Some Kolmogorov - Smirnov - Rényi Type Theorems of Probability

by

Miklos Csorgo, M.A.

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics,
McGill University,
Montreal.

May 1963.

ACKNOWLEDGEMENT

I am greatly indebted to Professor
W.A.O'N. Waugh for his help and criticism.

Table of Contents

Introduction and Summary.

In his fundamental paper (1953) [23], entitled "On the Theory of Order Statistics", A. Rényi developed a new method by means of which many important results of the theory of order statistics can be obtained with surprising simplicity. The essential novelty of his method is that it reduces the problems connected with order statistics to the study of sums of mutually independent random variables. Chapter 1 of this thesis contains a review of this method.

The above mentioned method has also enabled Rényi to give an interesting improvement of the Kolmogorov - Smirnov theorems. Let $F_n(x)$ denote the distribution function of a sample of size n drawn from a population having continuous distribution function $F(x)$. Kolmogorov [16] determined the limiting distribution of the supremum of $|F_n(x) - F(x)|$ and Smirnov [24] did the same for $F_n(x) - F(x)$. But it may be more significant to measure the relative deviation of $F_n(x)$ from $F(x)$; for example, if $F(x)$ is small, it may be more important to know something about $\sup |F_n(x) - F(x)| /F(x)$ and $\sup \{F_n(x) - F(x)\}/F(x)$ than the above mentioned deviations of the Kolmogorov - Smirnov theorems. Using his method and some generalized results of P. Erdős and M. Kac [5], A. Rényi has determined the limiting distribution of the supremum of the relative deviations $|F_n(x) - F(x)| /F(x)$ and $F_n(x) - F(x) /F(x)$ respectively. His results and the results of P. Erdős and M. Kac generalized by him are given in

Chapter 2 of this thesis (Theorems 1, ..., 8).

Introducing the weight function $1/F(x)$ in the Kolmogorov - Smirnov theorems we characterize the asymptotic behaviour of the relative deviation of the population distribution function and that of the sample. The theorems provide tests for verifying the hypothesis that a random sample of size n with empirical distribution function $F_n(x)$ has been drawn from a population having continuous distribution function $F(x)$. At the same time, as far as statistical considerations are concerned, we lose one of the convenient properties of the Kolmogorov - Smirnov theorems. Namely, we can use these theorems to construct confidence intervals for an unknown continuous distribution function $F(x)$. Having the limiting distribution of $\sup |F_n(x) - F(x)| / F(x)$ instead, we no longer have that property. We could retain the advantage of these new theorems that they measure relative deviation of the population distribution function and that of the sample and could regain the above mentioned confidence interval property of the old theorems if we used $1/F_n(x)$ instead of $1/F(x)$ as weight function. Thus in this way the idea arises of considering the limiting distribution of the quotients $\{F_n(x) - F(x)\} / F_n(x)$ and $|F_n(x) - F(x)| / F_n(x)$ respectively. One of the major objectives of this thesis was the derivation of these distributions using the method of A. Rényi and our results in this connection are given and proved in Chapter 3 of this thesis (Theorems 9, ..., 12).

Let $F_n(x)$ and $H_m(x)$ be the empirical distribution functions of two random samples of size n and m respectively from a population having continuous distribution function $F(x)$. Smirnov [25] determined the limiting distribution of the supremum of $|F_n(x) - H_m(x)|$ and $F_n(x) - H_m(x)$ respectively. Again, it may be more significant to measure the relative deviation of these two empirical distributions $F_n(x)$ and $H_m(x)$. Thus in this way the idea arises of considering the limiting distribution of the quotients $\{F_n(x) - H_m(x)\}/F(x)$ and $|F_n(x) - H_m(x)|\,/F(x)$ respectively. The second major objective of this thesis was the derivation of these distributions using the method of A. Rényi , some of the results of P. Erdős and M. Kac and 2 lemmas of our own. Our results in this connection are given and proved in Chapters 4 and 5 of this thesis (Theorems 13, ..., 18 and Lemmas 1 and 2).

I would like to express my sincere gratitude to Professor W.A.O'N. Waugh who supervised this work while he was at McGill University and special thanks are due to him for continuing to do so after leaving for a position with the University of Hull, England. His patient encouragement was invaluable in the execution of this thesis, his comments and criticisms led to many corrections and improvements. I also take this opportunity to thank him for his general guidance and help during my graduate years at McGill University. My thanks also go to Professors A. Evans (McGill), I. Guttman (formerly at McGill, presently at the University of Wisconsin),

A. Joffe (formerly at McGill, presently at the University of Montreal), Z.A. Melzak (formerly at McGill, presently at the University of British Columbia) for accasional but valuable discussions. I am also grateful to the audience of the Probability and Statistics Seminar of McGill University and University of Montreal for giving me the opportunity of talking three times on some of the results of this thesis in the last two years. This thesis has been written with the support of a National Research Council of Canada Studentship.

1. Presentation of A. Rényi's method in the theory of order statistics.

Consider the following special case : ζ is a random variable distributed according to the exponential law. That is ζ has the probability density function (written as p.d.f. from now on) : $f(x) = \lambda e^{- x}$ if $x > 0$, zero otherwise and cumulative density function (written as c.d.f. from now on) $F(x) = 1 - e^{-\lambda x}$ if $x \geq 0$, zero otherwise for $\lambda > 0$. Take a random sample of size n on ζ ; i.e. we have $\zeta_1, \zeta_2, \ldots, \zeta_n$ as mutually independent random variables with the same exponential distribution function. We shall need the following property of the exponential distribution : if ζ is an exponentially distributed random variable then, if $x > 0$ and $y \geq 0$, we have

$$P(\zeta < x+y \mid \zeta \geq y) = P(\zeta < x) \tag{1.1}$$

To verify (1.1) we have :

$$\begin{aligned} P(\zeta < x+y \mid \zeta \geq y) &= P(\zeta < x+y, \zeta \geq y)/P(\zeta \geq y) \\ &= P(y \leq \zeta < x+y)/P(\zeta \geq y) \\ &= \left\{ P(\zeta < x+y) - P(\zeta < y) \right\} / P(\zeta \geq y) \\ &= \left\{ F(x+y) - F(y) \right\} / \left\{ 1 - F(y) \right\} \end{aligned}$$

If we use the relation now that $F(x) = 1 - e^{- x}$, $x \geq 0$, $\lambda > 0$, then we get

$$\begin{aligned} P(\zeta < x+y \mid \zeta \geq y) &= \left\{ 1 - e^{-\lambda(x+y)} - (1 - e^{-\lambda y}) \right\} / e^{-\lambda y} \\ &= 1 - e^{-\lambda x} \\ &= P(\zeta < x), \text{ which was to be proved.} \end{aligned}$$

The converse of this statement is also true, that is property (1.1) uniquely characterizes the exponential distribution and it can be, therefore, used to derive certain distribution properties of a random sample taken on the random variable ζ distributed according to the exponential law.

To show converse of this statement, we have that (1.1) is equivalent to

$$\{F(x+y) - F(y)\}/\{1 - F(y)\} = F(x)$$

so we have

$$1 - \{F(x+y) - F(y)\}/\{1 - F(y)\} = 1 - F(x),$$

which is equivalent to the relation

$$\phi(x+y) = \phi(x)\,\phi(y), \tag{1.2}$$

where $\phi(x) = 1 - F(x)$ and it is known that, except for the trivial cases $\phi(x) = 0$ and $\phi(x) = 1$, the monotonic non-increasing functions which uniquely satisfy (1.2) have the form $\phi(x) = e^{-\lambda x}$, $x \geq 0$, $\lambda > 0$; i.e. $F(x) = 1 - e^{-\lambda x}$ as a consequence of (1.1).

The meaning of (1.1) becomes especially clear if we interpret the random variable ζ as the duration of time of the occurence of a random event. In this interpretation proposition (1.1) can be formulated as follows : if the waiting time for the occurence of the random event is distributed according to the exponential law and if we are given that the waiting time has not yet terminated at time y, then the duration of the further waiting time to the occurence is independent of y; i.e. of the waiting time that has already elapsed.

Going back to our random sample of size n on ζ, let us arrange the numbers $\zeta_1, \zeta_2, \ldots, \zeta_n$ in order of magnitude and use the notation

(1.3) $$\zeta_{(k)} = R_k(\zeta_1, \zeta_2, \ldots, \zeta_n) , \quad k = 1, 2, \ldots, n$$

where the function $R_k(X_1, X_2, \ldots, X_n)$ of the n variables $X_1, X_2, \ldots, X_n$ denotes the k^{th} of the values $X_1, X_2, \ldots, X_n$ in order of magnitude $(k = 1, 2, \ldots, n)$; thus e.g. $\zeta_{(1)} = \min\limits_{1 \le k \le n} \zeta_k$ and $\zeta_{(n)} = \max\limits_{1 \le k \le n} \zeta_k$. Using relation (1.1), the individual and joint distribution of the random variables of the order statistics $\zeta_{(1)} < \zeta_{(2)} < \ldots < \zeta_{(n)}$ can be easily determined. For this purpose we interpret the random variables ζ_k as random waiting times for the occurence of mutually independent random events. (This is going to be done for the sake of bye-passing lengthy analytical proofs.) Then $\zeta_{(k)}$ denotes the duration of time of the occurence of the random event finished as k^{th} of the n observations (k^{th} longest duration of time $= \zeta_{(k)}$).

We determine first the distribution of $\zeta_{(k+1)} - \zeta_{(k)}$. If we are given $\zeta_{(k)} = y$, then

(1.4) $$P(\zeta_{(k+1)} - \zeta_{(k)} > x \mid \zeta_{(k)} = y) = P(\zeta_{(k+1)} > x+y \mid \zeta_{(k)} = y),$$

where on the right side there stands the probability of the event that none of the $n-k$ happenings, being in progress at the moment y, finishes until the moment $x+y$. By virtue of (1.1), the value of this probability is

$$\{P(\zeta > x)\}^{n-k} = \{e^{-\lambda x}\}^{n-k} = e^{-(n-k)\lambda x}$$

and thus the conditional distribution function of $\zeta_{(k+1)} - \zeta_{(k)}$ with respect to the condition that $\zeta_{(k)} = y$ is

$$(1.5) \qquad P(\zeta_{(k+1)} - \zeta_{(k)} < x \mid \zeta_{(k)} = y) = 1 - e^{-(n-k)\lambda x} .$$

As (1.5) does not depend on y, it also gives the non-conditional distribution function of $\zeta_{(k+1)} - \zeta_{(k)}$. Indeed, by virtue of the theorem on total probability, we have

$$(1.6) \quad P(\zeta_{(k+1)} - \zeta_{(k)} < x) = \int_0^\infty P(\zeta_{(k+1)} - \zeta_{(k)} < x \mid \zeta_{(k)} = y) dP(\zeta_{(k)} < y)$$

$$= (1 - e^{-(n-k)\lambda x}) \int_0^\infty dP(\zeta_{(k)} < y)$$

$$= 1 - e^{-(n-k)\lambda x} .$$

Therefore the differences $\zeta_{(k+1)} - \zeta_{(k)}$ are themselves exponentially distributed with the mean value $\frac{1}{(n-k)\lambda}$ and thus the random variables

$$(1.7) \qquad \delta_{k+1} = (n-k)\,(\zeta_{(k+1)} - \zeta_{(k)}), \quad k = 0, 1, \ldots, n-1$$

are also exponentially distributed with the mean value $\frac{1}{\lambda}$.

(In the above relation $\zeta_{(0)} = 0$ by definition.)

It also follows from what has been said above that the variables $\delta_1, \delta_2, \ldots, \delta_n$ are mutually independent random variables. It is, namely, easy to see that the following relation holds

$$(1.8) \quad P(\zeta_{(k+1)} - \zeta_{(k)} < x \mid \zeta_{(1)} = y_1, \zeta_{(2)} - \zeta_{(1)} = y_2, \ldots, \zeta_{(k)} - \zeta_{(k-1)} = y_k)$$

$$= P(\zeta_{(k+1)} - \zeta_{(k)} < x), \quad k = 0, 1, \ldots, n-1 .$$

This is evident, as the above conditions in (1.8) mean that $\zeta_{(1)} = y_1$, $\zeta_{(2)} = y_1 + y_2$, ..., $\zeta_{(k)} = y_1 + y_2 + \dots + y_k$; i.e. they give the finishing instants of the first k happenings of the n observations which started simultaneously at the moment $t = 0$. These conditions imply that at the moment $t = y_1 + y_2 + \dots + y_k$ there are still $(n-k)$ waiting times incompleted and the probability of the finishing of at least one of them before the moment $t + x$ is given by (1.6). So we have the relation of (1.8) which in itself is a necessary and sufficient condition for the random variables $\zeta_{(k+1)} - \zeta_{(k)}$ (and therefore for $\delta_{k+1} = (n-k)(\zeta_{(k+1)} - \zeta_{(k)})$), $k = 0, 1, \dots, n-1$, to be mutually independent.

Using (1.7) the random variables $\zeta_{(k)}$ can be expressed in the form

$$(1.9) \qquad \zeta_{(k)} = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \dots + \frac{\delta_k}{n-k+1}, \quad k = 1, 2, \dots, n,$$

i.e. $\zeta_{(k)}$ can be expressed as summs of multiples of sequences of mutually independent, identically distributed random variables. An alternative way of saying this is : the random variables $\zeta_{(k)}$ form an additive Markov Chain. By virtue of (1.9) the distribution of any $\zeta_{(k)}$, further, the joint ditribution of any number of the random variables $\zeta_{(k)}$ can be determined in explicit form.

<u>The above method can be applied in general to the study of order statistics</u>. To show this, let ξ be any random variable having a continuous c.d.f. $F(x)$. Let $\xi_1, \xi_2, \dots, \xi_n$

be a random sample of size n on ξ; i.e. $\xi_1, \xi_2, \ldots, \xi_n$ are mutually independent random variables with the same c.d.f. $F(x)$. Let $\xi_{(1)} < \xi_{(2)} < \ldots < \xi_{(n)}$ be the order statistics based on the above sample, that is to say, we form the new random variables $\xi_{(k)} = R_k(\xi_1, \xi_2, \ldots, \xi_n)$.

The study of the random variables $\xi_{(k)}$ can be reduced to the special case where the random variables ξ_k are exponentially distributed (and therefore - by virtue of (1.9) - to the study of sums of mutually independent random variables), as follows : let us put

(1.10) $$\eta_k = F(\xi_k) \quad \text{and} \quad \zeta_k = \log \frac{1}{\eta_k}, \quad k = 1, 2, \ldots, n.$$

Then $\eta_{(k)} = F(\xi_{(k)})$ is the k^{th} of the random variables $\eta_1, \eta_2, \ldots, \eta_n$, i.e. $\eta_{(k)} = R_k(\eta_1, \eta_2, \ldots, \eta_n)$. Further, let us put

(1.11) $$\zeta_{(k)} = \log \frac{1}{\eta_{(n+1-k)}}, \quad k = 1, 2, \ldots, n.$$

As $\log \frac{1}{x}$ is a steadily decreasing function, we obtain

(1.12) $$\zeta_{(k)} = R_k(\zeta_1, \zeta_2, \ldots, \zeta_n), \quad k = 1, 2, \ldots, n,$$

i.e. $\zeta_{(k)}$ is the k^{th} of the random variables $\zeta_{(1)}, \zeta_{(2)}, \ldots, \zeta_{(n)}$ in order of magnitude. As we have assumed the variables ξ_k to be mutually independent it follows that $\zeta_1, \zeta_2, \ldots, \zeta_n$ are also mutually independent.

We want to show now that the random variables $\zeta_1, \zeta_2, \ldots, \zeta_n$ are distributed according to the exponential law. Let us investigate the distribution of the single random variable $\zeta_k = \log \frac{1}{\eta_k}$. Let $x = F^{-1}(y)$ be the inverse function of $y = F(x)$, $0 \leq y \leq 1$; thus we have

$$P(\zeta_k < x) = P(\log \frac{1}{F(\xi_k)} < x) = P(\xi_k > F^{-1}(e^{-x})), \quad \text{for } x \geq 0$$

i.e. $P(\zeta_k < x) = 1 - P(\xi_k < F^{-1}(e^{-x})) = 1 - F(F^{-1}(e^{-x}))$

$$= 1 - e^{-x}, \quad x \geq 0 .$$

Therefore $\zeta_1, \zeta_2, \ldots, \zeta_n$ are mutually independent, identically distributed, exponential random variables with common c.d.f. $1 - e^{-x}$, $x \geq 0$ and mean value 1.

In this way the random variables $\xi_{(k)}$ themselves can be expressed in the form

$$(1.13) \quad \xi_{(k)} = F^{-1}(e^{-\zeta_{(n+1-k)}}) = F^{-1}(e^{-(\frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ldots + \frac{\delta_{n+1-k}}{k})}),$$

$k = 1, 2, \ldots, n$, for

$\zeta_{(k)} = \log \frac{1}{F(\xi_{(n+1-k)})}$ implies that $\xi_{(n+1-k)} = F^{-1}(e^{-\zeta_{(k)}})$, therefore, by (1.9),

$$\xi_{(k)} = F^{-1}(e^{-\zeta_{(n+1-k)}}) = F^{-1}(e^{-(\frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ldots + \frac{\delta_{n+1-k}}{k})}),$$

where the random variables $\delta_1, \delta_2, \ldots, \delta_n$ are defined as in (1.7) and are mutually independent exponentially distributed random variables with c.d.f. $1 - e^{-x}$ $(x \geq 0)$ and mean value 1.

A consequence of (1.13) is that the random variables $\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(n)}$ form a Markov Chain. To show this, let us start with the relation

$$\frac{F(\xi_{(k+1)})}{F(\xi_{(k)})} = \frac{\eta_{(k+1)}}{\eta_{(k)}} = \frac{e^{-\zeta_{(n-k)}}}{e^{-\zeta_{(n+1-k)}}} = e^{\zeta_{(n+1-k)} - \zeta_{(n-k)}}$$

Put $\rho_k = e^{\zeta(n+1-k) - \zeta(n-k)}$, $k = 1, 2, \ldots, n$, where

$e^{-\zeta(0)} = \eta_{(n+1)} = 1$.

The random variables ρ_k are mutually independent since we have already seen that the differences $\zeta_{(n+1-k)} - \zeta_{(n+1)} = \frac{\delta_{(n+1-k)}}{k}$ are mutually independent. Let us also put $\phi(x, y) = F^{-1}(yF(x))$. Then, by (1.13) we have

$$\xi_{(k+1)} = F^{-1}(e^{-\zeta(n-k)}) = F^{-1}(e^{-\zeta(n+1-k)}\, e^{\zeta(n+1-k) - \zeta(n-k)})$$

$$= F^{-1}(F(\xi_{(k)})\rho_k) = \phi(\xi_{(k)}, \rho_k)\ ,\ \text{~~where~~}\ \xi_{(k)}$$

and ρ_k are independent random variables. This ~~implies~~ that the random variables $\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(n)}$ form a Markov Chain. For we have :

Theorem. Let $\xi_{n+1} = \phi(\xi_n, \rho_n)$, ~~where~~ $\phi(x, y)$ is an arbitrary two-variate continuous function; further ρ_n is independent of the random variables $\xi_1, \xi_2, \ldots, \xi_n$ $(n = 1, 2, \ldots)$. Then the random variables ξ_n form a Markov chain [22].

A.N. Kolmorgorov [16] was the first who remarked that the random variables $\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(n)}$, i.e. a sequence of order statistics, form a Markov Chain. A. Rényi's method presented here starts from this fundamental observation, but the possibilities implied by it could be developed only after having transformed the Markov Chain $\{\xi_{(k)}\}$ into an additive Markov Chain by means of the transformation

$$\zeta_{(k)} = \log \frac{1}{F(\xi(n+1-k))}\ .$$

In his paper [23] A. Rényi uses this method to prove the following theorem of Malmquist [19]

Theorem. The random variables $\left(\frac{\eta_{(k)}}{\eta_{(k+1)}}\right)^k$, $k = 1, \ldots, n$, are mutually independent and have the same uniform distribution in the interval $(0, 1)$.

He also builds up the theory of order statistics by means of the method of this section. Then he proves the theorems which we are going to present in the next section of this dissertation.

2. A. Rényi's theorems. Improvements of the Kolmogorov - Smirnov theorems.

Keeping the notation and assumptions of section 1 about the random variable ξ, let us define

$$(2.1) \qquad F_n(x) = \begin{cases} 0 , & \text{if } \; x < \xi_{(1)} \\ \frac{k}{n} , & \text{if } \; \xi_{(k)} \leq x < \xi_{(k+1)} \\ 1 , & \text{if } \; \xi_{(n)} \leq x \end{cases} ,$$

i.e. $F_n(x)$ is the distribution function of the sample, in other words, the frequency ratio of the values less than x in the sample.

A.N. Kolmogorov [16] proved a fundamental theorem giving a test for the hypothesis that a sample has been drawn from a population having a given continuous c.d.f. $F(x)$. By means of this test we can give confidence limits for unknown distribution functions. Kolmogorov's theorem is as follows :

$$(2.2) \qquad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{-\infty < x < +\infty} |F_n(x)-F(x)| < y\right) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2y^2}$$

if $y > 0$, zero otherwise.

This theorem considers the difference $|F_n(x) - F(x)|$ with the same weight, regardless to the value of $F(x)$; so e.g. the difference $|(F_n(x) - F(x)| = 0.01$ has the same weight at a point x with $F(x) = 0.5$ (where this difference is 2% of the value of $F(x)$) as at a point x with $F(x) = 0.01$ (where this difference is 100% of the value of $F(x)$). We can avoid this by considering the quotient

$\{|F_n(x) - F(x)|\}/F(x)$ instead of $|F_n(x) - F(x)|$, that is to say, by considering the relative error of $F_n(x)$. In this way the idea arises to consider the limiting distribution of the supremum of the quotient $\{|F_n(x) - F(x)|\}/F(x)$ which characterizes the relative deviation of the population distribution function and that of the sample.

A theorem similar to that of Kolmogorov's was proved by N.V. Smirnov concerning the one-sided deviation of the sample and population distribution functions. Smirnov's theorem is as follows :

$$(2.3) \qquad \lim_{n\to\infty} P(\sqrt{n} \sup_{-\infty<x<+\infty} (F_n(x) - F(x)) > y) = 1 - e^{-2y^2}$$

if $y > 0$, zero otherwise. A. Rényi [23] also considers the analogous problem for relative deviations.

In the course of solving these problems a natural limitation is to be adopted : as $F(x)$ can take on arbitrary small values, we are not going to consider the supremum of $\{F_n(x) - F(x)\}/F(x)$ or the supremum of $\{|F_n(x) - F(x)|\}/F(x)$ taken in the whole interval $-\infty < x < +\infty$. We restrict ourselves to an interval $x^{(a)} \leq x < +\infty$, where $F(x^{(a)}) = a > 0$. The value of a, however, can be an arbitrarily small positive value. A. Rényi proves the following results :

Theorem 1.

$$(2.4) \qquad \lim_{n\to\infty} P(\sqrt{n} \sup_{a\leq F(x)} \frac{F_n(x) - F(x)}{F(x)} < y)$$

$$= \begin{cases} \sqrt{\dfrac{2}{\pi}} \displaystyle\int_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}}\, dt & \text{if } \; y > 0 \\ 0 \quad \text{if} \quad y \leq 0 \end{cases}$$

$$= \Phi\left(y\sqrt{\frac{a}{1-a}}\right)$$

Theorem 2.

$$\lim_{n\to\infty} P\left(\sqrt{n} \sup_{a\leq F(x)} \frac{|F_n(x) - F(x)|}{F(x)} < y\right) \tag{2.5}$$

$$= \begin{cases} \dfrac{4}{\pi} \displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}\, e^{-\frac{(2k+1)^2\pi^2}{8}\frac{1-a}{ay^2}} & \text{if} \quad y > 0\ , \\ 0 \quad \text{if} \quad y \leq 0 \end{cases}$$

$$= L\left(y\sqrt{\frac{a}{1-a}}\right)$$

We may consider the limiting distribution of the supremum of $\{F_n(x) - F(x)\}/F(x)$ and of its absolute value taken in the interval $x^{(a)} \leq x \leq x^{(b)}$ respectively, where $F(x^{(a)}) = a > 0$, $F(x^{(b)}) = b < 1$, $(0 < a < b < 1)$. In this regard A. Rényi [23] proves the following results :

Theorem 3.

$$\lim_{n\to\infty} P\left(\sqrt{n} \sup_{a\leq F(x)\leq b} \frac{F_n(x) - F(x)}{F(x)} < y\right) \tag{2.6}$$

$$= \frac{1}{\pi} \int_{-\infty}^{y\sqrt{\frac{b}{1-b}}} e^{-\frac{u^2}{2}} \left(\int_0^{\left(\sqrt{\frac{b}{1-b}}\, y-u\right)\sqrt{\frac{a(1-b)}{b-a}}} e^{-\frac{t^2}{2}}\, dt \right) du, \quad -\infty < y < +\infty$$

$= N(y; a, b)$

Theorem 4.

$$(2.7) \quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \le F(x) \le b} \frac{|F_n(x) - F(x)|}{F(x)} < y\right)$$

$$= \begin{cases} \dfrac{4}{\pi} \displaystyle\sum_{k=0}^{\infty} \dfrac{(-1)^k}{2k+1} e^{-\frac{(2k+1)^2\pi^2(1-a)}{8ay^2}} E_k, & \text{if } y > 0, \\ 0, & \text{if } y \le 0, \end{cases}$$

$$\text{where } E_k = 1 - \frac{2}{\sqrt{2\pi}} \int_{y\sqrt{\frac{b}{1-b}}}^{\infty} e^{\frac{-u^2}{2}} du + \rho_k ,$$

$$\rho_k = \left(\sqrt{2\pi}\sqrt{\frac{a}{1-a}}\, y\right)^{-1} 2e^{-\frac{by^2}{2(1-b)}} \int_0^{(2k+1)\frac{\pi}{2}} e^{\frac{(1-b)u^2}{2by^2}} \sin u \, du$$

$= R(y: a, b)$

These theorems provide tests for verifying the hypothesis that a random sample of size n, say $\xi_1, \xi_2, \ldots, \xi_n$, has been drawn from a population having continuous c.d.f $F(x)$. The character of these tests consists in that they give a band around $F(x)$ in which, if the hypothesis is tue, the sample distribution function $F_n(x)$ has to lie with a certain probability and the width of this band is proportional at all its points x to $F(x)$.

All these theorems are proved using the method of section 1 and the results of some limiting distribution theorems generalizing some results of P. Erdős and M. Kac [5]. We are going to give here these results of P. Erdős and M. Kac too in the version given and proved by A. Rényi in his paper [23], for we shall also need them in the proofs of theorems proposed in sections 3, 4 and 5 of this dissertation. Toward this end, let a sequence be given consisting of the sets of random variables

$$\xi_{n,1}, \ \xi_{n,2}, \ \ldots, \ \xi_{n,N_n} \quad (n = 1, 2, \ldots).$$

Let us assume that the random variables $\xi_{n,k}$ have expectation 0 and finite variance, further, that the random variables having the same first index n $(n = 1, 2, \ldots)$ are mutually independent and satisfy Lindeberg's condition, that is to say, introducing the notations

$$F_{n,k}(x) = P(\xi_{n,k} < x) \ ; \quad S_{n,k} = \sum_{v=1}^{k} \xi_{n,v} \ ; \quad B_n^2 = D^2 S_{n,N_n} = \sum_{k=1}^{N_n} D^2 \xi_{n,k}$$

where $D^2(x) = \text{Variance } (x)$, we suppose

$$M \ \xi_{n,k} = \int_{-\infty}^{\infty} x \ dF_{n,k}(x) = 0 \ ,$$

and $$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{N_n} \int_{|x|>\epsilon B_n} x^2 \ dF_{n,k}(x) = 0 \ .$$

Concerning these sequences satisfying the above conditions A. Rényi proves [23] the following theorems.

Theorem 5.

$$(2.8)\quad \lim_{n\to\infty} P\left(\max_{1\le k\le N_n} S_{n,k} < xB_n\right) = \begin{cases} \sqrt{\frac{2}{\pi}} \displaystyle\int_0^x e^{-\frac{t^2}{2}}\,dt & \text{if } x > 0, \\ 0 \text{ if } x \le 0. \end{cases}$$

Theorem 6.

$$(2.9)\quad \lim_{n\to\infty} P\left(\max_{1\le k\le N_n} |S_{n,k}| < xB_n\right) = \begin{cases} \frac{4}{\pi} \displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} e^{-\frac{(2k+1)^2\pi^2}{8x^2}} & \text{if } x > 0, \\ 0 \text{ if } x \le 0. \end{cases}$$

Theorem 7.

$$(2.10)\quad \lim_{n\to\infty} P\left(-yB_n \le \min_{1\le k\le N_n} S_{n,k} \le \max_{1\le k\le N_n} S_{n,k} < xB_n\right)$$

$$= \begin{cases} \frac{4}{\pi} \displaystyle\sum_{k=0}^{\infty} \frac{1}{2k+1} e^{-\frac{(2k+1)^2\pi^2}{2(x+y)^2}} \sin (2k+1)\pi \frac{x}{x+y}, & \text{if } x > 0 \text{ and } y \ge 0 \\ 0 \text{ if either } x \le 0 \text{ or } y < 0. \end{cases}$$

Remark. In case $y = x$, Theorem 7 reduces to Theorem 6.

Theorem 8. Let $A_n^2 = D^2 S_{n,M_n}$ with $1 \le M_n < N_n$ and

$$\lim_{n\to\infty} \frac{A_n}{B_n} = \lambda \qquad (0 \le \lambda < 1).$$

Then

$$(2.11)\quad \lim_{n\to\infty} P\left(\max_{M_n<k\le N_n} |S_{n,k}| < yB_n\right)$$

$$= \begin{cases} \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} e^{-\frac{(2k+1)^2\pi^2}{8y^2}} \left(1 - \frac{2}{\sqrt{2\pi}} \int_{\frac{y}{\lambda}}^{\infty} e^{\frac{-u^2}{2}} du + \rho_k \right) \text{ if } y > 0, \\ 0 \quad \text{if} \quad y \leq 0 \end{cases}$$

where

$$\rho_k = \frac{2\lambda}{\sqrt{2\pi}y} e^{-\frac{y^2}{2\lambda^2}} \int_0^{(2k+1)\frac{\pi}{2}} e^{\frac{\lambda^2 u^2}{2y^2}} \sin u \, du .$$

Remark. In the special case of $M_n = 1$ (i.e. by Lindeberg's condition, for $\lambda = 0$), Theorem 8 is identical with Theorem 6.

Note. For the special case in which all the considered random variables $\xi_{n,k}$ have the smae distribution, Theorems 5 and 6 were proved by P. Erdős and M. Kac [5] . They remarked that their theorems can be proved under more general conditions.

Going back for a moment to Theorems 1 and 3 of this section we remark here that the statements of these theorems hold for $\sqrt{n} \sup_{a \leq F(x)} \{F(x) - F_n(x)\}/F(x)$ and

$\sqrt{n} \sup_{a \leq F(x) \leq b} \{F(x) - F_n(x)\}/F(x)$ respectively too. **From this** remark and Theorem 1 it follows that we have

Corollary 1.

(2.12) $\lim_{n\to\infty} P(\sup_{x(a) \leq x < +\infty} (F_n(x) - F(x)) < 0)$

$$= \lim_{n\to\infty} P\left(\sup_{x^{(a)} \le x < +\infty} (F(x) - F_n(x)) < 0 \right) = 0 ,$$

i.e. the probability of the event that the sample distribution function does not exceed the population distribution function, and vice versa, all along the interval $x^{(a)} \le x < +\infty$, tends to zero as $n \to \infty$.

On the other hand we have by the above remark and Theorem 3 the following

Corollary 2.

$$(2.13) \quad \lim_{n\to\infty} P\left(\sup_{x^{(a)} \le x \le x^{(b)}} (F_n(x) - F(x)) < 0 \right)$$

$$= \lim_{n\to\infty} P\left(\sup_{x^{(a)} \le x \le x^{(b)}} (F(x) - F_n(x)) < 0 \right)$$

$$= \frac{1}{\pi} \int_0^\infty e^{\frac{-u^2}{2}} \left(\int_0^{y\sqrt{\frac{a(1-b)}{b-a}}} e^{\frac{-t^2}{2}} \, dt \right) du ,$$

i.e. the probability of the event that the sample distribution function does not exceed the population distribution function, and vice versa, all along the interval in which the value of $F(x)$ lies between arbitrarily fixed values a and b $(0 < a < b < 1)$, remains positive in the limit.

This result is obviously important from the point of view of statistical practice (truncation problems).

The result that $\lim_{n\to\infty} P\left(\sup_{x^{(a)} \le x \le x^{(b)}} (F_n(x) - F(x)) < 0 \right)$ is positive was also proved by Gihman [8]; moreover he obtained that

$$(2.14)\quad \lim_{n\to\infty} P\left(\sup_{x(a)\leq x\leq x(b)} (F_n(x) - F(x)) < 0\right) = \frac{1}{\pi} \text{ arc sin} \sqrt{\frac{a(1-b)}{b(1-a)}} .$$

Gihman also mentioned that the result (2.14) has already been known to Gnedenko. The results of Corollary 2 and (2.14) are, of course, identical. Indeed, the result of Corollary 2 is twice the probability of the event that a random variable (x, y) with p.d.f. $\frac{1}{2\pi} \exp(-\frac{1}{2}(x^2 + y^2))$ lies in the infinite sector $0 < x < +\infty$, $0 < y < x\sqrt{\frac{a(1-b)}{b-a}}$, and this probability is equal to

$$(2.15)\quad 2 \cdot \frac{\text{arc tg}\sqrt{\frac{a(1-b)}{b-a}}}{2\pi} = \frac{1}{\pi} \text{ arc sin} \sqrt{\frac{a(1-b)}{b(1-a)}} ,$$

because of the circular symmetry of $\frac{1}{2\pi} \exp(-\frac{1}{2}(x^2 + y^2))$, the probability corresponding to an infinite sector of angle φ is equal to $\frac{\varphi}{2\pi}$.

3. The Kolmogorov - Smirnov theorems using $1/F_n(x)$ as weight function.

The intorduction of the weight function $1/F(x)$ in the theorems of Kolmogorov and Smirnov by A. Rényi characterizes the asymptotic behaviour of the relative deviation of the population distribution function and that of the sample. The theorems provide tests for verifying the hypothesis that a random sample of size n with c.d.f. $F_n(x)$ has been drawn from a population having continuous c.d.f. $F(x)$. At the same time, as far as statistical considerations are concerned, we lose one of the convenient properties of Kolmogorov's theorem, given in (2.2). Namely, we can use this theorem to construct confidence intervals for an unknown continuous c.d.f. $F(x)$ (e.g. $y_1 = F_n(x) - \frac{1.63}{n}$ and $y_2 = F_n(x) + \frac{1.63}{n}$ constitute a 99 % confidence interval for an unknown continuous c.d.f. $F(x)$; for tables see e.g. [22]). Having the limiting distribution of $\sup |F_n(x) - F(x)|/F(x)$ instead, we no longer have that property. We could retain the advantage of these new theorems that they measure the relative deviation of the population c.d.f. and that of the sample and could regain the above mentioned confidence interval property of the old theorems if we used $1/F_n(x)$ instead of $1/F(x)$ as weight function. Thus in this way the idea arises of considering the limiting distribution of the supremum of the quotients $\{F_n(x) - F(x)\}/F_n(x)$ and $|(F_n(x) - F(x)|/F_n(x)$. In examining the limiting distribution of these quotients a natural limitation on $F_n(x)$ is to be adopted. Namely, we restrict ourselves to the set of those x's for which we have $F_n(x) \geq a > 0$. The value of a can, however, be arbitrarily small.

Keeping the notation and assumptions of the previous chapters we are going to prove the following theorems :

Theorem 9.

$$(3.1) \quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leq F_n(x)} \frac{F_n(x) - F(x)}{F_n(x)} < y\right) = \Phi\left(y\sqrt{\frac{a}{1-a}}\right)$$

Theorem 9'.

$$(3.2) \quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leq F_n(x)} \frac{F(x) - F_n(x)}{F_n(x)} < y\right) = \Phi\left(y\sqrt{\frac{a}{1-a}}\right)$$

where, in both cases, $\Phi\left(y\sqrt{\frac{a}{1-a}}\right)$ stands for the function defined in (2.4) of chapter 2.

Theorem 10.

$$(3.3) \quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leq F_n(x)} \frac{|F_n(x) - F(x)|}{F_n(x)} < y\right) = L\left(y\sqrt{\frac{a}{1-a}}\right)$$

where $L\left(y\sqrt{\frac{a}{1-a}}\right)$ stands for the function defined in (2.5) of chapter 2.

Theorem 11.

$$(3.4) \quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leq F_n(x) \leq b} \frac{F_n(x) - F(x)}{F_n(x)} < y\right) = N(y;\ a,\ b)$$

Theorem 11'.

$$(3.5)\quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a\leq F_n(x)\leq b} \frac{F(x) - F_n(x)}{F_n(x)} < y\right) = N(y;\ a,\ b)$$

where, in both cases, $N(y;\ a,\ b)$ stands for the function defined in (2.6) of chapter 2 and, as there, a and b are such that $0 < a < b < 1$.

Theorem 12.

$$(3.6)\quad \lim_{n\to\infty} P\left(\sqrt{n} \sup_{a\leq F_n(x)\leq b} \frac{|F_n(x) - F(x)|}{F_n(x)} < y\right) = R(y;\ a,\ b)$$

where $R(y;\ a,\ b)$ stands for the function defined in (2.7) of chapter 2 and again we have $0 < a < b < 1$.

These theorems provide tests for verifying the hypothesis that a random sample of size n with c.d.f $F_n(x)$ has been drawn from a population having continuous c.d.f. $F(x)$. The character of these tests is that they give a band around $F(x)$ in which, if the hypothesis is true, the sample distribution function $F_n(x)$ has to lie with a certain probability and the width of this band is proportional at all its points to $F_n(x)$ instead of the previous proportionality to $F(x)$. Having these theorems we can also construct confidence intervals for an unknown continuous c.d.f. $F(x)$ using theorems 10 and 12, or lower and upper boundaries for an unknown continuous c.d.f. $F(x)$ using theorems 9 and 9' or 11 and 11'. For example if $a = 0.05$, using Theorem 10 and the table published by A. Rényi in [23] for the function $L(y\sqrt{\frac{a}{1-a}})$, we get

$Z_1 = F_n(x) - 8.5 \frac{F_n(x)}{\sqrt{n}}$ and $Z_2 = F_n(x) + 8.5 \frac{F_n(x)}{\sqrt{n}}$ as a 90 % confidence interval for $F(x)$ at all points x such that $F_n(x) \geq a = 0.05$ provided that n is large enough to make $Z_1 = F_n(x) - 8.5 \frac{F_n(x)}{\sqrt{n}} \geq 0$. In general, given the value of a and the probability level on which we would like to construct a confidence interval for $F(x)$, we can always decide how large a sample size is the minimum for these theorems to work at all. On the other hand, if we are given a sample of a certain size and have a desirable degree of confidence interval in mind we can get meaningful answers by manipulating the values of y and a, perhaps at the cost of getting no intervals around the first few order statistics if the fixed sample size in question would be a smaller one. To illustrate this point let us consider that we are given a sample of size 30 and that on the basis of this sample we would like to construct a 90 % confidence interval for $F(x)$; i.e. $F(x)$ is to lie in the interval : $F_n(x)\ (1 \pm \frac{y}{\sqrt{n}})$ with probability 0.90, given that $n = 30$. Taking $y = 5$ makes $F_{30}(x)\ (1 \pm \frac{y}{\sqrt{30}})$ positive on both sides. Using Theorem 10 and the table published by A. Renyi in [23] for the function $L(y \sqrt{\frac{a}{1-a}})$, we get probabilities 0.8088 and 0.9751 for $a = 0.1$ and $a = 0.2$ respectively. Interpolating, we take $a = 0.16$. We must now have $F_{30}(x) > 0.16$, i.e. $\frac{k}{30} > 0.16$ which implies $k > 4.8$. Thus we shall get confidence intervals for $F_{30}(x) > \frac{5}{30}$

that is from and to the right of the fifth order statistics.

The statements of corollaries 1 and 2 of chapter 2 hold to the theorems of this chapter too, mutatis mutandis. That is, instead of talking about restricting the values of $F(x)$ to some intervals for x, we would talk here about restricting the values of $F_n(x)$ in the sense of the theorems of this chapter.

Proof of Theorem 9.

To summarize our assumptions, let ξ be a random variable having continuous c.d.f. $F(x)$ and let $\xi_1, \xi_2, \ldots, \xi_n$ denote n independent observations on the random variable ξ, i.e. let $\xi_1, \xi_2, \ldots, \xi_n$ be n mutually independent random variables having the same continuous c.d.f. $F(x)$. The distribution function of this random sample is denoted by $F_n(x)$.

In keeping with the notation of previous sections, we put $\eta_k = F(\xi_k)$ and $\zeta_k = \log \frac{1}{\eta_k}$, further, $\eta_{(k)} = F(\xi_{(k)})$ and $\zeta_{(k)} = R_k(\zeta_1, \zeta_2, \ldots, \zeta_n)$ which is such that $\zeta_{(k)} = \log \frac{1}{\eta_{(n+1-k)}}$, $k = 1, 2, \ldots, n$. The random variables η_k are uniformly distributed in the interval $(0, 1)$ and, if $u = F(x)$, their sample distribution function is $G_n(u) = F_n(F^{-1}(u))$, where $x = F^{-1}(u)$ is the inverse function of $u = F(x)$.

Now the limiting distribution of the random variable $\sqrt{n} \sup_{a \leq F_n(x)} \frac{F_n(x) - F(x)}{F_n(x)}$ is identical with that of the random variable $\sqrt{n} \sup_{a \leq G_n(u)} \frac{G_n(u) - u}{G_n(u)}$. We also have

$$(3.7) \quad \sqrt{n} \sup_{a \leq G_n(u)} \frac{G_n(u) - u}{G_n(u)} = \sqrt{n} \max_{a \leq G_n(\eta_{(k)}+0)} \left(1 - \frac{\eta_{(k)}}{G_n(\eta_{(k)}+0)}\right)$$

$$= \sqrt{n} \max_{an \leq k \leq n} \left(1 - \frac{\eta_{(k)}}{k/n}\right),$$

for in the interval $\eta_{(k)} \leq u < \eta_{(k+1)}$ $G_n(u) = k/n$ and because we can disregard the possibility of having

$$\sup_{a \leq G_n(u)} \frac{G_n(u) - u}{G_n(u)} = \max_{a \leq G_n(\eta_{(k+1)}-0)} \left(1 - \frac{\eta_{(k+1)}}{G_n(\ (k+1)-0)}\right) = \max_{an \leq k \leq n} \left(1 - \frac{\eta_{(k+1)}}{k/n}\right)$$

for $\max_{an \leq k \leq n} \left(1 - \frac{\eta_{(k)}}{k/n}\right) \geq \max_{an \leq k \leq n} \left(1 - \frac{\eta_{(k+1)}}{k/n}\right)$.

So it is sufficient to examine the limiting distribution of the random variable

$$\sqrt{n} \max_{an \leq k \leq n} \left(1 - \frac{\eta_{(k)}}{k/n}\right)$$

which in turn is identical with that of the random variable

$$(3.8) \quad \sqrt{n} \max_{an \leq k \leq n} \left(-\log \frac{\eta_{(k)}}{k/n}\right) = \sqrt{n} \max_{an \leq k \leq n} \log \frac{k/n}{(k)}$$

Applying the theory of section 1 and using the notation introduced there we have

$$(3.9) \quad \log \frac{1}{\eta_{(k)}} = \sum_{v=k}^{n} \frac{\delta_{n+1-v}}{v}$$

where the δ_{n+1-v} are mutually independent exponentially distributed random variables with c.d.f. $1 - e^{-x}$ $(x \geq 0)$.

Therefore we have

$$M \log \frac{1}{\eta_{(k)}} = \sum_{v=k}^{n} \frac{1}{v}$$

(3.10)

$$D^2 \log \frac{1}{\eta_{(k)}} = \sum_{v=k}^{n} \frac{1}{v^2}$$

as the mean and variance of $\log \frac{1}{\eta_{(k)}}$ respectively.

Consider now the sequence of random variables

(3.11) $$\frac{\delta_{n+1-v} - 1}{v}, \quad v = k, \ldots, n$$

This sequence satisfies Lindeberg's condition and therefore we apply Theorem 5 of section 2 with

(3.12) $$\max_{an \leq k \leq n} \left(\log \frac{1}{\eta_{(k)}} - \sum_{v=k}^{n} \frac{1}{v}\right) = \max_{an \leq k \leq n} \sum_{k=v}^{n} \frac{\delta_{n+1-v} - 1}{v}$$

Therefore we have

(3.13) $$\lim_{n \to \infty} P\left(\max_{an \leq k \leq n} \left(\log \frac{1}{\eta_{(k)}} - \sum_{v=k}^{n} \frac{1}{v}\right) < z \sqrt{\sum_{an \leq k \leq n} \frac{1}{k^2}}\right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{t^2}{2}} dt, \quad z > 0 .$$

Since, if $k \geq an$ and $0 < a < 1$, we have, by Euler's summation formula, that

(3.14) $$\sum_{v=k}^{n} \frac{1}{v} = \log n - \log k + o\left(\frac{1}{n}\right) = \log \frac{n}{k} + o\left(\frac{1}{n}\right)$$

and

$$(3.15) \quad \sqrt{\sum_{an\le k\le n} \frac{1}{k^2}} = \sqrt{\frac{1}{an} - \frac{1}{n}} + o\left(\frac{1}{n}\right) = \sqrt{\frac{1-a}{an}} + o\left(\frac{1}{n}\right),$$

from (3.13), (3.14) and (3.15) we deduce

$$(3.16) \quad \lim_{n\to\infty} P\left(\max_{an\le k\le n} \left(\log \frac{1}{\eta_{(k)}} - \log \frac{n}{k}\right) < z \sqrt{\frac{1-a}{an}} \right)$$

$$= \lim_{n\to\infty} P\left(\max_{an\le k\le n} \left(\log \frac{k/n}{(k)}\right) < z \sqrt{\frac{1-a}{an}} \right) = \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{t^2}{2}} dt ,$$

$z > 0$.

Letting $y = z\sqrt{\frac{1-a}{a}}$ we get

$$(3.17) \quad \lim_{n\to\infty} P\left(\sqrt{n} \max_{an\le k\le n} \log \frac{k/n}{\eta_{(k)}} < y \right) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}} dt$$

if $y > 0$, zero otherwise. This, by (3.8), completes the proof of Theorem 9.

Proof of Theorem 9'.

Repeating the argument of the first part of the proof of Theorem 9 we can show that the limiting distribution of the random variable

$$\sqrt{n} \sup_{a\le F_n(x)} \frac{F(x) - F_n(x)}{F_n(x)}$$

is identical with that of the random variable

$$(3.18) \quad \sqrt{n} \max_{an\le k\le n} \left(\frac{\eta_{(k+1)}}{k/n} - 1 \right)$$

which in turn has the same limiting distribution as the random variable

$$(3.19) \quad \sqrt{n} \max_{an\le k\le n} \log \frac{\eta_{(k+1)}}{k/n} = \sqrt{n} \max_{an\le k\le n} \left(-\log \frac{k/n}{\eta_{(k+1)}}\right), \quad \eta_{(n+1)}=1$$

Using again the notation and results of section 1 we consider

$$(3.20)\qquad \log \frac{1}{\eta_{(k+1)}} = \sum_{v=k+1}^{n} \frac{\delta_{n+1-v}}{v},$$

where the δ_{n+1-v} are mutually independent exponentially distributed random variables with c.d.f. $1 - e^{-x}$ $(x \geq 0)$. Therefore we have

$$(3.21)\qquad \begin{aligned} &M \log \frac{1}{\eta_{(k+1)}} = \sum_{v=k+1}^{n} \frac{1}{v}, \quad \text{where} \quad \sum_{v=n+1}^{n} \frac{1}{v} = 0, \\ &D^2 \log \frac{1}{\eta_{(k+1)}} = \sum_{v=k+1}^{n} \frac{1}{v^2}, \quad \text{where} \quad \sum_{v=n+1}^{n} \frac{1}{v} = 0, \end{aligned}$$

as the expected value and variance of $\log \frac{1}{\eta_{(k+1)}}$ respectively.

We consider now the sequence of random variables

$$(3.22)\qquad \frac{1 - \delta_{n+1-v}}{v}, \quad v = k+1, \ldots, n .$$

This sequence satisfies Lindeberg's condition and therefore we apply Theorem 5 of section 2 with

$$(3.23)\qquad \max_{an \leq k \leq n} \left(\sum_{v=k+1}^{n} \frac{1}{v} - \log \frac{1}{\eta_{(k+1)}} \right) = \max_{an \leq k \leq n-1} \sum_{v=k+1}^{n} \frac{1 - \delta_{n+1-v}}{v}$$

Therefore we have

$$(3.24)\qquad \lim_{n \to \infty} P\left(\max_{an \leq k \leq n} \sum_{v=k+1}^{n} \frac{1}{v} - \log \frac{1}{\eta_{(k+1)}} \right) < z \sqrt{\sum_{an \leq k \leq n-1} \frac{1}{(k+1)^2}} \;)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{t^2}{2}} \, dt \quad \text{if} \quad z > 0 \, .$$

Using Euler's summation formula with $k \geq an$ and $0 < a < 1$ we have that

$$(3.25) \qquad \sum_{v=k+1}^{n} \frac{1}{v} = \log n - \log (k+1) + o(\tfrac{1}{n}) = \log \frac{n}{k+1} + o(\tfrac{1}{n})$$

and

$$(3.26) \qquad \sqrt{\sum_{an \leq k \leq n-1} \frac{1}{(k+1)^2}} = \sqrt{\frac{1}{an+1} - \frac{1}{n} + o(\tfrac{1}{n})} = \sqrt{\frac{n-an-1}{(an+1)n} + o(\tfrac{1}{n})}$$

But $\log \frac{n}{k+1} = \log \frac{1}{\frac{k}{n} + \frac{1}{n}}$ and $\frac{n-an-1}{(an+1)n} = \frac{1-a-\frac{1}{n}}{an(1+\frac{1}{an})}$,

that is when n is large we have

$$(3.27) \qquad \sum_{v=k+1}^{n} \frac{1}{v} = \log \frac{n}{k}$$

and

$$(3.28) \qquad \sqrt{\sum_{an \leq k \leq n-1} \frac{1}{(k+1)^2}} = \sqrt{\frac{1-a}{an}} \, .$$

So from (3.24) we conclude that

$$(3.29) \qquad \lim_{n \to \infty} P\left(\max_{an \leq k \leq n} \left(\log \frac{n}{k} - \log \frac{1}{(k+1)}\right) < z \sqrt{\frac{1-a}{an}} \right)$$

$$= \lim_{n \to \infty} P\left(\max_{an \leq k \leq n} \left(-\log \frac{k/n}{(k+1)}\right) > z \sqrt{\frac{1-a}{an}} \right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{t^2}{2}} \, dt \, , \quad \text{if} \quad z > 0 \, .$$

Putting $y = z \sqrt{\frac{1-a}{an}}$ we get

$$(3.30)\quad \lim_{n\to\infty} P\left(\sqrt{n} \max_{an\le k\le n} \left(-\log \frac{k/n}{\eta_{(k+1)}}\right) < y\right) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}}\,dt$$

if $y > 0$, zero otherwise. By (3.19) this proves Theorem 9'.

Because of some similarities to the above two proofs let us turn now to the proofs of Theorems 11 and 11'.

Proof of Theorem 11.

From the proof of Theorem 9 (namely from (3.12)) it is clear that here we will have to consider the limiting distribution of the random variable

$$(3.31)\quad \omega = \sqrt{n} \max_{an\le k\le bn} \left(\log \frac{1}{\eta_{(k)}} - \sum_{v=k}^{n} \frac{1}{v}\right) = \sqrt{n} \max_{an\le k\le bn} \sum_{v=k}^{n} \frac{\delta_{n+1-v}^{-1}}{v}$$

which may be written as the sum of two independent random variables ω_1 and ω_2 where

$$(3.32)\quad \omega_1 = \sqrt{n} \sum_{bn\le k\le n} \frac{\delta_{n+1-k}^{-1}}{k}$$

$$(3.33)\quad \omega_2 = \sqrt{n} \max_{an\le k\le bn} \sum_{k} \frac{\delta_{n+1-k}^{-1}}{k}$$

By the Lindeberg form of the central limit theorem, in the limit ω_1 is a normally distributed random variable with standard deviation $\sqrt{\frac{1-b}{b}}$, for, by (3.15), the standard deviation of $\sum_{bn\le k\le n} \frac{\delta_{n+1-k}}{k}$ is equal to $\sqrt{\frac{1-b}{bn}} + o(\frac{1}{n})$.

Further, from the proof of Theorem 9, we can see that

$$(3.34)\quad \lim_{n\to\infty} P(\omega_2 \sqrt{b} < z) = \sqrt{\frac{2}{\pi}} \int_0^{z\sqrt{\frac{a}{b-a}}} e^{-\frac{t^2}{2}}\,dt\ , \quad z > 0\ ,$$

for if $an \leq k \leq bn$, $0 < a < b$ then

$$(3.35)\quad D^2 \sum_{an\leq k\leq bn} \frac{\delta_{n+1-k}}{k} = \sum_{an\leq k\leq bn} \frac{1}{k^2} = \frac{1}{an} - \frac{1}{bn} + o\left(\frac{1}{n}\right)$$

and so

$$(3.36)\quad D \sum_{an\leq k\leq bn} \frac{\delta_{n+1-k}}{k} = \sqrt{\frac{b-a}{abn}} + o\left(\frac{1}{n}\right)$$

are the variance and standard deviation of the random variable $\sum_{an\leq k\leq bn} \frac{\delta_{n+1-k}}{k}$.

If, in (3.34), we let $y = z/\sqrt{b}$ then

$$(3.37)\quad \lim_{n\to\infty} P(\omega_2 < y) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{ab}{b-a}}} e^{-\frac{t^2}{2}}\,dt\ , \quad y > 0\ .$$

Taking into account that ω_1 and ω_2 are ~~independent~~ random variables it follows from (3.32), (3.37) and convolution that

$$(3.38)\quad \lim_{n\to\infty} P(\omega < y) = \frac{1}{\pi}\sqrt{\frac{b}{1-b}} \int_{-\infty}^{y} e^{-\frac{bu^2}{2(1-b)}} \int_0^{(y-u)\sqrt{\frac{ab}{b-a}}} e^{-\frac{t^2}{2}}\,dt\,du\ .$$

This completes the proof of Theorem 11.

Proof of Theorem 11'.

From the proof of Theorem 9' (namely from (3.19)) it is clear that we are considering here the limiting distribution of the random variable

$$(3.39)\quad \sqrt{n} \max_{an \le k \le bn} \log \frac{\eta(k+1)}{k/n} = \sqrt{n} \max_{an \le k \le bn} \left(-\log \frac{k/n}{\eta(k+1)} \right)$$

which in turn has the same limiting distribution as the random variable

$$(3.40)\quad \mathcal{T} = \sqrt{n} \max_{an \le k \le bn} \left(\sum_{v=k+1}^{n} \frac{1}{v} - \log \frac{1}{\eta(k+1)} \right)$$

$$= \sqrt{n} \max_{an \le k \le bn} \sum_{v=k+1}^{n} \frac{1 - \delta_{n+1-v}}{v}$$

in comparison with (3.23). The right hand side of (3.40) can be written as the sum of two independent random variables $\mathcal{T}_1$ and $\mathcal{T}_2$ where

$$(3.41)\quad \mathcal{T}_1 = \sqrt{n} \sum_{bn \le k \le n-1} \frac{1 - \delta_{n-k}}{k+1}$$

and

$$(3.42)\quad \mathcal{T}_2 = \sqrt{n} \max_{an \le k \le bn} \sum_{k} \frac{1 - \delta_{n-k}}{k+1}$$

By the Lindeberg form of the central limit theorem in the limit $\mathcal{T}_1$ is a normally distributed random variable with standard deviation $\sqrt{\frac{1-b}{b}}$, for we have already seen that

$$D^2 \sum_{bn \le k \le n-1} \frac{\delta_{n-k}}{k+1} = \sum_{bn \le k \le n-1} \frac{1}{(k+1)^2} = \sqrt{\frac{1-b}{bn}} + o\left(\frac{1}{n}\right)$$

Further, from the proof of Theorem 9', we can see that

$$(3.43)\quad \lim_{n\to\infty} P(\mathcal{T}_2\sqrt{b} < z) = \sqrt{\frac{2}{\pi}} \int_0^{z\sqrt{\frac{a}{b-a}}} e^{-\frac{t^2}{2}}\, dt \quad , \quad z > 0 \,,$$

for if $an \leq k \leq bn$, $0 < a < b$ then

$$(3.44) \quad D^2\Big(\sum_{an \leq k \leq bn} \frac{\delta_{n-k}}{k+1} \Big) = \sum_{an \leq k \leq bn} \frac{1}{(k+1)^2}$$

$$= \frac{1}{an+1} - \frac{1}{bn+1} + o(\tfrac{1}{n}) \text{ , by ~~Euler's~~ formula}$$

$$= \frac{b-a}{abn(1+\frac{1}{an})(1+\frac{1}{bn})} + o(\tfrac{1}{n}) .$$

That is when n is large we have

$$(3.45) \quad D\Big(\sum_{an \leq k \leq bn} \frac{\delta_{n-k}}{k+1} \Big) = \sqrt{\frac{b-a}{abn}} .$$

From (3.42) and (3.43) we deduce, on letting $y = z/\sqrt{b}$, that

$$(3.46) \quad \lim_{n\to\infty} P(\mathcal{T}_2 < y) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{ab}{b-a}}} e^{-\frac{t^2}{2}}\, dt , \quad y > 0 .$$

Considering further that $\mathcal{T}_1$ and $\mathcal{T}_2$ independent random variables it follows from (3.41), (3.43) and convolution that

$$(3.47) \quad \lim_{n\to\infty} P(\mathcal{T} < y) = \frac{1}{\pi}\sqrt{\frac{b}{1-b}} \int_{-\infty}^{y} e^{-\frac{bu^2}{2(1-b)}} \int_0^{(y-u)\sqrt{\frac{ab}{b-a}}} e^{-\frac{t^2}{2}}\, dt\, du .$$

This completes the proof of Theorem 11'.

What remains now is to prove Theorems 10 and 12 of this section.

Proof of Theorem 10.

First we observe that the limiting distribution of the random variable $\sqrt{n} \sup_{a \leq F_n(x)} |F_n(x) - F(x)|/F_n(x)$ is identical with that of the random variable

$\sqrt{n} \sup_{a \leq G_n(u)} |G_n(u) - G(u)|/G_n(u)$ where again $G_n(u) = F_n(F^{-1}(u))$,

and $x = F^{-1}(u)$ is the inverse function of $u = F(x)$.

We recall that $G_n(u) = k/n$ if $\eta_{(k)} \leq u < \eta_{(k+1)}$ and in particular that $G_n(\eta_{(k)}+0) = G_n(\eta_{(k+1)}-0) = k/n$. Therefore we have

$$(3.48) \quad \sqrt{n} \sup_{a \leq G_n(u)} \frac{|G_n(u) - u|}{G_n(u)}$$

$$= \sqrt{n} \max_{an \leq k < n} \left(\left| 1 - \frac{\eta_{(k)}}{G_n(\eta_{(k)}+0)} \right|, \left| 1 - \frac{\eta_{(k+1)}}{G_n(\eta_{(k+1)}-0)} \right| \right)$$

$$= \sqrt{n} \max_{an \leq k < n} \left(\left| 1 - \frac{\eta_{(k)}}{k/n} \right|, \left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| \right)$$

Now if $\eta_{(k+1)} < k/n$, then

$$(3.49) \quad \left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| = 1 - \frac{\eta_{(k+1)}}{k/n} < 1 - \frac{\eta_{k+1}}{\frac{k+1}{n}} = \left| 1 - \frac{\eta_{(k+1)}}{\frac{k+1}{n}} \right|$$

On the other hand, if $\eta_{(k+1)} \geq k/n$ then

$$\left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| = \frac{\eta_{(k+1)}}{k/n} - 1 = \frac{\eta_{(k+1)}+C}{\frac{k+1}{n}} - 1$$

where $\dfrac{C}{\eta_{(k+1)}} = \dfrac{1/n}{k/n} = \dfrac{1}{k}$, that is $C = \dfrac{\eta_{(k+1)}}{k}$, and

therefore

$$\left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| = \frac{\eta_{(k+1)}}{\frac{k+1}{n}} - 1 + \frac{n\,\eta_{(k+1)}}{k(k+1)}$$

$$\leq \left| 1 - \frac{\eta_{(k+1)}}{\frac{k+1}{n}} \right| + \frac{n\,\eta_{(k+1)}}{k^2} \quad , \quad an \leq k \leq n$$

But $\eta_{(k+1)} = F(\xi_{(k+1)}) \leq 1$ and therefore, if $\eta_{(k+1)} \geq k/n$, we have

$$(3.50) \quad \left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| \leq \left| 1 - \frac{\eta_{(k+1)}}{\frac{k+1}{n}} \right| + \frac{1}{a^2 n}$$

Consequently, in either case

$$(3.51) \quad \left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| \leq \left| 1 - \frac{\eta_{(k+1)}}{\frac{k+1}{n}} \right| + \frac{1}{a^2 n}$$

By definition $\eta_{(n+1)} = 1$ and $\frac{k+1}{n} = 1$ if $k = n$, therefore

$$(3.52) \quad \max_{an \leq k \leq n} \left| 1 - \frac{\eta_{(k+1)}}{\frac{k+1}{n}} \right| \leq \max_{an \leq k \leq n} \left| 1 - \frac{\eta_{(k)}}{k/n} \right|$$

and from (3.51) and (3.52) we get

$$(3.53) \quad \max_{an \leq k \leq n} \left(\left| 1 - \frac{\eta_{(k)}}{k/n} \right|, \left| 1 - \frac{\eta_{(k+1)}}{k/n} \right| \right) \leq \max_{an \leq k \leq n} \left| 1 - \frac{\eta_{(k)}}{k/n} \right| + \frac{1}{a^2 n}$$

From (3.48) and (3.53) we conclude that

$$(3.54) \quad \sqrt{n} \max_{an \leq k \leq n} \left| 1 - \frac{\eta_{(k)}}{k/n} \right| \leq \sqrt{n} \sup_{a \leq G_n(u)} \frac{|G_n(u) - u|}{G_n(u)}$$

$$\leq \sqrt{n} \max_{an \leq k \leq n} \left| 1 - \frac{\eta_{(k)}}{k/n} \right| + \frac{1}{a^2 \sqrt{n}}$$

and this implies that the limiting distribution of the random variable $\sqrt{n} \sup_{a \leq F_n(x)} |F_n(x) - F(x)|/F_n(x)$ is identical with that of the random variable

$$(3.55)\quad \sqrt{n} \max_{an\le k\le n} \left| 1 - \frac{\eta_{(k)}}{k/n} \right| = \sqrt{n} \max_{an\le k\le n} \left| \frac{\eta_{(k)}}{k/n} - 1 \right|$$

which in turn has the same limiting distribution as the random variable

$$(3.56)\quad \sqrt{n} \max_{an\le k\le n} \left| \log \frac{\eta_{(k)}}{k/n} \right| = \sqrt{n} \max_{an\le k\le n} \left| \log \frac{k/n}{\eta_{(k)}} \right|$$

From the proof of Theorem 9 it is clear that the limiting distribution of the random variable of (3.56) is identical with that of the random variable

$$(3.57)\quad \max_{an\le k\le n} \left| \log \frac{1}{\eta_{(k)}} - \sum_{v=k}^{n} \frac{1}{v} \right| = \max_{an\le k\le n} \left| \sum_{v=k}^{n} \frac{\delta_{n+1-v}-1}{v} \right|$$

Applying Theorem 6 of section 2 with (3.59) and using the results of (3.14) and (3.15) we get

$$(3.58)\quad \lim_{n\to\infty} P\left(\sqrt{n} \max_{an\le k\le n} \left| \log \frac{k/n}{\eta_{(k)}} \right| < y\right) = L\left(y\sqrt{\frac{a}{1-a}}\right) .$$

This by (3.56) completes the proof of Theorem 10.

Proof of Theorem 12.

Using the method of the proof of Theorem 10 we can show here that the limiting distribution of the random variable $\sqrt{n} \sup_{a\le F_n(x)\le b} |F_n(x) - F(x)|/F_n(x)$ is identical with that of the random variable

$$(3.59)\quad \sqrt{n} \max_{an\le k\le bn} \left| \frac{\eta_{(k)}}{k/n} - 1 \right|$$

and therefore it is also identical with the limiting distribution of the random variable

$$(3.60) \quad \sqrt{n} \max_{an \le k \le bn} \left| \log \frac{\eta_{(k)}}{k/n} \right| = \sqrt{n} \max_{an \le k \le bn} \left| \log \frac{k/n}{\eta_{(k)}} \right| ,$$

which in turn has the same limiting distribution as the random variable

$$(3.61) \quad \max_{an \le k \le bn} \left| \log \frac{1}{\eta_{(k)}} - \sum_{v=k}^{n} \frac{1}{v} \right| = \max_{an \le k \le bn} \left| \sum_{v=k}^{n} \frac{\delta_{n+1-v}-1}{v} \right|$$

Let us define

$$(3.62) \quad S_{n,n+1-k} = \sum_{v=k}^{n} \frac{\delta_{n+1-v}-1}{v}$$

$$(3.63) \quad B_n^2 = D^2 \left(\sum_{an \le k \le n} \frac{\delta_{n+1-k}-1}{k} \right) = \sum_{an \le k \le n} \frac{1}{k^2}$$

$$(3.64) \quad A_n^2 = D^2 \left(\sum_{bn \le k \le n} \frac{\delta_{n+1-k}-1}{k} \right) = \sum_{bn \le k \le n} \frac{1}{k^2}$$

By (3.15) we have

$A_n = \sqrt{\frac{1-b}{bn}} + o(\frac{1}{n})$ and $B_n = \sqrt{\frac{1-a}{an}} + o(\frac{1}{n})$ and so

$$(3.65) \quad \lambda = \lim_{n\to\infty} \frac{A_n}{B_n} = \sqrt{\frac{a(1-b)}{b(1-a)}}$$

Thus we can apply Theorem 8 of section 2 with $S_{n,n+1-k}$ defined in (3.62) and with (3.63), (3.64), (3.65) and $1 \le M_n = n+1-bn < N_n = n+1-an$. Therefore we have

$$(3.66) \quad \lim_{n\to\infty} P(\max_{an \le k \le bn} |S_{n,n+1-k}| < zB_n)$$

$$= \lim_{n\to\infty} P\left(\max_{an\leq k\leq bn} \left| \log \frac{k/n}{\eta_{(k)}} \right| < z \sqrt{\frac{1-a}{an}} \right)$$

$$= \lim_{n\to\infty} P\left(\sqrt{n} \max_{an\leq k\leq bn} \left| \log \frac{k/n}{\eta_{(k)}} \right| < z \sqrt{\frac{1-a}{a}} \right)$$

which is equal to the statement of Theorem 8 of chapter 2. Letting $y = z\sqrt{\frac{1-a}{a}}$ we get

$$(3.67) \quad \lim_{n\to\infty} P\left(\sqrt{n} \max_{an\leq k\leq bn} \left| \log \frac{k/n}{\eta_{(k)}} \right| < y \right) = R(y;\, a,\, b)$$

This, by (3.60) and (3.59) completes the proof of Theorem 12.

4. Some particular cases of the Smirnov limit theorems for empirical distributions using $1/F(x)$ as weight function.

Let $\xi_{11}, \xi_{12}, \ldots, \xi_{1n}$ be independent observations on the random variable ξ_1, and let $\xi_{21}, \xi_{22}, \ldots, \xi_{2m}$ be independent observations on the random variable ξ_2. Assume that the random variables $\xi_{11}, \xi_{12}, \ldots, \xi_{1n}, \xi_{21}, \xi_{22}, \ldots, \xi_{2m}$ are mutually independent and, also, assume that the random **variables** ξ_{1i} and ξ_{2j} have continuous c.d.f - s $F(x)$ and $H(x)$ respectively, which are unknown. Let $F_n(x)$ and $H_m(x)$ be the sample (empirical) distribution functions of $\xi_{11}, \xi_{12}, \ldots, \xi_{1n}$ and $\xi_{21}, \xi_{22}, \ldots, \xi_{2m}$ respectively. Smirnov (1939) [25] proved the following two theorems :

If $F(x) = H(x)$, then

$$(4.1) \qquad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{-\infty<x<+\infty} (F_n(x) - H_m(x)) < y\right) = 1 - e^{-2y^2}$$

if $y > 0$, zero otherwise, and

$$(4.2) \qquad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{-\infty<x<+} |F_n(x) - H_m(x)| < y\right) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2y^2}$$

if $y > 0$, zero otherwise, where, in both cases, $\lim\limits_{(m,n;\rho)}$ is to mean the limit as $m \to \infty$ and $n \to \infty$ in such a way that $\frac{m}{n} \to \rho$. (The problem of determining the distributions of the respective statistics for finite values of n and m was solved by Gnedenko and Korolyuk (1951) [10] on the assumption that $n = m$.)

These results are used to test the statistical hypothesis that two random samples come from the same unknown population. Even if $F(x)$, the hypothetical c.d.f of the two random samples in question, were known we would not get more information out of the above theorems for they consider the differences $(F_n(x) - H_m(x))$ and $|F_n(x) - H_m(x)|$ with the same weight, regardless to the value of $F(x)$. Thus in this way the idea arises of considering the limiting distribution of the supremum of the quotients $(F_n(x) - H_m(x))/F(x)$ and $|F_n(x) - H_m(x)|/F(x)$. In examining the limiting distribution of these quotients a natural limitation on $F(x)$ is to be adopted. Namely, we restrict ourselves to an interval $x^{(t)} \leq x < +\infty$, where $F(x^{(t)}) = t > 0$. The value of t, however, can be an arbitrarily small positive value.

In this connection we are going to examine the limiting distribution of the quotient $(F_n(x) - H_m(x))/F(x)$ on the assumption that the number $t > 0$ is such that both $F_n(x)$ and $H_m(x)$ are greater than zero and less than one when $F(x^{(t)}) = t$. (Later on we are going to relax these conditions.) The following theorems are going to be proved :

Theorem 13. If $F(x) = H(x)$ and $n, m \to \infty$ so that $\frac{m}{n} \to \rho$ then

$$\lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x)} \frac{F_n(x) - H_m(x)}{F(x)} < y\right) = \begin{cases} \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{cd\,(1+\rho)}{c(1-d)\rho + d\,(1-c)}}} e^{-\frac{v^2}{2}}\,dv &, y > 0 \\ 0 &, y \leq 0 \end{cases}$$

$$= \Phi(y;\ c,\ d,\ \rho)$$

for all values of t, $0 < t < 1$, so that when $F(x^{(t)}) = t$

then $\begin{cases} F_n(x) = d & \text{with } 0 < d < 1 \\ H_m(x) = c & \text{with } 0 < c < 1\ . \end{cases}$

Theorem 14. If $F(x) = H(x)$ and $n, m \to \infty$ so that $\frac{m}{n} \to \rho$ then

$$(4.4) \quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x) \leq 1} \frac{F_n(x) - H_m(x)}{F(x)} < y\right)$$

$$= \frac{1}{\pi}\sqrt{\frac{f(1-e)\rho + e(1-f)}{ef(1+\rho)}} \int_{-\infty}^{y} e^{-\frac{ef(1+\rho)u^2}{2[f(1-e)\rho+e(1-f)]}} \int_0^{(y-u)\sqrt{\frac{edcf(1+\rho)}{(e-d)cf\rho+(f-c)ed}}} e^{-\frac{v^2}{2}}\, dv\, du\ , \qquad -\infty < y < +\infty,$$

$$= N(y;\ d,\ c,\ e,\ f,\ \rho)$$

for all t and l with $0 < t < l < 1$ where the other parameters satisfy the following conditions :

when $F(x^{(t)}) = t$ then $\begin{cases} F_n(x) = d & \text{with } 0 < d < 1 \\ H_m(x) = c & \text{with } 0 < c < 1 \end{cases}$

and when $F(x^{(l)}) = l$ then $\begin{cases} F_n(x) = e \\ H_m(x) = f \end{cases}$ where l being less than one, at most one of e and f can be equal to 1. If one of e and f is equal to 1, the appropriate one is replaced by 1 in $N(y;\ d,\ c,\ e,\ f,\ \rho)$.

To prove these theorems we need only the assumption that $F(x)$ is continuous but the statistical applicability of them requires the knowledge of $F(x)$. Just as when introducing the weight function $1/F(x)$ to $(F_n(x) - F(x))$ and

$|F_n(x) - F(x)|$ we no longer have the possibility of constructing confidence intervals for unknown continuous distribution functions; that is the applicability of the Kolmogorov - Smirnov theorems using the weight function $1/F(x)$ depends upon the knowledge of $F(x)$ but thereby they measure the asymptotic behaviour of the relative deviation of the population distribution function and that of the sample.

Assuming then that $F(x)$ is a known continuous c.d.f., the above theorems provide tests for verifying the statistical hypothesis that two random samples come from the same population with c.d.f. $F(x)$. From the proofs of these theorems it will become clear that we do not actually have to know $F(x)$ completely in order to apply them. It will be seen that if we can estimate the numerical values of $F(x)$ at all the sample points of the pooled sample of size $n+m$, gained by pooling the two samples of size n and m, it will be sufficient for the application of these theorems. The character of these tests is that they give upper bounds below which, if the hypothesis is true, $(F_n(x) - H_m(x))/F(x)$ must lie with probabilities $\Phi(y; c, d, \rho)$ and $N(y; c, d, e, f, \rho)$ respectively for given values of c, d, e, f, ρ and y. We shall always have to keep in mind though that the above theorems have only an asymptotic character and do not allow for the influence of the number of observations, whereas this influence may be very considerable if the numbers n and m are small. A successful Gnedenko and Korolyuk type examination [10] of this problem would answer these difficulties.

Proof of Theorem 13. Part A.

We have to examine the asymptotic behaviour of the random variable

$$(4.5) \qquad \sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x)} \frac{F_n(x) - H_m(x)}{F(x)}$$

Let us pool the two random samples $\xi_{11}, \xi_{12}, \ldots, \xi_{1n}$ and $\xi_{21}, \xi_{22}, \ldots, \xi_{2m}$ of size n and m respectively. Let this new sample of size $n+m$ be $\xi_1, \xi_2, \ldots, \xi_{n+m}$, which, on the assumption that both samples come from the same population with continuous c.d.f. $F(x)$ and on the assumption that the random variables of the two samples are mutually independent, is a random sample of size $n+m$ from a population with continuous c.d.f. $F(x)$.

Let the order statistic of the random sample of size n be $\xi_{1(1)} < \xi_{1(2)} < \ldots < \xi_{1(n)}$ and that of the random sample of size m be $\xi_{2(1)} < \xi_{2(2)} < \ldots < \xi_{2(m)}$ and also the order statistic of the pooled random sample of size $n+m$ be $\xi_{(1)} < \xi_{(2)} < \ldots < \xi_{(n+m)}$; that is using the notation of chapter 1 we have

$$(4.6) \qquad \begin{aligned} \xi_{1(i)} &= R_i(\xi_{11}, \xi_{12}, \ldots, \xi_{1n}) , \quad i = 1, 2, \ldots, n \\ \xi_{2(j)} &= R_j(\xi_{21}, \xi_{22}, \ldots, \xi_{2m}) , \quad j = 1, 2, \ldots, m \\ \xi_{(k)} &= R_k(\xi_1, \xi_2, \ldots, \xi_{n+m}) , \quad k = 1, 2, \ldots, n+m \end{aligned}$$

We shall also need the transformed forms of these random variables introduced in chapter 1 and adapted to the present situation as

$$\eta_{1i} = F(\xi_{1i}) \quad \text{and} \quad \zeta_{1i} = \log \frac{1}{\eta_{1i}} , \quad i = 1, 2, \ldots, n$$

$$(4.7) \quad \eta_{2j} = F(\xi_{2j}) \quad \text{and} \quad \zeta_{2j} = \log \frac{1}{\eta_{2j}} , \quad j = 1, 2, \ldots, m$$

$$\eta_{k} = F(\xi_{k}) \quad \text{and} \quad \zeta_{k} = \log \frac{1}{\eta_{k}} , \quad k = 1, 2, \ldots, n+m$$

and the corresponding order statistics

$$\eta_{1(i)} = F(\xi_{1(i)}) \quad \text{and} \quad \zeta_{1(i)} = \log \frac{1}{\eta_{1(n+1-i)}}, \quad i = 1, 2, \ldots, n$$

$$(4.8) \quad \eta_{2(j)} = F(\xi_{2(j)}) \quad \text{and} \quad \zeta_{2(j)} = \log \frac{1}{\eta_{2(m+1-j)}}, \quad j = 1, 2, \ldots, m$$

$$\eta_{(k)} = F(\xi_{(k)}) \quad \text{and} \quad \zeta_{(k)} = \log \frac{1}{\eta_{(n+m+1-k)}}, \quad k = 1,2,\ldots,n+m$$

where $F(x)$ is the continuous c.d.f. of the population from which the two random samples of size n and m and, therefore, also the pooled sample of size $n+m$ come.

Let us denote the empirical distribution function of the pooled sample of size $n+m$ by $F_{n+m}(x)$; that is we have

$$F_{n+m}(x) = \begin{cases} 0 , & \text{if } x < \xi_{(1)} \\ \dfrac{k}{n+m} , & \text{if } \xi_{(k)} \leq x < \xi_{(k+1)} \\ 1 , & \text{if } \xi_{(n+m)} \leq x \end{cases}$$

We remark here that $\xi_{(k)} = \xi_{1(i)}$ or $\xi_{(k)} = \xi_{2(j)}$ and that $k = i+j$; if $\xi_{(k)} = \xi_{1(i)}$ then we have $\xi_{2(j)} < \xi_{1(i)}$ and if $\xi_{(k)} = \xi_{2(j)}$ then $\xi_{1(i)} < \xi_{2(j)}$.

The random variables η_{1i}, η_{2j} and η_k of (4.7) are uniformly distributed in the interval $(0, 1)$ and, if $u = F(x)$, their sample distribution functions are

$$(4.9) \quad G_n(u) = F_n(F^{-1}(u)), \quad K_m = H_m(F^{-1}(u)) \quad \text{and}$$

$$G_{n+m}(u) = F_{n+m}(F^{-1}(u))$$

respectively, where $x = F^{-1}(u)$ is the inverse function of $u = F(x)$.

Now the limiting distribution of the random variable $\sup\limits_{t \leq F(x)} \dfrac{F_n(x) - H_m(x)}{F(x)}$ is identical with that of the random variable

$$(4.10) \qquad \sup_{t \leq u \leq 1} \frac{G_n(u) - K_m(u)}{u}$$

and the limiting distribution of the random variable of (4.10) is identical with that of the random variable

$$(4.11) \qquad \sup_{t \leq G_{n+m}(u) \leq 1} \frac{G_n(u) - K_m(u)}{u} \quad ;$$

that is to prove Theorem 13 it is sufficient to prove that

$$(4.12) \quad \lim_{(m,n;\ )} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq G_{n+m}(u) \leq 1} \frac{G_n(u) - K_m(u)}{u} < y\right) = \Phi(y;c,d,\rho)$$

To see this last step, that is the identity of (4.10) and (4.11), let us consider the event $|G_{n+m}(u) - u| \leq \epsilon$, i.e. the event $-\epsilon \leq G_{n+m}(u) - u \leq \epsilon$. From $0 < t \leq u \leq 1$ it follows that $|G_{n+m}(u) - t| \leq \epsilon$ or $|G_{n+m}(u) - t| \geq \epsilon$. In the first case there is nothing to prove and the second case can only result from $G_{n+m}(u) - t \geq \epsilon$. It follows then from $G_{n+m}(u) - u \leq \epsilon$ and $G_{n+m}(u) - t \geq \epsilon$ that $t \leq G_{n+m}(u) - \epsilon \leq u$ and thus we have

$$(4.13) \qquad \sup_{t+\epsilon \leq G_{n+m}(u)} \frac{G_n(u) - K_m(u)}{u} \leq \sup_{t \leq u} \frac{G_n(u) - K_m(u)}{u}$$

Let A be the event that $\sup_{t \leq u} \frac{G_n(u) - K_m(u)}{u} < y \sqrt{\frac{n+m}{nm}}$

and A' be the event that $\sup_{t+\epsilon \leq G_{n+m}(u)} \frac{G_n(u) - K_m(u)}{u} < y \sqrt{\frac{n+m}{nm}}$.

Then, by (4.13), $A \subseteq A'$ and if we let B be the event $|G_{n+m}(u) - u| \leq \epsilon$ then $AB \subseteq A'B$. But $A = A\overline{B} \cup AB \subseteq \overline{B} \cup A'B \subseteq \overline{B} \cup A'$, where $\overline{B}$ denotes the complementary event of B. Therefore, $P(A) \leq P(\overline{B}) + P(A')$, that is

$$(4.14) \qquad P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq u} \frac{G_n(u) - K_m(u)}{u} < y\right) \leq$$

$$P(\,|G_{n+m}(u) - u| > \epsilon\,) + P\left(\sqrt{\frac{nm}{n+m}} \sup_{t+\epsilon \leq G_{n+m}(u)} \frac{G_n(u) - K_m(u)}{u} < y\right)$$

Similarly, it can be shown that

$$(4.15) \qquad P\left(\sqrt{\frac{nm}{n+m}} \sup_{t-\epsilon \leq G_{n+m}(u)} \frac{G_n(u) - K_m(u)}{u} < y\right) \leq$$

$$P(\,|G_{n+m}(u) - u| > \epsilon\,) + P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq u} \frac{G_n(u) - K_m(u)}{u} < y\right)$$

Since $\lim_{n,m \to \infty} P(\,|G_{n+m}(u) - u| > \epsilon\,) = 0$, $\epsilon > 0$, it follows from (4.14) and (4.15) that

$$\overline{\lim_{(m,n;\rho)}} \; P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \le u} \frac{G_n(u) - K_m(u)}{u} < y\right) \le \Phi(y;c+\epsilon',d+\epsilon'',\ )$$

and

$$\underline{\lim_{(m,n;\rho)}} \; P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \le u} \frac{G_n(u) - K_m(u)}{u} < y\right) \ge \Phi(y;c-\epsilon',d-\epsilon'',\ )$$

if the statement of (4.12) is true, where ϵ' and ϵ'' are possible changes induced in the values of c and d by changing the value of t to $t \pm \epsilon$. Now ϵ can be chosen arbitrarily small and, as a result of that, ϵ' and ϵ'' are also made arbitrarily small. Also, the integral is a continuous function of its upper limit and so it follows that

$$(4.16) \qquad \lim_{(m,n;\rho)} \; P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \le u} \frac{G_n(u) - K_m(u)}{u} < y\right) = \Phi(y;c,d,\rho)$$

on the condition that (4.12) is true.

The identity of the random variables of (4.10) and (4.11) is explained in the following heuristic considerations too. According to (4.10) we consider the limiting distribution of the random variable $\sup_{t \le u \le 1} \frac{G_n(u) - K_m(u)}{u}$. Let us take the general case when t is positive, and arbitrarily small. If both $G_n(u)$ and $K_m(u)$ are zero for some subset of u in $t \le u \le 1$, then we have started the examination of the above random variable too soon and, in that subset of u, we get no real information on the behaviour of this random variable. The real examination starts when at least one of $G_n(u)$ and $K_m(u)$ is greater than zero and this is implied by the condition that the c.d.f. of the pooled sample $G_{n+m}(u)$ is such that $t \leq G_{n+m}(u) \leq 1$ for some however small number $t > 0$. This implies that we have the following random variable to start with : $\sup\limits_{t \leq G_{n+m}(u) \leq 1} \dfrac{G_n(u) - K_m(u)}{u}$. In (4.11) we have this random variable with the restriction that when $t = G_{n+m}(u)$ both $G_n(u)$ and $K_m(u)$ are positive as a result of the original assumptions of Theorem 13. This was the way we arrived at the idea of examining the identity of (4.10) and (4.11) in the above analytic way which culminates in the statement of (4.12). The proof of the equivalence of (4.10) and (4.11) is free of the restrictions of Theorem 13 on t and this enables us to attempt the relaxation of them later in this thesis.

To prove (4.12) we consider again its random variable

$$\sup_{t \leq G_{n+m}(u) \leq 1} \frac{G_n(u) - K_m(u)}{u} \tag{4.17}$$

If, in general, $t \leq G_{n+m}(u) \leq 1$ then at least one of the empirical distribution functions $G_n(u)$ and $K_m(u)$ is greater than zero and less than one at u when $t = G_{n+m}(u)$. Let $d = G_n(u)$ and $c = K_m(u)$ when $t = G_{n+m}(u)$ and, for the sake of Theorem 13, let us assume that both d and c are greater than zero and less than one. Thus examining the random variable of (4.12) for the set of u's for which we have $t \leq G_{n+m}(u) \leq 1$ also means the examination of this random variable for the set of u's for which $d \leq G_n(u) \leq 1$

and $c \leq K_m(u) \leq 1$ simultaneously. We are going to express this by writing (4.17) in the following form

$$(4.18) \qquad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u}$$

where $I = \{u : t \leq G_{n+m}(u) \leq 1,\ d \leq G_n(u) \leq 1,\ c \leq K_m(u) \leq 1\}$.

Now the value of $G_n(u) - K_m(u)$ changes only when the value of u passes a value $\eta_{(k)}$ of the pooled sample. This we express by writing (4.18) as

$$(4.19) \qquad \sup_{u\ I} \frac{G_n(u) - K_m(u)}{u}$$

$$= \max_{S} \left\{ \frac{G_n(\eta_{(k)}+0) - K_m(\eta_{(k)}+0)}{\eta_{(k)}},\ \frac{G_n(\eta_{(k+1)}-0) - K_m(\eta_{(k+1)}-0)}{\eta_{(k+1)}} \right\}$$

$$= \max_{S} \left\{ \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}},\ \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k+1)}} \right\}$$

where $S = \left\{k, i, j : t \leq \frac{k}{n+m} \leq 1,\ d \leq \frac{i}{n} \leq 1,\ c \leq \frac{j}{m} \leq 1\right\}$ and where, if $\frac{i}{n} - \frac{j}{m} > 0$, the expression $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ is used and, if $\frac{i}{n} - \frac{j}{m} < 0$, the expression $(\frac{i}{n} - \frac{j}{m})/\eta_{(k+1)}$ is used to find maximum. If $\frac{i}{n} - \frac{j}{m} = 0$, it is irrelevant which one of them is used. Moreover, having $t(n+m) \leq k \leq n+m$ and as a result of that, $nd \leq i \leq n$, $mc \leq j \leq m$, the maximum of the above expression is at least zero and can be found through examination of $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ in the indicated regions for k, i and j. Therefore we have

$$(4.20) \quad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \max_{S} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}$$

Thus, the examination of the random variable of (4.12) boils down to the examination of the order statistic of the pooled sample for $t(n+m) \leq k \leq n+m$ and, thereby, to the examination of the order statistics of the original two samples for $dn \leq i \leq n$ and $cm \leq j \leq m$ respectively, changing the value of $\frac{i}{n} - \frac{j}{m}$ of (4.20) for a given k according to the possibilities of having an $\eta_{1(i)}$ or an $\eta_{2(j)}$ in $\eta_{(k)}$, the k^{th} order statistic of the pooled sample. In any given practical situation, that is when we are having two random samples of size n and m, this maximum statistic of (4.20) is easily found and we are going to show now that its limiting distribution is given by (4.12).

Proof of Theorem 13. Part B : a Lemma.

The present form of the right hand side of the random variable of (4.20) does not lend itself to the method of A. Rényi, presented in chapter 1 and used so far in proofs of theorems of this thesis and which we also would like to continue using in proving this theorem. Toward this end, we are going to introduce a random variable which will always be greater than or equal to $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ of (4.20) for any given k and, therefore, its maximum will also have this property in relation to the maximum of $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ of (4.20) in the indicated regions for k, i and j. Also, we are going to introduce a random variable which will always be less than or equal to $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ of (4.20) for any given k and, therefore, its maximum will also have this property in relation to the maximum of $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ of (4.20) in the indicated regions for k, i and j. The form of these new random variables will be adaptable to the method of A. Rényi and this will enable us to derive their limiting distribution and that, in turn, will enable us to derive the limiting distribution of the right hand side of (4.20) and, thereby, to prove relation (4.12) which was shown to be sufficient for the proof of Theorem 13. In this connection we are going to verify

Lemma 1.

$$(4.21) \quad \max_{S} \left[\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right] \geq \max_{S} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}$$

$$(4.22) \quad \max_{S} \left[\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} \right] \leq \max_{S} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}$$

where $\eta_{1(i)}$, $\eta_{1(i+1)}$, $\eta_{2(j)}$, $\eta_{2(j+1)}$ and $\eta_{(k)}$ are as in (4.8) and S was defined in (4.19).

Proof of relation (4.21).

(a) Let $\frac{i}{n} - \frac{j}{m} > 0$ when $\eta_{(k)} = \eta_{1(i)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \geq \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i)}}, \quad \text{for } \eta_{(k)} = \eta_{1(i)} \text{ and so}$$

$\eta_{1(i)} < \eta_{2(j+1)}$. We have equality when $\eta_{(k)} = \eta_{1(i)}$ such that there is no $\eta_{2(j)}$ before $\eta_{1(i)}$; that is ~~when~~ $k = i$. In this case relation (a) above becomes $\dfrac{\frac{i}{n}}{\eta_{1(i)}} - 0 = \dfrac{\frac{i}{n} - 0}{\eta_{1(i)}}$.

This equality is impossible when we have the restriction of Theorem 13 on t, but we shall need this property of Lemma 1 that it remains valid when t is an arbitrarily small positive number.

(b) Let $\frac{i}{n} - \frac{j}{m} > 0$ when $\eta_{(k)} = \eta_{2(j)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} > \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j)}}, \quad \text{for } \eta_{(k)} = \eta_{2(j)} \text{ and so}$$

$\eta_{1(i)} < \eta_{2(j)} < \eta_{2(j+1)}$.

(c) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{1(i)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} > \frac{\frac{i}{m} - \frac{j}{m}}{\eta_{1(i)}}, \quad \text{for } \eta_{(k)} = \eta_{1(i)} \text{ and so}$$

$\eta_{1(i)} < \eta_{2(j+1)}$.

(d) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{2(j)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} > \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j)}}, \quad \text{for } \eta_{(k)} = \eta_{2(j)} \text{ and so}$$

$\eta_{1(i)} < \eta_{2(j)} < \eta_{2(j+1)}$.

(e) Let $\frac{i}{n} - \frac{j}{m} = 0$ when $\eta_{(k)} = \eta_{1(i)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} > \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i)}} = 0 \;,\quad \text{for} \quad \eta_{(k)} = \eta_{1(i)} \quad \text{and so}$$

$\eta_{1(i)} < \eta_{2(j+1)}$.

(f) Let $\frac{i}{n} - \frac{j}{m} = 0$ when $\eta_{(k)} = \eta_{2(j)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} > \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j)}} = 0 \;,\quad \text{for} \quad \eta_{(k)} = \eta_{2(j)} \quad \text{and so}$$

$\eta_{1(i)} < \eta_{2(j)} < \eta_{2(j+1)}$

Relations (a), (b), (c), (d), (e) and (f) imply statement of (4.21).

Proof of relation (4.22).

(a) Let $\frac{i}{n} - \frac{j}{m} > 0$ when $\eta_{(k)} = \eta_{1(i)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} < \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i)}} \;,\quad \text{for} \quad \eta_{(k)} = \eta_{1(i)} \quad \text{and so}$$

$\eta_{2(j)} < \eta_{1(i)} < \eta_{1(i+1)}$.

(b) Let $\frac{i}{n} - \frac{j}{m} > 0$ when $\eta_{(k)} = \eta_{2(j)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} < \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j)}} \;,\quad \text{for} \quad \eta_{(k)} = \eta_{2(j)} \quad \text{and so}$$

$\eta_{2(j)} < \eta_{1(i+1)}$.

(c) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{1(i)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} < \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i)}}\,, \quad \text{for } \eta_{(k)} = \eta_{1(i)} \text{ and so}$$

$\eta_{2(j)} < \eta_{1(i)} < \eta_{1(i+1)}$.

(d) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{2(j)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} \leq \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j)}}\,, \quad \text{for } \eta_{(k)} = \eta_{2(j)} \text{ and so}$$

$\eta_{2(j)} < \eta_{1(i+1)}$. We have equality when $\eta_{(k)} = \eta_{2(j)}$ such that there is no $\eta_{1(i)}$ before $\eta_{2(j)}$; i.e. when $k = j$. In this case relation (d) above becomes $0 - \frac{\frac{j}{m}}{\eta_{2(j)}} = \frac{0 - \frac{j}{m}}{\eta_{2(j)}}$.

This equality is impossible when we have the restriction of Theorem 13 on t, but we shall need this property of Lemma 1 that it remains valid when t is an arbitrarily small positive number.

(e) Let $\frac{i}{n} - \frac{j}{m} = 0$ when $\eta_{(k)} = \eta_{1(i)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} < \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i)}} = 0\,, \quad \text{for } \eta_{(k)} = \eta_{1(i)}$$

and so $\eta_{2(j)} < \eta_{1(i)} < \eta_{1(i+1)}$.

(f) Let $\frac{i}{n} - \frac{j}{m} = 0$ when $\eta_{(k)} = \eta_{2(j)}$. Then

$$\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} < \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j)}} = 0\,, \quad \text{for } \eta_{(k)} = \eta_{2(j)} \text{ and so}$$

$\eta_{2(j)} < \eta_{1(i+1)}$.

Relations (aO, (b), (c), (d), (e) and (f) imply statement of (4.22) and this completes the proof of the above Lemma.

Proof of Theorem 13. Part C.

Taking the entities of the above Lemma let us introduce the following notations. Let

$$A \text{ be the event } \sqrt{\frac{nm}{n+m}} \max_S \left(\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right) < y ,$$

$$B \text{ be the event } \sqrt{\frac{nm}{n+m}} \max_S \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} < y , \text{ and}$$

$$C \text{ be the event } \sqrt{\frac{nm}{n+m}} \max_S \left[\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} \right] < y .$$

Relations (4.21) and (4.22) of the above Lemma imply that $A \subseteq B \subseteq C$ and therefore we have $P(A) \leq P(B) \leq P(C)$, that is

$$(4.23) \quad P\left(\sqrt{\frac{nm}{n+m}} \max_S \left(\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right) < y \right) \leq P\left(\sqrt{\frac{nm}{n+m}} \max_S \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} < y \right)$$

and

$$(4.24) \quad P\left(\sqrt{\frac{nm}{n+m}} \max_S \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} < y \right) \leq P\left(\sqrt{\frac{nm}{n+m}} \max_S \left[\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} \right] < y \right)$$

Now we are going to show that

$$(4.25) \quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \max_S \left[\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right] < y \right) = \Phi(y;c,d,\rho)$$

and that

$$(4.26)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}}\ \max_S \left[\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}}\right] < y\right) = \Phi(y;c,d,\rho)$$

Assuming for a moment that statements of (4.25) and (4.26) are true it follows from (4.23) and (4.24) that

$$(4.27)\quad \underline{\lim}_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}}\ \max_S \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} < y\right) \geq \Phi(y;c,d,\rho)$$

and

$$(4.28)\quad \overline{\lim}_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}}\ \max_S \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} < y\right) \leq \Phi(y;\ c,\ d,\rho)$$

and, in turn, (4.27) and (4.28) imply that

$$(4.29)\quad \lim_{(m,n;\ )} P\left(\sqrt{\frac{nm}{n+m}}\ \max_S \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} < y\right) = \Phi(y;c,d,\rho)$$

But, by (4.20), the random variable of (4.29) is equivalent to that of (4.12) and, as we have already stated, it is sufficient to prove (4.12) in order to prove Theorem 13. This means that we have proved Theorem 13 provided that the statements of (4.25) and (4.26) are true.

To prove (4.25) we note that the limiting distribution of its random variable is identical with that of the random variable

$$(4.30)\quad \max_S \left(\log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}}\right),\quad \eta_{2(m+1)} = 1$$

Using the notation and results of chapter 1 we consider

$$(4.31)\quad \log \frac{1}{\eta_{1(i)}} = \sum_{v=i}^{n} \frac{\delta_{n+1-v}}{v}$$

and

$$(4.32) \qquad \log \frac{1}{\eta_{2(j+1)}} = \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}}{s}$$

where the δ_{n+1-v}, δ_{m+1-s} are mutually independent exponentially distributed random variables with c.d.f. $1 - e^{-x}$, $x \geq 0$. Therefore we have

$$(4.33) \qquad \begin{aligned} M \log \frac{1}{\eta_{1(i)}} &= \sum_{v=i}^{n} \frac{1}{v} \\ D^2 \log \frac{1}{\eta_{1(i)}} &= \sum_{v=i}^{n} \frac{1}{v^2} \end{aligned}$$

as the mean and variance of (4.31) respectively and

$$(4.34) \qquad \begin{aligned} M \log \frac{1}{\eta_{2(j+1)}} &= \sum_{s=j+1}^{m} \frac{1}{s}, \quad \text{where} \quad \sum_{s=m+1}^{m} \frac{1}{s} = 0 \\ D^2 \log \frac{1}{\eta_{2(j+1)}} &= \sum_{s=j+1}^{m} \frac{1}{s^2}, \quad \text{where} \quad \sum_{s=m+1}^{m} \frac{1}{s^2} = 0, \end{aligned}$$

as the expected value and variance of (4.32) respectively.

Consider now the sequences of random variables

$$(4.35) \qquad \begin{aligned} &\frac{\delta_{n+1-v}-1}{v}, \quad v = i, \ldots, n \\ &\frac{\delta_{m+1-s}-1}{s}, \quad s = j+1, \ldots, m \end{aligned}$$

These sequences satisfy Lindeberg's condition (given in chapter 2) and considering these two sequences of mutually independent random variables as one sequence, the random variables of this sequence are again mutually independent and satisfy Lindeberg's condition. Therefore we apply

Theorem 5 of chapter 2 with

$$(4.36)\quad \max_S \left\{ \left[\left(\log \frac{1}{\eta_{1(i)}} - \sum_{v=i}^{n} \frac{1}{v} \right) - \left(\log \frac{1}{\eta_{2(j+1)}} - \sum_{s=j+1}^{m} \frac{1}{s} \right) \right] \right\}$$

$$= \max_S \left(\sum_{v=i}^{n} \frac{\delta_{n+1-v}^{-1}}{v} - \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}^{-1}}{s} \right)$$

as $\max S_{n+m,k}$ of this theorem. Therefore we have

$$(4.37)\quad \lim_{(m,n;\rho)} P\left(\max_S \left[\left(\log \frac{1}{\eta_{1(i)}} - \sum_{v=i}^{n} \frac{1}{v} \right) - \left(\log \frac{1}{\eta_{2(j+1)}} - \sum_{s=j+1}^{m} \frac{1}{s} \right) \right] < z \sqrt{ \sum_{nd \le i \le n} \frac{1}{i^2} + \sum_{mc \le j \le m-1} \frac{1}{(j+1)^2} } \right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{v^2}{2}} dv \ , \quad \text{if} \quad z > 0 \ , \quad \text{zero otherwise.}$$

But, if $i \ge dn$ and $0 < d < 1$, we have by Euler's summation formula that

$$(4.38)\quad \sum_{v=i}^{n} \frac{1}{v} = \log n - \log i + o(\tfrac{1}{n}) = \log \frac{n}{i} + o(\tfrac{1}{n})$$

and

$$(4.39)\quad \sqrt{\sum_{nd \le i \le n} \frac{1}{i^2}} = \sqrt{\frac{1}{dn} - \frac{1}{n}} + o(\tfrac{1}{n}) = \sqrt{\frac{1-d}{dn}} + o(\tfrac{1}{n})$$

Similarly, if $j \ge cm$ and $0 < c < 1$, we have by Euler's summation formula that

$$(4.40)\quad \sum_{s=j+1}^{m} \frac{1}{s} = \log m - \log (j+1) + o(\tfrac{1}{m}) = \log \frac{m}{j+1} + o(\tfrac{1}{m})$$

and

$$(4.41)\quad \sqrt{\sum_{mc \le j \le m-1} \frac{1}{(j+1)^2}} = \sqrt{\frac{1}{cm+1} - \frac{1}{m}} + o(\tfrac{1}{m}) = \sqrt{\frac{m-cm-1}{(cm+1)m}} + o(\tfrac{1}{m})$$

But $\log \frac{m}{j+1} = \log \frac{1}{\frac{j}{m} + \frac{1}{m}}$ and $\frac{m-cm-1}{(cm+1)m} = \frac{1-c-\frac{1}{m}}{cm(1+\frac{1}{cm})}$, that

is when m is large we have

$$(4.40)' \qquad \sum_{s=j+1}^{m} \frac{1}{s} = \log \frac{m}{j}$$

and

$$(4.41)' \qquad \sqrt{\sum_{mc \leq j \leq m-1} \frac{1}{(j+1)^2}} = \sqrt{\frac{1-c}{cm}}$$

Therefore, using (4.38), (4.39), (4.40), (4.41), (4.40)' and (4.41)', it follows from (4.37) that

$$(4.42) \quad \lim_{(m,n;\rho)} P\left(\max_S \left[\left[\log \frac{1}{\eta_{1(i)}} - \log \frac{n}{i}\right]\right] - \left[\left[\log \frac{1}{\eta_{2(j+1)}} - \log \frac{m}{j}\right]\right] < z\sqrt{\frac{1-d}{dn} + \frac{1-c}{cm}}\right)$$

$$= \lim_{(m,n;\rho)} P\left(\max_S \left(\log \frac{i/n}{\eta_{1(i)}} - \log \frac{j/m}{\eta_{2(j+1)}}\right) < z\sqrt{\frac{n+m}{nm}}\sqrt{\frac{c(1-d)\frac{m}{n+m} + d(1-c)\frac{n}{n+m}}{dc}}\right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{v^2}{2}} dv \quad \text{if} \quad z > 0 \ , \quad \text{zero otherwise.}$$

Letting $y = z\sqrt{\frac{c(1-d)\frac{m}{n+m} + d(1-c)\frac{n}{n+m}}{dc}}$ and taking into consideration that $\frac{m}{n} \to \rho$ as $n, m \to \infty$ we get

$$(4.43) \quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \max_S \left(\log \frac{i/n}{\eta_{1(i)}} - \log \frac{j/m}{\eta_{2(j+1)}}\right) < y\right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{dc(1+\rho)}{c(1-d)\rho + d(1-c)}}} e^{-\frac{v^2}{2}}\, dv \quad \text{if} \quad y > 0 \;, \quad \text{zero otherwise.}$$

This, by (4.30), proves the assertion of (4.25). The assertion of (4.26) can be proved in exactly the same way. Having thus proved (4.25) and (4.26), taking into consideration (4.27) and (4.28) we also verified (4.29) and this, by the remarks after (4.29), proves Theorem 13.

If in the above theorem we put $n = m$ then $\rho = 1$ and Theorem 13 becomes

Theorem 13'. If $F(x) = H(x)$ then

$$(4.44) \quad \lim_{n\to\infty} P\left(\sqrt{\frac{n}{2}} \sup_{t \leq F(x)} \frac{F_n(x) - H_n(x)}{F(x)} < y\right) = \begin{cases} \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{2cd}{c(1-d)+d(1-c)}}} e^{-\frac{v^2}{2}}\, dv & \text{if} \quad y>0 \\ 0 & \text{if} \quad y \leq 0 \end{cases}$$

for all values of, $0 < t < 1$, so that when

$$F(x^{(t)}) = t \quad \text{then} \begin{cases} F_n(x) = d & \text{with} \quad 0 < d < 1 \;, \\ H_n(x) = c & \text{with} \quad 0 < c < 1 \;. \end{cases}$$

<u>Remarks on and some generalizations of Theorems 13 and 13'</u>. If we have two random samples of size n and m such that all the observations of one are less than all the observations of the other then Theorems 13 and 13' are not applicable in their present forms, for then the set of x's for which $t \leq F(x)$, $0 < t < 1$, so that when $t = F(x^{(t)})$ we have $d = F_n(x)$ and $c = H_m(x)$ such that both d and c are greater than zero and less than one is empty. On page 47

we made the remark that the proof of the equivalence of (4.10) and (4.11) was free of the restrictions of Theorem 13 on t, and that this would enable us to attempt the relaxation of them. This is what we are going to do now.

We consider here the possibility of dropping the restriction that both $d = F_n(x)$ and $c = H_m(x)$ are less than one when both of them are greater than zero at $x^{(t)}$ with $t = F(x^{(t)})$. Since we have $0 < t < 1$, only one of $d = F_n(x)$ and $c = H_m(x)$ can be equal to one when $t = F(x^{(t)})$ or, making use of the above remark regarding the equivalence of (4.10) and (4.11), when $t = G_{n+m}(u)$ at most one of the values $d = G_n(u)$ and $c = K_m(u)$ is equal to one. Repeating the argument of (4.18), (4.19) and (4.20) we have again

$$(4.45) \qquad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \max_{S} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}$$

where $I = \left\{u : t \leq G_{n+m}(u) \leq 1,\ d \leq G_n(u) \leq 1,\ c \leq K_m(u) \leq 1\right\}$,

$$S = \left\{k, i, j : t \leq \frac{k}{n+m} \leq 1,\ d \leq \frac{i}{n} \leq 1,\ c \leq \frac{j}{m} \leq 1\right\} \text{ and}$$

where we still assume, as in Theorems 13 and 13', that both of the values d and c are greater than zero but do not exclude the possibility of having one of them equal to one. Assuming then that both $d = \frac{i}{n}$ and $c = \frac{j}{m}$ are greater than zero when $t = \frac{k}{n+m}$ $(t > 0)$ we can have the following three mutually exclusive possibilities :

(i) $d = \frac{i}{n} < 1$ and $c = \frac{j}{m} < 1$

or

$$\text{(ii)} \quad d = \frac{i}{n} = 1 \quad \text{and} \quad c = \frac{j}{m} < 1$$

or

$$\text{(ii)} \quad d = \frac{i}{n} < 1 \quad \text{and} \quad c = \frac{j}{m} = 1$$

Introducing $S' = \left\{ k, j \ : \ t \leq \frac{k}{n+m} \leq 1, \quad c \leq \frac{j}{m} \leq 1 \right\}$ and $S'' = \left\{ k, i \ : \ t \leq \frac{k}{n+m} \leq 1, \quad d = \frac{i}{n} \leq 1 \right\}$ for cases (ii) and (iii) respectively, (4.45) can be written as

$$(4.46) \quad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \begin{cases} \max\limits_{S} \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}, & \text{if } d < 1, \quad c < 1 \\[2ex] \max\limits_{S'} \dfrac{1 - \frac{j}{m}}{\eta_{(k)}}, & \text{if } d = 1, \quad c < 1 \\[2ex] \max\limits_{S'''} \dfrac{\frac{i}{n} - 1}{\eta_{(k)}}, & \text{if } d < 1, \quad c = 1 \end{cases}$$

Case (i) is handled by Theorems 13 and 13' in their present forms.

Case (ii), where both d and c are greater than zero and $d = 1$, $c < 1$, can be stated as

$$(4.47) \quad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \max_{S'} \frac{1 - \frac{j}{m}}{\eta_{(k)}}$$

In terms of Lemma 1 this implies that we are to examine the limiting distribution of the random variable

$$(4.48) \qquad \max_{S'} \left(\frac{1}{\eta_{1(n)}} - \frac{j/m}{\eta_{2(j+1)}} \right)$$

and

$$(4.49) \qquad \max_{S'}\left(\frac{1}{\eta_{1(n+1)}} - \frac{j/m}{\eta_{2(j)}}\right) = \max_{S'}\left(1 - \frac{j/m}{\eta_{2(j)}}\right)$$

where the last identity follows from $\eta_{1(n+1)} = 1$.

The limiting distribution of (4.48) is equivalent to that of

$$(4.50) \qquad \max_{S'}\left(\log \frac{1}{\eta_{1(n)}} - \log \frac{j/m}{\eta_{2(j+1)}}\right)$$

From now on the argument follows that of examining the limiting distribution of (4.30). The statement of (4.31) becomes

$$(4.51) \qquad \log \frac{1}{\eta_{1(n)}} = \frac{\delta_1}{n},$$

and the statements of (4.32) and (4.34) remain exactly the same while (4.33) becomes

$$(4.52) \qquad \begin{aligned} M \ \log \frac{1}{\eta_{1(n)}} &= \frac{1}{n} \\ D^2 \ \log \frac{1}{\eta_{1(n)}} &= \frac{1}{n^2} \end{aligned}$$

as the expected mean and variance of (4.51) respectively. Thus in the limit $\log \frac{\delta_1}{n}$ has mean and variance equal to zero. In the light of this (4.36) becomes

$$(4.53) \qquad \max_{S'}\left\{\log \frac{1}{\eta_{1(n)}} - \left(\log \frac{1}{\eta_{2(j+1)}} - \sum_{s=j+1}^{m} \frac{1}{s}\right)\right\}$$

$$= \max_{S'} \left(\frac{\delta_1}{n} - \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}^{-1}}{s} \right)$$

while (4.42) becomes

$$(4.54) \quad \lim_{(m,n;\rho)} P\left(\max_{S'} \left[\log \frac{1}{\eta_{1(n)}} - \log \frac{j/m}{\eta_{2(j+1)}}\right] < z\sqrt{\frac{1-c}{cm}}\right)$$

$$= \lim_{(m,n;\rho)} P\left(\max_{S'} \left[\log \frac{1}{\eta_{1(n)}} - \log \frac{j/m}{\eta_{2(j+1)}}\right] < z\sqrt{\frac{n+m}{nm}}\sqrt{\frac{1-c}{c}\,\frac{n}{n+m}}\right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^z e^{-\frac{v^2}{2}}\, dv \;, \quad \text{if } z > 0 \;, \text{ zero otherwise.}$$

Letting $y = z\sqrt{\frac{1-c}{c}\,\frac{n}{n+m}}$ and taking into account that $\frac{m}{n} \to \rho$ as $n, m \to \infty$ we get

$$(4.55) \quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \max_{S'} \left[\log \frac{1}{\eta_{1(n)}} - \log \frac{j/m}{\eta_{2(j+1)}}\right] < y\right)$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{c(1+\rho)}{1-c}}} e^{-\frac{v^2}{2}}\, dv \quad \text{if } y > 0 \;, \text{ zero otherwise.}$$

This, by (4.50), is the limiting distribution of (4.48). It is easily seen that the limiting distribution of (4.49) is the same as that of (4.48). It follows then from Lemma 1 and using the argument of (4.23), (4.24), (4.25), (4.26), (4.27), (4.28) and (4.29) that the limiting distribution of the random variable of (4.47) is as stated in (4.55).

Case (iii), where both d and c are greater than zero and $d < 1$, $c = 1$, can be stated as

$$(4.56) \quad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \max_{S''} \frac{\frac{i}{n} - 1}{\eta_{(k)}} \;.$$

where S'' was defined in (4.46). An argument similar to that of case (ii) shows that the limiting distribution of (4.56) is as follows

$$(4.57)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{u\in I} \frac{G_n(u) - K_m(u)}{u} < y\right)$$
$$= \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{d(1+\rho)}{(1-d)\rho}}} e^{-\frac{v^2}{2}}\, dv \,, \text{ if } y > 0 \,, \text{ zero otherwise.}$$

A look at (4.3) shows that if $d = 1$, $c < 1$, that is in case (ii), then $\Phi(y;\, c,\, d=1,\rho)$ = right hand side of (4.55) and when $d < 1$ and $c = 1$, that is in case (iii), then $\Phi(y;\, c=1,\, d,\rho)$ = right hand side of (4.57). Thus Theorem 13 and therefore Theorem 13' can be extended to the case when $d = 1$ or $c = 1$, both of them are being greater than zero. Equating d or c in $\Phi(y;\, c,\, d,\rho)$ to 1 would not be valid without the verification of statements (4.55) and (4.57), for the proof of Theorem 13 relies on the fact that we have c and d such that $0 < c < 1$, $0 < d < 1$, when using Euler's summation formula. So we have the following extension of Theorem 13:

Theorem 15. If $F(x) = H(x)$ and $n, m \to \infty$ so that $\frac{m}{n} \to \rho$ then

$$(4.58)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t\leq F(x)} \frac{F_n(x) - H_m(x)}{F(x)} < y\right) = \Phi(y;c,d,\rho)$$

for all values of t, $0 < t < 1$, so that when

$F(x^{(t)}) = t$ then $\begin{cases} F_n(x) = d & \text{with } 0 < d \leq 1 \\ H_m(x) = c & \text{with } 0 < c \leq 1 \end{cases}$. If any one of d and c is equal to 1, we put $d = 1$ or $c = 1$ in $\Phi(y;\, c,\, d, \rho)$. Since we have $0 < t < 1$, at most one of the values d and c can be equal to 1.

A similar extension of Theorem 13' is obvious. Were we able to derive now the limiting distribution of the random variable of (4.11), which is given as

$$(4.59) \qquad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u},$$

for any arbitrarily small positive number t, we would have a complete generalization of Theorems 13 and 13'. This is what we are going to attempt next.

We recall first that $I = \{u : t \leq G_{n+m}(u) \leq 1,\ d \leq G_n(u) \leq 1,\ c \leq K_m(u) \leq 1\}$, $0 < t < 1$. When $t = G_{n+m}(u)$ then at least one (but not necessarily both) of the sample distribution functions $G_n(u)$ and $K_m(u)$ is greater than zero. So far, in theorems 13, 13' and 15, we have imposed the restriction on d and c, which are the values of $G_n(u)$ and $K_m(u)$ respectively for the value of u for which $t = G_{n+m}(u)$, that both of them are greater than zero. We are going to drop this restriction now. Thus we are going to examine the limiting distribution of the random variable

$$(4.60) \qquad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u},$$

where, as before, $I = \{u : t \leq G_{n+m}(u) \leq 1,\ d \leq G_n(u) \leq 1,$

$c \leq K_m(u) \leq 1\}$, $0 < t < 1$, and d and c are such that

$$\left.\begin{aligned} d &= G_n(u) \\ c &= K_m(u) \end{aligned}\right\} \text{ for the value of } u \text{ for which } t = G_{n+m}(u)$$

and, since $0 < t < 1$, at least one of c and d is non-zero.

We again have

$$(4.61) \quad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \max_{S} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}},$$

where, as before, $S = \left\{k,i,j : t \leq \frac{k}{n+m} \leq 1,\ d \leq \frac{i}{n} \leq 1,\ c \leq \frac{j}{m} \leq 1\right\}$

but with the above explained relaxation of previous conditions on d and c. Due to this relaxation of conditions on d and c we have the following possibilities for them when $t = \frac{k}{n+m}$:

(i) $d = \frac{i}{n} > 0$ and $c = \frac{j}{m} > 0$

(ii) $d = \frac{i}{n} = 0$ and $c = \frac{j}{m} > 0$

(iii) $d = \frac{i}{n} > 0$ and $c = \frac{j}{m} = 0$

These are three mutually exclusive possibilities for d and c. Introducing $S^* = \left\{k,i,j : t \leq \frac{k}{n+m} \leq 1,\ 0 \leq \frac{i}{n} \leq 1,\ c \leq \frac{j}{m} \leq 1\right\}$ and $S^{**} = \left\{k,i,j : t \leq \frac{k}{n+m} \leq 1,\ d \leq \frac{i}{n} \leq 1,\ 0 \leq \frac{j}{m} \leq 1\right\}$ for cases (ii) and (iii) respectively (4.61) can be written as

$$(4.62) \quad \sup_{u \in I} \frac{G_n(u) - K_m(u)}{u} = \begin{cases} \max\limits_{S} \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}, & \text{if } d > 0,\ c > 0 \\[2ex] \max\limits_{S^*} \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}, & \text{if } d = 0,\ c > 0 \\[2ex] \max\limits_{S^{**}} \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}, & \text{if } d > 0,\ c = 0 \end{cases}$$

Case (i), where both d and c are greater than zero, is handled by Theorems 13, 13' and 15.

Let us consider now case (ii), where $d = 0$, $c > 0$ when $t = \frac{k}{n+m}$. The statements (4.21) and (4.22) of Lemma 1 are valid in this case too but A. Rényi's method cannot be immediately used to derive the limiting distribution of the random variables

$$\max_{S^*}\left(\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}}\right), \quad \max_{S^*}\left(\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}}\right)$$

which, according to the statement of Lemma 1, are to be examined when $d = 0$, $c > 0$. To handle this problem let $t' = \frac{k}{n+m}$ be the smallest positive number such that $d' = \frac{i}{n}$, $c' = \frac{j}{m}$ and both of them are greater than zero. In this situation the appropriate case of (4.62) can be written as

$$(4.63) \quad \max_{S^*} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} = \max\left\{\max_{S_1^*} \frac{0 - \frac{j}{m}}{\eta_{(k)}}, \ \max_{S_2^*} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}\right\}$$

where $S_1^* = \left\{k, j \ : \ t \leq \frac{k}{n+m} < t', \ c \leq \frac{j}{m} < c'\right\}$ and

$S_2^* = \left\{k, i, j \ : \ t' \leq \frac{k}{n+m} \leq 1, \ d' \leq \frac{i}{n} \leq 1, \ c' \leq \frac{j}{m} \leq 1\right\}$. Now having $t'(n+m) \leq k \leq n+m$ and, as a result of that, $nd' \leq i \leq n$, $mc' \leq j \leq m$ in S_2^*, the second random variable of the right hand side of (4.63) is at least zero while the first random variable of it is always negative. We can pick, therefore, the second random variable of the right hand side of (4.63) as the one which is going to provide maximum for us. Thus (4.63) can be written as

$$(4.63)' \quad \max_{S^*} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} = \max_{S_2^*} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}},$$

where, as given above, t' in S_2^* is the smallest positive number such that both d' and c' in S_2^* are greater than zero. This means that we reduced case (ii) to case (i) and Theorems 13, 13' and 15 hold with $t = t'$, $d = d'$ and $c = c'$.

Concerning case (iii), where $d > 0$, $c = 0$, using the argument of case (ii), mutatis mutandis, the appropriate case of (4.62) can be written, analogously to (4.63), as

$$(4.64) \quad \max_{S^{**}} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} = \max \left\{ \max_{S_1^{**}} \frac{\frac{i}{n} - 0}{\eta_{(k)}}, \max_{S_2^{**}} \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right\}$$

where $S_1^{**} = \left\{ k, i : t \leq \frac{k}{n+m} \leq t'', d \leq \frac{i}{n} \leq d'' \right\}$ and

$S_2^{**} = \left\{ k, i, j : t'' \leq \frac{k}{n+m} \leq 1, d'' \leq \frac{i}{n} \leq 1, c'' \leq \frac{j}{m} \leq 1 \right\}$ and

where $t'' = \frac{k}{n+m}$ is the smallest positive number such that

$d'' = \frac{i}{n}$, $c'' = \frac{j}{m}$ and both of them are greater than zero. Unfortunately, there seems to be no way of choosing any one of the random variables of the right hand side of (4.64) as maximum of the two. Theoretically speaking either one of them can turn out to be the maximum of the two or when they would be equal either one would be satisfactory for deriving the limiting distribution of the left hand side of (4.64). In a given practical example we could of course spot the appropriate one for examination and to handle these possibilities we could derive here the possible limiting distributions for the appropriate situations. To state things exactly we can have the following three mutually exclusive possibilities for (4.64):

$$(4.65) \quad \max_{S^{**}} \frac{\frac{1}{n} - \frac{1}{m}}{\eta_{(k)}} = \begin{cases} \max\limits_{S_1^{**}} \dfrac{\frac{1}{n}}{(k)} \quad \ldots \quad (1) \quad \text{or} \\[2ex] \max\limits_{S_2^{**}} \dfrac{\frac{1}{n} - \frac{1}{m}}{(k)} \quad \ldots \quad (2) \quad \text{or} \\[2ex] \text{either one of } (1) \text{ and } (2) \\ \text{when they are equal.} \end{cases}$$

The limiting distribution of the random variable of (2) is handled by the extended form of Theorem 13, that is Theorem 15 holds with $t = t''$, $d = d''$ and $c = c''$. In case of (1) we would have to consider the possibilities of having : (a) $d < 1$ and $d'' < 1$, (b) $d = 1$ and, therefore, $d'' = 1$, (c) $d < 1$ and $d'' = 1$. Thus the limiting distribution of the random variable of (1) would have three different forms. Namely in case (a) Theorem 3 of chapter 2 would hold with $a = d$ and $b = d''$, in case (c) Theorem 1 of chapter 2 would hold with $a = d$ and in case (b) we would have to examine the "limiting" distribution of $\max\limits_{S_1^{**}} \dfrac{\frac{1}{n}}{\eta_{(k)}} = \dfrac{1}{\eta_{(k)}}$ and could possibly use the exact distribution of $\log \dfrac{1}{\eta_{(k)}}$ for decision problems which was proved to be the exponential law with c.d.f. $1 - e^{-x}$. In case of equality of the random variables of (1) and (2) the limiting distribution can be taken as that of either of them, and would no doubt best be taken as the more convenient one, which is the one of case (2) handled by Theorem 15 as explained above. All these limiting distributions are conditional ones corresponding to the specified possible situations.

It is clear that, because of these conditional limiting distribution statements, an attempt to give a completely general form of Theorem 13 would become very cumbersome. We are going to propose instead a convention which will enable us to formulate the desired generalization of Theorem 13 in a relatively simple manner.

We are trying to derive the limiting distribution of $\sup_{t \leq F(x)} (F_n(x) - H_m(x))/F(x)$ or, equivalently, that of

$\sup_{u \in I} (G_n(u) - K_m(u))/u$ for any arbitrarily small positive number t, thereby relaxing our conditions on d and c as stated in (4.60). We did not succeed in this attempt because of the difficulties encountered in case (iii) of (4.62), after successfully handling cases (i) and (ii) of (4.62). This troublesome case (iii) would reduce to well behaving case (ii) if we would adapt the following convention. If in (4.62), when $t = \frac{k}{n+m}$, we would have $d = \frac{i}{n} > 0$ and $c = \frac{j}{m} = 0$, that is case (iii), let us write the appropriate statement of (4.62) as

$$(4.66) \quad \sup_{u \in I} \frac{K_m(u) - G_n(u)}{u} = \max_{S^{**}} \frac{\frac{j}{m} - \frac{i}{n}}{\eta_{(k)}}$$

which amounts to interchanging $F_n(x)$ and $H_m(x)$ in the proposed extended statement of Theorem 13; that is instead of starting with $\sup_{t \leq F(x)} (F_n(x) - H_m(x))/F(x)$ we would start with $\sup_{t \leq F(x)} (H_m(x) - F_n(x))/F(x)$ in our attempt to generalize Theorem 13. This is not a restriction, for we can set up our original random variable with $F_n(x)$ and $H_m(x)$ in any order in it. Only because of the method of proof we are trying to use here we would want them the way given in (4.66). Having got (4.66), instead of (4.64) we could have the following relation

$$(4.67) \quad \max_{S^{**}} \frac{\frac{j}{m} - \frac{i}{n}}{\eta_{(k)}} = \max\left\{ \max_{S_1^{**}} \frac{0 - \frac{i}{n}}{\eta_{(k)}}, \ \max_{S_2^{**}} \frac{\frac{j}{m} - \frac{i}{n}}{\eta_{(k)}} \right\}$$

where, again, S_1^{**}, S_2^{**}, t'', d'' and c'' are as defined in (4.64). Following the argument of (4.63)' we can write (4.67) as

$$(4.67)' \quad \max_{S^{**}} \frac{\frac{j}{m} - \frac{i}{n}}{\eta_{(k)}} = \max_{S_2^{**}} \frac{\frac{j}{m} - \frac{i}{n}}{\eta_{(k)}}$$

and theorems 13, 13' and 15 hold with $t = t''$, $d = d''$ and $c = c''$ in this modified form of case (iii) of (4.62).

We have, therefore, through Theorem 15, the following generalized form of Theorem 13.

Theorem 16. If $F(x) = H(x)$ and $n, m \to \infty$ so that $\frac{m}{n} \to \rho$ then

$$(4.68)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t\leq F(x)} \frac{F_n(x) - H_m(x)}{F(x)} < y\right) =$$

$\Phi(y; c, d, \rho)$, if $d > 0$, $c > 0$ for all values of t, $0 < t < 1$, so that when

$$F(x^{(t)}) = t \text{ then } \begin{cases} F_n(x) = d & \text{with } 0 < d \leq 1 \\ H_m(x) = c & \text{with } 0 < c \leq 1 \end{cases}, \text{ where,}$$

since $0 < t < 1$, at most one of d and c can be equal to 1. If any one of d and c is equal to 1 then we put $d = 1$ or $c = 1$ in $\Phi(y; c, d, \rho)$; or

$$(4.69)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t\leq F(x)} \frac{F_n(x) - H_m(x)}{F(x)} < y\right)$$

$$= \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t'\leq F(x)} \frac{F_n(x) - H_m(x)}{F(x)} < y\right)$$

$$= \Phi(y; c', d', \rho) , \text{ if } d = 0, c > 0$$

when $F(x^{(t)}) = t$ and where $F(x^{(t')}) = t'$, $0 < t' < 1$, is the smallest positive number such that $\begin{cases} d' = F_n(x) & \text{with } 0 < d' \leq 1 \\ c' = H_m(x) & \text{with } 0 < c' \leq 1 \end{cases}$,

where, since $0 < t' < 1$, at most one of d' and c' can be equal to 1. If any one of d' and c' is equal to 1 then we put $d' = 1$ or $c' = 1$ in $\Phi(y; c', d', \rho)$; or

$$(4.70)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t\leq F(x)} \frac{H_m(x) - F_n(x)}{F(x)} < y\right)$$

$$= \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t''\leq F(x)} \frac{H_m(x) - F_n(x)}{F(x)} < y\right)$$

$$= \Phi(y; c'', d'', \rho) , \text{ if } d > 0, c = 0 \text{ when } F(x^{(t)}) = t$$

and where $F(x^{(t'')}) = t''$, $0 < t'' < 1$, is the smallest positive number such that $\begin{cases} d'' = F_n(x) & \text{with } 0 < d'' \leq 1 \\ c'' = H_m(x) & \text{with } 0 < c'' \leq 1 \end{cases}$, where, since $0 < t'' < 1$, at most one of d'' and c'' can be equal to 1. If any one of d'' and c'' is equal to 1 then we put $d'' = 1$ or $c'' = 1$ in $\Phi(y;\, c'', d'', \rho)$.

We could have, of course, started the discussion of Theorem 13 with $\sup_{t \leq F(x)} (H_m(x) - F_n(x))/F(x)$ instead of $\sup_{t \leq F(x)} (F_n(x) - H_m(x))/F(x)$ and would have arrived at the same generalized statements of Theorem 16 as given above but (4.68) where we would have $F_n(x)$ and $H_m(x)$ in reverse to their present order which is irrelevant to the statement being made there anyway.

Proof of Theorem 14.

Repeating the argument of the proof of Theorem 13 it can be shown that it will be sufficient to derive the limiting distribution of

$$(4.71) \qquad \sup_{t \leq G_{n+m}(u) \leq 1} \frac{G_n(u) - K_m(u)}{u}, \quad 0 < t < 1 < 1,$$

in order to prove Theorem 14. If, in general, $0 < t \leq G_{n+m}(u) \leq 1 < 1$ then at least one of the empirical distribution functions $G_n(u)$ and $K_m(u)$ is greater than zero at u where $t = G_{n+m}(u)$ and, also, at least one of them is less than 1 at u where $G_{n+m}(u) = 1$. Let again, as before, $d = G_n(u)$ and $c = K_m(u)$ when $t = G_{n+m}(u)$ and let $e =$

$G_n(u)$, $f = K_m(u)$ when $G_{n+m}(u) = 1$. For the sake of Theorem 14 we make the assumption that both d and c are greater than zero and less than 1. As a result of this assumption both e and f are greater than zero. Otherwise we do not make any restriction on e and f, that is we do not exclude the possibility of having one of them equal to 1. Thus, examining the random variable of (4.71) for the set of u's for which we have $t \leq G_{n+m}(u) \leq 1$ also means the examination of this random variable for the set of u's for which $d \leq G_n(u) \leq e$ and $c \leq K_m(u) \leq f$ simultaneously. This we express by writing (4.71) in the following form

$$(4.72) \qquad \sup_{u \in U} \frac{G_n(u) - K_m(u)}{u} ,$$

where $U = \{u : t \leq G_{n+m}(u) \leq 1,\ d \leq G_n(u) \leq e,\ c \leq K_m(u) \leq f\}$

Analogously to (4.19), (4.72) can be written as

$$(4.73) \qquad \sup_{u \in U} \frac{G_n(u) - K_m(u)}{u} = \max_T \left\{ \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} , \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k+1)}} \right\} ,$$

where $T = \{k, i, j : t \leq k/(n+m) \leq 1,\ d \leq i/n \leq e,\ c \leq j/m \leq f\}$ and where, if $i/n - j/m > 0$, the expression $(i/n - j/m)/\eta_{(k)}$ is used and, if $i/n - j/m < 0$, the expression $(i/n - j/m)/\eta_{(k+1)}$ is used to find maximum. If $\frac{i}{n} - \frac{j}{m} = 0$, it is irrelevant which one of them is used. We cannot say here, as we did in case of (4.19), that the maximum of the above expression is at least zero, for we have here $t(n+m) \leq k \leq 1(n+m)$ and as a result of that $dn \leq i \leq en$, $cm \leq j \leq fm$ which implies that we may never have $\frac{i}{n} - \frac{j}{m} = 0$ and, as a consequence of this, if $\frac{i}{n} - \frac{j}{m} < 0$ when examining $\eta_{(k)}$, the ordered statistics of the pooled samples, for $t(n+m) \leq k \leq l(n+m)$ we need $(\frac{i}{n} - \frac{j}{m})/\eta_{(k+1)}$ to find indicated maximum in (4.73). An extended form of Lemma 1 will enable us, though, to derive the limiting distribution of right hand side of (4.73). In this connection we are going to prove

Lemma 2.

$$(4.74)\quad \max_{T}\left(\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}}\right) \geq \max_{T}\left\{\frac{\frac{i}{n}-\frac{j}{m}}{\eta_{(k)}},\ \frac{\frac{i}{n}-\frac{j}{m}}{\eta_{(k+1)}}\right\}$$

$$(4.75)\quad \max_{T}\left(\frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}}\right) \leq \max_{T}\left\{\frac{\frac{i}{n}-\frac{j}{m}}{\eta_{(k)}},\ \frac{\frac{i}{n}-\frac{j}{m}}{\eta_{(k+1)}}\right\}$$

where $\eta_{1(i)}$, $\eta_{1(i+1)}$, $\eta_{2(j)}$, $\eta_{2(j+1)}$ and $\eta_{(k)}$ are as defined in (4.8) and T is as defined in (4.73).

Proof of relation (4.74).

If $\frac{i}{n} - \frac{j}{m} > 0$, then to find maximum of right hand side of (4.74) we use $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ and thus the first two steps, (a) and (b), of the proof of (4.21) apply here too. So we take the case when $\frac{i}{n} - \frac{j}{m} < 0$, i.e. when $(\frac{i}{n} - \frac{j}{m})/\ _{(k+1)}$ of (4.74) is used to find maximum.

(c) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{1(i)}$ and let $\eta_{(k+1)} = \eta_{1(i+1)}$.

Then $\dfrac{\frac{i}{n}}{\eta_{1(i)}} - \dfrac{\frac{j}{m}}{\eta_{2(j+1)}} > \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i+1)}}$, for $\eta_{(k)} = \eta_{1(i)}$,

$\eta_{(k+1)} = \eta_{1(i+1)}$ imply that $\eta_{1(i)} < \eta_{1(i+1)} < \eta_{2(j+1)}$.

(c)' Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{1(i)}$ and let $\eta_{(k+1)} = \eta_{2(j+1)}$.

Then $\dfrac{\frac{i}{n}}{\eta_{1(i)}} - \dfrac{\frac{j}{m}}{\eta_{2(j+1)}} > \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j+1)}}$, for $\eta_{1(i)} < \eta_{2(j+1)}$.

(d) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{2(j)}$ and let $\eta_{(k+1)} = \eta_{1(i+1)}$.

Then $\dfrac{\frac{i}{n}}{\eta_{1(i)}} - \dfrac{\frac{j}{m}}{\eta_{2(j+1)}} > \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i+1)}}$, for $\eta_{(k)} = \eta_{2(j)}$,

$\eta_{(k+1)} = \eta_{1(i+1)}$ imply that $\eta_{1(i)} < \eta_{2(j)} < \eta_{1(i+1)} < \eta_{2(j+1)}$.

(d)' Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{2(j)}$ and let $\eta_{(k+1)} = \eta_{2(j+1)}$.

Then $\dfrac{\frac{i}{n}}{\eta_{1(i)}} - \dfrac{\frac{j}{m}}{\eta_{2(j+1)}} > \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j+1)}}$, for $\eta_{1(i)} < \eta_{2(j)} < \eta_{2(j+1)}$.

The next two steps, (e) and (f), of the proof of (4.21) apply here too. The quoted relations of the proof of (4.21) and (c), (c)', (d) and (d)' above prove (4.74).

Proof of relation (4.75).

If $\frac{i}{n} - \frac{j}{m} > 0$, then to find maximum of right hand side of (4.75) we use $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ and thus the first two steps, (a) and (b), of the proof of (4.22) apply here too. So we take the case when $\frac{i}{n} - \frac{j}{m} < 0$, that is when $(\frac{i}{n} - \frac{j}{m})/\eta_{(k+1)}$ of (4.75) is used to find maximum.

(c) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{1(i)}$ and let $\eta_{(k+1)} = \eta_{1(i+1)}$. Then $\dfrac{\frac{i}{n}}{\eta_{1(i+1)}} - \dfrac{\frac{j}{m}}{\eta_{2(j)}} < \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i+1)}}$, for $\eta_{(k)} = \eta_{1(i)}$ implies that $\eta_{2(j)} < \eta_{1(i)} < \eta_{1(i+1)}$.

(c)' Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{1(i)}$ and let $\eta_{(k+1)} = \eta_{2(j+1)}$. Then $\dfrac{\frac{i}{n}}{\eta_{1(i+1)}} - \dfrac{\frac{j}{m}}{\eta_{2(j)}} < \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j+1)}}$, for $\eta_{(k+1)} = \eta_{2(j+1)}$ implies that $\eta_{2(j+1)} < \eta_{1(i+1)}$.

(d) Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{2(j)}$ and let $\eta_{(k+1)} = \eta_{1(i+1)}$. Then $\dfrac{\frac{i}{n}}{\eta_{1(i+1)}} - \dfrac{\frac{j}{m}}{\eta_{2(j)}} < \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{1(i+1)}}$, for $\eta_{(k)} = \eta_{2(j)}$ and $\eta_{(k+1)} = \eta_{1(i+1)}$ imply that $\eta_{2(j)} < \eta_{1(i+1)}$.

(d)' Let $\frac{i}{n} - \frac{j}{m} < 0$ when $\eta_{(k)} = \eta_{2(j)}$ and let $\eta_{(k+1)} = \eta_{2(j+1)}$. Then $\dfrac{\frac{i}{n}}{\eta_{1(i+1)}} - \dfrac{\frac{j}{m}}{\eta_{2(j)}} < \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{2(j+1)}}$, for $\eta_{(k)} = \eta_{2(j)}$ and $\eta_{(k+1)} = \eta_{2(j+1)}$ imply that

$\eta_{2(j+1)} < \eta_{1(i+1)}$ and naturally $\eta_{2(j)} < \eta_{2(j+1)}$.

The next two steps, (e) and (f), of the proof of (4.22) apply here too. The quoted relations of the proof of (4.22) and (c), (c)', (d) and (d)' above prove (4.75).

Lemma 2 implies that the random variables which have been used to prove Theorem 13 can also be used to derive the limiting distribution of (4.71). Expressions, analogous to (4.23), (4.24), (4.25), (4.26), (4.27), (4.28) and (4.29) imply that the proof of Theorem 14 can be accomplished by examining the limiting distribution of the left hand side of (4.74) and (4.75) respectively. We consider first the left hand side of (4.74) that is we examine the asymptotic behaviour of

$$(4.76) \qquad \max_{T}\left(\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}}\right),$$

where, repeating the conditions of Theorem 14, we make the assumption that both d and c of T are greater than zero and less than 1. Because we have : $0 < t < l < 1$, at most one of the values e and f of T can be equal to 1. Under these conditions (4.76) can have three mutually exclusive forms which we indicate by writing it as

$$(4.77) \quad \max_{T}\left(\frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}}\right) = \begin{cases} \max\limits_{T} \dfrac{\frac{i}{n}}{\eta_{1(i)}} - \dfrac{\frac{j}{m}}{\eta_{2(j+1)}}, & \text{if } e,f<1 \ \ldots \quad (1) \\[2ex] \max\limits_{T'} \dfrac{\frac{i}{n}}{\eta_{1(i)}} - \dfrac{\frac{j}{m}}{\eta_{2(j+1)}}, & \text{if } e=1, f<1 \ldots \quad (2) \end{cases}$$

$$\left| \max_{T''} \frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right. \text{, if } e<1\text{, } f=1 \ldots (3)$$

where T is as defined in (4.73) and where

$$T' = \left\{k,i,j : t \leq \frac{k}{n+m} \leq 1,\ d \leq \frac{i}{n} \leq 1,\ c \leq \frac{j}{m} \leq f\right\},$$

$$T'' = \left\{k,i,j : t \leq \frac{k}{n+m} \leq 1,\ d \leq \frac{i}{n} \leq e,\ c \leq \frac{j}{m} \leq 1\right\}.$$

We examine first case (1), where both e and f are less than 1. Because of the assumption that both d and c are greater than zero and less than 1 it is assured that both e and f are greater than zero. The limiting distribution of the random variable of (4.77) in this case is identical with that of the random variable

$$(4.78) \quad \max_{T} \left(\log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right), \quad \eta_{2(m+1)} = 1 .$$

This, analogously to (4.36), can be written as

$$(4.79) \quad \gamma = \max_{T}\left\{\left(\log \frac{1}{\eta_{1(i)}} - \sum_{v=i}^{n} \frac{1}{v}\right) - \left(\log \frac{1}{\eta_{2(j+1)}} - \sum_{s=j+1}^{m} \frac{1}{s}\right)\right\}$$

$$= \max_{T} \left(\sum_{v=i}^{n} \frac{\delta_{n+1-v}-1}{v} - \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}-1}{s}\right)$$

which, in turn may be written as the sum of two independent random variables α and β where

$$(4.80) \quad \alpha = \sum_{en\leq i\leq n} \frac{\delta_{n+1-i}-1}{i} - \sum_{fm\leq j\leq m-1} \frac{\delta_{m-j}-1}{j+1}$$

and

$$(4.81)\quad \beta = \max_{T} \sum_i \frac{\delta_{n+1-i}^{\;-1}}{i} - \sum_j \frac{\delta_{m-j}^{\;-1}}{j+1} \quad .$$

It is clear from (4.39), (4.41) and (4.41)' that, as $n, m \to \infty$ so that $\frac{m}{n} \to \rho$, the standard deviation of $\sqrt{\frac{nm}{n+m}}\,\alpha$ is given by

$$(4.82)\qquad \sqrt{\frac{f(1-e)\rho + e(1-f)}{ef(1+\rho)}}$$

and, by the Lindeberg form of the central limit theorem in the limit, as $n, m \to \infty$ so that $\frac{m}{n} \to \rho$, $\sqrt{\frac{nm}{n+m}}\,\alpha$ is a normally distributed random variable with standard deviation as given in (4.82).

Considering (4.81), the variance of $\sum_i \frac{\delta_{n+1-i}}{i}$, if $dn \leq i \leq en$, $0 < d < e < 1$ and using Euler's summation formula, is given by

$$(4.83)\quad D^2\left(\sum_{dn \leq i \leq en} \frac{\delta_{n+1-i}}{i}\right) = \sum_{dn \leq i \leq en} \frac{1}{i^2} = \frac{1}{dn} - \frac{1}{en} + o\left(\frac{1}{n}\right)$$

$$= \frac{e-d}{edn} + o\left(\frac{1}{n}\right)$$

and, similarly, the variance of $\sum_j \frac{\delta_{m-j}}{j+1}$, if $cm \leq j \leq fm$, $0 < c < f < 1$, is given by

$$(4.84)\quad D^2\left(\sum_{cm \leq j \leq fm} \frac{\delta_{m-j}}{j+1}\right) = \frac{f-c}{cfm} + o\left(\frac{1}{m}\right) ,$$

which, mutatis mutandis, can be seen from calculations of (3.44). Thus the standard deviation of

$$\sqrt{\frac{nm}{n+m}}\left(\sum_{dn\le i\le en}\frac{\delta_{n+1-i}^{-1}}{i}-\sum_{cm\le j\le fm}\frac{\delta_{m-j}^{-1}}{j+1}\right),$$

as $n, m \to \infty$ so that $\frac{m}{n}\to\rho$, is given by

$$\sqrt{\frac{(e-d)cf\rho+(f-c)ed}{edcf(1+\rho)}}\,, \tag{4.85}$$

on using (4.83) and (4.84). From the proof of Theorem 13 it can be seen that, when applying Theorem 5 of chapter 2 with $\sqrt{\frac{nm}{n+m}}\,\beta$ as $\max S_{ne+mf,k}$ of this theorem, we have

$$\lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}}\,\beta < y\right)=\sqrt{\frac{2}{\pi}}\int_0^{y\sqrt{\frac{edcf(1+\rho)}{(e-d)cf\rho+(f-c)ed}}} e^{-\frac{v^2}{2}}\,dv\,, \quad \text{if} \quad y>0, \tag{4.86}$$

zero otherwise.

Considering further that α and β are independent it follows from (4.82), (4.86) and convolution that

$$\lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}}\,\gamma < y\right)=\frac{1}{\pi}\sqrt{\frac{f(1-e)\rho+e(1-f)}{ef(1+\rho)}}\int_{-\infty}^{y} e^{-\frac{ef(1+\rho)u^2}{2[f(1-e)\rho+e(1-f)]}}\int_0^{(y-u)\sqrt{\frac{edcf(1+\rho)}{(e-d)cf\rho+(f-c)ed}}} e^{-\frac{v^2}{2}}\,dv\,du \tag{4.87}$$

where we have : $-\infty < y < +\infty$, and this, by (4.78), is the limiting distribution of left hand side of (4.74) for case (1) of (4.77).

The limiting distribution of left hand side of (4.75) for case (1) of (4.77) can be derived in exactly the same way with the same result. This completes the proof of Theorem 14 in the case when both e and f are less than 1.

Concerning case (2) of (4.77), where $e = 1$, $f < 1$, it can easily be seen from (4.52) that, as $n, m \to \infty$ so that $\frac{m}{n} \to \rho$, $\sqrt{\frac{nm}{n+m}}\,\alpha$ is a normally distributed random variable with standard deviation

$$\sqrt{\frac{(1-f)}{f(1+\rho)}}\,, \tag{4.88}$$

and the limiting distribution of $\sqrt{\frac{nm}{n+m}}\,\beta$ has standard deviation

$$\sqrt{\frac{(1-d)cf\rho + (f-c)d}{dcf(1+\rho)}}\,, \tag{4.89}$$

also easily seen from (4.83) and (4.84).

Similarly, in case (3) of (4.77), where $e < 1$, $f = 1$, as $n, m \to \infty$ so that $\frac{m}{n} \to \rho$, $\sqrt{\frac{nm}{n+m}}\,\alpha$ is a normally distributed random variable with standard deviation

$$\sqrt{\frac{(1-e)\rho}{e(1+\rho)}}\,, \tag{4.90}$$

while the limiting distribution of $\sqrt{\frac{nm}{n+m}}\,\beta$ has standard deviation

$$\sqrt{\frac{(e-d)c\rho + (1-c)ed}{edc(1+\rho)}} \tag{4.91}$$

Therefore, (4.88), (4.89), (4.90) and (4.91) imply that in (4.87) we have proved Theorem 14 with the understanding that we put $e = 1$ when $e = 1$, $f < 1$ and $f = 1$ when $e < 1$, $f = 1$ in the right hand side expression of (4.87), that is in $N(y;\, d, c, e, f, \rho)$ of the statement of Theorem 14. Also, if $n = m$ we put $\rho = 1$ in (4.87) and then we have a special form of Theorem 14.

Just as in the case of Theorem 13, we could examine here too the possibility of having one of d and c equal to zero or having one of them equal to one and the possible combinations of these two cases. To any one of these possibilities for the values of d and c there would belong some possible combinations for the values of e and f. Because of these numerous possible combinations, difficulties arise which are similar to those of cases (ii) and (iii) of p. 66 and which cannot be similarly overcome when trying to examine their limiting behaviour. I have therefore not attempted the problem of generalizing Theorem 14 in this direction.

5. Remarks and some results concerning the problems of deriving the limiting distribution of the supremum of $|F_n(x) - H_m(x)|/F(x)$.

Repeating the argument leading to (4.19), mutatis mutandis, we get

$$(5.1)\quad \sup_{u \in I} \left| \frac{G_n(u) - K_m(u)}{u} \right| = \max_S \left\{ \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right|, \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k+1)}} \right| \right\}$$

But $\max\limits_S \left| \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right| > \max\limits_S \left| \dfrac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k+1)}} \right|$ and, therefore, we get from (5.1) that

$$(5.2)\quad \sup_{u \in I} \left| \frac{G_n(u) - K_m(u)}{u} \right| = \max_S \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right| .$$

From the statement of Lemma 1 of chapter 4 it is clear that

$$(5.3)\quad \max_S \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right| \leq \max_S \left\{ \left| \frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right|, \left| \frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} \right| \right\}$$

Here we have the problem of deciding which one of the random variables of the right hand side of (5.3) is going to be maximum and there seems to be no way of doing this. If we were able to derive the limiting distribution of the right hand side of (5.3) then, through (5.3), we could have a statement regarding $|F_n(x) - H_m(x)|/F(x)$ as follows

$$(5.4)\quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x)} \frac{|F_n(x) - H_m(x)|}{F(x)} < y \right) \geq$$

$$\lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}}\; R_{n,m} < y\right)$$

where $R_{n,m}$ stands for the right hand side of (5.3). Then we still would have the problem of finding an upper bound for

$\overline{\lim}_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x)} \frac{|F_n(x) - H_m(x)|}{F(x)} < y\right)$ and if this upper bound would turn out to be equal to right hand side of (5.4) we would have a new limiting distribution at our disposal. The derivation of this upper bound in question would require a random variable which would be less than or equal to $(\frac{i}{n} - \frac{j}{m})/\eta_{(k)}$ when positive and greater than or equal to it when negative. Thus its absolute value would be less than or equal to $\left|\frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}}\right|$, and if we wanted to use A. Rényi's method of proof we would have to have it in a form adaptable to this method. I have not succeeded in finding such a random variable and have settled on trying to derive statements like (5.4). Such statements would be of some interest as "at least" probability statements. In this connection we are going to prove the following theorems :

Theorem 17. If $F(x) = H(x)$ and $n,m \to \infty$ so that $\frac{m}{n} \to \rho$ then

$$(5.5) \quad \underline{\lim}_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x)} \frac{|F_n(x) - H_m(x)|}{F(x)} < y\right)$$

$$\geq \begin{cases} \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}\, e^{-\frac{(2k+1)^2\pi^2}{8}\frac{c(1-d)\rho + d(1-c)}{y^2 cd(1+\rho)}}, & \text{if } y > 0, \\ 0, & \text{if } y \leq 0 \end{cases}$$

$$= L\left(y\sqrt{\frac{cd(1+\rho)}{c(1-d)\rho + d(1-c)}}\right),$$

for all values of t, $0 < t < 1$, so that when

$F(x^{(t)}) = t$ then $\begin{cases} F_n(x) = d & \text{with } 0 < d \leq 1 \\ H_m(x) = c & \text{with } 0 < c \leq 1 \end{cases}$. If any one of d and c is equal to 1, we put $d = 1$ or $c = 1$ in $L\left(y\sqrt{\dfrac{cd(1+\rho)}{c(1-d)\rho + d(1-c)}}\right)$. Since we have $0 < t < 1$, at most one of the values d and c can be equal to 1.

Theorem 18. If $F(x) = H(x)$ and $n, m \to \infty$ so that $\frac{m}{n} \to \rho$ then

$$(5.6) \quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x) \leq 1} \frac{|F_n(x) - H_m(x)|}{F(x)} < y\right)$$

$$\geq \begin{cases} \dfrac{4}{\pi} \displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} e^{-\frac{(2k+1)^2\pi^2}{8} \frac{c(1-d)\rho + d(1-c)}{y^2 cd(1+\rho)}} \cdot E_k, & \text{if } y > 0, \\ 0, & \text{if } y \leq 0 \end{cases}$$

where $E_k = 1 - \dfrac{2}{\sqrt{2\pi}} \displaystyle\int_{y\sqrt{\frac{ef(1+\rho)}{f(1-e)\rho + e(1-f)}}}^{\infty} e^{-\frac{v^2}{2}} dv + \rho_k$

and $\rho_k = \left(\sqrt{2\pi} \sqrt{\dfrac{cd(1+\rho)}{c(1-d)\rho + d(1-c)}}\, y\right)^{-1} 2e^{\frac{y^2 ef(1+\rho)}{2[f(1-e)\rho + e(1-f)]}} \displaystyle\int_0^{(2k+1)\pi/2} e^{\frac{[f(1-e)\rho + e(1-f)]v^2}{2y^2 ef(1+\rho)}} \sin v \, dv$

$$= R(y; d, c, e, f, \rho),$$

for all t and l with $0 < t < l < 1$ where the other parameters satisfy the following conditions :

when $F(x^{(t)}) = t$ then $\begin{cases} F_n(x) = d & \text{with } 0 < d < 1 \\ H_m(x) = c & \text{with } 0 < c < 1 \end{cases}$

and when $F(x^{(l)}) = l$ then $\begin{cases} F_n(x) = e \\ H_m(x) = f \end{cases}$ where, l being less than one, at most one of e and f can be equal to 1. If one of e and f is equal to 1, the appropriate one is replaced by 1 in $R(y;\, d,\, c,\, e,\, f,\, \rho)$.

These theorems provide tests for verifying the hypothesis that 2 random samples of size n and m respectively have been drawn from a population having continuous c.d.f. $F(x)$. The character of these tests consists in that they give a band in which, if the hypothesis is true, $\sup |F_n(x) - H_m(x)|$ has to lie with <u>at least</u> that much probability as given in (5.5) and (5.6) and the width of this band is proportional at all its points x to $F(x)$. It is quite likely that in (5.5) we have

$$\lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup_{t \leq F(x)} \frac{|F_n(x) - H_m(x)|}{F(x)} < y\right) = L\left(y\sqrt{\frac{cd(1+\rho)}{c(1-d)\rho+d(1-c)}}\right)$$

and in (5.6) we have $\lim\limits_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \sup\limits_{t \leq F(x) \leq 1} \frac{|F_n(x) - H_m(x)|}{F(x)} < y\right)$

$= R(y;\, d,\, c,\, e,\, f,\, \rho)$. As it was already mentioned above, one would have to be able to show that $\limsup\limits_{(m,n;\rho)} P(\,.\,)$ statements of (5.5) and (5.6) are bounded above by $L(\,.\,)$ and $R(\,.\,;\,\ldots)$ respectively in order to prove these theorems in such forms.

Proof of Theorem 17.

To prove Theorem 17 we will have to derive the limiting distribution of right hand side of (5.3) and for that we would need the maximum of its two random variables. As we have already remarked after (5.3), there seems to be no way of choosing this desired maximum random variable. It so happens though that the limiting distribution of the two random variables in question is the same. This enables us to say that it is sufficient to examine the asymptotic behaviour of any one of them for if the one we pick would not happen to be the maximum one we would have to choose the other one and it would provide us with the same limiting c.d.f., Keeping this argument in mind let us assume that in (5.3) the following occurs :

$$(5.7) \qquad \max_S \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right| \leq \max_S \left| \frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| , \quad \eta_{2(m+1)} = 1 .$$

Let A be the event $\sqrt{\frac{nm}{n+m}} \max_S \left| \frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| < y$,

and B be the event $\sqrt{\frac{nm}{n+m}} \max_S \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right| < y$.

It follows then from (5.7) that $A \subseteq B$ and, therefore, we have $P(A) \leq P(B)$, which immediately implies Theorem 17 if the limiting distribution of the event A is given by $L\left(y \sqrt{\frac{cd(1+\rho)}{c(1-d)\rho+d(1-c)}}\right)$ of (5.5)

Let us consider then the right hand side of (5.7) where, for the time being, we assume that both d and c of S are less than 1, where S is as it was defined in (4.19). Now the limiting distribution of the right hand side of (5.7) is identical with that of the random variable

$$(5.8)\qquad \max_{S} \left| \log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| ,$$

which, in turn, has the same limiting distribution as

$$(5.9)\qquad \max_{S} \left| \sum_{v=i}^{n} \frac{\delta_{n+1-v}-1}{v} - \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}-1}{s} \right| ,$$

as can be seen from (4.31), (4.32), (4.33), (4.34), (4.35) and (4.36). Applying Theorem 6 of chapter 2 with (5.9) as $\max |S_{n+m,k}|$ of this theorem and using the results of (4.39), (4.41) and (4.41)' we get

$$(5.10)\qquad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \max_{S} \left| \log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| < y\right)$$

$$= L\left(y \sqrt{\frac{cd(1+\rho)}{c(1-d)\rho + d(1-c)}}\right)$$

where $L(.)$ is as it was defined in (5.5).

It is clear from (4.48), (4.50), (4.51), (4.52), (4.53) and (4.56) that the statement of (5.10) remains valid when one of d and c is equal to 1. This, with the remark that the limiting distribution of the second random variable of the right hand side of (5.3) is also given by (5.10), completes the proof of Theorem 17.

An attempt to generalize Theorem 17 on the lines of Theorem 16 of chapter 4, that is when we would want to allow d or c to be equal to zero, fails because of the absolute sign of $|F_n(x) - H_m(x)|$ of this theorem.

Proof of Theorem 18.

From (4.73) it is clear that we have

$$(5.11)\qquad \sup_{u\in U} \frac{G_n(u) - K_m(u)}{u} = \max_T \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right|$$

where U and T are as defined in (4.72) and (4.73) and where, for the time being, we assume that both e and f of U and T are less than 1. Now by Lemma 2 of chapter 4 we have

$$(5.12)\qquad \max_T \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \leq \max_T \left\{ \left| \frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| , \left| \frac{\frac{i}{n}}{\eta_{1(i+1)}} - \frac{\frac{j}{m}}{\eta_{2(j)}} \right| \right\}$$

where again we cannot though decide which one of the two random variables is going to give us maximum but here too we can say that both of them have the same limiting distribution. As a matter of fact, both of them have the asymptotic distribution as given in (5.6) by $R(y; d, c, e, f, \rho)$. Repeating the argument of (5.7) let us assume that in (5.12) the following happens :

$$(5.13)\qquad \max_T \left| \frac{\frac{i}{n} - \frac{j}{m}}{\eta_{(k)}} \right| \leq \max_T \left| \frac{\frac{i}{n}}{\eta_{1(i)}} - \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| , \quad \eta_{2(m+1)} = 1 .$$

From the argument immediately after (5.7), mutatis mutandis, it follows that in order to prove Theorem 18 it is sufficient to prove that the limiting distribution of the right hand side of (5.13) is given by $R(y; d, c, e, f, \rho)$ of (5.6).

Now the limiting distribution of the right hand side of (5.13) is identical with that of the random variable

$$(5.14) \qquad \max_{T} \left| \log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| ,$$

which, in turn, has the same limiting distribution as

$$(5.15) \qquad \max_{T} \left| \sum_{v=i}^{n} \frac{\delta_{n+1-v}^{-1}}{v} - \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}^{-1}}{s} \right|$$

and this is easily seen from (4.79).

Let us define

$$(5.16) \qquad S_{n+m;n+m+1-k:n+1-i,m-j} = \sum_{v=i}^{n} \frac{\delta_{n+1-v}^{-1}}{v} - \sum_{s=j+1}^{m} \frac{\delta_{m+1-s}^{-1}}{s}$$

$$(5.17) \qquad B_{n+m}^{2} = D^{2}\left(\sum_{dn\leq i\leq n} \frac{\delta_{n+1-i}^{-1}}{i} - \sum_{cm\leq j\leq m-1} \frac{\delta_{m-j}^{-1}}{j+1} \right)$$

$$= \sum_{dn\leq i\leq n} \frac{1}{i^2} + \sum_{cm\leq j\leq m-1} \frac{1}{(j+1)^2}$$

$$(5.18) \qquad A_{n+m}^{2} = D^{2}\left(\sum_{en\leq i\leq n} \frac{\delta_{n+1-i}^{-1}}{i} - \sum_{fm\leq j\leq m-1} \frac{\delta_{m-j}^{-1}}{j+1} \right)$$

$$= \sum_{en\leq i\leq n} \frac{1}{i^2} + \sum_{fm\leq j\leq m-1} \frac{1}{(j+1)^2}$$

By (4.39), (4.41) and (4.41)' we have

$$A_{n+m} = \sqrt{\frac{1-e}{en} + \frac{1-f}{fm}} + o\left(\frac{1}{n}\right) + o\left(\frac{1}{m}\right)$$

$$B_{n+m} = \sqrt{\frac{1-d}{dn} + \frac{1-c}{cm}} + o\left(\frac{1}{n}\right) + o\left(\frac{1}{m}\right) ,$$

and so

$$(5.19) \quad \lim_{(m,n;\rho)} \frac{A_{n+m}}{B_{n+m}} = \sqrt{\frac{dcf(1-e)\rho + dce(1-f)}{efc(1-d)\rho + efd(1-c)}} = \lambda$$

as $m, n \to \infty$ so that $\frac{m}{n} \to \rho$. Thus we can apply Theorem 8 of chapter 2 with (5.16), (5.17), (5.18), (5.19) and with $1 \leq M_n = n+1-en < N_n = n+1-dn$ and $1 \leq M_m = m+1-fm < N_m = m+1-cm$. Therefore we have, by (5.14) - (5.19), that

$$(5.20) \quad \lim_{(m,n;\rho)} P\left(\max_T \left| \log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| < z\sqrt{\frac{n+m}{nm}} \sqrt{\frac{c(1-d)\frac{m}{n+m} + d(1-c)\frac{n}{n+m}}{dc}} \right)$$

being equal to the statement of Theorem 8 of chapter 2. Letting

$$y = z\sqrt{\frac{c(1-d)\frac{m}{n+m} + d(1-c)\frac{n}{n+m}}{dc}}$$ and $n, m \to \infty$ so that $\frac{m}{n} \to \rho$ we get

$$(5.21) \quad \lim_{(m,n;\rho)} P\left(\sqrt{\frac{nm}{n+m}} \max_T \left| \log \frac{\frac{i}{n}}{\eta_{1(i)}} - \log \frac{\frac{j}{m}}{\eta_{2(j+1)}} \right| < y \right)$$

$$= R(y; d, c, e, f, \rho),$$

where $R(y; d, c, e, f, \rho)$ is as it was defined in (5.6)

It is clear from (4.48), (4.50), (4.51), (4.52) and their application, mutatis mutandis, to (5.16), (5.17), (5.18) and (5.19) that the statement of (5.21) remains valid when one of e and f is equal to 1. This, with the remark that the limiting distribution of the second random variable of the right hand side of (5.12) is also given by (5.21), completes the proof of Theorem 17.

Bibliography.

[1] Anderson, T.W. - Darling, D.A., Asymptotic theory of certain "goodness of fit" criteria on stochastic processes, Ann. Math. Stat., Vol. 23 (1952), pp. 193 - 212.

[2] Donsker, M.D., Justification and extension of Doob's heuristic approach to the Kolmogorov - Smirnov theorems, Ann. Math. Stat., Vol 23 (1952), pp. 277 - 281.

[3] Doob, J. L., Stochastic Processes, John Wiley, New York (1953).

[4] , Heuristic approach to the Kolmogorov - Smirnov theorems, Ann. Math. Stat., Vol 20 (1949), pp. 393 - 403.

[5] Erdős, P. - Kac, M., On certain limit theorems of the theory of probability, Bull. Amer. Math. Soc., Vol 52 (1946), pp. 292 - 302.

[6] Feller, W., An Introduction to Probability Theory and its Applications, Vol 1, second edition, John Wiley, New York (1957).

[7] , On the Kolmogorov - Smirnov limit theorems for empirical distributions, Ann. Math. Stat., Vol 19 (1948), pp. 177 - 189.

[8] Gihman, I.I., Ob empiričeskoi funkcii raspredelenija slučaje gruppirovki dannych, Dokl. Akad. Nauk SSSR,

Vol. 82 (1952), pp. 837 - 840.

[9] Glivenko, V.I., Sulla determinazione empirica delle leggi di probabilita, Giorn. Ist. Ital. Attuari, Vol 4 (1933), pp. 92 - 99.

[10] Gnedenko, B.V. - Korolyuk, V.S., On the maximum discrepancy between two empirical distributions, Dokl. Akad. Nauk S.S.S.R., Vol 4 (1951), pp. 525 - 528.

[11] Gnedenko, B.V. - Kolmogorov, A.N., Limit Distributions for Sums of Random Variables (translated from the 1949 Russian edition by K.L. Chung; with an appendix by J.L. Doob), Addison - Wesley Publishing Company, Reading, Mass. (1954).

[12] Hajós, G. - Rényi, A., Elementary proofs of some basic facts concerning order statistics, Acta Math. Acad. Sci. Hung., Vol 5 (1954), pp. 1 - 6.

[13] Halmos, P.R., Measure Theory, D. Van Nostrand, Princeton, N.J. (1950).

[14] Kaplansky, I. - Hewitt, E. - Hall, M., Jr. - Fortet, R., Some Aspects of Analysis and Probability, John Wiley, New York (1958).

[15] Kolmogorov, A.N., Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergeb. Math. No. 3, Berlin (1933) (English translation by N. Morrison, Chelsea, New York (1950)).

[16] , Sulla determinazione empirico di una legge di distribuzione, Giornale dell' Instituto Italiano d. Attuari, Vol 4 (1933), pp. 83 - 91.

[17] , Confidence limits for an unknown distribution function, Ann. Math. Stat., Vol 12 (1941), pp. 461 - 463.

[18] Loève, M., Probability Theory, D. Van Nostrand, Princeton, N.J. (1960).

[19] Malmquist, S., On a property of order statistics from a rectangular distribution, Skand. Aktuerietidsskrift, Vol. 33 (1950), pp. 214 - 222.

[20] Massey, F.J., A note on the estimation of a distribution function by confidence limits, Ann. Math. Stat., Vol. 21 (1950), pp. 116 - 119.

[21] , Distribution table for the deviation between two sample cumulatives, Ann. Math. Stat., Vol 23 (1952), pp. 435 - 441.

[22] Rényi, A., Valószínűségszámítás (Probability Theory, in Hungarian), Tankönyvkiadó, Budapest (1954).

[23] , On the theory of order statistics, Acta Math. Acad. Sci. Hung., Vol 4 (1953) pp. 190 - 231.

[24] Smirnov, N.V., Ob uklonenijah empiričeskoi krivoi raspredelenija, Recueil Mathematique (Matematičeskii Sbornik), N.S. Vol 6 (1939), pp. 3 - 26.

[25] , On the estimation of the discrepancy between empirical curves of distribution for two independent samples, Bulletin Mathématique de l'Université de Moscow, Vol. 2 (1939), No. 2, 3 - 14.

[26] Wald, A., Limit distribution of the maximum and minimum of successive cumulative sums of random variables, Bull. Amer. Math. Soc., Vol. 53 (1947), pp. 142 - 153.

[27] , On the distribution of the maximum of successive cumulative sums of independently, but not identically distributed chance variables, Bull. Amer. Math. Soc., Vol. 54 (1948), pp. 422 - 430.

[28] Wilks, S.S., Order statistics, Bull. Amer. Math. Soc., Vol. 54 (1948), pp. 6 - 50.

[29] , Mathematical Statistics, John Wiley, New York (1962).