu, \Sigma)$, where the mean μ is an arbitrary k -dimensional vector and the covariance matrix Σ is a (non-singular) positive definite matrix. Let $\hat{\theta}_n = (\bar{X}_n, S_n)$ be the maximum likelihood estimators of μ and Σ respectively. One may consider the multivariate process

$$\hat{\beta}_n(x) := n^{1/2} \{F_n(x) - \Phi(x, \hat{\theta}_n)\}, \quad x \in \mathbb{R}^{(k)},$$

where $F_n(\cdot)$ is the empirical distribution function of the multivariate sample, (i.e., $F_n(x)$ is the proportion of the X_j , $1 \leq j \leq n$, which satisfy $X_j \leq x$ in the usual partial ordering of $\mathbb{R}^{(k)}$) and $\Phi(\cdot, \hat{\theta}_n)$ is as above with parameter $\hat{\theta}_n$. Burke et al. (1979) and later Koziol (1983) studied the process $\hat{\beta}_n(\cdot)$. Burke et al. (1979) proved the weak convergence of $\hat{\beta}_n(\cdot)$ to a Gaussian Process $D(x)$ with mean zero and covariance matrix depending on the unknown parameters. This result has no immediate practical application since little is known about the distribution of its functionals of interest.

Now a procedure is called invariant if the statistics involved are invariant under affine transformations of the observed values (i.e., if the statistics T verify that given any non-singular $k \times k$

matrix A and any k -dimensional vector b , $T(AX_1+b, AX_2+b, \dots, AX_n+b) = T(X_1, X_2, \dots, X_n)$ for any random sample $X_1, \dots, X_n$. In such a case, we may do all our calculations as if the true parameters of the underlying distribution were $\theta_0 = (0, \Pi)$ where 0 is the k -dimensional zero vector and Π is the $k \times k$ unit matrix.

Let us denote by X^T the $1 \times k$ transpose vector of the $k \times 1$ column vector X . We recall that $\hat{\theta}_n = (\bar{X}_n, S_n)$, defined by

$$(4.1.1) \quad \begin{aligned} \bar{X}_n &:= n^{-1} \sum_{j=1}^n X_j, \quad X_j \in \mathbb{R}^{(k)} \\ S_n &:= n^{-1} \sum_{j=1}^n (X_j - \bar{X}_n) \cdot (X_j - \bar{X}_n)^T, \end{aligned}$$

are the maximum likelihood estimators of the mean and covariance of the population, respectively. We know that any positive definite $k \times k$ matrix S can be expressed as the product of a non-singular matrix by its transpose. Let us denote this non-singular matrix by $S^{\frac{1}{2}}$, i.e., for any positive definite matrix S , let $S^{\frac{1}{2}}$ be such that $S = S^{\frac{1}{2}} \cdot (S^{\frac{1}{2}})^T$. Now we define the Mahalanobis distances of the X -sample by

$$(4.1.2) \quad \tilde{Y}_j = \psi(X_j, \hat{\theta}_n) := (X_j - \bar{X}_n)^T S_n^{-1} (X_j - \bar{X}_n), \quad j=1, 2, \dots, n.$$

The $\tilde{Y}_j$'s are invariant statistics. When $\theta_0 = (\mu, \Sigma)$ is the true parameter of the underlying distribution we have

$$(4.1.3) \quad Y_j = \psi(X_j, \theta_0) := (X_j - \mu)^T \Sigma^{-1} (X_j - \mu), \quad j=1, 2, \dots, n,$$

where under the assumption of joint normality the Y_j 's represent a

random sample from a chi-squared distribution function with k -degrees of freedom.

Some invariant tests for multivariate normality have been introduced previously. There are several techniques with differing sensitivities to different types of departures. Mardia (1970, 1974, 1975) proposed multivariate measures of skewness and kurtosis, b_{1k} and b_{2k} respectively. The statistic $nb_{1k}/6$ is asymptotically distributed as $\chi^2_{(r_k)}$ (chi-squared with r_k degrees of freedom) and b_{2k} is asymptotically distributed as $N(\mu_k, \rho_k^2)$ (normal with mean μ_k and variance ρ_k^2), where r_k , μ_k and ρ_k^2 depend on the number k of coordinates of the random vector X .

Malkovich and Afifi (1973) proposed a generalization of the univariate Shapiro-Wilk's SW statistic. This technique is based on the characterization of the multivariate normal distribution in terms of univariate normality of all linear combinations of the variables. Let C be any k -dimensional vector, and let $Z_j = C^T \cdot X_j$ and $Z_{1:n} < \dots < Z_{n:n}$ the corresponding order statistics. Define

$$SW(C) := \frac{[\sum_j a_j (Z_{j:n} - \bar{Z}_n)]^2}{\sum_j (Z_{j:n} - \bar{Z}_n)^2}$$

where the a_j 's are weights, the same coefficients which were tabulated in Shapiro and Wilk (1965). The statistic proposed by Malkovich and Afifi (1973) is

$$SW^* = \min_C SW(C)$$

The procedure is "lower-tail" (i.e., small values of SW^* indicate non-normality). This test is also coordinate free (invariant), but direct numerical evaluation of the statistic SW^* is quite difficult and precise sampling and asymptotical distributions of SW^* were not established so far.

Then there are the *radius and angles* graphical procedures. These have been suggested by Andrews et al. (1973) among others, and are based on the following spherical coordinate transformation: for $Z := \Sigma^{-1/2}(X - \mu) \stackrel{D}{=} N_{(k)}(0, I)$, consider

$$Z \rightarrow (Y, \Theta),$$

where, under the assumption of joint normality, $Y = (X - \mu)^T \Sigma^{-1/2} (X - \mu) \stackrel{D}{=} \chi_{(k)}^2$ and Θ is a $(k-1)$ -dimensional vector of angles (corresponding to Z) which is uniformly distributed on the unit hypersphere S_{k-1} (see Mardia et al. (1979), Chapter 15). The estimated $\tilde{Y}_j$'s are asymptotically distributed as $\chi_{(k)}^2$. The graphical procedure often recommended was to plot the ordered $\tilde{Y}_j$ against the expected order statistics for samples of size n from the $\chi_{(k)}^2$ distribution. A departure from linearity in the plot would indicate non-normality (see Gnanadesikan, 1977). Probability plots were also recommended for the estimated angles.

Hawkins (1981) proposed statistics based also on the squared-radius $\tilde{Y} = \psi(X, \hat{\theta}_n)$. They could be useful as descriptive statistics but there is no asymptotic theory for them. Koziol (1982) has made

serious efforts to provide an invariant formal test based on the Mahalanobis distance of the sample. However his arguments as to the weak convergence of his process do not appear to be correct. More recently, S. Csörgő (1984) proposed a test based on the asymptotic behaviour of the multivariate "studentised" empirical characteristic function. He defined a maximal deviation statistic $M_n^{(k)}$ and obtained an upper bound for the asymptotic distribution of $M_n^{(k)}$. This is the first goodness of fit test proven consistent against all multivariate alternatives with finite mean and positive definite covariance matrix. The result is a large sample, conservative, Kolmogorov-type test. Also it involves a considerable amount of computing effort.

Cox and Small (1978) suggested that a significance test based on the graphical procedure illustrated above (for the Mahalanobis distances $\tilde{Y}_j$'s) would be useful. Koziol (1982) studied a process based on the equivalent consideration of comparing departures of the "empirical distribution function" of the $\tilde{Y}_j$'s from their postulated chi-squared with k-degrees of freedom distribution function $F_{(k)}(\cdot)$, namely

$$(4.1.4) \quad W_n(s) := n^{1/2}(\tilde{F}_n(s) - F_{(k)}(s)), \quad 0 \leq s < +\infty.$$

Thus if an asymptotic theory for $W_n(\cdot)$ were established, we could reduce the more complicated problem of multivariate testing when the parameters are unknown into a univariate formulation with the added advantage of invariance. Of course, we would be testing only a characteristic of multivariate normality, namely, that the statistics

$Y_j = (X_j - \mu)^T \Sigma^{-1} (X_j - \mu)$ are distributed as chi-squared with k -degrees of freedom rv.

Koziol's (1982) line of proof for the convergence of $W_n(\cdot)$ to a Gaussian Process $W(\cdot)$ does not seem to be correct (we shall explain in detail later). He does not find the limiting expression of $W_n(\cdot)$, though by an educated guess he obtains the covariance structure for $W(\cdot)$:

$$(4.1.5) \quad \text{cov}(W(s), W(t)) = F_{(k)}(s \wedge t) - F_{(k)}(s) \cdot F_{(k)}(t) \\ - (2/k) s f_{(k)}(s) t f_{(k)}(t),$$

where $f_{(k)}(\cdot) = F'_{(k)}(\cdot)$ is the probability density function of the chi-squared with k degrees of freedom rv., and s, t are positive real numbers. He also defines a Cramér-von Mises type statistic of the form

$$J_n = J_n^M(k) := \int_0^M W_n^2(s) dF_{(k)}(s), \quad M > 0$$

which, if we assume the weak convergence of $W_n(\cdot)$ (in $\mathcal{D}[0, +\infty)$) to be correct, converges in distribution to the rv.

$$(4.1.6) \quad J = J_n^M(k) := \int_0^M W^2(s) dF_{(k)}(s), \quad \text{for all } M > 0.$$

Koziol produces a set of asymptotic critical values for $J^M(k)$ with $k = 2, 6$ and 10 . Since he did not prove weak convergence of $W_n(\cdot)$ nor convergence in distribution of $J_n^M(k)$, he computed his asymptotic values only on the assumption that $W(\cdot)$ is Gaussian and has the

covariance structure defined in (4.1.5). We cannot, at this stage, comment very much about these values except for the fact that he compares them against the values of the statistic $J_n^M(k)$, when, $J_n^M(k)$ is computed for n independent k -variate normal vectors drawn at random by simulation; he concludes that one could reasonably use the asymptotic critical value for J_n for testing purposes where k is small or n is large. The calculation of $J_n^M(k)$ yields a formula similar to the classical one for the Cramér-von Mises statistic; let the "uniform empirical process" $\tilde{E}_n(\cdot)$ be defined as

$$\tilde{E}_n(\cdot) = \begin{cases} 0 & \text{if } Z_{1:n} > y \\ j/n & \text{if } Z_{j:n} \leq y < Z_{j+1:n} \\ 1 & \text{if } Z_{n:n} \leq y \end{cases}$$

where $Z_{j:n} = F_{(k)}(\tilde{Y}_{j:n})$ and $\tilde{Y}_{j:n}$ are the ordered Mahalanobis distances of the sample. Hence for M large enough,

$$\begin{aligned} J_n^M(k) &= n \int_0^1 (\tilde{E}_n(y) - y)^2 dy \\ &= \sum_{j=1}^n (Z_{j:n} - (j-0.5)/n)^2 + 1/12n. \end{aligned}$$

In this chapter we propose to find an asymptotic representation for $W_n(\cdot)$ and to obtain the asymptotic distribution of two Cramér-von Mises type functionals of $W_n(\cdot)$, which can be used to test multivariate normality. One of these functionals, the J_n above, was introduced by Malkovich (1971) and studied by Kosiol (1982). In particular, we first prove the convergence in distribution of certain type of continuous

linear functionals of $W_n(\cdot)$ to those of the process defined by

$$(4.1.7) \quad W(s) := B(F_{(k)}(s)) + sf_{(k)}(s) \int_0^{+\infty} (u/k-1) dB(F_{(k)}(u)),$$

$$0 \leq s < +\infty,$$

where $B(\cdot)$ is a Brownian Bridge, $F_{(k)}$ is the distribution function of the chi-squared with k -degrees of freedom rv., and $f_{(k)}(\cdot) = F'_{(k)}(\cdot)$. We also prove that for every $M > 0$,

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} P\left\{ \sup_{0 \leq s \leq M} |W_n(s)| > \varepsilon \right\} = 0,$$

and the weak convergence of the continuous process defined by the Fourier Transform operator applied to $W_n(\cdot)$, that is, we obtain

$$(4.1.8) \quad F(W_n)(\cdot) - F(\tilde{W}_n)(\cdot) \xrightarrow{D[0,+\infty)} 0$$

where $\tilde{W}_n$ is an auxiliary process such that $\tilde{W}_n \xrightarrow{D[0,+\infty)} W$. We also obtain the convergence in distribution of the following Cramér-von Mises type statistics: for all $k \geq 2$, and all $M > 0$, we have

$$(4.1.9) \quad L_n = L_n^M(k) := \int_0^M W_n^2(s) ds \xrightarrow{D} L = L^M(k) := \int_0^M W^2(s) ds$$

and

$$J_n = J_n^M(k) := \int_0^M W_n^2(s) dF_{(k)}(s) \xrightarrow{D} J = J^M(k) := \int_0^M W^2(s) dF_{(k)}(s)$$

where $W(\cdot)$ is the process defined in (4.1.7). Due to the Karhunen-Loève expansion of a Gaussian Process (see M. Csörgő and Révész, 1981b, p. 249), the random variables $L^M(k)$ and $J^M(k)$ can be expressed

as an infinite sum of weighted independent chi-squared rv. with one degree of freedom. We shall obtain tables for $L^M(k)$ and $J^M(k)$ in the near future and the computations will be based on this infinite sum representation.

Our aim is to test the composite hypothesis, for $k \geq 2$:

H_0 : X is distributed as multivariate normal $N_{(k)}(\mu, \Sigma)$ for some (μ, Σ) , Σ positive definite.

Let us denote by $E(X)$ and $\text{cov}(X, Y)$ the respective expectation of the random variable X and covariance matrix defined by the random vectors X and Y . Let H_1 be the alternative hypothesis:

H_1 : the law of $Y = (X - \mu)^T \Sigma^{-1} (X - \mu)$ is given by any distribution function $G_{(k)}$ such that $G_{(k)} \neq F_{(k)}$ on a set of positive Lebesgue measure in $0 < x < +\infty$, for some $(\mu, \Sigma) = (E(X), \text{cov}(X, Y))$ with positive definite Σ , and such that $Y_{n:n} = o_p(n^{1-\epsilon})$ for any $0 < \epsilon < 1/2$.

We propose two different tests, one for each statistic $L_n^M(k)$ and $J_n^M(k)$, where we reject H_0 vs. H_1 if the value of the corresponding statistic is too large, and we prove that the tests are consistent.

We also define another statistic

$$M_n = M_n(k) := \int_0^{+\infty} W_n(s) dF_{(k)}(s), \quad k \neq 3,$$

and prove that under the null hypothesis H_0 , M_n converges in distribution to a normal random variable with mean zero and variance

$$\sigma_k^2 = 1/12 - (2/k) \{(k-1)!\}^2 / \{\Gamma(k/2)\}^4 2^{2k}, \quad k \neq 3.$$

Now let H_2 be the alternative hypothesis, for $k \geq 2$, $k \neq 3$,

H_2 : The law of $Y = (X - \mu)^T \Sigma^{-1} (X - \mu)$ is given by $G_{(k)}$ for some $(\mu, \Sigma) =$

$(E(X), \text{cov}(X, X))$ with positive definite Σ , such that

$\int_0^{+\infty} \{G_{(k)}(s) - F_{(k)}(s)\} dF_{(k)}(s) \neq 0$, and such that $Y_{n:n} = o_p(n^{1-\epsilon})$ for any $0 < \epsilon < 1/2$, where $Y_{n:n}$ is the largest order statistic of the Y-sample.

Then we reject H_0 vs. H_2 if the absolute value $|M_n(k)|$ is too large. We prove that this test is also consistent. Of course, the Cramér-von Mises type statistics $L_n^M(k)$ and $J_n^M(k)$ should provide a better test than the $M_n(k)$, but the practical implementation of the $M_n(k)$ -test is so simple that we decided to include it in this work.

Consider now the statistic $J_n^M(k)$. Malkovich and Afifi (1973) and later Giorgi and Fattorini (1976) compared the empirical power of SW^* , $J_n^M(k)$ and other statistics, when testing H_0 against a certain group of non-normal alternatives. These power studies are not yet extensive but the conclusions are similar to the univariate case, i.e., SW^* shows the greatest power for the alternatives considered, $J_n^M(k)$ had nearly the same power of SW^* and in general, other tests are no better than SW^* (for any n). (For the univariate power studies see Shapiro et al. (1968)).

The univariate Shapiro-Wilk statistic SW (see Shapiro and Wilk, 1965) "feel the shape" of the density function $\phi(\cdot)$ especially on the tails, due to the weights " a_j " (see the definition of SW^*), and this sensitivity on the tails appears to be extended to the multivariate case. Now, the statistics L_n and J_n can be expressed as

$$\begin{aligned}
 (4.2.5) \quad \hat{\theta}_n - \theta_0 &= (n^{-1} \sum_j X_j, n^{-1} \sum_j (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T - \mathbb{I}) \\
 &= (n^{-1} \sum_j X_j, n^{-1} \sum_j (X_j X_j^T - \mathbb{I})) + O(n^{-1} \log \log n) \quad \text{a.s.}
 \end{aligned}$$

by the law of the iterated logarithm for sums of i.i.d.r. vectors with mean zero and finite variance. Let us denote by H_n the vector in the first term above, i.e.,

$$\begin{aligned}
 (4.2.6) \quad n^{1/2}(\hat{\theta}_n - \theta_0) &= (n^{-1/2} \sum_j X_j, n^{-1/2} \sum_j (X_j X_j^T - \mathbb{I})) + O(n^{-1/2} \log \log n) \\
 &= H_n + O(n^{-1/2} \log \log n) \quad \text{a.s.},
 \end{aligned}$$

where H_n is a random vector converging in distribution to a vector normally distributed with zero vector mean and covariance matrix with elements 0, 1 and 2. (This is by the Central Limit Theorem).

We also have

$$(4.2.7) \quad n^{1/2}(\tilde{Y}_j - Y_j) = (-2X_j, -X_j X_j^T)^T H_n + O(n^{-1/2} \log n \cdot \log \log n) \quad \text{a.s.}$$

Indeed, by a two-term Taylor expansion

$$\begin{aligned}
 n^{1/2}(\tilde{Y}_j - Y_j) &= n^{1/2} \{ \psi(X_j, \hat{\theta}_n) - \psi(X_j, \theta_0) \} = \psi_\theta(X_j, \theta_0)^T n^{1/2}(\hat{\theta}_n - \theta_0) + \frac{n^{1/2}}{2} \psi_{\theta\theta}(X_j, \xi_n)^T \\
 &\quad \{ (\hat{\theta}_n - \theta_0) \cdot (\hat{\theta}_n - \theta_0)^T \},
 \end{aligned}$$

where $|\xi_n - \theta_0| \leq |\hat{\theta}_n - \theta_0|$ and $\psi_\theta(\cdot, \cdot)$ and $\psi_{\theta\theta}(\cdot, \cdot)$ are the first and second order partial derivatives of $\psi(\cdot, \cdot)$ with respect to θ . Koziol (1982b) proved that if we express $\psi(\cdot, \cdot)$ as

Neuhaus considers a univariate version of the process $\hat{\beta}_n(\cdot)$ defined at the beginning of this section, i.e. for $x \in \mathbb{R}$,

$$\begin{aligned}\hat{\beta}_n(x) &= n^{1/2}[F_n(x) - F(x, \hat{\theta}_n)] \\ &= n^{1/2}[F_n(x) - F(x, \theta_0)] - n^{1/2}(\hat{\theta}_n - \theta_0) \frac{\partial}{\partial \theta} F(x, \xi),\end{aligned}$$

where ξ is an intermediate point between $\hat{\theta}_n$ and θ_0 (θ_0 is the true parameter of the distribution). Here θ_0 is unknown and estimated by $\hat{\theta}_n$ whereas in $F_{(k)}(\cdot)$ the only parameter k is fixed and known. In the $\hat{\beta}_n$ process it is assumed that $\frac{\partial}{\partial \theta} F(x, \cdot)$ has sufficiently nice properties and that $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges weakly. Durbin's process is more similar to our $W_n(\cdot)$: let $x_1, \dots, x_n$ be a sample of independent observations from a distribution with continuous $\frac{\partial}{\partial \theta} F(x, \theta)$ depending on a vector θ of p parameters; let

$$x_n(t) := n^{1/2}[\hat{F}_n(t, \hat{\theta}_n) - t], \quad 0 \leq t \leq 1,$$

where $\hat{F}_n(t, \hat{\theta}_n)$ is the proportion of values x_i such that $F(x_i, \hat{\theta}_n) \leq t$. Durbin proves that there is a sequence of random functions $t_n(\cdot)$ such that

$$\hat{F}_n(\cdot, \hat{\theta}_n) = F_n(t_n(\cdot), \theta_0),$$

where $F_n(t, \theta_0)$ is the proportion of values x_i such that $F(x_i, \theta_0) \leq t$, and such that

$$t_n(t) - t = -n^{1/2}(\hat{\theta}_n - \theta_0) \frac{\partial F}{\partial \theta}(x, \xi),$$

where ξ is an intermediate point between $\hat{\theta}_n$ and θ_0 and $x = F^{-1}(t, \theta_0)$. Hence

$$\begin{aligned} X_n(t) &= n^{1/2}[F_n(t_n(t), \theta_0) - t] \\ &= \alpha_n(t_n(t)) - n^{1/2}(\hat{\theta}_n - \theta_0) \frac{\partial}{\partial \theta} F(x, \xi), \quad 0 \leq t \leq 1, \end{aligned}$$

where $\alpha_n(\cdot)$ is the uniform empirical process. Both Neuhaus and Durbin's processes can be expressed as sums of independent identically distributed random variables with mean zero and finite variance evaluated at a random function or vector. With the strong approximation theory available nowadays, it is readily seen how $X_n(\cdot)$ or $\beta_n(\cdot)$ converge, (cf. Burke et al. (1979)).

In the case of $W_n(\cdot)$, a sequence like Durbin's $t_n(\cdot)$ would have an approximate form

$$s_n = \psi(\psi^{-1}(s, \hat{\theta}_n), \theta_0)$$

if such function could be well defined. The idea is to find a function that maps $\tilde{Y}_j = \psi(X_j, \hat{\theta}_n)$ into $Y_j = \psi(X_j, \theta_0)$. We run into trouble when we want to calculate the difference $n^{1/2}(s_n - s)$: let us write

$$\begin{aligned} W_n(s) &= n^{1/2}(\tilde{F}_n(s, \hat{\theta}_n) - F_{(k)}(s)) = n^{-1/2} \sum_{j=1}^n \{I_s(\tilde{Y}_j) - F_{(k)}(s)\} \\ &= n^{-1/2} \sum_{j=1}^n \{I_s(\tilde{Y}_j) - I_s(Y_j)\} + n^{-1/2} \sum_{j=1}^n \{I_s(Y_j) - F_{(k)}(s)\}, \end{aligned}$$

where $I_s(\cdot)$ is the set function of $0 \leq t \leq s$. Let us denote the second term above by $\alpha_n(F_{(k)}(s))$. Koziol considered the auxiliary

process $W_n^*(\cdot)$ defined by

$$(4.1.10) \quad W_n^*(s) - \alpha_n(F_{(k)}(s)) := -n^{1/2}(\hat{\theta}_n - \theta_0) \frac{\partial F_{(k)}}{\partial \theta} \{\psi(x, \theta)\} \Big|_{\theta=\theta_0}$$

when $F_{(k)}(\cdot)$ is expressed implicitly as a function of s by

$s = \psi(x, \theta_0) = x^T \cdot x$. But $W_n^*(\cdot)$ is not univocally determined,

even up to distribution, since it yields different covariance structures according to the choice of x . Indeed, if we write

$$V_n(s) := W_n^*(s) - \alpha_n(F_{(k)}(s)), \quad 0 \leq s < +\infty,$$

the covariance of $V_n(\cdot)$, calculated by Koziol (1982b), is given by

$$\text{cov}(V_n(s_1), V_n(s_2)) = f_{(k)}(s_1) f_{(k)}(s_2) [4x_1^T x_2 + 2(x_1^T x_2)^2],$$

where $x_1^T \cdot x_1 = s_1$ and $x_2^T \cdot x_2 = s_2$ and $f_{(k)}(\cdot) = F'_{(k)}(\cdot)$. We

immediately see that the covariance varies with the angle between x_1 and x_2 and even if we leave the angle fixed we do not obtain the appropriate covariance function for the limiting process of $V_n(\cdot)$.

There is no right choice of the vector x such that $s = \psi(x, \theta_0) = x^T \cdot x$ in the definition (4.1.10). Now consider

$$(4.1.11) \quad \begin{aligned} Z_n(s) &:= W_n(s) - n^{-1/2} \sum_{j=1}^n \{I_s(Y_j) - F_{(k)}(s)\} \\ &= n^{-1/2} \sum_{j=1}^n \{I_s(\tilde{Y}_j) - I_s(Y_j)\} \end{aligned}$$

where

$$I_s(x) := \begin{cases} 1 & \text{if } x \leq s \\ 0 & \text{otherwise.} \end{cases}$$

That is, for a fixed outcome of our probability space, $Z_n(\cdot)$ is a step function. Moreover, it is a combination of set functions for sets whose Lebesgue measure goes to zero as $n \rightarrow \infty$. What Koziol tried to do when he defined $W_n^*(\cdot)$ in (4.1.10); was to apply the mean value theorem to the difference

$$Z_n(\cdot) = W_n(\cdot) - \alpha_n(F_{(k)}(\cdot)) .$$

This however cannot be done in the usual sense of differentiation. The classical method to calculate the limiting form of a random function like $Z_n(\cdot)$ is to consider it as a generalized function. This is what we do in the following section.

In the sequel we shall use the following notation: we write for the sequence of random variables X_n

$$X_n = O_p(\delta_n) \text{ as } n \rightarrow \infty ,$$

where δ_n is a sequence of positive numbers, if and only if

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} P(|X_n|/\delta_n > \epsilon) = 0 .$$

Hence, when $\delta_n = 1 \forall n$, the statement $X_n = O_p(1)$ is equivalent to saying that the sequence of random variables X_n is tight (in the sense of Billingsley (1968)). A sequence of random functions Y_n in $\mathcal{D}(-\infty, +\infty)$ converges weakly to Y in $\mathcal{D}(-\infty, +\infty)$, i.e.,

$$Y_n(\cdot) \xrightarrow{\mathcal{D}[-\infty, +\infty]} Y(\cdot)$$

if and only if for every $M < \infty$, $Y_n(\cdot) \xrightarrow{\mathcal{D}[-M, M]} Y(\cdot)$.

4.2. Calculation of the Limit of $Z_n(\cdot)$. Given the random vector sample $X_1, X_2, \dots, X_n$ from a normal distribution $N_{(k)}(\mu, \Sigma)$ where μ is a k -dimensional vector and Σ is a $k \times k$ positive definite matrix, our process $W_n(\cdot)$ depends on the statistics $\tilde{Y}_j = \psi(X_j, \hat{\theta}_n)$ defined in (4.1.1). These statistics are invariant under affine transformations of the observed vectors X_j . We recall, from Section 4.1, that

$$\Sigma^{-1/2}(X_j - \mu) \stackrel{D}{=} N_{(k)}(0, I)$$

where 0 is the k -zero vector and I is the $k \times k$ identity matrix. We can assume, without loss of generality, that the true parameter of the distribution is $\theta_0 = (0, I)$.

Let $Z_n(\cdot)$ be the process defined in (4.1.11). For a fixed element of our probability space, let

$$(4.2.1) \quad R_j(s) := I_s(\tilde{Y}_j) - I_s(Y_j), \quad 0 \leq s < +\infty,$$

where $I_s(\cdot)$ is the set function of $0 \leq t \leq s$, as defined in the previous section. Then $R_j(\cdot)$ is a right continuous function taking values $-1, 0$ or 1 and hence $Z_n(\cdot)$ is a right continuous step function since

$$(4.2.2) \quad Z_n(s) = n^{-1/2} \sum_{j=1}^n R_j(s) \quad 0 \leq s < +\infty.$$

If we define for $s < 0$,

$$Z_n(s) := 0,$$

we have the following

THEOREM 4.2.1. $Z_n(\cdot)$ is a right continuous step function defined on the real line such that

$$(4.2.3) \quad Z_n(\cdot) \in L^1(-\infty, +\infty) \text{ and } Z_n(\cdot) \in L^2(-\infty, +\infty).$$

The theorem will follow from the fact that $Z_n(\cdot)$ vanishes outside the interval $0 \leq s \leq Y_{n:n} + r_n$, where $Y_{n:n}$ denotes the largest order statistic of the sample $Y_1, \dots, Y_n$ and r_n is a random variable defined by

$$r_n := \max_{1 \leq j \leq n} |\tilde{Y}_j - Y_j|.$$

In order to prove this, the following two lemmas are useful. First, for $M > 0$ we define

$$r_n^M := \max\{|\tilde{Y}_j - Y_j| : Y_j \leq M, j=1, 2, \dots, n\}.$$

We have

LEMMA 4.1.1. Let r_n and r_n^M be as above for $M > 0$. Then

$$(4.2.4) \quad \begin{cases} n^{1/2} r_n = O((\log n \log \log n)^{1/2}) \text{ a.s.} \\ n^{1/2} r_n \leq Y_{n:n}^{1/2} O_p(1) = O_p((\log n)^{1/2}) \\ n^{1/2} r_n^M \leq M^{1/2} O_p(1) = O_p(1) \end{cases}$$

Proof: By definition of $\hat{\theta}_n$ (see 4.1.1), which is the maximum likelihood estimate of θ_0 , and $\theta_0 = (0, \mathbb{I})$, we have

$$\begin{aligned}
 (4.2.5) \quad \hat{\theta}_n - \theta_0 &= (n^{-1} \sum_j X_j, n^{-1} \sum_j (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T - \mathbb{I}) \\
 &= (n^{-1} \sum_j X_j, n^{-1} \sum_j (X_j X_j^T - \mathbb{I})) + O(n^{-1} \log \log n) \quad \text{a.s.}
 \end{aligned}$$

by the law of the iterated logarithm for sums of i.i.d.r. vectors with mean zero and finite variance. Let us denote by H_n the vector in the first term above, i.e.,

$$\begin{aligned}
 (4.2.6) \quad n^{1/2}(\hat{\theta}_n - \theta_0) &= (n^{-1/2} \sum_j X_j, n^{-1/2} \sum_j (X_j X_j^T - \mathbb{I})) + O(n^{-1/2} \log \log n) \\
 &= H_n + O(n^{-1/2} \log \log n) \quad \text{a.s.},
 \end{aligned}$$

where H_n is a random vector converging in distribution to a vector normally distributed with zero vector mean and covariance matrix with elements 0, 1 and 2. (This is by the Central Limit Theorem).

We also have

$$(4.2.7) \quad n^{1/2}(\tilde{Y}_j - Y_j) = (-2X_j, -X_j X_j^T)^T H_n + O(n^{-1/2} \log n \cdot \log \log n) \quad \text{a.s.}$$

Indeed, by a two-term Taylor expansion

$$\begin{aligned}
 n^{1/2}(\tilde{Y}_j - Y_j) &= n^{1/2} \{ \psi(X_j, \hat{\theta}_n) - \psi(X_j, \theta_0) \} = \psi_\theta(X_j, \theta_0)^T n^{1/2}(\hat{\theta}_n - \theta_0) + \frac{n^{1/2}}{2} \psi_{\theta\theta}(X_j, \xi_n)^T \\
 &\quad \{ (\hat{\theta}_n - \theta_0) \cdot (\hat{\theta}_n - \theta_0)^T \},
 \end{aligned}$$

where $|\xi_n - \theta_0| \leq |\hat{\theta}_n - \theta_0|$ and $\psi_\theta(\cdot, \cdot)$ and $\psi_{\theta\theta}(\cdot, \cdot)$ are the first and second order partial derivatives of $\psi(\cdot, \cdot)$ with respect to θ . Kozlowski (1982b) proved that if we express $\psi(\cdot, \cdot)$ as

$$\psi(x, \theta) = (x - \theta_1)^T \theta_2^{-1} (x - \theta_1)$$

where $\theta = (\theta_1, \theta_2)$, $\theta_1 \in \mathbb{R}^{(k)}$, θ_2 a $k \times k$ matrix, then

$$\psi_{\theta}(x_j, \theta_0) = \frac{\partial}{\partial \theta} \psi(x_j, \theta) \Big|_{\theta=\theta_0}$$

$$= (-2\theta_2^{-1}(x_j - \theta_1), -\theta_2^{-1}(x_j - \theta_1)(x_j - \theta_1)^T \theta_2^{-1}) \Big|_{\theta=\theta_0}$$

$$= (-2x_j, -x_j \cdot x_j^T)$$

We know that $x_j = O(\log n)^{1/2}$ a.s. (see 3.2.4) and ξ_n is bounded, then

$$(4.2.9) \quad \frac{n^{1/2}}{2} \psi_{\theta\theta}(x_j, \xi_n)^T \{(\hat{\theta}_n - \theta_0)(\hat{\theta}_n - \theta_0)^T\} = O(n^{-1/2} \log n \cdot \log \log n) \text{ a.s.}$$

Combining (4.2.5); (4.2.8) and (4.2.9) we obtain (4.2.7). Let us denote

by $\|\cdot\|$, the Euclidean norm in $\mathbb{R}^{k+k^2}$ ($\|x\| = \|(x_1, \dots, x_{k+k^2})\| =$

$(\sum_i x_i^2)^{1/2}$). Now (4.2.7) implies

$$n^{1/2} r_n = \max_{1 \leq j \leq n} n^{1/2} |\tilde{y}_j - y_j| \leq \max_{1 \leq j \leq n} \|(-2x_j, -x_j \cdot x_j^T)\| \cdot \|H_n\| + O(n^{-1/2} \log n \cdot \log \log n) \text{ a.s.}$$

Also by the law of the iterated logarithm

$$H_n = O((\log \log n)^{1/2}) \text{ a.s.}$$

This together with the fact that there is $C_1 > 0$ such that

$$\max_{1 \leq j \leq n} \|(-2x_j, x_j \cdot x_j^T)\| \leq C_1 \max_{1 \leq j \leq n} y_j^{1/2} \leq C_1 y_{n:n}^{1/2}$$

imply

$$n^{1/2} r_n = O(\log n \cdot \log \log n)^{1/2} \text{ a.s. as } n \rightarrow \infty$$

22

OF / DE

and

$$n^{\frac{1}{2}} r_n = Y_{n:n}^{\frac{1}{2}} O_p(1) \text{ a.s. as } n \rightarrow \infty.$$

For $M > 0$, with a similar argument we also obtain

$$n^{\frac{1}{2}} r^M = O_p(1) \text{ as } n \rightarrow \infty,$$

which proves the lemma.

LEMMA 4.2.2. Let $R_j(\cdot)$ be as in (4.2.1). Then for every

$s \geq 0$,

$$(4.2.10) \quad Y_j \notin (s-r_n, s+r_n) \text{ implies } R_j(s) = 0$$

and

$$(4.2.11) \quad Y_j \notin (s-r_n^M, s+r_n^M) \text{ implies } R_j(s) = 0 \text{ for all } 0 \leq s \leq M.$$

Proof. If $Y_j < s$ and $Y_j \notin (s-r_n, s+r_n)$, then

$$\tilde{Y}_j = Y_j + (\tilde{Y}_j - Y_j) \leq Y_j + r_n < (s-r_n) + r_n = s,$$

so that $R_j(s) = 1-1 = 0$. If $Y_j > s$ and $Y_j \notin (s-r_n, s+r_n)$, then

$$\tilde{Y}_j \geq Y_j - r_n > s + r_n - r_n = s,$$

so that $R_j(s) = 0$. This proves (4.2.10). Statement (4.2.11) can be proved similarly.

Proof of Theorem 4.2.1. By Lemma 4.2.2, if $s > Y_{n:n} + r_n$,

for every $j = 1, 2, \dots, n$, we have

$$Y_j + r_n \leq Y_{n:n} + r_n < s,$$

that is,

$$Y_j \notin (s-r_n, s+r_n) .$$

Hence for $j = 1, 2, \dots, n$ we have

$$R_j(s) = 0$$

and

$$Z_n(s) = n^{-1/2} \sum_{j=1}^n R_j(s) = 0 .$$

As a consequence of Lemma 4.2.1 and 4.2.2 we can obtain bounds for the process $Z_n(\cdot)$. These are given in the following theorem.

THEOREM 4.2.2. We have for every $M > 0$,

$$(4.2.12) \quad \sup_{0 \leq s \leq M} |Z_n(s)| = O_p(1) \text{ as } n \rightarrow \infty, \text{ and}$$

$$\sup_{0 \leq s \leq +\infty} |Z_n(s)| = O_p[(\log n)^{1/2}] \text{ as } n \rightarrow \infty .$$

Proof. From Lemma 4.2.2 we have

$$|Z_n(s)| \leq n^{-1/2} \sum_{j=1}^n |R_j(s)| \leq n^{-1/2} \#\{j : Y_j \in (s-r_n, s+r_n)\} .$$

We define

$$X(s, n) := \#\{j : Y_j \in (s-r_n, s+r_n)\}$$

and

$$X^M(s, n) := \#\{j : Y_j \in (s-r_n^M, s+r_n^M)\} .$$

Then by the definition of the empirical distribution function of the sample $Y_1, \dots, Y_n$, we have

$$X(s, n) = n\{F_n(s+r_n) - F_n(s-r_n)\} .$$

We also have

$$F_n(s) = F_{(k)}(s) + n^{-1/2} \alpha_n(F_{(k)}(s)), \quad 0 \leq s < +\infty,$$

where $\alpha_n(\cdot)$ is the uniform empirical process and $F_n(\cdot) = E_n(F_{(k)}(\cdot))$.

Then

$$X(s, n) = n\{F_{(k)}(s+r_n) - F_{(k)}(s-r_n) + n^{-1/2}[\alpha_n(F_{(k)}(s+r_n)) - \alpha_n(F_{(k)}(s-r_n))]\}.$$

Now, there exists an almost sure representation of $\alpha_n(\cdot)$ in terms of a Kiefer Process $K(\cdot, \cdot)$ (see (1.1.4) in Chapter I) so we can write

$$X(s, n) = n\{F_{(k)}(s+r_n) - F_{(k)}(s-r_n)\} + \{K(F_{(k)}(s+r_n), n) - K(F_{(k)}(s-r_n), n)\} + O((\log n)^2) \quad \text{a.s.}$$

Also, by the Modulus of Continuity of the Kiefer Process (Lemma 1.1.7) we have

$$|K(F_{(k)}(s+r_n), n) - K(F_{(k)}(s-r_n), n)| \leq 2F'_{(k)}(\xi_s)(nr_n \log \frac{1}{r_n})^{1/2} \quad \text{a.s.}$$

where ξ_s is a point in $(s-r_n, s+r_n)$ and $F'_{(k)}(\cdot) = f_{(k)}(\cdot)$. Hence

$$X(s, n) \leq 2nf_{(k)}(s_1)r_n + 2 \sup_t |f_{(k)}(t)| (nr_n \log \frac{1}{r_n})^{1/2} + O((\log n)^2) \quad \text{a.s.}$$

where $s_1 \in (s-r_n, s+r_n)$.

By Lemma 4.2.1 we have that

$$(r_n \log 1/r_n)^{1/2} = o(1) \quad \text{a.s.}$$

Thus

$$|Z_n(s)| \leq n^{-1/2} X(s, n) \leq 2f_{(k)}(s_1)n^{1/2}r_n + o(1) \quad \text{a.s.}$$

where $s_1 \in (s-r_n, s+r_n)$. Therefore

$$\sup_{0 \leq s \leq +\infty} |Z_n(s)| = O_p((\log n)^{1/2}) \text{ as } n \rightarrow \infty.$$

Similarly, we can prove that for $M > 0$,

$$\sup_{0 \leq s \leq M} |Z_n(s)| \leq \sup_{0 \leq s \leq M} n^{-1/2} X^M(s, n) \leq 2f_{(k)}(s_1) n^{1/2} r_n^M + o(1) \text{ a.s.,}$$

where s_1 is as above. Therefore

$$\sup_{0 \leq s \leq M} |Z_n(s)| = O_p(1) \text{ as } n \rightarrow \infty.$$

Here the constant in the $O_p(\cdot)$ depends on M .

REMARK 4.2.1. These bounds for $Z_n(\cdot)$ are not needed in the sequel, but they are interesting by themselves and they help us understand the process $Z_n(\cdot)$ better.

Now we shall calculate the expression of the limit of $Z_n(\cdot)$. Let us denote by T_n the random variable with mean zero and variance $2/k$ defined by

$$(4.2.13) \quad T_n := n^{-1/2} \sum_{j=1}^n (Y_j/k - 1)$$

where $Y_1, \dots, Y_n$ is the random sample defined in (4.1.3), $Y_j = X_j^T \cdot X_j$ with $\theta_0 = (0, \Pi)$, distributed as a chi-squared with k -degrees of freedom. Hence by the Central Limit Theorem

$$(4.2.14) \quad T_n \xrightarrow{D} N(0, 2/k).$$

Let us also denote by

$$\alpha_n(F_{(k)}(s)) = n^{-1} \sum_{j=1}^n \{I_s(Y_j) - F_{(k)}(s)\}.$$

We prove first the following

LEMMA 4.2.3. We have, for every $n = 1, 2, \dots$

$$(4.2.15) \quad \text{cov}(T_n, \alpha_n(F_{(k)}(s))) = -(2/k) s f_{(k)}(s),$$

where $f_{(k)}(\cdot) = \frac{d}{ds} F_{(k)}(\cdot)$.

Proof. By independence of the random sample $Y_1, \dots, Y_n$,

$$\begin{aligned} \text{cov}(T_n, \alpha_n(F_{(k)}(s))) &= n^{-1} \sum_{j=1}^n \text{cov}(Y_j/k, I_s(Y_j)) \\ &= E(I_s(Y_1) \cdot Y_1/k) - F_{(k)}(s) \\ &= \int_0^s \frac{t}{k} f_{(k)}(t) dt - \int_0^s f_{(k)}(t) dt \\ &= (k (k/2) 2^{k/2})^{-1} \left\{ \int_0^s t^{k/2} e^{-t/2} dt \right. \\ &\quad \left. - k \int_0^s t^{k/2-1} e^{-t/2} dt \right\}. \end{aligned}$$

Integrating by parts,

$$\int_0^s t^{k/2} e^{-t/2} dt = -2t^{k/2} e^{-t/2} \Big|_0^s - k \int_0^s t^{k/2-1} e^{-t/2} dt,$$

or

$$\int_0^s t^{k/2} e^{-t/2} dt - k \int_0^s t^{k/2-1} e^{-t/2} dt = -2s^{k/2} e^{-s/2},$$

which implies (4.2.15).

Now let $g(s) := s f_{(k)}(s)$. We have

THEOREM 4.2.3. Let $h(\cdot)$ be any differentiable real or complex valued function such that $h(\cdot)$ and $h'(\cdot) = \frac{d}{ds} h(\cdot)$ are bounded in $0 \leq s < +\infty$. Then we have

$$(4.2.16) \quad \int_0^{+\infty} z_n(s) h(s) ds = \frac{1}{n} \sum_{j=1}^n h(Y_j) (2X_j, X_j \cdot X_j^T)^T H_n + (\sup_s |h| + \sup_s |h'|) O\left(\frac{\log n \cdot \log \log n}{n^{\frac{1}{2}}}\right) \text{ a.s.}$$

where the constant in the $O(\cdot)$ does not depend on $h(\cdot)$ and H_n is the $k+k^2$ -dimensional vector defined in (4.2.6). We also have

$$(4.2.17) \quad \int_0^{+\infty} z_n(s) h(s) ds = T_n \int_0^{+\infty} g(s) h(s) ds + o_h(1) \text{ a.s.}$$

where $o_h(1)$ converges almost surely to zero as $n \rightarrow \infty$, but the constant in $o_h(\cdot)$ depends on $h(\cdot)$.

Proof. Let us consider first a real valued $h(\cdot)$ defined on the positive real line, and satisfying the conditions of the theorem. By definition of $R_j(\cdot)$ in (4.2.1),

$$R_j(s) = \begin{cases} 0 & \text{if } \tilde{Y}_j \wedge Y_j \leq s \text{ or } Y_j \vee \tilde{Y}_j > s \\ -1 & \text{if } Y_j \leq s < \tilde{Y}_j \\ 1 & \text{if } \tilde{Y}_j \leq s < Y_j. \end{cases}$$

Then by the mean value theorem for integrals,

$$\begin{aligned} n^{\frac{1}{2}} \int R_j(s) h(s) ds &= -n^{\frac{1}{2}} \int_{Y_j}^{\tilde{Y}_j} h(s) ds \\ &= -n^{\frac{1}{2}} h(\xi_j) (\tilde{Y}_j - Y_j), \end{aligned}$$

where $|\xi_j - Y_j| \leq |\tilde{Y}_j - Y_j|$. Now by a one term Taylor expansion, the expression above can be written as

$$-h(Y_j) n^{\frac{1}{2}} (\tilde{Y}_j - Y_j) - h'(\eta_j) n^{\frac{1}{2}} (\xi_j - Y_j) (\tilde{Y}_j - Y_j),$$

where η_j is an intermediate point between ξ_j and Y_j . Applying (4.2.7) to the first term above and (4.2.4) to the second term, we have

$$\begin{aligned} n^{\frac{1}{2}} \int R_j(s) h(s) ds &= h(Y_j) (2X_j, X_j \cdot X_j^T)^T \cdot H_n + |h|_{\infty} O(n^{-\frac{1}{2}} \log n \cdot \log \log n) + \\ &+ |h'|_{\infty} O(n^{-\frac{1}{2}} \log n \cdot \log \log n) \quad \text{a.s.}, \end{aligned}$$

where by $|\cdot|_{\infty}$ we denote the sup-norm of the function within the bars, and the constant in the $O(\cdot)$ does not depend on h . Hence

$$\begin{aligned} \int_0^{+\infty} Z_n(s) h(s) ds &= n^{-1} \sum_{j=1}^n (n^{\frac{1}{2}} \int_0^{+\infty} R_j(s) h(s) ds) \\ &= n^{-1} \sum_{j=1}^n h(Y_j) (2X_j, X_j \cdot X_j^T)^T \cdot H_n + (|h|_{\infty} + |h'|_{\infty}) O(n^{-\frac{1}{2}} \log n \cdot \log \log n) \quad \text{a.s.} \end{aligned}$$

which proves (4.2.16) for real valued h .

Let $h(\cdot)$ be a complex valued function. Then

$$h(\cdot) = h_1(\cdot) + ih_2(\cdot)$$

where $h_1(\cdot)$ are real valued functions. We write

$$\begin{aligned} \int_0^{+\infty} Z_n(s) h(s) ds &= \int_0^{+\infty} Z_n(s) h_1(s) ds + i \int_0^{+\infty} Z_n(s) h_2(s) ds \\ &= \frac{1}{n} \sum_{j=1}^n \{h_1(Y_j) + i h_2(Y_j)\} (2X_j, X_j \cdot X_j^T)^T \cdot H_n + C_1 (|h|_\infty + |h'|_\infty) O(n^{-1/2} \log n \cdot \log \log n) \\ &\quad \text{a.s.} \end{aligned}$$

where $C_1 > 0$ is a constant such that

$$|x| + |y| \leq C_1 (x^2 + y^2)^{1/2}$$

for every pair (x, y) in the real plane.

Now, by the law of the iterated logarithm, we obtain

$$(4.2.18) \quad n^{-1/2} \sum_{j=1}^n h(Y_j) (2X_j, X_j \cdot X_j^T)^T = E[h(X^T \cdot X) (2X, X \cdot X^T)] + O(n^{-1/2} (\log \log n)^{1/2}) \quad \text{a.s.}$$

where the constant in $O(\cdot)$ depends on $h(\cdot)$. Let us calculate this vector of expectations, element by element. For each $i = 1, 2, \dots, k$,

$$E[h(X^T \cdot X) X^i] = \int_{\mathbb{R}^k} h(X^T \cdot X) X^i d\phi(X) = 0$$

by the symmetry of the multivariate normal distribution. For each $i \neq \ell$, $i, \ell = 1, \dots, k$

$$E[h(X^T \cdot X) X^i \cdot X^\ell] = \int_{\mathbb{R}^k} h(X^T \cdot X) X^i \cdot X^\ell d\phi(X) = 0$$

also by symmetry (or, alternatively, by calculation through transformation to polar coordinates). For $i = \ell$,

$$\begin{aligned} (4.2.19) \quad E[h(X^T \cdot X) (X^i)^2] &= \int_{\mathbb{R}^k} \phi(X^T \cdot X) (X^i)^2 d\phi(X) \\ &= (1/k) \int_0^{+\infty} s f_{(k)}(s) h(s) ds. \end{aligned}$$

This last integral was calculated by transformation to polar coordinates in Koziol (1982b), i.e. by the following transformation:

$$x^1 = s^{\frac{1}{2}} \cos \gamma_1$$

$$x^2 = s^{\frac{1}{2}} \sin \gamma_1 \cos \gamma_2$$

$$x^{k-1} = s^{\frac{1}{2}} \sin \gamma_1 \cdot \sin \gamma_2 \dots \sin \gamma_{k-2} \cos \gamma_{k-1}$$

$$x^k = s^{\frac{1}{2}} \sin \gamma_1 \cdot \sin \gamma_2 \cdot \dots \cdot \sin \gamma_{k-1}$$

The H_n in (4.2.16) is the $k + k^2$ -vector

$$H_n = (n^{-\frac{1}{2}} \sum_{j=1}^n x_j, n^{-\frac{1}{2}} \sum_{j=1}^n (x_j \cdot x_j^T - I)) = O(\log \log n)^{\frac{1}{2}} \text{ a.s.}$$

and the (i, i) coordinate is given by

$$H_n^{ii} = n^{-\frac{1}{2}} \sum_{j=1}^n \{(x_j^i)^2 - 1\}$$

Therefore

$$\begin{aligned} (4.2,20) \quad E[h(x^T \cdot x) (2x, x \cdot x^T)]^T \cdot H_n &= k^{-1} \sum_{i=1}^k \int_0^{+\infty} sf_{(k)}(s) h(s) ds \cdot H_n^{ii} \\ &= k^{-1} \sum_{i=1}^k (n^{-\frac{1}{2}} \sum_{j=1}^n \{(x_j^i)^2 - 1\}) \int_0^{+\infty} sf_{(k)}(s) h(s) ds \\ &= n^{-\frac{1}{2}} \sum_{j=1}^n \{y_j / k - 1\} \int_0^{+\infty} g(s) h(s) ds \\ &= T_n \int_0^{+\infty} g(s) h(s) ds \end{aligned}$$

Combining (4.2.18) and (4.2.20) we obtain (4.2.17) and the theorem is proved.

COROLLARY 4.2.1. Let $W_n(\cdot)$ be the process defined in (4.1.4) with $k \neq 3$. Then

$$(4.2.21) \quad M_n(k) := \int_0^{+\infty} W_n(s) dF_{(k)}(s) \xrightarrow{D} N(0, \sigma_k^2)$$

where

$$(4.2.22) \quad \sigma_k^2 := 1/12 - \frac{(2/k)((k-1)!)^2}{\{\Gamma(k/2)\}^4 2^{2k}}.$$

Proof. Let $k \neq 3$. Then the function

$$h(s) := f_{(k)}(s)$$

satisfies the conditions of Theorem 4.2.3. By (4.1.11),

$$W_n(s) = \alpha_n(F_{(k)}(s)) + Z_n(s), \quad 0 \leq s < +\infty,$$

and

$$\begin{aligned} \int_0^{+\infty} W_n(s) dF_{(k)}(s) &= \int_0^{+\infty} \alpha_n(F_{(k)}(s)) dF_{(k)}(s) + \int_0^{+\infty} Z_n(s) f_{(k)}(s) ds \\ &= n^{-1/2} \sum_{j=1}^n \int_0^{+\infty} (I_s(Y_j) - F_{(k)}(s)) dF_{(k)}(s) + n^{-1/2} \sum_{j=1}^n (Y_j/k-1) \int_0^{+\infty} s f_{(k)}^2(s) ds + o(1) \text{ a.s.} \end{aligned}$$

Let us call

$$(4.2.23) \quad V_j := \int_0^{+\infty} (I_s(Y_j) - F_{(k)}(s)) dF_{(k)}(s) + (Y_j/k-1) \int_0^{+\infty} s f_{(k)}^2(s) ds.$$

Since the Y_j 's are i.i.d.r.v. (distributed like chi-squared with k degrees of freedom), the V_j 's are i.i.d.r.v. with mean zero and variance given by

$$\begin{aligned}
 (4.2.24) \quad E(V_j^2) &= E\left\{\int_0^{+\infty} (I_s(Y_j) - F_{(k)}(s)) dF_{(k)}(s)\right\}^2 \\
 &\quad + 2 \operatorname{cov}\left(\int_0^{+\infty} (I_s(Y_j) - F_{(k)}(s)) dF_{(k)}(s), (Y_j/k-1)\delta_k\right) \\
 &\quad + E(Y_j/k - 1)^2 \delta_k^2
 \end{aligned}$$

where

$$\begin{aligned}
 \delta_k &:= \int_0^{+\infty} s f_{(k)}^2(s) ds \\
 &= \{[\Gamma(k/2)]^2 2^k\}^{-1} \int_0^{+\infty} s^{k-1} e^{-s} ds \\
 &= (k-1)! / \{[\Gamma(k/2)]^2 2^k\}.
 \end{aligned}$$

Now, let $U_j := F_{(k)}(Y_j)$, then

$$\begin{aligned}
 (4.2.25) \quad E\left\{\int_0^{+\infty} (I_s(Y_j) - F_{(k)}(s)) dF_{(k)}(s)\right\}^2 &= E\left\{\int_0^1 (I_Y(U_j) - y) dy\right\}^2 \\
 &= E\left\{\int_0^1 dx \int_0^1 dy (I_Y(U_j) - y) (I_X(U_j) - x)\right\} \\
 &= \int_0^1 dx \int_0^1 dy E\{I_{X \wedge Y}(U_j) - xy\} \\
 &= \int_0^1 dx \int_0^1 dy (x \wedge y - xy) = 1/12.
 \end{aligned}$$

We also have

$$\begin{aligned}
 (4.2.26) \quad \operatorname{cov}\left\{\int_0^{+\infty} (I_s(Y_j) - F_{(k)}(s)) dF_{(k)}(s), Y_j/k-1\right\} \\
 &= \int_0^{+\infty} \operatorname{cov}\{I_s(Y_j), Y_j/k-1\} dF_{(k)}(s) \\
 &= -(2/k)\delta_k.
 \end{aligned}$$

Combining (4.2.24), (4.2.25), (4.2.26) and recalling that

$$E[(Y_1/k - 1)^2] = 2/k,$$

we obtain

$$\begin{aligned} E(V_j^2) &= 1/12 + 2(-2/k)\delta_k^2 + (2/k)\delta_k^2 \\ &= 1/12 - (2/k)\delta_k^2 \\ &= \sigma_k^2. \end{aligned}$$

We have then,

$$\int_0^{+\infty} W_n(s) dF_{(k)}(s) = n^{-1/2} \sum_{j=1}^n V_j + o(1) \text{ a.s.}$$

where the V_j 's are i.i.d.r.v. with mean zero and variance σ_k^2 , hence

$$M_n(k) = \int_0^{+\infty} W_n(s) dF_{(k)}(s) \xrightarrow{D} N(0, \sigma_k^2)$$

where $N(0, \sigma_k^2)$ denotes a normal random variable with mean zero and variance σ_k^2 .

REMARK 4.2.2. Theorems 4.2.2 and 4.2.3 imply that the $[0, M]$ -sup-norm of the process $Y_n(\cdot) := Z_n(\cdot) - T_n g(\cdot)$ is an $O_p(1)$ and the convergence in distribution of certain type of continuous functionals to zero. Although we cannot prove tightness of the process $Y_n(\cdot)$ with our tools, if we also had tightness of $Y_n(\cdot)$ in the interval $0 \leq s \leq M$, $M > 0$, then we could obtain weak convergence. This is due to the following theorem:

THEOREM 4.2.4. Assume that $Y_n(\cdot)$ is a sequence of processes in $L^1(-\infty, +\infty) \cap D$, where D is the space of right-continuous functions with left hand limits on $(-\infty, +\infty)$. Assume that for every $M > 0$, we have

- (i) $\sup_{-M \leq s \leq M} |Y_n(s)| = O_p(1)$
- (ii) $\{Y_n(\cdot); -M \leq s \leq M\}$ is tight
- (iii) For every test function $h \in C_0^\infty(-\infty, +\infty) = \{\text{space of infinitely differentiable functions with compact support}\}$,

$$\int Y_n(s)h(s)ds = o_p(1).$$

Then the process $Y_n(\cdot)$ converges weakly to zero in $D(-\infty, +\infty)$.

Proof. Let $M > 0$. Tightness of the process $Y_n(\cdot)$ in $-M \leq s \leq M$ implies, by Prohorov's Theorem, that for every subsequence (n') there exists a subsubsequence (n'') such that

$$Y_{n''}(\cdot) \xrightarrow{D[-M, M]} Y(\cdot)$$

for some random measurable function $Y(\cdot)$, i.e. convergence in distribution, of the sequence $Y_{n''}(\cdot)$. This implies

$$\sup_{-M \leq s \leq M} |Y_{n''}(s)| \xrightarrow{D} \sup_{-M \leq s \leq M} |Y(s)|$$

where by condition (i),

$$\sup_{-M \leq s \leq M} |Y(s)| < +\infty \text{ a.s.}$$

Then for every $h \in C_0^\infty[-M, M]$

$$\int_{-M}^M |hY| ds \leq \sup_{-M \leq s \leq M} |Y(s)| \int_{-M}^M |h(s)| ds < +\infty \text{ a.s.,}$$

and weak convergence of $Y_{n''}(\cdot)$ implies weak convergence of its continuous functionals, hence

$$\int_{-M}^M Y_{n''}(s)h(s)ds \xrightarrow{D} \int_{-M}^M Y(s)h(s)ds.$$

This together with (iii) imply

$$(4.2.27) \quad \int_{-M}^M Y(s)h(s)ds = 0 \text{ a.s. for every } h \in C_0^\infty[-M, M].$$

Now for each different $h(\cdot)$ there is a new set Ω_h in the probability space with $P(\Omega_h) = 1$ where (4.2.27) holds. Let $g \geq 0$ be an infinitely differentiable function with support in $[-M, M]$ such that

$$\int_{-M}^M g(s)ds = 1.$$

Let $g_m(s) := m^{-1}g(s/m)$ and let

$$h_m(\cdot) := \operatorname{sgn} Y * g_m(\cdot) = \int \operatorname{sgn} Y(s)g_m(\cdot-s)ds$$

where $\operatorname{sgn} Y$ denotes the sign function of $Y(\cdot)$, i.e.,

$$\operatorname{sgn} Y(s) = \begin{cases} 1 & \text{if } Y(s) \geq 0 \\ 0 & \text{if } Y(s) < 0. \end{cases}$$

Let Ω_m be the set with probability one such that for every element $\omega \in \Omega_m$ we have

$$\int_{-M}^M Y(s)h_m(s)ds = 0.$$

Consider the set $\Omega_0 := \bigcap_{m=1}^{\infty} \Omega_m$. Then

$$P(\Omega_0) = \lim_{r \rightarrow \infty} P\left(\bigcap_{m=1}^r \Omega_m\right) = 1.$$

Hence for every $\omega \in \Omega_0$ and $m \geq 1$ we have

$$\int_{-M}^M Y(s) h_m(s) ds = 0$$

which, by the Lebesgue Convergence Theorem implies

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \int_{-M}^M Y(s) h_m(s) ds = \int_{-M}^M \lim_{m \rightarrow \infty} Y(s) h_m(s) ds \\ &= \int_{-M}^M |Y(s)| ds. \end{aligned}$$

Therefore

$$|Y(s)| = 0 \text{ a.e. in } s, \text{ almost surely in } \omega.$$

We have obtained that for every subsequence $Y_{n'}(\cdot)$ of the process $Y_n(\cdot)$ there exists a subsubsequence $Y_{n''}(\cdot)$ such that $Y_{n''}(\cdot)$ converges weakly to zero. Then the whole sequence converges weakly to zero, which proves the theorem.

§4.3. Weak Convergence of the Cramér-von Mises type functionals
of the process $W_n(\cdot)$. In the preceding section we proved the convergence of certain linear functionals of $W_n(\cdot)$. Now we are interested in the weak convergence of L^2 -norms of $W_n(\cdot)$. We do not go about it in the usual way, that is, we do not first prove weak convergence of the process $Z_n(\cdot)$ and hence that of

$$W_n(\cdot) = \alpha_n(F_{(k)}(\cdot)) + Z_n(\cdot).$$

Instead, we shall first prove weak convergence of the continuous process defined by the Fourier Transform of $Z_n(\cdot)$, i.e.,

$$\hat{Z}_n(\cdot) = F(Z_n)(\cdot)$$

and then we deduce the convergence in distribution of the $L^2[-M, M]$ -norm of $\hat{Z}_n(\cdot)$, for all $M > 0$. Then we obtain the convergence in distribution of the $L^2[-M, M]$ -norm of $Z_n(\cdot)$ and hence that of $W_n(\cdot)$.

We know that $Z_n(\cdot)$ is an absolutely integrable function for every outcome of our probability space. So the Fourier Transform of $Z_n(\cdot)$ is well defined as

$$(4.3.1) \quad \hat{Z}_n(t) := (2\pi)^{-1/2} \int_0^{+\infty} Z_n(s) e^{-its} ds$$

for every $-\infty < t < +\infty$, and $\hat{Z}_n(\cdot)$ is a continuous function in the real line. Hence $\hat{Z}_n(\cdot)$ is a continuous process.

THEOREM 4.3.1. Let $\hat{Z}_n(\cdot)$ be the continuous process defined above. Then

$$(4.3.2) \quad \hat{Z}_n(\cdot) - T_n \hat{g}(\cdot) \xrightarrow{\mathcal{D}[0, +\infty)} 0$$

where T_n is the random variable defined in (4.2.13) and

$$\hat{g}(t) := (2\pi)^{-1/2} \int_0^{+\infty} sf_{(k)}(s) e^{-its} ds$$

that is, the process $F(Z_n - T_n g)(\cdot)$ converges weakly in $\mathcal{D}[0, +\infty)$ to 0.

Proof. For $-\infty < t < +\infty$ let us define

$$h_t(s) := (2\pi)^{-1/2} e^{-its}, \quad -\infty < s < +\infty,$$

and apply Theorem 4.2.3 to $h_t(\cdot)$. We obtain

$$\begin{aligned} (4.3.3) \quad \hat{Z}_n(t) &= \int_0^{+\infty} Z_n(s) h_t(s) ds \\ &= T_n \int_0^{+\infty} s f_{(k)}(s) h_t(s) ds + o_t(1) \quad \text{a.s.}, \end{aligned}$$

where the constant in $o_t(\cdot)$ depends on t , and $\lim_{n \rightarrow \infty} o_t(1) = 0$ a.s..

Hence the finite dimensional distributions of the process

$$\{\hat{Z}_n - T_n \hat{g}\}(\cdot)$$

converge to zero, and also the sequence of random variables defined by

$$\hat{Z}_n(0) - T_n \hat{g}(0)$$

is tight. If we show now tightness of $\hat{Z}_n(\cdot)$, the theorem is proved.

For this purpose we apply a one-term Taylor expansion to the real and imaginary parts of the difference

$$\begin{aligned} e^{-itu} - e^{-isu} &= (\cos tu - \cos su) - i(\sin tu - \sin su) \\ &= (t-s)(-u \sin(\theta_u u)) - i(t-s)(u \cos(\rho_u u)) \\ &= -(t-s)u[\sin(\theta_u u) - i \cos(\rho_u u)] \end{aligned}$$

where $0 < \theta_u < 1$ and $0 < \rho_u < 1$ and both numbers depend on u .

Hence

$$\begin{aligned}
 (4.3.4) \quad (2\pi)^{\frac{1}{2}} |\hat{Z}_n(t) - \hat{Z}_n(s)| &= |n^{-\frac{1}{2}} \sum_j \{-n^{\frac{1}{2}} \int_{Y_j}^{\tilde{Y}_j} (e^{-itu} - e^{-isu}) du\}| \\
 &\leq n^{-\frac{1}{2}} \sum_j |t-s| \cdot |n^{\frac{1}{2}} \int_{Y_j}^{\tilde{Y}_j} u [\sin(\theta_u u) - i \cos(\rho_u u)] du| \\
 &\leq |t-s| \cdot n^{-1} \sum_j |n^{\frac{1}{2}} \int_{Y_j}^{\tilde{Y}_j} 2u du| \\
 &= |t-s| \cdot n^{-1} \sum_j n^{\frac{1}{2}} |\tilde{Y}_j - Y_j| \cdot |\tilde{Y}_j + Y_j|.
 \end{aligned}$$

Now we recall that

$$\tilde{Y}_j = Y_j + (\tilde{Y}_j - Y_j) \leq Y_j + r_n,$$

where $r_n = O[(n^{-1} \log n \cdot \log \log n)^{\frac{1}{2}}]$ a.s., by Lemma 4.2.1. Therefore

$$\tilde{Y}_j + Y_j \leq 2Y_j + r_n$$

and since $n^{\frac{1}{2}} |\tilde{Y}_j - Y_j| = O_p[(\log n)^{\frac{1}{2}}]$, by Lemma 4.2.1, the expression in (4.3.4) is smaller or equal to

$$\begin{aligned}
 &|t-s| \left(\frac{1}{n} \sum_{j=1}^n |n^{\frac{1}{2}} 2Y_j (\tilde{Y}_j - Y_j)| + O_p(n^{-\frac{1}{2}} \log n \cdot \log \log n) \right) \\
 &\leq |t-s| \left(\frac{1}{n} \sum_{j=1}^n \|2Y_j (-2X_j, -X_j \cdot X_j^T)\| \cdot \|H_n\| + O_p(n^{-\frac{1}{2}} \log n \cdot \log \log n) \right),
 \end{aligned}$$

where H_n is the vector defined in (4.2.6). We know that $H_n = O_p(1)$ as $n \rightarrow \infty$, and by the strong law of large numbers

$$(4.3.5) \quad n^{-1} \sum_{j=1}^n \|2Y_j (-2X_j, -X_j \cdot X_j^T)\| = O(1) \quad \text{a.s.}$$

Hence

$$(2\pi)^{1/2} |\hat{Z}_n(t) - \hat{Z}_n(s)| \leq |t-s| O_p(1) \text{ as } n \rightarrow \infty.$$

Given $\varepsilon > 0$ and $\eta > 0$ there exists $\delta > 0$ small enough and $n_0 = n_0(\varepsilon, \eta)$ such that for all $n \geq n_0$,

$$\begin{aligned} (4.3.6) \quad & P\left(\sup_{|s-t| < \delta} (2\pi)^{1/2} |\hat{Z}_n(t) - \hat{Z}_n(s)| > \varepsilon\right) \\ & \leq P\left(\sup_{|t-s| < \delta} |t-s| O_p(1) > \varepsilon\right) \\ & \leq P(O_p(1) > \varepsilon/\delta) < \eta. \end{aligned}$$

Statement (4.3.6) together with (4.3.3) and tightness of the process at $t = 0$ imply weak convergence of the continuous process $\hat{Z}_n(\cdot) - T_n \hat{g}(\cdot)$ in $\mathcal{D}[-M, M]$ for all $M > 0$.

COROLLARY 4.3.1. Let $\|Z_n - T_n g\|_2^M$ denote the L^2 -norm of the random function $(Z_n - T_n g)(\cdot)$ in $-M < t < M$, i.e.

$$\|Z_n(\cdot) - T_n g(\cdot)\|_2^M = \int_{-M}^M |Z_n(s) - T_n s f_{(k)}(s)|^2 ds,$$

where T_n is the random variable with mean zero and variance $2/k$ defined in (4.2.13). Then

$$(4.3.7) \quad \|Z_n - T_n g\|_2^M = o_p(1) \text{ as } n \rightarrow \infty.$$

Proof. Since $Z_n(\cdot) \in L^2(-\infty, +\infty)$, by Theorem 4.2.1, then $\hat{Z}_n(\cdot) \in L^2(-\infty, +\infty)$. By Theorem 4.3.1 all continuous functionals of the process $\hat{Z}_n(\cdot) - T_n \hat{g}(\cdot)$ converge in distribution to zero. In particular, for every $M > 0$,

$$(4.3.8) \quad \int_{-M}^M |\hat{Z}_n(t) - T_n \hat{g}(t)|^2 dt = o_p(1).$$

Moreover, applying (4.2.16) with $h_t(\cdot)$ as in Theorem 4.3.1, we have, for all $M > 0$,

$$\int_{-M}^M |\hat{Z}_n(t) - \frac{1}{n} \sum_j \frac{1}{(2\pi)^{1/2}} e^{-iy_j t} (2x_j, x_j, x_j^T)^T \cdot H_n|^2 dt = o(1) \text{ a.s.}$$

By standard Fourier Techniques we have for all $M > 0$,

$$\int_{-M}^M |Z_n - T_n g|^2 dt = o_p(1)$$

which is the statement of the theorem.

Consider the auxiliary process

$$(4.3.10) \quad \tilde{W}_n(s) := \alpha_n(F_{(k)}(s)) + T_n s f_{(k)}(s) \quad 0 \leq s < +\infty,$$

where as before,

$$\alpha_n(F_{(k)}(s)) = n^{1/2} (F_n(s) - F_{(k)}(s)) = n^{1/2} \sum_{j=1}^n \{I_s(Y_j) - F_{(k)}(s)\}$$

and T_n is the random variable defined in (4.2.13). We prove the following

THEOREM 4.3.2. Let $G(\cdot, \cdot)$ be a two-parameter Gaussian Process defined by

$$(4.3.11) \quad G(s, n) := n^{-1/2} K(F_{(k)}(s), n) + s f_{(k)}(s) \int_0^{+\infty} (t/k - 1) d_t \{n^{-1/2} K(F_{(k)}(t), n)\},$$

where $K(\cdot, \cdot)$ a Kiefer Process. $G(\cdot, \cdot)$ has mean and covariance functions

$$(4.3.12) \quad \begin{cases} E[G(s,n)] = 0 \\ E[G(s,n) \cdot G(t,m)] = (n \wedge m) (nm)^{-1/2} \{ F_{(k)}(s \wedge t) - F_{(k)}(t) \\ - \frac{2}{k} t f_{(k)}(t) s f_{(k)}(s) \} , \end{cases}$$

respectively. Then

$$(4.3.13) \quad \sup_{0 \leq s < +\infty} |\tilde{W}_n(s) - G(s,n)| = O(n^{-\varepsilon}) \quad \text{a.s.}$$

for some $\varepsilon > 0$. Also for each n fixed, and $g(s) := sf_{(k)}(s)$,

$$(4.3.14) \quad G(\cdot, n) \stackrel{D}{=} W(\cdot) := B(F_{(k)}(\cdot)) + g(\cdot) \int_0^{+\infty} (t/k - 1) d_t B(F_{(k)}(t))$$

where $\stackrel{D}{=}$ stands for equality of all finite-dimensional distributions and $B(\cdot)$ is a Brownian Bridge.

Proof. Let us consider again the random sample $Y_1, \dots, Y_n$ with distribution function $F_{(k)}(\cdot)$. The sequence $\hat{\mu}_n := n^{-1} \sum_{j=1}^n Y_j/k$ is a sequence of estimators of the parameter $\mu := E(Y)/k = 1$ where Y is a chi-squared with k -degrees of freedom rv. Therefore

$$n^{1/2}(\hat{\mu}_n - 1) = n^{-1/2} \sum_{j=1}^n \{Y_j/k - 1\}$$

has the form (i) of Lemma 1.1.9 with $\varepsilon_{1n} = 0$. Moreover the distribution function $F_{(k)}(\cdot) = F(\cdot, k)$ and the sequence $\hat{\mu}_n$ satisfy all the conditions of Lemma 1.1.9. Indeed, we observe first that the function

$$l(s, k) := s/k - 1$$

is a measurable real valued function such that $E[l(Y_j, k)] = 0 \quad \forall j = 1, 2, \dots, n$, and that $E[l(Y_j, k)]^2 = E(Y_j/k - 1)^2 = 2/k \quad \forall j = 1, 2, \dots, n$.

Now let us write $Q(\cdot) := \text{inv } F_{(k)}(\cdot)$. The density-quantile function $f_{(k)}Q(\cdot)$ behaves in the tail near 1 like

$$(4.3.15) \quad f_{(k)}Q(y) \sim c(1-y) \quad \text{as } y \rightarrow 1, c > 0,$$

(cf. Parzen, 1979) and since $Q(0) = 0$,

$$\lim_{y \downarrow 0} (y \log \log 1/y)^{1/2} |Q(y)/k - 1| = 0$$

and from (4.3.15),

$$\lim_{y \uparrow 1} ((1-y) \log \log 1/(1-y))^{1/2} |Q(y)/k - 1| = 0$$

which are conditions (v) of Lemma 1.1.9. We also have

$$\lim_{y \downarrow 0} y \frac{d}{dy} Q(y) = \lim_{y \downarrow 0} \frac{y}{fQ(y)} = 0$$

and

$$\lim_{y \uparrow 1} (1-y) \frac{d}{dy} Q(y) = \lim_{y \uparrow 1} \frac{(1-y)}{fQ(y)} = c > 0$$

which are conditions (vi) of the Lemma. All the conditions of Lemma 1.1.9 are verified, hence there exists a Kiefer Process $K(\cdot, \cdot)$

such that for every positive s ,

$$\begin{aligned} (4.3.16) \quad & |\tilde{W}_n(s) - n^{-1/2} K(F_{(k)}(s), n) - s f_{(k)}(s) \int_0^{+\infty} \ell(t, k) d_t (n^{-1/2} K(F_{(k)}(t), n))| \\ & \leq |\alpha_n(F_{(k)}(s)) - n^{-1/2} K(F_{(k)}(s), n)| + \sup_{0 \leq s < +\infty} s f_{(k)}(s) |n^{-1/2} \sum_{j=1}^n (Y_j/k - 1) \\ & \quad - \int_0^{+\infty} \ell(t, k) d_t (n^{-1/2} K(t, n))| \\ & = O(n^{-1/2} (\log n)^2) + \|g(\cdot)\|_{\infty} O(n^{-\epsilon}) = O(n^{-\epsilon}) \quad \text{a.s.} \end{aligned}$$

for some $\varepsilon > 0$, where the constant in the $O(\cdot)$ does not depend on s . Then

$$(4.3.17) \quad \sup_{0 \leq s < +\infty} |\tilde{W}_n(s) - G(s, n)| = O(n^{-\varepsilon}) \quad \text{a.s..}$$

The covariance function can be calculated directly from the expression of $G(\cdot, \cdot)$, where by the usual covariance formula for the Kiefer Process, we have

$$(4.3.18) \quad E(n^{-1/k} K(F_{(k)}(s), n) \cdot m^{-1/k} K(F_{(k)}(t), m)) = (nm)^{-1/k} (n \wedge m) \{F_{(k)}(s \wedge t) - F_{(k)}(s) F_{(k)}(t)\},$$

$$(4.3.19) \quad E(n^{-1/k} K(F_{(k)}(s), n) \cdot \int_0^{+\infty} (u/k-1) d_n \{m^{-1/k} K(F_{(k)}(u), m)\}) \\ = (nm)^{-1/k} (n \wedge m) \int_0^s (u/k-1) dF_{(k)}(u) \\ = -(2/k) (nm)^{-1/k} (n \wedge m) s f_{(k)}(s),$$

and finally,

$$(4.3.20) \quad E(\int_0^{+\infty} (u/k-1) d_n \{n^{-1/k} K(F_{(k)}(u), n)\} \cdot \int_0^{+\infty} (u/k-1) d_n \{m^{-1/k} K(F_{(k)}(u), m)\}) \\ = (nm)^{-1/k} (n \wedge m) (2/k).$$

Combining the last four lines we obtain the covariance formula specified in (4.3.12).

In the sequel we obtain the weak convergence of a Cramér-von Mises statistic of $W_n(\cdot)$. Let us call, for $k \geq 2$, and $M > 0$,

$$(4.3.21) \quad L_n = L_n^M(k) := \int_0^M W_n^2(s) ds$$

and

$$(4.3.22) \quad L = L^M(k) := \int_0^M W^2(s) ds .$$

THEOREM 4.3.3. The process $W_n(\cdot)$ defined in (4.1.4) is a function of $L^2[0, M]$, for every $n=1,2,\dots$; $W(\cdot)$ defined in (4.3.14) is also in $L^2[0, M]$ and the L^2 -norm $L_n^{1/2}$, of $W_n(\cdot)$, converges in distribution to the L^2 -norm $L^{1/2}$, of $W(\cdot)$, that is,

$$(4.3.23) \quad \left\{ \int_0^M W_n^2(s) ds \right\}^{1/2} \xrightarrow{D} \left\{ \int_0^M W^2(s) ds \right\}^{1/2} .$$

REMARK 4.3.1. The random variable $L = \int_0^M W^2(s) ds$ can be expressed as the infinite sum of weighted independent chi-squared random variables with one degree of freedom. This is a consequence of the Karhunen-Loève expansion of a Gaussian Process and the fact that $\int_0^{+\infty} \{E(W(s)W(t))\}^2 ds dt < +\infty$. For details we refer to M. Csörgő and Révész, 1981b, p.249.

Proof of Theorem 4.3.3.: We have already seen that if we define $Z_n(\cdot)$ as zero on the negative real line, then $Z_n \in L^2(-\infty, +\infty)$ a.s.. Now consider $\alpha_n(F_{(k)}(\cdot))$ defined as zero for $s < 0$. We observe that

$$\alpha_n(F_{(k)}(s)) = n^{1/2}(1 - F_{(k)}(s)) \quad \text{for } s \geq Y_{n:n} .$$

Hence

$$\int_{Y_{n:n}}^{+\infty} \alpha_n^2(F_{(k)}(s)) ds = n \int_{F_{(k)}(Y_{n:n})}^1 (1-y)^2 dQ(y) .$$

where $Q(\cdot) = F_{(k)}^{-1}(\cdot)$. Then the integral above converges a.s. because of the tail conditions of $Q(\cdot)$ (conditions (v) and (vi). of Lemma 1.1.9). Similarly, M. Csörgő and Révész (in Lemma 5.5.5, 1981b) prove

$$\int_0^{+\infty} B^2(F_{(k)}(s)) ds = \int_0^1 B^2(y) dQ(y) < +\infty \quad \text{a.s.}$$

where $Q(\cdot)$ is a quantile function satisfying the conditions (v) of Lemma 1.1.9. Then for every $M > 0$,

$$W_n(\cdot) = \alpha_n(F_{(k)}(\cdot)) + Z_n(\cdot) \in L^2[0, M],$$

$$\tilde{W}_n(\cdot) = \alpha_n(F_{(k)}(\cdot)) + T_n g(\cdot) \in L^2[0, M],$$

and

$$W(\cdot) = B(F_{(k)}(\cdot)) + g(\cdot) \int_0^{+\infty} (u/k-1) dB(F_{(k)}(u)) \in L^2[0, M] \quad \text{a.s.}$$

We have

$$\|W_n - \tilde{W}_n\|_2^M = \|Z_n - T_n g\|_2^M = o_p(1) \quad \text{as } n \rightarrow \infty,$$

by Corollary 4.3.1. Also

$$(4.3.24) \quad \|W_n\|_2^M = \|(W_n - \tilde{W}_n) + \tilde{W}_n\|_2^M \\ = \{(\|W_n - \tilde{W}_n\|_2^M)^2 + (\|\tilde{W}_n\|_2^M)^2 + 2\langle W_n - \tilde{W}_n, \tilde{W}_n \rangle_2\}^{1/2}$$

where $\langle \cdot, \cdot \rangle_2$ denotes the inner product in $L^2[0, M]$. By Schwartz' inequality,

$$|\langle \tilde{W}_n, W_n - \tilde{W}_n \rangle_2| \leq \|\tilde{W}_n\|_2^M \cdot \|W_n - \tilde{W}_n\|_2^M = o_p(1) \cdot o_p(1) = o_p(1).$$

Hence

$$(4.3.25) \quad \|w_n\|_2^M = \|\tilde{w}_n\|_2^M + o_p(1).$$

Then by Theorem 4.3.2,

$$\|w_n\|_2^M \xrightarrow{D} \|w\|_2^M,$$

which proves the theorem.

Now consider the Cramér-von Mises type statistic proposed by Koziol (1982). For $k \geq 2$, and for $M > 0$ let

$$J_n = J_n^M(k) := \int_0^M w_n^2(s) dF_{(k)}(s).$$

We want to prove its weak convergence to the random variable defined by

$$J = J^M(k) := \int_0^M w^2(s) dF_{(k)}(s).$$

We have

THEOREM 4.3.4. Let $w_n(\cdot)$ be the process defined in (4.1.4).

Then with $k \geq 2$, $M > 0$,

$$(4.3.26) \quad \int_0^M w_n^2(s) dF_{(k)}(s) \xrightarrow{D} \int_0^M w^2(s) dF_{(k)}(s),$$

where

$$w(s) = B(F_{(k)}(s)) + sf_{(k)}(s) \int_0^{+\infty} (u/k-1) dB(F_{(k)}(s)), \quad 0 \leq s < +\infty,$$

with $B(\cdot)$ a Brownian Bridge and covariance function given by (4.3.12).

Here again the statistic $J_n^M(k)$ can be expressed as an infinite sum of weighted independent chi-squared with one degree of freedom rv.

Proof. Consider the auxiliary process $\tilde{W}_n(\cdot)$ defined in (4.3.10). We have, as in the proof of the previous theorem,

$$\int_0^M |W_n(s) - \tilde{W}_n(s)|^2 dF_{(k)}(s) \leq \sup_s |f_{(k)}(s)| (\|W_n - \tilde{W}_n\|_2^M)^2 = o_p(1)$$

Hence, by an argument similar to the one used to prove (4.3.25), we have

$$(4.3.27) \quad \int_0^M W_n^2(s) dF_{(k)}(s) = o_p(1) + \int_0^M \tilde{W}_n^2(s) dF_{(k)}(s)$$

Then by Theorem 4.3.2,

$$J_n^M(k) = \int_0^M W_n^2(s) dF_{(k)}(s) \xrightarrow{p} \int_0^M \tilde{W}_n^2(s) dF_{(k)}(s),$$

which is the statement of the theorem.

§4.4. On Assessing Multivariate Normality. Let $N_{(k)}(\mu, \Sigma)$ denote the k -dimensional normal distribution with mean μ and covariance matrix Σ , where $k \geq 2$. We are interested in testing the composite hypothesis

H_0 : The law of X is $N_{(k)}(\mu, \Sigma)$ with some $\mu \in \mathbb{R}^{(k)}$ and some positive definite Σ , ($|\Sigma| \neq 0$).

In the preceding section we defined two Gramér-von Mises statistics $L_n^M(k)$ and $J_n^M(k)$ for $k \geq 2$, and proved that under the null hypothesis

H_0 they converge in distribution to

$$L^M(k) = \int_0^M w^2(s) ds$$

and

$$J^M(k) = \int_0^M w^2(s) dF_{(k)}(s)$$

respectively, where $W(\cdot)$ is the process defined by (4.3.14). Now

let H_1 be the alternative hypothesis

H_1 : Let (μ, Σ) be the mean and covariance matrix of the random vector X

for some $\mu \in R^{(k)}$ and positive definite matrix Σ . The law of

$Y = (X - \mu)^T \Sigma^{-1} (X - \mu)$ is given by some distribution function $G_{(k)}(\cdot)$

such that $G_{(k)}(\cdot) \neq F_{(k)}(\cdot)$ for a set of positive Lebesgue measure

on $0 \leq s < \infty$, and such that if $Y_{n:n}$ denotes the largest order

statistic of a Y -sample of size n then $Y_{n:n} = o_p(n^{1-\epsilon})$ for any

$0 < \epsilon < 1/2$.

We have, for all $M > 0$,

THEOREM 4.4.1. Under H_1 , $L_n^M \xrightarrow{P} \infty$ and $J_n^M \xrightarrow{P} \infty$. Hence, when testing H_0 against H_1 with either L_n^M or J_n^M , the test is consistent.

Proof. Given a random vector sample $X_1, \dots, X_n$ from any multivariate law with mean μ and covariance Σ , we calculate

$$\tilde{Y}_j = \psi(X_j, \hat{\theta}_n).$$

where

$$\hat{\theta}_n = (n^{-1} \sum_j X_j, n^{-1} \sum_j (X_j - \bar{X}_n)(X_j - \bar{X}_n)^T).$$

Let $Y_{n:n}$ be the largest order statistic of the Y -sample and assume the conditions of the alternative hypothesis H_1 . If $W_n \in L^2[0, M]$ on a set of positive probability, consider $Z_n(\cdot)$ defined as in (4.1.11) hence $Z_n \in L^2(0, +\infty)$ and

$$(4.4.1) \quad \frac{1}{n} \int_0^{+\infty} Z_n^2(s) ds = o_p(1).$$

Indeed, following the line of proof of Lemma 4.2.1, we have

$$n^{1/2}(\tilde{Y}_j - Y_j) = (-2X_j, X_j \cdot X_j^T)^T \cdot H_n + O(n^{-1/2} \log \log n \cdot Y_{n:n}) \text{ a.s.}$$

where H_n is as in (4.2.6). Hence $H_n = O_p(1)$ as $n \rightarrow \infty$ and by hypothesis,

$$\max_{1 \leq j \leq n} |\tilde{Y}_j - Y_j| \leq n^{-1/2} Y_{n:n} \cdot \|H_n\| + o_p(1).$$

Here $\|\cdot\|$ denotes the euclidean norm in $R^{(k+k^2)}$. Thus

$$\max_j |\tilde{Y}_j - Y_j| = o_p(1).$$

Therefore

$$\begin{aligned} \frac{1}{n} \int_0^{+\infty} Z_n^2(s) ds &= \frac{1}{n} \left\{ \frac{1}{n} \sum_{j=1}^n n^{1/2} \int_{Y_j}^{\tilde{Y}_j} Z_n(s) ds \right\} \\ &\leq \frac{1}{n} \sum_{j=1}^n \left[\sup_{1 \leq j \leq n} |Z_n(s)| / n^{1/2} \right] \cdot |\tilde{Y}_j - Y_j| \\ &= n^{-1/2} \sup_s |Z_n(s)| \cdot o_p(1). \end{aligned}$$

But

$$|Z_n(s)| \leq \frac{1}{n^{1/2}} \sum_{j=1}^n |R_j(s)| \leq n^{1/2}$$

since $R_j(s)$ takes only 0, -1 or 1 values, then

$$n^{-1/2} |Z_n(s)| \leq 1 \quad \forall s,$$

which proves (4.4.1). Now let $F_n(\cdot)$ be the empirical distribution function of the Y-sample with distribution function $G_{(k)}$. Then by Glivenko-Cantelli,

$$(4.4.2) \quad \lim_{n \rightarrow \infty} \sup_s |F_n(s) - F_{(k)}(s)| = \sup_s |G_{(k)}(s) - F_{(k)}(s)| > 0 \quad \text{a.s.}$$

in at least a set of positive Lebesgue measure in $0 \leq s < +\infty$.

Now we assumed $W_n \in L^2[0, M]$, hence $W_n - Z_n \in L^2[0, M]$ and

$$\begin{aligned} (4.4.3) \quad \|W_n\|_2^M &\geq \|W_n - Z_n\|_2^M - \|Z_n\|_2^M \\ &= \left\{ \left(n \int_0^M (F_n(s) - F_{(k)}(s))^2 ds \right)^{1/2} - \left(\int_0^M Z_n^2(s) ds \right)^{1/2} \right\} \\ &= n^{1/2} \left\{ \int_0^M (F_n(s) - F_{(k)}(s))^2 ds - \left(\frac{1}{n} \int_0^M Z_n^2(s) ds \right)^{1/2} \right\} \\ &= n^{1/2} \left\{ \int_0^M (F_n(s) - F_{(k)}(s))^2 ds + o_p(1) \right\}. \end{aligned}$$

Combining (4.4.1), (4.4.2) and (4.4.3) we obtain

$$L_n^{1/2} = \|W_n\|_2^M \xrightarrow{P} \infty.$$

Similarly, we can prove that

$$J_n^{\frac{1}{2}P} \rightarrow +\infty,$$

and the theorem follows.

The statistic M_n defined in (4.2.21) provides us with another test. Corollary (4.2.1) states that under the null hypothesis H_0 ($k \neq 3$) we have

$$M_n = M_n(k) = \int_0^{+\infty} W_n(s) dF_{(k)}(s) \xrightarrow{D} N(0, \sigma_k^2)$$

where σ_k^2 is as in (4.2.22). The advantage of an M_n -test is that we do not need new tables. For this test, we reject H_0 if $|M_n|$ is too large and we prove that the test is consistent against a certain type of alternatives. We consider

H_2 : The law of $Y = (X-\mu)^T \Sigma^{-1} (X-\mu)$ is given by any $G_{(k)}(\cdot)$ such that

$$(i) \int_0^{+\infty} [G_{(k)}(s) - F_{(k)}(s)] dF_{(k)}(s) \neq 0;$$

$$(ii) Y_{n:n} = o_p(n^{1-\epsilon}) \text{ for any } 0 < \epsilon < 1/2,$$

for some $(\mu, \Sigma) = (EX, \text{cov}(X, X))$, $|\Sigma| \neq 0$.

We have

THEOREM 4.4.2. Under H_2 , $|M_n| \xrightarrow{P} +\infty$.

Proof. By the Glivenko-Cantelli theorem we have, as in Theorem

4.4.1,

$$\lim_{n \rightarrow \infty} \int_0^{+\infty} (F_n(s) - F_{(k)}(s)) dF_{(k)}(s) = \int_0^{+\infty} (G_{(k)}(s) - F_{(k)}(s)) dF_{(k)}(s) \quad \text{a.s.}$$

where $F_n(\cdot)$ is the empirical distribution function of the Y-sample with distribution function $G_{(k)}(\cdot)$. We also obtain, given condition (ii) of H_2 ,

$$\frac{1}{n^{1/2}} \int_0^{+\infty} Z_n(s) dF_{(k)}(s) = \left\{ \frac{1}{n} \sum_{j=1}^n f_{(k)}(Y_j) (2X_j, X_j \cdot X_j^T)^T \right\} \cdot O\left(\frac{\log \log n}{n}\right)^{1/2} + o_p(1)$$

a.s.

$$= o_p(1)$$

since the factor within $\{\cdot\}$ is bounded a.s. by the law of large numbers. Hence

$$|M_n| \geq n^{1/2} \left\{ \left| \int_0^{+\infty} (F_n(s) - F_{(k)}(s)) dF_{(k)}(s) \right| - o_p(1) \right\} \xrightarrow{P} +\infty.$$

This proves the theorem.

§4.5. An Example. Practical implementation of the statistic M_n for assessing multivariate normality is simple. Given $X_1, \dots, X_n$, random k -dimensional vectors, calculate first $\bar{X}_n$ and S_n and then the $\tilde{Y}_j = \psi(X_j, \theta_n)$:

$$\tilde{Y}_j = (X_j - \bar{X}_n)^T S_n^{-1} (X_j - \bar{X}_n).$$

Then compute $Z_j = F_{(k)}(\tilde{Y}_j)$ and order the values Z_j as $Z_{1:n} < \dots < Z_{n:n}$. Hence

$$M_n = \int_0^{+\infty} W_n(s) dF_{(k)}(s) = n^{1/2} \int_0^1 (\tilde{E}_n(y) - y) dy,$$

where

$$\tilde{E}_n(y) = \begin{cases} 0 & \text{if } y < z_{1:n} \\ j/n & \text{if } z_{j:n} \leq y < z_{j+1:n} \\ 1 & \text{if } z_{n:n} > y. \end{cases}$$

We write $z_{0:n} = 0$ and $z_{n+1:n} = 1$. Then we have

$$\begin{aligned} M_n &= n^{-1} \sum_{j=0}^n \int_{z_{j:n}}^{z_{j+1:n}} (j/n - y) dy \\ &= n^{-1} \left\{ \sum_{j=1}^n (j/n) (z_{j+1:n} - z_{j:n}) - 1/2 \right\}. \end{aligned}$$

As an illustration of this procedure, consider the Iris Setosa data of 50 quadrivariate observations consisting of the variables sepal length, sepal width, petal length and petal width. This data was collected originally by Anderson (1935) and subsequently analyzed by Fisher (1936). It is commonly believed that, in some form or another, this data is from a quadrivariate normal distribution.

The test proposed by Rincón-Gallardo et al. (1979) is based on transformations of the original sample into a sample of uniform $U(0,1)$ rv.'s. They studied the original Iris Setosa data at the upper 10% significance level and they did not reject the hypothesis of quadrivariate normality.

Kozioł (1982, 1982b, 1983) considered the Cramér-von Mises test statistic $J_n(4)$ (see (4.1.6)) and analyzed the logarithms of the original "Iris Setosa" observations. He did not reject either the

the null hypothesis, at the upper 10% significance level. We stress again that we are going to obtain our own tables for $J(k)$, $k \geq 2$, and compare our critical values with Koziol (1982)'s.

S. Csörgő (1984) studied a statistic $M_n^{(4)}$ and obtained an upper bound for the asymptotic distribution of $M_n^{(4)}$. With this test, he rejected the hypothesis of quadrivariate normality of both the original and the logarithmic data at the upper 1% level. Since the asymptotic distribution function of $M_n^{(4)}$ is difficult to calculate in closed form S. Csörgő provides a table of upper bounds $Z_4(\alpha)$ of the asymptotic $(1-\alpha)$ percentage points of $M_n^{(4)}$ and the significance level may be too large for the corresponding $Z_4(\alpha)$ -value (here $P(M_n^{(4)} \geq Z_4(\alpha)) \leq \alpha$), and therefore this test may be highly sensitive.

In our case, the statistic $M_n^{(4)}/\sigma_4$ has as asymptotic distribution the standard normal distribution function $\Phi(\cdot)$. Upon reference to tables for $\Phi(\cdot)$ we see that for $\alpha = 0.05$ and $\frac{\alpha}{2} = 0(N \geq Z_{\alpha/2})$, $Z_{\alpha/2} = 1.96$, where $N = N(0,1)$, and for $\alpha = 0.10$, $Z_{\alpha/2} = 1.645$. The original data yields a value of $N = 1.7920$ and the logarithmic data yields $N = 1.6741$. Hence at the $\alpha = 0.05$ significance level we see no reason to question the null hypothesis for either the original or the transformed data. However, at the upper 10% significance level, the raw data yields an N value that is too large and even the transformed (logarithmic) data, with $N = 1.6741$, might suggest a departure from an underlying quadrivariate normal structure. We suspect that the Rincón-Gallardo test is not very powerful.

References

- AITKIN, M.A. (1972). A class of tests for multivariate normality based on linear functions of order statistics. Unpublished manuscript.
- AKHIEZER and GLAZMAN (1978). Theory of Linear Operators in Hilbert Space. Vol. 1. Frederick Ungar Publishing Co. New York, third printing.
- ALY, E.E. (1983). Strong Approximations of the Q-Q processes. Preprint.
- ALY, E.-E.A.A. (1983b). A note on the asymptotic theory of Lilliefors' empirical process. Technical Report Series of the Laboratory for Research in Statistics and Probability, Carleton University and the University of Ottawa, No.3, July (1983).
- ALY, E.E. (1983c). Some limit theorems for uniform and exponential spacings. Canad. J. Statist., 11, 211-219.
- ALY, E.E., BLEUER, S. and CSÖRGÖ, M. (1984). On the asymptotic theory of the Kumar-Pathak test for Normality. In Three Papers on Composite Goodness of Fit. Technical Report Series of the Laboratory for Research in Statistics and Probability. No.22, 0-15.
- ALY, E.E. and CSÖRGÖ, M. (1985). Quadratic Nuisance-parameter-free goodness-of-fit tests in the presence of location and scale parameters. The Canadian J. of Statist., Vol. 13, No. 1.
- ANDERSON, E. (1935). The irises of the Gaspé Peninsula. Bull. Amer. Iris Soc. 59, 2-5.

- ANDREWS, D.F., GNANADESIKAN, R. and WARNER, J.L. (1981). Transformations of multivariate data. Biometrics 27, 825-840.
- ANDREWS, D.F., GNANADESIKAN, R. and WARNER, J.L. (1972). Methods for assessing multivariate normality. Bell Laboratories Memorandum.
- ANDREWS, D.F., GNANADESIKAN, R. and WARNER, J.L. (1973). Methods for assessing multivariate normality. In Multivariate Analysis III (P.R. Krishnaiah, ed.), pp.95-116, Academic Press, New York.
- Billingsley (1968). Convergence of Probability Measures. J. Wiley, New York.
- BURKE, M.D., CSÖRGÖ, M., CSÖRGÖ, S., and RÉVÉSZ, P. (1979). Approximations of the Empirical Process when Parameters are Estimated, Annals of Probability, 9, 790-810.
- CHUNG, C. (1984). WINNER PACK (Package) to be published by the Geological Survey of Canada.
- COX, D.R. (1968). Notes on some aspects of regression analysis. J. Roy Statist. Soc. Ser A 131, 265-279.
- COX, D.R. and SMALL, N.J.H. (1978). Testing multivariate normality. Biometrika 65, 263-272.
- CSÖRGÖ, M. KOMLÓS, J., MAJOR, P., RÉVÉSZ, P. and TUSNÁDY, G. (1974). On the empirical process when parameters are estimated. Trans Seventh Prague Conf. 1974, Academia, Prague, 1977, 87-97.
- CSÖRGÖ, M. (1983). Quantile processes with statistical applications. (Regional Conference Series on Appl. Math.) S.I.A.M. Philadelphia.
- CSÖRGÖ, M. and RÉVÉSZ, P. (1975). Some notes on the empirical distribution function and the quantile process. In Coll. Math. Soc. J. Bolyai 11, Limit Theorems of Probability Theory. (P. Révész, Ed.), Keszthely, Hungary 1974, North Holland, Amsterdam-London, 59-71.

CSÖRGÖ, M., RÉVÉSZ, P. (1978). Strong approximations of the quantile process. Ann. Statist. 6, 882-894.

CSÖRGÖ, M. and RÉVÉSZ, P. (1980). Quantile Processes and Sums of Weighted Spacings for Composite Goodness-of-Fit I-II. Carleton Mathematical Lecture Note No.27-1980.

CSÖRGÖ, M., RÉVÉSZ, P. (1981a). Two approaches to constructing simultaneous confidence bounds for quantiles. Carleton Mathematical Series, No. 176.

CSÖRGÖ, M. and RÉVÉSZ, P. (1981b). Strong approximations in probability and statistics. Academic Press, New York.

CSÖRGÖ, M. and RÉVÉSZ, P. (1981c). Quantile Processes and sums of weighted spacings for composite goodness-of-fit. In Statistics and Related Topics, Proceedings of the International Symposium on Statistics and Related Topics held in Ottawa, Canada, May 5-7, 1980. North Holland Publishing Company, Amsterdam-New York-Oxford.

CSÖRGÖ, M., CSÖRGÖ, S., HORVÁTH, L., RÉVÉSZ, P. (1982). On weak and strong approximations of the quantile process. To appear in Proceedings of the Seventh Conference on Probability Theory (Brasov, Aug 29 - Sept 4, 1982).

DAHIYA, R.C. and GURLAND, J. (1973). A test of fit for bivariate distributions. J. Roy. Statist. Soc. Ser. B 35, 452-465.

DAY, N.R. (1969). Divisive cluster analysis and a test for multivariate normality. Bull. Inst. Internat. Statist. 43, 110-112.

DARLING, D.A. (1955). The Cramér-Smirnov test in the parametric case. Ann. Math. Statist. 26, 1-20.

- DEWET, T., VENTER, J.H. (1972). Asymptotic distributions of certain tests criteria of normality. S. Afr. Statist. J. 6, 135-149.
- DEWET, T., VENTER, J.H. (1973). Asymptotic distributions for quadratic forms with applications to test of fit. Ann. Statist. 1, 380-387.
- DOKSUM, K.A. (1974). Empirical probability plots and statistical inference for nonlinear models in the two sample case. Ann. Statist. 2, 267-277.
- DOKSUM, K.A., SIEVERS, G.L. (1976). Plotting with confidence: graphical comparisons of two populations. Biometrika 63, 421-434.
- DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. 1, 279-90.
- DURBIN, J. (1976). Kolmogorov-Smirnov tests when parameters are estimated. In Empirical Distributions and Processes (P. Gaenssler and P. Revesz, Ed.) Lecture Notes in Mathematics 566, 33-44, Springer-Verlag, Berlin.
- FISHER, R.A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179-188. (Also paper 32 in Contributions to Mathematical Statistics, R.A. Fisher, Wiley, New York).
- GIORGI, G.M. and FATTORINI, L. (1976). An empirical study of some tests for multivariate normality. Quaderni dell'Istituto di Statistica No.20 (8 pages). Siena.
- GNANADESIKAN, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. Wiley, New York.

- GNANADESIKAN, R. and KETTERING, J.R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. Biometrics 28, 81-124.
- HAWKINS, D.M. (1981). A new test for multivariate normality and homoscedasticity. Technometrics 23, 105-110.
- HEALY, M.J.R. (1968). Multivariate normal plotting. Appl. Statist. 17, 157-161.
- HENSLER, G.L., MEHROTA, K.G. and MICHALEK, J.E. (1977). A goodness of fit test for multivariate normality. Comm. Statist. Theor. Meth. A 6, 33-41.
- KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. Ann. Math. Statist. 26, 189-211.
- KESSEL, D.L. and FUKUNAGA, K. (1972). A test for multivariate normality with unspecified parameters. Unpublished report, Purdue Univ.
- KOMLÓS, J., MAJOR, P., TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and sample D.F. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111-131.
- KOZIOL, J.A. (1982). A class of invariant procedures for assessing multivariate normality. Biometrika 69, 423-427.
- KOZIOL, J.A. (1982b). A class of Invariant Procedures for Assessing Multivariate Normality. Technical Report, Department of Mathematics and Medicine, University of California at San Diego, La Jolla, California.

- KOZIOL, J.A. (1983). On assessing multivariate normality. J. Roy. Statist. Soc. Ser. B. 45, 358-361.
- KUMAR, A., PATHAK, P.K. (1977). Sufficiency and tests of goodness of fit. Scand. J. Statist. 4, 39-43.
- LA RICCIA, V.N. and MASON, D.M. (1985). Optimal goodness of fit tests for location/scale families of distributions based on the sum of squares of L-statistics. Ann. Statist. Vol. 13, No. 1, to appear.
- LILLIEFORS, H.W. (1967). "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, 62, 399-402.
- MALKOVICH, J.F. (1971). Tests for multivariate normality. Unpublished Ph.D. thesis, Univ. of California, Los Angeles.
- MALKOVICH, J.F. and AFIFI, A.A. (1973). On tests for multivariate normality. J. Amer. Statist. Assoc. 68, 176-179.
- MARDIA, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519-530.
- MARDIA, K.V. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. Sankhyā A 36, 115-128.
- MARDIA, K.V. (1975). Assessment of multinormality and the robustness of Hotelling's T^2 test. J. Roy. Statist. Soc. Ser. C. 24, 163-171.
- MARDIA, K.V. (1980). Tests of univariate and multivariate normality. In Handbook of Statistics 1. Analysis of Variance (P.R. Krishnaiah, Ed.), pp. 279-320. North-Holland, Amsterdam.

MARDIA, K.V., KENT, J.T. and BIBBY, J.M. (1979). Multivariate Analysis.
London: Academic Press.

MARDIA, K.V. and ZEMROCH, P.J. (1975). Algorithm AS84. Measures of
multivariate skewness and kurtosis. J. Roy. Statist. Soc. Ser. C
24, 262-265.

MOORE, D.S. and STUBBLEBINE, J.B. (1981). Chi-square tests for
multivariate normality with applications to common stock prices.
Commun. Statist., Theor. Meth. A 10, 713-738.

NEUHAUS, G. (1974). Asymptotic properties of the Cramér-von Mises
statistic when parameters are estimated. In Proc. Prague Symp.
Asymptotic Statist. 1973 (J. Hajek, Ed.) 2, 257-297. Charles
Univ. Prague.

NEUHAUS, G. (1976). Weak convergence under contiguous alternatives of
the empirical process when parameters are estimated: the D_k
approach. In Empirical Distributions and Processes, Lecture Notes
in Mathematics 566, Ed. P. Gaenssler and P. Révész, pp. 68-82.
Berlin: Springer-Verlag.

PARZEN, E. (1979). Nonparametric statistical data modeling. J. Amer.
Statist. Assoc. 74, 105-121.

PETTITT, A.N. (1979). Testing for bivariate normality using the
empirical distribution function. Commun. Statist. Theor. Math. A
8, 699-712.

RINCÓN-GALLARDO, S., QUESENBERY, C.P. and O'REILLY, F.J. (1979).
Conditional probability integral transformations and goodness-of-fit
tests for multivariate normal distributions. Ann. Statist. 7 1052-10

- SARKADI, K. and TUSNÁDY, G. (1977). Testing for normality and for the exponential distribution. In Proceedings Fifth Conference on Probability Theory 1974, Brasov (B. Bereanu, M. Josifescu, G. Popescu, eds.), pp. 99-118. Editura Academici, Bucharest.
- SHAPIRO, S.S. and FRANCIA, R.S. (1972). An approximate analysis of variance tests for normality. J. Amer. Statist. Assoc. 67 215-216.
- SHAPIRO, S.S. and WILK, M.B. (1965). An analysis of variance tests for normality (complete samples). Biometrika 52, 591-611.
- SHAPIRO, S.S., WILK, M.B., and CHEN, H.J. (1968). A comparative study of various tests for normality. J. Amer. Statist. Assoc. 63 1343-1372.
- SWITZER, P. (1976). Confidence procedures for two sample problems. Biometrika 63, 13-25.
- WAGLE, B. (1968). Multivariate beta distribution and a test for multivariate normality. J. Roy. Statist. Soc. Ser. B. 30 511-536.
- WEISS, L. (1958). A test of fit for multivariate distributions. Ann. Math. Statist. 29, 595-599.
- WELLNER, J.A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. Z. Wahrscheinlichkeitstheorie verw. Gebiete 45, 73-88.
- WILK, M.B. and GNANADESIKAN, R. (1968). Probability plotting methods for the analysis of data. Biometrika 55, 1-17.
- YANG, S.-S. (1981). Linear combinations of concomitants of order statistics with applications to testing and estimation. Ann. Inst. Statist. Math. 33, 463-470.

ADDITIONAL REFERENCES

- CSÖRGÖ, S. (1984). Testing by the Empirical characteristic Function,
· six papers by S. Csörgö and C.R. Heathcote, *Technical Report*
Series of the Laboratory for Research in Statistics and
Probability. No. 41, September 1984.
- MARDIA, K.V. Tests of univariate and multivariate normality. *Handbook*
of Statistics, North-Holland, 1980, 279-320, Vol. 1.
- MATTHEWS, J.N.S. Robust Methods in Assessment of Multivariate Normality,
Appl.-Statist., 1984, 33 272-277.
- ROYSTON, J.P. Some techniques for assessing multivariate normality
based on the Shapiro-Wilk W", *Appl. Statistic*, 1983, 32, 121-33.

END

250286

FIN