1 2

OF/DE

29068

NATIONAL LIBRARY
OTTAWA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR.....Dr. Murray David Burke
TITLE OF THESIS.....Weak and Strong Approximations of the Quantile
.....and Multivariate Empirical Processes when
.....Parameters are Estimated; and Some Distribution-
Free k-Sample Tests.
UNIVERSITY.....Carleton University
DEGREE FOR WHICH THESIS WAS PRESENTED.....Ph.D.
YEAR THIS DEGREE GRANTED.....August 1976

Permission is hereby granted to THE NATIONAL LIBRARY
OF CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and
neither the thesis nor extensive extracts from it may be
printed or otherwise reproduced without the author's
written permission.

(Signed).....*Murray David Burke*

PERMANENT ADDRESS:

.....Department of Mathematics
.....University of British Columbia
.....VANCOUVER, British Columbia

DATED.....*September 1, 1976*

NL-91 (10-68)

Supervisor: Miklos Csorgo

INFORMATION TO USERS

THIS DISSERTATION HAS BEEN
MICROFILMED EXACTLY AS RECEIVED

This copy was produced from a microfiche copy of the original document. The quality of the copy 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.

PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Canadian Theses Division
Cataloguing Branch
National Library of Canada
Ottawa, Canada K1A 0N4

AVIS AUX USAGERS

LA THESE A ETE MICROFILMEE
TELLE QUE NOUS L'AVONS RECUE

Cette copie a été faite à partir d'une microfiche du document original. La qualité de la copie dépend grandement de la qualité de la thèse soumise pour le microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

NOTA BENE: La qualité d'impression de certaines pages peut laisser à désirer. Microfilmée telle que nous l'avons reçue.

Division des thèses canadiennes
Direction du catalogage
Bibliothèque nationale du Canada
Ottawa, Canada K1A 0N4

WEAK AND STRONG APPROXIMATIONS OF THE QUANTILE AND MULTIVARIATE
EMPIRICAL PROCESSES, WHEN PARAMETERS ARE ESTIMATED;
AND SOME DISTRIBUTION-FREE k-SAMPLE TESTS

by

MURRAY DAVID BURKE, B.Sc.(Hons.), M.A.

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

Carleton University

Ottawa, Canada

August, 1976

© Murray David Burke 1976

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

"Weak and Strong Approximations of the Quantile and Multivariate
Empirical Processes, When Parameters are Estimated; and Some
Distribution-free k-sample Tests",
submitted by Murray David Burke, B.Sc.(Hons.), M.A., in partial
fulfillment of the requirements for the degree of Doctor of Philosophy.

Miklos Csorgo

.....
Dr. M. Csörgö, Thesis Supervisor

Donald A. Dawson

.....
Dr. D.A. Dawson, Chairman, Department of Mathematics

Peter J. Bickel

.....
Dr. P. J. Bickel, External Examiner

Carleton University.

August, 1976

ABSTRACT

Some statistical consequences of the recent results of M. Csörgö - P. Révész and of J. Komlós - P. Major - G. Tusnády, concerning almost-sure representations of the sample quantile process and the univariate and multivariate empirical processes, in terms of Gaussian processes, are studied. These results are useful for goodness-of-fit statistical tests, when the underlying distribution function $F(\cdot)$ is completely specified. However, in most cases, only the form of F is known while some parameters $\underline{\theta}$ of F are not specified. Suppose that $\underline{\theta}$ is estimated by $\hat{\underline{\theta}}_n$, a sequence of maximum likelihood estimators based on the sample of size n . Strong approximations by certain Gaussian processes, using the above mentioned results, of the quantile process $\hat{Q}_n(s) = \sqrt{n}[Q_n(s) - F^{-1}(s; \hat{\underline{\theta}}_n)]$ and the multivariate empirical process $\hat{E}_n(\underline{x}) = \sqrt{n}[F_n(\underline{x}) - F(\underline{x}; \hat{\underline{\theta}}_n)]$ are obtained, when the unknown parameters $\underline{\theta}$ of F are estimated by $\hat{\underline{\theta}}_n$. These Gaussian processes depend, in general, on the true unknown value of $\underline{\theta}$. This difficulty is overcome by considering, instead, properly modified versions of these Gaussian processes, obtained by replacing the true values of the parameters with their corresponding estimators. It is also proved that if $\underline{\theta}$ is estimated by a sequence of maximum likelihood estimators based on a randomly chosen half of the sample, then the resulting multivariate empirical process behaves asymptotically as if the unknown parameters $\underline{\theta}$ were completely specified.

The strong approximation methodology can be used to handle a more general class of estimators than maximum likelihood. In probability representations by certain Gaussian processes of the quantile and multivariate empirical processes are obtained, when the parameters of the underlying distribution function are estimated so that they satisfy the fairly general (A1) assumptions of Durbin. These representations are obtained under a null hypothesis and also under a general sequence of alternatives. The Gaussian processes depend, in general, on the true unknown value of the estimated parameters. The same methods, mentioned above, are used to overcome this difficulty.

Strong approximations are also obtained of the k -sample quantile and the k -sample multivariate empirical processes, in terms of Gaussian processes. The Kolmogorov-Smirnov and Cramér-von Mises statistics, based on these k -sample processes, are discussed and their rates of convergence to appropriate Gaussian process functionals are also studied.

Expressions for the asymptotic power of the Kolmogorov-Smirnov and Cramér-von Mises statistics, based on the one and k -sample quantile process and the one and k -sample multivariate empirical process, are obtained against various classes of alternative distribution functions. The above mentioned strong approximation results are used in the proofs.

Throughout the dissertation, emphasis is placed on the strong approximation methodology used.

ACKNOWLEDGEMENTS

I wish to thank Professor Miklós Csörgő, my thesis supervisor, who welcomed me into his world of research. His diligent and patient guidance and his unselfish sharing of ideas made this dissertation possible.

I also wish to thank Professor Donald Solitar of York University who helped and encouraged me to pursue doctoral studies at Carleton University.

Mrs. G. Murray, who typed this thesis, deserves a great deal of credit for her fast and very accurate work.

I wish to thank the National Research Council of Canada for their Postgraduate Scholarship support.

Lastly, I wish to thank my wife, Anne, whose love and constant encouragement made my doctoral student days very happy.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
FOREWORD	viii
I. PRELIMINARIES..	1
1. The Empirical Process.	1
2. The Quantile Process.	7
II. STRONG APPROXIMATIONS OF THE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES WHEN PARAMETERS ARE ESTIMATED.	14
3. Introduction.	14
4. A Strong Approximation of Maximum Likelihood Estimators.	15
5. Strong Approximations of the Estimated Multivariate Empirical Process.	19
6. Strong Approximations of the Estimated Quantile Process.	30
III. WEAK APPROXIMATIONS OF THE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES WHEN PARAMETERS ARE ESTIMATED.	35
7. Introduction.	35
8. Weak Approximations of the Estimated Multivariate Empirical Process.	36
9. The Maximum Likelihood Estimation Case.	48
10. Weak Approximations of the Estimated Quantile Process.	50

IV. STRONG APPROXIMATIONS OF THE k -SAMPLE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES AND SOME DISTRIBUTION-FREE

k -SAMPLE TESTS. 58

11. Introduction. 58

12. The k -sample Univariate Empirical Process. 58

13. The k -sample Quantile Process. 63

14. The k -sample Multivariate Empirical Process. 73

V. THE ASYMPTOTIC POWER OF SOME STATISTICS BASED ON THE ONE AND SEVERAL SAMPLE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES. 78

15. Introduction. 78

16. The One-sample Kolmogorov-Smirnov Tests. 80

17. The k -sample Kolmogorov-Smirnov Tests. 93

18. The One-sample Cramér-von Mises and Integral Tests. 101

19. The k -sample Cramér-von Mises and Integral Tests. 106

20. The One-sample Kolmogorov-Smirnov-type Tests for the Quantile Process. 109

21. The k -sample Kolmogorov-Smirnov-type Tests for the Quantile Process. 113

22. The One-sample Cramér-von Mises-type and Integral Tests for the Quantile Process. 115

23. The k -sample Cramér-von Mises-type and Integral Tests for the Quantile Process. 119

CONCLUDING REMARKS 120

REFERENCES 124

ADDITIONAL REFERENCES 131

FOREWORD

A classic problem in Statistics is that of goodness-of-fit. The study of the empirical distribution function $F_n(x)$ based on a random sample of size n and its "closeness" relationship with the true theoretical one $F(x)$ began with Kolmogorov (1933). Doob (1949) was the first to see the connection between the empirical process $\sqrt{n}[F_n(x) - F(x)]$ and Brownian Motion. His "heuristic approach" was later justified by Donsker (1952). They have shown that $\sqrt{n}[F_n(x) - F(x)]$ converges in distribution to a Brownian Bridge, if F is continuous. Consequently, statistics such as the Kolmogorov-Smirnov and the Cramér-von Mises, based on the empirical process can be viewed asymptotically, in a unified way, as continuous functionals of the Brownian Bridge process. For illumination of these remarks as well as further results and references, we refer to Billingsley (1968) and to Durbin (1973b).

These results can be used to handle goodness-of-fit problems when the true distribution function $F(x)$ is completely specified. However, in most cases, one only knows the form of the distribution function, $F(x; \theta)$, while some parameters θ are not specified. One approach to this problem, which is taken in this exposition, is to estimate the unknown nuisance parameters θ from the given random sample of size n and substitute these estimators (denoted by $\hat{\theta}_n$) in the distribution function. We call the process

$\sqrt{n}[F_n(x) - F(x; \hat{\theta}_n)]$ the estimated empirical process. Study of the estimated empirical process was begun by I. I. Gihman (1953 a and b, 1954), who developed a framework for conditional limit theorems involving functionals of independent random variable sequences, using solutions of partial differential equations to identify some asymptotic conditional probabilities in question. He used his theory to identify a number of asymptotic conditional probabilities for the estimated empirical process. Subsequently, Darling (1955) studied the Cramér-von Mises test in this parametric case. Kac, Kiefer and Wolfowitz (1955) considered tests of normality when both the mean and the variance are estimated and they computed tables for the limiting distribution of the Cramér-von Mises and Kolmogorov-Smirnov statistics, when parameters are estimated. The type of estimation used in the papers quoted above is that of maximum likelihood. More recently, Durbin (1973a) considered a wider class of estimators in his paper on weak convergence of the estimated empirical process. Durbin (1973a) (see also Burke (1975)) proves that the estimated empirical process, under a sequence of alternatives, converges in distribution to a Gaussian process and gives its mean and covariance.

Kiefer (1972) and subsequently Komlós, Major and Tusnády (1975) have obtained almost-sure representations, with rates of convergence, of the empirical process in terms of Gaussian processes.

This enabled Csörgö, Komlós, Major, Révész and Tusnády ([14], to appear) to introduce a new methodology to handle the estimated empirical process. Using these new (strong approximation) techniques, they obtained an almost-sure representation, with rates of convergence, of the univariate empirical process $\sqrt{n}[F_n(x) - F(x; \hat{\theta}_n)]$ in terms of a Gaussian process, when one parameter θ is estimated by a sequence $\hat{\theta}_n$ of maximum likelihood estimators. Subsequent to Csörgö and Révész (1975) and to [14], Csörgö and Révész ([11], [12], to appear) have obtained similar almost-sure representations of the multivariate empirical process and also of the (univariate) sample quantile process, when $F(\cdot)$ is continuous and satisfies some regularity conditions. In this dissertation, we shall elaborate on and explore further the techniques in [14] for the multivariate empirical process and the sample quantile process, when a vector of unknown parameters is estimated. We shall also show that Durbin's wider class of estimators can be accommodated by the strong approximation techniques.

Another classic problem concerning goodness-of-fit is whether two or more samples come from the same or different populations. Study of this problem was initiated by N. V. Smirnov (1939). J. Kiefer (1959) obtained the limiting distribution of the k -sample analogues of the Kolmogorov-Smirnov and the Cramér-von Mises statistics in the univariate case, where these tests are distribution-free.

For further references and results concerning multivariate observations, we refer to Bickel (1969). In this exposition, we shall show how the strong approximation results of Komlós, Major, Tusnády (1975) and Csörgő and Révész ([11], [12], to appear) can be used to obtain almost-sure and in-probability representations of the k -sample quantile and multivariate empirical processes in terms of Gaussian process. These techniques enable us to study the asymptotic power of statistics based on the one and several sample quantile and multivariate empirical process in a natural and unified way.

This dissertation is divided into five chapters. In Chapter I, we introduce some notation and quote the theorems which will be used in the sequel.

Chapter II deals with obtaining almost-sure representations of the estimated multivariate empirical process $\sqrt{n}[F_n(\underline{x}) - F(\underline{x}; \hat{\underline{\theta}}_n)]$ in terms of Gaussian processes, when $\hat{\underline{\theta}}_n$ is the maximum likelihood estimator of the unknown vector $\underline{\theta}$. We begin by showing that $\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_0)$ can be almost-surely represented by a sum of independent and identically distributed random vectors. We obtain similar almost-sure representations for the estimated quantile process. The Gaussian process representations depend, in general, on the true unknown value of $\underline{\theta}$. Some useful ways of overcoming this are discussed.

Our Chapter III is concerned with obtaining in-probability representations of the estimated multivariate empirical process in

terms of Gaussian processes, when Durbin's (1973a) wider class of estimators are used. These representations are obtained under a slightly more general sequence of alternatives than Durbin's. Again, the Gaussian representations depend, in general, on the true unknown value of θ and ways of overcoming this are examined. The same program is carried out for the estimated quantile process.

In Chapter IV, we obtain almost-sure representations of the k -sample quantile and multivariate empirical processes. Statistics, used to test whether several independent samples have the same distribution function, either known or unknown, are discussed. In dealing with the k -sample quantile process, empirical density functions are employed to obtain distribution-free tests, when the true common distribution function is unknown.

Our Chapter V contains a unified treatment (via strong approximations) of the consistency and the asymptotic power of the Kolmogorov-Smirnov, Cramér-von Mises and integral test statistics, based on the one and k -sample quantile and multivariate empirical processes.

We wish to put our emphasis not solely on the particular results obtained, but also on the strong approximation methodology used.

The results in Chapters II, III and IV are joint with M. Csörgő, unless explicitly indicated otherwise. Chapter V is based on the author's paper ([8], to appear). The results of these chapters are believed to be new.

August, 1976

I. PRELIMINARIES

1. The Empirical Process

Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed (i.i.d.) univariate random variables with a continuous, completely specified, distribution function (d.f.) $F(x) = P\{X_1 \leq x\}$. Further, let $F_n(x)$ denote the empirical d.f. based on the outcomes of $X_1, X_2, \dots, X_n$, that is, $F_n(x)$ is the proportion of the $X_i \leq x$ ($i = 1, 2, \dots, n$). The empirical process $E_n(x)$ is defined by

$$(1.1) \quad E_n(x) = \sqrt{n}[F_n(x) - F(x)] , \quad x \in R .$$

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

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

J. Kiefer (1972) first constructed the process which bears his name, in his paper on Skorohod embedding of the empirical process. Both of these processes can be represented in terms of Wiener processes:

$$B(s) \stackrel{d}{=} W(s) - s W(1), \quad 0 \leq s \leq 1;$$

$$K(s,y) \stackrel{d}{=} W(s,y) - s W(1,y), \quad 0 \leq s \leq 1, y \geq 0,$$

where $\{W(s) : 0 \leq s \leq 1\}$ is a standard Wiener process with $E W(s) = 0$ and $E W(s) W(t) = s \wedge t$, and $W(s,y)$ is a standard two-dimensional Wiener process with $E W(s,y) = 0$ and $E W(s,y_1) W(t,y_2) = (y_1 \wedge y_2)(s \wedge t)$. Note that for fixed $y > 0$, $B(s) \stackrel{d}{=} y^{-\frac{1}{2}} K(s,y)$.

Komlós, Major and Tusnády (1975) have shown that $E_n(x)$, defined by (1.1), is "near" to a sequence of Brownian Bridges and a Kiefer process. We state their result:

Theorem 1A. (Komlós, Major, Tusnády, 1975).

If the underlying probability space is rich enough, one can define a Brownian Bridge $\{B_n(s) : 0 \leq s \leq 1\}$ for each $n \in N$, and a Kiefer process $\{K(s,y) : 0 \leq s \leq 1, y \geq 0\}$ such that

$$P\left\{\sup_{x \in R} |E_n(x) - B_n(F(x))| > n^{-\frac{1}{2}}(C \log n + z)\right\} < K e^{-\lambda z}$$

for all z , where C, K and λ are positive absolute constants;

$$P\left\{\sup_{1 \leq j \leq n} \sup_{x \in R} |\sqrt{j} E_j(x) - K(F(x), j)| > (C \log n + z) \log n\right\} < K e^{-\lambda z}$$

for all z and $n \in N$, where C, K and λ are positive absolute constants. These inequalities, in turn, imply

$$(1.2) \quad \sup_{x \in R} |E_n(x) - B_n(F(x))| \stackrel{a.s.}{=} O\{n^{-\frac{1}{2}} \log n\}$$

and

$$(1.3) \quad \sup_{x \in R} |\sqrt{n} E_n(x) - K(F(x), n)| \stackrel{a.s.}{=} O\{\log^2 n\}.$$

Remark 1.1. By the phrase "the underlying probability space is rich enough", we mean that an independent sequence of Wiener processes, independent of the originally given i.i.d. sequence $\{X_n\}$, can be constructed on the assumed probability space. If this is the case, one can construct the appropriate Gaussian processes on the probability space of the sequence $\{X_n\}$ so that (1.2) and (1.3) hold. In the sequel, it will be assumed that the underlying probability spaces are rich enough in the above sense.

Also, for the univariate empirical process, we have the following inequality by Dvoretzky, Kiefer and Wolfowitz (1956):

there is a universal constant C such that

$$(1.4) \quad P\left\{ \sup_{x \in R} |E_n(x)| \geq z \right\} \leq C \exp \{-2z^2\},$$

for all $n \in N$ and $z \geq 0$.

A multivariate version of (1.4) was proved by Kiefer (1961).

Now, let $\underline{X}_1, \underline{X}_2, \dots$ be a sequence of i.i.d. d -dimensional random vectors with a continuous d.f. $F(\underline{x}) = P\{\underline{X}_1 \leq \underline{x}\}$, where " $\leq$ " is the usual partial order relation of R^d . Let $F_n(\underline{x})$

denote the empirical d.f. based on $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$, that is,
 $F_n(\underline{x})$ is the proportion of the $\underline{X}_i \leq \underline{x}$ ($i = 1, 2, \dots, n$). Finally,
 we define the multivariate empirical process by

$$(1.5) \quad E_n(\underline{x}) = \sqrt{n}[F_n(\underline{x}) - F(\underline{x})], \quad \underline{x} \in R^d.$$

M. Csörgö and P. Révész have shown that $E_n(\underline{x})$ is "near"
 to properly modified versions of the following Gaussian processes:

- (c) a Brownian Bridge $B(\underline{s})$ ($\underline{s} \in I^d$), where $I = [0, 1]$,
 with $E B(\underline{s}) = 0$ and covariance function
 $E B(\underline{s}) B(\underline{t}) = (s_1 \wedge t_1)(s_2 \wedge t_2) \dots (s_d \wedge t_d) - s_1 s_2 \dots s_d t_1 t_2 \dots t_d$,
 where $\underline{s} = (s_1, s_2, \dots, s_d)$, $\underline{t} = (t_1, t_2, \dots, t_d)$;
 (d) a Kiefer process $K(\underline{s}, y)$ ($\underline{s} \in I^d$, $0 \leq y < \infty$), with
 $E K(\underline{s}, y) = 0$ and covariance function
 $E K(\underline{s}, y_1) K(\underline{t}, y_2) = (y_1 \wedge y_2) \cdot E B(\underline{s}) \cdot B(\underline{t})$.

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

$$B(\underline{s}) \stackrel{\mathcal{D}}{=} W(\underline{s}) - s_1 s_2 \dots s_d W(1, 1, \dots, 1);$$

$$K(\underline{s}, y) \stackrel{\mathcal{D}}{=} W(\underline{s}, y) - s_1 s_2 \dots s_d W(1, 1, \dots, 1, y),$$

where $\{W(\underline{s}); \underline{s} \in I^d\}$ has mean zero and covariance function
 $E W(\underline{s}) W(\underline{t}) = (s_1 \wedge t_1)(s_2 \wedge t_2) \dots (s_d \wedge t_d)$, and $\{W(\underline{s}, y);$
 $\underline{s} \in I^d, 0 \leq y < \infty\}$ has mean zero and covariance function

$$E W(\underline{s}, y_1) W(\underline{t}, y_2) = (y_1 \wedge y_2) (s_1 \wedge t_1) (s_2 \wedge t_2) \dots (s_d \wedge t_d)$$

Note that for fixed $y > 0$, $B(\underline{s}) \stackrel{\text{def}}{=} y^{-\frac{1}{2}} K(\underline{s}, y)$.

We shall first state some notation which will be used in the sequel. For a random vector $\underline{Y} = (Y_1, Y_2, \dots, Y_d)$ with continuous d.f. $F(\underline{y}) = F(y_1, y_2, \dots, y_d)$, set

$$\begin{aligned} F_1(y_1) &= P\{Y_1 \leq y_1\} \quad \text{and} \\ (1.6) \quad F_i(y_i | y_1, y_2, \dots, y_{i-1}) &= P\{Y_i \leq y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}\} \\ &\quad (i = 2, 3, \dots, d). \end{aligned}$$

For the d.f. F , define the transformations T_i by

$$\begin{aligned} T_i(y_1, y_2, \dots, y_i) &= (F_1(y_1), F_2(y_2 | y_1), \dots, F_i(y_i | y_1, y_2, \dots, y_{i-1})) \\ &\quad (i = 1, 2, \dots, d). \end{aligned}$$

These transformations were introduced by Paul Lévy (see also Rosenblatt (1952)). Further, for $\underline{x} = (x_1, x_2, \dots, x_d)$, define

$$(1.7) \quad D_{\underline{x}} = \{(a_1, a_2, \dots, a_d) \in \mathbb{R}^d : a_i \leq x_i, \quad i = 1, 2, \dots, d\};$$

$$\begin{aligned} (1.8) \quad TD_{\underline{x}} = TFD_{\underline{x}} &= \{(b_1, b_2, \dots, b_d) \in I^d : b_i \leq F_i(x_i | x_1, \dots, x_{i-1}), \\ &\quad i = 1, 2, \dots, d\}. \end{aligned}$$

We have

Theorem 1B. (Csörgö and Révész [11], to appear).

Suppose that the following conditions are satisfied:

- (i) the functions $F_i(y_i | y_1, \dots, y_{i-1})$ ($i = 1, 2, \dots, d$) are continuous, strictly monotone d.f. in y_i , for any fixed $y_1, y_2, \dots, y_{i-1}$;
- (ii) the functions $F_i(y_i | T_{i-1}^{-1}(t_1, t_2, \dots, t_{i-1}))$ ($i = 2, 3, \dots, d$) are differentiable with respect to $t_1, t_2, \dots, t_{i-1}$ over the interior of I^{i-1} and

$$\sum_{j=1}^{i-1} \int_{I^{i-1}} \left| \frac{\partial}{\partial t_j} F(y_i | T_{i-1}^{-1}(t_1, t_2, \dots, t_{i-1})) \right| dt_1 dt_2 \dots dt_{i-1} \leq C,$$

where C is a positive constant.

Then, one can define a sequence $\{B_n(\underline{s}) : \underline{s} \in I^d\}$ of Brownian Bridges and a Kiefer process $\{K(\underline{s}, y) : \underline{s} \in I^d, y \geq 0\}$ such that

$$\sup_{\underline{x} \in R^d} |E_n(\underline{x}) - B_n(TD_{\underline{x}})| \stackrel{a.s.}{=} O\left\{n^{-\frac{1}{2(d+1)}} (\log n)^{3/2}\right\}$$

and

$$\sup_{\underline{x} \in R^d} |\sqrt{n} E_n(\underline{x}) - K(TD_{\underline{x}}, n)| \stackrel{a.s.}{=} O\left\{n^{\frac{(d+1)}{2(d+2)}} \log^2 n\right\},$$

where $TD_{\underline{x}} = TFD_{\underline{x}}$ is defined by (1.8).

Remark 1.2. Conditions (i) and (ii) of Theorem 1B enabled Csörgö and Révész to approximate the sets $TD_{\underline{x}}$ ($\underline{x} \in R^d$) by rectangles in I^d , uniformly in $\underline{x}$. The Brownian Bridge measure and the Kiefer measure were defined on these sets in the inclusion-exclusion way.

In fact, $B(TD_{\underline{x}})$ is a representation of the Gaussian process $G(\underline{x}) = B(TD_{\underline{x}})$ (the existence of which was first proved by Dudley (1966))

in terms of a Brownian Bridge measure. The covariance is given by

$$\begin{aligned} E B(TD_{\underline{x}}) B(TD_{\underline{y}}) &= \lambda(TD_{\underline{x}} \cap TD_{\underline{y}}) - \lambda(TD_{\underline{x}}) \cdot \lambda(TD_{\underline{y}}) \\ &= F(\underline{x} \wedge \underline{y}) - F(\underline{x}) F(\underline{y}), \end{aligned}$$

where λ is the Lebesgue measure on I^d and $\underline{x} \wedge \underline{y} = (x_1 \wedge y_1, x_2 \wedge y_2, \dots, x_d \wedge y_d)$.

Remark 1.3. Many multivariate d.f. satisfy the conditions (i) and (ii) of Theorem 1B. One example is the multivariate normal.

2. The Quantile Process.

Suppose $X_1, X_2, \dots$ is a sequence of i.i.d. univariate random variables with a continuous d.f. $F(x)$. Define the quantile function $Q_n(s)$, $0 < s < 1$, by

$$(2.1) \quad Q_n(s) = X_{(j)} \quad \text{if} \quad \frac{j-1}{n} < s \leq \frac{j}{n} \quad (j = 1, 2, \dots, n),$$

and the quantile process $q_n(s)$ by

$$(2.2) \quad q_n(s) = \sqrt{n} [Q_n(s) - F^{-1}(s)] \quad (0 < s < 1),$$

where $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are the order statistics based on the first n observations and $F^{-1}(s)$ is the inverse function of F .

In the sequel, we shall require the following strong approximation theorems for the quantile process:

Theorem 2A. (Csörgö and Révész [12], to appear).

Suppose that $F(x)$ is twice differentiable on (a, b) , where $a = \sup \{x : F(x) = 0\}$ and $b = \inf \{x : F(x) = 1\}$ and that $F' = f \neq 0$ on (a, b) .

Assume that the following conditions hold:

$$(2.3) \quad \sup_{\log n \leq k \leq n - \log n} \sup_{\xi \in [a_{nk}, b_{nk}]} \frac{f[F^{-1}(\frac{k}{n})]}{f[F^{-1}(\xi)]} = O(1),$$

$$\text{with } a_{nk} = \frac{k - [\frac{k}{n}(1 - \frac{k}{n})n \log n]^{\frac{1}{2}}}{n}, \quad b_{nk} = \frac{k + [\frac{k}{n}(1 - \frac{k}{n})n \log n]^{\frac{1}{2}}}{n}$$

and for some $C > 0$,

$$(2.4) \quad \sup_{x \in R} F(x)[1 - F(x)] \left| \frac{f'(x)}{(f(x))^2} \right| \leq C.$$

Then, one can define a sequence $\{B_n(s) : 0 \leq s \leq 1\}$ of Brownian Bridges such that

$$(2.5) \quad \sup_{s \in A(n)} |f[F^{-1}(s)] q_n(s) - B_n(s)| \stackrel{a.s.}{=} O\{n^{-\frac{1}{2}} \log n\},$$

where $A(n)$ is the closed interval $[\frac{\log n}{n}, 1 - \frac{\log n}{n}]$.

If it is also assumed that

$$(2.6) \quad x \cdot f(x) \rightarrow 0 \text{ as } x \rightarrow +\infty, \text{ and } x f(x) \rightarrow 0 \text{ as } x \rightarrow -\infty,$$

then for any $\epsilon > 0$,

$$(2.7) \quad \sup_{0 < s < 1} |f[F^{-1}(s)] q_n(s) - B_n(s)| \stackrel{a.s.}{=} O\{\sqrt{n} w_n(\epsilon), n^{-\frac{1}{2}} \log n\},$$

where $w_n(\epsilon) = |F^{-1}(n^{-2-\epsilon})| \cdot f[F^{-1}(n^{-1} \log n)]$.

Remark 2.1. If $f(\cdot)$ is bounded away from zero and infinity and $|f'(\cdot)|$ is bounded on (a,b) , assumed now to be finite, then conditions (2.3), (2.4) and (2.6) hold and $w_n(\epsilon) = O(n^{-2-\epsilon})$. Consequently, (2.7) implies

$$(2.8) \quad \sup_{0 < s < 1} |f[F^{-1}(s)] q_n(s) - B_n(s)| \stackrel{a.s.}{=} O\{n^{-\frac{1}{2}} \log n\}.$$

Remark 2.2. Condition (2.3) can be difficult to verify in particular situations. It is satisfied by a very large class of d.f.. The following conditions are easier to check and together imply (2.3):

$$(2.9) \quad \frac{|f'(x)|}{(f(x))^2} [F(x)(1 - F(x))]^{3/2} \leq C,$$

$$|f''(x)| \leq C, \quad \frac{F(x)[1 - F(x)]}{f(x)} \leq C,$$

for some constant $C > 0$. For a proof of this assertion, we refer to Csörgö and Révész ([12], to appear). We note that the (2.9) conditions are immediately justifiable for a number of densities; e.g. the exponential, gamma and normal laws.

Remark 2.3. To make statement (2.8) more concrete, we note that

$$\sqrt{n} w_n(\epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ if } \lim_{|x| \rightarrow \infty} x^\alpha f(x) = 0, \text{ for some } \alpha > 3.$$

Given that the density f is positive and bounded away from zero on a finite interval, one can obtain a strong approximation of $q_n(s)$ in terms of a Kiefer process:

Theorem 2B.

Suppose that F is twice differentiable on a finite interval (a,b) and that $F'(x) = f(x) \geq \varepsilon > 0$, for $x \in (a,b)$ and zero otherwise.

Then, if $|f'(x)| \leq C < \infty$ for $x \in (a,b)$, one can define a Kiefer process $\{K(s,y) : 0 < s < 1, y \geq 0\}$ such that

$$\sup_{0 < s < 1} |\sqrt{n} q_n(s) \cdot f[F^{-1}(s)] - K(s,n)| \stackrel{\text{a.s.}}{=} O\{(n \log \log n)^{\frac{1}{4}} (\log n)^{\frac{1}{2}}\}.$$

As pointed out by Csörgö and Révész ([12], to appear), Theorem 2B is a direct consequence of Theorem 1A and

Theorem 2C. (Kiefer, 1970).

Under the conditions of Theorem 2B,

$$\limsup_{n \rightarrow \infty} n^{\frac{1}{4}} (\log \log n)^{-\frac{1}{4}} (\log n)^{-\frac{1}{2}} \sup_{0 < s < 1} |q_n(s) + E_n(F^{-1}(s)) \cdot [f(F^{-1}(s))]^{-1}|$$

$$\stackrel{\text{a.s.}}{=} 2^{-\frac{1}{4}}.$$

Remark 2.4. A disadvantage of strong approximations in terms of a sequence of Brownian Bridges is that one does not know anything about their joint distribution in n . Consequently, only "in-probability" and "in-distribution" limit laws can be proved from them for $E_n(x)$ and $q_n(s)$. In spite of their weaker rates of convergence (e.g. Theorem 2B versus Theorem 2A), an advantage of strong approximations in terms of Kiefer processes is that one can produce strong laws for $E_n(x)$ and $q_n(s)$ by establishing the same for the Kiefer

process. For example, knowing the law of the iterated logarithm for the Kiefer process:

$$\limsup_{n \rightarrow \infty} \sup_{0 \leq s \leq 1} \left| \frac{K(s, n)}{\sqrt{2n \log \log n}} \right| \stackrel{\text{a.s.}}{=} \frac{1}{2},$$

one can deduce the same for $q_n(s)$, via Theorem 2B (cf. Csörgö and Révész [12]):

$$\limsup_{n \rightarrow \infty} \sup_{0 \leq s \leq 1} \left| \frac{q_n(s) \cdot f[F^{-1}(s)]}{\sqrt{2 \log \log n}} \right| \stackrel{\text{a.s.}}{=} \frac{1}{2},$$

and for $E_n(x)$, via Theorem 1A:

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| \frac{E_n(x)}{\sqrt{2 \log \log n}} \right| \stackrel{\text{a.s.}}{=} \frac{1}{2}.$$

It is clear that Theorems 2A, 2B and 2C involve the density f of P . In Chapters IV and V of this dissertation, we will be estimating the density f by empirical density functions. Consider the following two definitions taken from P. Révész (1972):

Definition 2.1. (Uniform Partition). Let

$$\dots < x_{-2}(n) < x_{-1}(n) < x_0(n) < x_1(n) < \dots$$

be a partition of the real line such that

$$a_n = x_{i+1}(n) - x_i(n) \quad (i = 0, \pm 1, \pm 2, \dots)$$

and $\{a_n\}$ is a decreasing sequence of real numbers. Define the empirical density $f_n^{(1)}(x)$ by

$$f_n^{(1)}(x) = \frac{\#_n \{x_i(n); x_{i+1}(n)\}}{n \cdot a_n}, \quad \text{if } x_i(n) \leq x < x_{i+1}(n),$$

where $\#_n \{a; b\}$ is the number of the observations $X_1, X_2, \dots, X_n$ lying in the interval $[a, b)$.

For simplicity, denote $\#_n \{x_i(n); x_{i+1}(n)\}$ by $\#(i, n)$, $(i = 0, \pm 1, \pm 2, \dots)$.

Definition 2.2. (Smoothed empirical density based on the uniform partition).

Using the partition of Definition 2.1, define $f_n^{(2)}(x)$ by

$$\begin{aligned} f_n^{(2)}(x) &= (n a_n)^{-1} \left\{ \#(i, n) + \left| \frac{\#(i+1, n) - \#(i, n)}{a_n} \right| \cdot \left| x - \frac{x_{i+1}(n) + x_i(n)}{2} \right| \right\} \\ &= f_n^{(1)}(x_i(n)) + a_n^{-1} \cdot \{ f_n^{(1)}(x_{i+1}(n)) - f_n^{(1)}(x_i(n)) \} \cdot \\ &\quad \left\{ x - \frac{x_{i+1}(n) + x_i(n)}{2} \right\}, \\ \text{if } \frac{x_i(n) + x_{i+1}(n)}{2} &\leq x < \frac{x_{i+1}(n) + x_{i+2}(n)}{2}. \end{aligned}$$

One can obtain the following strong consistency result:

Theorem 2D. (Révész, 1972). Suppose that $f(x)$ is a uniformly continuous density function and there exists a $\delta > 0$, such that

$$\int_{-\infty}^{+\infty} |x|^{\delta} \cdot f(x) dx < \infty .$$

Further, let $a_n \geq w(n) \frac{\log n}{n}$, where $w(n)$ is a function tending to $+\infty$ as $n \rightarrow +\infty$. Then,

$$\sup_{x \in \mathbb{R}} |f_n^{(r)}(x) - f(x)| \xrightarrow{\text{a.s.}} 0 ,$$

as $n \rightarrow +\infty$, where $r = 1, 2$.

Remark 2.5. There are many possible definitions for empirical density functions. For a survey of these definitions, we refer to Rosenblatt (1956) and for results on their strong consistency, to Révész (1972). For results on strong approximations of density functions by empirical density functions, we refer to Bickel and Rosenblatt (1973) and to Révész ([43], to appear).

II. STRONG APPROXIMATIONS OF THE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES WHEN PARAMETERS ARE ESTIMATED

3. Introduction.

The results stated in Chapter I concern strong approximations of the quantile and multivariate empirical processes, when the underlying d.f. $F(\underline{x})$ is completely specified. However, in most cases, only the form $F(\underline{x}; \underline{\theta})$ of the d.f. is specified, while some parameters $\underline{\theta}$ of F are not known. In this chapter, we shall estimate $\underline{\theta}$ by $\{\hat{\theta}_n\}$, a sequence of maximum likelihood estimators (m.l.e.). If $\underline{\theta}$ is one-dimensional, Ibragimov and Has'minskii (1973) have shown that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can be almost-surely represented as a sum of i.i.d. random variables.* Section 4 extends their result to the case when $\underline{\theta}$ and the observations are both multidimensional. Using the results of Chapter I, in Sections 5 and 6, we obtain almost-sure representations of $\sqrt{n}[F_n(\underline{x}) - F(\underline{x}; \hat{\theta}_n)]$ and $\sqrt{n}[Q_n(s) - F^{-1}(s; \hat{\theta}_n)]$ in terms of Gaussian processes.

As to the nature of our results, we wish to put the emphasis on the strong approximation techniques in general, rather than on the particular rates gained as a result of the numerous regularity conditions imposed. Indeed, better rates might be obtained using our techniques, in case of any particular distribution and/or estimator.

The results of this chapter appear also in Burke and Csörgö ([5], to appear).

* θ_0 is the unknown true value of θ , that is, the X_j have true d.f. $F(\cdot; \theta_0)$.

We shall use the following notations:

(i) The norm $|| \cdot ||$ defined on R^p will denote the usual Euclidean norm.

(ii) For a function $g(\underline{\theta})$, where $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_p) \in R^p$, $\frac{\partial}{\partial \underline{\theta}} f(\underline{\theta}_0)$ will denote the vector $(\frac{\partial}{\partial \theta_1} g(\underline{\theta}), \frac{\partial}{\partial \theta_2} g(\underline{\theta}), \dots, \frac{\partial}{\partial \theta_p} g(\underline{\theta}))^t$, evaluated at $\underline{\theta} = \underline{\theta}_0$, and $\frac{\partial^2}{\partial \underline{\theta}^2} g(\underline{\theta})$ will denote the vector $(\frac{\partial^2}{\partial \theta_1^2} g(\underline{\theta}), \frac{\partial^2}{\partial \theta_2^2} g(\underline{\theta}), \dots, \frac{\partial^2}{\partial \theta_p^2} g(\underline{\theta}))^t$, evaluated at $\underline{\theta} = \underline{\theta}_0$.

(iii) For the particular function $l(\underline{x}, \underline{\theta}) = \log f(\underline{x}; \underline{\theta})$, let $l'_{\underline{\theta}}(\underline{x}, \underline{\theta})$ denote the vector $\frac{\partial}{\partial \underline{\theta}} l(\underline{x}, \underline{\theta})$ and let $l''_{\underline{\theta}\underline{\theta}}(\underline{x}, \underline{\theta})$ denote the matrix $(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\underline{x}, \underline{\theta}))$.

(iv) For a matrix or vector $\underline{V} = (v_{ij})$, let $|\underline{V}|$ denote the matrix $(|v_{ij}|)$, $\int \underline{V}$ denote $(\int v_{ij})$, let $\underline{V}^\delta$ denote (v_{ij}^δ) and let $\underline{V}^t$ denote the transpose of $\underline{V}$.

4. A Strong Approximation of Maximum Likelihood Estimators.

Let $\underline{X}_1, \underline{X}_2, \dots$ be a sequence of i.i.d. d-dimensional random vectors with an absolutely continuous d.f. $F(\cdot; \underline{\theta})$, where $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ is a vector of unknown parameters. Extending the technique for univariate θ by I. A. Ibragimov and R.Z. Has'minskii (1973), we shall estimate the vector $\underline{\theta}$ by $\hat{\underline{\theta}}_n$, a sequence of maximum likelihood estimators (m.l.e.) based on the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$. Let $\underline{\theta}_0$ stand for the theoretical true value of $\underline{\theta}$, that is, suppose $\underline{X}_1, \underline{X}_2, \dots$ are distributed according to $F(\cdot; \underline{\theta}_0)$.

Under certain regularity conditions, we shall approximate

$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} I(\theta_0)$ by a sequence of i.i.d. random vectors, where $I(\theta)$ is the Fisher information matrix.

Consider the following assumptions:

(A1) (i) $F(\underline{x}; \underline{\theta})$ has a density function $f(\underline{x}; \underline{\theta})$.

(ii) The parameter set ν is an open cube (bounded or unbounded) in R^p .

(iii) The density $f(\underline{x}; \underline{\theta})$ is measurable with respect to $\mathcal{X} \times \xi$, where $\mathcal{X}$ is the collection of Borel subsets of R^d and ξ is a σ -algebra of measurable subsets of ν .

(iv) If $\underline{\theta} \neq \underline{\theta}'$, then $\int_{R^d} |f(\underline{x}; \underline{\theta}) - f(\underline{x}; \underline{\theta}')| d\underline{x} > 0$.

(v) There is a positive number δ such that

$$\sup_{\underline{\theta}} \|\underline{\theta} - \underline{\theta}_0\|^\delta \int_{R^d} [f(\underline{x}; \underline{\theta}) f(\underline{x}; \underline{\theta}_0)]^{\frac{1}{2}} d\underline{x} < \infty.$$

(vi) All second partial derivatives of the function

$l(\underline{x}, \underline{\theta}) = \log f(\underline{x}; \underline{\theta})$, with respect to the components of $\underline{\theta}$, exist.

(vii) First and second partial differentiation of

$\int_{R^d} f(\underline{x}; \underline{\theta}) d\underline{x}$, with respect to the components of

$\underline{\theta}$, can be taken under the integral sign.

(viii) The matrix of functions $E_{\underline{\theta}} |l''_{\underline{\theta}\underline{\theta}}(\underline{X}_1, \underline{\theta})|^{1+\delta}$ and the vector $E_{\underline{\theta}} |l'_{\underline{\theta}}(\underline{X}_1, \underline{\theta})|^{2+\delta}$ have bounded components, for some $\delta > 0$, on every compact subset of ν .
 (If $A = (a_{ij})$, $|A|^{1+\delta}$ is the matrix $(|a_{ij}|^{1+\delta})$ and $E_{\underline{\theta}} |A|^{1+\delta}$ is the matrix $(E_{\underline{\theta}} |a_{ij}|^{1+\delta})$.)

(ix) for some $\beta > 0$,

$$||l'_{\underline{\theta}\underline{\theta}}(\underline{x}, \underline{\theta}_2) - l'_{\underline{\theta}\underline{\theta}}(\underline{x}, \underline{\theta}_1)|| < H(\underline{x}) \cdot ||\underline{\theta}_2 - \underline{\theta}_1||^{\beta},$$

and $E_{\underline{\theta}} H(\underline{X}_1)$ exists, for each $\underline{\theta} \in \nu$.

We note that (A1) (i) to (v) are the multivariate versions of the univariate conditions I and III of Ibragimov and Has'minski (1972).

We have

Theorem 4.1. If the (A1) assumptions are satisfied, then for some $\epsilon > 0$,

$$\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_0)^t \cdot I(\underline{\theta}_0) = n^{-1/2} \sum_{j=1}^n l'_{\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) \stackrel{a.s.}{=} O(n^{-\epsilon}).$$

Proof. Using the identity $\sum_{j=1}^n l'_{\underline{\theta}}(\underline{X}_j, \hat{\underline{\theta}}_n) = 0$ and the vector form of Taylor's Theorem,

$$\begin{aligned} S_n &= n^{-1/2} \sum_{j=1}^n l'_{\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) = n^{-1/2} \sum_{j=1}^n [l'_{\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) - l'_{\underline{\theta}}(\underline{X}_j, \hat{\underline{\theta}}_n)] \\ &= n^{-1/2} (\underline{\theta}_0 - \hat{\underline{\theta}}_n)^t \cdot \sum_{j=1}^n l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_j), \end{aligned}$$

* $E_{\underline{\theta}}$, resp. $P_{\underline{\theta}}$, denotes the expectation, resp. the probability measure, associated with the d.f. $F(\cdot; \underline{\theta})$.

where the vectors $\underline{\theta}_j$ satisfy $||\underline{\theta}_0 - \underline{\theta}_j|| < ||\underline{\theta}_0 - \hat{\underline{\theta}}_n||$.

By adding and subtracting $E_{\underline{\theta}}[l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0)]$ and $l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0)$, and since $E_{\underline{\theta}}[l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0)] = -I(\underline{\theta}_0)$, we obtain

$$S_n = \sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_0)^t \left[I(\underline{\theta}_0) + n^{-1} \sum_{j=1}^n \{ l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) - E_{\underline{\theta}}[l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0)] \} \right] \\ + n^{-\frac{1}{2}}(\underline{\theta}_0 - \hat{\underline{\theta}}_n)^t \cdot \sum_{j=1}^n \{ l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_j) - l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) \}.$$

Denote the second and third summands in the above expression by J_2 and J_3 , respectively. Then,

$$J_2 = n^{\frac{1}{2}-\varepsilon}(\underline{\theta}_0 - \hat{\underline{\theta}}_n)^t n^{2\varepsilon-1} \sum_{j=1}^n \{ l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) - E_{\underline{\theta}} l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) \}.$$

By (A1) (viii), each component of $l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0)$ possesses the absolute moment of order $1+\delta$, which implies that for $2\varepsilon < \delta(1+\delta)^{-1}$,

$$n^{2\varepsilon-1} \sum_{j=1}^n \{ l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) - E_{\underline{\theta}} l''_{\underline{\theta}\underline{\theta}}(\underline{X}_j, \underline{\theta}_0) \} \xrightarrow{a.s.} 0, \text{ as } n \rightarrow \infty,$$

in view of the obvious multivariate generalization of Marcinkiewicz's Theorem (see Loève (1963), 4^o of §16.4).

Ibragimov and Has'minskiĭ (1972) have shown that for any $\eta > 0$, there exists an n_0 such that for $n \geq n_0$, $E_{\underline{\theta}} ||\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_0)||^{\eta}$ exists and converges to the η^{th} moment of $N(0, I(\underline{\theta}_0)^{-1})$, and hence is bounded in n . By Markov's inequality (2.11.1 of Rényi (1970)),

$$\begin{aligned} P_{\underline{\theta}} \{ n^{\frac{1}{2}-2\epsilon} ||\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0|| > n^{-\epsilon} \} &= P \{ \sqrt{n} ||\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0|| > n^{\epsilon} \} \\ &\leq n^{-k\epsilon} E_{\underline{\theta}} ||\sqrt{n}(\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0)||^k. \end{aligned}$$

By the Borel-Cantelli lemma, on choosing $k > \epsilon^{-1}$,

$$P_{\underline{\theta}} \left\{ \overline{\lim}_{n \rightarrow \infty} n^{\frac{1}{2}-2\epsilon} ||\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0|| > n^{-\epsilon} \right\} = 0, \text{ that is } J_2 \stackrel{\text{a.s.}}{=} o(n^{-\epsilon}).$$

$$\text{By (A1) (vi) to (ix), we obtain } ||J_3|| < n^{-\frac{1}{2}} ||\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0||^{1+\beta} \sum_{j=1}^n H(\underline{X}_j).$$

By the SLLN, $\sum_{j=1}^n n^{-1} H(\underline{X}_j) \xrightarrow{\text{a.s.}} E_{\underline{\theta}} H(\underline{X}_1) < \infty$. In an argument similar to that used for J_2 , we obtain for $2\epsilon < \beta$,

$$P_{\underline{\theta}} \{ \sqrt{n} ||\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0||^{1+\beta} > n^{-\epsilon} \} \leq n^{-k\epsilon} E_{\underline{\theta}} \{ \sqrt{n} ||\hat{\underline{\theta}}_{\underline{n}} - \underline{\theta}_0||^{(1+\beta)k} \}.$$

On choosing k so that $k > \epsilon^{-1}$, we get, by the Borel-Cantelli lemma, $J_3 \stackrel{\text{a.s.}}{=} o(n^{-\epsilon})$, and thus the theorem is proved. $\square$

5. Strong Approximations of the Estimated Multivariate Empirical Process

Let $\underline{X}_j = (X_{j1}, \dots, X_{jd})$ ($j = 1, 2, \dots$) be an i.i.d. sequence of d -dimensional random vectors with an absolutely continuous d.f. $F(\underline{x}, \underline{\theta})$, where $\underline{\theta}$ is a p -dimensional vector of unknown parameters. Suppose $\underline{\theta}$ is estimated by $\hat{\underline{\theta}}_{\underline{n}}$, a sequence of maximum likelihood estimators based on the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{\underline{n}}$. We shall assume that the (A1) assumptions of Section 4 hold and shall also consider:

(A2) (i) The vectors $\frac{\partial}{\partial \underline{\theta}} F(\underline{x}; \underline{\theta})$ and $\frac{\partial^2}{\partial \underline{\theta}^2} F(\underline{x}; \underline{\theta})$ of partial derivatives exist and are bounded on $R^d \times \nu$.

(ii) $l'_{\underline{\theta}}(\underline{x}, \underline{\theta}) = \frac{\partial}{\partial \underline{\theta}} \log f(\underline{x}; \underline{\theta})$ is bounded on $R^d \times \nu$ and is absolutely continuous with respect to $\underline{x}$.

(iii) The partial derivatives $\frac{\partial^s}{\partial x_{i(1)} \dots \partial x_{i(s)}} l'_{\underline{\theta}}(\underline{x}, \underline{\theta})$, with respect to the components of $\underline{x}$, are bounded functions of $\underline{x}$, uniformly in $\underline{\theta} \in \nu$, where

$1 \leq i(1) < i(2) < \dots < i(s) \leq d$. Note that when

$d = 1$, the derivative above is just $\frac{d}{dx} l'_{\underline{\theta}}(\underline{x}, \underline{\theta})$.

Let $F_n(\underline{x})$ be the empirical d.f. based on the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ and let

$$(5.1) \quad \hat{E}_n(\underline{x}) = \sqrt{n}[F_n(\underline{x}) - F(\underline{x}; \hat{\theta}_n)] .$$

Define

$$(5.2) \quad G = G(TD_{\underline{x}, n}; \underline{\theta}_0) = K(TD_{\underline{x}, n}) - \left\{ \int_{\underline{x} \in R^d} l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0) dK(TD_{\underline{x}, n}) \right\}^t I(\underline{\theta}_0)^{-1} .$$

$$\frac{\partial}{\partial \underline{\theta}} F(\underline{x}, \underline{\theta}_0) ,$$

where K is the Kiefer Process of Theorem A, $TD_{\underline{x}}$ is defined by

(1.8) with $F = F(\underline{x}; \underline{\theta}_0)$ and $I(\underline{\theta})$ is the Fisher information matrix.

Also, let

$$(5.3) \quad Z_n = Z_n(TD_{\underline{x}}; \underline{\theta}_0) \\ = B_n(TD_{\underline{x}}) - \left\{ \int_{\underline{x} \in R^d} l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0) d_{\underline{x}} B_n(TD_{\underline{x}}) \right\}^t \cdot I(\underline{\theta}_0)^{-1} \frac{\partial}{\partial \underline{\theta}} F(\underline{x}, \underline{\theta}_0)$$

where B_n is the sequence of B.B.'s of Theorem A. We obtain

Theorem 5.1. Suppose that the conditions of Theorem 1B, (A1) of Section 3, and (A2) are fulfilled. Then, one can construct a Kiefer process $K(\underline{s}, n)$ and a sequence $B_n(\underline{s})$ of Brownian Bridges such that

$$(5.4) \quad \sup_{\underline{x}} |\sqrt{n} \hat{E}_n(\underline{x}) - G(TD_{\underline{x}}, n; \underline{\theta}_0)| \stackrel{a.s.}{=} O(n^\gamma \log^2 n),$$

for some γ , $0 < \gamma < 1/2$, where G is defined by (5.2), $\hat{E}_n$ is defined by (5.1), and

$$(5.5) \quad \sup_{\underline{x}} |\hat{E}_n(\underline{x}) - Z_n(TD_{\underline{x}}, \underline{\theta}_0)| \stackrel{a.s.}{=} O(n^{-\epsilon} (\log n)^{3/2}),$$

for some $\epsilon > 0$, where Z_n is defined by (5.3).

Before we give the proof of Theorem 5.1, let us define

$$\epsilon_n(\underline{x}) = \sqrt{n}(F_n(\underline{x}) - F(\underline{x}; \underline{\theta}_0)) - B_n(TD_{\underline{x}})$$

and

$$\delta_n(\underline{x}) = n[F_n(\underline{x}) - F(\underline{x}, \underline{\theta}_0)] - K(TD_{\underline{x}}, n),$$

where $B_n(\underline{t})$ is the sequence of Brownian Bridges and $K(\underline{t}, n)$ is

the Kiefer process of Theorem 1B. Then,

$$\sup_{\underline{x}} |\varepsilon_n(\underline{x})| \stackrel{\text{a.s.}}{=} O\left(n^{-\frac{1}{2(d+1)}} (\log n)^{3/2}\right)$$

and

$$\sup_{\underline{x}} |\delta_n(\underline{x})| \stackrel{\text{a.s.}}{=} O\left(n^{\frac{d+1}{2(d+2)}} \log^2 n\right).$$

We obtain

Lemma 5.1. Under the conditions of Theorem 5.1,

$$(5.6) \quad \int l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0) d_{\underline{x}} \varepsilon_n(\underline{x}) \stackrel{\text{a.s.}}{=} O\left(n^{\frac{-1}{(d+1)(d+2)}} (\log n)^{3/2}\right)$$

and

$$(5.7) \quad \int l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0) d_{\underline{x}} \delta_n(\underline{x}) \stackrel{\text{a.s.}}{=} O\left(n^{\frac{d^2 + 3d + 1}{2(d+1)(d+2)}} \log^2 n\right).$$

Proof of Lemma 5.1. Let $N = n^{\frac{1}{2(d+1)(d+2)}}$ and $A_N = [-N, N]^d$.

Then, expanding the vector of stochastic integrals by parts, we obtain

$$\begin{aligned} & \int_{A_N} l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0) d_{\underline{x}} \varepsilon_n(\underline{x}) \\ &= [\dots [[l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0) \cdot \varepsilon_n(\underline{x})]_{x_1=-N}^N]_{x_2=-N}^N \dots]_{x_d=-N}^N \\ &+ \sum_{i=1}^d (j(1), j(2), \dots, j(i)) \int_{-N}^N \dots \int_{-N}^N \int_{-N}^N (-1)^i \\ &\cdot \left[\frac{\partial^i l'_{\underline{\theta}}(\underline{x}, \underline{\theta}_0)}{\partial x_{j(1)} \dots \partial x_{j(i)}} \cdot \varepsilon_n(\underline{x}) \right] dx_{j(1)} \dots dx_{j(i)}, \end{aligned}$$

where in the integrand, the components x_s ($x_s \neq x_{j(r)}$, $r=1,2,\dots,i$) are evaluated from $-N$ to N , and where the summation is taken over all possible vectors $(j(1),\dots,j(i))$ of integers for which $1 \leq j(1) < j(2) < \dots < j(i) \leq d$. For example, if $d = 2$,

$$\begin{aligned} \int_{A_N} &= \left[l_{\theta}'(\underline{x}, \underline{\theta}_0) \cdot \epsilon_n(\underline{x}) \right]_{x_1=-N}^N \Big|_{x_2=-N}^N \\ &- \int_{-N}^N \left[\frac{\partial}{\partial x_1} l_{\theta}'(\underline{x}, \underline{\theta}_0) \cdot \epsilon_n(\underline{x}) \right]_{x_2=-N}^N dx_1 \\ &- \int_{-N}^N \left[\frac{\partial}{\partial x_2} l_{\theta}'(\underline{x}, \underline{\theta}_0) \cdot \epsilon_n(\underline{x}) \right]_{x_1=-N}^N dx_2 \\ &+ \int_{-N}^N \int_{-N}^N \frac{\partial^2}{\partial x_1 \partial x_2} l_{\theta}'(\underline{x}, \underline{\theta}_0) \cdot \epsilon_n(\underline{x}) dx_1 dx_2. \end{aligned}$$

Evaluating these integrals, we obtain

$$\begin{aligned} \int_{A_N} \text{a.s.} \sigma \left[n^{\frac{-1}{2(d+1)}} (\log n)^{3/2} \right] &+ \sum_{i=1}^d \sum_{(j(1),\dots,j(i))} (-1)^i \\ &O \left(n^{\frac{-1}{2(d+1)}} (\log n)^{3/2} \cdot N^i \right) \\ \text{a.s.} \sigma \left(n^{\frac{-1}{(d+1)(d+2)}} (\log n)^{3/2} \right). \end{aligned}$$

$$\text{Also, } \int_{A_N} \text{a.s.} \sigma \left(n^{\frac{-3}{2(d+1)(d+2)}} (\log n)^{3/2} \right)$$

by expanding the stochastic integral over appropriate regions whose union is $\bar{A}_N$ (the complement of A_N). For example, over the region $R_{1N} = (-\infty, -N) \times [-N, N]^{d-1}$,

$$\int_{R_{1N}} a.s. O\left(n^{\frac{-3}{2(d+1)(d+2)}} (\log n)^{3/2}\right), \text{ while, over } R_{2N} = (-\infty, -N)^2 \times [-N, N]^{d-2},$$

$$\int_{R_{2N}} a.s. O\left(n^{\frac{-2}{(d+1)(d+2)}} (\log n)^{3/2}\right). \text{ Thus, (5.6) is proved. By}$$

a similar calculation, (5.7) is proved. □

Proof of Theorem 5.1. Using Theorem 1B and the Taylor expansion of $F(\underline{x}, \underline{\theta})$, we obtain

$$\begin{aligned} \sqrt{n} \hat{E}_n(\underline{x}) &= n[F_n(\underline{x}) - F(\underline{x}; \hat{\underline{\theta}}_n)] \\ &= n[F_n(\underline{x}) - F(\underline{x}; \underline{\theta}_0)] - n[F(\underline{x}; \hat{\underline{\theta}}_n) - F(\underline{x}; \underline{\theta}_0)] \\ &= K(TD_{\underline{x}}, n) + \delta_n(\underline{x}) \\ &= n(\hat{\underline{\theta}}_n - \underline{\theta}_0)^t \cdot \frac{\partial}{\partial \underline{\theta}} F(\underline{x}; \underline{\theta}_0) - \frac{n}{2} [(\hat{\underline{\theta}}_n - \underline{\theta}_0)^2]^t \cdot \frac{\partial^2}{\partial \underline{\theta}^2} F(\underline{x}; \underline{\theta}_0) \end{aligned}$$

where $\|\underline{\theta}_0 - \underline{\theta}'_n\| < \|\underline{\theta}_0 - \hat{\underline{\theta}}_n\|$. By Theorem 4.1, $n(\hat{\underline{\theta}}_n - \underline{\theta}_0)^2$ a.s. $O(n^\beta)$, for any $\beta > 0$. So $\sqrt{n} \hat{E}_n(\underline{x}) = K(TD_{\underline{x}}, n) + n(\hat{\underline{\theta}}_n - \underline{\theta}_0)^t$.

$$\frac{\partial}{\partial \underline{\theta}} F(\underline{x}; \underline{\theta}_0) + \delta'_n(\underline{x}), \text{ where } \sup_{\underline{x}} |\delta'_n(\underline{x})| \text{ a.s. } O\left(n^{\frac{d+1}{2(d+2)}} \log^2 n\right).$$

$$\text{By Theorem 4.1, } \sqrt{n} \hat{E}_n(\underline{x}) = K(TD_{\underline{x}}, n) - \sum_{j=1}^n \ell'_{\underline{\theta}}(\underline{x}_j, \underline{\theta}_0) \cdot I(\underline{\theta}_0)^{-1} \cdot \frac{\partial}{\partial \underline{\theta}} F(\underline{x}, \underline{\theta}_0) + \delta''_n(\underline{x}).$$

where for the ϵ of Theorem 4.1, if $\gamma = \max \left(\frac{d+1}{2(d+2)}, \frac{1}{2} - \epsilon \right)$,

then $0 < \gamma < \frac{1}{2}$ and $\sup_{\underline{x}} |\delta_n''(\underline{x})| \stackrel{a.s.}{=} O(n^\gamma \log^2 n)$.

$$\text{Now, } \sum_{j=1}^n \ell_{\underline{\theta}}'(X_j, \underline{\theta}_0) = \int \ell_{\underline{\theta}}'(\underline{x}, \underline{\theta}_0) d_{\underline{x}} n F_n(\underline{x})$$

$$= \int \ell_{\underline{\theta}}'(\underline{x}, \underline{\theta}_0) d_{\underline{x}} n [F_n(\underline{x}) - F(\underline{x}; \underline{\theta}_0)] , \text{ by (A1) (vii) ,}$$

$$= \int \ell_{\underline{\theta}}'(\underline{x}, \underline{\theta}_0) d_{\underline{x}} K(TD_{\underline{x}}, n) - \int \ell_{\underline{\theta}}'(\underline{x}, \underline{\theta}_0) d_{\underline{x}} \delta_n(\underline{x}) .$$

Thus, by Lemma 5.1, statement (5.4) is proved. Statement (5.5) is proved in the similar way. $\square$

Remark 5.1. Note that the process $P(n) = \left\{ \int \ell_{\underline{\theta}}'(\underline{x}; \underline{\theta}_0) d_{\underline{x}} K(TD_{\underline{x}}, n) \right\}^t I(\underline{\theta}_0)^{-1}$ is a vector-valued Gaussian process with mean zero and covariance $E P(n_1) P(n_2)^t = \min(n_1, n_2) \cdot I(\underline{\theta}_0)^{-1}$. Thus, G , defined by (5.2), has covariance function given by

$$\begin{aligned} (5.8) \quad & E[G(TD_{\underline{x}}, n_1, \underline{\theta}_0) G(TD_{\underline{y}}, n_2, \underline{\theta}_0)] \\ &= \min(n_1, n_2) \cdot \{ F(\underline{x} \wedge \underline{y}; \underline{\theta}_0) - F(\underline{x}, \underline{\theta}_0) F(\underline{y}, \underline{\theta}_0) \\ &\quad - \frac{\partial}{\partial \underline{\theta}} F(\underline{x}; \underline{\theta}_0)^t \cdot I(\underline{\theta}_0)^{-1} \cdot \frac{\partial}{\partial \underline{\theta}} F(\underline{y}, \underline{\theta}_0) \} , \end{aligned}$$

where $\underline{x} \wedge \underline{y} = (\min(x_1, y_1), \min(x_2, y_2), \dots, \min(x_d, y_d))$.

The covariance function for Z_n is given by (5.8) with the factor $\min(n_1, n_2)$ removed.

The process G defined by (5.2) depends not only on F , but also on θ_0 , the true value of θ . So, in general, (5.4) cannot be used to test the composite hypothesis

$$H : F = F(\underline{x}, \theta) .$$

To give an approximate solution to this problem, let

$$TD_{\underline{x}} = \{(b_1, b_2, \dots, b_d) \in [0,1]^d : b_1 \leq F_1(x_1 | x_1, x_2, \dots, x_{i-1}; \hat{\theta}_n)\}$$

and define $\hat{G} = G(TD_{\underline{x}}, n, \theta_0)$ by (5.2) but with θ_0 replaced by $\hat{\theta}_n$ everywhere it appears (and hence $TD_{\underline{x}}$ is replaced by $\hat{TD}_{\underline{x}}$).

We obtain

Theorem 5.2. Assuming the conditions of Theorem 5.1 and that the $(p \times p)$ -dimensional matrix $I(\theta)$ is uniformly continuous in $\theta \in \nu$, we have

$$\sup_{\underline{x}} |G(TD_{\underline{x}}, n; \theta_0) - G(\hat{TD}_{\underline{x}}, n; \hat{\theta}_n)| \stackrel{a.s.}{=} o(1)$$

and hence

$$\sup_{\underline{x}} |\sqrt{n} \hat{E}_n(\underline{x}) - G(\hat{TD}_{\underline{x}}, n; \hat{\theta}_n)| \stackrel{a.s.}{=} O(n^\gamma) ,$$

for some γ ($0 < \gamma < \frac{1}{2}$).

The result follows from the fact that $\|\hat{\theta}_n - \theta_0\| \stackrel{a.s.}{=} o(1)$ and the continuous sample path properties of the Gaussian processes involved.

Theorem 4.2 states that $\hat{G}$, which does not depend on the unknown true value of θ , is just as good an approximation to $\sqrt{n} \hat{E}_n(\underline{x})$ as G is.

J. Durbin (1973b) has observed that if the m.l.e. of θ is based on a randomly chosen half of the sample, the resulting univariate estimated empirical process converges in distribution to a Brownian Bridge process. This line of thinking was begun by K. C. Rao (1972). We shall prove the result here using our strong approximation techniques. Csörgö, Komlós, Major, Révész and Tusnády ([14], to appear) proved the result in the case when the observations are univariate and θ one-dimensional. We shall extend their proof for multivariate observations and multidimensional θ .

Let $\bar{\theta}_n$ be the m.l.e. of θ based on a randomly chosen half of the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$. Without loss of generality, we can assume it is the first half: $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{[n/2]}$. We have

Theorem 5.3. Under the conditions of Theorem 5.1, one can define a Kiefer process $\bar{K}(\underline{s}, n)$ such that

$$\sup_{\underline{x}} |n[F_n(\underline{x}) - F(\underline{x}; \bar{\theta}_n)] - \bar{K}(T\underline{D}_{\underline{x}}, n)| \stackrel{a.s.}{=} O(n^\gamma),$$

for some γ , $0 < \gamma < \frac{1}{2}$ and consequently,

$$\sqrt{n} [F_n(\underline{x}) - F(\underline{x}; \bar{\theta}_n)] \xrightarrow{D} B(TD_{\underline{x}}),$$

where $B(\underline{s})$ is a Brownian Bridge ($\underline{s} \in I^d$) and $TD_{\underline{x}}$ is defined by (1.8) with $F = F(\underline{x}; \theta_0)$.

Proof. Let $F_n^{(1)}(\underline{x})$, resp. $F_n^{(2)}(\underline{x})$, be the empirical d.f. based on the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{[n/2]}$, resp. $\underline{X}_{[n/2]+1}, \dots, \underline{X}_n$ and let $K_1(\underline{s}, n)$ and $K_2(\underline{s}, n)$ be the Kiefer processes for which

$$\sup_{\underline{x}} \left| \frac{n}{2} [F_n^{(1)}(\underline{x}) - F(\underline{x}; \theta_0)] - K_1(TD_{\underline{x}}, n/2) \right| \stackrel{a.s.}{=} o\{\log^2(n/2)\};$$

$$\sup_{\underline{x}} \left| \frac{n}{2} [F_n^{(2)}(\underline{x}) - F(\underline{x}; \theta_0)] - K_2(TD_{\underline{x}}, n/2) \right| \stackrel{a.s.}{=} o\{\log^2(n/2)\}.$$

Without loss of generality, we can assume that K_2 is independent from K_1 and also from the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{[n/2]}$. Then we have

$$\begin{aligned} n[F_n(\underline{x}) - F(\underline{x}; \bar{\theta}_n)] &= \frac{n}{2}[F_n^{(1)}(\underline{x}) - F(\underline{x}; \theta_0)] + \frac{n}{2}[F_n^{(2)}(\underline{x}) - F(\underline{x}; \theta_0)] \\ &\quad - n[F(\underline{x}; \bar{\theta}_n) - F(\underline{x}; \theta_0)] \\ &= K_1(TD_{\underline{x}}, n/2) + K_2(TD_{\underline{x}}, n/2) - n[F(\underline{x}; \bar{\theta}_n) \\ &\quad - F(\underline{x}; \theta_0)] + R_{1n}(\underline{x}), \end{aligned}$$

where $\sup_{\underline{x}} |R_{1n}(\underline{x})| = o\{\log^2(n/2)\}$.

As in the proof of Theorem 5.1, we obtain

$$\begin{aligned} n[F_n(\underline{x}) - F(\underline{x}; \underline{\theta}_n)] &= K_1(TD_{\underline{x}}, n/2) + K_2(TD_{\underline{x}}, n/2) \\ &- 2\left\{ \int \dot{L}_{\underline{\theta}}'(\underline{x}, \underline{\theta}_0) d_{\underline{x}} K_1(TD_{\underline{x}}, n/2) \right\}^t \cdot I(\underline{\theta}_0)^{-1} \cdot \\ &\frac{\partial}{\partial \underline{\theta}} F(\underline{x}, \underline{\theta}_0) + R_{2n}(\underline{x}), \end{aligned}$$

where $\sup_{\underline{x}} |R_{2n}(\underline{x})| \stackrel{a.s.}{=} O(n^\gamma)$, for some γ , $0 < \gamma < \frac{1}{2}$. The process

$$\begin{aligned} \bar{K}(TD_{\underline{x}}, n) &= K_1(TD_{\underline{x}}, n/2) + K_2(TD_{\underline{x}}, n/2) \\ &- 2\left\{ \int \dot{L}_{\underline{\theta}}'(\underline{x}, \underline{\theta}_0) d_{\underline{x}} K_1(TD_{\underline{x}}, n/2) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} F(\underline{x}; \underline{\theta}_0) \end{aligned}$$

is clearly a Gaussian process with mean zero and, by calculation, covariance

$$E \bar{K}(TD_{\underline{x}}, n) \bar{K}(TD_{\underline{y}}, m) = (n \wedge m) (F(\underline{x} \wedge \underline{y}; \underline{\theta}_0) - F(\underline{x}; \underline{\theta}_0) F(\underline{y}; \underline{\theta}_0)).$$

Hence $\bar{K}$ is a Kiefer process and the theorem is proved. $\square$

Remark 5.2. In the univariate case ($d=1$), the Kolmogorov-Smirnov-type statistics based on the process $\sqrt{n}[F_n(x) - F(x; \underline{\theta}_n)]$ become asymptotically distribution-free. The same is true in the multivariate case if $F(\underline{x}; \underline{\theta}_0)$ is a product d.f. (i.e. $F(\underline{x}; \underline{\theta}_0)$ is the product of its marginals).

As observed by Durbin (1973b), the asymptotic distribution-freeness of $\sqrt{n}[F_n(x) - F(x; \bar{\theta}_n)]$ in the univariate case thus gained is somewhat illusory. A randomization has been introduced; the value of $\bar{\theta}_n$ depends on the particular half-sample chosen. With this randomization, the empirical process discussed in Theorem 5.3 behaves as if the unknown parameters were completely specified.

Remark 5.3. There has been some work done on the computation of statistics based on the empirical process discussed in Theorem 5.1. In particular, we refer to Kac, Kiefer and Wolfowitz (1955), Durbin (1975) and to Stephens (1974, 1976).

6. Strong Approximations of the Estimated Quantile Process.

Let $X_1, X_2, \dots$ be an i.i.d. sequence of univariate random variables with an absolutely continuous d.f. $F(x; \theta)$, where θ is a p -dimensional vector of unknown parameters. Again, let θ be estimated by $\hat{\theta}_n$, a sequence of m.l.e. based on $X_1, X_2, \dots, X_n$.

Let $Q_n(s)$ be the quantile function based on $X_1, X_2, \dots, X_n$, defined by (2.1), and let

$$(6.1) \quad \hat{q}_n(s) = \sqrt{n}[Q_n(s) - F^{-1}(s; \hat{\theta}_n)] , \text{ for } s \in (0, 1) ,$$

where $F^{-1}(s; \theta)$ is the inverse function of $F(x; \theta)$. Consider the following assumption:

(A3) The vectors $\frac{\partial}{\partial \underline{\theta}} F^{-1}(s; \underline{\theta})$ and $\frac{\partial^2}{\partial \underline{\theta}^2} F^{-1}(s; \underline{\theta})$ of partial derivatives exist and are bounded on $(0, 1) \times v$.

Let $M = M(s, n; \underline{\theta}_0)$ be the Gaussian process defined by

$$(6.2) \quad M(s, n; \underline{\theta}_0) =$$

$$\frac{K(s, n)}{f[F^{-1}(s; \underline{\theta}_0); \underline{\theta}_0]} - \left\{ \int_{x \in R} \underline{l}'_{\underline{\theta}}(x, \underline{\theta}_0) d_x K(F(x; \underline{\theta}_0), n) \right\}^t I(\underline{\theta}_0)^{-1} \cdot \frac{\partial}{\partial \underline{\theta}} F^{-1}(s; \underline{\theta}_0)$$

where $K(s, n)$ is the Kiefer process of Theorem 2B. We obtain

Theorem 6.1. Let assumptions (A1), (A3) and (A2) (ii) and (iii), with $d = 1$, be satisfied. Then, under the conditions of Theorem 2B,

$$\sup_{0 < s < 1} |\sqrt{n} \hat{q}_n(s) - M(s, n; \underline{\theta}_0)| \stackrel{a.s.}{=} O(n^\gamma),$$

for some γ ($0 < \gamma < \frac{1}{2}$),

where $M(s, n; \underline{\theta}_0)$ is defined by (6.2).

We shall first state

Lemma 6.1. Assume that the conditions of Theorem 6.1 are satisfied.

Let $\epsilon_n(x) = \sqrt{n} E_n(x) - K(F(x); n)$, where E_n and K are defined as in Theorem 1A. Then,

$$\int \underline{l}'_{\underline{\theta}}(x, \underline{\theta}_0) d_x \epsilon_n(x) \stackrel{a.s.}{=} O(n^\gamma),$$

for some γ' ($0 < \gamma' < \frac{1}{2}$).

The proof of this lemma is the same as that for Lemma 5.1 with $d = 1$.

Proof of Theorem 6.1. As in the proof of Theorem 5.1, using Theorems 1A and 2B,

$$\begin{aligned} \sqrt{n} \hat{q}_n(t) &= n[Q_n(s) - F^{-1}(s; \underline{\theta}_0)] - n[F^{-1}(s; \hat{\underline{\theta}}_n) - F^{-1}(s; \underline{\theta}_0)] \\ &= \sqrt{n} q_n(s) - n(\hat{\underline{\theta}}_n - \underline{\theta}_0)^t \cdot \frac{\partial}{\partial \underline{\theta}} F^{-1}(s; \underline{\theta}_0) - \\ &\quad \left(\frac{n}{2} (\hat{\underline{\theta}}_n - \underline{\theta}_0)^2 \right)^t \cdot \frac{\partial^2}{\partial \underline{\theta}^2} F^{-1}(s; \underline{\theta}_0') \\ &= \sqrt{n} q_n(s) - \left\{ \int \underline{\ell}'_{\underline{\theta}}(x, \underline{\theta}_0) d_x \sqrt{n} E_n(x) \right\}^t \cdot I(\underline{\theta}_0)^{-1} \cdot \frac{\partial}{\partial \underline{\theta}} F^{-1}(s; \underline{\theta}_0) \\ &\quad - \left[\frac{n}{2} (\hat{\underline{\theta}}_n - \underline{\theta}_0)^2 \right]^t \cdot \frac{\partial^2}{\partial \underline{\theta}^2} F^{-1}(s; \underline{\theta}_0') , \text{ where } ||\underline{\theta}_0' - \underline{\theta}_0|| < ||\hat{\underline{\theta}}_n - \underline{\theta}_0||. \end{aligned}$$

Hence,

$$\begin{aligned} \sqrt{n} \hat{q}_n(s) &= \frac{K(s, n)}{f[F^{-1}(s; \underline{\theta}_0); \underline{\theta}_0]} \\ &\quad - \left\{ \int \underline{\ell}'_{\underline{\theta}}(x, \underline{\theta}_0) d_x K(F(x; \underline{\theta}_0), n) \right\}^t I(\underline{\theta}_0)^{-1} \cdot \\ &\quad \frac{\partial}{\partial \underline{\theta}} F^{-1}(s; \underline{\theta}_0) + \epsilon'_n(s) , \end{aligned}$$

where $\sup_{0 < s < 1} |\epsilon'_n(s)| \stackrel{a.s.}{=} o(n^\gamma)$, with $0 < \gamma < \frac{1}{2}$, and hence the theorem. □

Remark 6.1. As to the covariance function of $M(s, n; \underline{\theta}_0)$ of (6.2), it is similar to that of G , given in Remark 5.1.

As in Section 5, the Gaussian process $M(s, n; \underline{\theta}_0)$ depends not only on F , but also on $\underline{\theta}_0$, the theoretical true value of $\underline{\theta}$. So, in general, Theorem 6.1 cannot be used to test the composite hypothesis

$$H : F = F(x; \underline{\theta}) .$$

A partial solution is the following: let $\hat{M} = \hat{M}(s, n, \hat{\underline{\theta}}_n)$ be defined as in (6.2) but with $\underline{\theta}_0$ replaced by $\hat{\underline{\theta}}_n$ everywhere it appears. Then, we obtain

Theorem 6.2. Under the conditions of Theorem 6.1 and assuming that the $(p \times p)$ -dimensional matrix $I(\underline{\theta})$ is uniformly continuous in $\underline{\theta} \in \nu$, we obtain

$$\sup_{0 < s < 1} |M(s, n; \hat{\underline{\theta}}_n) - M(s, n; \underline{\theta}_0)| = o_p(1) ,$$

and hence

$$\sup_{0 < s < 1} |\sqrt{n} \hat{q}_n(s) - M(s, n, \hat{\underline{\theta}}_n)| = o_p(n^\gamma) ,$$

for some γ ($0 < \gamma < \frac{1}{2}$).

The proof follows from the fact that $\|\hat{\underline{\theta}}_n - \underline{\theta}_0\| \stackrel{a.s.}{=} o(1)$ and the continuity of the Gaussian processes involved.

Remark 6.2. Note that, for each n ,

$$n^{-\frac{1}{2}} M(s, n; \theta_0) \stackrel{D}{=} \frac{B(s)}{f[F^{-1}(s; \theta_0); \theta_0]} - \left\{ \int_0^s l'_{\theta}(x, \theta_0) d_x B(F(x; \theta_0)) \right\}.$$

$$I(\theta_0)^{-1} \cdot \frac{\partial}{\partial \theta} F^{-1}(s; \theta_0)$$

where $B(\cdot)$ is a Brownian Bridge on $(0, 1)$. This follows from the fact that, for each n ,

$$\frac{1}{\sqrt{n}} K(s, n) \stackrel{D}{=} B(s).$$

III WEAK APPROXIMATIONS OF THE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES WHEN PARAMETERS ARE ESTIMATED

7. Introduction.

In Chapter II, our aim was to obtain almost-sure representations of the estimated quantile and multivariate empirical processes in terms of Gaussian processes. The methodology used in Chapter II can accommodate a more general method of estimation (than m.l.e.) with less restrictive conditions. However, the Gaussian processes, thus gained, are in-probability representations. In the present chapter, we shall assume that the estimators of the unknown parameters satisfy the fairly general (A1)-conditions of Durbin (1973a), who studied the weak convergence of the estimated, univariate empirical process. We shall study the estimated quantile and multivariate empirical processes under slightly more general alternative hypotheses than Durbin's (1973a).

In our Section 8, we obtain in-probability representations of the estimated multivariate empirical process, under the general sequence of alternatives defined by (8.3). The corresponding results for the univariate case and for Durbin's (1973a) alternatives are mentioned as special cases.

The maximum likelihood case is discussed in Section 9. In Section 10, the same program (as in Section 8) is carried out for

the estimated (univariate) quantile process.

The results of this chapter also appear in Burke and Csörgö [6] and in Csörgö and Burke [15]. The same notation as in Chapter II will be used.

8. Weak Approximations of the Estimated Multivariate Empirical Process.

Let $\underline{X}_1, \underline{X}_2, \dots$ be a sequence of i.i.d. d -dimensional random vectors with a continuous d.f. F . Consider the family of d.f. given by $\{H(\underline{x}; \underline{\beta}_0, \underline{\theta}) : \underline{\theta} \in \Omega\}$, where $\underline{\beta}_0$ is a p_1 -dimensional vector of specified parameters and $\underline{\theta}$ is a p_2 -dimensional vector of unknown parameters, but known to belong to a subset Ω of R^{p_2} . One is often interested in testing whether the true d.f. F of the i.i.d. sequence $\{\underline{X}_n\}$ belongs to $\{H(\underline{x}; \underline{\beta}_0, \underline{\theta}) : \underline{\theta} \in \Omega\}$, that is, we wish to test the following null hypothesis:

$$(8.1) \quad H_0 : F \in \{H(\cdot; \underline{\beta}_0, \underline{\theta}) : \underline{\theta} \in \Omega\}.$$

Suppose that the vector $\underline{\theta}$ is estimated by $\{\hat{\underline{\theta}}_n\}$, a sequence of estimators based on the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$. Let

$$(8.2) \quad \hat{E}_n(\underline{x}) = \sqrt{n} [F_n(\underline{x}) - H(\underline{x}; \underline{\beta}_0, \hat{\underline{\theta}}_n)]$$

be the estimated empirical process, where $F_n(\underline{x})$ is the empirical d.f. based on $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$. We shall obtain in-probability representations of $\hat{E}_n(\underline{x})$ under the null hypothesis (8.1) and also

under a sequence of alternative hypotheses, given by

$$(8.3) \quad A_n : P \in \{G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}) : \underline{\theta} \in \Omega\},$$

where the family $\{G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}) : \underline{\theta} \in \Omega\}$ has the same parametric structure as $H(\cdot; \underline{\beta}_0, \underline{\theta})$ of (8.1). For example, $\underline{\beta}$ and/or $\underline{\theta}$ could be vectors of location or scale parameters of the d.f. H and G_n . We shall assume that the following conditions are satisfied, under the null hypothesis H_0 of (8.1):

$$(C1) \quad \sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_0) = n^{-\frac{1}{2}} \sum_{j=1}^n l(\underline{x}_j, \underline{\beta}_0, \underline{\theta}_0) + \varepsilon_{on},$$

where $\underline{\theta}_0$ is the true unknown value of $\underline{\theta}$ and that, for a random observation $\underline{x}$,

$$(i) \quad E\{l(\underline{x}, \underline{\beta}_0, \underline{\theta}_0) \mid H_0\} = \underline{0};$$

$$(ii) \quad \varepsilon_{on} \xrightarrow{P} \underline{0} \text{ as } n \rightarrow \infty;$$

$$(iii) \quad L(\underline{\beta}_0) = E\{l(\underline{x}, \underline{\beta}_0, \underline{\theta}_0)l(\underline{x}, \underline{\beta}_0, \underline{\theta}_0)^t \mid H_0\}$$

is a non-negative definite matrix;

$$(iv) \quad l \text{ is a continuous function of } (\underline{\beta}, \underline{\theta})^t \in v, \text{ uniformly in } \underline{x};$$

$$(v) \quad \text{the partial derivatives } \frac{\partial^s}{\partial x_{1(1)} \cdots \partial x_{1(s)}} l(\underline{x}, \underline{\beta}, \underline{\theta}),$$

with respect to the components of $\underline{x}$ are uniformly bounded in $\underline{x} \in R^d$ and $(\underline{\beta}, \underline{\theta})^t \in v$, where v is the closure of a given neighborhood of $(\underline{\beta}_0, \underline{\theta}_0)^t$.

Concerning the sequence $\{A_n\}$ of alternatives defined by (8.3), we will assume that

(C2) (i) $\sqrt{n}(\underline{\beta}_n - \underline{\beta}_0) \rightarrow \gamma$ as $n \rightarrow \infty$, for some p_1 -dimensional vector γ ;

$$(ii) \sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_0) = n^{-1/2} \sum_{j=1}^n \ell_n(\underline{x}_j, \underline{\beta}_n, \underline{\theta}_0) + A \cdot \gamma + \varepsilon_{1n},$$

where the vector-valued functions ℓ_n are measurable and uniformly bounded in n , $\underline{x}$ and $(\underline{\beta}, \underline{\theta})^t \in \nu$, and for a random observation $\underline{x}$,

$$(iii) E\{\ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) \mid A_n\} = \underline{0};$$

(iv) A is a given finite matrix of order $p_2 \times p_1$;

$$(v) \varepsilon_{1n} \xrightarrow{P} \underline{0} \text{ as } n \rightarrow \infty;$$

(vi) the matrix $L_n(\underline{\beta}_n) = E\{\ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) \cdot \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0)^t \mid A_n\}$ is non-negative definite and converges to $L(\underline{\beta}_0)$ as $n \rightarrow \infty$;

(vii) ℓ_n is a continuous function of $(\underline{\beta}, \underline{\theta})^t \in \nu$, uniformly in $\underline{x}$;

(viii) the partial derivatives $\frac{\partial^s}{\partial x_{i(1)} \dots \partial x_{i(s)}} \ell_n(\underline{x}, \underline{\beta}, \underline{\theta})$ are uniformly bounded in $\underline{x} \in R^d$, $(\underline{\beta}, \underline{\theta})^t \in \nu$, and n , where $1 \leq i(1) < i(2) < \dots < i(s) \leq d$;

$$(ix) \sup_{\underline{x}} ||\ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_n) - \ell(\underline{x}, \underline{\beta}_0, \underline{\theta}_0)|| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Remark 8.1. Conditions (C2) (i) to (vi), with $G_n = H$ and $\ell_n = \ell$ for each n , were the (A1)-assumptions studied by Durbin (1973a), in the univariate case.

Maximum likelihood estimators often satisfy conditions (C1) and (C2). In most cases, conditions less restrictive than the (A1)-assumptions of our Section 4 are required. For example, the conditions stated in Cramér (1946, p.500) can be modified to handle alternatives like those of (8.3). The maximum likelihood case is discussed in Section 9.

Concerning the d.f. and the nature of the alternatives involved, we will assume the following:

(C3) (i) the d.f. $H(\underline{x}; \underline{\beta}, \underline{\theta})$ and $G_n(\underline{x}; \underline{\beta}, \underline{\theta})$ satisfy the conditions of Theorem 1B, for each $(\underline{\beta}, \underline{\theta})^t \in \nu$.

(ii) $H(\underline{x}; \underline{\beta}, \underline{\theta})$ and $G_n(\underline{x}; \underline{\beta}, \underline{\theta})$ are continuous in $(\underline{\beta}, \underline{\theta})^t \in \nu$, uniformly in $\underline{x} \in R^d$;

(iii) the vectors of partial derivatives $\frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} H(\underline{x}; \underline{\beta}, \underline{\theta})$ and $\frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} G_n(\underline{x}; \underline{\beta}, \underline{\theta})$ exist, are continuous functions of $(\underline{\beta}, \underline{\theta})^t \in \nu$, uniformly in $\underline{x} \in R^d$, and are uniformly

bounded in $\underline{x} \in R^d$, $(\underline{\beta}, \underline{\theta})^t \in v$ and n ;

(iv) for $\underline{\theta}$ satisfying $(\underline{\beta}_0, \underline{\theta})^t \in v$,

$$\sup_{\underline{x}} |\sqrt{n}[G_n(\underline{x}; \underline{\beta}_0, \underline{\theta}) - H(\underline{x}; \underline{\beta}_0, \underline{\theta})] - w(\underline{x}, \underline{\theta})| \rightarrow 0 \text{ as}$$

$n \rightarrow \infty$, where $w(\underline{x}, \underline{\theta})$ is uniformly bounded and a continuous function of $\underline{\theta}$, $(\underline{\beta}_0, \underline{\theta})^t \in v$, uniformly in $\underline{x} \in R^d$;

$$(v) \sup_{\underline{x}} \left\| \frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} G_n(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) - \frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \right\| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Remark 8.2. In the univariate case, condition (C3) (i) reduces to the requirement that the d.f. H and G_n be continuous.

We obtain

Theorem 8.1. Under the sequence $\{A_n\}$ of alternatives, defined by (8.3), and assuming that conditions (C2) and (C3) are satisfied, one can construct a Gaussian process $\mathcal{J} = \mathcal{J}(\underline{x}, n; G_n, \underline{\beta}_n, \underline{\theta}_0)$ such that

$$(8.4) \sup_{\underline{x} \in R^d} |\hat{E}_n(\underline{x}) - \mathcal{J}(\underline{x}, n; G_n, \underline{\beta}_n, \underline{\theta}_0)| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty,$$

where $\mathcal{J}$ is defined by

$$\begin{aligned}
 (8.5) \quad \mathcal{J} &= \mathcal{J}(\underline{x}, n; G_n, \underline{\beta}_n, \underline{\theta}_0) = n^{-\frac{1}{2}} K(TG_n D_{\underline{x}}, n) \\
 &- \left\{ \int \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) d_{\underline{x}} n^{-\frac{1}{2}} K(TG_n D_{\underline{x}}, n) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} G_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) \\
 &- (A \cdot \gamma)^t \frac{\partial}{\partial \underline{\theta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) + \gamma^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) + w(\underline{x}, \underline{\theta}_0),
 \end{aligned}$$

$K(\cdot, \cdot)$ is the Kiefer process of Theorem 1B, $\hat{E}_n(\underline{x})$ is defined by

(8.2), $TG_n D_{\underline{x}}$ is defined by (1.8) with $F = G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0)$.

Corollary 8.1.1. Assuming the conditions of Theorem 8.1 are satisfied, under $\{A_n\}$,

$$(8.6) \quad \sup_{\underline{x} \in R^d} |\hat{E}_n(\underline{x}) - \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0)| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty,$$

where $\mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0)$ is defined by

$$\begin{aligned}
 (8.7) \quad \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) &= n^{-\frac{1}{2}} K(THD_{\underline{x}}, n) \\
 &- \left\{ \int \ell(\underline{x}, \underline{\beta}_0, \underline{\theta}_0) d_{\underline{x}} n^{-\frac{1}{2}} K(THD_{\underline{x}}, n) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}, \underline{\beta}_0, \underline{\theta}_0) \\
 &- (A \cdot \gamma)^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}, \underline{\beta}_0, \underline{\theta}_0) + \gamma^t \cdot \frac{\partial}{\partial \underline{\beta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \\
 &+ w(\underline{x}; \underline{\theta}_0),
 \end{aligned}$$

$K(\cdot, \cdot)$ is the Kiefer process of Theorem 1B, $THD_{\underline{x}}$ is defined by (1.8) with $F = H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)$, and $\hat{E}_n(\underline{x})$ is defined by (8.2).

Remark 8.3. In the case where $G_n = H$ and $\ell_n = \ell$, that is, under the Durbin (1973a) alternatives, the resultant Gaussian process is defined by (8.7), but with $w(\underline{x}, \underline{\theta}_0)$ equal to zero.

Theorem 8.2. Under the null hypothesis H_0 , defined by (8.1), and assuming that (C1) and (C3) (i) to (iii) concerning H are satisfied, we have

$$\sup_{\underline{x} \in R^d} |\hat{E}_n(\underline{x}) - \mathcal{J}_0(\underline{x}, n)| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty,$$

where $\mathcal{J}_0(\underline{x}, n)$ is defined by (8.7) but with γ and $w(\underline{x}, \underline{\theta}_0)$ equal to zero, and $\hat{E}_n(\underline{x})$ is defined by (8.2).

We shall require

Lemma 8.1. For the Kiefer process $K(\cdot, \cdot)$ of Theorem 1B, let $\epsilon_{2n}(\underline{x}) = \sqrt{n}[F_n(\underline{x}) - G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0)] - n^{-1/2} K(TG_n D_{\underline{x}}, n)$. Then, under the sequence $\{A_n\}$ of alternatives,

$$\int \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) d_{\underline{x}} \epsilon_{2n}(\underline{x}) \xrightarrow{\text{a.s.}} 0 \text{ as } n \rightarrow \infty.$$

The proof of Lemma 8.1 is the same as that for Lemma 5.1.

Proof of Theorem 8.1. Under the sequence $\{A_n\}$ of alternatives, using Theorem 1B and the Taylor expansion of $G_n(\underline{x}; \underline{\beta}, \underline{\theta})$ and $H(\underline{x}; \underline{\beta}, \underline{\theta})$, we obtain

$$\begin{aligned}
 \hat{E}_n(\underline{x}) &= \sqrt{n} [F_n(\underline{x}) - H(\underline{x}; \underline{\beta}_0, \hat{\theta}_n)] \\
 &= \sqrt{n} [F_n(\underline{x}) - G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0)] \\
 &\quad + \sqrt{n} [G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) - G_n(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)] \\
 &\quad + \sqrt{n} [G_n(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) - H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)] \\
 &\quad - \sqrt{n} [H(\underline{x}; \underline{\beta}_0, \hat{\theta}_n) - H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)] \\
 &= n^{-\frac{1}{2}} K(TG_n D_{\underline{x}}, n) + \epsilon_{3n}(\underline{x}) \\
 &\quad + \sqrt{n} (\underline{\beta}_n - \underline{\beta}_0)^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n(\underline{x}; \underline{\beta}_n', \underline{\theta}_0) + w(\underline{x}, \underline{\theta}_0) \\
 &\quad - \sqrt{n} (\hat{\theta}_n - \underline{\theta}_0)^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_n^*) \\
 &= n^{-\frac{1}{2}} K(TG_n D_{\underline{x}}, n) + \gamma^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n(\underline{x}; \underline{\beta}_n', \underline{\theta}_0) + w(\underline{x}, \underline{\theta}_0) \\
 &\quad - \left\{ n^{-\frac{1}{2}} \sum_{j=1}^n \ell_n(\underline{X}_j; \underline{\beta}_n, \underline{\theta}_0) + A \cdot \gamma \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_n^*) + \epsilon_{4n}(\underline{x}),
 \end{aligned}$$

where $\epsilon_{3n}(\underline{x}) = \epsilon_{2n}(\underline{x}) + \sqrt{n} [G_n(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) - H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)] - w(\underline{x}, \underline{\theta}_0)$,

$$\begin{aligned}
 \epsilon_{4n}(\underline{x}) &= \epsilon_{3n}(\underline{x}) + (\sqrt{n}(\underline{\beta}_n - \underline{\beta}_0) - \gamma)^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n(\underline{x}; \underline{\beta}_n', \underline{\theta}_0) \\
 &\quad - \epsilon_{1n}^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_n^*), \text{ and}
 \end{aligned}$$

$$(8.8) \quad ||\underline{\beta}_n' - \underline{\beta}_0|| \leq ||\underline{\beta}_n - \underline{\beta}_0||, \quad ||\underline{\theta}_n^* - \underline{\theta}_0|| \leq ||\hat{\theta}_n - \underline{\theta}_0||.$$

By (C3) (iv) and Theorem 1B, $\sup_{\underline{x}} |\epsilon_{3n}(\underline{x})| \xrightarrow{a.s.} 0$.

By (C2) (i) and the continuity and boundedness properties (C3) (iii) of the partial derivatives,

$$\sup_{\underline{x}} |(\sqrt{n}(\underline{\beta}_n - \underline{\beta}_0) - \gamma)^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Finally, by (C2) (v) and (C3) (iii), we have

$$\sup_{\underline{x}} |\epsilon_{1n}^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_n^*)| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty \text{ and hence}$$

$$\sup_{\underline{x}} |\epsilon_{4n}(\underline{x})| \xrightarrow{P} 0, \text{ as } n \rightarrow \infty.$$

Because of the continuity property (C3) (iii), (C3)(v) and (8.8), we obtain

$$\begin{aligned} \hat{E}_n(\underline{x}) &= n^{-\frac{1}{2}} K(TG_n D_{\underline{x}}, n) + \gamma^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) + w(\underline{x}, \underline{\theta}_0) \\ &\quad - \left\{ n^{-\frac{1}{2}} \sum_{j=1}^n \ell_n(\underline{X}_j, \underline{\beta}_n, \underline{\theta}_0) + A \cdot \gamma \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) + \epsilon_{5n}(\underline{x}), \end{aligned}$$

where $\sup_{\underline{x}} |\epsilon_{5n}(\underline{x})| \xrightarrow{P} 0, \text{ as } n \rightarrow \infty$.

$$\text{Now, since } \sum_{j=1}^n \ell_n(\underline{X}_j, \underline{\beta}_n, \underline{\theta}_0) = \int \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) d_{\underline{x}} n F_n(\underline{x})$$

$$= \int \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) d_{\underline{x}} n [F_n(\underline{x}) - G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0)], \text{ by (C2) (iii),}$$

we obtain under A_n ,

$$\begin{aligned} \hat{E}_n(\underline{x}) &= \mathcal{G}(\underline{x}, n; G_n, \underline{\beta}_n, \underline{\theta}_0) + \epsilon_{5n}(\underline{x}) \\ &+ \left\{ \int \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) d_{\underline{x}} \epsilon_{2n}(\underline{x}) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) . \end{aligned}$$

By Lemma 8.1 and (C3) (iii) ,

$$\sup_{\underline{x}} \left| \left\{ \int \ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) d_{\underline{x}} \epsilon_{2n}(\underline{x}) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} G_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) \right| \xrightarrow{\text{a.s.}} 0 \text{ as}$$

$n \rightarrow \infty$ and hence (8.4) is proved. It is clear from (8.5) that

$\mathcal{G}(\underline{x}, n; G_n, \underline{\beta}_n, \underline{\theta}_0)$ is a Gaussian process. □

Proof of Corollary 8.1.1. Since $\underline{\beta}_n \rightarrow \underline{\beta}_0$, as $n \rightarrow \infty$, by the continuity and convergence conditions (C3) , by (C2) (ix) , and finally by the continuous sample path properties of the Kiefer process, the result follows. □

The proof of Theorem 8.2 is similar to that of Theorem 8.1.

In fact, Theorem 8.2 can be considered a special case of Theorem 8.1 ($G_n = H$, $\ell_n = \ell$ and $\underline{\beta}_n = \underline{\beta}_0$).

Remark 8.4. The Gaussian process $\mathcal{G}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0)$, defined by (8.7) , has mean

$$\begin{aligned} E \mathcal{G}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) &= \gamma^t \cdot \frac{\partial}{\partial \underline{\beta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) + w(\underline{x}, \underline{\theta}_0) \\ &- (A \cdot \gamma)^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) . \end{aligned}$$

and covariance function

$$\begin{aligned} E \mathcal{Y}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) \cdot \mathcal{Y}(\underline{y}, m; H, \underline{\beta}_0, \underline{\theta}_0) \\ = ((n \wedge m) / \sqrt{nm}) \cdot \{H(\underline{x} \wedge \underline{y}; \underline{\beta}_0, \underline{\theta}_0) - H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)H(\underline{y}; \underline{\beta}_0, \underline{\theta}_0) \\ - J(\underline{x})^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{y}; \underline{\beta}_0, \underline{\theta}_0) - J(\underline{y})^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \\ + \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)^t \cdot L(\underline{\theta}_0) \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{y}; \underline{\beta}_0, \underline{\theta}_0)\}, \end{aligned}$$

$$\text{where } J(\underline{x}) = \int_{D_{\underline{x}}} l(\underline{z}, \underline{\beta}_0, \underline{\theta}_0) dH(\underline{z}, \underline{\beta}_0, \underline{\theta}_0).$$

We have, for each n , $\mathcal{Y}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) \stackrel{D}{=} \mathcal{Y}(\underline{x}, 1; H, \underline{\beta}_0, \underline{\theta}_0)$
and since $K(\underline{s}, 1) \stackrel{L}{=} B(\underline{s})$, a Brownian Bridge, it follows that
 $K(\text{THD}_{\underline{x}}, 1) \stackrel{D}{=} B(\text{THD}_{\underline{x}})$ and hence under $\{A_n\}$,

$$\begin{aligned} \hat{E}_n(\underline{x}) \stackrel{D}{\rightarrow} B(\text{THD}_{\underline{x}}) - \left\{ \int l(\underline{x}, \underline{\beta}_0, \underline{\theta}_0) d_{\underline{x}} B(\text{THD}_{\underline{x}}) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \\ - (A \cdot \gamma)^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) + \gamma^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) + w(\underline{x}, \underline{\theta}_0). \end{aligned}$$

It is clear that the Gaussian processes defined by (8.5) and (8.7) depend on $\underline{\theta}_0$; the true unknown value of $\underline{\theta}$. Thus, in general, Theorem 8.2 and Corollary 8.1.1 cannot be applied to test the composite hypotheses (8.1) and (8.3). However, we can remedy the situation as follows: define

$$\hat{\mathcal{Y}} = \mathcal{Y}(\underline{x}, n; H, \underline{\beta}_0, \hat{\underline{\theta}}_n)$$

by (8.7) , but with $\underline{\theta}_0$ replaced by $\hat{\underline{\theta}}_n$ everywhere it appears, (hence $\text{THD}_{\underline{x}}$ becomes $\hat{\text{THD}}_{\underline{x}}$, the set defined by (1.8) with $F = H(\underline{x}; \underline{\beta}_0, \hat{\underline{\theta}}_n)$. Then, we have

Theorem 8.3. Under the conditions of Theorem 8.1,

$$\sup_{\underline{x}} | \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) - \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \hat{\underline{\theta}}_n) | \xrightarrow{P} 0 ,$$

as $n \rightarrow \infty$, and hence under $\{A_n\}$,

$$\sup_{\underline{x}} | \hat{E}_n(\underline{x}) - \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \hat{\underline{\theta}}_n) | \xrightarrow{P} 0 ,$$

as $n \rightarrow \infty$, where $\hat{E}_n(\underline{x})$ is defined by (8.2) .

If we assume that the conditions of Theorem 8.2 are satisfied, then under the null hypothesis H_0 ,

$$\sup_{\underline{x}} | \hat{E}_n(\underline{x}) - \hat{\mathcal{J}}(\underline{x}, n) | \xrightarrow{P} 0 ,$$

as $n \rightarrow \infty$, where $\hat{\mathcal{J}}(\underline{x}, n)$ is defined by (8.7) , with $\underline{\theta}_0$ replaced by $\hat{\underline{\theta}}_n$ and with γ and $w(\underline{x}, \underline{\theta}_0)$ equal to zero.

The proof of Theorem 8.3 follows from $\hat{\underline{\theta}}_n \xrightarrow{P} \underline{\theta}_0$ and the continuity properties (uniformly in $\underline{x}$) of the functions and the Kiefer process involved in $\mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta})$.

Remark 8.5. Theorem 8.3 states that $\hat{\theta}$, which does not depend on θ_0 , the true unknown value of θ , is as good an approximation to $\hat{E}_n(\underline{x})$ as $\hat{\theta}$ is.

Similar statements to our Theorems 8.1, 8.2 and 8.3 and our Corollary 8.1.1 can be made using the sequence of Brownian Bridges of Theorem 1B, instead of the Kiefer process. However, our approach gives one the joint distribution of $\hat{\theta}$ in $\underline{x}$ and n , while a corresponding construction in terms of a Brownian Bridge sequence would give one a Gaussian process in $\underline{x}$, for each n .

9. The Maximum Likelihood Estimation Case.

As mentioned in Remark 8.1, maximum likelihood estimators often satisfy conditions (C1) and (C2). In this case, we have

$$\ell(\underline{x}, \underline{\beta}_0, \underline{\theta}_0) = \frac{\partial}{\partial \underline{\theta}} \log h(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \cdot I(\underline{\theta}_0; \underline{\beta}_0)^{-1}$$

and

$$\ell_n(\underline{x}, \underline{\beta}_n, \underline{\theta}_0) = \frac{\partial}{\partial \underline{\theta}} \log g_n(\underline{x}; \underline{\beta}_n, \underline{\theta}_0) \cdot I_n(\underline{\theta}_0; \underline{\beta}_n)^{-1},$$

where h and g_n are the density functions of H and G_n , respectively. $I(\underline{\theta}; \underline{\beta}_0)$, respectively $I_n(\underline{\theta}; \underline{\beta}_n)$, is the Fisher information matrix, with respect to h , respectively g_n , and $\underline{\theta}$.

To be specific, $I(\underline{\theta}; \underline{\beta}_0) = E\left\{\frac{\partial}{\partial \underline{\theta}} \log h(\underline{X}_1; \underline{\beta}_0, \underline{\theta}) \cdot \frac{\partial}{\partial \underline{\theta}} \log h(\underline{X}_1; \underline{\beta}_0, \underline{\theta})^t | H_0\right\}$,

and $I_n(\underline{\theta}, \underline{\beta}_n) = E\left\{\frac{\partial}{\partial \underline{\theta}} \log g_n(\underline{X}_1; \underline{\beta}_n, \underline{\theta}) \cdot \frac{\partial}{\partial \underline{\theta}} \log g_n(\underline{X}_1; \underline{\beta}_n, \underline{\theta})^t | A_n\right\}$.

Further, let $I_{21} = E\left\{\frac{\partial}{\partial \underline{\theta}} \log h(\underline{X}_1; \underline{\beta}_0, \underline{\theta}_0) \cdot \frac{\partial}{\partial \underline{\beta}} \log h(\underline{X}_1; \underline{\beta}_0, \underline{\theta}_0)^t | H_0\right\}$,

then, as in Durbin (1973a) , $L_n(\underline{\beta}_n) = L_n(\underline{\theta}_0; \underline{\beta}_n)$, $L(\underline{\beta}_0) = I(\underline{\theta}_0; \underline{\beta}_0)$, and $A = I(\underline{\theta}_0; \underline{\beta}_0)^{-1} \cdot I_{21}$.

In this maximum likelihood case, the Gaussian process $\mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0)$, defined by (8.7) with the above substitutions, has mean

$$E \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) = \gamma^t \cdot \frac{\partial}{\partial \underline{\beta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) + w(\underline{x}; \underline{\theta}_0) - (I(\underline{\theta}_0; \underline{\beta}_0)^{-1} \cdot I_{21} \cdot \gamma)^t \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)$$

and the covariance function simplifies to

$$E \mathcal{J}(\underline{x}, n; H, \underline{\beta}_0, \underline{\theta}_0) \cdot \mathcal{J}(\underline{y}, m; H, \underline{\beta}_0, \underline{\theta}_0) = ((n \wedge m) / \sqrt{nm}) \cdot \{H(\underline{x} \wedge \underline{y}; \underline{\beta}_0, \underline{\theta}_0) - H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \cdot H(\underline{y}; \underline{\beta}_0, \underline{\theta}_0) - \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)^t \cdot I(\underline{\theta}_0; \underline{\beta}_0)^{-1} \cdot \frac{\partial}{\partial \underline{\theta}} H(\underline{y}; \underline{\beta}_0, \underline{\theta}_0)\} ,$$

since $J(\underline{x}) = \frac{\partial}{\partial \underline{\theta}} H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0) \cdot I(\underline{\theta}_0; \underline{\beta}_0)^{-1}$, if we assume that

$$\frac{\partial}{\partial \underline{\theta}} \int h(\underline{x}; \underline{\beta}, \underline{\theta}) d\underline{x} = \int \frac{\partial}{\partial \underline{\theta}} h(\underline{x}; \underline{\beta}, \underline{\theta}) d\underline{x} , \text{ for } (\underline{\beta}, \underline{\theta})^t \in \mathcal{V} .$$

Now, suppose that $\bar{\underline{\theta}}_n$ is the maximum likelihood estimator of $\underline{\theta}$, based on a randomly chosen half of the sample $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$. Then, we obtain

Theorem 9.1. Suppose that the conditions (C1) and (C3) are satisfied in this maximum likelihood estimator context. Then under the null hypothesis H_0 of (8.1), one can define a Kiefer process $\bar{K}(s, n)$ such that

$$\sup_{\underline{x}} |\sqrt{n} [F_n(\underline{x}) - H(\underline{x}; \underline{\beta}_0, \underline{\theta}_n)] - n^{-1/2} \bar{K}(\text{THD}_{\underline{x}}, n)| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, where $\text{THD}_{\underline{x}}$ is defined by (1.8) with $F = H(\underline{x}; \underline{\beta}_0, \underline{\theta}_0)$.

The proof (and the construction of $\bar{K}$) is the same as that for Theorem 5.3. Remarks 5.2 and 5.3 are valid in this context.

10. Weak Approximations of the Estimated Quantile Process.

In this section, suppose that $X_1, X_2, \dots$ is a sequence of i.i.d. univariate ($d=1$) random variables with continuous d.f. F . Again, we wish to test the null hypothesis H_0 of (8.1) against the sequence $\{A_n\}$ of alternatives, defined by (8.3). Let

$$(10.1) \quad \hat{q}_n(s) = \sqrt{n} [Q_n(s) - H^{-1}(s; \underline{\beta}_0, \hat{\underline{\theta}}_n)] \quad , \quad 0 < s < 1,$$

where $\hat{\underline{\theta}}_n$ is a sequence of estimators of the unknown vector $\underline{\theta}$, H^{-1} is the inverse function of H , and $Q_n(s)$ is the quantile function based on the sample $X_1, X_2, \dots, X_n$, defined by (2.1).

We shall study the asymptotic behavior of the estimated quantile process $\hat{q}_n(s)$, defined by (10.1), under H_0 and under the sequence of alternatives A_n . Under H_0 , we shall assume that

conditions (C1) are valid with x univariate, while under A_n , we shall assume that (C2) are valid (with x univariate). Concerning the d.f. and the nature of the alternatives involved, we shall assume the following:

(C4) (i) the d.f. $H(x, \underline{\beta}, \underline{\theta})$ and $G_n(x; \underline{\beta}, \underline{\theta})$ satisfy the conditions of Theorem 2B, for each $(\underline{\beta}, \underline{\theta})^t \in v$;

(ii) $H(x; \underline{\beta}, \underline{\theta})$ and $G_n(x; \underline{\beta}, \underline{\theta})$ are continuous in $(\underline{\beta}, \underline{\theta})^t \in v$, uniformly in x ;

(iii) the densities $h(x; \underline{\beta}, \underline{\theta})$ and $g_n(x; \underline{\beta}, \underline{\theta})$ are continuous and bounded, uniformly in $(\underline{\beta}, \underline{\theta})^t \in v$ and x ;

(iv) the vectors $\frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} H^{-1}(s; \underline{\beta}, \underline{\theta})$ and $\frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} G_n^{-1}(s; \underline{\beta}, \underline{\theta})$ exist, are uniformly bounded in $(\underline{\beta}, \underline{\theta})^t \in v$, $s \in (0,1)$ and n , and are continuous in $(\underline{\beta}, \underline{\theta})^t \in v$, uniformly in s ;

(v) for $\underline{\theta}$, with $(\underline{\beta}_0, \underline{\theta})^t \in v$,

$$\sup_{0 < s < 1} |\sqrt{n}[G_n^{-1}(s; \underline{\beta}_0, \underline{\theta}) - H^{-1}(s; \underline{\beta}_0, \underline{\theta})] - u(s, \underline{\theta})| \rightarrow 0$$

as $n \rightarrow \infty$, where $u(s, \underline{\theta})$ is uniformly bounded in s and $\underline{\theta}$, and $-u$ is a continuous function of $\underline{\theta}$, uniformly in $s \in (0,1)$;

$$(vi) \sup_{0 < s < 1} \left| \left| \frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} G_n^{-1}(s; \underline{\beta}_0, \underline{\theta}_0) - \frac{\partial}{\partial(\underline{\beta}, \underline{\theta})^t} H^{-1}(s; \underline{\beta}_0, \underline{\theta}_0) \right| \right| \rightarrow 0$$

and

$$\sup_x |g_n(x; \underline{\beta}_0, \underline{\theta}_0) - h(x; \underline{\beta}_0, \underline{\theta}_0)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

We obtain

Theorem 10.1. Suppose that conditions (C2) and (C4) are satisfied.

Then, under the sequence $\{A_n\}$ of alternatives defined by (8.3),

$$(10.2) \sup_{0 < s < 1} |\hat{q}_n(s) - \eta(s, n; G_n, \underline{\beta}_n, \underline{\theta}_n)| \xrightarrow{P} 0, \text{ as } n \rightarrow \infty,$$

where η is the Gaussian process defined by

$$\begin{aligned} \eta(s, n; G_n, \underline{\beta}_n, \underline{\theta}_n) &= n^{-\frac{1}{2}} K(s, n) / g_n[G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_n); \underline{\beta}_n, \underline{\theta}_n] \\ &\quad - \left\{ \int \ell_n(x, \underline{\beta}_n, \underline{\theta}_n) d_x n^{-\frac{1}{2}} K(G_n(x; \underline{\beta}_n, \underline{\theta}_n), n) \right\}^t \cdot \\ (10.3) \quad &\cdot \frac{\partial}{\partial \underline{\theta}} G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_n) - (A \cdot \gamma)^t \cdot \frac{\partial}{\partial \underline{\theta}} G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_n) \\ &\quad + \gamma^t \cdot \frac{\partial}{\partial \underline{\theta}} G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_n) + u(s, \underline{\theta}_n), \end{aligned}$$

and where K is the Kiefer process of Theorem 2B and $\hat{q}_n(s)$ is defined by (10.1).

Corollary 10.1.1. Under the conditions of Theorem 10.1,

$$\sup_{0 < s < 1} |\eta(s, n; G_n, \underline{\beta}_n, \underline{\theta}_n) - \eta(s, n; H, \underline{\beta}_0, \underline{\theta}_0)| \xrightarrow{P} 0$$

and hence

$$\sup_{0 < s < 1} |\hat{q}_n(s) - \eta(s, n; H, \underline{\beta}_0, \underline{\theta}_0)| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, where $\eta(s, n; H, \underline{\beta}_0, \underline{\theta}_0)$ is defined by (10.3) but with G_n, G_n^{-1} and g_n replaced by H, H^{-1} and h , and $\underline{\beta}_n$ replaced by $\underline{\beta}_0$.

Theorem 10.2. Suppose that conditions (C1) and (C4) are satisfied.

Then under the null hypothesis H_0 of (8.1),

$$\sup_{0 < s < 1} |\hat{q}_n(s) - \eta_0(s, n)| \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty,$$

where $\eta_0(s, n) = n^{-\frac{1}{2}} K(s, n) / h[H^{-1}(s; \underline{\beta}_0, \underline{\theta}_0); \underline{\beta}_0, \underline{\theta}_0]$

$$- \left\{ \int \lambda(x, \underline{\beta}_0, \underline{\theta}_0) d_x n^{-\frac{1}{2}} K(H(x; \underline{\beta}_0, \underline{\theta}_0), n) \right\}^t \cdot \frac{\partial}{\partial \underline{\theta}} H^{-1}(s; \underline{\beta}_0, \underline{\theta}_0),$$

and $\hat{q}_n(s)$ is defined by (10.1).

Proof of Theorem 10.1. Under A_n , using Theorem 2B and the Taylor

expansion of G_n^{-1} and H^{-1} , we obtain (as in the proof of Theorem 8.1)

$$\begin{aligned} \hat{q}_n(s) &= \sqrt{n} [Q_n(s) - H^{-1}(s; \underline{\beta}_0, \hat{\underline{\theta}}_n)] \\ &= \sqrt{n} [Q_n(s) - G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_0)] + \sqrt{n} [G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_0) - G_n^{-1}(s; \underline{\beta}_0, \underline{\theta}_0)] \\ &\quad + \sqrt{n} [G_n^{-1}(s; \underline{\beta}_0, \underline{\theta}_0) - H^{-1}(s; \underline{\beta}_0, \underline{\theta}_0) - \sqrt{n} [H^{-1}(s; \underline{\beta}_0, \hat{\underline{\theta}}_n) - H^{-1}(s; \underline{\beta}_0, \underline{\theta}_0)]] \\ &= n^{-\frac{1}{2}} K(s, n) / g_n[G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_0); \underline{\beta}_n, \underline{\theta}_0] + \varepsilon_{6n}(s) \\ &\quad + \sqrt{n} (\underline{\beta}_n - \underline{\beta}_0)^t \cdot \frac{\partial}{\partial \underline{\beta}} G_n^{-1}(s; \underline{\beta}_n, \underline{\theta}_0) + u(s, \underline{\theta}_0) \end{aligned}$$

$$\begin{aligned}
 & - \sqrt{n}(\hat{\theta}_n - \theta_0)^t \cdot \frac{\partial}{\partial \theta} H^{-1}(s; \beta_0, \theta_n^*) \\
 & = n^{-\frac{1}{2}} K(s, n) / g_n [G_n^{-1}(s; \beta_n, \theta_0); \beta_n, \theta_0] \\
 & \quad + \gamma^t \cdot \frac{\partial}{\partial \beta} G_n^{-1}(s; \beta_n', \theta_0) + u(s, \theta_0) \\
 & = \{n^{-\frac{1}{2}} \sum_{j=1}^n l_n(X_j, \beta_n, \theta_0) + A \cdot \gamma\}^t \cdot \frac{\partial}{\partial \theta} H^{-1}(s; \beta_0, \theta_n^*) + \epsilon_{7n}(s),
 \end{aligned}$$

where (like in the proof of Theorem 8.1) $\sup_{0 < s < 1} |\epsilon_{6n}(s)| \xrightarrow{a.s.} 0$ and

$\sup_{0 < s < 1} |\epsilon_{7n}(s)| \xrightarrow{P} 0$, as $n \rightarrow \infty$, and

$$\|\beta_n' - \beta_0\| \leq \|\beta_n - \beta_0\|, \quad \|\theta_n^* - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|.$$

Again, we have $\sum_{j=1}^n l_n(X_j, \beta_n, \theta_0) = \int l_n(x, \beta_n, \theta_0) d_x n F_n(x)$

$= \int l_n(x, \beta_n, \theta_0) d_x n [F_n(x) - G_n(x; \beta_n, \theta_0)]$, under A_n . By using

Theorem 1A, we have that

$$n^{-\frac{1}{2}} \sum_{j=1}^n l_n(X_j, \beta_n, \theta_0) = \int l_n(x, \beta_n, \theta_0) d_x n^{-\frac{1}{2}} K(G_n(x; \beta_n, \theta_0)$$

$+ \int l_n(x, \beta_n, \theta_0) d_x \epsilon_{2n}(x)$, where by Lemma 8.1,

$$\int l_n(x, \beta_n, \theta_0) d_x \epsilon_{2n}(x) \xrightarrow{a.s.} 0, \quad \text{as } n \rightarrow \infty.$$

Finally, by (C4) (iv) and (vi), we have

$\sup_s \left\| \frac{\partial}{\partial \theta} H^{-1}(s; \beta_0, \theta_0) - \frac{\partial}{\partial \theta} G_n^{-1}(s; \beta_n, \theta_0) \right\| \rightarrow 0$, as $n \rightarrow \infty$,
and hence the theorem. □

Proof of Corollary 10.1.1. Since $\beta_n \rightarrow \beta_0$, as $n \rightarrow \infty$, by the continuity and convergence conditions (C3), (C2) (ix), and finally by the continuous sample path properties of the Kiefer process, the result follows. □

The proof of Theorem 10.2 is the same as that of Theorem 10.1. Theorem 10.2 is a special case of Theorem 10.1 ($G_n = H$, $l_n = 1$ and $\beta_n \rightarrow \beta_0$).

Remark 10.1. We note that h , respectively g_n , is assumed (by Theorem 2B) to be bounded away from zero on its interval of positivity (a, b) , respectively (a_n, b_n) . Hence the Gaussian processes defined in Theorems 10.1 and 10.2, and in Corollary 10.1.1 are well-defined for $s \in (0, 1)$. (We are not dividing by zero). The interval (a_n, b_n) of positivity for g_n need not coincide with (a, b) , but by condition (C4) (vi), $a_n \rightarrow a$ and $b_n \rightarrow b$.

As to the covariance of η and η_0 , it is similar to that given in Remark 8.4.

As in Section 8, the limiting Gaussian processes of Theorem 10.2 and Corollary 10.1.1 depend, in general, on θ_0 , the true unknown value of θ . In order to test the composite hypotheses (8.1) and (8.3), define

$$\hat{\eta} = \eta(s, n; H, \beta_0, \hat{\theta}_n)$$

by (10.3), but with θ_0 replaced by $\hat{\theta}_n$ everywhere it appears.

Then, we obtain

Theorem 10.3. Under the conditions of Theorem 10.1,

$$\sup_{0 < s < 1} |\eta(s, n; H, \underline{\beta}_0, \hat{\theta}_n) - \eta(s, n; H, \underline{\beta}_0, \underline{\theta}_0)| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, and hence under $\{A_n\}$,

$$\sup_{0 < s < 1} |\hat{q}_n(s) - \eta(s, n; H, \underline{\beta}_0, \hat{\theta}_n)| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, where $\hat{q}_n(s)$ is defined by (10.1).

If we assume that the conditions of Theorem 10.2 are satisfied, then under H_0 ,

$$\sup_{0 < s < 1} |\hat{q}_n(s) - \hat{\eta}(s, n)| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, where $\hat{\eta}(s, n)$ is defined as η_0 in Theorem 10.2 but with $\underline{\theta}_0$ replaced by $\hat{\theta}_n$ everywhere it appears.

The proof of Theorem 10.3 follows from $(\hat{\theta}_n - \underline{\theta}_0) \xrightarrow{P} 0$ and the continuity properties (uniformly in s) of the functions and the Kiefer process involved.

Remark 10.2. Theorem 10.3 states that $\hat{\eta}$, which does not depend on the true unknown value $\underline{\theta}_0$ of $\underline{\theta}$, is as good an approximation to $\hat{q}_n(s)$ as η is.

As stated in Section 9, maximum likelihood estimators often satisfy (C1) and (C2) , and one can make the same substitutions for l , l_n and A , in this case. However, the covariance function of η does not simplify as did the covariance function of γ in Section 9.

IV STRONG APPROXIMATIONS OF THE k -SAMPLE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES AND SOME DISTRIBUTION-FREE k -SAMPLE TESTS

11. Introduction

The aim of this chapter is to propose and study the rate of convergence of a number of univariate distribution-free tests, which are based on the k -sample univariate empirical process (Section 12) and on the k -sample quantile process (Section 13). Statistics based on the k -sample multivariate empirical process are discussed in Section 14. The results of this chapter can also be found in Burke and Csörgö ([7], to appear), except for Theorem 14.2 which is in Burke ([8], to appear).

12. The k -sample Univariate Empirical Process

Let X_{ji} ($1 \leq i \leq n(j)$) be k sequences of i.i.d. random variables each with d.f. $F^j(x)$ ($1 \leq j \leq k$). A classical problem is to test whether these k samples come from a common population with d.f. $F(x)$. Often, we do not know what that common d.f. might be. Thus we wish to test the following null hypotheses:

$$(12.1) \quad H_0 : F^1 = F^2 = \dots = F^k ;$$

$$(12.2) \quad H^0 : F^1 = F^2 = \dots = F^k = F .$$

Let $F_{n(j)}^j(x)$ denote the empirical d.f. based on the outcomes of the j^{th} sample $X_{j1}, X_{j2}, \dots, X_{j, n(j)}$. For each vector $N = (n(1), n(2), \dots, n(k))$ of positive integers, we define the k -sample empirical process $S_N(x)$ by

$$(12.3) \quad S_N(x) = \sum_{j=1}^k c_j(N, x) \sqrt{n(j)} F_{n(j)}^j(x),$$

where the coefficients $c_j(N, x)$ satisfy

$$(12.4) \quad \sum_{j=1}^k c_j(N, x) \sqrt{n(j)} = 0, \text{ for each } N, \text{ and}$$

$$\sup_N \sup_x |c_j(N, x)| < \infty.$$

The main result of this section is

Theorem 12.1. On the null hypotheses (12.1) and (12.2), assuming that the true common d.f. F is continuous and that condition (12.4) is satisfied, there exists k independent sequences $\{B_{n(j)}^j\}$ of Brownian Bridges and k independent Kiefer processes K_j such that

$$\sup_{x \in R} |S_N(x) - \sum_{j=1}^k c_j(N, x) B_{n(j)}^j(F(x))|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} ([n(j)]^{-\frac{1}{2}} \log n(j)) \right\}$$

and

$$\sup_{x \in R} |S_N(x) - \sum_{j=1}^k [n(j)]^{-\frac{1}{2}} c_j(N, x) K_j(F(x), n(j))|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} ([n(j)]^{-\frac{1}{2}} \log^2 n(j)) \right\},$$

where $S_N(x)$ is defined by (12.3).

Proof. From (12.4), $S_N(x) = \sum_{j=1}^k c_j(N, x) \sqrt{n(j)} [F_{n(j)}^j(x) - F(x)]$.

For each j , by Theorem 1A, we can construct a sequence $\{B_{n(j)}^j(t)\}$ of Brownian Bridges ($t \in I$), such that

$$\sup_x |\sqrt{n(j)} [F_{n(j)}^j(x) - F(x)] - B_{n(j)}^j(F(x))|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ [n(j)]^{-\frac{1}{2}} \log n(j) \right\}.$$

Since the k samples are independent, the k sequences $\{B_{n(j)}^j\}$ can be constructed independently. So, since $\sup_N \sup_x |c_j(N, x)| < \infty$, our first assertion is proved. The proof of the second assertion of Theorem 12.1 is similar and hence the theorem. $\square$

Corollary 12.1.1. Suppose that the coefficients c_j depend on x only through F and write $c_j = c_j(N, F(x))$. Then, under the conditions of Theorem 12.1, we have

$$\left| \int_{x \in R} S_N^2(x) dF(x) - \int_{t \in I} \left[\sum_{j=1}^k c_j(N, t) [n(j)]^{-\frac{1}{2}} K_j(t, n(j)) \right]^2 dt \right|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} ([n(j)]^{-\frac{1}{2}} \log^2 n(j)) \right\},$$

and

$$\left| \sup_{x \in \mathbb{R}} |S_N(x)| - \sup_{t \in I} \left| \sum_{j=1}^k c_j(N, t) [n(j)]^{-\frac{1}{2}} K_j(t, n(j)) \right| \right|$$

$$\stackrel{\text{a.s.}}{=} O \left\{ \max_{1 \leq j \leq k} ([n(j)]^{-\frac{1}{2}} \log^2 n(j)) \right\},$$

where the $K_j(\cdot, \cdot)$ ($1 \leq j \leq k$) are defined in Theorem 12.1. The same statements can be made in terms of the sequences $\{B_{n(j)}^j\}$ and for many other statistics which are asymptotically distribution-free in the usual one-sample case.

Proof. The results follow from Theorem 12.1 and the transformation $t = F(x)$, since F is a continuous d.f.. □

Corollary 12.1.2. Suppose that the coefficients c_j depend only on j and N and write $c_j = c_j(N)$. Assume that $\{c_j(N)\}$ satisfies, in addition to (12.4), $\sum_{j=1}^k c_j^2(N) = 1$, for all N . Then, under the conditions of Theorem 2.1,

$$\sup_x |S_N(x) - B_N(F(x))| \stackrel{\text{a.s.}}{=} O \left\{ \max_{1 \leq j \leq k} (n(j)^{-\frac{1}{2}} \log n(j)) \right\},$$

and consequently

$$S_N(x) \rightarrow B(F(x)) \text{ as all } n(j) \rightarrow \infty,$$

where $B_N(t) = \sum_{j=1}^k c_j(N) B_{n(j)}^j(t)$ is a Brownian Bridge for each N and $B(t)$ is a Brownian Bridge.

The proof follows from Theorem 12.1 and the fact that $\sum c_j^2(N) = 1$ implies that $B_N(t)$ is a Brownian Bridge for each N . (For each N , the j Brownian Bridges $B_{n(j)}^j(t)$ are independent). Hence one can deduce that the continuous functionals of $S_N(x)$ converge to the corresponding functionals of $B_N(F(x))$. In particular, we have

Corollary 12.1.3. Under the null hypothesis H_0 of (12.1) and assuming the conditions of Corollary 12.1.2, we obtain

$$\left| \sup_{x \in R} S_N(x) - \sup_{t \in I} B_N(t) \right| \stackrel{a.s.}{\rightarrow} \sigma\{g(N)\},$$

$$\left| \sup_{x \in R} |S_N(x)| - \sup_{t \in I} |B_N(t)| \right| \stackrel{a.s.}{\rightarrow} \sigma\{g(N)\},$$

$$\left| \int_{x \in R} S_N^2(x) dF_N(x) - \int_0^1 B_N^2(t) dt \right| \stackrel{a.s.}{\rightarrow} \sigma\{g(N)\}$$

and

$$\left| \int_{x \in R} S_N(x) dF_N(x) - \int_0^1 B_N(t) dt \right| \stackrel{a.s.}{\rightarrow} \sigma\{g(N)\},$$

where $B_N(t) = \sum_{j=1}^k c_j(N) B_{n(j)}^j(t)$, a Brownian Bridge for each N ,

$F_N(x) = \sum_{j=1}^k c_j^2(N) F_{n(j)}^j(x)$, and $g(N) = \max_{1 \leq j \leq k} (n(j)^{-\frac{1}{2}} \log n(j))$.

Remark 12.1. Since

$$P\left\{ \sup_{t \in I} B(t) < y \right\} = 1 - \exp\{-2y^2\}, \text{ if } y > 0,$$

and

$$P\left\{ \sup_{t \in I} |B(t)| < y \right\} = \sum_{k=-\infty}^{\infty} (-1)^k \exp \{-2k^2 y^2\}, \quad \text{if } y > 0,$$

then

$$P\left\{ \sup_{x \in R} S_N(x) < y \right\} \xrightarrow{d} 1 - \exp \{-2y^2\}, \quad \text{if } y > 0,$$

and

$$P\left\{ \sup_{x \in R} |S_N(x)| < y \right\} \xrightarrow{d} \sum_{k=-\infty}^{\infty} (-1)^k \exp \{-2k^2 y^2\}, \quad \text{if } y > 0$$

as all $n(j) \rightarrow \infty$.

The univariate k -sample empirical process was previously studied by Kiefer (1959), where some useful examples concerning the $c_j(N)$ coefficients are given. The same considerations are also valid in the multivariate case of Section 14 of this chapter. M. Kac (1951) and Anderson and Darling (1952), in the one-sample case, used weight functions similar to our $c_j(N, x)$. The latter are useful when, for example, one would wish to give more weight to the tails of the distribution functions.

13. The k -sample Quantile Process.

Let X_{j1} ($1 \leq i \leq n(j)$) be univariate i.i.d. random variables with d.f. $F^j(x)$ ($1 \leq j \leq k$). As in Section 12, we wish to test whether these k -samples come from a common d.f. F , either

known or unknown. That is, we wish to test the null hypotheses H_0 of (12.1) and H^0 of (12.2).

Let us assume that $F(\cdot)$ has a differentiable density function $f(\cdot)$ and that the regularity conditions of Theorem 2A hold.

For $N = (n(1), n(2), \dots, n(k))$, consider the functions $\{c_j(N, t)\}_{j=1}^k$ of N and $t \in I$, the unit interval, and define

$$(13.1) \quad Z_N(t) = f(F^{-1}(t)) \sum_{j=1}^k c_j(N, t) \sqrt{n(j)} Q_{n(j)}^j(t), \text{ for } 0 < t < 1,$$

where $Q_{n(j)}^j$ is the quantile function of the $X_{j1}, \dots, X_{j, n(j)}$.

Consider the following conditions on the $c_j(N, t)$:

$$(13.2) \quad \sum_{j=1}^k c_j(N, t) \sqrt{n(j)} = 0, \text{ for each } N, \text{ and}$$

$$\sup_N \sup_{t \in I} |c_j(N, t)| < \infty.$$

We obtain

Theorem 13.1. On the null hypothesis H^0 of (12.2), assuming conditions (2.3), (2.4) and (13.2), there exists k independent sequences $\{B_{n(j)}^j(t)\}_{j=1}^k$ of Brownian Bridges such that

$$(13.3) \quad \sup_{t \in A(N)} |Z_N(t) - \sum_{j=1}^k c_j(N, t) B_{n(j)}^j(t)| \stackrel{a.s.}{=} O\left(\max_{1 \leq j \leq k} \left(\frac{\log n(j)}{\sqrt{n(j)}}\right)\right)$$

where $Z_N(t)$ is defined by (13.1), and $A(N)$ is the interval

$$\left[\max_{1 \leq j \leq k} \left(\frac{\log n(j)}{n(j)} \right), 1 - \max_{1 \leq j \leq k} \left(\frac{\log n(j)}{n(j)} \right) \right].$$

If we also assume (2.6), then

$$(13.4) \quad \sup_{0 < t < 1} |Z_N(t) - \sum_{j=1}^k c_j(N, t) B_{n(j)}^j(t)|$$

$$\stackrel{\text{a.s.}}{=} \left\{ \max_{1 \leq j \leq k} \left(\frac{\log n(j)}{\sqrt{n(j)}} \right), \sqrt{n(j)} \omega_{n(j)}(\epsilon) \right\},$$

for any $\epsilon > 0$, where (cf. Remark 2.1)

$$(13.5) \quad \omega_{n(j)}(\epsilon) = |F^{-1}(n(j))^{-2-\epsilon}| \cdot f[F^{-1}(n(j))^{-1} \log n(j)] .$$

Proof. As in the proof of Theorem 12.1,

$$Z_N(t) = \sum_{j=1}^k c_j(N, t) f[F^{-1}(t)] \sqrt{n(j)} \{Q_{n(j)}^j(t) - F^{-1}(t)\}$$

$$= \sum_{j=1}^k c_j(N, t) \{B_{n(j)}^j(t) + R_j(t, n(j))\},$$

$$\text{where } \sup_{t \in A(n(j))} |R_j(t, n(j))| \stackrel{\text{a.s.}}{=} O\left(\frac{\log n(j)}{\sqrt{n(j)}}\right),$$

$$\text{with } A(n(j)) = \left[\frac{\log n(j)}{n(j)}, 1 - \frac{\log n(j)}{n(j)} \right].$$

Thus, by (13.2)

$$\sup_{t \in A(N)} |Z_N(t) - \sum_{j=1}^k c_j(N, t) B_{n(j)}^j(t)|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} \left(\frac{\log n(j)}{\sqrt{n(j)}} \right) \right\}.$$

Similarly, if we assume condition (2.6), (13.4) can also be proved and we have the theorem. $\square$

Corollary 13.1.1. Suppose that the functions $c_j(N, t)$ depend only on j and the vector N , and write $c_j(N, t) = c_j(N)$. Assume that $\{c_j(N)\}_{j=1}^k$ satisfies, in addition to (13.2), $\sum_{j=1}^k [c_j(N)]^2 = 1$.

Then, assuming (2.3) and (2.4),

$$(3.6) \quad \sup_{t \in A(N)} |Z_N(t) - B_N(t)| \stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} \left(\frac{\log n(j)}{\sqrt{n(j)}} \right) \right\},$$

where $B_N(t) = \sum_{j=1}^k c_j(N) B_{n(j)}^j(t)$ and consequently, $B_N(t) \stackrel{d}{=} B(t)$, a Brownian Bridge on I , for each N .

Remark 13.1. In Corollary 13.1.1, if we also assume (2.6) and that

$$\max_{1 \leq j \leq k} (\sqrt{n(j)} \omega_{n(j)}(\epsilon)) \rightarrow 0 \quad (\text{cf Remark 2.1}) \text{ where } \omega_{n(j)}(\epsilon) \text{ is defined}$$

by (13.5), then (3.6) is valid if the supremum is taken over $(0, 1)$

and consequently $Z_N(t) \stackrel{d}{\rightarrow} B(t)$, as all $n(j) \rightarrow \infty$.

Remark 13.2. It is both important and desirable that $Z_N(t)$, defined by (13.1), depends on the proposed d.f. F (via the factor $f(F^{-1}(t))$), when we wish to test the null hypothesis, H^0 of (12.2),

by the Kolmogorov-Smirnov, Cramér-von Mises, and similar statistics.

In order to test the null hypothesis H_0 of (12.1) without knowing what F is, we must estimate the factor $f(F^{-1}(t))$ of (13.1). In this context, we find that the more restrictive conditions of Theorem 2B on f , as compared to those of Theorem 2A, are more convenient. We estimate $F^{-1}(t)$ by $Q_n(t)$:

Lemma 13.1. Under the conditions of Theorem 2B,

$$\sup_{t \in I} |Q_n(t) - F^{-1}(t)| \stackrel{\text{a.s.}}{=} O\left(\frac{(\log n)^{\frac{1}{2}}}{n}\right).$$

Proof. From Theorem 2C, we obtain

$$\sup_{t \in I} |q_n(t) + E_n(F^{-1}(t)) \cdot \frac{1}{f(F^{-1}(t))}| \stackrel{\text{a.s.}}{=} O\left(\frac{(\log \log n)^{\frac{1}{2}} (\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right).$$

Hence

$$\sup_{t \in I} |[Q_n(t) - F^{-1}(t)] + \frac{[F_n(F^{-1}(t)) - t]}{f(F^{-1}(t))}| \stackrel{\text{a.s.}}{=} O\left(\frac{(\log \log n)^{\frac{1}{2}} (\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right)$$

which, when combined with (1.4), yields

$$\begin{aligned} P\left\{ \sup_{t \in I} |Q_n(t) - F^{-1}(t)| > \epsilon - O\left(\frac{(\log \log n)^{\frac{1}{2}} (\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right) \right\} \\ \leq C \exp \left\{ -2 \left(\epsilon - O\left[\frac{(\log \log n)^{\frac{1}{2}} (\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right] \right)^2 \cdot n \right\} \\ \leq C \exp \left\{ -2\epsilon^2 n + 4\epsilon \cdot O[n^{\frac{1}{2}} (\log \log n)^{\frac{1}{2}} (\log n)^{\frac{1}{2}}] \right\}. \end{aligned}$$

Let $\varepsilon = \delta \cdot \left(\frac{\log n}{n}\right)^{\frac{1}{2}}$.

Then $\left(\frac{\log n}{n}\right)^{\frac{1}{2}} \cdot O[n^{\frac{1}{2}}(\log \log n)^{\frac{1}{2}}(\log n)^{\frac{1}{2}}] \rightarrow 0$ and for large n ,

$$P\left\{\sup_{t \in I} |Q_n(t) - F^{-1}(t)| > \delta \left(\frac{\log n}{n}\right)^{\frac{1}{2}}\right\} \leq C n^{-2\delta^2}.$$

Since $\sum_{n=1}^{\infty} C n^{-2\delta^2} < \infty$, for $\delta^2 > \frac{1}{2}$, the result follows. $\square$

To estimate f , we must consider an appropriate empirical density function. Let $f_{j,n}^{(1)}(x)$, resp. $f_{j,n}^{(2)}(x)$, denote the empirical density function of Definition 2.1, resp. Definition 2.2, based on the j^{th} sample $X_{j1}, X_{j2}, \dots, X_{j,n(j)}$.

By using Theorem 2B in the proof of Theorem 3.1, instead of Theorem 2A, we can obtain

Theorem 13.1*. On the null hypothesis H^0 of (12.2) and assuming the conditions of Theorem 2B on $F(\cdot)$, there exist k independent Kiefer processes K_j such that

$$\sup_{0 < t < 1} |Z_N(t) - \sum_{j=1}^k c_j(N, t) [n(j)]^{-\frac{1}{2}} K_j(t, n(j))|$$

$$\stackrel{\text{a.s.}}{=} O\left\{\max_{1 \leq j \leq k} ([n(j)]^{-\frac{1}{2}} [\log \log n(j)]^{\frac{1}{2}} [\log n(j)]^{\frac{1}{2}})\right\},$$

where $Z_N(t)$ is defined by (13.1).

Using the above results, we obtain

Theorem 13.2. On the null hypothesis H_0 of (12.1), assuming conditions (13.2) and those of Theorems 2B and 2D, there exist k independent Kiefer processes K_j such that

$$(13.7) \sup_{0 < t < 1} |Z_N(t)/f[F^{-1}(t)] - \sum_{j=1}^k [n(j)]^{-\frac{1}{2}} c_j(N, t) \cdot K_j(t, n(j))/f_{j,n}^{(r)}[Q_{n(j)}^j(t)]|$$

$$\xrightarrow{P} 0$$

as all $n(j) \rightarrow \infty$, where $Z_N(t)$ is defined by (13.1) and $f_{j,n}^{(r)}$ is an empirical density function defined above ($r = 1, 2$).

Proof. Since f is bounded away from zero on (a, b) , by Theorem 13.1*, (13.7) is almost-surely true if $f_{j,n}^{(r)}[Q_{n(j)}^j(t)]$ is replaced by $f[F^{-1}(t)]$ everywhere it occurs in (13.7). Thus, it suffices to show that

$$\sup_{0 < t < 1} \left| \sum_{j=1}^k [n(j)]^{-\frac{1}{2}} c_j(N, t) K_j(t, n(j)) \cdot \left(\frac{1}{f[F^{-1}(t)]} - \frac{1}{f_{j,n}^{(r)}[Q_{n(j)}^j(t)]} \right) \right|$$

converges in probability to zero. This latter statement follows from our Lemma 13.1, Theorem 2D, the assumption that f is continuous and bounded away from zero, the fact that $[n(j)]^{-\frac{1}{2}} K_j(t, n(j)) \stackrel{D}{=} B_j(t)$, a Brownian Bridge for each j , and the inequality

$$\sup_{0 < t < 1} |f_{j,n}^{(r)}[Q_{n(j)}^j(t)] - f[F^{-1}(t)]|$$

$$\leq \sup_t |f_{j,n}^{(r)}[Q_{n(j)}^j(t)] - f[Q_{n(j)}^j(t)]| + \sup_t |f[Q_{n(j)}^j(t)] - f[F^{-1}(t)]|,$$

and hence the theorem. □

Remark 13.3. We note that $Z_N(t)/f[F^{-1}(t)]$ does not depend on f or F (after cancellation); therefore Theorem 13.2 is of interest from the point of view of H_0 of (12.1). However, the resulting sum of Kiefer processes of (13.7) to which $Z_N(t)/f[F^{-1}(t)]$ converges in distribution is complicated from a distributional point of view. With this observation and (12.1) in mind, we shall prove our next theorem.

Theorem 13.3. Suppose that the coefficients $c_j(N, t)$ depend only on j and the vector N , and write $c_j(N, t) = c_j(N)$. Assume that

$\{c_j(N)\}$ satisfies, in addition to (13.2), $\sum_{j=1}^k [c_j(N)]^2 = \sum_{j=1}^k c_j^2(N) = 1$,

for each N . Define $Z_N^{(r)}(t) = \left(\sum_{j=1}^k c_j^2(N) f_{j, n(j)}^{(r)}[Q_{n(j)}^j(t)] \right)$

$\left(\sum_{j=1}^k c_j(N) \sqrt{n(j)} Q_{n(j)}^j(t) \right)$. Then, on the null hypotheses (12.1)

and (12.2), assuming the conditions of Theorems 2B and 2D, we have

$$\sup_{0 < t \leq 1} |Z_N^{(r)}(t) - B_N(t)| \xrightarrow{P} 0 \text{ as all } n(j) \rightarrow \infty, \text{ where}$$

$$B_N(t) = \sum_{j=1}^k c_j(N) [n(j)]^{-1/2} K_j(t, n(j)), \text{ a Brownian Bridge for each } N,$$

with K_j a Kiefer process as in Theorem 13.1*.

Proof. It suffices to show that

$$\sup_{0 < t < 1} |Z_N^{(r)}(t) - Z_N(t)| \xrightarrow{P} 0$$

and this, in turn, follows from

$$\sup_{0 < t < 1} \left| \sum_{j=1}^k c_j^2(N) (f_{j,n}^{(r)}[Q_{n(j)}^j(t)] - f[F^{-1}(t)]) \right| \xrightarrow{a.s.} 0 \quad \text{and}$$

$$\sum_{j=1}^k c_j(N) \sqrt{n(j)} Q_{n(j)}^j(t) \xrightarrow{d} B(t)/f[F^{-1}(t)] = B_N(t)/f[F^{-1}(t)].$$

The latter statement follows from Theorem 13.1*, noting that f is continuous and bounded away from zero, while the former is obtained from the inequality stated in the proof of Theorem 13.2, and hence the Theorem. $\square$

One can immediately deduce

Corollary 13.3.1. Under the conditions of Theorem 13.3, we have

$Z_N^{(r)}(t) \xrightarrow{d} B(t)$, a Brownian Bridge.

Proof. Since $\sum_j c_j^2(N) = 1$, $B_N(t) = B(t)$ for each N . $\square$

The following Corollary contains statistics based on $Z_N^{(r)}(t)$ of Theorem 13.3. The asymptotic power of these statistics will be studied in Chapter V.

Corollary 13.3.2. Under the conditions of Theorem 13.3, we have

$$\sup_t Z_N^{(r)}(t) - \sup_t B_N(t) \xrightarrow{P} 0,$$

$$\sup_t |Z_N^{(r)}(t)| - \sup_t |B_N^{(r)}(t)| \xrightarrow{P} 0,$$

$$\int [Z_N^{(r)}(t)]^2 dt - \int B_N^2(t) dt \xrightarrow{P} 0,$$

and

$$\int Z_N^{(r)}(t) dt - \int B_N(t) dt \xrightarrow{P} 0, \text{ as all } n(j) \rightarrow \infty.$$

It perhaps is worthwhile to state the following Corollary to Theorem 2B, in the one-sample case. The asymptotic power of the statistics cited here will be studied in our Chapter V.

Corollary 2B.1. Under the conditions of Theorems 2B and 2D, there exists a Kiefer process K such that

$$(13.8) \quad \sup_{0 < t < 1} |f_n^{(r)}(Q_n(t)) q_n(t) - n^{-\frac{1}{2}} K(t, n)| \xrightarrow{P} 0$$

as $n \rightarrow \infty$, where $f_n^{(r)}$ is the empirical density function based on the sample $X_1, X_2, \dots, X_n$, and hence

$$\sup_{0 < t < 1} [f_n^{(r)}(Q_n(t)) q_n(t)] - \sup_{0 < t < 1} [n^{-\frac{1}{2}} K(t, n)] \xrightarrow{P} 0,$$

$$(13.9) \quad \sup_t |f_n^{(r)}(Q_n(t)) q_n(t)| - \sup_t |n^{-\frac{1}{2}} K(t, n)| \xrightarrow{P} 0,$$

$$\int [f_n^{(r)}(Q_n(t)) q_n(t)]^2 dt - \int n^{-1} K^2(t, n) dt \xrightarrow{P} 0,$$

and

$$\int f_n^{(r)}(Q_n(t)) q_n(t) dt - \int n^{-1/2} K(t,n) dt \xrightarrow{P} 0,$$

as $n \rightarrow \infty$.

Remark 13.4. Similar results can be obtained using other definitions of empirical density functions, if the conclusion of Theorem 2D holds. In fact, we need only require that the convergence in Theorem 2D be in-probability, since our results here are in-probability. For a survey of these definitions, we refer to Rosenblatt (1956) and for results on their strong consistency, to Révész (1972).

14. The k-sample Multivariate Empirical Process.

Using Theorem 1B instead of Theorem 1A in the proof, we are able to extend Theorem 13.1 to multidimensions.

Let $\underline{X}_{j1} \dots (1 \leq i \leq n(j))$ be k sequences of i.i.d. d -dimensional random vectors each with continuous d.f. $F^j(\underline{x})$ ($1 \leq j \leq k$). Like in Section 12, for each vector $N = (n(1), n(2), \dots, n(k))$ of positive integers, we define the k -sample d -dimensional empirical process by

$$(14.1) \quad S_N(\underline{x}) = \sum_{j=1}^k c_j(N, \underline{x}) \sqrt{n(j)} F_{n(j)}^j(\underline{x}),$$

where $F_{n(j)}^j(\underline{x})$ is the empirical d.f. based on the j^{th} sample

$\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{j,n(j)}$, and where the coefficients satisfy (12.4) with $\underline{x} \in R^d$. We obtain

Theorem 14.1. On the null hypothesis (12.2), assuming that the true common d.f. F satisfies the conditions of Theorem 1B and that (12.4) is satisfied, there exists k independent sequences $\{B_{n(j)}^j(\underline{t})\}$, $\underline{t} \in I^d$ of Brownian Bridges and k independent Kiefer processes $K_j(\underline{t}, n)$, $\underline{t} \in I^d$, such that

$$\sup_{\underline{x} \in R^d} |S_N(\underline{x}) - \sum_{j=1}^k c_j(N, \underline{x}) B_{n(j)}^j(TD_{\underline{x}})|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} ([n(j)]^{\frac{-1}{2(d+1)}} [\log n(j)]^{3/2}) \right\}$$

and

$$\sup_{\underline{x} \in R^d} |S_N(\underline{x}) - \sum_{j=1}^k [n(j)]^{-\frac{1}{2}} c_j(N, \underline{x}) K_j(TD_{\underline{x}}, n(j))|$$

$$\stackrel{\text{a.s.}}{=} O\left\{ \max_{1 \leq j \leq k} ([n(j)]^{\frac{-1}{2(d+2)}} \log^2 n(j)) \right\}.$$

The proof is the same as that for Theorem 12.1.

We immediately have the following Corollary:

Corollary 14.1.1. Suppose that the coefficients c_j depend only on j and N , and write $c_j = c_j(N)$. In addition to (12.4) let us assume that $\sum_{j=1}^k c_j^2(N) = 1$. Then, under the conditions of Theorem 14.1,

$$\left| \sup_{\underline{x}} S_N(\underline{x}) - \sup_{\underline{x}} B_N(\text{TFD}_{\underline{x}}) \right| \stackrel{\text{a.s.}}{=} o\{g(N)\},$$

$$\left| \sup_{\underline{x}} |S_N(\underline{x})| - \sup_{\underline{x}} |B_N(\text{TFD}_{\underline{x}})| \right| \stackrel{\text{a.s.}}{=} o\{g(N)\},$$

$$\left| \int S_N^2(\underline{x}) dF_N(\underline{x}) - \int B_N^2(\text{TFD}_{\underline{x}}) dF_N(\underline{x}) \right| \stackrel{\text{a.s.}}{=} o\{g(N)\},$$

and

$$\left| \int S_N(\underline{x}) dF_N(\underline{x}) - \int B_N(\text{TFD}_{\underline{x}}) dF_N(\underline{x}) \right| \stackrel{\text{a.s.}}{=} o\{g(N)\},$$

where $g(N) = \max_{1 \leq j \leq k} \left(n(j)^{\frac{-1}{2(d+1)}} [\log n(j)]^{3/2} \right),$

$$B_N(\text{TFD}_{\underline{x}}) = \sum_{j=1}^k c_j(N) B_{n(j)}^j(\text{TFD}_{\underline{x}}), \text{ a Theorem 1B-type Brownian Bridge}$$

$$\text{measure for each } N, \quad F_N(\underline{x}) = \sum_{j=1}^k c_j^2(N) F_{n(j)}^j(\underline{x}), \text{ and } \text{TFD}_{\underline{x}} = \text{TD}_{\underline{x}}$$

is defined by (1.8).

Remark 14.1. In the general multivariate case, one cannot say that the statistics cited in Corollary 14.1.1 are asymptotically

distribution-free. However, if $F(\underline{x})$ is a product d.f., that is,

$F(\underline{x})$ is product of its marginals, then $\sup_{\underline{x}} S_N(\underline{x}), \sup_{\underline{x}} |S_N(\underline{x})|,$

$\int S_N^2(\underline{x}) dF_N(\underline{x})$ and $\int S_N(\underline{x}) dF_N(\underline{x})$ have limiting distributions

equal to those of $\sup_{\underline{t} \in I^d} B(\underline{t}), \sup_{\underline{t}} |B(\underline{t})|, \int_{I^d} B^2(\underline{t}) d\underline{t}$ and

$\int_{I^d} B(\underline{t}) d\underline{t}$, respectively. For $d \geq 2$, the distribution of

$\sup_{\underline{t} \in I^d} B(\underline{t})$ and $\sup_{\underline{t} \in I^d} |B(\underline{t})|$ is not yet known. For $d = 2$,

Durbin (1970) has tabulated 10, 5 and 1% significance points of

$\int_{I^2} B^2(\underline{t}) d\underline{t}$ and he has obtained the distribution of $\int_{I^2} B(\underline{t}) d\underline{t}$.

For the integral of the d -dimensional Brownian Bridge, we obtain

Theorem 14.2. $\int_{\underline{t} \in I^d} B(\underline{t}) d\underline{t}$ is normally distributed with mean zero and variance $(3^d - 2^{2d})$.

Proof. Let $\underline{t}_1, \underline{t}_2, \dots, \underline{t}_n$ be a random sample of size n from a uniform I^d population, where $\underline{t}_1 = (s_1, t_{12}, t_{13}, \dots, t_{1d})$.

Then, $\bar{F}_n = \int F_n(\underline{t}) d\underline{t}$ is distributed as the mean of n independent random variables each of which is distributed as the product of d independent $(0,1)$ random variables. This follows from induction.

Assume the result is true for $(d-1)$ -dimensional random vectors from a uniform I^{d-1} population. (cf. Durbin (1970) for the case $d=2$).

Suppose, for a given s that exactly r of $\underline{t}_1 = (s_1, t_{12}, \dots, t_{1d})$ have their first coordinate $s_1 \leq s$ and denote their remaining $(d-1)$ -coordinates of these r points by $\underline{t}_{rj} = (t_{rj2}, \dots, t_{rjd})$ ($1 \leq j \leq r$). Then, by the induction hypothesis, it follows that

$$\begin{aligned} G_n(s) &= \int F_n(\underline{t}) dt_2 \dots dt_d \\ &= \frac{1}{n} \sum_{j=1}^r (1 - t_{rj2})(1 - t_{rj3}) \dots (1 - t_{rjd}). \end{aligned}$$

Let us order the n points $\underline{t}_1 = (s_1, t_{12}, \dots, t_{1d})$ by

$s_1 \leq s_2 \leq \dots \leq s_n$ and hence

$$\bar{F}_n = \int G_n(s) ds$$

$$= \frac{1}{n} \sum_{r=0}^n \sum_{j=1}^r (1 - t_{rj2}) \dots (1 - t_{rjd}) (s_{r+1} - s_r)$$

$$= \frac{1}{n} \sum_{r=0}^n (1 - t_{rr2}) \dots (1 - t_{rrd}) (1 - s_r),$$

by expanding the summation.

Thus, we obtain $E(\bar{F}_n) = 2^{-d}$, and $E(\bar{F}_n^2) = 3^{-d}$. By the Central Limit Theorem for i.i.d. r.v.'s, we have

$\int \sqrt{n} (F_n(t) - t_1 \dots t_d) dt = \sqrt{n} (\bar{F}_n - 2^{-d})$ is asymptotically normal with mean zero and variance $(3^{-d} - 2^{-2d})$ and hence the theorem. □

V THE ASYMPTOTIC POWER OF SOME STATISTICS BASED ON THE ONE AND
SEVERAL SAMPLE QUANTILE AND MULTIVARIATE EMPIRICAL PROCESSES

15. Introduction.

The Kolmogorov-Smirnov and Cramér-von Mises test statistics are useful for goodness-of-fit purposes. Their realm of applicability does not lie solely with goodness-of-fit, but in the present chapter we shall conduct our study in those terms. Using the strong approximation results of Chapters I and IV, we shall obtain expressions for the bounds of the asymptotic power of the Kolmogorov-Smirnov, Cramér-von Mises and integral test statistics against various classes of alternative distribution functions. We wish to put our emphasis not solely on the specific classes of alternatives considered, but also on the strong approximation techniques used. The reader might envisage other classes of alternatives for which these techniques would be applicable.

In the general multivariate case, very little is known as to the computation of the various bounds for the asymptotic power of these statistics or even their distribution under the null hypothesis. The difficulty lies in the fact that the limiting distribution of the statistics involved is, most of the time, not distribution-free and even when it is, the distribution problem of the resulting stochastic processes is not immediately solvable.

This chapter, in part, is an attempt to provide a theoretical framework for future statistical studies.

The following corollary is easily obtained from Theorem 1B:

Corollary 1B.1. Under the conditions of Theorem 1B, we have

$$(15.1) \quad \left| \sup_{\underline{x} \in R^d} E_n(\underline{x}) - \sup_{\underline{x} \in R^d} B_n(TFD_{\underline{x}}) \right| \stackrel{a.s.}{=} O\{g(n)\},$$

$$(15.2) \quad \left| \sup_{\underline{x}} |E_n(\underline{x})| - \sup_{\underline{x}} |B_n(TFD_{\underline{x}})| \right| \stackrel{a.s.}{=} O\{g(n)\},$$

$$(15.3) \quad \left| \int E_n^2(\underline{x}) dF_n(\underline{x}) - \int B_n^2(TFD_{\underline{x}}) dF_n(\underline{x}) \right| \stackrel{a.s.}{=} O\{g(n)\},$$

and

$$(15.4) \quad \left| \int E_n(\underline{x}) dF_n(\underline{x}) - \int B_n(TFD_{\underline{x}}) dF_n(\underline{x}) \right| \stackrel{a.s.}{=} O\{g(n)\},$$

where $\{B_n\}$ is defined in Theorem 1B, $F_n(\underline{x})$ is the empirical d.f.,

$$\text{and } g(n) = n^{\frac{-1}{2(d+1)}} (\log n)^{3/2}.$$

The other test statistics we will consider are given in (13.9), Corollary 13.3.2, and Corollary 14.1.1. Although the integral test statistic (15.4) is not as sensitive as the Cramér-von Mises' (15.3), it is appropriate for tests against one-sided alternatives and it possesses desirable additive and computability (cf. Theorem 14.2) properties.

Our Section 16 deals with obtaining bounds for the

asymptotic power of the one-sample Kolmogorov-Smirnov tests against various classes of alternative d.f. sequences defined by (16.1). In the univariate case, these are the classes studied by Quade (1965).

Section 17 carries out the same program for the k-sample Kolmogorov-Smirnov tests. The classes of alternative sequences, defined by (17.3) are k-sample analogues of those of (16.1).

The asymptotic power of the Cramér-von Mises and integral tests are studied in Section 18. The classes of alternative sequences include the case when the null hypothesis is $F = F(\underline{x}; \underline{\theta}_0)$ and the alternatives are of the form $G_n = F(\underline{x}; \underline{\theta}_n)$ where $\underline{\theta}$ is a vector of parameters of the d.f. F and (18.4) is satisfied.

The extension of Section 18 to the k-sample situation is carried out in Section 19.

The same program is carried out in Sections 20 through 23 for the one and several sample quantile process.

16. The One-sample Kolmogorov-Smirnov Tests.

If $H(\underline{x})$ is some completely specified d.f. satisfying the conditions of Theorem 1B, we may reject the null hypothesis that $F = H$ at level α , if

$$\sup_{\underline{x} \in R^d} \sqrt{n} (F_n(\underline{x}) - H(\underline{x})) \geq d_n^+(\alpha) \quad \text{or if}$$

$$\sup_{\underline{x} \in R^d} \sqrt{n} |F_n(\underline{x}) - H(\underline{x})| \geq d_n(\alpha),$$

where $d_n^+(\alpha)$ and $d_n(\alpha)$ are defined by

$$P\left\{\sup_{\underline{x}} E_n(\underline{x}) \geq d_n^+(\alpha) \mid F = H\right\} = \alpha \quad \text{and}$$

$$P\left\{\sup_{\underline{x}} |E_n(\underline{x})| \geq d_n(\alpha) \mid F = H\right\} = \alpha.$$

These tests are called the one and two-sided Kolmogorov-Smirnov (K-S) tests.

Remark 16.1: In Sections 16 through 19, it will be assumed that all distribution functions satisfy the conditions of Theorem 1B.

The power of the two-sided K-S test against some alternative d.f. G is defined as $P\left\{\sup_{\underline{x} \in R^d} \sqrt{n} |F_n(\underline{x}) - H(\underline{x})| \geq d_n(\alpha) \mid F = G\right\}$.

The power of the one-sided K-S test against G is defined similarly.

Remark 16.2. The two-sided K-S test is consistent against all alternatives $F = G$ for which $|H(\underline{x}_0) - G(\underline{x}_0)| \neq 0$ for some $\underline{x}_0 \in R^d$:

$$\begin{aligned} \sqrt{n} (F_n(\underline{x}) - H(\underline{x})) &= \sqrt{n} (F_n(\underline{x}) - G(\underline{x})) + \sqrt{n} (G(\underline{x}) - H(\underline{x})) \\ &= B_n(\text{TGD}_{\underline{x}}) + \sqrt{n} (G(\underline{x}) - H(\underline{x})) + R_n(\underline{x}), \end{aligned}$$

where $\{B_n\}$ is the sequence of Brownian Bridges defined in Theorem 1B with $\sup_{\underline{x}} |R_n(\underline{x})| \stackrel{a.s.}{=} O\{n^{-1/[2(d+1)]} (\log n)^{3/2}\}$. By Corollary

1B.1, there exists a function $d(\alpha)$ such that $\sup_{\alpha} |d_n(\alpha) - d(\alpha)| \rightarrow 0$

as $n \rightarrow \infty$, $d(\alpha)$ is obtained by solving the equation

$P\{\sup_{\underline{x}} |B(\text{THD}_{\underline{x}})| \geq d(\alpha)\} = \alpha$ and hence

$$\begin{aligned} & \lim_{n \rightarrow \infty} P\left\{\sup_{\underline{x} \in R^d} \sqrt{n} |F_n(\underline{x}) - H(\underline{x})| > d_n(\alpha)\right\} \\ &= \lim_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} |B(\text{TGD}_{\underline{x}}) - \sqrt{n} (H(\underline{x}) - G(\underline{x}))| \geq d(\alpha)\right\} \\ &\geq \lim_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} \sqrt{n} |H(\underline{x}) - G(\underline{x})| \geq d(\alpha) + \sup_{\underline{x}} |B(\text{TGD}_{\underline{x}})|\right\} \\ &= 1, \text{ since } \sup_{\underline{x}} |H(\underline{x}) - G(\underline{x})| > 0. \end{aligned}$$

Remark 16.3. The one-sided K-S test is consistent against all

alternatives G (satisfying the conditions of Theorem 1B and) for

which $G(\underline{x}_0) - H(\underline{x}_0) = \epsilon > 0$ for some $\underline{x}_0 \in R^d$. By Corollary 1B.1,

$\lim_{n \rightarrow \infty} \sup_{\alpha} |d_n^+(\alpha) - d(\alpha)| = 0$, for $d^+(\alpha)$ satisfying

$P\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d^+(\alpha)\} = \alpha$, and we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} \sqrt{n} (F_n(\underline{x}) - H(\underline{x})) \geq d_n^+(\alpha) \mid F = G\right\} \\ &= \lim_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} [B_n(\text{TGD}_{\underline{x}}) - \sqrt{n} (H(\underline{x}) - G(\underline{x}))] \geq d_n^+(\alpha)\right\} \\ &\geq P\left\{\sup_{\underline{x}} B(\text{TGD}_{\underline{x}}) + \lim_{n \rightarrow \infty} \sqrt{n} \epsilon \geq d_n^+(\alpha)\right\} \\ &= 1, \text{ since } \sqrt{n} \epsilon \rightarrow \infty. \end{aligned}$$

2 OF/DE 2

OF/DE

Now, suppose $\{G_n\}$ is a sequence of alternative distribution functions. Then, the asymptotic power against $\{G_n\}$ is defined as

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{\underline{x} \in R^d} \sqrt{n} |F_n(\underline{x}) - H(\underline{x})| \geq d_n(\alpha) \mid F = G_n \right\} \text{ and}$$

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{\underline{x} \in R^d} \sqrt{n} (F_n(\underline{x}) - H(\underline{x})) \geq d_n^+(\alpha) \mid F = G_n \right\} \text{ for the}$$

two and one-sided K-S tests, respectively.

The remainder of this section is devoted to finding expressions for the bounds of the asymptotic power of the K-S tests against various classes of alternatives. Let

$$C_n(\underline{x}) = \sqrt{n} (H(\underline{x}) - G_n(\underline{x})).$$

Consider the following classes of alternatives:

(a) $\mathcal{K}_\Delta$, for $\Delta > 0$: all $\{G_n\}$ such that

$$\sup_{\underline{x} \in R^d} |C_n(\underline{x})| = \Delta,$$

(b) $\mathcal{K}_\Delta(\tau)$, for $\Delta > 0$ and $\tau \in R^d$: all $\{G_n\} \in \mathcal{K}_\Delta$

$$\text{such that } \limsup_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \inf_{\underline{x} \in S(\tau, \delta)} C_n(\underline{x}) = -\Delta,$$

(16.1)

where $S(\tau, \delta)$ is the open ball in R^d with centre τ and radius δ ;

(c) $\mathcal{K}_\Delta^*$, for $\Delta > 0$: $\bigcup_{\tau \in R^d} \mathcal{K}_\Delta(\tau)$;

(d) $\mathcal{K}_\Delta^+(\tau)$, for $\Delta > 0$ and $\tau \in R^d$: all $\{G_n\} \in \mathcal{K}_\Delta(\tau)$

such that $C_n(\underline{x}) \leq 0$ for all $\underline{x}$ and $n \geq n_0$, for some positive integer n_0 ;

$$(e) \mathcal{K}_{\Delta}^{+}, \text{ for } \Delta > 0 : \bigcup_{\tau \in R^d} \mathcal{K}_{\Delta}^{+}(\tau);$$

These are multivariate versions of the (univariate) classes studied in Quade (1965).

The least upper bound for the asymptotic power of the one-sided K-S test, against a class $\mathcal{K}$ at level α , will be denoted by $U^{+}(\alpha, \mathcal{K})$, that is

$$U^{+}(\alpha, \mathcal{K}) = \sup_{\{G_n\} \in \mathcal{K}} \limsup_{n \rightarrow \infty} P\left\{ \sup_{\underline{x} \in R^d} (E_n(\underline{x}) - C_n(\underline{x})) \geq d_n^{+}(\alpha) \right\}.$$

The greatest lower bound will be denoted by

$$L^{+}(\alpha, \mathcal{K}) = \inf_{\{G_n\} \in \mathcal{K}} \liminf_{n \rightarrow \infty} P\left\{ \sup_{\underline{x} \in R^d} (E_n(\underline{x}) - C_n(\underline{x})) \geq d_n^{+}(\alpha) \right\}.$$

The asymptotic least upper bound $U(\alpha, \mathcal{K})$ and greatest lower bound $L(\alpha, \mathcal{K})$ for the two-sided test may be defined analogously. We obtain

Theorem 16.1. Let $\mathcal{K}$ be any of the classes defined in (16.1).

Then,

$$U^{+}(\alpha, \mathcal{K}) = P\left\{ \sup_{\underline{x} \in R^d} B(\text{THD}_{\underline{x}}) \geq d^{+}(\alpha) - \Delta \right\}.$$

Proof. By Corollary 1B.1, we obtain

$$\limsup_{n \rightarrow \infty} P\left\{ \sup_{\underline{x} \in R^d} [E_n(\underline{x}) - C_n(\underline{x})] \geq d_n^{+}(\alpha) \right\}$$

$$\begin{aligned}
 &= \limsup_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} [B_n(TG_n D_{\underline{x}}) - C_n(\underline{x})] \geq d_n^+(\alpha)\right\} \\
 &\leq \limsup_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} B_n(TG_n D_{\underline{x}}) - \inf_{\underline{x}} C_n(\underline{x}) \geq d_n^+(\alpha)\right\} \\
 &= P\left\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d_n^+(\alpha) - \Delta\right\},
 \end{aligned}$$

since $\lim_{n \rightarrow \infty} \sup_{\underline{x}} |G_n(\underline{x}) - H(\underline{x})| = 0$ implies that

$$\lim_{n \rightarrow \infty} \sup_{\underline{x}} |\lambda(\text{THD}_{\underline{x}}) - \lambda(TG_n D_{\underline{x}})| = 0 \quad (\text{where } \lambda \text{ is the Lebesgue}$$

measure on I^d), and hence we have an upper bound. To prove it is the least upper bound, define $\{G_n\}$ so that

$$\begin{aligned}
 C_n(\underline{x}) &= -\sqrt{n} F(\underline{x}), \quad \text{if } 0 < F(\underline{x}) \leq \Delta/\sqrt{n}; \\
 &= -\Delta, \quad \text{if } \Delta/\sqrt{n} \leq F(\underline{x}) \leq 1 - \Delta/\sqrt{n}; \\
 &= \sqrt{n}(F(\underline{x}) - 1), \quad \text{if } 1 - \Delta/\sqrt{n} \leq F(\underline{x}) < 1.
 \end{aligned}$$

Then, $\{G_n\} \in \mathcal{K}$ for any class $\mathcal{K}$ of (16.1), $\lim_{n \rightarrow \infty} \sup_{\underline{x}} [B(TG_n D_{\underline{x}}) - C_n(\underline{x})] = \sup_{\underline{x}} B(\text{THD}_{\underline{x}}) + \Delta$, and hence the result. □

Theorem 16.2. We have

$$(16.2) \quad L^+(\alpha, \mathcal{K}_\Delta) = P\left\{\sup_{\underline{x} \in R^d} B(\text{THD}_{\underline{x}}) \geq d^+(\alpha) + \Delta\right\},$$

$$(16.3) \quad L^+(\alpha, \mathcal{K}_\Delta(\tau)) = P\left\{\sup_{\underline{x} \in R^d} [B(\text{THD}_{\underline{x}}) - f_\tau(\underline{x})] \geq d^+(\alpha)\right\},$$

$$\text{where } f_\tau(\underline{x}) = \Delta, \quad \text{if } \underline{x} \neq \tau \\ = -\Delta, \quad \text{if } \underline{x} = \tau,$$

$$(16.4) \quad L^+(\alpha, \mathcal{K}_\Delta^*) = \inf_{\tau \in R^d} L^+(\alpha, \mathcal{K}_\Delta(\tau)),$$

$$(16.5) \quad L^+(\alpha, \mathcal{K}_\Delta^+(\tau)) = P\left\{\sup_{\underline{x} \in R^d} [B(\text{THD}_{\underline{x}}) - g_\tau(\underline{x})] \geq d^+(\alpha)\right\},$$

$$\text{where } g_\tau(\underline{x}) = 0, \quad \text{if } \underline{x} \neq \tau \\ = -\Delta, \quad \text{if } \underline{x} = \tau,$$

$$(16.6) \quad L^+(\alpha, \mathcal{K}_\Delta^+) = \inf_{\tau \in R^d} L^+(\alpha, \mathcal{K}_\Delta^+(\tau)).$$

Remark 16.4. In the univariate situation ($d = 1$), the K-S statistic is distribution-free and each of the bounds can be explicitly evaluated. It is known that $P\left\{\sup_{t \in I} B(t) \geq \lambda\right\} = \exp(-2\lambda^2)$ and so, for example $L^+(\alpha, \mathcal{K}_\Delta) = \exp\{-2(d^+(\alpha) + \Delta)^2\}$ and $U^+(\alpha, \mathcal{K}) = \exp\{-2(d^+(\alpha) - \Delta)^2\}$. For evaluation of the other (univariate) bounds, we refer to Quade (1965). In the multivariate situation, if $H(\underline{x})$ is the product of its marginals, the K-S statistic is also distribution free but the exact distribution of $\sup_{\underline{t} \in I^d} B(\underline{t})$ is not yet known.

Proof of Theorem 16.2. Assertions (16.4) and (16.6) are clearly

true. To prove (16.2), we have for any $\{G_n\} \in \mathcal{K}_\Delta$,

$$\begin{aligned} \liminf_{n \rightarrow \infty} P \left\{ \sup_{\underline{x} \in R^d} (E_n(\underline{x}) - C_n(\underline{x})) \geq d_n^+(\alpha) \right\} \\ = \liminf_{n \rightarrow \infty} P \left\{ \sup_{\underline{x}} (B_n(TG_n D_{\underline{x}}) - C_n(\underline{x})) \geq d_n^+(\alpha) \right\} \\ \geq \liminf_{n \rightarrow \infty} P \left\{ \sup_{\underline{x}} B_n(TG_n D_{\underline{x}}) - \sup_{\underline{x}} C_n(\underline{x}) \geq d_n^+(\alpha) \right\} \\ \geq P \left\{ \sup_{\underline{x}} B(THD_{\underline{x}}) - \Delta \geq d^+(\alpha) \right\}, \end{aligned}$$

which is a lower bound. To prove it is the greatest lower bound,

define $\{G_n\}$ so that

$$\begin{aligned} C_n(\underline{x}) &= \sqrt{n} F(\underline{x}), & \text{if } 0 < F(\underline{x}) < \Delta/\sqrt{n} \\ &= \Delta, & \text{if } \Delta/\sqrt{n} \leq F(\underline{x}) \leq 1 - \Delta/\sqrt{n} \\ &= \sqrt{n} (1 - F(\underline{x})), & \text{if } 1 - \Delta/\sqrt{n} \leq F(\underline{x}) < 1. \end{aligned}$$

Then $\{G_n\} \in \mathcal{K}_\Delta$ and since $\sup_{0 < F(\underline{x}) < 1} |C_n(\underline{x}) - \Delta| \rightarrow 0$, we have

$$\sup_{\underline{x}} [B(TG_n D_{\underline{x}}) - C_n(\underline{x})] \xrightarrow{P} \sup_{\underline{x}} B(THD_{\underline{x}}) - \Delta \text{ and hence (16.2) is proved.}$$

Now, for any $\{G_n\} \in \mathcal{K}_\Delta(\tau)$, the following is clear:

$$\begin{aligned} \liminf_{n \rightarrow \infty} P \left\{ \sup_{\underline{x}} (E_n(\underline{x}) - C_n(\underline{x})) \geq d_n^+(\alpha) \right\} \\ = \liminf_{n \rightarrow \infty} P \left\{ \sup_{\underline{x}} (B_n(TG_n D_{\underline{x}}) - C_n(\underline{x})) \geq d_n^+(\alpha) \right\} \end{aligned}$$

$$\geq P\left\{\sup_{\underline{x}} [B(\text{THD}_{\underline{x}}) - f_{\tau}(\underline{x})] \geq d^+(\alpha)\right\},$$

since $\liminf_{n \rightarrow \infty} C_n(\underline{x}) \geq f_{\tau}(\underline{x})$, and so we have a lower bound. To prove it is the greatest lower bound, define $\{G_n\}$ so that

$$\begin{aligned} C_n(\underline{x}) &= \sqrt{n} F(\underline{x}), & \text{if } 0 < F(\underline{x}) \leq \Delta/\sqrt{n} \\ &= \sqrt{n} \|\underline{x} - \tau\| - \Delta, & \text{if } \underline{x} \in S(\tau, 2\Delta/\sqrt{n}) \\ &= \sqrt{n} (1 - F(\underline{x})), & \text{if } 1 - \Delta/\sqrt{n} \leq F(\underline{x}) < 1 \\ &= \Delta, & \text{elsewhere.} \end{aligned}$$

Then for sufficiently large n , $\{G_n\}$ is well-defined and $\{G_n\} \in \mathcal{K}_{\Delta}(\tau)$. Then $\limsup_{n \rightarrow \infty} \sup_{\underline{x}} |C_n(\underline{x}) - f_{\tau}(\underline{x})| = 0$, hence

$$\sup_{\underline{x}} [B_n(\text{TG}_{n, \underline{x}}^D) - C_n(\underline{x})] \xrightarrow{P} \sup_{\underline{x}} [B(\text{THD}_{\underline{x}}) - f_{\tau}(\underline{x})]$$

and so (16.3) is proved.

To prove (16.5), for any $\{G_n\} \in \mathcal{K}_{\Delta}^+(\tau)$, we have

$$\begin{aligned} &\liminf_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} (E_n(\underline{x}) - C_n(\underline{x})) \geq d_n^+(\alpha)\right\} \\ &= \liminf_{n \rightarrow \infty} P\left\{\sup_{\underline{x}} [B_n(\text{TG}_{n, \underline{x}}^D) - C_n(\underline{x})] \geq d_n^+(\alpha)\right\} \\ &\geq P\left\{\sup_{\underline{x}} [B(\text{THD}_{\underline{x}}) - g_{\tau}(\underline{x})] \geq d^+(\alpha)\right\}, \end{aligned}$$

which is a lower bound. To prove it is the greatest lower bound, define $\{G_n\}$ so that

$$C_n(\underline{x}) = \sqrt{n} \left| \|\underline{x} - \tau\| - \Delta \right|, \text{ if } \underline{x} \in S(\tau, \Delta/\sqrt{n})$$

$$= 0 \quad \text{elsewhere.}$$

Then, $\{G_n\} \in \mathcal{K}_\Delta^+(\tau)$ and $\lim_{n \rightarrow \infty} \sup_{\underline{x}} |C_n(\underline{x}) - g_\tau(\underline{x})| = 0$,

hence $\sup_{\underline{x}} [B_n(TG_n D_n) - C_n(\underline{x})] \xrightarrow{P} \sup_{\underline{x}} [B(\text{THD}_n) - g_\tau(\underline{x})]$

and the theorem is proved. □

For the two-sided K-S test, we obtain the following:

Theorem 16.3. We have

$$(16.7) \quad U(\alpha, \mathcal{K}_\Delta) = U(\alpha, \mathcal{K}_\Delta(\tau)) = U(\alpha, \mathcal{K}_\Delta^*)$$

$$= P\left\{ \sup_{\underline{x} \in R_d} |B(\text{THD}_n)| \geq d(\alpha) - \Delta \right\}$$

$$(16.8) \quad U(\alpha, \mathcal{K}_\Delta^+(\tau)) = U(\alpha, \mathcal{K}_\Delta^+)$$

$$= 1 - P\{-d(\alpha) + \Delta \leq B(\text{THD}_n) \leq d(\alpha), \text{ for all } \underline{x}\}$$

Proof. $\lim_{n \rightarrow \infty} \sup P\left\{ \sup_{\underline{x} \in R_d} |B_n(TG_n D_n) - C_n(\underline{x})| \geq d_n(\alpha) \right\}$

$$\leq \lim_{n \rightarrow \infty} \sup P\left\{ \sup_{\underline{x}} |B_n(TG_n D_n)| + \sup_{\underline{x}} |C_n(\underline{x})| \geq d_n(\alpha) \right\}$$

$$= P\left\{ \sup_{\underline{x}} |B(\text{THD}_n)| \geq d(\alpha) - \Delta \right\}$$

which is an upper bound. To prove it is the least upper bound, define $\{G_n\}$ so that

$$C_n(\underline{x}) = \Delta \cdot \sin(\sqrt{n} F(\underline{x})) \cdot [1 - F(\underline{x})]/\Delta.$$

Then, the sets $\{\underline{x} \in R^d \mid C_n(\underline{x}) = \Delta, \text{ for some } n \in N\}$ and $\{\underline{x} \in R^d \mid C_n(\underline{x}) = -\Delta, \text{ for some } n \in N\}$ are both dense in R^d . Thus, $\{G_n\} \in \mathcal{K}_\Delta(\tau)$, for every τ , and hence $\{G_n\} \in \mathcal{K}_\Delta^*$ and $\mathcal{K}_\Delta$. We have

$$\lim_{n \rightarrow \infty} \sup_{\underline{x} \in R^d} |B_n(TG_n D_n \underline{x}) - C_n(\underline{x})| = \sup_{\underline{x} \in R^d} |B(\text{THD}_\underline{x})| + \lim_{n \rightarrow \infty} \sup_{\underline{x} \in R^d} |C_n(\underline{x})|$$

and hence (16.7) is proved.

The proof that the right side of (16.8) is an upper bound follows as above. It is the least upper bound since if we choose $\{G_n\}$ so that $C_n(\underline{x}) = -\Delta \sin^2(\sqrt{n} F(\underline{x})[1 - F(\underline{x})]/2\Delta)$, as in the proof of (16.7), then $\{G_n\} \in \mathcal{K}_\Delta^+(\tau)$, for all τ , hence $\{G_n\} \in \mathcal{K}_\Delta^+$, and

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{\underline{x}} P\{\sup_{\underline{x}} |B_n(TG_n D_n \underline{x}) - C_n(\underline{x})| \geq d_n(\alpha)\} \\ = 1 - P\{\Delta - d(\alpha) \leq B(\text{THD}_\underline{x}) \leq d(\alpha), \text{ for all } \underline{x} \in R^d\} \end{aligned}$$

and hence the theorem. □

Again in the univariate case ($d = 1$), one can obtain explicit expressions for the bounds in Theorem 16.3 and we refer to Quade (1965). As in Quade (1965) for $d = 1$, we were not able to obtain exact expressions for the greatest lower bounds of the two-sided K-S test. However, we do obtain the following

Theorem 16.4. Let $\beta = P\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d(\alpha)\}$. Then, for any of the classes $\mathcal{K}$ of (16.1), we have

$$(16.9) \quad L(\alpha, \mathcal{K}) \geq L^+(\beta, \mathcal{K}) \geq L^+(\alpha/2, \mathcal{K}).$$

For each $\tau \in R^d$, let $f_\tau(\underline{x})$ and $g_\tau(\underline{x})$ be defined as in Theorem 16.2. Then,

$$\begin{aligned} 1 - P\{-d(\alpha) + f_\tau(\underline{x}) \leq B(\text{THD}_{\underline{x}}) \leq d(\alpha) + f_\tau(\underline{x}), \text{ for all } \underline{x}\} \\ \leq L(\alpha, \mathcal{K}_\Delta(\tau)) \\ (16.10) \quad \leq L(\alpha, \mathcal{K}_\Delta^+(\tau)) \\ \leq 1 - P\{-d(\alpha) \leq B(\text{THD}_{\underline{x}}) \leq d(\alpha) + g_\tau(\underline{x}), \text{ for all } \underline{x}\}. \end{aligned}$$

Proof. Since for any $\{G_n\} \in \mathcal{K}$,

$$\begin{aligned} P\{\sup_{\underline{x}} |B(\text{TG}_{n, \underline{x}}) - C_n(\underline{x})| \geq d_n(\alpha)\} \\ \geq P\{\sup_{\underline{x}} [B(\text{TG}_{n, \underline{x}}) - C_n(\underline{x})] \geq d_n(\alpha)\}, \end{aligned}$$

the first inequality of (16.9) holds. Now

$$\alpha = P\{\sup_{\underline{x}} |B(\text{THD}_{\underline{x}})| \geq d(\alpha)\}$$

$$= P\left\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d(\alpha), \text{ or } \inf_{\underline{x}} B(\text{THD}_{\underline{x}}) \leq -d(\alpha)\right\}$$

$$\leq 2 P\left\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d(\alpha)\right\}$$

= 2δ and hence (16.9) follows.

To prove (16.10), choose $\{G_n\}$ as in the proof of part (16.3) of Theorem 16.2. Then, $\{G_n\} \in \mathcal{K}_{\Delta}(\tau)$ and

$$\lim_{n \rightarrow \infty} \sup_{\underline{x}} |C_n(\underline{x}) - f_{\tau}(\underline{x})| = 0. \text{ Hence}$$

$$L(\alpha, \mathcal{K}_{\Delta}(\tau)) \geq P\left\{\sup_{\underline{x}} |B(\text{THD}_{\underline{x}}) - f_{\tau}(\underline{x})| \geq d(\alpha)\right\}$$

which equals the left side of (16.10).

$$\text{Now for any } \{G_n\} \in \mathcal{K}_{\Delta}^+(\tau), \text{ we have } \lim_{n \rightarrow \infty} \sup_{\underline{x}} C_n(\underline{x}) \leq g_{\tau}(\underline{x}) \leq 0$$

for all $\underline{x} \in \mathbb{R}^d$. If we define $\{G_n\}$ as in the proof of (16.5) of Theorem 16.2, then $\lim_{n \rightarrow \infty} \sup_{\underline{x}} |C_n(\underline{x}) - g_{\tau}(\underline{x})| = 0$. Hence

$$L(\alpha, \mathcal{K}_{\Delta}^+(\tau))$$

$$= 1 - \sup_{\{G_n\} \in \mathcal{K}_{\Delta}^+(\tau)} \lim_{n \rightarrow \infty} \sup_{\underline{x}} P\{-d_n(\alpha) + C_n(\underline{x}) < B(\text{TG}_n^D \underline{x}) < d_n(\alpha) + C_n(\underline{x}), \text{ for all } \underline{x}\}$$

$$\leq 1 - \sup_{\{G_n\}} \lim_{n \rightarrow \infty} \sup_{\underline{x}} P\{-d_n(\alpha) < B(\text{TG}_n^D \underline{x}) < d_n(\alpha) + C_n(\underline{x}), \text{ for all } \underline{x}\}$$

$$\leq 1 - \lim_{n \rightarrow \infty} \sup P\{-d_n(\alpha) < B(TG_n D_{\underline{x}}) < d_n(\alpha) + g_r(\underline{x})\},$$

for all $\underline{x}$ }

which equals the right side of (16.10) and hence the theorem. □

17. The k-sample Kolmogorov-Smirnov Tests

Throughout this section, we will be using Corollary 14.1.1.

If $H(\underline{x})$ is some distribution function satisfying the conditions of Theorem 1B, we may reject the null hypothesis (12.2) at level α , if

$$\sup_{\underline{x} \in R^d} S_N(\underline{x}) \geq d_N^+(\alpha) \quad \text{or if}$$

$$\sup_{\underline{x} \in R^d} |S_N(\underline{x})| \geq d_N(\alpha), \quad \text{for large values of } n(j),$$

where $d_N^+(\alpha)$ and $d_N(\alpha)$ are defined by

$$P\left\{\sup_{\underline{x}} S_N(\underline{x}) \geq d_N^+(\alpha) \mid H^0\right\} = \alpha \quad \text{and}$$

$$P\left\{\sup_{\underline{x}} |S_N(\underline{x})| \geq d_N(\alpha) \mid H^0\right\} = \alpha,$$

$S_N(\underline{x})$ is defined by (12.3), and the c_j coefficients do not depend on $\underline{x}$ and satisfy (12.4) and $\sum_{j=1}^k c_j^2(N) = 1$, for each N .

These are the one and two-sided, k-sample K-S tests. The power of the two-sided, k-sample K-S test against the k alternative d.f.

$\{G^j\}_{j=1}^k$ is defined as

$$P\left\{ \sup_{\underline{x} \in R^d} |S_N(\underline{x})| \geq d_N(\alpha) \mid F^j = G^j \right\}.$$

The power for the one-sided, k -sample K-S test against $\{G^j\}$ is defined analogously. For the remainder of this section, we shall suppress the phrase k -sample.

Remark 17.1. By $N \rightarrow \infty$, we shall mean that all $n(j) \rightarrow \infty$ in some specified way. For example, one may require $N \rightarrow \infty$ so that $c_j(N) \rightarrow c_j$, for some constants c_j ($1 \leq j \leq k$).

Remark 17.2. The two-sided K-S test is consistent against all alternatives $A : F^j \neq G^j$ (cf. Remark 17.3) for which

$$(17.1) \quad \sup_{\underline{x} \in R^d} \left| \sum_{j=1}^k c_j(N) \sqrt{n(j)} G^j(\underline{x}) \right| \rightarrow \infty, \text{ as } N \rightarrow \infty.$$

$$\begin{aligned} \text{Under } A, \quad S_N(\underline{x}) &= \sum_{j=1}^k c_j(N) \sqrt{n(j)} [F_{n(j)}^j(\underline{x}) - G^j(\underline{x})] + \sum_{j=1}^k c_j(N) \sqrt{n(j)} G^j(\underline{x}) \\ &= \sum_{j=1}^k c_j(N) B_{n(j)}(TG^j D_{\underline{x}}) + \sum_{j=1}^k c_j(N) \sqrt{n(j)} G^j(\underline{x}) + R_N(\underline{x}), \end{aligned}$$

where the Brownian Bridges $B_{n(j)}$ are defined in Theorem 14.1 and

$$\sup_{\underline{x}} |R_N(\underline{x})| = o.p. \left\{ \max_{1 \leq j \leq k} (n(j))^{\frac{-1}{2(d+1)}} [\log n(j)]^{3/2} \right\}.$$

Hence, $\lim_{N \rightarrow \infty} P\{\sup_{\underline{x}} |S_N(\underline{x})| \geq d_N(\alpha)\}$

$$\geq \lim_{N \rightarrow \infty} P\{\sup_{\underline{x}} \left| \sum_j c_j(N) \sqrt{n(j)} G^j(\underline{x}) \right| \geq d_N(\alpha) +$$

$$\sup_{\underline{x}} \left| \sum_j c_j(N) B_{n(j)}(TG^j D_{\underline{x}}) \right| \}$$

$$\geq \lim_{N \rightarrow \infty} P\{\sup_{\underline{x}} \left| \sum_j c_j(N) \sqrt{n(j)} G^j(\underline{x}) \right|$$

$$\geq d_N(\alpha) + \sup_{\underline{x}} \left| \sum_{j=1}^k B_{n(j)}(TG^j D_{\underline{x}}) \right| \}$$

$$= 1, \text{ by (17.1).}$$

The last inequality holds because $\sum c_j^2(N) = 1$ implies that $|c_j(N)| \leq 1$ for each N and j .

Remark 17.3. As seen in Example 17.1 below, a particular (k-sample) two-sided K-S test may not be consistent against all alternatives.

Consistency against an alternative depends on the way in which all the $n(j)$ approach infinity. In a particular situation, if one wishes one's test to be consistent against a certain class $\mathcal{J}$ of alternatives, one should make the $n(j)$ approach infinity and define the coefficients $c_j(N)$ so that (17.1) is satisfied for all $\{G^j\} \in \mathcal{J}$.

On the other hand, given the coefficients $c_j(N)$ and the

way in which all $n(j) \rightarrow \infty$, the class of alternatives, against which the two-sided K-S test is consistent, will contain the class whose members satisfy (17.1).

Example 17.1. Let $k = 4$, $c_1(N) = c_3(N) = \frac{1}{2}$, $c_2(N) = c_4(N) = -\frac{1}{2}$ and consider the alternative $A : G^1 = G^2 \neq G^3 = G^4$. Then, letting $n(1) = n(2) = n(3) = n(4) = n$, we have

$\sum_{j=1}^4 c_j(N) \sqrt{n(j)} G^j(\underline{x}) = 0$. This particular test would not be

consistent against the alternative A , if all $n(j) \rightarrow \infty$ in such a way that $n(1) = n(2) = n(3) = n(4)$.

If, in addition, we assume that $G^1 = G^2 \geq G^3 = G^4$, then on letting $N \rightarrow \infty$ so that $n(1) = n(4)$ and $n(2) = n(3) = \log n(1)$,

we have $\sum_{j=1}^4 c_j(N) \sqrt{n(j)} G^j(\underline{x}) \rightarrow \infty$ for those $\underline{x}$'s satisfying

$G^1(\underline{x}) > G^4(\underline{x})$ and hence (17.1) is satisfied. On the other hand,

if we define the coefficients to be $c_1(N) = c_2(N) = \frac{1}{2}$ and

$c_3(N) = c_4(N) = -\frac{1}{2}$, then on letting $N \rightarrow \infty$ so that $n(1) = n(2) = n(3) = n(4)$, (17.1) is again satisfied.

Remark 17.4. As in Remark 17.2, the one-sided K-S test is consistent against alternatives $\{G^j\}$ for which

$$(17.2) \quad \sup_{\underline{x}} \left[\sum_{j=1}^k c_j(N) \sqrt{n(j)} G^j(\underline{x}) \right] \rightarrow \infty \quad \text{as } N \rightarrow \infty.$$

Example 17.2. If k is even, an appropriate test is that which gives equal weight to each sample. Suppose that for each j , the $c_j(N)$ coefficient does not change sign as $N \rightarrow \infty$. Suppose also that $|c_1(N) \sqrt{n(1)}| = |c_j(N) \sqrt{n(j)}|$, for all j ($1 \leq j \leq k$) and $c_1(N) \rightarrow c_1 \neq 0$ as $N \rightarrow \infty$. Then (17.1) is satisfied if

$$\left| \sum_{j=1}^k [\text{sign } c_j(N)] G^j(\underline{x}_0) \right| > 0 \text{ for some } \underline{x}_0 \in R^d : \text{ we have}$$

$$\sup_{\underline{x}} \left| \sum_{j=1}^k c_j(N) \sqrt{n(j)} G^j(\underline{x}) \right| = |c_1(N) \sqrt{n(1)}| \sup_{\underline{x}} \left| \sum_{j=1}^k [\text{sign } c_j(N)] G^j(\underline{x}) \right|$$

which converges to infinity as $N \rightarrow \infty$, since $c_1(N) \rightarrow c_1 \neq 0$ and $n(1) \rightarrow \infty$. Similarly, for the one-sided test, (17.2) is satisfied

if $\sum_{j=1}^k [\text{sign } c_j(N)] G^j(\underline{x}_0) > 0$ for some $\underline{x}_0 \in R^d$. The classical

Smirnov two-sample test, where $c_1(N) = (n(2)/[n(1) + n(2)])^{1/2}$ and $c_2(N) = - (n(1)/[n(1) + n(2)])^{1/2}$, is one instance of the above.

Now suppose $\{G_{n(j)}^j\}$ are k sequences of alternative distributions. Then the asymptotic power of the two-sided K-S test against $\{G_{n(j)}^j\}$ is defined as

$$\lim_{N \rightarrow \infty} P \left\{ \sup_{\underline{x} \in R^d} |S_N(\underline{x})| \geq d_N(\alpha) \mid F^j = G_{n(j)}^j \right\}.$$

For notational convenience, we shall write $G_{n(j)}$ for $G_{n(j)}^j$.

The asymptotic power of the one-sided K-S test against $\{G_{n(j)}\}$ is defined similarly.

Let $C_N(\underline{x}) = \sum_{j=1}^k c_j(N) \sqrt{n(j)} G_{n(j)}(\underline{x})$. Consider the

following classes of alternative sequences which are analogues of (16.1):

(a) $\mathcal{J}_\Delta$, for $\Delta > 0$: all $\{G_{n(j)}\}$ such that

$$\sup_{\underline{x}} |C_N(\underline{x})| \leq \Delta;$$

(17.3) (b) $\mathcal{J}_\Delta(\tau)$, for $\Delta > 0$ and $\tau \in R^d$: all $\{G_{n(j)}\} \in \mathcal{J}_\Delta$ such that

$$\limsup_{\delta \rightarrow 0} \limsup_{N \rightarrow \infty} \sup_{\underline{x} \in S(\tau, \delta)} C_N(\underline{x}) = \Delta,$$

where $S(\tau, \delta)$ is defined in (16.1);

(c) $\mathcal{J}_\Delta^*$, for $\Delta > 0$: $\bigcup_{\tau \in R^d} \mathcal{J}_\Delta(\tau)$;

(d) $\mathcal{J}_\Delta^+(\tau)$, for $\Delta > 0$ and $\tau \in R^d$: all $\{G_{n(j)}\} \in \mathcal{J}_\Delta(\tau)$ such that $C_N(\underline{x}) \geq 0$ for all $\underline{x} \in R^d$ and those N involved in $N \rightarrow \infty$ (cf. Remark 17.1);

(e) $\mathcal{J}_\Delta^+$, for $\Delta > 0$: $\bigcup_{\tau \in R^d} \mathcal{J}_\Delta^+(\tau)$.

Define the least upper bound of the asymptotic power of the one-sided K-S test, against a class $\mathcal{J}$ of alternatives at level α , by

$$\begin{aligned} U^+(\alpha, \mathcal{J}) &= \sup_{\{G_{n(j)}\} \in \mathcal{J}} \limsup_{N \rightarrow \infty} P\left\{\sup_{\underline{x}} S_N(\underline{x}) \geq d_N^+(\alpha) \mid F^j = G_{n(j)}\right\} \\ &= \sup_{\mathcal{J}} \limsup_{N \rightarrow \infty} P\left\{\sup_{\underline{x}} \left[\sum_j c_j(N) E_{n(j)}^j(\underline{x}) + C_N(\underline{x})\right] \geq d_N^+(\alpha)\right\} \end{aligned}$$

where $E_{n(j)}^j(\underline{x})$ is the (true) empirical process based on the j^{th} sample $X_{j1}, X_{j2}, \dots, X_{jn(j)}$. The bounds $L^+(\alpha, \delta)$, $U(\alpha, \delta)$ and $L(\alpha, \delta)$ can be defined similarly.

Let $H(\underline{x})$ denote the theoretical true common d.f. under the null hypothesis H^0 of (12.2). The analogous results of Theorems 16.1 through 16.4 are valid for the k -sample classes defined by (17.3):

Theorem 17.1. For the one-sided k -sample K-S tests, we obtain

$$U^+(\alpha, \delta) = P\left\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d^+(\alpha) - \Delta\right\};$$

$$L^+(\alpha, \delta_{\Delta}) = P\left\{\sup_{\underline{x}} B(\text{THD}_{\underline{x}}) \geq d^+(\alpha) + \Delta\right\};$$

$$L^+(\alpha, \delta_{\Delta}(\tau)) = P\left\{\sup_{\underline{x}} [B(\text{THD}_{\underline{x}}) - f_{\tau}(\underline{x})] \geq d^+(\alpha)\right\};$$

$$L^+(\alpha, \delta_{\Delta}^*) = \inf_{\tau \in R^d} L^+(\alpha, \delta_{\Delta}(\tau));$$

$$L^+(\alpha, \delta_{\Delta}^+(\tau)) = P\left\{\sup_{\underline{x}} [B(\text{THD}_{\underline{x}}) - g_{\tau}(\underline{x})] \geq d^+(\alpha)\right\};$$

$$L^+(\alpha, \delta_{\Delta}^+) = \inf_{\tau \in R^d} L^+(\alpha, \delta_{\Delta}^+(\tau)).$$

For the two-sided k -sample K-S tests, we obtain

$$U(\alpha, \delta_{\Delta}) = U(\alpha, \delta_{\Delta}(\tau)) = U(\alpha, \delta_{\Delta}^*)$$

$$= P\left\{\sup_{\underline{x}} |B(\text{THD}_{\underline{x}})| \geq d(\alpha) - \Delta\right\},$$

$$U(\alpha, \mathcal{J}_{\Delta}^{+}(\tau)) = U(\alpha, \mathcal{J}_{\Delta}^{+})$$

$$= 1 - P\{-d(\alpha) + \Delta \leq B(\text{THD}_{\underline{x}}) \leq d(\alpha) \text{ , for all } \underline{x}\} ;$$

$$L(\alpha, \mathcal{J}) \geq L^{+}(\beta, \mathcal{J}) \geq L^{+}(\alpha/2, \mathcal{J}) ,$$

where β is defined in Theorem 16.4;

$$1 - P\{-d(\alpha) + f_{\tau}(\underline{x}) \leq B(\text{THD}_{\underline{x}}) \leq d(\alpha) + f_{\tau}(\underline{x}) \text{ , for all } \underline{x}\}$$

$$\leq L(\alpha, \mathcal{J}_{\Delta}(\tau))$$

$$\leq L(\alpha, \mathcal{J}_{\Delta}^{+}(\tau))$$

$$\leq 1 - P\{-d(\alpha) \leq B(\text{THD}_{\underline{x}}) \leq d(\alpha) + g_{\tau}(\underline{x}) \text{ , for all } \underline{x}\} ,$$

where $\mathcal{J}$ is any of the classes defined by (17.3) and where $f_{\tau}(\underline{x})$ and $g_{\tau}(\underline{x})$ are defined in Theorem 16.2.

Proof. The proofs are the same as those in Section 16 with the following obvious changes of notation: $\{G_{n(j)}\}$ instead of $\{G_n\}$, $C_N(\underline{x})$ instead of $C_n(\underline{x})$, $N \rightarrow \infty$ instead of $n \rightarrow \infty$ and $\mathcal{J}$ instead of $\mathcal{K}$. □

Remark 17.5. The statements of Remark 16.3 regarding the evaluation of the bounds are also valid in the k-sample situation.

18. The One-sample Cramér-von Mises and Integral Tests.

We shall study the asymptotic power of the Cramér-von Mises (C-M) test statistic $\int E_n^2(\underline{x}) dF_n(\underline{x})$ and the integral test statistic $\int E_n(\underline{x}) dF_n(\underline{x})$. If $H(\underline{x})$ is a completely specified d.f. satisfying the conditions of Theorem 1B, we may reject the null hypothesis $F = H$ at level α , if

$$\int E_n^2(\underline{x}) dF_n(\underline{x}) \geq c_n(\alpha) \quad \text{or if}$$

$$\int E_n(\underline{x}) dF_n(\underline{x}) \geq c_n^+(\alpha) \quad \text{for large } n,$$

where $E_n(\underline{x}) = \sqrt{n} (F_n(\underline{x}) - H(\underline{x}))$ and where $c_n(\alpha)$ and $c_n^+(\alpha)$ are defined by

$$P \left\{ \int E_n^2(\underline{x}) dF_n(\underline{x}) \geq c_n(\alpha) \mid F = H \right\} = \alpha \quad \text{and}$$

$$P \left\{ \int E_n(\underline{x}) dF_n(\underline{x}) \geq c_n^+(\alpha) \mid F = H \right\} = \alpha.$$

Remark 18.1. The Cramér-von Mises (C-M) test is consistent against all alternatives $F = G$ for which $|H(\underline{x}_0) - G(\underline{x}_0)| \neq 0$ for some $\underline{x}_0 \in \mathbb{R}^d$. The proof is similar to that of Remark 16.2, since $H \neq G$ implies that $\int [G(\underline{x}) - H(\underline{x})]^2 dG(\underline{x}) > 0$ and hence $n \int [G(\underline{x}) - H(\underline{x})]^2 dF_n(\underline{x}) \rightarrow \infty$ as $n \rightarrow \infty$. Similarly, the integral test is consistent against those alternatives $F = G$ for which $\int [G(\underline{x}) - H(\underline{x})] dG(\underline{x}) = \varepsilon > 0$.

Now for a sequence $\{G_n\}$ of alternative distribution functions such that $\lim_{n \rightarrow \infty} \sup_{\underline{x}} |G_n(\underline{x}) - H(\underline{x})| = 0$, let $C_n(\underline{x})$ be

defined as in Section 16. Consider the following classes of alternatives

$$(a) \mathcal{L}_\Delta^2, \text{ for } \Delta > 0: \text{ all } \{G_n\} \text{ such that}$$

$$\lim_{n \rightarrow \infty} \int [C_n(\underline{x})]^2 dF_n(\underline{x}) \leq \Delta;$$

(18.1)

$$(b) \mathcal{L}_\Delta, \text{ for } \Delta > 0: \text{ all } \{G_n\} \text{ such that}$$

$$\lim_{n \rightarrow \infty} \left| \int C_n(\underline{x}) dF_n(\underline{x}) \right| \leq \Delta;$$

$$(c) \mathcal{L}_\Delta^+, \text{ for } \Delta > 0: \text{ all } \{G_n\} \in \mathcal{L}_\Delta \text{ for which}$$

$$C_n(\underline{x}) \leq 0 \text{ for all } \underline{x} \in R^d \text{ and sufficiently large } n.$$

Let $\mathcal{L}$ be any of the classes defined by (18.1). We define the least upper bound of the asymptotic power of the C-M test, against $\mathcal{L}$ at level α , to be

$$U(\alpha, \mathcal{L}) = \sup_{\{G_n\} \in \mathcal{L}} \limsup_{n \rightarrow \infty} P\left\{ \int n(F_n(\underline{x}) - H(\underline{x}))^2 dF_n(\underline{x}) \geq c_n(\alpha) \mid F = G_n \right\}.$$

The greatest lower bound for the C-M test $L(\alpha, \mathcal{L})$, the least upper bound $U^+(\alpha, \mathcal{L})$ and the greatest lower bound $L^+(\alpha, \mathcal{L})$ for the integral test are defined similarly.

$$\text{We have } \sup_{\alpha} |c_n(\alpha) - c(\alpha)| \rightarrow 0 \text{ and } \sup_{\alpha} |c_n^+(\alpha) - c^+(\alpha)| \rightarrow 0$$

as $n \rightarrow \infty$, by Corollary 1B.1 (Section 15), where $c(\alpha)$ and $c^+(\alpha)$ are defined by $P\left\{ \int B^2(\text{THD}_{\underline{x}}) dH(\underline{x}) \geq c(\alpha) \right\} = \alpha$ and $P\left\{ \int B(\text{THD}_{\underline{x}}) dH(\underline{x}) \geq c^+(\alpha) \right\} = \alpha$. One can obtain the following bounds for the asymptotic

power of the C-M and integral tests:

Theorem 18.1.

$$(18.2) \quad U(\alpha, \mathcal{L}_\Delta^2) \leq P\{(|\int B^2(\text{THD}_{\underline{x}}) dH(\underline{x})|^\frac{1}{2} - \Delta^\frac{1}{2})^2 \geq c(\alpha)\} ;$$

$$(18.3) \quad U^+(\alpha, \mathcal{L}_\Delta) = U^+(\alpha, \mathcal{L}_\Delta^+) = P\{\int B(\text{THD}_{\underline{x}}) dH(\underline{x}) \geq c^+(\alpha) - \Delta\} .$$

Theorem 18.2. We have

$$L(\alpha, \mathcal{L}_\Delta^2) \geq P\{(|\int B^2(\text{THD}_{\underline{x}}) dH(\underline{x})|^\frac{1}{2} - \Delta^\frac{1}{2})^2 \geq c(\alpha)\} ;$$

$$L^+(\alpha, \mathcal{L}_\Delta) = P\{\int B(\text{THD}_{\underline{x}}) dH(\underline{x}) \geq c^+(\alpha) + \Delta\} ;$$

$$L^+(\alpha, \mathcal{L}_\Delta^+) = \alpha .$$

Proof of Theorem 18.1. Statement (18.2) follows from the triangle inequality for L_2 norms, while (18.3) follows from integration. $\square$

The proof of Theorem 18.2 is the same as that for Theorem 18.1.

One class of alternative sequences closely related to those of (18.1) is the following:

let $F(\underline{x}) = F(\underline{x}; \underline{\theta})$, where $\underline{\theta} = (\theta_1, \dots, \theta_p)$ is a vector of parameters and suppose $H(\underline{x}) = F(\underline{x}; \underline{\theta}_0)$. Let $\mathcal{L}_{\underline{\theta}}$ be the class of sequences $\{G_n\}$ such that $G_n(\underline{x}) = F(\underline{x}; \underline{\theta}_n)$, where the parameters $\{\underline{\theta}_n\}$ satisfy

$$(18.4) \quad \lim_{n \rightarrow \infty} \sqrt{n} (\underline{\theta}_n - \underline{\theta}_0) = \gamma, \text{ for some vector } \gamma .$$

Denote the vector $((\partial/\partial\theta_1) F(\underline{x}; \underline{\theta}), (\partial/\partial\theta_2) F(\underline{x}; \underline{\theta}), \dots, (\partial/\partial\theta_p) F(\underline{x}; \underline{\theta}))$ of partial derivatives, evaluated at $\underline{\theta} = \underline{\theta}_0$, by $(\partial/\partial\theta) F(\underline{x}; \underline{\theta}_0)$ and assume that it exists. Assume also that $(\partial/\partial\theta) F(\underline{x}; \underline{\theta})$ is continuous in $\underline{\theta} \in v$, uniformly in $\underline{x} \in R^d$, where v is the closure of a given neighborhood of $\underline{\theta}_0$. Then, we obtain

$$\begin{aligned} \sqrt{n}(F_n(\underline{x}) - F(\underline{x}; \underline{\theta}_0)) &= \sqrt{n}(F_n(\underline{x}) - F(\underline{x}, \underline{\theta}_n)) + \sqrt{n}(F(\underline{x}; \underline{\theta}_n) - F(\underline{x}; \underline{\theta}_0)) \\ &= B_n(TG_n D_{\underline{x}}) + R_n(\underline{x}) + \sqrt{n}(\underline{\theta}_n - \underline{\theta}_0)(\partial/\partial\theta) F(\underline{x}; \underline{\theta}_n^*), \end{aligned}$$

by Taylor's formula, where $\|\underline{\theta}_n^* - \underline{\theta}_0\| \leq \|\underline{\theta}_n - \underline{\theta}_0\|$ and

$$\sup_{\underline{x}} |R_n(\underline{x})| \stackrel{\text{a.s.}}{=} O\{n^{-\frac{1}{2(d+1)}} (\log n)^{3/2}\}. \text{ By (18.4), we obtain}$$

$$\sup_{\underline{x}} |\sqrt{n}(F_n(\underline{x}) - H(\underline{x})) - [B_n(THD_{\underline{x}}) + \gamma(\partial/\partial\theta) F(\underline{x}; \underline{\theta}_0)]| \xrightarrow{P} 0,$$

under the sequence $\{G_n(\underline{x}) = F(\underline{x}; \underline{\theta}_n)\}$ of alternatives. We have shown

Theorem 18.3. Assume that $(\partial/\partial\theta) F(\underline{x}; \underline{\theta})$ is continuous in $\underline{\theta} \in v$, uniformly in $\underline{x} \in R^d$. For any sequence $\{G_n(\underline{x}) = F(\underline{x}; \underline{\theta}_n)\} \in \mathcal{C}_{\theta}$, on putting $F_0 = H = F(\cdot; \underline{\theta}_0)$, we obtain the following expressions for the asymptotic powers:

$$\begin{aligned} &P\{([\int B^2(TF_0 D_{\underline{x}}) dF_0(\underline{x})]^{1/2} + \Delta^{1/2})^2 \geq c(\alpha)\} \\ (18.5) \quad &\geq P\{[\int [B(TF_0 D_{\underline{x}}) + \gamma(\partial/\partial\theta) F(\underline{x}; \underline{\theta}_0)]^2 dF_0(\underline{x}) \geq c(\alpha)\} \\ &= \lim_{n \rightarrow \infty} P\{\int n[F_n(\underline{x}) - F_0(\underline{x})]^2 dF_n(\underline{x}) \geq c_n(\alpha) \mid F = G_n\} \\ &\geq P\{([\int B^2(TF_0 D_{\underline{x}}) dF_0(\underline{x})]^{1/2} - \Delta^{1/2})^2 \geq c(\alpha)\}, \end{aligned}$$

where $\int [\gamma \cdot (\partial/\partial \theta) F(\underline{x}; \underline{\theta}_0)]^2 dF_0(\underline{x}) = \Delta$; for the integral test, we obtain

$$(18.6) \quad \lim_{n \rightarrow \infty} P\left\{ \int \sqrt{n} (F_n(\underline{x}) - F_0(\underline{x})) dF_n(\underline{x}) \geq c_n^+(\alpha) \mid F = G_n \right\} \\ = P\left\{ \int B(TF_0 D_{\underline{x}}) dF_0(\underline{x}) + \int \gamma \cdot (\partial/\partial \theta) F(\underline{x}; \underline{\theta}_0) dF_0(\underline{x}) \geq c^+(\alpha) \right\}.$$

Example 18.1. The integral test, in the univariate case, can be used to test for a normal d.f. with mean μ_0 and variance σ^2 against sequences of alternative normal d.f. whose variances are fixed at σ^2 and whose means μ_n satisfy (18.4) with $\gamma < 0$:

$$\int B(F(x; \mu_0)) dF(x; \mu_0) + \gamma \cdot \int (\partial/\partial \mu) F(x; \mu_0) dF(x; \mu_0) \\ = Y - \gamma/(2\sigma\sqrt{\pi}) \sim N(-\gamma/(2\sigma\sqrt{\pi}), 1/12).$$

This follows from the fact that $Y = \int_0^1 B(t) dt \sim N(0, \frac{1}{12})$.

However, if the variances of the alternatives also change and satisfy (18.4), then $\int (\partial/\partial \sigma^2) F(\underline{x}; \underline{\theta}_0) dF(\underline{x}; \underline{\theta}_0) = 0$, where $\underline{\theta}_0 = (\mu_0, \sigma_0^2)$, and hence the integral test is not sensitive to such changes in the variances of these normal alternatives.

The preceding is also true in the fairly general situation, where $F(\underline{x}; \underline{\theta})$ is the product of its marginals, $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ is a vector of location parameters, F has density f , and the range of f does not depend on $\underline{\theta}$:

$$(\partial/\partial\theta_1) F(\underline{x}; \underline{\theta}) = -f(\underline{x}; \underline{\theta}) \quad \text{for each } \theta_1 \quad (1 \leq i \leq p)$$

and hence $\int \gamma \cdot (\partial/\partial\theta) F(\underline{x}; \underline{\theta}_0) dF(\underline{x}; \underline{\theta}_0) = -(\gamma_1 + \gamma_2 + \dots + \gamma_p) \cdot \int f^2(\underline{x}; \underline{\theta}_0) d\underline{x}$ which is positive if all the γ_i are negative, where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$. Hence we obtain

$$\begin{aligned} & \int B(TF_0 D_{\underline{x}}) dF(\underline{x}; \underline{\theta}_0) + \int \gamma (\partial/\partial\theta) F(\underline{x}; \underline{\theta}_0) dF(\underline{x}; \underline{\theta}_0) \\ &= \int_{t \in I^d} B(t) dt - (\gamma_1 + \dots + \gamma_p) \cdot \int f^2(\underline{x}; \underline{\theta}_0) d\underline{x}, \end{aligned}$$

which is normally distributed with mean $-(\gamma_1 + \dots + \gamma_p) \cdot \int f^2(\underline{x}; \underline{\theta}_0) d\underline{x}$ and variance $(3^{-d} - 2^{-2d})$ (cf. Theorem 14.2).

Example 18.2. Let $f(x; \theta) = \theta^{-1} \exp(-x/\theta)$ be the exponential density. Then, $\int (d/d\theta) F(x; \theta_0) dF(x; \theta_0) = -(2\theta_0^2)^{-1}$ and hence $\int B(F(x; \theta_0)) dF(x; \theta_0) + \gamma \cdot \int (d/d\theta) F(x; \theta_0) dF(x; \theta_0)$ is normally distributed with mean $-\gamma(2\theta_0^2)^{-1}$ and variance $1/12$.

19. The k-sample Cramér-von Mises and Integral Tests.

We shall study the asymptotic power of the (k-sample) C-M test statistic $\int S_N^2(\underline{x}) dF_N(\underline{x})$ and the (k-sample) integral statistic $\int S_N(\underline{x}) dF_N(\underline{x})$, defined in Corollary 14.1.1. The analogous statements made in Remarks 17.1, 17.2, 17.3 and 17.4 and in Examples 17.1 and 17.2 are valid in the present context.

Again by $N \rightarrow \infty$, we mean that all $n(j)$ ($1 \leq j \leq k$)

approach infinity in some specified manner. The C-M test is consistent against all alternatives $\{G^j\}$ for which $\int (\sum c_j(N) \sqrt{n(j)} G^j(\underline{x}))^2 dF_N(\underline{x}) \rightarrow \infty$ as $N \rightarrow \infty$; the integral test is consistent against all $\{G^j\}$ for which $\int \sum c_j(N) \sqrt{n(j)} G^j(\underline{x}) dF_N(\underline{x}) \rightarrow \infty$ as $N \rightarrow \infty$.

If $F = H$ is the true common d.f. under the null hypothesis, let $c_N(\alpha)$ and $c_N^+(\alpha)$ be defined by

$$P\left\{\int S_N^2(\underline{x}) dF_N(\underline{x}) \geq c_N(\alpha) \mid F = H\right\} = \alpha \quad \text{and}$$

$$P\left\{\int S_N(\underline{x}) dF_N(\underline{x}) \geq c_N^+(\alpha) \mid F = H\right\} = \alpha.$$

Then $\sup_{\alpha} |c_N(\alpha) - c(\alpha)| \rightarrow 0$ and $\sup_{\alpha} |c_N^+(\alpha) - c^+(\alpha)| \rightarrow 0$ as $N \rightarrow \infty$, where $c(\alpha)$ and $c^+(\alpha)$ are defined in Section 18.

If $\{G_{n(j)}^j = G_{n(j)}\}$ are k sequences of alternative d.f., the asymptotic power of the C-M and integral tests are defined in the usual way. Let $C_N(\underline{x})$ be defined as in Section 17 and suppose that $\lim_{n(j) \rightarrow \infty} \sup_{\underline{x}} |G_{n(j)}(\underline{x}) - H(\underline{x})| = 0$ for all j . Consider the following classes of alternative sequences which are k -sample analogues of (18.1):

$$(a) \mathcal{M}_{\Delta}^2, \text{ for } \Delta > 0 : \text{ all } \{G_{n(j)}\} \text{ such that}$$

$$\lim_{N \rightarrow \infty} \int C_N^2(\underline{x}) dF_N(\underline{x}) \leq \Delta;$$

(19.1)

$$(b) \mathcal{M}_{\Delta}, \text{ for } \Delta > 0 : \text{ all } \{G_{n(j)}\} \text{ such that}$$

$$\lim_{N \rightarrow \infty} \left| \int C_N(\underline{x}) dF_N(\underline{x}) \right| \leq \Delta ;$$

(c) $\mathcal{M}_\Delta^+$, for $\Delta > 0$: all $\{G_{n(j)}\} \in \mathcal{M}_\Delta$ for which
 $C_N(\underline{x}) \geq 0$ for all $\underline{x}$ and all those N involved in
 $N \rightarrow \infty$.

We obtain

Theorem 19.1.

$$U(\alpha, \mathcal{M}_\Delta^2) \leq P\{([\int B^2(\text{THD}_{\underline{x}}) dH(\underline{x})]^\frac{1}{2} + \Delta^\frac{1}{2})^2 \geq c(\alpha)\}$$

$$U(\alpha, \mathcal{M}_\Delta) = U^+(\alpha, \mathcal{M}_\Delta^+) = P\{\int B(\text{THD}_{\underline{x}}) dH(\underline{x}) \geq c^+(\alpha) - \Delta\} ;$$

$$L(\alpha, \mathcal{M}_\Delta^2) \geq P\{([\int B^2(\text{THD}_{\underline{x}}) dH(\underline{x})]^\frac{1}{2} - \Delta^\frac{1}{2})^2 \geq c(\alpha)\} ;$$

$$L^+(\alpha, \mathcal{M}_\Delta) = P\{\int B(\text{THD}_{\underline{x}}) dH(\underline{x}) \geq c^+(\alpha) + \Delta\} ;$$

$$L^+(\alpha, \mathcal{M}_\Delta^+) = \alpha .$$

The proof is the same as that for Theorems 18.1 and 18.2.

We can also obtain the k -sample analogue of Theorem 18.3:
 Suppose $H(\underline{x}) = F(\underline{x}; \underline{\theta}_0)$ and $G_{n(j)}(\underline{x}) = F(\underline{x}; \frac{\underline{\theta}^j}{n(j)})$ and assume
 the regularity conditions of Theorem 18.3 on F . Suppose that

$$\lim_{N \rightarrow \infty} \sum_{j=1}^k c_j(N) \sqrt{n(j)} \frac{\underline{\theta}^j}{n(j)} = \gamma$$

for some p -dimensional vector γ .

Theorem 19.2. Assuming the above stated conditions, expression (18.5) (resp. (18.6)) for the asymptotic power of the one-sample C-M test (resp. the integral test) is valid in this k-sample case.

The proof is the same as that of Theorem 18.3.

The k-sample versions of Examples 18.1 and 18.2 are also valid.

20. The One-sample Kolmogorov-Smirnov-type Tests for the Quantile Process.

We define the one-sided Kolmogorov-Smirnov-type test for the quantile process (Q-K-S) as: reject the null hypothesis $F = H$ at level α , if

$$\sup_{0 < t < 1} f_n^{(k)}(Q_n(t)) \sqrt{n} [Q_n(t) - H^{-1}(t)] \geq b_n^+(\alpha),$$

for large n , where $b_n^+(\alpha)$ is defined by

$$P\left\{ \sup_t f_n^{(k)}(Q_n(t)) q_n(t) \geq b_n^+(\alpha) \mid F = H \right\} = \alpha,$$

where $f_n^{(k)}(Q_n(t)) q_n(t)$ is the quantile process defined in Corollary 2B.1 (cf. (13.8) and (13.9)). The two-sided Q-K-S test is defined similarly.

Remark 20.1. In the remaining sections, we shall assume that all d.f. satisfy the conditions of Corollary 2B.1, stated in Section 13.

The two-sided Q-K-S test is consistent against all alternatives $F \neq G$ for which $g(G^{-1}(t_0)) > 0$ and $H^{-1}(t_0) - G^{-1}(t_0) \neq 0$ for some $t_0 \in (0,1)$, where $g = G'$. Similarly, the one-sided Q-K-S test is consistent against alternatives $F \neq G$ for which $g(G^{-1}(t_0)) > 0$ and $H^{-1}(t_0) - G^{-1}(t_0) < 0$ for some $t_0 \in (0,1)$. The proofs are the same as those in Remarks 16.2 and 16.3.

We note that the interval (a,b) of positivity (mentioned in Theorem 2B) for g need not be the same as that for $H' = h$.

Let $C_n^*(t) = g_n[G_n^{-1}(t)] \cdot \sqrt{n} [H^{-1}(t) - G_n^{-1}(t)]$; where $G_n' = g_n$ and consider the following classes of alternatives which are analogues of (16.1).

(a) $2\mathcal{K}_\Delta$, for $\Delta > 0$: all $\{G_n\}$ such that

$$\sup_{0 < t < 1} |C_n^*(t)| = \Delta;$$

(b) $2\mathcal{K}_\Delta(\tau)$, for $\Delta > 0$ and $0 < \tau < 1$: all $\{G_n\} \in 2\mathcal{K}_\Delta$ such that

$$(20.1) \quad \limsup_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \inf_{t \in S(\tau, \delta)} C_n^*(t) = -\Delta,$$

where $S(\tau, \delta)$ is the open interval $(\tau - \delta, \tau + \delta)$;

(c) $2\mathcal{K}_\Delta^*$, for $\Delta > 0$: $\bigcup_{0 < \tau < 1} 2\mathcal{K}_\Delta(\tau)$;

(d) $2\mathcal{K}_\Delta^+(\tau)$, for $\Delta > 0$ and $0 < \tau < 1$: all $\{G_n\} \in 2\mathcal{K}_\Delta(\tau)$ such that $C_n^*(t) \leq 0$ for all t and n sufficiently large;

$$(e) \mathcal{K}_{\Delta}^{+}(\tau), \text{ for } \Delta > 0: \bigcup_{0 < t < 1} \mathcal{K}_{\Delta}^{+}(\tau).$$

The least upper bound for the asymptotic power of the one-sided Q-K-S test, against a class $\mathcal{K}$ at level α , is defined as

$$U^{+}(\alpha, \mathcal{K}) = \sup_{\{G_n\} \in \mathcal{K}} \limsup_{n \rightarrow \infty} P\left\{ \sup_{0 < t < 1} [F_n^{(t)}(Q_n(t)) - C_n^{*}(t)] \geq b_n^{+}(\alpha) \right\}.$$

The one-sided, greatest lower bound $L^{+}(\alpha, \mathcal{K})$, the two-sided, least upper bound $U(\alpha, \mathcal{K})$ and the two-sided, greatest lower bound $L(\alpha, \mathcal{K})$ are defined in the same way. Let $b^{+}(\alpha)$ and $b(\alpha)$ be defined by

$$P\left\{ \sup_t B(t) \geq b^{+}(\alpha) \right\} = \alpha = P\left\{ \sup_t |B(t)| \geq b(\alpha) \right\}.$$

For the one-sided Q-K-S test we state

Theorem 20.1.

$$U^{+}(\alpha, \mathcal{K}) = \exp [-2(\Delta - b^{+}(\alpha))^2] \text{ for } \Delta \leq b^{+}(\alpha),$$

$$= 1 \text{ for } \Delta \geq b^{+}(\alpha),$$

where $\mathcal{K}$ is any of the classes defined by (20.1);

$$L^{+}(\alpha, \mathcal{K}_{\Delta}) = \exp [-2(\Delta + b^{+}(\alpha))^2];$$

$$(20.2) \quad L^{+}(\alpha, \mathcal{K}_{\Delta}(\tau)) = P\left\{ \sup_{0 < t < 1} [B(t) - f_{\tau}(t)] \geq b^{+}(\alpha) \right\},$$

where $f_{\tau}(t) = \Delta$, if $t \neq \tau$
 $= -\Delta$, if $t = \tau$;

$$(20.3) \quad L^+(\alpha, 2\kappa_{\Delta}^*) = \inf_{0 < \tau < 1} L^+(\alpha, 2\kappa_{\Delta}(\tau)) ;$$

$$(20.4) \quad L^+(\alpha, 2\kappa_{\Delta}^+(\tau)) = P\left\{ \sup_{0 < t < 1} [B(t) - g_{\tau}(t)] \geq b^+(\alpha) \right\} ,$$

where $g_{\tau}(t) = 0$, if $t \neq \tau$
 $= -\Delta$, if $t = \tau$;

$$(20.5) \quad L^+(\alpha, 2\kappa_{\Delta}^+) = \inf_{0 < \tau < 1} L^+(\alpha, 2\kappa_{\Delta}^+(\tau)) .$$

The proof of Theorem 20.1 will be omitted since it is similar to the proofs of Theorems 16.1 and 16.2. $F(\underline{x})$, in the definitions of the various $C_n(\underline{x})$, is replaced by t in the definitions of the $C_n^*(t)$ here. Evaluation of (20.2) through (20.5) can be found in Quade (1965).

For the two-sided Q-K-S test, we have

Theorem 20.2.

$$\begin{aligned} U(\alpha, 2\kappa_{\Delta}) &= U(\alpha, 2\kappa_{\Delta}(\tau)) = U(\alpha, 2\kappa_{\Delta}^*) \\ &= 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp[-2k^2(\Delta - b(\alpha))^2] \quad \text{for } \Delta < b(\alpha) \\ &= 1 \quad \quad \quad \text{for } \Delta \geq b(\alpha) \end{aligned}$$

$$(20.6) \quad U(\alpha, 2K_{\Delta}^{+}(\tau)) = U(\alpha, 2K_{\Delta}^{+}) \\ = 1 - P\{-b(\alpha) + \Delta \leq B(t) \leq b(\alpha), \text{ for all } t\}.$$

Let $\beta = \exp [-2(d(\alpha))^2]$; then

$$L(\alpha, 2K) \geq L^{+}(\beta, 2K) \geq L^{+}(\alpha/2, 2K),$$

where $2K$ is any of the classes defined by (20.1); and lastly we have

$$(20.7) \quad 1 - P\{-b(\alpha) + f_{\tau}(t) \leq B(t) \leq b(\alpha) + f_{\tau}(t), \text{ for all } t\} \\ \leq L(\alpha, 2K_{\Delta}(\tau)) \\ \leq L(\alpha, 2K_{\Delta}^{+}(\tau)) \\ \leq 1 - P\{-d(\alpha) \leq B(t) \leq d(\alpha) + g_{\tau}(t), \text{ for all } t\}.$$

The proof of Theorem 20.2 is similar to the proofs of Theorems 16.3 and 16.4 and hence will be omitted. Again, evaluation of the expressions in (20.6) and (20.7) can be found in Quade (1965).

21. The k-sample Kolmogorov-Smirnov-type Tests for the Quantile Process.

The k-sample Q-K-S tests are defined as: reject the null hypothesis H_0 of (12.1) at level α , if

$$\sup_{0 < t < 1} Z_n^{(r)}(t) \geq b_N^+(\alpha) \quad \text{or if}$$

$$\sup_{0 < t < 1} |Z_N^{(r)}(t)| \geq b_N(\alpha) \quad \text{for large values of the } n(j)$$

($1 \leq j \leq k$), where $b_N^+(\alpha)$ and $b_N(\alpha)$ are defined by

$$P\left\{\sup_t Z_N^{(r)}(t) \geq b_N^+(\alpha) \mid H_0\right\} = \alpha \quad \text{and} \quad P\left\{\sup_t |Z_N^{(r)}(t)| \geq b_N(\alpha) \mid H_0\right\} = \alpha,$$

and where $Z_N^{(r)}(t)$ is defined in Theorem 13.3.

Remark 21.1. The Remarks stated in Section 17 are also valid in this k-sample quantile process case. The two-sided, k-sample Q-K-S test is consistent against all alternatives $\{G_j\}_{j=1}^k$ for which

$$\left| \sum_{j=1}^k c_j(N) \sqrt{n(j)} G_j^{-1}(t_0) \right| \rightarrow \infty \quad \text{as } N \rightarrow \infty \quad \text{and} \quad g_j[G_j^{-1}(t_0)] \neq 0$$

for some t_0 , $0 < t_0 < 1$ (cf. (17.1)) while the one-sided, k-sample Q-K-S test is consistent against those alternatives $\{G_j\}_{j=1}^k$ for

$$\sum_j c_j(N) \sqrt{n(j)} G_j^{-1}(t_0) \rightarrow \infty \quad \text{as } N \rightarrow \infty \quad \text{and} \quad g_j(G_j^{-1}(t_0)) \neq 0$$

for some t_0 , $0 < t_0 < 1$ (cf. (17.2)).

As in Section 17, the asymptotic power of the k-sample Q-K-S tests against a sequence $\{G_{n(j)}\}$ of alternative d.f. depends on the summation

$$C_N^*(t) = \left(\sum_{j=1}^k c_j^2(N) g_{n(j)}[G_{n(j)}^{-1}(t)] \right) \left(\sum_{j=1}^k c_j(N) \sqrt{n(j)} G_{n(j)}^{-1}(t) \right),$$

where $G_{n(j)}' = g_{n(j)}$.

Again, we could define the classes 2δ of alternative sequences, which are k -sample versions of the classes $2K$ (cf. (20.1)); and obtain

Theorem 21.1. The bounds obtained in Theorems 20.1 and 20.2 are valid for the asymptotic power of the one and two-sided, k -sample Q-K-S tests against the classes 2δ of alternative sequences (analogues of (20.1)).

22. The One-sample Cramer-von Mises-type and Integral Tests for the Quantile Process.

The Cramér-von Mises-type test for the quantile process (Q-C-M) can be defined as: reject the null hypothesis $F = H$ at level α , if $\int [F_n^{(k)}(Q_n(t)) \sqrt{n} (Q_n(t) - H^{-1}(t))]^2 dt \geq e_n(\alpha)$ for large values of n , where $e_n(\alpha)$ is defined by $P\{\int [F_n^{(k)}(Q_n(t)) \sqrt{n} (Q_n(t) - H^{-1}(t))]^2 dt \geq e_n(\alpha) \mid F = H\} = \alpha$. The integral test and $e_n^+(\alpha)$ are defined similarly.

The Q-C-M test is consistent against all alternatives $F = G$ for which $g(G^{-1}(t_0)) > Q$ and $H^{-1}(t_0) - G^{-1}(t_0) \neq 0$ for some $t_0 \in (0,1)$; the integral test is consistent against those alternatives $F = G$ for which $\int g(G^{-1}(t)) \cdot [G^{-1}(t) - H^{-1}(t)] dt$ is strictly larger than zero. The proof is the same as that stated in Remarks 16.2 and 18.1.

For a sequence $\{G_n\}$ of alternative d.f. such that

$$\lim_{n \rightarrow \infty} \sup_{0 < t < 1} |G_n^{-1}(t) - H^{-1}(t)| = 0, \text{ let } C_n^*(t) \text{ be defined as}$$

in Section 20 and consider the following quantile versions of the class defined by (18.1):

- (a) $2\mathcal{L}_\Delta^2$, for $\Delta > 0$: all $\{G_n\}$ such that
- $$\lim_{n \rightarrow \infty} \int [C_n^*(t)]^2 dt \leq \Delta;$$
- (b) $2\mathcal{L}_\Delta^+$, for $\Delta > 0$: all $\{G_n\}$ such that
- $$(22.1) \quad \lim_{n \rightarrow \infty} \left| \int C_n^*(t) dt \right| \leq \Delta;$$
- (c) $2\mathcal{L}_\Delta^+$, for $\Delta > 0$: all $\{G_n\} \in 2\mathcal{L}_\Delta$ such that
- $$C_n^*(t) \leq 0, \text{ for all } t \in (0,1) \text{ and sufficiently large } n.$$

The least upper bound of the Q-C-M test against a class $2\mathcal{L}$ at level α is defined as:

$$U(\alpha, 2\mathcal{L}) = \sup_{\{G_n\} \in 2\mathcal{L}} \limsup_{n \rightarrow \infty} P\left\{ \int [f_n^{(t)}(Q_n(t)) \sqrt{n}(Q_n(t) - H^{-1}(t))]^2 dt \geq e_n(\alpha) \mid F = G_n \right\}.$$

The greatest lower bound $L(\alpha, 2\mathcal{L})$ for the Q-C-M test, the least upper bound $U^+(\alpha, 2\mathcal{L})$ and the great lower bound $L^+(\alpha, 2\mathcal{L})$ for the integral test are defined similarly. Define $e(\alpha)$ and $e^+(\alpha)$

by $P\{\int B^2(t)dt \geq e(\alpha)\} = \alpha$ and $P\{\int B(t)dt \geq e^+(\alpha)\}$. Note that $\int B(t)dt \sim N(0, \frac{1}{12})$ (cf. Theorem 14.2). As in Section 18, we obtain

Theorem 22.1.

$$U(\alpha, 2\epsilon_{\Delta}^2) \leq P\{([\int B^2(t)dt]^{\frac{1}{2}} + \Delta^{\frac{1}{2}})^2 \geq e(\alpha)\};$$

$$U^+(\alpha, 2\epsilon_{\Delta}^2) = U^+(\alpha, 2\epsilon_{\Delta}^{+2}) = 1 - \Phi(\sqrt{12} (e^+(\alpha) - \Delta));$$

$$L(\alpha, 2\epsilon_{\Delta}^2) = P\{([\int B^2(t)dt]^{\frac{1}{2}} - \Delta^{\frac{1}{2}})^2 \geq e(\alpha)\};$$

$$L^+(\alpha, 2\epsilon_{\Delta}^2) = 1 - \Phi(\sqrt{12} (e^+(\alpha) + \Delta));$$

$$L^+(\alpha, 2\epsilon_{\Delta}^2) = \alpha,$$

where $\Phi(\cdot)$ is the normal (0,1) d.f..

Suppose now that $F(\cdot) = F(\cdot; \underline{\theta})$ where $\underline{\theta}$ is a p -dimensional vector of parameters. Denote $((\partial/\partial\theta_1) F^{-1}(t; \underline{\theta}), \dots, (\partial/\partial\theta_p) F^{-1}(t; \underline{\theta}))$, evaluated at $\underline{\theta}_0$, by $(\partial/\partial\theta) F^{-1}(t; \underline{\theta}_0)$.

Let $H(x) = F(x; \underline{\theta}_0)$ and let $2\epsilon_{\theta}$ be the class of alternatives $\{G_n(x) = F(x; \underline{\theta}_n)\}$ defined in Section 18, where $\{\underline{\theta}_n\}$ satisfies (18.4). Assuming that $(\partial/\partial\theta) F^{-1}(t; \underline{\theta})$ is continuous in $\underline{\theta} \in \nu$ (ν defined in Section 18) uniformly in $t \in (0,1)$, then

$$\sup_t |f_n^{(t)}(Q_n(t)) \sqrt{n}(Q_n(t) - F^{-1}(t; \underline{\theta}_0)) - [n^{-\frac{1}{2}}K(t,n) + h(H^{-1}(t)) \cdot$$

$$\gamma \cdot (\partial/\partial\theta) F^{-1}(t; \underline{\theta}_0)]| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty.$$

under the sequence of alternatives $\{G_n\}$, where $K(t,n)$ is the Kiefer process of Corollary 2B.1, stated in Section 13. Analogous to Theorem 18.3, we obtain

Theorem 22.2. Assume that $(\partial/\partial\theta) F^{-1}(t; \theta)$ is continuous in $\theta \in \nu$ uniformly in $t \in (0,1)$. Under any sequence $\{G_n(x) = F(x; \theta_n)\} \in \mathcal{L}_{\theta_0}$, with $H = F(\cdot; \theta_0)$, we have

$$\begin{aligned} & P\{([\int B^2(t)dt]^{\frac{1}{2}} + \Delta^{\frac{1}{2}})^2 \geq e(\alpha)\} \\ & \geq P\{[\int B(t) - h(H^{-1}(t)) \gamma \cdot (\partial/\partial\theta) F^{-1}(t; \theta_0)]^2 dt \geq e(\alpha)\} \\ & = \lim_{n \rightarrow \infty} P\{[\int f_n^{(t)}(Q(t)) \sqrt{n}(Q_n(t) - H^{-1}(t))]^2 dt \geq e_n(\alpha) \mid F = G_n\} \\ & \geq P\{([\int B^2(t)dt]^{\frac{1}{2}} - \Delta^{\frac{1}{2}})^2 \geq e(\alpha)\}, \end{aligned}$$

where $\int [h(H^{-1}(t)) \gamma \cdot (\partial/\partial\theta) F^{-1}(t; \theta_0)]^2 dt = \Delta$.

For the integral test, we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} P\{[\int f_n^{(t)}(Q(t)) \sqrt{n}(Q_n(t) - H^{-1}(t))]^2 dt \geq e_n(\alpha) \mid F = G_n\} \\ & = 1 - \phi(e(\alpha) - \int h(H^{-1}(t)) \gamma \cdot (\partial/\partial\theta) F^{-1}(t; \theta_0) dt), \end{aligned}$$

where $\phi(\cdot)$ is the standard $N(0,1)$ d.f.

23. The k-sample Cramer-von Mises-type and Integral Tests for the Quantile Process.

The k-sample Q-C-M and integral tests, as defined in Corollary 13.3.2, can be used to test the homogeneous null hypothesis H_0 of (12.1). The k-sample Q-C-M test is consistent against those alternatives $\{G_j\}_{j=1}^k$ for which $\int (\sum_j c_j^2(N) g_j[G_j^{-1}(t)])^2 (\sum_j c_j(N) \sqrt{n(j)} G_j^{-1}(t))^2 dt \rightarrow \infty$; the k-sample integral test is consistent against those $\{G_j\}$ for which

$$\int (\sum_j c_j^2(N) g_j[G_j^{-1}(t)]) (\sum_j c_j(N) \sqrt{n(j)} G_j^{-1}(t)) dt \rightarrow \infty \text{ as } N \rightarrow \infty.$$

One can define the classes $2\mathcal{M}$ of alternative sequences for the k-sample quantile process which would be analogous to the classes $\mathcal{M}$ of Section 19. The corresponding results of Theorems 22.1 and 22.2 can be proved for the classes $2\mathcal{M}_\Delta^2$, $2\mathcal{M}_\Delta$, $2\mathcal{M}_\Delta^+$ and $2\mathcal{M}_\emptyset$.

CONCLUDING REMARKS

The central problem of this dissertation is the strong approximation of empirical processes by appropriate Gaussian processes and applications of these to the asymptotic theory of goodness-of-fit tests for composite hypotheses in the one and several sample cases.

The results of Chapters II and III are extensions of and elaborations on the strong approximation methodology as developed in the works of Csörgő, Komlós, Major, Révész and Tusnády cited in the References. In [14], Csörgő, Komlós, Major, Révész and Tusnády have introduced a new methodology to tackle the problem of obtaining almost-sure representations of the univariate empirical process, when one parameter of the underlying distribution function is estimated by a sequence of maximum likelihood estimators. Their new methodology hinges on strong approximations of the univariate empirical process. The first of these was given by Kiefer (1972), followed by Csörgő and Révész (1975) and Komlós, Major and Tusnády (1975), the latter containing a best and an almost best rate of convergence. As to estimating a parameter, [14] uses results of Ibragimov and Has'minskiĭ (1973). The multivariate version of Kiefer's 1972 result was proved by Csörgő and Révész [11] and it was achieved via using a different construction and a different method of proof than those of the available one-dimensional results ([35], [37]) and germinated from the Csörgő-Révész (1975) paper. Our Chapter II uses the multivariate strong approximation result of Csörgő and Révész [11] and our multivariate extension of the Ibragimov-Has'minskiĭ (1973) theorem (cf. Section 4) and it further develops the methodology of [14]. Our results are believed to be new in both the one and multi-dimensional cases.

Chapter III is a new approach, which combines the methods used in [14] and Chapter II with Durbin (1973a)'s more general (than maximum likelihood) class of estimators. The results gained here are again new, even in the

univariate case. As to the univariate X_1 and multi-dimensional θ type results of Chapter II and III, they could have been proved already after Kiefer's 1972 paper, but the first attempt only came after Durbin's 1973 paper, while the multivariate X_1 and θ type results in these chapters could only have come after [11].

The k-sample empirical process has been frequently studied using weak convergence methods. Our almost-sure representations, with rates of convergence, in Chapter IV are again believed to be new.

The results of Chapters II, III and IV are not proved for the sole purpose of using them in Chapter V. Hopefully they are going to have many other possible applications. Chapter V is intended to be a "first statistical study" of the power properties of the one and k-sample tests using the strong approximation methodology. In the multivariate and the quantile cases, the results are believed to be new. The multivariate results obtained could have possibly been done using weak convergence methods (e.g. Neuhaus (1971) in the Additional References), but to our knowledge, they have not been. It appears that such an approach would employ cumbersome proofs (e.g. proving "tightness"). I feel that the proofs in Chapter V, using strong representations, are straightforward and natural from an aesthetic point of view, although their full strength is not needed to achieve this.

It should also be noted that the entire investigation of the quantile process, namely when parameters are estimated, the k-sample problem and power studies (Sections 6, 10, 13, 20, 21, 22 and 23 of this thesis), is believed to be new and are based on the new methodology of Csörgő and Révész in [12].

Concerning the nomenclature "Kiefer process": (cf. (b) of page 1), Professor P.J. Bickel has kindly pointed out that Pyke (1968) and Bickel and Wichura (1971) (cited in the Additional References) studied the process

which we call the Kiefer process in a weak convergence context. It appears to us that Pyke only used the Brownian Bridge part of it, in the sense of Dudley (1966). Also, Müller (1970) proved that the univariate empirical process converges weakly to this Kiefer process. The name "Kiefer process" originated with Csörgő and Révész (1975). The nomenclature is appropriate because Kiefer (1972) was the first to construct this process in his Skorohod-type embedding of the univariate empirical process in order to prove a Strassen-type law of strong invariance principle in terms of it. Also, Kiefer (1970) (in the Additional References) lists this latter problem as an open one.

I wish to thank Professor Bickel for calling to my attention the Čibisov (1965) paper, which made an extensive study of goodness-of-fit tests up to that time, and also to Brillinger (1969) and Breiman (1968), who obtained in-probability representations of the univariate empirical process. I was aware of the latter two, and it was not my intention to ignore their important contributions on representations of the empirical process; this dissertation germinated from the 1975 results stated above (in which they are prominently mentioned) and from [14]. In the References, an attempt was made to include all those papers whose results have directly or indirectly influenced my work, and I did not attempt to include all the other important results. The above papers are cited in the Additional References.

For more recent studies concerning empirical density functions, we refer to Rosenblatt (1971) (suggested by Professor Bickel) and to Wegman (1972a and b) (suggested by Professor B.B. Winter).

Certain areas of possible future investigation arise from this dissertation. A study and tabulation of the asymptotic distribution of statistics based on the estimated empirical and quantile processes, using the representations obtained in our Chapters II and III, require further

significant breakthroughs to get any handle on the statistical problems they generate. It would be desirable to develop a procedure, for example, to obtain, in a unified way using these representations, exact distributions (if possible) or numerical tables for the Kolmogorov-Smirnov and Cramér-von Mises functionals of them.

The problem of computing statistics such as the Kolmogorov-Smirnov and Cramér-von Mises of the multivariate empirical process is still open, even when the statistics are asymptotically distribution-free under the null hypothesis.

A comparison of the univariate empirical process tests versus the quantile process tests, discussed in Chapter V, needs to be done. Further power studies, based on Chapter V, are needed. I wish to thank Professor Bickel for these latter two suggestions.

REFERENCES

- [1] ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. Ann. Math. Statist. 23 193-212.
- [2] BICKEL, P. J. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. Ann. Math. Statist. 40 1-23.
- [3] BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. Ann. Statist. 1 1071-1095. Errata (1975) 3 1370.
- [4] BILLINGSLEY, P. (1968). Convergence of Probability Measures. John Wiley, New York.
- [5] BURKE, M. D. and CSÖRGÖ, M. Strong approximations of the estimated quantile and multivariate empirical processes. (submitted for publication).
- [6] BURKE, M. D. and CSÖRGÖ, M. Weak approximations of the estimated sample quantile and multivariate empirical processes via strong approximation methods. (submitted for publication).
- [7] BURKE, M. D. and CSÖRGÖ, M. Strong approximations of the k-sample quantile and multivariate empirical processes. (submitted for publication).

- [8] BURKE, M. D. On the asymptotic power of some statistics based on the one and several sample quantile and multivariate empirical processes. (submitted for publication).
- [9] BURKE, M. D. (1975). Weak convergence of the multivariate empirical process when parameters are estimated. Technical Report. Carleton Mathematical Series No. 126.
- [10] CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton Univ. Press, Princeton N.J..
- [11] CSÖRGÖ, M. and RÉVÉSZ, P. A strong approximation of the multivariate empirical process. Studia Sci. Math. Hungar. (to appear).
- [12] CSÖRGÖ, M. and RÉVÉSZ, P. Strong approximations of the quantile process. Ann. Statist. (to appear).
- [13] CSÖRGÖ, M. and RÉVÉSZ, P. (1975). A new method to prove Strassen type laws of invariance principle. II. Z. Wahrscheinlichkeitstheorie verw. Gebiete 31 261-269.
- [14] CSÖRGÖ, M., KOMLÓS, J., MAJOR, P., RÉVÉSZ, P. and TUSNÁDY, G. On the empirical process when parameters are estimated. Proc. Seventh Prague Conf., 1974. (to appear).
- [15] CSÖRGÖ, M. and BURKE, M. D. Weak approximations of the empirical process when parameters are estimated. Contributions to

Empirical Distributions and Multivariate Empirical Processes. Lecture Notes in Mathematics. Springer-Verlag.
(to appear).

- [16] DARLING, D. A. (1955). The Cramér-Smirnov test in the parametric case. Ann. Math. Statist. 26 1-20.
- [17] DONSKER, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statist. 23 277-281.
- [18] DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statist. 20 393-403.
- [19] DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. Illinois J. Math. 10 109-126.
- [20] DURBIN, J. (1970). Asymptotic distributions of some statistics based on the bivariate sample distribution function. Nonparametric Techniques in Statistical Inference, edited by M. L. Puri. Cambridge Univ. Press. 435-451.
- [21] DURBIN, J. (1973a). Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. 1 279-290.

- [22] DURBIN, J. (1973b). Distribution Theory for Tests Based on the Sample Distribution Function. Regional Conference Series in Applied Math. 9 S.I.A.M., Philadelphia.
- [23] DURBIN, J. (1975). Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika 62 5-22.
- [24] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. 27 642-669.
- [25] GIEMAN, I. I. (1953a). On some limit theorems for conditional distributions and on problems of mathematical statistics connected with them. Ukrain. Math. J. 5 413-433.
English: Sel. Transl. in Math. Statist. Prob. (IMS-AMS) 2 5-26.
- [26] GIEMAN, I. I. (1953b). Some notes to the goodness-of-fit criteria of A. N. Kolmogorov. Doklady (Lectures) of the Academy of Sciences of USSR. 91 715-718. (In Russian).
- [27] GIEMAN, I. I. (1954). On a use of a theory of ω^2 criteria. Kiev Scientific Notes of the University. Math. Sbornik 8 51-60. (In Ukrainian).

- [28] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1972). Asymptotic behavior of statistical estimators in the smooth case. Theor. Probability Appl. 17 445-462.
- [29] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1973). On the approximation of statistical estimators by sums of independent variables. Soviet Math. Dokl. 14 883-887.
- [30] KAC, M. (1951). On some connections between probability theory and differential and integral equations. Second Berkeley Symp. Prob. Statist., Univ. of California Press. 180-215.
- [31] 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.
- [32] KIEFER, J. (1959). k-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests. Ann. Math. Statist. 30 420-447.
- [33] KIEFER, J. (1961). On large deviations of the empirical d.f. of vector chance variables and a law of the iterated logarithm. Pacific J. Math. 11 649-660.
- [34] KIEFER, J. (1970). Deviations between the sample quantile process and the sample DE. Nonparametric Techniques in Statistical Inference, ed. M.L. Puri, Cambridge Univ. Press. 299-311

- [35] KIEFER, J. (1972). Skorohod embedding of multivariate R.V.'s and the sample D.F.. Z. Wahrscheinlichkeitstheorie verw. Gebiete 24 1-35.
- [36] KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. Giorn. Ist. Ital. Attuari. 4 83-91.
- [37] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and the sample D.F. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32 111-131.
- [38] LOÈVE, M. (1963). Probability Theory. 3rd edition, van Nostrand-Reinhold, Princeton, N.J..
- [39] QUADE, D. (1965). On the asymptotic power of the one-sample Kolmogorov-Smirnov tests. Ann. Math. Statist. 36 1000-1018.
- [40] RAO, K. C. (1972). The Kolmogoroff, Cramér-von Mises, chisquare statistics for goodness-of-fit tests in the parametric case. Abstract 133-6. Bulletin I.M.S. 1 87.
- [41] RÉNYI, A. (1970). Foundations of Probability. Holden-Day, San Francisco, Calif...
- [42] RÉVÉSZ, P. (1972). On empirical density functions. Period. Math. Hungar. 2 82-110.

- [43] RÉVÉSZ, P. On multivariate empirical density functions.
(to appear).
- [44] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. Ann. Math. Statist. 23 470-472.
- [45] ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of density functions. Ann. Math. Statist. 27 832-837.
- [46] SMIRNOV, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bul. Math. de l'Universite de Moscou 2 fasc. 2.
- [47] STEPHENS, M. A. (1974). EDF statistics for goodness-of-fit and some comparisons. J. Amer. Statist. Assoc. 69 730-737.
- [48] STEPHENS, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. Ann. Statist. 4 357-369.

ADDITIONAL REFERENCES

- [49] BICKEL, P.J. and WICHURA, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. 42 1656-1670.
- [50] BREIMAN, L. (1968). Probability, Addison-Wesley.
- [51] BRILLINGER, D.R. (1969). An asymptotic representation of the sample df. Bull. AMS 75 545-547.
- [52] CIBISOV, D.M. (1965). An investigation of the asymptotic power of the tests of fit. Theor. Probability Appl. 10 No. 2 421-437.
- [53] KIEFER, J. (1970). Old and new methods for studying order statistics and sample quantiles. Nonparametric Techniques in Statistical Inference, ed. M.L. Puri, Cambridge Univ. Press. 349-357.
- [54] MULLER, D.W. (1970). On Glivenko-Cantelli convergence. Z. Wahrscheinlichkeitstheorie verw. Gebiete. 16 196-210.
- [55] NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. Ann. Math. Statist. 42 1285-1295.
- [56] PYKE, R. (1968). Weak convergence of the empirical process with random sample size. Proc. Cambridge Philos. Soc. 64 155-160.
- [57] ROSENBLATT, M. (1971). Curve estimates. Ann. Math. Statist. 42 1815-1842.
- [58] WEGMAN, E.J. (1972a). Nonparametric probability density estimation: I. A summary of available methods. Technometrics 14 533-546.
- [59] WEGMAN, E.J. (1972b). Nonparametric probability density estimation: II. A comparison of density estimation methods. J. Statist. Computation and Computer Simulation. 1 225-245.

END

10

103

77

FIN