# ASYMPTOTIC THEORY OF SOME NONPARAMETRIC TESTS

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

Shashikala B. Sukhatme

1960

This is to certify that the

thesis entitled

ASYMPTOTIC THEORY OF SOME NONPARAMETRIC TESTS

presented by

SHASHIKALA B. SUKHATME

has been accepted towards fulfillment
of the requirements for

PH. D. degree in STATISTICS

Charles H Kraft
Major professor

Date August 3, 1960

O-169

THE

# ASYMPTOTIC THEORY OF SOME NONPARAMETRIC TESTS

By

Shashikala B. Sukhatme

A THESIS

Submitted to the School for Advanced Graduate
Studies of Michigan State University in
partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics

1960

## ACKNOWLEDGMENTS

It is with pleasure that I express my sincere gratitude to Professor Gopinath Kallianpur for suggesting the problem treated in Part II, for his stimulating advice and guidance throughout the entire work. My sincere thanks are also due to Professor Ingram Olkin for suggesting the problem treated in Part I and for his keen interest and encouragement during the progress of the work. Thanks are also due to Mrs. Barbara A. Johnson for typing the manuscript.

The author appreciates the financial support of the Office of Ordnance Research.

THE

ASYMPTOTIC THEORY OF SOME

NONPARAMETRIC TESTS

By

Shashikala B. Sukhatme

AN ABSTRACT

Submitted to the School for Advanced Graduate
Studies of Michigan State University in
partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics

Year 1960

APPROVED Charles H Kraft

SHASHIKALA B. SUKHATME ABSTRACT

This thesis consists of two parts in which two different problems are treated. Part I deals with some nonparametric tests for location and scale parameters in a mixed model of discrete and continuous variables. In Part II we consider asymptotic theory of modified Cramér-Smirnov test statistics in parametric case.

The following problem is studied in Sections 1 - 6 which constitute Part I. Let $Z_1, \ldots, Z_N$ with $Z_i = (X_i, Y_i)$ be independent observations from a bivariable population. Assume that the random variable $X$ takes only two values 1 and 0 with probabilities $p$ and $1 - p$ respectively. Let $P(Y \leq y | X = j) = F_j(y)$, $j = 0, 1$. We consider the problem of testing the hypothesis $H$: $F_1 = F_0$ against the alternative $A$: $F_1 \neq F_0$ where $F_1$ and $F_0$ are assumed to have the same functional form except that they differ either in the location or the scale parameter. Two sample median test and Wilcoxon test have been considered for testing the differences in location while two sample rank test and run test are studied for the differences in scale. The problem has also been generalized to the case when the random variable $X$ has a multinomial distribution. In the case when $p$ is unknown, the test statistics are modified by replacing $p$ by its usual estimator and we investigated whether the tests based on the modified statistics are asymptotically distribution-free.

In Part II consisting of Sections 7 - 10, we consider the following problem. Let $X_1, \ldots, X_n$ be $n$ independent observations with a continuous distribution function $G(x)$ . For testing the hypothesis H: $G(x) = F(x, \theta)$ where the functional form of $F$ is known, but the value of $\theta \in \mathcal{I}$, an open interval in $R^1$ is unknown, Darling modified Cramér-Smirnov $\omega_n^2$ test by replacing $\theta$ by its estimate $\hat{\theta}_n$ obtained from the sample. He obtained the asymptotic distribution of the modified test statistic under the hypothesis and studied its properties. In this part we extend Darling's results to the case when $\theta = (\theta_1, \theta_2)$ is a point belonging to an open interval in $R^2$ . We obtain the asymptotic distribution of the modified Cramér-Smirnov test statistic under the hypothesis. The limiting distribution is found to depend on the properties of the estimators of $(\theta_1, \theta_2)$ . Two different cases are considered according as the estimators are superefficient or regular, jointly efficient in the sense defined by Cramér. As the characteristic function of the limiting distribution is the Fredholm determinant of a symmetric, bounded, positive definite kernel of a particular form, methods of finding the Fredholm determinants of such kernels are given. Lastly we study the k-sample Cramér-Smirnov test in parametric case for testing the hypothesis of goodness of fit and investigate its asymptotic distribution under the hypothesis.

## TABLE OF CONTENTS

Part I

# NONPARAMETRIC TESTS FOR LOCATION AND SCALE PARAMETERS IN A MIXED MODEL OF DISCRETE AND CONTINUOUS VARIABLES

## 1. Introduction

Let $Z_1, \ldots, Z_n$ where $Z_i = (X_i, Y_i)$ be $N$ independent observations from a bivariate distribution. We assume that the random variable $X$ takes only two values, 1 and 0 with $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$. Let the conditional distribution of $Y$ given $X = j$ $(j = 0, 1)$, be $P(Y \leq y \mid X = j) = F_j(y)$. The problem considered is that of testing the hypothesis $H$: $F_1 = F_0$ against the alternative $A$: $F_1 \neq F_0$.

We divide the observations $Z_1, \ldots, Z_N$ into two groups according as the observed value of $X$ is 1 or 0. Let $U_1, U_2, \ldots, U_n$, $(n > 0)$ and $V_1, \ldots, V_{N-n}$ denote those values of $Y$ for which the corresponding $X$ is observed to be 1 and 0, respectively. Since $U_1, \ldots, U_n$ and $V_1, V_2, \ldots, V_{N-n}$ are independent, the problem of testing the hypothesis $H$ is equivalent to the problem of testing the hypothesis that the two independent samples come from the same population. However, the problem differs from the usual two sample problems in that, the number of observations in each of the two samples is a random variable.

In what follows, we assume that $F_1$ and $F_0$ are absolutely continuous, having density functions $F_1'$ and $F_0'$, respectively. We further assume that $F_1$ and $F_0$ have the same functional form except that they differ either in the location or in the scale parameter.

Several two sample nonparametric tests have been proposed for testing the differences in location, especially those by Wilcoxon [1], Mood [2], Wald and Wolfowitz [3], and Lehmann [4]. More recently some nonparametric tests have been proposed for testing differences in dispersion by Mood [5], Sukhatme [6, 7], and Kamat [8].

In Section 2 we consider the median test and in Section 3, the two sample Wilcoxon test, with reference to the problem considered here. In Section 4 we generalize the median test to a c-sample problem. In Section 5 we consider Mood's rank test for testing differences in dispersion, and Section 6 is devoted to the run test. For convenience of exposition, the cases where $p$ is known or unknown are treated separately. In the former, both the exact and asymptotic properties are investigated. In the latter case, the various test statistics are modified by replacing $p$ by its usual estimator $\hat{p}$, and we investigate whether the test based on the modified statistics is asymptotically distribution-free.

## 2. Two Sample Median Test

Without any loss of generality we assume that the sample size $N$ is odd, say $N = 2k + 1$. Let $\tilde{W}$ denote the median of the combined sample of $U_i$'s and $V_j$'s, and let $m$ be the number of $U_i$'s which are less than $\tilde{W}$. The hypothesis $H$: $F_1 = F_0$ is rejected, if $m$ is either too large or too small. First consider the case when the distribution of $X$ is known, i.e. when $p$ is known. In Section 2.1, the exact distribution of $m$ is derived and in Section 2.2, its limiting distribution both under the hypothesis and the alternative is obtained. The consistency of the test is proved in Section 2.3, and its asymptotic efficiency with respect to the corresponding parametric test based on the correlation coefficient is determined in Section 2.4. The case when $p$ is not known is dealt with in Section 2.5, where it is shown that the test based on the statistic $(m - k\hat{p}) / \sqrt{k\hat{p}\hat{q}}$ is asymptotically distribution-free.

## 2.1 Joint and Marginal Distributions of $m$ and $\tilde{W}$.

Henceforth $f(\cdot)$ denotes the probability density function of the random variables written in the parentheses. We first prove the following lemma which gives joint distribution of $m$ and $\tilde{W}$.

Lemma 2.1.1. The joint distribution of $m$ and $\tilde{W}$ is

$$(2.1.1) \qquad f(m, \tilde{w}) = \frac{(2k+1)! \, [p\, F_1(\tilde{w})]^m [qF_0(\tilde{w})]^{k-m}}{m! \, (k-m)! \, k!}$$

$$[1-p\, F_1(\tilde{w}) - q\, F_0(\tilde{w})]^k \, [p\, F_1'(\tilde{w}) + q\, F_0'(\tilde{w})] \, ,$$

$$m = 0, 1, 2, \ldots, k \, .$$

Proof. Observing that $n$ is a binomial random variable $b(N, p)$, we have

$$(2.1.2) \qquad f(n, m, \tilde{w}) = f(m, \tilde{w}|n) \, f(n) \, ,$$

where

$$(2.1.3) \qquad f(n) = \binom{N}{n} p^n \, q^{N-n}$$

and from Mood [2],

(2.1.4) $f(m, \tilde{w}|n) =$

$$= \frac{n!\,[F_1(\tilde{w})]^m\,[1-F_1(\tilde{w})]^{n-m-1}}{m!\,(n-m-1)!} \cdot \frac{(N-n)!\,[F_0(\tilde{w})]^{k-m}\,[1-F_0(\tilde{w})]^{N-n-k+m}\,F_1'(\tilde{w})}{(k-m)!\,(N-n-k+m)!}$$

$$+ \frac{n!\,[F_1(\tilde{w})]^m\,[1-F_1(\tilde{w})]^{n-m}}{m!\,(n-m)!} \cdot \frac{(N-n)!\,[F_0(\tilde{w})]^{k-m}\,[1-F_0(\tilde{w})]^{N-n-k+m-1}\,F_0'(\tilde{w})}{(k-m)!\,(N-n-k+m-1)!}$$

To obtain $f(m, \tilde{w})$, sum (2.1.2) over all admissible values of $n$, recalling that $N = 2k + 1$,

$$f(m, \tilde{w}) = \frac{N!\,[pF_1(\tilde{w})]^m\,[qF_0(\tilde{w})]^{k-m}}{m!\,(k-m)!\,k!} \times$$

$$\left\{ \sum_{n=m+1}^{N} \frac{k!\,p^{n-m}\,q^{N-n-k+m}\,[1-F_1(\tilde{w})]^{n-m-1}\,[1-F_0(\tilde{w})]^{N-n-k+m}\,F_1'(\tilde{w})}{(n-m-1)!\,(N-n-k+m)!} \right.$$

$$\left. + \sum_{n=m}^{N} \frac{k!\,p^{n-m}\,q^{N-n-k+m}\,[1-F_1(\tilde{w})]^{n-m}\,[1-F_0(\tilde{w})]^{N-n-k+m-1}\,F_0'(\tilde{w})}{(n-m)!\,(N-n-k+m-1)!} \right\}$$

$$= \frac{(2k+1)!\,[pF_1(\tilde{w})]^m\,[qF_0(\tilde{w})]^{k-m}}{m!\,(k-m)!\,k!} \times$$

$$[1 - pF_1(\tilde{w}) - qF_0(\tilde{w})]^k\,[pf_1'(\tilde{w}) + qF_0'(\tilde{w})] \;. \quad ||$$

Under the hypothesis $H$: $F_1 = F_0$, (2.1.1) reduces to,

(2.1.5) $$f(m, \tilde{w}) = \frac{(2k+1)!}{m!\,(k-m)!\,k!}\; p^m\, q^{k-m}\,[F(\tilde{w})]^k\,[1-F(\tilde{w})]^k\,F'(\tilde{w}) \;,$$

$m = 0, 1, \ldots, k$ .

To obtain the marginal distribution $f(m)$ integrate (2.1.5) over the domain $0 \leq F(\tilde{w}) \leq 1$ :

(2.1.6) $$f(m) = \binom{k}{m} p^m q^{k-m} ,$$

so that $m$ is a binomial random variable $b(k, p)$ . The marginal distribution $f(w)$ is obtained by summing (2.1.5) over $m$ :

$$f(\tilde{w}) = \frac{(2k+1)!}{k!\,k!} [F(\tilde{w})]^k [1-F(\tilde{w})]^k F'(\tilde{w}) , \qquad -\infty < \tilde{w} < \infty .$$

## 2.2 Asymptotic Distributions

Let $\xi$ denote the median of the distribution of $Y$ , i.e. $\xi$ is the root (assumed unique) of the equation

(2.2.1) $$pF_1(\xi) + qF_0(\xi) = 1/2 .$$

As before let $m$ denote the number of $U_i$'s that are less than $\tilde{W}$ , the sample median of the combined sample of $U_i$'s and $V_j$'s . The following theorem gives the asymptotic joint distribution of $m$ and $\tilde{W}$ .

Theorem 2.2.1. Let

$$\nu = \frac{m - NpF_1(\xi)}{\sqrt{NpF_1(\xi)}} , \qquad \eta = \sqrt{N}\,(\tilde{W} - \xi) ,$$

where $\xi$ satisfies (2.2.1). Assume that in some neighbourhood of the density function $f_i(x) = F_i'(x)$ $(i = 0, 1)$ has a continuous derivative. Then the asymptotic joint distribution of $(\nu, \eta)$ is bivariate normal with zero mean vector and covariance matrix $\Sigma = (\sigma_{ij})$ given by

$$\sigma_{11} = \sigma^2(\nu) = qF_0(\xi) + \frac{q[pf_1^2(\xi)F_0(\xi) + qf_0^2(\xi)F_1(\xi)]}{2F_1(\xi)[pf_1(\xi)+qf_0(\xi)]^2},$$

$$\sigma_{22} = \sigma^2(\eta) = \frac{1}{4[pf_1(\xi)+qf_0(\xi)]^2},$$

$$\sigma_{12} = \sigma_{21} = \mathrm{cov}(\nu,\eta) = \frac{pq[f_1(\xi)F_0(\xi) - f_0(\xi)F_1(\xi)]}{\sqrt{2pF_1(\xi)}\,[pf_1(\xi)+qf_0(\xi)]^2}.$$

**Proof.** Throughout this proof for simplicity write $F_i(\xi) = F_i$ and $f_i(\xi) = f_i$. The joint probability density function of $m$ and $\tilde{w}$ is given by (2.1.1). Substitution of expressions for $\nu$ and $\eta$ in (2.1.1) yields

(2.2.2) $$f(m,\tilde{w}) = \frac{N!\,p^m q^{k-m}}{m!\,(k-m)!\,k!}\left[F_1\left(\xi+\frac{\eta}{\sqrt{N}}\right)\right]^m\left[F_0\left(\xi+\frac{\eta}{\sqrt{N}}\right)\right]^{k-m}\times$$

$$\left[1 - pF_1\left(\xi+\frac{\eta}{\sqrt{N}}\right) - qF_0\left(\xi+\frac{\eta}{\sqrt{N}}\right)\right]^k\times$$

$$\left[pf_1\left(\xi+\frac{\eta}{\sqrt{N}}\right) + qf_0\left(\xi+\frac{\eta}{\sqrt{N}}\right)\right].$$

Expand $F_i\left(\xi+\frac{\eta}{\sqrt{N}}\right)$ in Taylor's series about $\xi$:

$$F_i\left(\xi+\frac{\eta}{\sqrt{N}}\right) = F_i + \frac{\eta}{\sqrt{N}}f_i + o\left(\frac{\eta^2}{N}\right), \quad i = 0,1;$$

$$1 - pF_1\left(\xi+\frac{\eta}{\sqrt{N}}\right) - qF_0\left(\xi+\frac{\eta}{\sqrt{N}}\right) = \frac{1}{2} - \frac{\eta}{\sqrt{N}}(pf_1+qf_0) - o\left(\frac{\eta^2}{N}\right).$$

Using these expansions in (2.2.2) and noting that $N = 2k + 1$ , we get

$$f(m,\tilde{w}) = \left\{ \frac{N\,(2k)!}{k!\,k!\,2^{2k}} \right\} \left\{ \frac{k!}{m!\,(k-m)!} (2pF_1)^m (2qF_0)^{k-m} \right\} \times$$

$$\left\{ \left[ 1 + \frac{\eta}{\sqrt{N}} \frac{f_1}{F_1} + o\left(\frac{\eta^2}{N}\right) \right]^m \left[ 1 + \frac{\eta}{\sqrt{N}} \frac{f_0}{F_0} + o\left(\frac{\eta^2}{N}\right) \right]^{k-m} \times \right.$$

$$\left. \left[ 1 - \frac{2\eta}{\sqrt{N}} (pf_1 + qf_0) + o\left(\frac{\eta^2}{N}\right) \right]^k \right\} \left\{ \left\{ pf_1\left(\xi + \frac{\eta}{\sqrt{N}}\right) + qf_0\left(\xi + \frac{\eta}{\sqrt{N}}\right) \right\}^2 \right\}$$

$$\equiv \{A_1\} \{A_2\} \{A_3\} \{A_4\} .$$

Let $S = \{ (v, \eta) : \; a \le v \le b \, , \; c \le \eta \le d \}$ where a, b, c, d, are finite. Now using Stirling's formula for $n!$ we have

$$A_1 \sim \frac{N}{\sqrt{k}} \frac{1}{\sqrt{\pi}} .$$

$A_2$ does not depend on $\eta$ and due to convergence of binomial distribution to normal, uniformly in $S$ we have

$$A_2 \sim \frac{1}{\sqrt{8\pi k p q F_1 F_0}} \exp\left( \frac{-v^2}{4qF_0} \right) .$$

Now consider

$$\log A_3 = (NpF_1 + v \;\; NpF_1) \;\; \log\left( 1 + \frac{\eta}{\sqrt{N}} \frac{f_1}{F_1} + o\left(\frac{\eta^2}{N}\right) \right)$$

$$+ \;(k - Npf_1 - v \;\; NpF_1) \;\; \log\left( 1 + \frac{\eta}{\sqrt{N}} \frac{f_0}{F_0} + o\left(\frac{\eta^2}{N}\right) \right)$$

$$+ \; k \log\left( 1 - \frac{2\eta}{\sqrt{N}} (pf_1 + qf_0) - o\left(\frac{\eta^2}{N}\right) \right) .$$

Using series expansion for $\log(1 + x)$ it is seen that for all $(v, \eta) \in S$

$$\log A_3 = \frac{-\eta^2}{2}\left[ [illegible] + 2(pf_1 + qf_0) \right] + [illegible] \left( \frac{f_1}{F_1} - \frac{f_0}{F_0} \right) + o(1) .$$

Using continuity of $f_1$ at [illegible] it follows that for all values of $m$ and $\tilde{W}$ for which $(v, \eta) \in S$

$$f(m, \tilde{W}) \sim \frac{[illegible]}{[illegible]} \exp -\frac{1}{2}\left\{ \eta^2 \left[ [illegible] + 2(pf_1 + qf_0) \right] \right.$$

$$\left. + \frac{v^2}{2[illegible]} - [illegible] \left( \frac{f_1}{F_1} - \frac{f_0}{F_0} \right) \right\} .$$

Now after making the transformation from $(m, \tilde{W}) \longrightarrow (v, \eta)$ we get

$$P\{ a \leq v \leq b , c \leq \quad \leq d \}$$

$$\sim \frac{(pf_1 + qf_0)}{[illegible]} \int_c^d \exp -\frac{1}{2}\left[ \eta^2 \{ [illegible] + 2(pf_1 + qf_0) \} \right] \times$$

$$\left\{ \sum_{a \leq v \leq b} \frac{1}{[illegible]} \exp\left[ -\frac{v^2}{2[illegible]} + [illegible] \left( \frac{f_1}{F_1} - \frac{f_0}{F_0} \right) \right] \right\} d\eta .$$

Hence $P\{ a \leq v \leq b, c \leq \eta \leq d \} \longrightarrow \int_a^b \int_c^d f(v, \eta) dv d\eta$ ,

where $f(v, \eta)$ is the density function of the bivariate normal distribution stated in the theorem. ||

Remark. It follows from the above theorem, that when the hypothesis $H$: $F_1 = F_0$ is true, $\sigma_{12} = 0$, so that $\nu$ and $\eta$ are asymptotically independent. $\nu$ is asymptotically $\mathcal{N}(0, q)$ and $\eta$ is asymptotically $\mathcal{N}(0, 1/4f^2(\xi))$.

## 2.3 Consistency of the Test

Consider a two sided test of the hypothesis $H$: $F_1 = F_0$ against $F_1 \neq F_0$, for which the critical region is given by $|(m-kp) / (kpq)^{1/2}| > t_{N,\alpha}$. The sequence $t_{N,\alpha}$ is chosen so that $\lim_{N \to \infty} t_{N,\alpha} = t_\alpha$, where $t_\alpha$ satisfies $1 - \Phi(t_\alpha) = \alpha/2$, and $\Phi(t)$ is the standardized normal distribution function. Then the power of the test is given by

$$P\left\{ \left| \frac{m-kp}{\sqrt{kpq}} \right| > t_{N,\alpha} \middle| F_1 \neq F_0 \right\}$$

$$= 1 - P\left\{ -\frac{t_{N,\alpha}}{2\sqrt{F_1(\xi)F_0(\xi)}} + \frac{kp(1-2pF_1(\xi))}{2\sqrt{kpq\,F_1(\xi)F_0(\xi)}} \right.$$

$$\left. < \frac{m-kpF_1(\xi)}{2\sqrt{kpq\,F_1(\xi)F_0(\xi)}} < \frac{t_{N,\alpha}}{2\sqrt{F_1(\xi)F_0(\xi)}} + \frac{kp(1-2F_1(\xi))}{2\sqrt{kpq\,F_1(\xi)F_0(\xi)}} \right\}.$$

If $F_1(\xi) \neq 1/2$, the power approaches $1$, as $N \to \infty$, and hence the test is consistent.

For alternatives $F_1 > F_0$, $(F_1 < F_0)$ we reject H if m is too large (small), and prove in a similar manner that the test is consistent if $F_1(\xi) > 1/2$, $(F_1(\xi) < 1/2)$.

2.4 Asymptotic Efficiency of the Test

Definition: Given two tests of the same size of the same statistical hypotheses, the relative efficiency of the first test with respect to the second is given by the ratio $n_2/n_1$, where $n_1$ is the sample size of the first test required to achieve the same power for a given alternative, A, as is achieved by the second test with respect to A, when using a sample of size $n_2$. For a sequence of alternatives changing with n in such a way that as $n \longrightarrow \infty$ the power of the corresponding sequences of the tests converge to some number less than 1, the relative asymptotic efficiency of the first test with respect to the second is defined as the limit of the corresponding ratios $n_2/n_1$.

Let $F_1(y) = F_0(y - \theta)$, then $H: F_1 = F_0$ is equivalent to $H: \theta = 0$. For alternatives $\theta > 0$, we evaluate the relative asymptotic efficiency of the median test with respect to the corresponding parametric test when $F_1$ and $F_0$ are normal, with means $\mu_1$ and $\mu_0$, respectively, and a common variance $\sigma^2$. Then $F_1(y) = F_0(y)$ if and only if $\mu_1 = \mu_0$, which is equivalent to $\rho = \rho(X, Y) = 0$, where $\rho$ is the correlation coefficient between X and Y, Tate [11]. Let $\theta = (\mu_1 - \mu_0)/\sigma$, then we are testing the hypothesis $\theta = 0$ against $\theta > 0$. The test is based on the sample correlation coefficient r, defined by

$$r = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}}{\left[\sum_{i=1}^{N}(X_i-\bar{X})^2 \cdot \sum_{i=1}^{N}(Y_i-\bar{Y})^2\right]^{1/2}}$$

Tate [11] proved that $r$ is asymptotically normally distributed with asymptotic mean and variance given by

$$(2.4.1) \quad \mu_\theta(r) = \theta\left(\frac{pq}{1+pq\theta^2}\right)^{1/2}, \quad \sigma_\theta^2(r) = \frac{4(1+pq\theta^2)-\theta^2(6pq-1)}{4N(1+pq\theta^2)^3} .$$

The critical region for this test is given by $r\sqrt{N} > t'_{N',\alpha}$, where $\{t'_{N',\alpha}\}$ is such that $\lim_{N'\to\infty} t'_{N',\alpha} = t_\alpha$ and $\Phi(t_\alpha) = 1-\alpha$.

The power of the test is given by

$$\beta'_{N'}(\theta) = P_\theta\left\{r > \frac{t'_{N',\alpha}}{\sqrt{N'}}\right\} = 1 - P_\theta\left\{\frac{r-\mu_\theta(r)}{\sigma_\theta(r)} > \frac{t'_{N',\alpha}-\sqrt{N'}\,\mu_\theta(r)}{\sqrt{N'}\,\sigma_\theta(r)}\right\} .$$

Since $r$ is asymptotically normally distributed,

$$\lim_{N'\to\infty} \beta'_{N'}(\theta) = 1 - \Phi\left(\lim_{N'\to\infty} \frac{t'_{N',\alpha} - \sqrt{N'}\,\mu_\theta(r)}{\sqrt{N'}\,\sigma_\theta(r)}\right)$$

Now for a sequence of alternatives $\{\theta'_{N'}\}$ where

$$\theta'_{N'} = \delta'/\sqrt{N'}, \quad \delta' > 0, \quad \lim_{N'\to\infty} 1/\left[\sqrt{N'}\,\sigma_{\theta'_{N'}}(r)\right] = 1$$

and

$$r = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}}{\left[\sum_{i=1}^{N}(X_i-\bar{X})^2 \cdot \sum_{i=1}^{N}(Y_i-\bar{Y})^2\right]^{1/2}}$$

Tate [11] proved that $r$ is asymptotically normally distributed with asymptotic mean and variance given by

$$(2.4.1) \qquad \mu_\theta(r) = \theta\left(\frac{pq}{1+pq\theta^2}\right)^{1/2}, \quad \sigma_\theta^2(r) = \frac{4(1+pq\theta^2)-\theta^2(6pq-1)}{4N(1+pq\theta^2)^3} .$$

The critical region for this test is given by $r\sqrt{N} > t'_{N',\alpha}$, where $\left\{t'_{N',\alpha}\right\}$ is such that $\lim_{N' \to \infty} t'_{N',\alpha} = t_\alpha$ and $\Phi(t_\alpha) = 1 - \alpha$.

The power of the test is given by

$$\beta'_{N'}(\theta) = P_\theta\left\{r > \frac{t'_{N',\alpha}}{\sqrt{N'}}\right\} = 1 - P_\theta\left\{\frac{r-\mu_\theta(r)}{\sigma_\theta(r)} > \frac{t'_{N',\alpha}-\sqrt{N'}\,\mu_\theta(r)}{\sqrt{N'}\,\sigma_\theta(r)}\right\} .$$

Since $r$ is asymptotically normally distributed,

$$\lim_{N' \to \infty} \beta'_{N'}(\theta) = 1 - \Phi\left(\lim_{N' \to \infty} \frac{t'_{N',\alpha} - \sqrt{N'}\,\mu_\theta(r)}{\sqrt{N'}\,\sigma_\theta(r)}\right)$$

Now for a sequence of alternatives $\left\{\theta'_{N'}\right\}$ where

$$\theta'_{N'} = \delta'/\sqrt{N'}, \; \delta' > 0, \quad \lim_{N' \to \infty} 1/\left[\sqrt{N'}\,\sigma_{\theta'_{N'}}(r)\right] = 1$$

and

$$\lim_{N' \to \infty} \left[ \mu_{\theta'_{N'}}(r) \Big/ \sigma_{\theta'_{N'}}(r) \right] = \delta' \sqrt{pq} \quad .$$

Therefore

$$\text{(2.4.2)} \qquad \lim_{N' \to \infty} \beta'_{N'}(\theta'_{N'}) = \Phi(-t_\alpha + \delta' \sqrt{pq} \quad ) \quad .$$

Now for the median test under consideration the critical region for testing $H: \theta = 0$, against the alternative $\theta > 0$, is given by $(m - kp)/\sqrt{kpq} < t_{N,\alpha}$ where $\lim_{N \to \infty} t_{N,\alpha} = -t_\alpha$ with $\Phi(-t_\alpha) = \alpha$.

The power of the test is obtained from

$$\beta_N(\theta) = P_\theta\left\{ \frac{m - kp}{\sqrt{kpq}} < -t_{N,\alpha} \right\}$$
$$= P_\theta\left\{ \frac{m - NpF_1(\xi)}{\sigma_\theta(m)} < \frac{-t_{N,\alpha}\sqrt{kpq} - (NpF_1(\xi) - kp)}{\sigma_\theta(m)} \right\}$$

where

$$\text{(2.4.3)} \qquad \sigma_\theta^2(m) = NpF_0(\xi-\theta)\left(qF_0(\xi-\theta) + \frac{[\text{illegible}]}{2F_0(\xi-\theta)[pF_0(\xi-\theta)+qF_0(\xi)]^2}\right) .$$

By Theorem 2.2.1

$$\lim_{N \to \infty} \beta_N(\theta) = \Phi\left( \lim_{N \to \infty} \frac{-t_{N,\alpha}\sqrt{kpq} - (NpF_1(\xi) - kp)}{\sigma_\theta(m)} \right) .$$

For a sequence of alternatives $\{\theta_N\}$, with $\theta_N = \delta/\sqrt{N}$,

$\delta > 0$, $\lim\limits_{N \to \infty} \dfrac{\sqrt{kpq}}{\sigma_{\theta_N}(m)} = 1$ . Also since $\xi$ satisfies

$pF_0(\xi - \theta) + qF_0(\xi) = 1/2$ ,

$$\frac{d\xi}{d\theta} = \frac{pf_0(\xi-\theta)}{pf_0(\xi-\theta)+qf_0(\xi)} .$$

Hence for the sequence $\{\theta_N\}$

$$NpF_0(\xi-\delta/\sqrt{N}) = Np\left[F_0(\xi)+\frac{\delta}{\sqrt{N}}f_0(\xi)(p-1)+o\left(\frac{\delta^2}{N}\right)\right]$$

so that

$$\lim_{N \to \infty} \frac{NpF_0(\xi-\delta/\sqrt{N})-kp}{\sigma_{\theta_N}(m)} = \delta f_0(\xi)\sqrt{2pq}$$

which yields

$$(2.4.4) \qquad \lim_{N \to \infty} \beta_N(\theta_N) = \Phi(-t_\alpha + \delta f_0(\xi)\sqrt{2pq}) .$$

The two sequences $\{\theta_N\}$ and $\{\theta'_{N'}\}$ will be the same if $N'/N = \delta'^2/\delta^2$ . From (2.4.2) and (2.4.4) it is seen that $\lim\limits_{N \to \infty} \beta_N(\theta_N) = \lim\limits_{N \to \infty} \beta'_{N'}(\theta'_{N'})$ only if $\delta'/\delta = \sqrt{2}\ f_0(\xi)$ . Hence the required efficiency is given by $e(M, r) = 2f_0^2(\xi) = 1/\pi$ .

## 2.5 Case When p is Unknown

The theory developed so far is not applicable when $p$ in unknown. The usual estimate for $p$ is $\hat{p} = n/N$ , and we consider the test based on the statistic $(m-k\hat{p}) / (k\hat{p}\hat{q})^{1/2}$ . We now show that the test based on this statistic is asymptotically distribution free.

__Theorem 2.5.1.__ Under the hypothesis $H: F_1 = F_0$ , the statistic $(m-k\hat{p}) / (k\hat{p}\hat{q})^{1/2}$ is asymptotically normally distributed with mean zero and variance $1/2$ .

__Proof.__ Since $\text{plim}\, \hat{p} = p$ , by an application of Slutsky's theorem [9, p. 255], $\text{plim}\, (\hat{p}\hat{q}/pq)^{1/2} = 1$ . Hence the limiting distribution of $(m - k\hat{p})/\sqrt{k\hat{p}\hat{q}}$ is the same as that of $(m - k\hat{p})/\sqrt{kpq}$ . Write

$$\frac{m-k\hat{p}}{\sqrt{kpq}} = \frac{m-kp}{\sqrt{kpq}} - \frac{k(\hat{p}-p)}{\sqrt{kpq}} \equiv T_1 - T_2 .$$

The asymptotic joint distribution of $(T_1, T_2)$ is bivariate normal $\mathcal{N}(0, \Sigma)$ with $\Sigma = (\sigma_{ij})$ where $\sigma_{11} = 1$ , $\sigma_{12} = \sigma_{21} = \sigma_{22} = 1/2$ . Hence the required result follows. ||

## 3. __Two Sample Wilcoxon Test__

As before, let $Z_i = (X_i, Y_i)$ , $i = 1, 2, \ldots, N$ , be $N$ independent observations from a bivariate population, where $X$ assumes only two values, 1 and 0 with probabilities $p$ and $q = 1 - p$ respectively. The test statistics may then be defined as,

$$\bar{U}_N = \frac{1}{N(N-1)} \sum_{i \neq j = 1}^{N} H(Z_i, Z_j) \quad ,$$

where,

$$H(Z_i, Z_j) = \begin{cases} 1, \text{ if } X_i = 1, X_j = 0 \text{ and } Y_i < Y_j , \\ 0, \text{ otherwise} . \end{cases}$$

If $U_1, U_2, \ldots, U_n$ denote those $Y$ observations for which the corresponding values of $X$ are observed to be $1$, and $V_1, V_2, \ldots, V_{N-n}$ the remaining observations on $Y$, then $N(N-1)\,\bar{U}_N$ is the total number of pairs $(U_i, V_j)$ such that $U_i < V_j$. The hypothesis $H\colon\ F_1 = F_0$ is rejected if $\bar{U}_N$ is either too large or too small.

In Sections 3.1 and 3.2 we obtain the mean and variance of $\bar{U}_N$, and the exact sampling distribution of $\bar{U}_N$ under the hypothesis. The asymptotic distribution of $\bar{U}_N$, both under the hypothesis and the alternative, in the case when $p$ is known, is obtained in Section 3.3. In Section3.4 we prove consistency of the test, and in Section 3.5 find its asymptotic efficiency. Lastly, Section 3.6 deals with the case when $p$ is unknown, where it is shown that the test statistic, with $p$ replaced by its estimate $\hat{p}$, does not yield an asymptotically distribution-free test.

## 3.1 Mean and Variance of $\bar{U}_N$.

$$
\begin{aligned}
(3.1.1) \qquad E_p(\bar{U}_N) &= E_p\, H(Z_i, Z_j) = P\left\{X_i = 1,\ X_j = 0 \ \text{ and } \ Y_i < Y_j\right\} \\
&= P\left\{X_i = 1,\ X_j = 0\right\}\ P\left\{\, Y_i < Y_j \mid X_i = 1,\ X_j = 0\right\} \\
&= pq \int F_1(y)\, dF_0(y) \ .
\end{aligned}
$$

To compute the variance of $\bar{U}_N$, write $\bar{U}_N$ as

$$
(3.1.2) \qquad \bar{U}_N = \frac{1}{N(N-1)} \sum_{j=1}^{N-n} \sum_{i=1}^{n} \phi(U_i, V_j) \ ,
$$

where

$$\phi(u, v) = \begin{cases} 1, & \text{if } u < v, \\ 0, & \text{otherwise}. \end{cases}$$

Squaring (3.1.2), and taking expected values, we obtain the conditional moment:

(3.1.3) $$N^2(N-1)^2 E_p(\bar{U}_N^2 | n) = E_p \sum_{j=1}^{N-n} \sum_{i=1}^{n} \phi(U_i, V_j)$$

$$+ E_p \sum_{j=1}^{N-n} \sum_{i \neq k=1}^{n} \phi(U_i, V_j)\phi(U_k, V_j) + E_p \sum_{j \neq k=1}^{N-n} \sum_{i=1}^{n} \phi(U_i, V_j)\phi(U_i, V_k)$$

$$+ E_p \sum_{j \neq k=1}^{N-n} \sum_{i \neq r=1}^{n} \phi(U_i, V_j)\phi(U_r, V_k)$$

$$= n(N-n) P\{U_i < V_j\} + n(n-1)(N-n) P\{U_i < V_j, U_k < V_j\}$$

$$+ n(N-n)(N-n-1) P\{U_i < V_j, U_i < V_k\}$$

$$+ n(n-1)(N-n)(N-n-1) P\{U_i < V_j, U_r < V_k\}$$

$$= n(N-n) \int F_1 dF_0 + n(n-1)(N-n) \int F_1^2 dF_0$$

$$+ n(N-n)(N-n-1) \int [1-F_0]^2 dF_1 + n(n-1)(N-n)(N-n-1) \left[\int F_1 dF_0\right]^2.$$

Since $n$ has a binomial distribution, $b(N, p)$,

$$(3.1.4)\qquad \begin{aligned} &En(N-n) = N(N-1)pq\ ,\ En(n-1)(N-n) = N(N-1)(N-2)p^2q\ ,\\ &En(N-n)(N-n-1) = N(N-1)(N-2)pq^2\ ,\\ &En(n-1)(N-n)(N-n-1) = N(N-1)(N-2)(N-3)p^2q^2\ .\end{aligned}$$

Using (3.1.3) and (3.1.4), the unconditional moment is

$$N^2(N-1)^2E_p\bar{U}_N^2 = N^2(N-1)^2E[E_p(\bar{U}_N^2|n)] = N(N-1)pq\int F_1dF_0 + N(N-1)(N-2)p^2q\int F_1^2dF_0$$
$$+ N(N-1)(N-2)pq^2\int [1-F_0]^2dF_1 + N(N-1)(N-2)(N-3)p^2q^2\left[\int F_1dF_0\right]^2 .$$

Hence

$$(3.1.5)\qquad \sigma_p^2\ (\bar{U}_N) = \frac{pq}{N(N-1)}\left[\int F_1dF_0 + (N-2)p\int F_1^2dF_0 + (N-2)q\int (1-F_0)^2dF_1 - 2pq(2N-3)\left(\int F_1dF_0\right)^2\right] .$$

In particular, under $H$: $F_1 = F_0$, (3.1.1) and (3.1.5) reduce to

$$(3.1.6)\qquad E_p(\bar{U}_N|H) = (pq)/2\ ,$$

$$(3.1.7)\qquad \sigma_p^2\ (\bar{U}_N|H) = \frac{pq}{N(N-1)}\left[\frac{(2N-1)}{6} - \frac{(2N-3)}{2}\ pq\right] .$$

## 3.2 Distribution of $\bar{U}_N$

Define

$$(3.2.1)\qquad T_N = N(N-1)\ \bar{U}_N = \left\{\begin{aligned}&\text{number of pairs } (Z_i, Z_j) \text{ such that}\\ &X_i = 1,\ X_j = 0 \text{ and } Y_i < Y_j\end{aligned}\right\} .$$

$T_N$ takes values 0, 1, ..., k , where $k = \max n(N-n) = [N^2/4]$ , and $[x]$ denotes the largest integer $\leq x$ . Let $T_{n,N-n}$ denote the value of $T_N$ when n is fixed. Clearly $T_{n,N-n}$ takes values 0, 1, 2, ..., $n(N-n)$ , and

$$(3.2.2) \qquad P\{T_N = t\} = \sum_{n=0}^{N} P\{n \text{ of the } X_i\text{'s} = 1,\ N-n \text{ of } X_i\text{'s} = 0, \text{ and } T_{n,N-n} = t\}$$

$$= \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} P\{T_{n,N-n} = t\} .$$

Mann and Whitney [12] have shown that the probability $P\{T_{n,N-n} = t\}$ satisfies the following recurrence relation:

$$(3.2.3) \qquad P\{T_{n,N-n} = t\} = \frac{n}{N} P\{T_{n-1,N-n} = t\} + \frac{(N-n)}{N} P\{T_{n,N-n-1} = t-n\}$$

Substituting (3.2.3) in (3.2.2) we get

$$(3.2.4) \qquad P\{T_N = t\} = p \sum_{n=1}^{N} \binom{N-1}{n-1} p^{n-1} q^{N-n} P\{T_{n-1,N-n} = t\}$$

$$+ q \sum_{n=0}^{t} \binom{N-1}{n} p^n q^{N-n-1} P\{T_{n,N-n-1} = t-n\}$$

$$= p P\{T_{N-1} = t\} + q \sum_{n=0}^{t} \binom{N-1}{n} p^n q^{N-n-1} P\{T_{n,N-n-1} = t-n\} .$$

(3.2.4) is a recurrence relation for $P\{T_N = t\}$, from which we can find the distribution of $T_N$ for all $N$. It is easy to prove by induction from (3.2.4) that,

$$P\{T_N = 0\} = \sum_{r=0}^{N} p^{N-r}q^r, \qquad \text{for all } N. \tag{3.2.5}$$

The probability distribution of $T_N$ obtained by using (3.2.4) is given below for $N = 2, 3, 4, 5$.

| | $t$ | $P\{T_N = t\}$ |
|---|---|---|
| | 0 | $p^2 + pq + q^2$ |
| $N = 2$ | 1 | $pq$ |
| | 2 | 0 |
| | 0 | $p^3 + p^2q + pq^2 + q^3$ |
| | 1 | $p^2q + pq^2$ |
| $N = 3$ | 2 | $p^2q + pq^2$ |
| | 3 | 0 |
| | 0 | $p^4 + p^3q + p^2q^2 + pq^3 + q^4$ |
| | 1 | $p^3q + p^2q^2 + pq^3$ |
| $N = 4$ | 2 | $p^3q + 2p^2q^2 + pq^3$ |
| | 3 | $p^3q + p^2q^2 + pq^3$ |
| | 4 | $p^2q^2$ |
| | 0 | $p^5 + p^4q + p^3q^2 + p^2q^3 + pq^4 + q^5$ |
| | 1 | $p^4q + p^3q^2 + p^2q^3 + pq^4$ |
| | 2 | $p^4q + 2p^3q^2 + 2p^2q^3 + pq^4$ |
| $N = 5$ | 3 | $p^4q + 2p^3q^2 + 2p^2q^3 + pq^4$ |
| | 4 | $p^4q + 2p^3q^2 + 2p^2q^3 + pq^4$ |
| | 5 | $p^3q^2 + p^2q^3$ |
| | 6 | $p^3q^2 + p^2q^3$ |

## 3.3 Asymptotic Distribution of $\bar{U}_N$

Let $Z_1, \ldots, Z_N$ be $N$ independent and identically distributed random variables; let $\phi\,(z_{\alpha_1}, z_{\alpha_2}, \ldots z_{\alpha_s})$, $s < N$, be a real valued symmetric function of its arguments. Hoeffding [13] defines a U-statistic as follows:

$$(3.3.1) \qquad U(z_1, \ldots, z_N) = \frac{1}{\binom{N}{s}} \sum_{C_n} \phi(z_{\alpha_1}, z_{\alpha_2}, \ldots, z_{\alpha_s}) ,$$

where $C_n$ means that the summation is over all combinations, $(\alpha_1, \alpha_2, \ldots, \alpha_s)$, of $s$ integers chosen from $(1, 2, \ldots, N)$; and proves the following theorem on the asymptotic normality of a U-statistic,

Theorem 3.3.1. (Hoeffding)

Let $Z_1, \ldots, Z_N$ be independent and identically distributed random variables. Let $U(z_1, \ldots z_N)$ be a U-statistic. If $E\,[\phi(z_{\alpha_1}, \ldots z_{\alpha_s})]^2 < \infty$, then as $N \longrightarrow \infty$, the limiting distribution of $(U_N - EU_N) / \sigma(U_N)$ is $\mathcal{N}(0, 1)$.

Clearly $\bar{U}_N$ is a U-statistic and hence $[\bar{U}_N - E_p(\bar{U}_N)] / \sigma_p(\bar{U}_N)$ is asymptotically $\mathcal{N}(0, 1)$, both under the hypothesis $H$ as well as under the alternative.

## 3.4 Consistency of the Test

Consider the two-sided test of the hypothesis $H$: $F_1 = F_0$ against $A$: $F_1 \neq F_0$, with critical region, $|[\bar{U}_N - E_p\bar{U}_N] / \sigma_p(\bar{U}_N)| > t_{N,\alpha}$. The sequence $\{t_{N,\alpha}\}$ is chosen so that $\lim_{N \to \infty} t_{N,\alpha} = t_\alpha$, where $t_\alpha$

satisfies $1 - \Phi(t_\alpha) = \alpha/2$ . The power of the test is given by

$$P\left\{\left|\frac{\bar{U}_N - E_0(\bar{U}_N)}{\sigma_0(\bar{U}_N)}\right| > t_\alpha \,\middle|\, F_1 \neq F_0\right\} = 1 - P\left\{-t_\alpha < \frac{\bar{U}_N - E_0(\bar{U}_N)}{\sigma_0(\bar{U}_N)} < t_\alpha \,\middle|\, F_1 \neq F_0\right\} .$$

Proceeding as in Section 2.3, if $\int F_1 dF_0 \neq 1/2$ , the power tends to $1$ , as $N \longrightarrow \infty$ , and hence the test is consistent. In a similar manner it can be verified, that the test is consistent when $F_1 > F_0$ or $F_1 < F_0$ .

### 3.5 Asymptotic Efficiency of the Test

We now find the asymptotic efficiency of the test based on $\bar{U}_N$ with respect to the parametric test based on the sample correlation coefficient between $X$ and $Y$ , described in Section 2.4. We have seen in Section 3.3 that $\bar{U}_N$ is asymptotically normally distributed both under the hypothesis and the alternative. Proceeding as in Section 2.4 it can be proved that the required relative asymptotic efficiency is given by

$$(3.5.1) \qquad e(\bar{U}_N, r) = \frac{3\left[\int f_0^2(y)dy\right]^2}{1 - 3pq} = \frac{3}{4\pi(1 - 3pq)} .$$

The asymptotic efficiency given by (3.5.1) is a maximum, namely $3/\pi$ when $pq = 1/4$ , and is a minimum, namely $3/4\pi$ when $pq = 0$ .

## 3.6 Case When $p$ is Unknown

We now estimate $p$ by its usual estimate $\hat{p} = n/N$ and consider the test based on the statistic, $[\bar{U}_N - E_{\hat{p}}\bar{U}_N] \,/\, \sigma_{\hat{p}}(\bar{U}_N)$, where $E_{\hat{p}}(\bar{U}_N)$ and $\sigma_{\hat{p}}(\bar{U}_N)$ are obtained by replacing $p$ and $q$ by $\hat{p}$ and $\hat{q}$, respectively, in (3.1.6) and (3.1.7). It is interesting to note that this test is not asymptotically distribution-free, in that it depends on the distribution of $X$.

<u>Theorem 3.6.1</u>. Under the hypothesis $H$: $F_1 = F_0$, the limiting distribution of the statistic $[\bar{U}_N - E_{\hat{p}}(\bar{U}_N)] \,/\, \sigma_{\hat{p}}(\bar{U}_N)$, is normal with mean zero and variance $\left[1 - \frac{3}{4}\,\frac{(1-2p)^2}{(1-3pq)}\right]$.

<u>Proof</u>. Because $\operatorname{plim} \hat{p} = p$, by an application of Slutsky's theorem, we obtain $\operatorname{plim} (\hat{p}\hat{q}/pq)^{1/2} = 1$, which implies that $\operatorname{plim} \sigma_{\hat{p}}(\bar{U}_N) = \sigma_p(\bar{U}_N)$. Hence by a theorem of Cramer [9, p. 254], it follows that the asymptotic distribution of $[\bar{U}_N - E_{\hat{p}}(\bar{U}_N)]/\sigma_{\hat{p}}(\bar{U}_N)$ is the same as that of $[\bar{U}_N - E_{\hat{p}}\bar{U}_N]/\sigma_p(\bar{U}_N)$. Write

$$\frac{\bar{U}_N - E_{\hat{p}}(\bar{U}_N)}{\sigma_p(\bar{U}_N)} = \frac{\bar{U}_N - E_p(\bar{U}_N)}{\sigma_p(\bar{U}_N)} - \frac{E_{\hat{p}}(\bar{U}_N) - E_p(\bar{U}_N)}{\sigma_p(\bar{U}_N)}$$

Since $\hat{p}\hat{q} - pq = (\hat{p} - p)(1 - 2p) - (\hat{p} - p)^2$, we have

$$(3.6.1)\quad \frac{\bar{U}_N - E_{\hat{p}}(\bar{U}_N)}{\sigma_{\hat{p}}(\bar{U}_N)} = \frac{\bar{U}_N - E_p(\bar{U}_N)}{\sigma_p(\bar{U}_N)} - \frac{(\hat{p}-p)(1-2p)}{2\left\{\frac{pq}{N(N-1)}\left[\frac{(2N-1)}{6} - \frac{(2N-3)pq}{2}\right]\right\}^{1/2}}$$

$$+ \frac{(\hat{p}-p)^2}{2\left\{\frac{pq}{N(N-1)}\left[\frac{(2N-1)}{6} - \frac{(2N-3)pq}{2}\right]\right\}^{1/2}} .$$

As $[\sqrt{N}\ (\hat{p} - p)]/\sqrt{pq}$ is bounded in probability and $\operatorname{plim} |\hat{p} - p| = 0$, the third term in (3.6.1) tends in probability to zero. By Hoeffding's Theorem [13, Theorem 7.2] the asymptotic joint distribution of the first two terms in (3.6.1) is bivariate normal $\mathcal{N}(0, \Sigma)$ where $\Sigma = (\sigma_{ij})$ with $\sigma_{11} = 1$ and $\sigma_{12} = \sigma_{21} = \sigma_{22} = [3(1 - 2p)^2]/[4(1 - 3pq)]$. This proves the theorem. ||

## 4. c - Sample Problem

Let $Z = (Y, X_1, X_2, \ldots, X_c)$ have a $(c + 1)$ variate distribution, where $X_j = 0$ or $1$,

$$\sum_{j=1}^{c} X_j = 1 ,\quad P\{X_j = 1\} = p_j ,\quad P\{X_j = 0\} = q_j = 1 - p_j ,\quad \text{and}$$

$P\{Y \leq y \mid X_j = 1\} = F_j(y)$, $j = 1, 2, \ldots, c$. The distribution functions $F_1, \ldots, F_c$ are absolutely continuous. On the basis of $N$ independent observations $Z_i = (Y_i, X_{1i}, X_{2i}, \ldots, X_{ci})$ $i = 1, 2, \ldots, N$, the hypothesis $H_c$: $F_1 = \ldots = F_c$ is to be tested. For this purpose divide the observations $Z_1, Z_2, \ldots, Z_N$ into $c$ sets according as $X_{ji} = 1$, $j = 1, 2, \ldots, c$. Let $U_{j1}, U_{j2}, \ldots, U_{jn_j}$ ($n_j > 0$ for each $j$, $\sum_{j=1}^{c} n_j = N$) denote those $Y_i$'s for which the corresponding $X_{ji} = 1$. The problem then reduces to that of testing the hypothesis that the $c$ independent samples of $U_{ji}$'s $(i = 1, \ldots, n_j, j = 1, \ldots, c)$ come from the same distribution, where the sample sizes $n_1, \ldots, n_c$ are random variables having a multinomial distribution with parameters $p_1, p_2, \ldots, p_c$.

We assume that the $F_j$'s differ only in location. Let $F_j(y) = F(y + \theta_j)$, $j = 1, 2, \ldots, c$ for some arbitrary choice of real numbers $\theta_1, \theta_2, \ldots, \theta_c$. Clearly $\theta_j = 0$ for all $j$ yields the hypothesis $H_c$. Further we denote by $H_N$ the hypothesis that specifies that $F_j(y) = F(y + \theta_j/\sqrt{N})$, $j = 1, 2, \ldots, c$, and for some pair $(i, j)$ $\theta_i \neq \theta_j$. It is known that the median test is sensitive to translation-type alternatives, so in this Section we generalize the two sample median test developed in Section 2 to the c-sample problem under consideration.

### 4.1 c-Sample Median Test

Let $\tilde{W}$ denote the median of the combined sample of $U_{ji}$'s $(i = 1, 2, \ldots n_j, j = 1, 2, \ldots c)$ and $m_j$ the number of $U_{ji}$'s $(i = 1, 2, \ldots n_j)$ that are less than $\tilde{W}$. We assume $N = 2k + 1$. Clearly $\sum_{j=1}^{c} m_j = k$. The test statistic proposed for testing the hypothesis $H_c$: $F_1 = F_2 = \dots = F_c$, may then be defined as

$$(4.1.1) \qquad M = \sum_{j=1}^{c} \left( \frac{m_j - k p_j}{\sqrt{N p_j}} \right)^2 ,$$

when $p_1, p_2, \dots, p_c$ are known, and as

$$(4.1.2) \qquad \hat{M} = \sum_{j=1}^{c} \left( \frac{m_j - k \hat{p}_j}{\sqrt{N \hat{p}_j}} \right)^2 ,$$

where $\hat{p}_j = n_j / N$, when $p_1, p_2, \dots, p_c$ are unknown. The test consists in rejecting the hypothesis $H_c$ if $M(\hat{M})$ is either too large or too small. Note that $M$ defined by (4.1.1) is different from the statistic defined for usual c-sample median test, e.g., see Andrews [14].

In Section 4.2 we find joint distribution of $m_1, \dots, m_c$ and $\tilde{W}$ and in Section 4.3 the limiting distribution of $M$. In Section 4.4 the relative asymptotic efficiency of the median test based on $M$ with respect to the corresponding parametric test based on multiple correlation coefficient is evaluated. Section 4.5 deals with the case when $p_1, p_2, \dots, p_c$ are unknown and gives the asymptotic distribution of $\hat{M}$ under the hypothesis $H_c$, from which we conclude that the test based on $\hat{M}$ is asymptotically distribution-free.

4.2 Joint Distribution of $m_1, m_2, \dots, m_c$ and $\tilde{W}$

Lemma 4.2.1. The joint distribution of $m_1, m_2, \dots, m_c$ and $\tilde{W}$ is

(4.2.1) $\quad f(m_1, m_2, \ldots, m_c, \tilde{w})$

$$= \frac{N! \prod_{j=1}^{c} [p_j F_j(\tilde{w})]^{m_j}}{\prod_{j=1}^{c} m_j!\ \ k!} \left[1 - \sum_{j=1}^{c} p_j F_j(\tilde{w})\right]^{N-k} \sum_{j=1}^{c} p_j f_j(\tilde{w}),$$

where $m_1, \ldots, m_c$ is a partition of $k$, $\sum_{j=1}^{c} m_j = k$.

<u>Proof</u>. As in Section 2.1 the conditional probability density of $m_1, m_2, \ldots, m_c$ and $\tilde{W}$ for fixed values of $n_1, \ldots, n_c$ is

(4.2.2) $\quad f(m_1, \ldots, m_c, \tilde{w} | n_1, n_2, \ldots, n_c)$

$$= \left[\sum_{j=1}^{c} \frac{n_j - m_j}{1 - F_j(\tilde{w})} f_j(\tilde{w})\right] \left[\prod_{j=1}^{c} \binom{n_j}{m_j} [F_j(\tilde{w})]^{m_j} [1 - F_j(\tilde{w})]^{n_j - m_j}\right].$$

$n_1, n_2, \ldots, n_c$ have the multinomial distribution $\mathcal{M}(N; p_1, p_2, \ldots, p_c)$ given by

(4.2.3) $$f(n_1, n_2, \ldots, n_c) = \frac{N!}{n_1! n_2! \cdots n_c!} p_1^{n_1} p_2^{n_2} \cdots p_c^{n_c}.$$

Hence using (4.2.2) and (4.2.3) we obtain

$f(m_1, m_2, \ldots, m_c \tilde{w})$

$$= \sum_{n_1, n_2, \ldots n_c} f(m_1, m_2, \ldots, m_c, \tilde{w} \mid n_1, n_2, \ldots, n_c)\, f(n_1, n_2, \ldots, n_c)$$

$$= \frac{N!}{\prod_{j=1}^{c} m_j!} \sum_{i=1}^{c} \frac{b_i F_i'(w)}{b_i(1-F_i(w))} \sum_{n_1, n_2, \ldots, n_c} \left\{ \left[ \frac{n_i - m_i}{\prod_{j=1}^{c} (n_j - m_j)!} \prod_{j=1}^{c} [b_j F_j(w)]^{m_j} \times \prod_{j=1}^{c} [b_j(1 - F_j(w))]^{n_j - m_j} \right] \right\} ,$$

where for each $i$ the summation is over all partitions $(n_1, n_2, \ldots, n_c)$ of $N$ such that $n_i \geq m_i + 1$ , $n_j > m_j (j \neq i)$ , which gives the required distribution (4.2.1). ||

Summing (4.2.1) over $m_1, m_2, \ldots, m_c$ we obtain the marginal distribution

$$f(w) = \frac{N!}{k!\,k!} \left[ \sum_{j=1}^{c} b_j F_j(w) \right]^k \left[ 1 - \sum_{j=1}^{c} b_j F_j(w) \right]^k \left[ \sum_{j=1}^{c} b_j F_j'(w) \right] .$$

Under $H_c$: $F_1 = F_2 = \ldots = F_c = F$ , integration over the domain $0 \leq F_j(w) \leq 1$ , in (4.2.1) yields the distribution of $m_1, m_2, \ldots, m_c$ :

$$(4.2.4) \qquad f(m_1, m_2, \ldots, m_c) = \frac{k!}{m_1!\, m_2! \cdots m_c!} p_1^{m_1} \cdots p_c^{m_c} ,$$

which is the multinomial distribution $\mathcal{M}_c(k; p_1, \ldots, p_c)$

## 4.3 Asymptotic Distribution of $M$

We first prove the following lemma which gives the limiting joint distribution of $m_1, m_2, \ldots, m_c$ and $\tilde{W}$.

**Lemma 4.3.1.** Let

$$\nu_j = \frac{m_j - Np_jF_j(\xi)}{\sqrt{Np_jF_j(\xi)}} \quad , \quad j = 1, 2, \ldots, c \; ; \quad \eta = \sqrt{N}(\tilde{W} - \xi)$$

where $\xi$ is such that

$$(4.3.1) \qquad \sum_{j=1}^{c} p_jF_j(\xi) = 1/2 \, .$$

Assume that in some neighbourhood of $\xi$ the density function $F_j'(y) = f_j(y)$ $(j = 1, 2, \ldots, c)$ has a continuous derivative. Then the asymptotic joint distribution of $\nu_1, \nu_2, \ldots, \nu_{c-1}$ and $\eta$ is c-variate normal distribution with zero mean vector and covariance matrix $\Sigma$ given by $\Sigma^{-1} = \Lambda = (\lambda_{ij})$ where

$$\lambda_{ii} = 1 + \frac{p_iF_i(\xi)}{p_cF_c(\xi)}, \quad i = 1, 2, \ldots, (c - 1) \, ,$$

$$\lambda_{cc} = \sum_{i=1}^{c} \frac{p_i f_i^2(\xi)}{F_i(\xi)} + 2\left[\sum_{i=1}^{c} p_i f_i(\xi)\right]^2 ,$$

$$\lambda_{ij} = \frac{\sqrt{p_i p_j f_i(\xi) f_j(\xi)}}{p_c F_c(\xi)} , \quad i \neq j = 1, 2, \cdots, (c-1) ;$$

$$\lambda_{ic} = f_i(\xi)\sqrt{\frac{p_i}{F_c(\xi)}} - \frac{f_c(\xi)}{F_c(\xi)} \sqrt{p_i f_i(\xi)} , \quad i = 1, 2 \cdots (c-1) .$$

**Proof.** Throughout this proof for convenience set $F_i = F_i(\xi)$ and $f_i = f_i(\xi)$ . Using Taylor's expansion about $\xi$ :

$$F_j(\tilde{w}) = F_j\left(\xi + \frac{\eta}{\sqrt{n}}\right) = F_j + \frac{\eta}{\sqrt{n}} f_j + o\left(\frac{\eta^2}{n}\right), \quad j = 1, 2, \cdots, c ,$$

$$1 - \sum p_j F_j(\tilde{w}) = \frac{1}{2} - \frac{\eta}{\sqrt{n}} \sum_{j=1}^{c} p_j f_j - o\left(\frac{\eta^2}{n}\right) ,$$

and substituting these in (4.2.1) we get

(4.3.2) $$f(m_1, m_2, \ldots, m_c, \tilde{w}) = \left\{ \frac{n(2k)!}{k!\,k!\,2^{2k}} \right\} \left\{ \frac{k!}{\prod_{i=1}^{c} m_i!} \prod_{i=1}^{c} (2p_i F_i)^{m_i} \right\} \times$$

$$\left\{ \prod_{i=1}^{c} \left[1 + \frac{\eta}{\sqrt{n}} \frac{f_i}{F_i} + o\left(\frac{\eta^2}{n}\right)\right]^{m_i} \left(1 - \frac{2\eta}{\sqrt{n}} \sum_{i=1}^{c} p_i f_i - o\left(\frac{\eta^2}{n}\right)\right)^{k} \right\} \times$$

$$\left\{ \sum_{i=1}^{c} p_i f_i\left(\xi + \frac{\eta}{\sqrt{n}}\right) \right\} .$$

$$\equiv \{A_1\} \cdot \{A_2\} \cdot \{A_3\} \cdot \{A_4\} .$$

Note that $\nu_i$ satisfy the relation

$$(4.3.3) \qquad \sum_{i=1}^{c} \nu_i \sqrt{p_i F_i} = 0 .$$

Now consider the region $S$ defined by

$$S = \left\{ (\nu_1, \ldots, \nu_{c-1}, \eta) : a_1 \le \nu_1 \le b_1, \; a_2 \le \nu_2 \le b_2 , \ldots, a_c \le \eta \le b_c \right\} .$$

Using Stirling's approximation for $n!$

$$A_1 \sim \frac{N}{\sqrt{k\pi}} .$$

$A_2$ is independent of $\eta$ and because of convergence of multinomial distribution to multivariate normal distribution, uniformly in $S$

$$A_2 \sim \left[ (2\pi)^{c-1} (2p_c F_c) \prod_{i=1}^{c-1} (2kp_i F_i) \right]^{-1/2} \times$$

$$\exp -\frac{1}{2} \left[ \sum_{i=1}^{c-1} \nu_i^2 \left(1 + \frac{p_i F_i}{p_c F_c}\right) + 2 \sum_{i<j=1}^{c-1} \frac{\nu_i \nu_j \sqrt{p_i p_j F_i F_j}}{p_c F_c} \right] .$$

Now consider

$$\log A_3 = \sum_{i=1}^{c} (\nu_i \sqrt{Np_i F_i} + Np_i F_i) \log\left(1 + \frac{\eta}{\sqrt{N}} \frac{f_i}{F_i} + o\left(\frac{\eta^2}{N}\right)\right) +$$

$$+ \ k \log \left[ 1 - \frac{2\eta}{\sqrt{N}} \sum_{i=1}^{c} p_i f_i - o\left(\frac{\eta^2}{N}\right) \right]$$

Using series expansion for $\log(1+x)$ it follows that, uniformly in S

$$\log A_3 = -\frac{\eta^2}{2}\left[\sum_{i=1}^{c} \frac{p_i f_i^2}{F_i} + 2\left(\sum_{i=1}^{c} p_i f_i\right)^2\right] + \sum_{i=1}^{c} \eta v_i f_i \left(\frac{p_i}{F_i}\right)^{1/2} + o(1).$$

Using the continuity of $f_i$ we have, uniformly in S

$$f(m_1, m_2, \ldots, m_c, \tilde{w}) \sim N\left(\sum_{i=1}^{c} p_i F_i\right)\left[k\pi(2\pi)^{c-1}(2p_c F_c)\prod_{i=1}^{c-1}(2kp_i F_i)\right]^{-1/2} X$$

$$\exp -\frac{1}{2}\left[\sum_{i=1}^{c-1} v_i^2\left(1+\frac{p_i F_i}{F_c}\right) + \eta^2\left\{\sum_{i=1}^{c} \frac{p_i f_i^2}{F_i} + 2\left(\sum_{i=1}^{c} p_i f_i\right)^2\right\}\right.$$

$$\left. + \sum_{i \neq j = 1}^{c-1} v_i v_j \frac{\sqrt{p_i p_j F_i F_j}}{p_c F_c} - 2\eta \sum_{i=1}^{c-1} v_i \left\{ f_i \left(\frac{p_i}{F_i}\right)^{1/2} - \frac{f_c}{F_c}(p_i F_i)^{1/2} \right\}\right].$$

Hence as in Theorem 2.2.1 it follows that

$$P \ a_1 \leq v_1 \leq b_1, \ a_2 \leq v_2 \leq b_2, \ \ldots, \ a_c \leq \eta \leq b_c$$

$$\longrightarrow \int_{a_1}^{b_1}\int_{a_2}^{b_2} \cdots \int_{a_c}^{b_c} f(v_1, v_2, \ldots v_{c-1}, \eta)\, dv_1\, dv_2 \ldots dv_{c-1}\, d\eta,$$

where $f(\nu_1, \ldots, \nu_{c-1}, \eta)$ is the probability density function of the multivariate normal distribution described in the present theorem. $||$

__Corollary__. Under the hypothesis $H_c$, the set $(\nu_1, \ldots, \nu_{c-1})$ and $\eta$ are asymptotically independent.

The following lemma gives the asymptotic joint distribution of $\nu_1, \ldots, \nu_{c-1}$ and $\eta$ under the hypothesis $H_N$ which specifies that

$$F_j(y) = F\left( y + \frac{\theta_j}{\sqrt{N}} \right), \quad j = 1, 2, \ldots, c .$$

__Lemma 4.3.2__. Under the hypothesis $H_N$ the asymptotic joint distribution of $(\nu_1, \ldots, \nu_{c-1}, \eta)$ is c-variate normal distribution given by

$$f(\nu_1, \nu_2, \ldots, \nu_{c-1}, \eta) = \frac{f(\xi_1)}{\pi^{c/2}\, 2^{(c-2)/2}\, p_c^{1/2}} \times$$

$$\exp \frac{-1}{2} \left[ +\eta^2 \text{[illegible]} + \sum_{i=1}^{c-1} \nu_i^2 \left( 1 + \frac{p_i}{p_c} \right) + \sum_{i \neq j = 1}^{c-1} \nu_i \nu_j \frac{\sqrt{p_i p_j}}{p_c} \right] .$$

__Proof__. It is similar to that of Lemma 4.3.1. $||$

Now we are in a position to obtain the limiting distribution of $M$ defined by (4.1.1) under the hypothesis $H_N$ .

__Theorem 4.3.1__. Under the hypothesis $H_N$ the asymptotic distribution of $2M$ is noncentral $\chi^2$ with $(c-1)$ degrees of freedom and noncentrality parameter

$$(4.3.4) \qquad \lambda = 2(F'(\xi))^2 \sum_{j=1}^{c} p_j (\theta_j - \bar{\theta})^2 ,$$

where $\bar{\theta} = \sum_{j=1}^{c} p_j \theta_j$ .

**Proof.** Write

$$u_i = \frac{m_i - k p_i}{\sqrt{N p_i}} = \frac{[m_i - N p_i F_i(\xi)]\sqrt{F_i(\xi)}}{\sqrt{N p_i F_i(\xi)}} - \frac{k p_i - N p_i F_i(\xi)}{\sqrt{N p_i}} , \quad i = 1, \ldots, c .$$

Under the hypothesis $H_N$ using Lemma 4.3.2 it follows that the asymptotic joint distribution of $(u_1, u_2, \ldots, u_c)$ is c-variate normal with means $\mu_i = \theta_i F'(\xi)\sqrt{p_i}$ , and covariance matrix $\Sigma = (\sigma_{ij})$ of rank $(c-1)$ where

$\sigma_{ii} = (1 - p_i)/2$, $i = 1, 2, \ldots, c$, $\sigma_{ij} = -\sqrt{p_i p_j} \,/\, 2$ , $i \neq j = 1, 2, \ldots, c$ .

Hence noting that

$\sum_{j=1}^{c} p_j \sqrt{u_j} = 0$ it follows that the limiting distribution of

$$2M = 2\left[ \sum_{i=1}^{c-1} u_i^2 \left(1 + \frac{p_i}{p_c}\right) + \sum_{i \neq j = 1}^{c-1} u_i u_j \frac{\sqrt{p_i p_j}}{p_c} \right]$$

is $\chi^2_{c-1}(\lambda)$ , where $\lambda$ is given by (4.3.4). ||

## 4.4 Asymptotic Efficiency

Let $F_j(y) = F(y + \theta_j)$, then $H_c$ is true when $\theta_j = 0$. We now find the relative asymptotic efficiency of the c-sample median test with respect to the corresponding parametric test, when $F_j (j = 1, 2, \ldots, c)$ is a normal distribution with mean $\theta_j$ and variance $\sigma^2$. The hypothesis $H_c$ is true if and only if $\rho^2_{Y(X_1, \ldots, X_c)} = 0$, (Olkin, Tate [15]) where $\rho_{Y(X_1, \ldots, X_c)}$ is the multiple correlation coefficient between Y and X. Let R denote the sample multiple correlation coefficient between Y and X. If

$$\bar{U}_{..} = \left(\sum_{i,j} U_{ji}\right)/N \quad , \quad \bar{U}_{j.} = \left(\sum_{i=1}^{n_j} U_{ji}\right)/n_j \quad ,$$

then

$$T^2 = \frac{R^2}{1-R^2} = \frac{\sum_{j=1}^{c} n_j (\bar{U}_{j.} - \bar{U}_{..})^2}{\sum_{j,i} (U_{ji} - \bar{U}_{..})^2 - \sum_{j=1}^{c} (\bar{U}_{j.} - \bar{U}_{..})^2} .$$

Also

$$\rho^2_{Y(X_1, \ldots, X_c)} = \frac{\sum_{j=1}^{c} \left[ (\theta_j - \bar{\theta})^2 p_j \right] / \sigma^2}{1 + \sum_{j=1}^{c} \left[ (\theta_j - \bar{\theta})^2 p_j \right] / \sigma^2} ,$$

where $\bar{\theta} = \sum_{j=1}^{c} p_j \theta_j$.

Following Fisher [16] it is seen that under the hypothesis

$H_N$: $F_j(y) = F\left(x + \frac{\theta_j}{\sqrt{N}}\right)$, $j = 1, 2, \ldots, c$, the asymptotic distribution of $(N - c)\, T^2 / (c - 1)$ is $\chi^2_{c-1}(\lambda')$ distribution where the noncentrality parameter is given by

$$\lambda' = \sum_{j=1}^{c} p_j (\theta_j - \bar{\theta})^2 / \sigma^2 .$$ Also it is proved (Theorem 4.3.1) that the limiting distribution of $2M$ is $\chi^2_{c-1}(\lambda)$, where noncentrality parameter $\lambda$ is given by

$$\lambda = 2[F'(\xi)]^2 \sum_{j=1}^{c} p_j (\theta_j - \bar{\theta})^2 .$$ Following Andrews [14], Hannan [17], since the two test statistics are asymptotically distributed as a noncentral $\chi^2$ variate with the same number of degrees of freedom, the asymptotic relative efficiency is given by the ratio of the two noncentrality parameters, i.e. the efficiency is found to be

$$e(M, R) = 2\,\sigma^2 [F'(\xi)]^2 = 1/\pi$$

## 4.5 Case When $p_1, p_2, \ldots, p_c$ Are Unknown.

In this case we estimate $p_j$ by $\hat{p}_j = n_j/N$, $j = 1, 2, \ldots, c$ and consider the test based on $\hat{M}$ defined by (4.1.2). It is interesting to note that the test of $H_c$ based on $\hat{M}$ is asymptotically distribution free.

**Theorem 4.5.1.** Under the hypothesis $H_c$, $4\hat{M}$ is asymptotically distributed as a $\chi^2$ variable with $c-1$ degrees of freedom.

**Proof.** Write

$$y_j = \frac{m_j - k\hat{p}_j}{\sqrt{n\hat{p}_j}} = (p_j/\hat{p}_j)^{1/2} \frac{(m_j - k\hat{p}_j)}{\sqrt{n p_j}} = (p_j/\hat{p}_j)^{1/2}\, w_j .$$

Let $v = (v_1, v_2, \ldots, v_c)$ and $w = (w_1, w_2, \ldots, w_c)$, then $v = wD$, where $D$ is a diagonal matrix with $\sqrt{p_j} / \sqrt{\hat{p}_j}$ as its diagonal elements. Since $\operatorname{plim} \hat{p}_j = p_j$, it follows that $\operatorname{plim} (\sqrt{p_j} / \sqrt{\hat{p}_j}) = 1$ and hence the matrix $D$ converges in probability (element-wise) to identity matrix. An application of a lemma of Chiang [18, Lemma 1] yields that the vectors $v$ and $w$ have the same limiting distribution. Noting that $w_j = (m_j - kp_j) / \sqrt{(2k+1)p_j} - k(\hat{p}_j - p_j) / \sqrt{(2k+1)p_j}$, it is seen that the asymptotic distribution of $w$ is c-variate normal with zero mean vector and covariance matrix $\Sigma = (\sigma_{ij})$ of rank $c - 1$ with $\sigma_{jj} = (1 - p_j)/4$, $j = 1, 2 \cdots c$, and $\sigma_{ij} = -\sqrt{p_i p_j} / 4$,

$i \neq j = 1, 2 \cdots c$. Noting that $\sum_{j=1}^{c} \sqrt{p_j}\, v_j$ converges in probability to zero as $N \longrightarrow \infty$, the asymptotic distribution of $v_1, v_2, \ldots, v_{c-1}$ is given by

$$f(v_1, \ldots, v_{c-1}) = \frac{1}{(2\pi)^{(c-1)/2} (1/4)^{(c-1)/2} p_c^{1/2}} \times$$

$$\exp -2 \left[ \sum_{i=1}^{c-1} v_i^2 \left(1 + \frac{p_i}{p_c}\right) + 2 \sum_{i \neq j = 1}^{c-1} \frac{v_i v_j \sqrt{p_i p_j}}{p_c} \right]$$

Hence

$$4\hat{M} = 4\left[\sum_{i=1}^{c-1} v_i^2 \left(1 + \frac{\hat{p}_i}{\hat{p}_c}\right) + \sum_{i\neq j=1}^{c-1} v_i v_j \frac{\sqrt{\hat{p}_i \hat{p}_j}}{\hat{p}_c}\right]$$

has the asymptotic distribution stated in the theorem. ||

## 5. Rank Test for Dispersion

Let $Z_1, Z_2, \ldots, Z_N$, where $Z_i = (X_i, Y_i)$, be $N$ independent observations from a bivariate population. We assume $P\{X = 1\} = p$, $P\{X = 0\} = q = (1 - p)$ ; $P(Y \leq y \mid X = j) = F_j(y)$ , $j = 0, 1$ . Let $U_1, U_2, \ldots, U_n$ $(n > 0)$ be those $Y$ observations for which the corresponding $X$ observations are $1$ , and $V_1, \ldots, V_{N-n}$ be the remaining $Y$ observations. Let $r_i$ denote the rank of the ith ordered $U$ observation in the combined sample of $U$'s and $V$'s . For testing the hypothesis $H$: $F_1 = F_0$ against the alternatives that $F_1$ and $F_0$ differ only in the scale parameter, we consider the test based on the statistic,

$$W = \sum_{i=1}^{n} \left(r_i - \frac{N+1}{2}\right)^2 ,$$

which is known to be sensitive for such alternatives. $H$ is rejected if $W$ is either too large or too small.

In Sections 5.1 and 5.2, the mean, and variance of $W$ , and the limiting distribution of $W$ are obtained, when $p$ is known, while in

Section 5.3 we deal with the case when $p$ in unknown.

## 5.1 Mean and Variance of $W$

Write $N^{(r)} = N(N-1) \cdots (N-r+1)$ . Since $n$ is a binomial random variable $b(N, p)$ ,

$$(5.1.1) \qquad E\left[ n^{(r)} (N-n)^{s} \right] = \sum_{n=r}^{N} \frac{N!\, p^{n} q^{N-n}}{(N-r)!\,(N-n-s)!} = p^{r} q^{s} N^{(r+s)} .$$

First we find the mean and variance of $W$ under the hypothesis $H\colon F_1 = F_0$ . It has been proved by Mood [5], that the conditional moments of $W$ for fixed $n$ are,

$$E_p(W|n) = n(N^2-1) / 12 , \quad \sigma_p^2(W|n) = n(N-n)(N+1)(N^2-4) / 180 .$$

Hence, using (5.1.1),

$$(5.1.2) \qquad E_p(W) = E[E_p(W|n)] = Np(N^2-1) / 12 ,$$

$$E[\sigma_p^2(W|n)] = pqN(N^2-1)(N^2-4) / 180 .$$

To find $\sigma_p^2(W)$ we note that $\sigma_p^2(w) = E[\sigma_p^2(W|n)] + \sigma^2[E_p(W|n)]$ . Hence

$$(5.1.3) \qquad \sigma_p^2(W) = pqN(N^2-1)(3N^2-7) / 240 .$$

Let,

$$M_{ij} = \int_{-\infty}^{+\infty} [F_0(y)]^i [F_1(y)]^j dF_1(y) .$$

To obtain $E_p(W)$ under the alternative note that

$$(5.1.4) \qquad W = \sum_{i=1}^{n} r_i^2 - (N+1) \sum_{i=1}^{n} r_i + \frac{n(N+1)^2}{4}$$

and use the following results proved by Sukhatme [ 7],

$$E\left(\sum_{i=1}^{n} r_i \,\middle|\, n\right) = n(N-n)\, M_{10} + [n(n+1)] \,/\, 2 \,,$$

$$E\left(\sum_{i=1}^{n} r_i^2 \,\middle|\, n\right) = 3n(N-n)\, M_{10} + n(N-n)^{(2)} M_{20} + 2n^{(2)}(N-n) M_{11} + \frac{1}{6}\, n(n+1)(2n+1) \,.$$

Also

$$E[n(n+1)] = Npq + N^2p^2 + Np \,,$$

and using (5.1.1),

$$E[n(n+1)(2n+1)] = E[2n^{(3)} + 9n^{(2)} + 6n] = 2N^{(3)}p^3 + 9N^{(2)}p^2 + 6\,N\,p \,,$$

$$(5.1.5) \quad E_p\left[\sum_{i=1}^{n} r_i\right] = E\left[E_p\left(\sum_{i=1}^{n} r_i \,\middle|\, n\right)\right] = N^{(2)}pqM_{10} + \frac{1}{2}\,[Npq + N^2p^2 + Np] \,,$$

$$(5.1.6) \quad E_p\left[\sum_{i=1}^{n} r_i^2\right] = E\left[E_p\left(\sum_{i=1}^{n} r_i^2 \,\middle|\, n\right)\right] = 3pqN^{(2)}M_{10} + pq^2N^{(3)}M_{20}$$

$$+ \; 2p^2qN^{(3)}M_{11} + \frac{1}{6}\,[2N^{(3)}p^3 + 9N^{(2)}p^2 + 6Np] \,.$$

After using (5.1.5), (5.1.6) we get

$$(5.1.7)\quad E_p(W) = E_p\left(\sum_{i=1}^{n} r_i^2\right) - (N+1)\, E_p\left(\sum_{i=1}^{n} r_i\right) + \frac{(N+1)^2}{4} E_p n$$

$$= \frac{p}{4} N(N-1)^2 + \frac{1}{6} p\, N(N-1)(N-2)[12\, pqM_{11} + 6\, q^2 M_{20} - 6\, qM_{10} + 2p^2 - 3p].$$

## 5.2 Asymptotic Distribution of W

We observe that,

$$r_i = 1 + \sum_{j=1}^{N-n} \phi(V_j, U_i) + \sum_{k=1}^{n} \phi(U_k, U_i)$$

$$\text{where } \phi(x, y) = \begin{cases} 1, & \text{if } x < y, \\ 0, & \text{otherwise.} \end{cases}$$

$$r_i^2 = 1 + \left[\sum_{j=1}^{N-n} \phi(V_j, U_i)\right]^2 + \left[\sum_{k=1}^{n} \phi(U_k, U_i)\right]^2$$

$$+ 2 \sum_{j=1}^{N-n} \phi(V_j, U_i) + 2 \sum_{k=1}^{n} \phi(U_k, U_i)$$

$$+ 2 \left[\sum_{j=1}^{N-n} \phi(V_j, U_i)\right]\left[\sum_{k=1}^{n} \phi(U_k, U_i)\right]$$

After defining

$$\phi(u, v, w) = \begin{cases} 1, & \text{if } u < w \text{ and } v < w , \\ 0, & \text{otherwise} , \end{cases}$$

we can write

$$\gamma_i^2 = 1 + 3\sum_{j=1}^{N-n} \phi(V_j, U_i) + 3\sum_{k=1}^{N-n} \phi(U_k, U_i) + \sum_{j \neq k=1}^{N-n} \psi(V_j, V_k, U_i) +$$

$$+ \sum_{j \neq k=1}^{n} \psi(U_j, U_k, U_i) + 2\sum_{k \neq i} \psi(V_j, U_k, U_i) .$$

Observing that

$$\sum_{k=1}^{n} \phi(U_k, U_i) = (i-1),$$

and

$$\sum_{j \neq k=1}^{n} \psi(U_j, U_k, U_i) = (i-1)(i-2) ,$$

we can write from (5.1.4)

$$W = -(N-2) \sum_{j=1}^{N-n} \sum_{i=1}^{n} \phi(V_j, U_i) + \sum_{i=1}^{n} \sum_{j \neq k=1}^{N-n} \psi(V_j, V_k, U_i)$$

$$+ 2\sum_{i \neq k=1}^{n} \sum_{j=1}^{N-n} \psi(V_j, U_k, U_i) + \frac{n(n+1)(2n+1)}{6} + \frac{n^2(n+1)^2}{4} - \frac{(N+1)n(n+1)}{2} .$$

Now define three functions H, K and L as

$$H(Z_i, Z_j) = \begin{cases} 1, & \text{if } X_i = 0 , X_j = 1 \text{ and } Y_i < Y_j , \\ 0, & \text{otherwise.} \end{cases}$$

$$K(Z_i, Z_j, Z_k) = \begin{cases} 1, & \text{if } X_i = 0,\ X_j = 0,\ X_k = 1 \text{ and } Y_i < Y_k,\ Y_j < Y_k, \\ 0, & \text{otherwise.} \end{cases}$$

$$L(Z_i, Z_j, Z_k) = \begin{cases} 1, & \text{if } X_i = 0,\ X_j = 1,\ X_k = 1 \text{ and } Y_i < Y_k,\ Y_j < Y_k, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly,

$$\sum_{j=1}^{(N-n)} \sum_{i=1}^{n} \phi(V_j, U_i) = \sum_{i \neq j=1}^{N} H(Z_j, Z_i),$$

$$\sum_{i=1}^{n} \sum_{j \neq k}^{N-n} \psi(V_j, V_k, U_i) = \sum_{i \neq j \neq k=1}^{N} K(Z_j, Z_k, Z_i),$$

$$\sum_{i \neq k=1}^{n} \sum_{j=1}^{N-n} \psi'(V_j, V_k, U_i) = \sum_{i \neq j \neq k=1}^{N} L(Z_j, Z_k, Z_i);$$

and hence,

$$\text{(5.2.1)} \quad \frac{W}{N^3} = \frac{N(N-1)}{N^2}\left[\bar{U}_N^{(1)} - \bar{U}_N^{(2)} - \bar{U}_N^{(3)}\right] + \frac{1}{12N^3}\left[2n(n+1)(2n+1) + 3n(N+1)^2 - 6(N+1)n(n+1)\right]$$

where $\bar{U}_N^{(1)}$, $\bar{U}_N^{(2)}$, $\bar{U}_N^{(3)}$ are U-statistics defined by,

$$\bar{U}_N^{(1)} = \frac{1}{N(N-1)} \sum_{j \neq i = 1}^{N} H(Z_j, Z_i)$$

$$\bar{U}_N^{(2)} = \frac{1}{N(N-1)(N-2)} \sum_{i \neq j \neq k = 1}^{N} K(Z_j, Z_k, Z_i) ,$$

$$\bar{U}_N^{(3)} = \frac{1}{N(N-1)(N-2)} \sum_{j \neq k \neq i = 1}^{N} L(Z_j, Z_k, Z_i) .$$

__Theorem 5.2.1.__ Let $T = W/N^3$ . The asymptotic distribution of $(T - E_p(T))/\sigma_p(T)$ is $\mathcal{N}(0, 1)$ both under the hypothesis as well as the alternative.

__Proof.__ Observe that the second term of (5.2.1) converges in probability to $p(4p^2 - 6p + 3)/12$ . By Hoeffding's theorem [13, Theorem 7.2] it follows that the asymptotic joint distribution of $\bar{U}_N^{(1)}$ , $\bar{U}_N^{(2)}$ , $\bar{U}_N^{(3)}$ is trivariate normal. The required result follows by an application of a theorem of Cramer [9, p. 254]. ||

5.3 __Case when $p$ is Unknown__

Here we estimate $p$ by $\hat{p} = n/N$ and consider the test based on $[T - E_{\hat{p}}(T)]/[\sigma_{\hat{p}}(T)$ , where $E_{\hat{p}}(T)$ and $\sigma_{\hat{p}}(T)$ are obtained from (5.1.2) and (5.1.3) by replacing $p$ by $\hat{p}$ and $q$ by $\hat{q}$ . It is interesting to note that this test is asymptotically distribution-free.

__Theorem 5.3.1.__ Under the hypothesis $H$: $F_1 = F_0$ , the asymptotic distribution of $[T - E_{\hat{p}}(T)]/\sigma_{\hat{p}}(T)$ is $\mathcal{N}(0, 4/9)$ .

**Proof.** Since $\operatorname{plim} \hat{p} = p$, the limiting distribution of $[T - E_{\hat{p}}(T)]/\sigma_{\hat{p}}(T)$ is the same as that of $[T - E_{\hat{p}}(T)]/\sigma_p(T)$ . Also

$$(5.3.1) \quad \frac{T - E_{\hat{p}}(T)}{\sigma_p(T)} = \frac{T - E_p(T)}{\sigma_p(T)} - \frac{E_{\hat{p}}(T) - E_p(T)}{\sigma_p(T)} \equiv a - b .$$

Note that after using expressions for $E_{\hat{p}}(T)$ , $E_p(T)$ and $\sigma_p(T)$ , $b$ in (5.3.1) can be written as

$$b = \frac{\sqrt{(5/9)N}\,(\hat{p} - p)}{\sigma_p(T)} + c \equiv b_1 + c .$$

where $c$ converges in probability to zero. Note that $a$ and $b_1$ are jointly asymptotically normally distributed with mean vector zero and covariance matrix $\Sigma = (\sigma_{ij})$ with $\sigma_{11} = 1$ , $\sigma_{12} = \sigma_{21} = \sigma_{22} = 5/9$ . Hence the theorem follows. ||

## 6. Two Sample Run Test

As before let $Z_i = (X_i, Y_i)$ , $i = 1, 2, \ldots, N$ be $N$ independent observations from a bivariate population where $X$ assumes only two values 1 and zero with probabilities $p$ and $1 - p = q$ respectively; and let $P\{Y \leq y \mid X = j\} = F_j(y)$ $(j = 0, 1)$ . Let $U_1, U_2, \ldots, U_n$ $(n > 0)$ be those $Y$ observations fro which the corresponding $X$ observations are one, and $V_1, \ldots, V_{N-n}$ be the remaining $Y$ observations. For testing the hypothesis $H$: $F_1(y) = F_0(y)$ , combine the two samples of $U$'s and $V$'s and arrange them in the order of magnitude. Here we consider the test based on $d$ , the total number of runs of $U$'s and $V$'s . The hypothesis $H$ is rejected if $d$ is too small. Mood [19] has given the exact sampling distribution of $d$ under the hypothesis $H$ when $p$ is known and further proved that under the hypothesis $H$, the asymptotic distribution of $[d - 2Npq]/[2\sqrt{Npq(1 - 3pq)}]$ is $\mathcal{N}(0, 1)$. These results are obtained by other authors, see, for example, Wishart and Hirshfeld [20], Iyer [21].

Here we consider the case when $p$ is unknown. Estimate $p$ by its usual estimator $\hat{p} = n/N$ and consider the test based on $[d - 2N\hat{p}\hat{q}]/[2\sqrt{N\hat{p}\hat{q}(1 - 3\hat{p}\hat{q})}]$. It is proved in the following theorem that this test is not asymptotically distribution-free in that the limiting distribution of the statistic depends on $p$.

<u>Theorem 6.1</u>. Under the hypothesis $H$: $F_1 = F_0$ the asymptotic distribution of $(d - 2N\hat{p}\hat{q})/2[\sqrt{N\hat{p}\hat{q}(1 - 3\hat{p}\hat{q})}]$ is normal with mean zero and variance $1 - (1 - 2p)^2/(1 - 3pq)$.

<u>Proof</u>. As in Theorem 3.6.1 the asymptotic distribution of $(d - 2N\hat{p}\hat{q})/2\sqrt{N\hat{p}\hat{q}(1 - 3\hat{p}\hat{q})}$ is the same as that of $(d - 2N\hat{p}\hat{q})/2\sqrt{Npq(1 - 3pq)}$. Since $\hat{p}\hat{q} - pq = (\hat{p} - p)(1 - 2p) - (\hat{p} - p)^2$ we can write

$$(6.1) \quad \frac{d - 2N\hat{p}\hat{q}}{2\sqrt{Npq(1-3pq)}} = \frac{d - Npq}{2\sqrt{Npq(1-3pq)}} - \frac{\sqrt{N}(\hat{p}-p)(1-2p)}{\sqrt{pq(1-3pq)}} + \frac{\sqrt{N}(\hat{p}-p)^2}{\sqrt{pq(1-3pq)}}$$

It can be shown that the asymptotic joint distribution of the first two terms in the above expression is $\mathcal{N}(0, \Sigma)$ with covariance matrix $\Sigma = (\sigma_{ij})$ where $\sigma_{11} = 1$, $\sigma_{12} = \sigma_{21} = \sigma_{22} = (1 - 2p)^2/(1 - 3pq)$. Also noting that the 3rd term in (6.1) converges in probability to zero the required theorem follows. ||

# Part II

# ASYMPTOTIC THEORY OF MODIFIED CRAMÉR-SMIRNOV TEST STATISTICS

## 7. Introduction

Let $X_1, \ldots, X_n$ be $n$ independent observations (random variables) from a population with continuous distribution function $G(x)$. For testing the hypothesis $H_0$: $G(x) = F(x)$ where $F(x)$ is some specified distribution function the following test was proposed by Cramér [1], Smirnov [2] and Von Mises [3]. The test statistic $\omega_n^2$ is defined as

$$\omega_n^2 = n \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 dF(x) ,$$

where $F_n(x)$ denotes the empirical distribution function of the sample i.e. $F_n(x) = \nu/n$, $\nu$ being the number of $X_i$ $(i = 1, 2, \ldots, n)$ that are less than $x$, $-\infty < x < +\infty$; and the hypothesis $H_0$ is rejected for large values of $\omega_n^2$. Properties of this test have been studied by various authors. Cramér in [4] suggested the idea of extending the theory of $\omega_n^2$ test to the case when the distribution function $F(x)$ is not completely specified, but depends on certain parameters that must be estimated from the sample. This extension was investigated by Darling [5] in the case when $F(x)$ depends on one parameter. He considered the following problem. Let $I$ be an open interval on the real line $R^1$ and assume that for every point $\theta \in I$, $F(x, \theta)$ is a distribution function. For testing the hypothesis $H_1$: $G(x) = F(x, \theta)$ , where the functional form of $F$ is known but the parameter $\theta$ is unknown, the modified $\omega_n^2$ criterion is defined as

$$C_{1n}^2 = n\int_{-\infty}^{+\infty}[F_n(x) - F(x, \hat{\theta}_n)]^2 dF(x, \hat{\theta}_n) ,$$

where $\hat{\theta}_n$ is an estimate of $\theta$ obtained from the sample. The hypothesis $H_1$ is rejected for large values of $C_{1n}^2$ . Under certain regularity conditions the asymptotic distribution of $C_{1n}^2$ is obtained in [5]. The limiting distribution depends on the properties of the estimator $\hat{\theta}_n$ .

Now assume that $I$ is an open set in $R^2$ , the two dimensional Euclidean space, and for every point $\theta = (\theta_1, \theta_2) \in I$ , $F(x, \theta)$ is a distribution function. Let $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$ be an estimate of $\theta$ . For testing the hypothesis $H$: $G(x) = F(x, \theta)$ for some unspecified $\theta \in I$ consider the test based on the statistic

$$C_n^2 = n\int_{-\infty}^{+\infty}[F_n(x) - F(x, \hat{\theta}_n)]^2 dF(x, \hat{\theta}_n) .$$

The hypothesis $H$ is rejected if $C_n^2$ is sufficiently large. Kac, Kiefer and Wolfowitz [6] considered the modified Cramér-Smirnov test based on $C_n^2$ when $F(x, \theta)$ is a normal distribution $N(x, \mu, \sigma^2)$ where both the mean $\mu$ and the variance $\sigma^2$ are unknown. Using the sample mean and the sample variance as estimates of $\mu$ and $\sigma^2$ they derived the asymptotic distribution of the test criterion. The methods used in the derivation do not seem to be general enough to obtain the limiting distribution when $F(x, \theta)$ is any arbitrary distribution function.

The object of this paper is to investigate the limiting distribution of $C_n^2$ when $F(x, \theta)$ is an arbitrary distribution function involving two unknown parameters and satisfying certain regularity conditions. As in one parameter case it will be seen that the asymptotic distribution of $C_n^2$ does depend on the properties of the estimators $\hat{\theta}_{1n}, \hat{\theta}_{2n}$. The limiting distribution of $C_n^2$ is derived by suitably combining the techniques of Darling and those of Kac, Kiefer and Wolfowitz [6].

We also study the modification of k-sample Cramér-Smirnow test for testing the hypothesis of goodness of fit. The k-sample problem is as follows. Let $n_j (j = 1, 2, \ldots, k)$ be fixed positive integers; and $X_{ji} (i = 1, 2, \ldots, n_j ; j = 1, \ldots, k)$ be independent random variables having unknown continuous distribution functions $G_j(x)$. Let $I$ be an open interval in $R^1$ so that for every $\theta \in I$, $F(x, \theta)$ is a distribution function. For testing the hypothesis $H_0$: $G_1(x) = G_2(x) = \cdots = G_k(x) = F(x, \theta_0)$, for some specified $\theta_0 \in I$, Kiefer [7] has considered various tests particularly k-sample Cramér-Smirnov test. The test statistic is defined as

$$\omega_n'^2 = \int_{-\infty}^{+\infty} \sum_{j=1}^{k} n_j \left[F_{n_j}^{(j)}(x) - F(x, \theta_0)\right]^2 dF(x, \theta_0),$$

where $n$ stands for the vector $n = (n_1, \ldots, n_k)$, and $F_{n_j}^{(j)}(x)$ is the empirical distribution function of the $j$th sample, that is $F_{n_j}^{(j)}(x) = (1/n_j)$ [number of $X_{ji} < x$, $i = 1, 2, \ldots, n_j$]. The hypothesis is rejected for large values of $\omega_n'^2$. Kiefer has obtained the limiting distribution of $\omega_n'^2$ under the hypothesis $H_0$ and has also tabulated it.

In this paper we consider the problem of testing the hypothesis $H_k$: $G_1(x) = \cdots = G_k(x) = F(x, \theta)$ , when the functional form of $F$ is known but $\theta \in I$ is unknown. To test the hypothesis $H_k$, the k-sample Cramér-Smirnov test statistic is modified as

$$C_n'^2 = \int_{-\infty}^{\infty} \sum_{j=1}^{k} n_j \left[ F_{n_j}^{(j)}(x) - F(x, \hat{\theta}_N) \right]^2 dF(x, \hat{\theta}_N),$$

where $N = \sum_{j=1}^{k} n_j$ , and $\hat{\theta}_N$ is an estimate of $\theta$ obtained by pooling together all the $k$ samples. The hypothesis $H_k$ is rejected if $C_n'^2$ is sufficiently large. Under certain regularity condition the limiting distribution of $C_n'^2$ is obtained when the hypothesis $H_k$ is true. As in the case of one sample problem the asymptotic distribution depends on the properties of the estimator $\hat{\theta}_N$ . These results can be extended to the case when the distribution function $F$ involves two parameters $\theta_1$, $\theta_2$ by using methods similar to those employed in one sample problem.

In Sections 8 and 9 we investigate the limiting distribution of the modified Cramér-Smirnov test statistic $C_n^2$ under the hypothesis $H$ in the case of one sample problem. Section 8 gives the asymptotic distribution of $C_n^2$ when the estimators $\hat{\theta}_1$, $\hat{\theta}_2$ are superefficient. In Section 9 the asymptotic distribution of $C_n^2$ is derived when $\hat{\theta}_1$, $\hat{\theta}_2$ are jointly efficient in the sense of Cramér [4]. The characteristic function of the limiting distribution is the Fredholm determinant of a symmetric positive definite kernel of a particular form. Theorems 9.5.1 and 9.5.2 give methods of obtaining the Fredholm determinant associated with such kernels. In Section 9.6 we study some properties of $C_n^2$ test and consider some consequences of the theory developed. In Section 10 we study the k-sample Cramér-Smirnov test in parametric case and investigate its asymptotic distribution when the hypothesis $H_k$ is true.

## 8. The Cramér-Smirnov Test in the Two-Parameter Case.

8.1 Let $X_1, X_2, \ldots, X_n$ be $n$ independent observations from a continuous distribution function $G(x)$. Assume that for every point $\theta = (\theta_1, \theta_2)$ belonging to an open interval $I$ in $R^2$, $F(x, \theta)$ is an absolutely continuous distribution function. For testing the hypothesis $H$: $G(x) = F(x, \theta)$ where the functional form of $F$ is known but $\theta$ is unspecified, the modified Cramér-Smirnov test criterion is defined as

$$C_n^2 = n\int_{-\infty}^{+\infty} [F_n(x) - F(x, \hat{\theta}_n)]^2 \, dF(x, \hat{\theta}_n) \tag{8.1.1}$$

where $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$ is an estimate of $\theta$ obtained from the sample. The hypothesis $H$ is rejected if $C_n^2$ is sufficiently large.

In the present section we consider the problem of finding the asymptotic distribution of $C_n^2$ when $\hat{\theta}_{1n}, \hat{\theta}_{2n}$ are superefficient estimators and also discuss the case when $\hat{\theta}_{1n}, \hat{\theta}_{2n}$ are regular estimators.

### 8.2 Case when $\hat{\theta}_1$ and $\hat{\theta}_2$ are Superefficient Estimators.

Henceforth for simplicity we write $\hat{\theta}_{1n}$ as $\hat{\theta}_1$ and $\hat{\theta}_{2n}$ as $\hat{\theta}_2$. Suppose that the hypothesis $H$ is true. Let $\theta$ denote the true unknown parameter vector, and $f(x, \theta)$ be the probability density function corresponding to $F(x, \theta)$. $\omega_n^2$ is defined as

(8.2.1) $$\omega_n^2 = n\int_{-\infty}^{+\infty}[F_n(x) - F(x,\theta)]^2 dF(x,\theta) .$$

Let $X_1', X_2', \ldots, X_n'$ be a rearrangement of the sample $X_1, X_2, \ldots, X_n$ so that $X_1' < X_2' < \ldots < X_n'$. Then $\omega_n^2$ and $C_n^2$ can be written as, see [4]

(8.2.2) $$\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(X_i',\theta) - \frac{(2i-1)}{2n}\right]^2 ,$$

(8.2.3) $$C_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(X_i',\hat{\theta}_n) - \frac{(2i-1)}{2n}\right]^2$$

<u>Theorem 8.2.1</u>. Assume that $\hat{\theta}_n$ and $F(x, \theta)$ satisfy:

(i) $\lim\limits_{n \to \infty} n\, E(\hat{\theta}_i - \theta_i)^2 = 0$, $i = 1, 2$.

(ii) For $\theta, \theta' \in I$,

$|F(x, \theta) - F(x, \theta')| < A(x)\, \delta(\theta, \theta')$,

where $\delta(\theta, \theta') = [(\theta_1 - \theta_1')^2 + (\theta_2 - \theta_2')^2]^{1/2}$, and $P\left(A^2(x) > A_0\right) = 0$ for some $A_0 < \infty$, where probability is according to the true distribution $F(x, \theta)$. Then $C_n^2 = \omega_n^2 + \delta_n$, where $\operatorname*{plim}\limits_{n \to \infty} \delta_n = 0$.

Proof. This theorem is a direct analogue of Theorem 2.1 of [5] and can be proved in a similar manner. //

Remark. When conditions (i) and (ii) of Theorem 8.2.1 are satisfied, the asymptotic distribution of $C_n^2$ and $\omega_n^2$ are the same.

## 8.3 Case when $\hat{\theta}_1$, $\hat{\theta}_2$ are Regular Estimators.

In general, condition (i) of Theorem 8.2.1 is not satisfied, so now we consider the case of regular estimation, Cramér [4, p. 479], where $\mathrm{Var}(\hat{\theta}_i) \geq A_i/n$, $(i = 1, 2)$ for some positive $A_i$. In many cases the estimates $\hat{\theta}_i$ are such that $\operatorname*{plim}_{n \to \infty} n^{1/2 - \delta}(\hat{\theta}_i - \theta_i) = 0$ for some $\delta$ such that $\frac{1}{2} > \delta > 0$. The following lemma which is a direct extension of Lemma 3.1 of [5] treats such cases.

Lemma 8.3.1. If

(i) for $\frac{1}{2} > \delta > 0$ $\quad \operatorname*{plim}_{n \to \infty} n^{1/2 - \delta}(\hat{\theta}_i - \theta_i) = 0$, $i = 1, 2$;

for almost all $x$

(ii) $\left| \frac{\partial^2}{\partial \theta_i^2} F(x, \theta) \right| < m_i(x)$,

(iii) $\left| \frac{\partial^2}{\partial \theta_1 \partial \theta_2} F(x, \theta) \right| < m_{12}(x)$,

(iv) $\left| \frac{\partial}{\partial \theta_i} f(x, \theta) \right| < h_i(x)$, $i = 1, 2$, where the functions $m_1(x)$, $m_2(x)$, $m_{12}(x)$, $h_1(x)$, $h_2(x)$ are square integrable, independent of $\theta$ and do not depend on the exceptional set. Then,

$$C_n^2 = C_n^{*2} + \delta_n, \tag{8.3.1}$$

where

$$C_n^{*2} = n\int_{-\infty}^{+\infty} \left[F_n(x) - F(x,\theta) - \sum_{i=1}^{2} \frac{\partial}{\partial\theta_i} F(x,\theta)\right]^2 f(x,\theta)\,dx \tag{8.3.2}$$

and $\operatorname*{plim}_{n \to \infty} \delta_n = 0$ .

<u>Proof</u>. Expand $F(x, \theta)$ and $f(x, \theta)$ in a Taylor's series about the true value $\theta$ :

$$F(x,\hat{\theta}_n) = F(x,\theta) + \sum_{i=1}^{2} (\hat{\theta}_i - \theta_i) \frac{\partial}{\partial\theta_i} F(x,\theta)$$

$$+ \frac{1}{2}\left[\sum_{i=1}^{2} (\hat{\theta}_i - \theta_i) q_i m_i(x) + 2(\hat{\theta}_1 - \theta_1)(\hat{\theta}_2 - \theta_2) q_{12} m_{12}(x)\right];$$

where $|q_i| < 1$ , $|q_{12}| < 1$ ;

$$f(x,\hat{\theta}_n) = f(x,\theta) + \sum_{i=1}^{2} (\hat{\theta}_i - \theta_i) \Delta_i h_i(x), \quad |\Delta_i| < 1, \quad i = 1, 2 .$$

Substitution of these expressions in (8.1.1) yields

$$(8.3.3) \quad C_n^2 = n\int_{-\infty}^{+\infty}\Big[F_n(x)-F(x,\theta)-\sum_{i=1}^{2}(\hat{\theta}_i-\theta_i)\frac{\partial}{\partial\theta_i}F(x,\theta)\Big]^2 f(x,\theta)\,dx$$

$$+\frac{n}{4}\int_{-\infty}^{+\infty}\Big[(\hat{\theta}_i-\theta_i)q_i m_i(x)+2(\hat{\theta}_1-\theta_1)(\hat{\theta}_2-\theta_2)q_{12}m_{12}(x)\Big]^2 f(x,\theta)\,dx$$

$$-n\int_{-\infty}^{+\infty}\Big[F_n(x)-F(x,\theta)-\sum_{i=1}^{2}(\hat{\theta}_i-\theta_i)\frac{\partial}{\partial\theta_i}F(x,\theta)\Big]\times$$

$$\Big[\sum_{i=1}^{2}(\hat{\theta}_i-\theta_i)^2 q_i m_i(x)+2(\hat{\theta}_1-\theta_1)(\hat{\theta}_2-\theta_2)q_{12}m_{12}(x)\Big]f(x,\theta)\,dx$$

$$+n\int_{-\infty}^{+\infty}\Big\{\Big[F_n(x)-F(x,\theta)-\sum_{i=1}^{2}(\hat{\theta}_i-\theta_i)\frac{\partial}{\partial\theta_i}F(x,\theta)-$$

$$-\frac{1}{2}\Big(\sum_{i=1}^{2}(\hat{\theta}_i-\theta_i)^2 q_i m_i(x)+2(\hat{\theta}_1-\theta_1)(\hat{\theta}_2-\theta_2)q_{12}m_{12}(x)\Big)\cdot$$

$$\Big[\sum_{i=1}^{2}(\hat{\theta}_i-\theta_i)\Delta_i h_i(x)\Big]\Big]\Big\}\,dx$$

Using the assumptions (i) - (iv) and that $\sup_x \; n|F_n(x) - F(x, \theta)|$ is bounded in probability, Kolmogorov [8], we find that each term except the first one in (8.3.3) tends in probability to zero as $n \longrightarrow \infty$. Hence the lemma follows. ||

Thus by Lemma 8.3.1 the problem of finding the asymptotic distribution of $C_n^2$ is equivalent to finding that of $C_n^{*2}$.

Now consider some transformations which are basic in the following work. Let

$$(8.3.4) \qquad u = F(x, \theta) , \quad u_j = F(X_j, \theta) \qquad j = 1, 2, \ldots, n .$$

By this transformation $x$ is defined implicitly as a function of $u$ and $\theta$, except possibly at a denumerable set of values of $u$, at which $x$ can be defined arbitrarily so as to make the function monotone non-decreasing. Define

$$(8.3.5) \qquad g_i(u) = \frac{\partial}{\partial \theta_i} F(x, \theta) , \quad 0 \le u \le 1 ; \quad i = 1, 2 ;$$

and the function $\psi_t(x)$ as

$$\psi_t(x) = \begin{cases} 1 , & \text{if } x < t , \\ 0 , & \text{if } x \geq t . \end{cases}$$

Then we can write

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \psi_x(X_j) = \frac{1}{n} \sum_{j=1}^{n} \psi_u(u_j) .$$

where u's are defined by (8.3.4) .

If $G_n(u) = (1/n)$(number of $u_j$'s less than $u$) and

(8.3.6) $$Z_n(u) = \sqrt{n}\,[G_n(u) - u] = \sqrt{n}\left[\frac{1}{n}\sum_{j=1}^{n}\psi_u(u_j) - u\right]$$

then using (8.3.1), $c_n^2$ can be written as

(8.3.7) $$c_n^2 = \int_0^1 Y_n^2(u)\,du + \delta_n ,$$

where

(8.3.8) $$Y_n(u) = Z_n(u) - \sum_{i=1}^{2}\sqrt{n}\,(\hat{\theta}_i - \theta_i)\,g_i(u) ;$$

and $\operatorname{plim}_{n \to \infty} \delta_n = 0$ .

The limiting form of the stochastic process $Y_n(u)$ defined by (8.3.8), required to obtain the asymptotic distribution of $c_n^2$ is given by the Lemma 8.3.2 below. This is an extension of Lemma 3.2 of [5] to the present case. Also note that Lemma 3.2 of [5] is proved under somewhat different conditions than those of the following lemma. For the time being consider the one parameter case studied by Darling. After writing $E\,Z_n(u)\sqrt{n}(\hat{\theta}_n - \theta)$ in a suitable form Darling arrived at the following two conditions. (Conditions (4) and (6) of Lemma 3.2 of [5]).

1) $\lim_{n \to \infty} n\,E(\hat{\theta}_n - \theta) = 0$ , i.e. $\hat{\theta}_n$ is "weakly unbiased".

2) $\lim_{n \to \infty} n\, u\, E \left\{ (\hat{\theta}_n - \theta) \,\middle|\, F(X_1, \theta) < u \right\} = h(u)$ , $0 < u < 1$ , and

$h(0) = h(1) = 0$ .

Instead of assuming the above two conditions for each of the estimators $\hat{\theta}_i$ we make assumption (iv) of the following lemma. There is an example of a distribution function $F(x, \theta)$ for which $\hat{\theta}$ is not weakly unbiased but at the same time $\lim_{n \to \infty} E\, Z_n(u) \sqrt{n}\,(\hat{\theta} - \theta) = h(u)$ exists and has the required properties. It will be seen in Section 9.6 that for the normal distribution $N(x, \mu, \sigma^2)$ the estimate

$s^2 = (1/n) \sum_{i=1}^{n} (x_i - \bar{x})^2$ for $\sigma^2$ is not weakly unbiased but at the same time $\lim_{n \to \infty} E\, Z_n(u) \sqrt{n}\,(s^2 - \sigma^2)$ exists.

<u>Lemma 8.3.2.</u> If

(i) $C_r^2 = \int_0^1 Y_n^2(u)\,du + \delta_n$ , where $\operatorname{plim}_{n \to \infty} \delta_n = 0$ , (i.e. we make the assumptions (i) - (iv) of Lemma 8.3.1)

(ii) $\sqrt{n}\,(\hat{\theta}_i - \theta_i)$ is a sum of independently and identically distributed random variables,

(iii) the asymptotic joint distribution of $(\sqrt{n}\,(\hat{\theta}_1 - \theta_1), \sqrt{n}\,(\hat{\theta}_2 - \theta_2))$ is normal with mean zero and nonsingular covariance matrix $\Sigma = (\sigma_{ij})$ ,

(iv) $\lim_{n \to \infty} E\, Z_n(u) \sqrt{n}\,(\hat{\theta}_i - \theta_i) = h_i(u)$ , $0 < u < 1$ , $h_i(0) = h_i(1) = 0$ , $i = 1, 2$ ,

Then $Y_n(u)$ converges in distribution to a Gaussian process $Y(u)$ , with mean zero and covariance function $\rho(u, v)$ given by

$$(8.3.9) \quad \rho(u, v) = \min(u, v) - uv - g_1(u)h_1(v) - g_1(v)h_1(u) - g_2(u)h_2(v) - g_2(v)h_2(u) + \sum_{i, j=1}^{2} \sigma_{ij} g_i(u) g_j(v) , \quad 0 \leq u, v \leq 1 .$$

Proof. The stochastic process $Z_n(u)$ converges in distribution to a Gaussian process which has mean zero and covariance function

$$(8.3.10) \quad K(u, v) = \min(u, v) - uv, \quad 0 \leq u, v \leq 1 ,$$

see for example [9]. Under the assumption (iii) the asymptotic distribution of $\sum_{i=1}^{2} \sqrt{n}\,(\hat{\theta}_i - \theta_i) g_i(u)$ is normal with mean zero and variance $\sum_{i, j=1}^{2} \sigma_{ij} g_i(u) g_j(v)$ . By multidimensional central limit theorem it follows that $Y_n(u)$ given by (8.3.8) converges in distribution to a Gaussian process with mean zero. To find the covariance function we have $\rho_n(u, v) = E(Y_n(u) Y_n(v))$

$$= E(Z_n(u)Z_n(v)) - E\left(Z_n(u) \sum_{i=1}^{2} \sqrt{n}\,(\hat{\theta}_i - \theta_i) g_i(v)\right)$$

$$- E\left(Z_n(v) \sum_{i=1}^{2} \sqrt{n}\,(\hat{\theta}_i - \theta_i) g_i(u)\right)$$

$$+ E\left(\sum_{i=1}^{2} \sqrt{n}\,(\hat{\theta}_i - \theta_i) g_i(u)\right)\left(\sum_{i=1}^{2} \sqrt{n}\,(\hat{\theta}_i - \theta_i) g_i(v)\right) .$$

Under the assumptions (i) - (iv) as $n \to \infty$, $\rho_n(u, v)$ tends to $\rho(u, v)$ given by (8.3.9) and the lemma follows. ||

## 9. Limiting Distribution of $C_n^2$ - Case of Efficient Estimators

9.1 In this Section we obtain the limiting distributions of $C_n^2$ defined by (8.1.1) when the estimators $\hat{\theta}_1$, $\hat{\theta}_2$ are regular, jointly efficient (or asymptotically jointly efficient) in the sense defined by Cramér [4, pp. 490-495]. It will be seen in Section 9.3 that the asymptotic distribution of $C_n^2$ is the distribution of the random variable

$C^2 = \int_0^1 Y^2(u)\, du$, where $Y(u)$ is a Gaussian process with mean zero

and covariance function $\rho(u, v)$ defined by (9.3.1). Section 9.5 gives two methods of finding the Fredholm determinant (F.D.) of the kernel $\rho(x, y)$ which is required to obtain the characteristic function of the limiting distribution. Lastly Section 9.6 deals with some properties of $C_n^2$ test and derives the results of [6] as a special case of the results given in this section

## 9.2 Case of Efficient Estimators

Following Cramér if we make a transformation from $(x_1, x_2, \ldots, x_n) \to (\theta_1, \theta_2, \xi_1, \ldots, \xi_{n-2})$ we have

$$\prod_{j=1}^{n} f(x_j, \theta)dx_j = g(\hat{\theta}_1, \hat{\theta}_2; \theta)h(\xi_1, \ldots, \xi_{n-2} | \hat{\theta}_1, \hat{\theta}_2)d\hat{\theta}_1 d\hat{\theta}_2 d\xi_1 \cdots d\xi_{n-2}$$

If $\hat{\theta}_1, \hat{\theta}_2$ are regular efficient estimators then $h(\xi_1, \ldots, \xi_{n-2}|\hat{\theta}_1, \hat{\theta}_2)$ is independent of $\theta_1, \theta_2$ and $g$ is such that

$$\text{(9.2.1)} \quad \frac{\partial}{\partial\theta_1} \log g = k_{11}(\hat{\theta}_1 - \theta_1) + k_{12}(\hat{\theta}_2 - \theta_2)$$

$$\text{(9.2.2)} \quad \frac{\partial}{\partial\theta_2} \log g = k_{21}(\hat{\theta}_1 - \theta_1) + k_{22}(\hat{\theta}_2 - \theta_2)$$

where $k_{ij}$ may depend on $\theta_1, \theta_2$ but are independent of $\hat{\theta}_1, \hat{\theta}_2$.
From (9.2.1), (9.2.2) differentiating each of them w.r.t. $\theta_1, \theta_2$ and taking expectations we obtain

$$\text{(9.2.3)} \quad k_{11} = nE\left(\frac{\partial}{\partial\theta_1} \log f(X,\theta)\right)^2, \quad k_{22} = nE\left(\frac{\partial}{\partial\theta_2} \log f(X,\theta)\right)^2$$

$$k_{12} = k_{21} = nE\left(\frac{\partial}{\partial\theta_1} \log f(X,\theta) \cdot \frac{\partial}{\partial\theta_2} \log f(X,\theta)\right).$$

Multiply (9.2.1) and (9.2.2) by $(\hat{\theta}_1 - \theta_1)$ , $(\hat{\theta}_2 - \theta_2)$ respectively and take expectations to obtain

(9.2.4) $\quad k_{11}\mathrm{Var}(\hat{\theta}_1) + k_{12}\mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2) = 1$

(9.2.5) $\quad k_{21}\mathrm{cov}(\hat{\theta}_2, \hat{\theta}_1) + k_{22}\mathrm{Var}(\hat{\theta}_2) = 1$ .

The covariance matrix $\Sigma = (\sigma_{ij})$ of $(\hat{\theta}_1, \hat{\theta}_2)$ is nonsingular if and only if

$$r^2 = \frac{\left[E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\cdot\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)\right]^2}{E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\right)^2 E\left(\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)^2} \neq 1 .$$

If $r^2 = 1$ , covariance matrix of $\hat{\theta}_1, \hat{\theta}_2$ is singular and $\hat{\theta}_1$ and $\hat{\theta}_2$ are linearly dependent. As they are unbiased estimates of $\theta_1$ and $\theta_2$ it follows that $\theta_1$ is a linear function of $\theta_2$ and then we are essnetially in a single parameter case. So henceforth we assume that $r^2 \neq 1$ . Now define

(9.2.6) $$\gamma = -\frac{E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\cdot\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)}{\left[E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\right)^2 \cdot E\left(\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)^2\right]^{1/2}} ,$$

(9.2.7) $$\sigma_{11} = \sigma_1^2 = \frac{1}{(1-\gamma^2)\,E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\right)^2} ,$$

$$\sigma_{22} = \sigma_2^{\,2} = \frac{1}{(1-r^2)\,E\left(\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)^2}\ , \qquad \sigma_{12} = \sigma_{21} = r\sigma_1\sigma_2 .$$

With this notation from (9.2.1), (9.2.2) we have

(9.2.8) $$(\hat{\theta}_1-\theta_1) = \frac{\sigma_1^{\,2}}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}\log f(X_j,\theta) + \frac{\sigma_{12}}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_2}\log f(X_j,\theta)\ ,$$

(9.2.9) $$(\hat{\theta}_2-\theta_2) = \frac{\sigma_2^{\,2}}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_2}\log f(X_j,\theta) + \frac{\sigma_{12}}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}\log f(X_j,\theta) .$$

For efficient estimators conditions (i) and (iv) of Lemma 8.3.1 are satisfied by assumptions of Cramér and we further assume that (ii) and (iii) hold. Now let

(9.2.10) $\gamma_i(u) = \sigma_i g_i(u)$ where $g_i(u)$ is defined by (8.3.5) .

The limiting form of the process $Y_n(u)$ given by (8.3.8) is obtained in the following lemma when $\hat{\theta}_1$, $\hat{\theta}_2$ are efficient estimators.

<u>Lemma 9.2.1</u>. If $\hat{\theta}_1$, $\hat{\theta}_2$ are regular, unbiased, jointly efficient estimators of $\theta_1$, $\theta_2$ , then the process $Y_n(u)$ given by (8.3.8) has mean zero and covariance function

$$(9.2.11) \qquad \rho(u, v) = \min(u, v) - uv - \varphi_1(u)\varphi_1(v) - \varphi_2(u)\varphi_2(v)$$

$$- r\varphi_1(u)\varphi_2(v) - r\varphi_1(v)\varphi_2(u),$$

where $\varphi_i(u)$ are defined by (9.2.10) and have the following properties.

(1) $\int_0^1 [\varphi_i'(u)]^2 du = 1/(1-r^2)$ (2) $\int_0^1 \varphi_1'(u)\varphi_2'(u) du = -r/(1-r^2)$.

Proof. From (9.2.8) and (9.2.9) it is seen that condition (ii) of Lemma 8.3.2 is satisfied. Since the asymptotic joint distribution of $\sqrt{n}\,(\hat{\theta}_1 - \theta_1)$, $\sqrt{n}\,(\hat{\theta}_2 - \theta_2)$ is normal $N(0, \Sigma)$ where the covariance matrix $\Sigma = (\sigma_{ij})$ is given by (9.2.7), the condition (iii) of Lemma 8.3.2 is satisfied. Let $h_{in}(u) = E(Z_n(u)\sqrt{n}\,(\hat{\theta}_i - \theta_i))$. Then proceeding as in Lemma 3.3 of [5] we can show that $h_{in}'(u) = n\,E\{(\hat{\theta}_i - \theta_i)|F(X_1, \theta) = u\} - n\,E\,(\hat{\theta}_i - \theta_i)$. As $\hat{\theta}_i$ is an unbiased estimator of $\theta_i$, $n\,E\,(\hat{\theta}_i - \theta_i) = 0$ and hence in the present case using (9.2.8) $h_{in}'(u)$ can be written as

$$h_{in}'(u) = nE\left\{\left(\frac{\sigma_1^2}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}\log f(X_j;\theta) + \frac{\sigma_{12}}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_2}\log f(X_j;\theta)\right)\Big| F(X_1,\theta)=u\right\}.$$

Since $X_1, X_2, \ldots, X_n$ are independently and identically distributed,

$$E\left(\frac{\partial}{\partial\theta_i}\log f(X_j,\theta)\Big| F(X_1,\theta)=u\right) = E\,\frac{\partial}{\partial\theta_i}\log f(X_j,\theta),$$

for $j = 2, 3, \ldots, n$ and $i = 1, 2$. Also as $F(X_1, \theta) = u$ is a condition on $X_1$,

$$E\left(\frac{\partial}{\partial\theta_i}\log f(X_1,\theta)\,\middle|\,F(X_1,\theta)=u\right)=\frac{\partial}{\partial\theta_i}\log f(X,\theta),\quad i=1,2\,.$$

Hence

$$h'_{1n}(u)=\sigma_1^2\frac{\partial}{\partial\theta_1}\log f(X,\theta)+\sigma_{12}\frac{\partial}{\partial\theta_2}\log f(X,\theta),$$

Similarly,

$$h'_{2n}(u)=\sigma_2^2\frac{\partial}{\partial\theta_2}\log f(X,\theta)+\sigma_{12}\frac{\partial}{\partial\theta_1}\log f(X,\theta)\,.$$

As $h'_{in}(U)$ $(i=1,2)$ is independent of $n$ we omit the subscript $n$.

From (8.3.5)

$$g'_i(u)=\frac{1}{f(x,\theta)}\frac{\partial}{\partial\theta_i}f(x,\theta)=\frac{\partial}{\partial\theta_i}\log f(x,\theta)\,,\quad i=1,2,$$ which

gives

(9.2.12) $h'_1(u)=\sigma_1^2 g'_1(u)+\sigma_{12}g'_2(u),\quad h'_2(u)=\sigma_2^2 g'_2(u)+\sigma_{12}g'_1(u)\,.$

Integrating (9.1.12) and noting $g_i(1)=g_i(0)=0$ we get

(9.2.13) $h_1(u)=\sigma_1^2 g_1(u)+\sigma_{12}g_2(u),\quad h_2(u)=\sigma_2^2 g_2(u)+\sigma_{12}g_1(u)\,.$

Thus, condition (iv) of Lemma 8.3.2 is satisfied. Substitution of (9.2.13) in (8.3.9) yields (9.2.11), which proves first part of the lemma. Now (1) and (2) follow as

$$\int_0^1 [\varphi_1'(u)]^2 du = \frac{\int_0^1 [g_1'(u)]^2 du}{(1-\gamma^2)\, E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\right)^2} = \frac{1}{1-\gamma^2}$$

and

$$\int_0^1 \varphi_1'(u)\varphi_2'(u)\,du = \frac{1}{1-\gamma^2} \cdot \frac{E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\cdot\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)}{\left[E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\right)^2 E\left(\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)^2\right]^{1/2}}$$

$$= -\gamma/(1-\gamma^2) \quad . \qquad //$$

## 9.3 Limiting Distribution of $C_n^2$

The following theorem proves that $C_n^2$ converges in distribution to

$C^2 = \int_0^1 Y^2(u)du$ , where $Y(u)$ , $0 \le u \le 1$ is a Gaussian process

with mean zero and covariance function $\rho(u, v)$ defined by (9.2.11). Also note that we have not made any auxiliary assumptions on the function $\varphi_1(u)$ used by Darling [5, p. 9].

**Theorem 9.3.1.** If $\hat{\theta}_1, \hat{\theta}_2$ are regular, unbiased jointly efficient estimators, then

$$\lim_{n\to\infty} P\{C_n^2 < x\} = P\left\{\int_0^1 Y^2(u)du < x\right\},$$

where $Y(u)$ is a Gaussian process with mean zero and covariance function

$$(9.3.1) \quad \rho(u, v) = \min(u, v) - uv - \varphi_1(u)\varphi_1(v) - \varphi_2(u)\varphi_2(v)$$
$$- r\varphi_1(u)\varphi_2(v) - r\varphi_1(v)\varphi_2(u) , \quad 0 \leq u , v \leq 1 .$$

Proof. Note that the functions $\varphi_i(u)$ defined by (9.2.10) are continuous and $\varphi_i(u) \in L_2(0, 1)$ , $i = 1, 2$ . By Lemma 9.2.1 the process $Y_n(u)$ given by (8.3.8) converges in distribution to a Gaussian process $Y(u)$ which has mean zero and covariance function $\rho(u, v)$ defined by (9.3.1). Write $\rho(u, v)$ as $\rho(u, v) = \min(u, v) - uv - \phi_1(u)\phi_1(v) - \phi_2(u)\phi_2(v)$ , where

$$\phi_1(u) = \sqrt{(1 - r^2)} \; \varphi_1(u) , \quad \text{and} \quad \phi_2(u) = r\varphi_1(u) + \varphi_2(u) .$$

By a method similar to that used in [6, pp. 195-197] we can get a Kac-Siegert representation, [10] for Gaussian process $Y(u)$ with mean zero and covariance function $\rho(u, v)$ and show that the sample functions of the process $Y(u)$ are continuous with probability one. Hence an application of Donsker's Theorem [11] gives the required result. ||

The characteristic function of the random variable

$C^2 = \int_0^1 Y^2(u)\,du$ is given by, see [9],

$$(9.3.2) \quad E\left\{\exp\left[it\int_0^1 Y^2(u)\,du\right]\right\} = \prod_{j=1}^{\infty}\left(1 - \frac{2it}{\mu_j}\right)^{-1/2} ,$$

where $\{\mu_j\}$ are the eigen values of the kernel $\rho(u, v)$ defined by (9.3.1) i.e. roots of the integral equation

$$g(u) = \mu \int_0^1 \rho(u, v)\, g(v)\, dv .$$

The expression on the right hand side of (9.3.2) is nothing but $[D\,(2\,it)]^{-1/2}$, where $D(\mu)$ denotes the Fredholm determinant (F.D.) associated with the kernel $\rho(u, v)$. Thus to obtain the characteristic function of the limiting distribution we have to find the F.D. of the kernel $\rho(u, v)$. We find this characteristic function in Section 9.5.

## 9.4 Case of Maximum-likelihood Estimators

Assume that all the conditions of Cramér [4, pp. 500-504] are satisfied. These conditions imply those of the Lemmas 8.3.1 and 8.3.2 except possibly condition (iv) of the latter. We assume that condition. Then by arguments similar to those used by Darling [5, Section 5] in the case when $\hat{\theta}_1$, $\hat{\theta}_2$ are maximum likelihood estimators, the asymptotic distribution of $C_n^2$ is given by Theorem 9.3.1.

## 9.5 Fredholm Determinant of the Kernel $\rho(x, y)$

This section gives two methods of finding the F.D. of positive definite kernels of special form which enable us to get the characteristic function of the limiting distribution of $C_n^2$.

Theorem 9.5.1. Let

(9.5.1) $$\rho(x, y) = K(x, y) - \phi_1(x)\,\phi_1(y) - \phi_2(x)\,\phi_2(y)\ , \quad 0 \le x,\ y \le 1\ ,$$

be a positive definite kernel, where $K(x, y)$ is a bounded symmetric, positive definite kernel over the unit square $0 \le x\ ,\ y \le 1$ and

$\psi_i(x) \in L_2(0, 1)$ , $i = 1, 2$ . Let the kernel $K(x, y)$ have simple eigen values $0 < \lambda_1 < \lambda_2 < \cdots$ and $f_1(x)$ , $f_2(x)$ ... be the corresponding normalized eigen functions of $K(x, y)$ , also let $d_1(\lambda)$ be the Fredholm determinant (F.D.) associated with $K(x, y)$ . Define

$$(9.5.2) \qquad \alpha_j = \int_0^1 \psi_1(x) f_j(x)\, dx \,, \quad \beta_j = \int_0^1 \psi_2(x) f_j(x)\, dx \,, \; j = 1, 2 \cdots ,$$

$$(9.5.3) \qquad c_i(g) = \int_0^1 \psi_i(x) g(x)\, dx \,,$$

$$(9.5.4) \qquad \rho_i(x, y) = K(x, y) - \psi_i(x)\, \psi_i(y) \,, \quad i = 1, 2 \,.$$

Let $\{\lambda_j^*\}$ $\left(\{\lambda_j^{**}\}\right)$ and $\{f_j^*(x)\}$ $\left(\{f_j^{**}(x)\}\right)$ denote respectively the eigen values (in the order of magnitude) and the corresponding normalized eigen functions of the kernel $\rho_1(x, y)$ $\left(\rho_2(x, y)\right)$ ; $\alpha_j^*$ $(\alpha_j^{**})$ , $\beta_j^*$ $(\beta_j^{**})$ be defined as in (9.5.2) with $f_j$ replaced by $f_j^*$ $(f_j^{**})$ .
Also define

$$(9.5.5) \qquad P_1(\lambda) = 1 + \lambda \sum_{j=1}^{\infty} \frac{\alpha_j^2}{1 - \lambda/\lambda_j} \,, \quad P_2(\lambda) = 1 + \lambda \sum_{j=1}^{\infty} \frac{\beta_j^2}{1 - \lambda/\lambda_j} \,, \; \lambda \neq \lambda_j \,.$$

$P_2^*(\lambda)$ $\left(P_1^{**}(\lambda)\right)$ are obtained by replacing $\beta_j(\alpha_j)$ by $\beta_j^*$ $(\alpha_j^{**})$ in $P_2(\lambda)$ $\left(P_1(\lambda)\right)$ . Then the F.D. $D(\lambda)$ associated with the kernel $\rho(x, y)$ is given by

(9.5.6) $\quad D(\lambda) = d_1(\lambda)\, P_1(\lambda)\, P_2^*(\lambda) = d_1(\lambda)\, P_2(\lambda)\, P_1^{**}(\lambda)$ .

Proof. We prove that $D(\lambda) = d_1(\lambda)\, P_1(\lambda)\, P_2^*(\lambda)$ . Since $\rho(x, y)$ is a positive definite kernel, $\rho_1(x, y)$ being the sum of two positive definite kernels is also positive definite. By theorem 6.2 of [5], the F.D. $D_1(\lambda)$ of the kernel $\rho_1(x, y)$ is $D_1(\lambda) = d_1(\lambda)\, P_1(\lambda)$ . Now we proceed to show that the F.D. associated with $\rho(x, y)$ is $D(\lambda) = D_1(\lambda)\, P_2^*(\lambda)$ .

The integral equation

(9.2.7) $$g(x) = \lambda \int_0^1 [K(x,y) - \psi_1(x)\psi_1(y) - \psi_2(x)\psi_2(y)]\, g(y)\,dy$$

can be written as

(9.5.8) $$g(x) = -\lambda\, c_2(g)\psi_2(g) + \lambda \int_0^1 \rho_1(x,y)\, g(y)\,dy \; .$$

Then we have

(9.5.9) $$g(x) = -\lambda\, c_2(g) \sum_{j=1}^{\infty} \frac{\beta_j^*}{1-\lambda/\lambda_j^*} f_j^*(x), \quad \lambda \neq \lambda_j^* \; ,$$

see, [12, p. 228]. As g appears on both sides of (9.5.9) it is not a solution of (9.5.8). Multiplying both sides of (9.5.9) by $\psi_2(x)$ and integrating we obtain

$$c_2(g)\left[1 + \lambda \sum_{j=1}^{\infty} \frac{{\beta_j^*}^2}{1-\lambda/\lambda_j^*}\right] = 0 \; , \quad i.e. \; c_2(g) P_2^*(\lambda) = 0 \, .$$

This implies that either $g$ is such that $c_2(g) = 0$ or $\lambda$ is a zero of $P_2^*(\lambda)$. When $\lambda \neq \lambda_j^*$ $c_2(g) \neq 0$, because if $c_2(g) = 0$, (9.5.8) is a homogeneous equation with a non zero solution for $\lambda \neq \lambda_j^*$. Therefore, only for those values of $\lambda$, which are either zeros of $P_2^*(\lambda)$ or are eigen values of the kernel $\rho_1(x, y)$, the equation (9.5.8) can have a solution i.e. $\lambda$ is a zero of $D_1(\lambda)\, P_2^*(\lambda)$. $P_2^*(\lambda)$ is analytic except for possible simple poles at $\lambda = \lambda_j^*$. Also $D_1(0)\, P_2^*(0) = 1$.

To prove that $D_1(\lambda)\, P_2^*(\lambda)$ is the F.D. of the kernel $\rho(x, y)$ we have to show that for any zero $\lambda = \bar{\lambda}$ of $D_1(\lambda)\, P_2^*(\lambda)$ there exists a solution $\bar{g}(x)$ of the integral equation (9.5.8) such that

$\int_0^1 \bar{g}^2(x)\, dx = 1$. In the course of the proof of Theorem 6.2 of [5] we observe that the zeros of $D_1(\lambda)$ are either simple or double. Let $\bar{\lambda}$ be a zero of $D_1(\lambda)\, P_2^*(\lambda)$. We have to consider the following three cases.

(i) $\bar{\lambda} \neq \lambda_j^*$ ;

(ii) $\bar{\lambda} = \lambda_j^*$, where $\lambda_j$ is a simple zero of $D_1(\lambda)$ ; $\beta_j^* = 0$.

(iii) $\bar{\lambda} = \lambda_j^*$, where $\lambda_j^*$ is a double root of $D_1(\lambda)$ say

$$\lambda_j^* = \lambda_{j+1}^*, \quad \beta_j^* = \beta_{j+1}^* = 0 .$$

Note that in case (ii) it is necessary that $\beta_j^* = 0$, because if $\beta_j^* \neq 0$,

$\bar{\lambda}$ cannot be a zero of $D(\lambda)$. Similarly in case (iii) it is necessary that $\beta_j^* = \beta_{j+1}^* = 0$.

In case (i) since $\bar{\lambda}$ is not a zero of $D_1(\lambda)$, it is such that $P_2^*(\bar{\lambda}) = 0$. Then

$$(9.5.10) \qquad \bar{g}(x) = \left[ -\sum_{j=1}^{\infty} \frac{\beta_j^*}{1-\bar{\lambda}/\lambda_j^*} f_j^*(x) \right] \Big/ \left[ \sum_{j=1}^{\infty} \frac{\beta_j^{*2}}{(1-\bar{\lambda}/\lambda_j^*)^2} \right]^{1/2}$$

is the solution of (9.5.8). As $\rho(x, y)$ is symmetric $\bar{\lambda}$ is real. Also since

$$P_2^{*\prime}(\lambda) = \sum_{j=1}^{\infty} \left( \frac{\beta_j^*}{1-\lambda/\lambda_j^*} \right)^2 > 0 , \qquad \text{for real } \lambda, \bar{\lambda}$$

is a simple zero of $P_2^*(\lambda)$. Thus for any $\bar{\lambda}$ under case (i) $\bar{g}(x)$ given by (9.5.10) satisfies (9.5.8).

In case (ii) we have two subcases. (a) $\bar{\lambda}$ is such that $D_1(\bar{\lambda}) = 0$, $P_2^*(\bar{\lambda}) \neq 0$. In this case $\bar{\lambda}$ is a simple zero of $D_1(\lambda)\, P_2^*(\lambda)$ and $f_j^*(x)$ satisfies (9.5.8). (b) If $D_1(\bar{\lambda}) = 0$, $P_2^*(\bar{\lambda}) = 0$, $D_1(\lambda)\, P_2^*(\lambda)$ has a double root at $\bar{\lambda} = \lambda_j^*$. In this case $f_j^*(x)$ and $\bar{g}(x)$ given by (9.5.10) are solutions of (9.5.8).

In case (iii) if $\bar{\lambda}$ is such that $D_1(\bar{\lambda}) = 0$, and $P_2^*(\bar{\lambda}) \neq 0$, $\bar{\lambda}$ is a double root of $D_1(\lambda)\, P_2^*(\lambda)$ and $f_j^*(x)$, $f_{j+1}^*(x)$ satisfy (9.5.8). If $\bar{\lambda}$ is a zero of $P_2^*(\lambda)$ and also $D_1(\bar{\lambda}) = 0$ then $\bar{\lambda}$ is a triple zero of $D_1(\lambda)\, P_2^*(\lambda)$. $f_j^*(x)$, $f_{j+1}^*(x)$ and $\bar{g}(x)$ given by (9.5.10) are the solutions of (9.5.8).

Thus for each zero of $D_1(\lambda)\ P_2^*(\lambda)$ we obtain solutions of appropriate multiplicity to the equation (9.5.8). Hence $D(\lambda) = d_1(\lambda)\ P_1(\lambda)\ P_2^*(\lambda)$ is the F.D. associated with $\rho(x, y)$. Writing the equation (9.5.7) as

$$g(x) = -\lambda c_1(g)\gamma_1(x) + \lambda\int_0^1 \rho_2(x,y)\,g(y)\,dy ,$$

and proceeding in the same manner as above we can show that $D(\lambda) = d_1(\lambda)\ P_2(\lambda)\ P_1^{**}(\lambda)$. This proves the theorem. ||

Even if Theorem 9.5.1 gives a method of obtaining the F.D. of $\rho(x, y)$, the method requires the laborious task of finding the eigen values and eigen functions of two kernels, namely $K(x, y)$ and $\rho_1(x, y)$ or $\rho_2(x, y)$. The following theorem which is a generalization of Theorem 6.2 of Darling [5], avoids the above mentioned difficulty by giving an expression for the F.D. of $\rho(x, y)$ for which only the eigen values and the eigen functions of the kernel $K(x, y)$ are needed. The proof of the theorem was suggested by Professor Gopinath Kallianpur.

__Theorem 9.5.2__. Let

$$\rho(x, y) = K(x, y) - \phi_1(x)\ \phi_1(y) - \phi_2(x)\ \phi_2(y)$$

by a positive definite kernel as described in Theorem 9.5.1. Then the F.D. of the kernel $\rho(x, y)$ is given by

(9.5.11) $\qquad D(\lambda) = d_1(\lambda)\Delta(\lambda)$ ,

where

(9.5.12) $$\Delta(\lambda) = \begin{vmatrix} P_1(\lambda) & Q(\lambda) \\ Q(\lambda) & P_2(\lambda) \end{vmatrix} ,$$

$P_1(\lambda)$ , $P_2(\lambda)$ being defined by (9.5.5) and $Q(\lambda)$ by

(9.5.13) $$Q(\lambda) = \lambda \sum_{j=1}^{\infty} \frac{\alpha_j \beta_j}{1-\lambda/\lambda_j} , \quad \lambda \neq \lambda_j$$

<u>Proof</u>. Write the integral equation (9.5.7) as

(9.5.14) $$g(x) = -\lambda \left[ c_1(g)\psi_1(x) + c_2(g)\psi_2(x) \right] + \lambda \int_0^1 K(x,y) g(y) dy .$$

Then

(9.5.15) $$g(x) = -\lambda c_1(g) \sum_{j=1}^{\infty} \frac{\alpha_j}{1-\lambda/\lambda_j} f_j(x) - \lambda c_2(g) \sum_{j=1}^{\infty} \frac{\beta_j}{1-\lambda/\lambda_j} f_j(x) .$$

Multiply (9.5.15) by $\phi_1(x)$ and $\phi_2(x)$ respectively and integrate to obtain

(9.5.16) $$\begin{cases} c_1(g)\, P_1(\lambda) + c_2(g)\, Q(\lambda) = 0 , \\ c_1(g)\, Q(\lambda) + c_2(g)\, P_2(\lambda) = 0 . \end{cases}$$

(9.5.16) is a system of homogeneous equations in $c_1(g)$ , $c_2(g)$ and has a non-zero solution if and only if $\Delta(\lambda) = 0$. If $\lambda \neq \lambda_j$ both $c_1(g)$ and $c_2(g)$ cannot be zero, because $c_1(g) = c_2(g) = 0$ implies that the equation (9.5.14) is homogeneous which cannot have non-trivial solution unless $d_1(\lambda) = 0$. Therefore the equation (9.5.14) has a solution only when either $\lambda$ is such that $\Delta(\lambda) = 0$ or $\lambda$ is a zero of $d_1(\lambda)$.

To prove that $D(\lambda) = d_1(\lambda)\Delta(\lambda)$ is the F.D. of the kernel $\rho(x, y)$ we show that

$$(9.5.17) \quad d_1(\lambda)\Delta(\lambda) = d_1(\lambda)\, P_1(\lambda)\, P_2^{*}(\lambda) = d_1(\lambda)\, P_2(\lambda)\, P_1^{**}(\lambda)\,.$$

It is sufficient to prove that zeros of $d_1(\lambda)\Delta(\lambda)$ and $d_1(\lambda)\, P_1(\lambda)\, P_2^{*}(\lambda) = d_1(\lambda)\, P_2(\lambda)\, P_1^{**}(\lambda)$ are the same. If $\bar{\lambda}$ is a zero of $d_1(\lambda)$ then it is a zero $d_1(\lambda)\Delta(\lambda)$, $d_1(\lambda)\, P_1(\lambda)\, P_2^{*}(\lambda)$ and also of $d_1(\lambda)\, P_2(\lambda)\, P_1^{**}(\lambda)$. Suppose that $\bar{\lambda}$ is a zero of $\Delta(\lambda)$ and $d_1(\bar{\lambda}) \neq 0$. Since $\Delta(\bar{\lambda}) = 0$ there exists a solution $(c_1(g), c_2(g))$ of (9.5.16) so that at least one of $c_i(g)$ $i = 1, 2$, is not zero. Without any loss of generality assume $c_1(g) \neq 0$. From the integral equation

$$g(x) = -\bar{\lambda}\, c_1(g)\, \psi_1(x) + \bar{\lambda} \int_0^1 \rho_2(x, y)\, g(y)\, dy$$ we have

$$g(x) = -\bar{\lambda}\, c_1(g) \sum_{j=1}^{\infty} \frac{\alpha_j^{**}}{1 - \bar{\lambda}/\lambda_j^{**}}\, f_j^{**}(x)\,.$$

Multiply this by $\psi_1(x)$ and integrate to obtain $c_1(g) P_1^{**}(\bar{\lambda}) = 0$. Since $c_1(g) \neq 0$, $P_1^{**}(\bar{\lambda}) = 0$, which implies that $d_1(\bar{\lambda}) P_2(\bar{\lambda}) P_1^{**}(\bar{\lambda}) = 0$. Thus we have proved that if $\bar{\lambda}$ is a zero of $d_1(\lambda)\Delta(\lambda)$ it is a zero of

$d_1(\lambda)P_1(\lambda)P_1^{**}(\lambda)$ .

Now we prove that a zero of the right hand side of (9.5.17) is a zero of $d_1(\lambda)\Delta(\lambda)$ . Here the following three cases arise.

Case (i) $\bar{\lambda} \neq \lambda_j$ and $\bar{\lambda}$ is such that $P_1(\bar{\lambda}) = 0$ , $P_2(\bar{\lambda}) = 0$ . By Schwarz's inequality $Q(\bar{\lambda}) = 0$ and hence $\Delta(\bar{\lambda}) = 0$ .

Case (ii) $\bar{\lambda} \neq \lambda_j$ , $P_1(\bar{\lambda}) \neq 0$ , $P_2(\bar{\lambda}) = 0$ . In this case also Schwarz's inequality yields $Q(\bar{\lambda}) = 0$ , hence $\Delta(\bar{\lambda}) = 0$ . Similarly when $P_1(\bar{\lambda}) = 0$ , $P_2(\bar{\lambda}) \neq 0$ , $\Delta(\bar{\lambda}) = 0$ .

Case (iii) $\bar{\lambda} \neq \lambda_j$ , $P_1(\bar{\lambda}) \neq 0$ , $P_2(\bar{\lambda}) \neq 0$ . Since $\bar{\lambda} \neq \lambda_j$ both $c_1(g)$ , $c_2(g)$ cannot be zero. Because if $c_1(g) = c_2(g) = 0$ equation (9.5.14) is homogeneous which cannot have a solution unless $\bar{\lambda} = \lambda_j$ . Without any loss of generality assume $c_2(g) \neq 0$ . Then from (9.5.16),

$$c_2(g)[1 - Q^2(\bar{\lambda})/P_1(\bar{\lambda})P_2(\bar{\lambda})] = 0 . \quad \text{As} \quad c_2(g) \neq 0 ,$$

$$Q^2(\bar{\lambda}) = P_1(\bar{\lambda})P_2(\bar{\lambda}) , \quad \text{and } \bar{\lambda} \text{ is a zero of } \Delta(\lambda) .$$

Hence a zero of $d_1(\lambda)P_2(\lambda)P_1^{**}(\lambda)$ is a zero of $d_1(\lambda)\Delta(\lambda)$ . This completes the proof. ||

<u>Corollary 1</u>. If $\phi_1(x) = f_m(x)/\sqrt{\lambda_m}$ , $\phi_2(x) = f_n(x)/\sqrt{\lambda_n}$ , then

$$D(\lambda) = \prod_{\substack{j=1 \\ j \neq m,n}}^{\infty} (1 - \lambda/\lambda_j)$$

<u>Proof</u>. In this case $Q(\lambda) = 0$ and $P_1(\lambda) = \lambda_m/(\lambda_m - \lambda)$ , $P_2(\lambda) = \lambda_n/(\lambda_n - \lambda)$ ; hence the result follows. ||

Corollary 2. The F.D. of the kernel $\rho(x, y)$ defined by (9.3.1) is

$$D(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}}\,\Delta(\lambda), \text{ where}$$

$$\Delta(\lambda) = \begin{vmatrix} 1+\lambda(1-r^2)\sum\limits_{j=1}^{\infty}\dfrac{a_j^2}{1-\lambda/\pi^2 j^2}, & \lambda\sqrt{1-r^2}\sum\limits_{j=1}^{\infty}\dfrac{ra_j^2+a_jb_j}{1-\lambda/\pi^2 j^2} \\ \lambda\sqrt{1-r^2}\sum\limits_{j=1}^{\infty}\dfrac{ra_j^2+a_jb_j}{1-\lambda/\pi^2 j^2}, & 1+\lambda\sum\limits_{j=1}^{\infty}\dfrac{(ra_j+b_j)^2}{1-\lambda/\pi^2 j^2} \end{vmatrix},$$

with $a_j = \sqrt{2}\int_0^1 \varphi_1(x)\sin(\pi jx)dx$, $b_j = \sqrt{2}\int_0^1 \varphi_2(x)\sin(\pi jx)dx$,

$j = 1, 2, \ldots$

Proof. Write $\phi_1(x) = \varphi_1(x)\sqrt{1-r^2}$, $\phi_2(x) = r\varphi_1(x) + \varphi_2(x)$, then (9.3.1) reduces to

$$\rho(x, y) = \min(x, y) - xy - \phi_1(x)\phi_1(y) - \phi_2(x)\phi_2(y).$$

Also for the kernel $K(x, y) = \min(x, y) - xy$

$$\lambda_j = \pi^2 j^2, \quad f_j(x) = \sqrt{2}\sin(\pi jx), \quad d_1(\lambda) = (\sin\sqrt{\lambda})/\sqrt{\lambda}.$$

Substitution of $\alpha_j = \sqrt{(1-r^2)}\,a_j$, $\beta_j = r\,a_j + b_j$, and $d_1(\lambda) = (\sin\sqrt{\lambda})/\sqrt{\lambda}$ in (9.5.11) yields the required result. ||

## 9.6 Some Properties of $C_n^2$ Test and Applications.

Cumulants of the limiting distribution: As in [9] it follows that the cumulants $\kappa_j$ of the asymptotic distribution of $C_n^2$ are given by

$$(9.6.1) \qquad \kappa_j = 2^{j-1}(j-1)! \sum_{n=1}^{\infty} \frac{1}{\mu_n^j}, \qquad j = 1, 2, \ldots$$

where $\{\mu_j\}$ are eigen values of $\rho(u, v)$. On account of Mercer's theorem [13] $\kappa_j$ can also be obtained from

$$\kappa_j = 2^{j-1}(j-1)! \int_0^1 \rho_j(u,u)\,du ,$$

where $\rho_j(u, v)$ is the jth interate of the kernel $\rho(u, v)$ i.e.

$$\rho_1(u, v) = \rho(u, v), \quad \rho_j(u, v) = \int_0^1 \rho_{j-1}(u, s)\rho(s, v)\,ds .$$

Hence the mean and the variance of the limiting distribution are obtained as

$$k_1 = \int_0^1 \rho(u,u)\,du$$

$$= \frac{1}{6} - \int_0^1 \varphi_1^2(u)\,du - \int_0^1 \varphi_2^2(u)\,du - 2r\int_0^1 \varphi_1(u)\varphi_2(u)\,du,$$

$$K_2 = 2\int_0^1\int_0^1 [\min(u,v) - uv - \varphi_1(u)\varphi_1(v) - \varphi_2(u)\varphi_2(v) - r\varphi_1(u)\varphi_2(v) - r\varphi_1(v)\varphi_2(u)]^2 \, du\, dv$$

$$= \frac{1}{45} + 2\left(\mu - \frac{1}{6}\right)^2 + 4(1-r^2)\int_0^1\int_0^1 \varphi_1(v)\varphi_2(v)\varphi_1(u)\varphi_2(u)\, du\, dv$$

$$- 4(1-r^2)\int_0^1\int_0^1 \varphi_2^2(v)\varphi_1^2(u)\, du\, dv - 8\int_0^1 (1-v)\varphi_1(v)\int_0^v u\varphi_1(u)\, du\, dv$$

$$- 8\int_0^1 (1-v)\varphi_2(v)\int_0^v u\varphi_2(u)\, du\, dv - 16r\int_0^1 (1-v)\varphi_2(v)\int_0^1 u\varphi_1(u)\, du\, dv .$$

When $\theta_1$, $\theta_2$ are both known, the Cramér-Smirnov test based on $\omega_n^2$ is used for testing the hypothesis $G(x) = F(x, \theta)$. The limiting distribution of $\omega_n^2$ is the distribution of the random variable

$\omega^2 = \int_0^1 W^2(u)\,du$, where $W(u)$ is a Gaussian process with mean zero and covariance function $\min(u, v) - uv$. Using Kac-Siegert representation [10] for the process $W(u)$, $\omega^2$ can be written as

$\omega^2 = \sum_{j=1}^{\infty}(G_j^2/\pi^2 j^2)$, where $G_1, G_2, \ldots$ are independently normally distributed with mean zero and variance 1.

When $\theta_1, \theta_2$ are unknown the limiting distribution of $C_n^2$ is the distribution of the random variable $C^2 = \int_0^1 Y^2(u)du$, where $Y(u)$ is a Gaussian process with mean zero and covariance function $\rho(u, v)$ given by (9.3.1). $C^2$ can be expressed as $C^2 = \sum_{j=1}^{\infty} (G_j^2 / \mu_j)$, where $\{\mu_j\}$ are eigen values of $\rho(u, v)$ and $G_1, G_2, \ldots$ are independently normally distributed with mean zero and variance 1 .

Note that $\varphi_1(x)\varphi_1(y) + \varphi_2(x)\varphi_2(y) + r\varphi_1(x)\varphi_2(y) + r\varphi_1(y)\varphi_2(x)$ is a positive definite kernel. Hence by maximum-minimum property of eigen values, [13, 14] it follows that the weights $1/\mu_j$ in $C^2$ are not greater than the weights $1/\pi^2 j^2$ in $\omega^2$. In the case when $\psi_1(x)$ and $\psi_2(x)$ are functions of the special form as described in Corollary 1 of Theorem 9.3.2, the number of terms in the infinite product for $D(\lambda)$ is reduced by 2 . This is analogous to reduction of degrees of freedom in the usual $\chi^2$ theory.

The cumulants of the distribution of $\omega^2$ are

$\kappa_j^{(0)} = 2^{j-1}(j-1)! \sum_{r=1}^{\infty}\left(1/\pi^2 r^2\right)^j$, while those of $C^2$ are given by (9.6.1). Since $1/\pi^2 j^2 \geq 1/\mu_j$, $\kappa_j^{(0)} \geq \kappa_j$ .

Scale and Location Parameters: A test is said to be asymptotically parameter-free if its limiting distribution under the hypothesis is in-

dependent of the unknown parameters. The $C_n^2$ test under investigation will be asymptotically parameter-free if $\rho(u, v)$, the covariance function involved in the asymptotic distribution of $C_n^2$, does not depend on the unknown parameters $\theta_1, \theta_2$ in $F(x, \theta_1, \theta_2)$. The following theorem shows that when $\theta_1$ is a location parameter and $\theta_2$ a scale parameter, $\rho(u, v)$ is independent of $\theta_1, \theta_2$ and hence the $C_n^2$ test is asymptotically parameter-free. In the case when the distribution depends on only one unknown parameter $\theta$, Darling has shown that if $\theta$ is the scale parameter or the location parameter the modified Cramér-Smirnov test is asymptotically parameter-free.

<u>Theorem 9.6.1.</u> If the distribution function $F$ is such that $F(x, \theta) = H((x - \theta_1)/\theta_2)$, $-\infty < \theta_1 < +\infty$, $\theta_2 > 0$, then $\rho(u, v)$ defined by (9.3.1) is independent of $\theta_1, \theta_2$.

<u>Proof.</u>

$$f(x,\theta) = H'\left(\frac{x-\theta_1}{\theta_2}\right) = \frac{1}{\theta_2} h\left(\frac{x-\theta_1}{\theta_2}\right). \qquad \text{Hence}$$

$$E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\right)^2 = \frac{1}{\theta_2^2}\int_{-\infty}^{+\infty}\left\{[-h'(y)]^2/h(y)\right\}dy,$$

$$E\left(\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)^2 = \frac{1}{\theta_2^2}\int_{-\infty}^{+\infty}\left\{[yh'(y)]^2/h(y)\right\}dy - \frac{1}{\theta_2^2},$$

and

$$E\left(\frac{\partial}{\partial\theta_1}\log f(X,\theta)\cdot\frac{\partial}{\partial\theta_2}\log f(X,\theta)\right)=\frac{1}{\theta_2^2}\int_{-\infty}^{+\infty}\{y[h'(y)]^2/h(y)\}dy\ .$$

$$r=-\frac{\int_{-\infty}^{+\infty}\{y[h'(y)]^2/h(y)\}dy}{\left\{\left[\int_{-\infty}^{+\infty}\{[h'(y)]^2/h(y)\}dy\right]\left[\int_{-\infty}^{+\infty}\{y[h'(y)]^2/h(y)\}dy\right]\right\}^{1/2}}$$

is independent of $\theta_1$, $\theta_2$ . Using these results

$$\varphi_1(u)\varphi_1(v)=\{h[H^{-1}(u)]h[H^{-1}(v)]\}\Big/\left\{(1-r^2)\int_{-\infty}^{+\infty}([h'(y)]^2/h(y))dy\right\}\ ,$$

$$\varphi_2(u)\varphi_2(v)=\{H^{-1}(u)H^{-1}(v)h[H^{-1}(v)]\,h[H^{-1}(u)]\}\Big/\left\{(1-r^2)\int_{-\infty}^{+\infty}([yh'(y)]^2/h(y))dy-1\right\}$$

$$r\varphi_1(u)\varphi_2(v)=\frac{rH^{-1}(v)h[H^{-1}(u)]h[H^{-1}(v)]}{(1-r^2)\left\{\left[\int_{-\infty}^{+\infty}\{[h'(y)]^2/h(y)\}dy\right]\left[\int_{-\infty}^{+\infty}\{[yh'(y)]^2/h(y)\}dy-1\right]\right\}^{1/2}}$$

$$r\varphi_1(v)\varphi_2(u)=\frac{rH^{-1}(u)h[H^{-1}(u)]h[H^{-1}(v)]}{(1-r^2)\left\{\left[\int_{-\infty}^{+\infty}\{[h'(y)]^2/h(y)\}dy\right]\left[\int_{-\infty}^{+\infty}\{[yh'(y)]^2/h(y)\}dy-1\right]\right\}^{1/2}}$$

are independent of $\theta_1$, $\theta_2$ and hence the result. ||

Case of Normal Distribution: In case $F(x, \theta)$ is a normal distribution $N(x, \mu, \sigma^2)$ with unknown mean $\mu$ and variance $\sigma^2$, we estimate $\mu$ by $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$, and $\sigma^2$ by $s^2 = (1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2$. In this case $r$, $\sigma_1^2$, $\sigma_2^2$ defined by (9.1.6) and (9.1.7) are found to be $r = 0$, $\sigma_1^2 = \sigma^2$, $\sigma_2^2 = 2\sigma^4$. Now we compute $\varphi_1(u)$, $\varphi_2(u)$ required to obtain $\rho(u, v)$ given by (9.3.1). Let $Y_j = (X_j - \mu)/\sigma$, $j = 1, 2, \ldots, n$. $Y_1, \ldots, Y_n$ is a sample of $n$ independent observations from $N(y, 0, 1)$. Let $0 \le u \le 1$ and

$$\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}, \quad \Phi(y) = \int_{-\infty}^{y} \phi(x)\,dx, \quad J(u) = \{y : u = \Phi(y)\}.$$

Now we find $h_1(u)$, $h_2(u)$ as

$$h_1(u) = \lim_{n\to\infty} E\left(\sqrt{n}\, Z_n(u)\, \bar{Y}\right) = \lim_{n\to\infty}\left[nuE\{\bar{Y} \mid Y_1 < J(u)\} - nuE\bar{Y}\right]$$

$$= \lim_{n\to\infty}\left\{nuE\left[\bar{Y} \mid Y_1 < J(u)\right]\right\} = \phi(J(u)).$$

$$h_2(u) = \lim_{n\to\infty} E\left[\sqrt{n}\, Z_n(u)(s^2-1)\right]$$

$$= \lim_{n\to\infty}\left[nuE\{(s^2-1) \mid Y_1 < J(u)\} - nuE(s^2-1)\right]$$

$$= \lim_{n\to\infty}\left[s^2 \mid Y_1 < J(u)\right] = J(u)\,\phi(J(u)).$$

Hence Using $\sigma_1^2 = 1$ , $\sigma_2^2 = 2$ , $\varphi_i(u) = h_i(u)/\sigma_i$ , $i = 1, 2$ we obtain $\varphi_1(u) = \phi(J(u))$ and $\varphi_2(u) = \frac{1}{\sqrt{2}} J(u)\,\phi(J(u))$ , and

$$\rho(u, v) = \min(u, v) - uv - \phi(J(u))\,\phi(J(u)) - \frac{J(u)\,J(v)}{2}\,\phi(J(u))\,\phi(J(v)) .$$

This was obtained in [6] by quite a different method.

## 10. k-Sample Cramér-Smirnov Test in the Parametric Case

10.1 In Sections 8 and 9 we considered Cramér-Smirnov test for one sample problem when the functional form of the underlying distribution was known but the parameters on which it depended were unknown. In this Section we propose to study the modification of k-sample Cramér-Smirnov test in parametric case.

Let $X_{ji}$ $(i = 1, 2, \ldots, n_j,\ j = 1, 2, \ldots, k)$ be independent random variables with continuous distribution function $G_j(x)$ . For every $\theta \in I$ , an open interval in $R^1$, let $F(x, \theta)$ be an absolutely continuous distribution function. For testing the hypothesis $H_k$: $G_1(x) = G_2(x) = \cdots = G_k(x) = F(x, \theta)$ , when the functional form of $F$ is known but $\theta$ is unknown, consider the test based on the statistic

$$C_n'^2 = \int_{-\infty}^{+\infty} \sum_{j=1}^{k} n_j \left[ F_{n_j}^{(j)}(x) - F(x, \hat{\theta}_N) \right]^2 dF(x, \hat{\theta}_N) , \tag{10.1.1}$$

where $N = \sum_{j=1}^{k} n_j$ , $n = (n_1, n_2, \ldots, n_k)$ , $F_{n_j}^{(j)}(x)$ is the empirical distribution function of $j$th sample, and $\hat{\theta}_N$ is an estimate of $\theta$ obtained from the pooled sample. The hypothesis $H_k$ is rejected if $C_n'^2$ is too large. The aim of this section is to find the asymptotic distribution of $C_n'^2$ under the hypothesis $H_k$. Here also, methods used in [5] and [6] are employed. Throughout this chapter it is assumed that when $N \to \infty$ each $n_j \to \infty$ $(j = 1, 2, \ldots, k)$ and $\lim_{N \to \infty} (n_j/N) = a_j$ exists.

We note that the asymptotic distribution depends on the properties of the estimator $\hat{\theta}_N$ and the characteristic function of the limiting distribution given in Section 10.4 involves $a_j$'s .

## 10.2 <u>Case When $\hat{\theta}_N$ is a Superefficient Estimator</u>

Suppose the hypothesis $H_k$ is true. Let $\theta$ be the true unknown value of the parameter and $f(x, \theta)$ be the probability density function corresponding to $F(x, \theta)$ . k-sample Cramér-Smirnov test statistic $\omega_n'^2$ is defined as ,

$$\omega_n'^2 = \int_{-\infty}^{+\infty} \sum_{j=1}^{k} n_j \left[ F_{n_j}^{(j)}(x) - F(x,\theta) \right]^2 f(x,\theta)\,dx .$$

Let $X_{j1}', X_{j2}', \ldots, X_{jn_j}'$ be a rearrangement of the $j$th sample $X_{j1}, X_{j2}, \ldots, X_{jn_j}$ so that $X_{j1}' < X_{j2}' < \cdots < X_{jn_j}'$ . Then $\omega_n'^2$ and $C_n'^2$ can be written as

$$\omega_n'^2 = \sum_{j=1}^{k} \frac{1}{12n_j} + \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[F(X_{ji}', \theta) - \frac{(2i-1)}{2n_j}\right]^2, \tag{10.2.1}$$

$$C_n'^2 = \sum_{j=1}^{k} \frac{1}{12n_j} + \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[F(X_{ji}', \hat{\theta}_N) - \frac{(2i-1)}{2n_j}\right]^2. \tag{10.2.2}$$

Theorem 10.2.1. Let $\hat{\theta}_N$ and $F(x, \theta)$ satisfy

(i) $\lim_{N \to \infty} N\, E(\hat{\theta}_N - \theta)^2 = 0$,

(ii) For $\theta, \theta' \in I$

$|F(x, \theta) - F(x, \theta')| \in A(x)|\theta - \theta'|$,

where $A$ is such that $P\{A^2(x) > A_0\} = 0$ for some $A_0 < \infty$, where the probability is according to true distribution $F(x, \theta)$.

(iii) $\lim_{N \to \infty} n_j/N = a_j$, $j = 1, 2, \ldots, k$.

Then $C_N'^2 = \omega_n'^2 + \delta_N$ where $\lim_{N \to \infty} \delta_N = 0$.

Proof. This theorem can be proved in a manner analogous to that of Theorem 2.1 of Darling [5]. ||

Remark. Under the conditions of Theorem 3.2.1 the limiting distribution of $C_n'^2$ is the same as that of $\omega_N'^2$ which is given in [7].

## 10.3 Case When $\hat{\theta}_N$ is a Regular Estimator.

Let $\hat{\theta}_N$ be a regular estimator in the sense of Cramér [4, p. 479]. In this case $\mathrm{Var}(\hat{\theta}_N) \geq A/N$ for some positive $A$. In general even if assumption (i) of Theorem 10.2.1 may not be true, in many cases we shall have for some $\delta$ such that $1/2 > \delta > 0$ $\lim_{N \to \infty} N^{1/2 - \delta}(\hat{\theta}_N - \theta) = 0$.

The following lemma enables us to write $C_n'^2$ in a suitable form that will be useful in obtaining the limiting distribution.

**Lemma 10.3.1.** Let

(i) $\lim_{N \to \infty} (n_j/N) = a_j$ , $j = 1, 2, \ldots, k$ .

(ii) For $1/2 > \delta > 0$ $\lim_{N \to \infty} N^{1/2 - \delta}(\hat{\theta}_N - \theta) = 0$ .

For almost all $x$ and all $\theta \in I$

(iii) $\left| \frac{\partial^2}{\partial \theta^2} F(x, \theta) \right| < g_0(x)$

(iv) $\left| \frac{\partial}{\partial \theta} f(x, \theta) \right| < g_1(x)$

where $g_0(x)$ , $g_1(x)$ are integrable functions and also independent of the exceptional set. Then

(10.3.1) $C_n'^2 = C_n'^{\,2} + \delta_N$ ,

where $\operatorname{plim}_{N \to \infty} \delta_N = 0$ , and

(10.3.2) $$C_n'^{*2} = \int_{-\infty}^{+\infty} \sum_{j=1}^{k} n_j \left[ F_{n_j}^{(j)}(x) - F(x,\theta) - (\hat{\theta}_N - \theta)\frac{\partial}{\partial\theta} F(x,\theta) \right]^2 f(x,\theta)\,dx .$$

<u>Proof</u>. Expand $F(x, \hat{\theta}_N)$ and $f(x, \hat{\theta}_N)$ in a Taylor's series around the true value $\theta$ :

$$F(x,\hat{\theta}_N) = F(x,\theta) + (\hat{\theta}_N - \theta)\frac{\partial}{\partial\theta}F(x,\theta) + \frac{1}{2}(\hat{\theta}_N - \theta)^2 \Delta_0 g_0(x),\ |\Delta_0| < 1,$$

$$f(x, \hat{\theta}_N) = f(x,\theta) + (\hat{\theta}_N - \theta)\Delta_1 g_1(x),\ |\Delta_1| < 1 .$$

Substitution of these expressions in (10.1.1) yields,

(10.3.3) $$C_n'^2 = \sum_{j=1}^{k} n_j \int_{-\infty}^{+\infty} \left[ F_{n_j}^{(j)}(x) - F(x,\theta) - (\hat{\theta}_N - \theta)\frac{\partial}{\partial\theta}F(x,\theta) \right]^2 f(x,\theta)\,dx$$

$$+ \sum_{j=1}^{k} (n_j/4) \int_{-\infty}^{+\infty} (\hat{\theta}_N - \theta)^4 \Delta_0^2 g_0^2(x) f(x,\theta)\,dx$$

$$- \sum_{j=1}^{k} n_j \int_{-\infty}^{+\infty} \left[ F_{n_j}^{(j)}(x) - F(x,\theta) - (\hat{\theta}_N - \theta)\frac{\partial}{\partial\theta}F(x,\theta) \right] (\hat{\theta}_N - \theta)^2 \Delta_0 g_0(x) f(x,\theta)\,dx$$

$$+ \sum_{j=1}^{k} n_j \int_{-\infty}^{+\infty} \left[ F_{n_j}^{(j)}(x) - F(x,\theta) - (\hat{\theta}_N - \theta)\frac{\partial}{\partial\theta}F(x,\theta) \right]^2 (\hat{\theta}_N - \theta)\Delta_1 g_1(x)\,dx$$

$+$

$$+\sum_{j=1}^{k}(n_j/4)\int_{-\infty}^{+\infty}(\hat{\theta}_N-\theta)^5\Delta_0^2 g_0^2(x)g_1(x)\,dx$$

$$-\sum_{j=1}^{k} n_j\int_{-\infty}^{+\infty}\Big[F_{n_j}^{(j)}(x)-F(x,\theta)-(\hat{\theta}_N-\theta)\frac{\partial}{\partial\theta}F(x,\theta)\Big](\hat{\theta}_N-\theta)^3\Delta_0\Delta_1 g_0(x)g_1(x)\,dx\,.$$

By an application of Kolmogorov's theorem, [8] under the assumptions (i) - (iv) of the lemma it follows that all the terms except the first one in (10.3.3) converge in probability to zero and hence the result follows. ||

Lemma 10.3.1 reduces the problem of finding the asymptotic distribution of $C_n'^2$ to obtaining that of $C_n'^{*2}$ given by (10.3.2).

The following transformations will be used in the sections to follow. As in Section 8.3 let $u = F(x, \theta)$ , $u_{ji} = F(X_{ji}, \theta)$ . Define as before $\psi_t(x) = 1$ if $x < t$ , $\psi_t(x) = 0$ , if $x \geq t$ . Then with probability one we have

$$F_{n_j}^{(j)}(x) = \frac{1}{n_j}\sum_{i=1}^{n_j}\psi_x(X_{ji}) = \frac{1}{n_j}\sum_{i=1}^{n_j}\psi_u(u_{ji})\,.$$

Also set

(10.3.4) $$Z_{n_j}^{(j)}(u) = \sqrt{n_j}\,[F_{n_j}^{(j)}(u)-u] = \sqrt{n_j}\Big[\frac{1}{n_j}\sum_{i=1}^{n_j}\psi_u(u_{ji})-u\Big]\,,$$

(10.3.5) $$g(u) = \frac{\partial}{\partial\theta}F(x,\theta).$$

Observe that $g(u)$ is in general a function of $\theta$. Employing these transformations, $C_n'^2$ can be written as

$$C_n'^2 = \sum_{j=1}^{k} \int_0^1 [Y_{n_j}^{(j)}(u)]^2 du + \delta_N \tag{10.3.6}$$

where $\underset{N \to \infty}{\text{plim}}\ \delta_N = 0$, and

$$Y_{n_j}^{(j)}(u) = Z_{n_j}^{(j)}(u) - \sqrt{n_j}\ (\hat{\theta}_N - \theta) . \tag{10.3.7}$$

The following lemma gives the limiting form of the stochastic process $Y_{n_j}^{(j)}(u)$. We note that this lemma is analogous to Lemma 3.2 of [5] which is proved under somewhat different assumption. For comparison of these one may refer back to comments before Lemma 8.3.2.

<u>Lemma 10.3.2</u>. Assume that

(i) $C_n'^2$ can be written as in (10.3.6),

(ii) $\sqrt{N}\ (\hat{\theta}_N - \theta)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 > 0$.

(iii) $\lim_{N \to \infty} E\ (Z_{n_j}^{(j)}(u)\ \sqrt{n_j}\ (\hat{\theta}_N - \theta)) = a_j h(u)$, where $h(u)$ is such that $h(1) = h(0) = 0$, and $a_j = \lim_{N \to \infty} (n_j/N)$.

Then the stochastic process $Y_n^{(j)}(u)$ given by (10.3.7) converges in

distribution to a Gaussian process $Y_j(u)$ which has mean zero and covariance function

$$(10.3.8) \quad \rho_j(u, v) = \min(u, v) - uv - a_j g(v)h(u) - a_j g(u)h(v) + a_j \sigma^2 g(u)g(v) .$$

Proof. It is known that the process $Z^{(j)}_{n_j}(u)$ converges in distribution to a Gaussian process $Z_j(u)$ which has mean zero and covariance function $K(u, v) = \min(u, v) - uv$, see [9]. From the assumption that $\lim_{N \to \infty} (n_j/N) = a_j$ and (ii) it follows that $\sqrt{n_j}\,(\hat{\theta}_N - \theta)$ is asymptotically $\mathcal{N}(0, a_j\sigma^2)$. Hence the process $Y^{(j)}_{n_j}(u)$ converges in distribution to a Gaussian process $Y_j(u)$ which has mean zero and the assumption (iii) yields that

$$\lim_{n_j \to \infty} \rho^{(j)}_{n_j}(u) = \lim_{n_j \to \infty} E\left[Y^{(j)}_{n_j}(u)\, Y^{(j)}_{n_j}(v)\right] = \rho_j(u, v)$$

given by (10.3.1) and hence the result. ||

To find the limiting distribution of $C^2_n$ the estimator $\hat{\theta}_N$ is specialized further in the next section.

## 10.4 Case of Efficient Estimator

Suppose that $\hat{\theta}_N$ is an unbiased, regular efficient estimator in the sense of [4]. Further we assume that Cramér's conditions [4, pp. 477-489] are satisfied and also conditions (i) and (iii) of Lemma 10.3.1 are fulfilled. Let

$$L = \prod_{j=1}^{k} \prod_{i=1}^{n_j} f(X_{ji}, \theta) \quad \text{and} \quad \sigma^2 = \left[ E\left(\frac{\partial}{\partial \theta} \log f(x, \theta)\right)^2 \right]^{-1} .$$

Since $\hat{\theta}_N$ is an efficient estimator,

(10.4.1) $$\frac{N}{\sigma^2}(\hat{\theta}_N - \theta) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log f(X_{ji}, \theta) \text{ , and variance of}$$

$(\sqrt{N}(\hat{\theta}_N - \theta)) = \sigma^2$, is independent of $N$.

To obtain $g(u)$ defined by (10.3.5) we proceed as in [5]. Write

$$h_{n_j}(u) = E\, L_{n_j}^{(j)}(u)\sqrt{n_j}(\hat{\theta}_N - \theta) = E\left[(\hat{\theta}_N - \theta)\sum_{i=1}^{n_j} \psi_u(u_{ji})\right] - n_j u E(\hat{\theta}_N - \theta)$$

Since under the hypothesis $H_k$, $u_{ji}$ $(i = 1, 2, \ldots, j_n$, $j = 1, 2, \ldots, k)$ are independently identically distributed each having uniform distribution on unit interval, $h_{n_j}(u)$ can be written as

$$h_{n_j}(u) = n_j u\, E\left\{ (\hat{\theta}_N - \theta) | u_{11} < u \right\} - n_j u\, E(\hat{\theta}_N - \theta) .$$

By Lemma 3.3 of [5] and due to unbiasedness of $\hat{\theta}_N$ using (10.4.1) we have

$$h'_{n_j}(u) = n_j\, E\left\{ (\hat{\theta}_N - \theta) | u_{11} = u \right\} =$$

$$= \frac{\sigma^2 n_j}{N} E\left\{ \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{\partial}{\partial\theta} \log f(X_{ji}, \theta) \,\Big|\, u_{11} = u \right\} .$$

Since $\frac{\partial}{\partial\theta} \log f(X_{ji}, \theta)$ are independently identically distributed when the hypothesis $H_k$ is true and $u_{11} = u$ is a condition on $X_{11}$,

(10.4.2) $$h_{n_j}(u) = \sigma^2 \frac{n_j}{N} \frac{\partial}{\partial\theta} \log f(x, \theta) .$$

From (10.3.5) and (10.4.2)

$$g'(u) = \frac{1}{f(x,\theta)} \frac{\partial}{\partial\theta} f(x, \theta) = \frac{\partial}{\partial\theta} \log f(x, \theta) = \frac{N}{n_j} \sigma^2 h'_{n_j}(u)$$

which gives after integration and noting $g(0) = g(1) = 0$, that

$$\lim_{N \to \infty} h_{n_j}(u) = \sigma^2 a_j g(u) .$$

Hence

$$\rho_j(u, v) = \lim_{n_j \to \infty} \rho^{(j)}_{n_j}(u, v)$$

$$= \lim_{n_j \to \infty} [E(Z^{(j)}_{n_j}(u) Z^{(j)}_{n_j}(v) - g(v)h_{n_j}(u) - g(u)h_{n_j}(v)$$

$$+ g(u)g(v)\ E\, n_j(\hat{\theta}_N - \theta)^2]$$

$$= \min(u, v) - uv - \sigma^2 a_j g(u)g(v) .$$

Thus in the case when $\hat{\theta}_N$ is an efficient estimator Lemma 10.3.2 yields,

<u>Lemma 10.4.1.</u> If $\hat{\theta}_N$ is an efficient estimator, the stochastic process $Y_{n_j}^{(j)}(u)$ given by (10.3.7) converges in distribution to a Gaussian process $Y_j(u)$ with mean zero and covariance function $\rho_j(u, v)$ defined by

$$(10.4.3) \quad \rho_j(u, v) = \min(u, v) - uv - a_j \varphi(u)\varphi(v) , \quad \text{where}$$

$$(10.4.4) \quad \varphi(u) = \sigma g(u) .$$

Now we are in a position to find the asymptotic distribution of $C_n'^2$. It is interesting to note that the characteristic function of the limiting distribution of $C_n'^2$ involves the proportions ($a_j$'s) in which the jth population $G_j$ is sampled. Further it might be observed that the limiting distribution of $\omega'^2$ obtained by Kiefer is independent of $a_j$'s.

<u>Theorem 10.4.1.</u> If $\hat{\theta}_N$ is an unbiased efficient estimator

$$\lim_{N \to \infty} P\left\{C_n'^2 < x\right\} = P\left\{\sum_{j=1}^{k} \int_0^1 Y_j^2(u)du < x\right\} ,$$

where $Y_j(u)(j = 1, 2, \ldots, k)$ are mutually independent Gaussian processes with zero means and covariance function $\rho_j(u, v)$ given by (10.4.3).

<u>Proof.</u> Observe that $\varphi(u)$ defined by (10.4.4) is a continuous function

and $\varphi \in L_2(0, 1)$ . Let $\{\lambda_k\}$ be the eigen values and $\{f_k(u)\}$ the corresponding normalized eigen functions of the kernel $K(u, v) = \min(u, v) - uv$ . Let

$$(10.4.5) \quad \alpha_j = \int_0^1 \varphi(u)\, f_j(u)du\,, \quad \alpha^2 = \sum_{j=1}^{\infty} \lambda_j\, \alpha_j^{\,2}\,.$$

By Lemma 1 of Kac, Kiefer, Wolfowitz, as $\rho_j(u, v)$ is a covariance function $a_j\, \alpha^2 \leq 1$ . Let $W(u)$ denote Kac-Siegert representation of a Gaussian process with mean zero and covariance function $K(u, v)$ . Then proceeding as in [6], it can be verified that

$$(10.4.6) \quad Y_j(u) = W(u) - \frac{[1 - \sqrt{(1 - a_j\, \alpha^2)}]}{\alpha^2}\, \varphi(u) \sum_{k=1}^{\infty} \lambda_k\, \alpha_k \int_0^1 W(u) f_k(u)du\,,$$

is a representation of a Gaussian process $Y_j(u)$ which has mean zero and covariance function $\rho_j(u, v)$ given by (10.4.3). Since the sample functions of the process $W(u)$ are continuous with probability one, and by an application of Lemma 2 of [6], $\sum_{k=1}^{\infty} \alpha_k f_k(u)$ converges uniformly to $\varphi(u)$ , the sample functions of $Y_j(u)$ are also continuous with probability one. By Donsker's theorem [11] the required result follows. ||

Now we obtain the characteristic function of the limiting distribution of $C_n'^2$. Let $M_j(t)$ denote the characteristic function of $\int_0^1 Y_j^2(u)du$, then the characteristic function $M(t)$ of the asymptotic distribution of $C_n'^2$ is given by

$$(10.4.7) \quad M(t) = \prod_{j=1}^{k} M_j(t) .$$

Let $\{\mu_{jr}\}$ denote the eigen values of the kernel $\rho_j(u, v)$. Then $M_j(t)$ is given by,

$$M_j(t) = \prod_{r=1}^{\infty} \left(1 - \frac{2it}{\mu_{jr}}\right)^{-1/2} = [D_j(2it)]^{-1/2} ,$$

where $D_j(\lambda)$ is the F.D. associated with the positive definite kernel $\rho_j(u, v)$. The F.D. $d_1(\lambda)$ of the kernel $K(u, v) = \min(u, v) - uv$ is $d_1(\lambda) = (\sin\sqrt{\lambda})/\sqrt{\lambda}$ and its eigen values $\lambda_r$ and eigen functions $f_r(x)$ are $\lambda_r = \pi^2 r^2$, $f_r(x) = \sqrt{2}\sin(\pi r x)$. Then by Theorem 6.2 of [5] we have

$$D_j(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}} \left[1 + a_j\lambda \sum_{r=1}^{\infty} \frac{\alpha_r^2}{1 - \lambda/\lambda_r}\right] , \qquad \lambda \neq \lambda_r$$

and $\alpha_r = \sqrt{2} \int \varphi(x) \sin(\pi r x) dx$ , $r = 1, 2, \ldots$ Putting $\lambda = 2it$ the characteristic function of the limiting distribution is obtained from (10.4.7). The characteristic function depends on $a_j$ , i.e., the proportion in which jth population is sampled.

# BIBLIOGRAPHY

## Part I

[1] F. Wilcoxon, "Individual Comparisons by Rankind Methods", Biometrics, Vol. I (1945), pp. 80-83.

[2] A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill, New York, 1950.

[3] A. Wald and J. Wolfowitz, "On a Test Whether Two Samples are from the Same Population", Ann. of Math. Stat., Vol. II (1940), pp. 147-162.

[4] E. L. Lehmann, "The Power of Rank Tests", Ann. of Math. Stat., Vol. 24 (1953), pp. 23-43.

[5] A. M. Mood, "On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests", Ann. of Math. Stat., Vol. 25 (1954), pp. 514-522.

[6] Balkrishna V. Sukhatme, "On Certain Two-Sample Nonparametric Tests for Variances", Ann. of Math. Stat., Vol. 28 (1927), pp. 188-194.

[7] Balkrishna V. Sukhatme, "A Two-Sample Distribution-Free Test for Comparing Variances", Biometrika, Vol. 45 (1958), pp. 544-550.

[8] A. R. Kamat, "A Two-Sample Distribution-Free Test", Biometrika, Vol. 43 (1956), pp. 377-388.

[9] H. Cramér, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

[10] D. A. S. Fraser, Nonparametric Methods in Statistics, John Wiley & Sons, New York, 1957.

[11] R. F. Tate, "Correlation between a Discrete and Continuous Variable. Point-Biserial Correlation", Ann. of Math. Stat., Vol. 25 (1954), pp. 603-607.

[12] H. B. Mann and D. R. Whitney, "On a Test Whether One of the Two Random Variables is Stochastically Larger than the Other", Ann. of Math. Stat., Vol. 18 (1947), pp. 50-60.

[13] W. Hoeffding, "A Class of Statistics with Asymptotically Normal Distribution", Ann. of Math. Stat., Vol. 19 (1948), pp. 293-325.

[14] Fred C. Andrews, "Asymptotic Behavior of Some Rank Tests for Analysis of Variance", Ann. of Math. Stat., Vol. 25 (1954), pp. 724-736.

[15] I. Olkin and R. F. Tate, "Multivariate Correlation Models with Mixed Continuous and Discrete Variables", (unpublished).

[16] R. A. Fisher, "The General Sampling Distribution of the Multiple Correlation Coefficient", Proc. Roy. Soc., London, Ser. A., Vol. 121 (1928), pp. 654-673.

[17] E. J. Hannan, "The Asymptotic Powers of Certain Tests Based on Multiple Correlations", Journal Roy. Stat. Soc., London, Ser. B., Vol. 18 (1956), pp. 227-233.

[18] Chin Long Chiang, "On Regular Best Asymptotically Normal Estimates", Ann. of Math. Stat., Vol. 27 (1956), pp. 336-351.

[19] A. M. Mood, "Distribution Theory of Runs", Ann. of Math. Stat., Vol. 11 (1940), pp. 367-392.

[20] J. Wishart and H. O. Hirshfeld, "A Theorem Concerning the Distribution of Joins Between Line Segments", London Math. Soc. Journal, Vol. 11 (1936), p. 227.

[21] P. V. K. Iyer, "Further Contributions to the Theory of Probability Distributions of Points on a Line - I", Journal Indian Soc. Agri. Stat., Vol. 2 (1950), pp. 141-160.

## Part II

[1] H. Cramér, "On the Composition of Elementary Errors", Skandinavisk Aktuarietidskrift, Vol. 11 (1928), pp. 141-180.

[2] N. Smirnov, "Sur la distribution de $\omega^2$ ", Comptes Rendus de l'Academie des Sciences, Paris, Vol. 202 (1936), pp. 449-452.

[3] R. von Mises, Wahrscheinlichkeitsrachnung, Deuticke, Leipzig and Wein, (1931).

[4] H. Cramer, Mathematical Methods of Statistics, Princeton University Press, 1946.

[5] D. A. Darling, "The Cramér-Smirnov Test in the Parametric Case", Ann. of Math. Stat., Vol. 26 (1955), pp. 1-20.

[6] M. Kac, J. Kiefer, and J. Wolfowitz, "On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods", Ann. of Math. Stat., Vol. 26 (1955), pp. 189-211.

[7] J. Kiefer, "K-Sample Analogue of the Kolmogorov-Smirnov and Cramer-von Mises Tests", Ann. of Math. Stat., Vol. 30 (1959), pp. 420-447.

[8] A. Kolmogorov, "Confidence Limits for an Unkown Distribution Function", Ann. of Math. Stat., Vol. 12 (1941), pp. 461-463.

[9] T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes", Ann. of Math. Stat., Vol. 23 (1952), pp. 193-212.

[10] M. Kac and A. J. F. Siegert, "An Explicit Representation of a Stationary Gaussian Process", Ann. of Math. Stat., Vol. 18 (1947), pp. 438-442.

[11] M. L. Donsker, "Justification and Extension of Doob's Heuristic Approach to the Kolmogorov-Smirnov Theorems", Ann. of Math. Stat., Vol. 23 (1952), pp. 277-281.

[12] E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, Cambridge University Press, 1927.

[13] F. Riesz and B. SZ. Nagy, Functional Analysis, Fredrick Ungar Publication Co., New York.

[14] R. Courant and D. Hilbert, Methods of Mathematical Physics, Interscience Publishers, New York, 1953.