

Summer Course (advanced) for Statisticians
May-June 1961

142

LECTURE NOTES No.—4

ORDER STATISTICS

by M. Sibuya, Institute of Statistical Mathematics, Japan

AND

NON-PARAMETRIC METHODS

by J. Sethuraman, Indian Statistical Institute

RESEARCH AND TRAINING SCHOOL
INDIAN STATISTICAL INSTITUTE
203 BARRACKPORE TRUNK ROAD
CALCUTTA-35

INDIAN STATISTICAL INSTITUTE
LIBRARY

Class No. SA.01

Book No. SCS.61.4

RESTRICTED

(3)

C-4718
15.4.64

ORDER STATISTICS

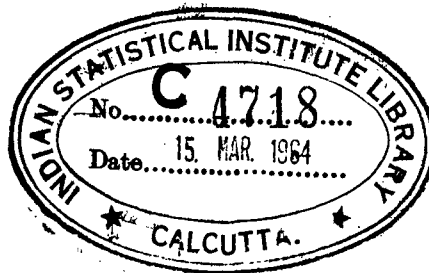
by

M. Sibuya

STATISTIC
LIBRARY

C O N T E N T S

	<u>Page</u>
1. Introduction	1
2. Moments of order statistics	3
3. Robustness of range	4
4. Systematic estimates	5
5. Limit distributions of extreme	8
6. Bivariate extreme statistics	10
7. Cutting out procedures for material with Poisson defects	12



Research and Training School
Indian Statistical Institute
203, Barrackpore Trunk Road,
Calcutta-35.

Rev-
SA-01
SCS.61.4

(3)

SOME TOPICS IN THE THEORY OF ORDER STATISTICS

1. Introduction

Take a sample $(X_1', X_2', \dots, X_n')$ from a population and rearrange the sample values according to their magnitudes

$$X_1 \leq X_2 \leq \dots \leq X_n$$

X_i is called the i -th order statistic, and the total of X_1, \dots, X_n is called the order statistics of the sample. The term order statistic in the wide sense, connotes some statistics which relate with the statistics above mentioned:

$$(X_1 + X_n)/2, (X_n - X_{n-1})/(X_n - X_1), \text{ etc.}$$

We can say that it is a combination of the techniques used in conventional statistics with those of rank order statistics.

In the first volume of the journal *Biometrika* (1901-02), Francis Galton presented the following problem. In Derby or in some race only the first and the second place winners are awarded the prizes. The sum of prize money is fixed. Then what is the rational ratio of the first to prize to the second? He suggested the ratio $(E(X_3) - E(X_1)) : (E(X_3) - E(X_2))$, where X_i is the i th order statistic of a normal sample. In the same issue of *Biometrika* Karl Pearson studied the ratio. This is one of the earliest works on the theory of order statistics.

The subsequent works of this field of statistics cover the various subjects, which may be classified in the following way.

(In most cases the normality of the population distribution is assumed).



- (i) Test for rejecting the outlying observation

Smirnov - Grubbs' test

$$(X_n - \bar{X})/s \quad s^2 = \sum(X_i - \bar{X})^2/n$$

Dixon's test

$$(X_n - X_{n-1})/(X_n - X_1)$$

- (ii) Test for normality

David-Hartley-Pearson $(X_n - X_1)/s$.

- (iii) Test for homogeneity, Multiple comparisons.

Studentized range $(\bar{X}_n - \bar{X}_1)/s$

Maximum F-ratio S_{\max}/S_{\min} .

- (iv) Short cut method

Use of range instead of standard deviation in statistical inferences

$\bar{X} - R$ chart, \bar{R}/d_2 , $(\bar{X} - \mu)/R$, $\bar{X} + CR$,

$(\bar{X} - L)/\bar{R}$, R_{\max}/R_{\min} , etc.

- (v) Systematic estimate

- (vi) Limit distribution of extreme statistics (large sample theory).

- (vii) Non-parametric methods

REFERENCES

- Wilks, S. S. : 'Order statistics', Bull. Amer. Math. Soc., 54 (1948).
Gumbel, E.J. : Statistics of Extremes, Columb. Univ. Press, (1958)
Greenberg, B.G. and A.E.Sarhan, 'Some applications of order statistics',
Bull. Inst. Intern. Stat., 30th Sess., (1957).
Sarhan, A.E. (ed).: Hand book on order statistics, Chicago Univ. Press
(to appear).

2. Moments of order statistics

The inverse function $x(F)$ of a cdf. $F(x)$ suitably defined at the discontinuous points is called a representative function. For most of usual distributions $x(F)$ is not simple, and numerical methods are necessary to obtain the moments.

$$E(X_n^i) = \int_0^1 nx^i(F)F^{n-1} dF$$

$$E(X_n X_{n-1}) = \int_0^1 \int_0^F n(n-1)x(F) x(G)(F-G)^{n-2} dGdF .$$

With the development of the electronic computers, many tables have been and will be published. For normal distribution H. Ruben (Biometrika 1954) and D. Teichrow (AMS, 1956) are useful.

If $x(F) = F$ the distribution is uniform on $(0, 1)$, the moments of order statistics are easily computed. From these values the asymptotic formulas for the moments of transformed variables are obtained.

C.E. Clark and G.T. Williams (AMS 1958)

$$p = \frac{m}{n+1} = 1 - q \quad x_p = x(p), \quad f = F'(x_p)$$

$$E(X_m) = x_p - \frac{f'}{2f^3} \frac{pq}{n+2} + \frac{3f' - ff''}{6f^5} \cdot \frac{2pq(q-p)}{(n+2)(n+3)} + O\left(\frac{1}{n^2}\right)$$

$$V(X_m) = \frac{1}{f^2} \frac{pq}{n+2} - \frac{f'}{f^4} \cdot \frac{2pq(q-p)}{(n+2)(n+3)} + O\left(\frac{1}{n^2}\right)$$

For example, the expected value of the median (say R) of N ranges in normal samples of size n can be computed from the formula:

$$E(R)/\sigma = dm + \frac{e}{N+2} + O\left(\frac{1}{N^2}\right), \text{ when } N \text{ is odd}$$

$$dm + \frac{e}{N+1} + O\left(\frac{1}{N^2}\right), \text{ when } N \text{ is even}$$

n	dm	e
2	0.95387	0.29519
3	1.588	0.162
4	1.978	0.124
5	2.257	0.108
6	2.472	0.098
7	2.645	0.093
8	2.791	0.090
9	2.916	0.086
10	3.024	0.084

3. Robustness of range

The effect of non-normality to the conventional statistic is an important problem for practical applications. Special attentions has been paid to the expected value of sample range. The values of $E(R)/\sigma$ for several distributions are shown below (Tsukibayashi, Rep. Stat. Appl. Res., JUSE 1958).

n	Rectang.	Triang.	Exp.	$\chi^2_{df=4}$	Normal	Upper bound
2	1.1547	1.1314	1.0000	1.0607	1.1284	1.1547
3	1.7321	1.6971	1.5000	1.5910	1.6926	1.7321
4	2.0785	2.0472	1.8333	1.9392	2.0588	2.0840
5	2.3094	2.2897	2.0833	2.1964	2.3259	2.3401
6	2.4744	2.4696	2.2833	2.3994	2.5344	2.5533
7	2.5981	2.6095	2.4500	2.5665	2.7044	2.7441
8	2.6943	2.7222	2.5929	2.7082	2.8472	2.9208
9	2.7713	2.8154	2.7179	2.8311	2.9700	3.0869
10	2.8343	2.8940	2.8290	2.9394	3.0775	3.2444
11	2.8868	2.9614	2.9290	3.0362	3.1729	3.3947
12	2.9312	3.0200	3.0199	3.1236	3.2585	3.5387

The expected values are affected little by the non-normality. They are close to each other and close to the upper limit, which is proved in the following theorem.

Theorem (Plackett-Moriguti): Whatever be the distribution,

$$0 < E(R)/\sigma \leq \sqrt{\frac{2(1-\varepsilon_n)}{2n-1}}$$

where $\varepsilon_n = \binom{2n-2}{n-1}^{-1}$

Equality holds when

$$x(F) = \sigma \sqrt{\frac{2n-1}{2(1-\varepsilon_n)}} F_1^{n-1} - (1-F)^{n-1} - \mu.$$

4. Systematic estimate

The distributions which belong to the class of distributions $\left\{ F(x-\mu/\sigma) : -\infty < \mu < \infty, 0 < \sigma < \infty \right\}$ are said to be of the same type.

The linear combinations of the order statistics of a sample from $F(x-\mu/\sigma)$, as estimates of μ or σ , are named systematic estimates.

Ex. Sample range, Midrange, Median, Arithmetic mean.

The systematic estimate which is unbiased and has the minimum variance is called the best s.e.

Best systematic estimate.

Let $X_1 < \dots < X_n$ be the ordered sample from the continuous cdf. $F(x)$, and let

$$E(X_i) = \alpha_i$$

$$V(X_i) = W_{ii}, V(X_i, X_j) = W_{ij}, [W_{ij}] = W.$$

Then if $Y_1 < \dots < Y_n$ is the ordered sample from the population with cdf. $F(\frac{y-\mu}{\sigma})$

$$E(Y_i) = \mu + \sigma \alpha_i$$

and the errors $\varepsilon_i = Y_i - E(Y_i)$ have the variance-covariance matrix $W\sigma^2$.

Put

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, P = \begin{pmatrix} \frac{1}{2} \alpha_1 \\ \vdots \\ \frac{1}{2} \alpha_n \end{pmatrix}, \theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$Y = P\theta + \varepsilon.$$

Then the best estimate is obtained by

$$\theta = (P' \begin{pmatrix} \square & \square \\ \square & \square \end{pmatrix} P)^{-1} P' \begin{pmatrix} \square \\ \square \end{pmatrix} Y$$

Ex. 1. Uniform distribution on $(-\sqrt{3}, \sqrt{3})$

$$\alpha_r = \frac{\sqrt{3}}{2} \frac{2r-n-1}{n+1}$$

$$W_{rs} = \frac{12r(n-s+1)}{(n+1)^2(n+2)}, \quad r \leq s$$

$$\begin{pmatrix} \square & \square \\ \square & \square \end{pmatrix} = \frac{(n+1)(n+2)}{12} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & & -1 & 2 \end{pmatrix}$$

$$\hat{\mu} = \frac{1}{2}(Y_1 + Y_n)$$

$$\hat{\sigma} = \frac{n+1}{2\sqrt{3}(n-1)} (Y_n - Y_1).$$

Ex. 2. Normal distribution, H. Godwin (Biometrika, 1949)

$$\hat{\mu} = \bar{Y} \quad \hat{\sigma} = \sum s_i Y_i$$

$$n = 5 \quad s_5, s_1 = \pm 0.37238$$

$$s_4, s_2 = \pm 0.13521$$

$$s_0 = 0$$

$$V(\hat{\sigma}) = 0.13332 \quad V(s/\hat{\sigma}_2) = 0.13177.$$

Ex. 3. Exponential, $F(x) = 1 - \exp.(-x)$

$$\hat{\mu} = (nY_1 - \bar{Y})/(n-1), \quad V(\hat{\mu}) = 1/n(n-1)$$

$$\hat{\sigma} = (\bar{Y} - Y_1)n/(n-1), \quad V(\hat{\sigma}) = 1/(n-1)$$

$$E(Y) = \hat{\mu} + \hat{\sigma} = \bar{Y}$$

Other examples

F.Downton (AMS, 1954) single triangle

A.E. Sarhan (AMS, 1954) double triangle, U shape
Parabolic, skewed.

The case of single parameter

$F(Y/\sigma)$

$$\hat{\sigma} = (\alpha' \begin{bmatrix} \bar{Y} \\ \alpha \end{bmatrix} \alpha)^{-1} \alpha' \begin{bmatrix} \bar{Y} \\ \alpha \end{bmatrix} Y.$$

In the normal sample of size 5, if upper 2 observations are lost.

$$\hat{\mu} = -0.06377 Y_1 + 0.14983 Y_2 + 0.91395 Y_3$$

$$\hat{\sigma} = -0.76958 Y_1 + 0.21212 Y_2 + 0.98170 Y_3$$

$$V(\hat{\mu}) = 0.61123$$

$$V(\hat{\sigma}) = 0.69571.$$

A.K. Gupta, (Biometrika 1952)

$$\hat{\mu} = \sum b_1 Y_i, \quad \hat{\sigma} = \sum c_i Y_i, \quad i = r_1 + 1, \dots, n - r_2$$

$$b_i = \frac{1}{n - r_1 - r_2} - \frac{\bar{u} (u_i - \bar{u})}{\sum_j (u_j - \bar{u})^2},$$

$$c_i = \frac{(u_i - \bar{u})}{\sum_j (u_j - \bar{u})^2}, \quad \bar{u} = \frac{1}{n - r_1 - r_2} \sum_{j=r_1+1}^{n-r_2} u_j$$

Theorem (Lloyd, Biometrika 1952)

$$V(\hat{\mu}) \leq \sigma^2/n$$

Equality is true when $W \underline{1} = \underline{1}$.

5. The limit distributions of the extreme observations

The study of the limit distribution of the largest value is nothing but the investigation of the asymptotic behaviours of the cdf $F^n(x)$ when n increases infinitely. But, to understand its meaning from the view point of statistics, the following characteristics are useful. The 'characteristic largest value', $u_n: F(u_n) = 1 - 1/n$.

The 'extremal intensity function', $\alpha_n: \alpha_n = \mu(u_n) = nf(u_n)$, where $\mu(x)$ is the 'intensity function'. $\mu(x) = f(x)/(1-F(x))$.

Three initial distributions; Exponential, Pareto, Limited distributions are typical.

1. Exponential

Initial distribution	$1 - \exp(-\alpha x), \alpha > 0, x \geq 0$
Characteristic largest value	$1/n \exp(\alpha u_n) = 1$
Asymptotic distribution	$\exp(-\exp \alpha(x - u_n))$

2. Pareto

Initial distribution	$1 - (z - \varepsilon)^{-k}, \varepsilon \geq 0, k \geq 1,$ $z > 1 + \varepsilon$
Characteristic largest value	$n(u_n - \varepsilon)^{-k} = 1$
Asymptotic distribution	$\exp(-(z - \varepsilon/u_n - \varepsilon)^{-k})$

3. Limited

Initial distribution	$1 - (w - x)^k, w > 0, k > 1$ $w - 1 < x \leq w$
Characteristic largest value	$n(w - u_n)^k = 1$
Asymptotic distribution	$\exp(-(w - x/w - u_n)^k)$

Such important distributions as logistic, normal, chi-square, logarithmic normal, belong to the exponential type, although for some of them the limit is approached slowly.

Three asymptotic distributions are derived from the stability condition (Fisher-Tippett, 1928)

$$F^n(x) = F(a_n x + b_n).$$

The necessary and sufficient conditions for the existence of three distributions of largest values are obtained by B. Gnedenko (1943).

The first asymptotic, the double exponential distribution, is fit to investigate many phenomena. For the applications, the probability paper is of great use.

6. Bivariate extreme statistics

Safe blast: When dynamite is used to blow-up rocks, many sticks of dynamite are charged in separate positions and exploded by electric detonators which are connected in a series circuit. An electric detonator consists of two parts, that is, the initiating explosive around an electric bridge and the larger charge of sensitive high explosive. Being heated by the bridge the initiating explosive begins to fire, and in next the larger charge is fired by flying sparks. The excitation time X , which is necessary for ignition, and the bursting time Y , the time from switching-on to the burst, are the main characteristics.

When n detonators are connected if X_{\max} is larger than Y_{\min} , at least one stick of dynamite misses fire. Then, the maker of detonators must control the process, so that the probability to miss fire is small enough.

Limit distribution of (X_{\max}, Y_{\max}) is discussed. The condition under which X_{\max} and Y_{\max} are mutually asymptotically independent is obtained. In the bivariate case, types of the limit joint distributions are various. The possible forms are found.

Definition: Dependence function $\bar{J}(G, H)$:

$$F(x, y) = \bar{J}(G(x), H(y))G(x)H(y)$$

Theorem

$$L(G, H) \leq \bar{J}(G, H) \leq U(G, H)$$

$$U(G, H) = \min(1/G, 1/H)$$

$$L(G, H) = \max(0, (G+H-1)/GH)$$

Theorem.

If the function $P(G(x), H(y)) = \Pr(X > x, Y > y)$ converges to zero as G and H approach 1 in such a way that

$$P(1-s, 1-s) = o(s)$$

X_{\max} and Y_{\max} are asymptotically independent, and vice versa.

Theorem.

If $P(1-s, 1-s) = s - o(s)$, the dependence function of (X_{\max}, X_{\min}) converges to $U(G, H)$.

Definition.

The distribution of (X_{\max}, Y_{\max}) is said to be stable with respect to dependence, if

$$\bar{J}^n(G^{1/n}, H^{1/n}) = \bar{J}(G, H)$$

for all n .



Theorem.

$(X \max, Y \max)$ is stable, if and only if

$$\int_0^{\infty} (G, H) = G^{\lambda(\log H / \log G)}$$

where $\lambda(\alpha)$, $0 \leq \alpha < \infty$ is a convex function.

7. Cutting out procedures for material with Poisson defects

Material with one-dimensional extension is sent to a market, after being cut out into parts of a unit length. If a cut out part contains more than c (say) defects on it, it is rejected. Four procedures are suggested.

Proc. 1. Simple cutting out. Starting from the end point we cut out the intervals of unit length, and inspecting defects we reject the intervals with more than c defects.

Proc. 2. Sequential cutting out. From the end point we measure the interval of unit length and count the number of defects. If it is less than or equal to c , we cut out the interval, and otherwise we move the origin of measurement to the position of first defect and count again the number of defects on the interval of unit length. In this way we continue to measure the interval of unit length until it does not contain more than c defects.

Proc. 3. Cutting out the interval of length l ($1 < l < 2$). We cut out intervals in the same way as in Proc. 1, but those of length l . If the defects are situated near the end points, we cut the end

parts to get a unit interval containing not more than c defects.

Proc. 4. Cutting out the interval of length h ($2 < h < 3$).

In the same way as in Proc. 3, we cut out, if possible two unit intervals to be accepted, and if it is not possible to get the two, we try to obtain an interval.

As a measure of performance of these procedures the yield, that is the ratio of the expected length of accepted parts ^{to} that of the original material, is computed for each procedure. The defects are assumed to be distributed according to the Poisson process with parameter λ .

Analysis

Proc. 1.

The yield is $P_1 = \sum_{m=0}^c e^{-\lambda} \lambda^m / m!$

Proc. 2.

Let X_i be the distance between the i th and $(i+1)$ th defects, and the event A_i be

$$A_i : T_i > 1 > T_i - 1, \text{ where } T_i = \max_{1 \leq j \leq i} \sum_{k=j}^{j+c} X_k.$$

Then the yield is

$$P_2 = [E \left\{ \sum_{j=0}^{i-1} X_j + 1; A_i \right\} \left\{ \Pr \{A_i\} \right\}]^{-1}$$

when $c=0$

$$P_2 = \lambda(e^\lambda - 1)$$

Proc. 3.

Let $0 = U_0 < U_1 < \dots < U_m < U_{m+1} = 1$ be the order statistics from the uniform distribution on $(0, 1)$.

$$p_3(m) = \begin{cases} 1 & m \leq c \\ \Pr \left\{ \max_{c+1 \leq j \leq m+1} (U_j - U_{j-c-1}) \geq 1/\ell \right\}, & m > c \end{cases}$$

The yield is

$$P_3 = \frac{1}{\ell} \sum_{m=0}^{\infty} p_3(m) \frac{e^{-\lambda \ell} (\lambda \ell)^m}{m!}$$

$$c = 0 \quad P_3 = e^{-\lambda} \left\{ \lambda + (1 - \lambda)/\ell \right\}.$$

$$c = 1 \quad P_3 = \frac{1}{\ell} \left[\left\{ \lambda^2(\ell - 1) + \lambda(2 - \ell) + 2 \right\} e^{-\lambda} - e^{-\lambda \ell} \right].$$

Proc. 4.

$$\text{The event } A_i : \begin{cases} U_{k+c+1} - U_k \leq 1/\ell, & k = 0, 1, \dots, i-1 \\ U_{i+c+1} - U_i > 1/\ell. \end{cases}$$

$$\text{The event } B_j : \begin{cases} U_k - U_{k-c-1} \leq 1/\ell & k = m+1, m, \dots, j+1, \\ U_j - U_{j-c-1} > 1/\ell \end{cases}$$

$$p_4(m) = \Pr \left\{ A_i \cap B_j \cap \left\{ U_j - U_i > 2/\ell \right\} \right\}$$

$$p_4'(m) = \Pr \left\{ A_i \cap B_j \cap \left\{ U_j - U_i \leq 2/\ell \right\} \right\}$$

$$P_4 = \frac{1}{\ell} \sum_{m=0}^{\infty} (2p_4(m) + p_4'(m)) \frac{e^{-\lambda \ell} (\lambda \ell)^m}{m!}$$

$c = 0$

$$P_4 = e^{-\lambda} \left\{ \lambda + \frac{1 - \lambda}{\ell} \right\} + e^{-2\lambda} \left\{ \lambda + \frac{1 - 2\lambda}{\ell} \right\}.$$

For $c = 0, 1, 2$ numerical values of the yields P_1, \dots, P_4 were computed. Some of them were obtained by Monte Carlo methods.

NON-PARAMETRIC METHODS

By

J.Sethuraman

C O N T E N T S

	Page
Chapter I Introduction	1
Part I Methods of Non-parametric Analysis	5-50
Chapter II General theories of inference	5
Chapter III Applications to non-parametric methods	25
Part II Limit Distributions	51-78
Chapter IV Limiting distributions	51
Part III Recent Work at the I.S.I.	79-124
Chapter V Estimates of the density function of a distribution distribution	79
Chapter VI Fixed interval analysis and fractile analysis	87
Tables	120
References	125

Research and Training School
Indian Statistical Institute
203 Barrackpore Trunk Road
Calcutta

CHAPTER I

INTRODUCTION

In recent years there has been an enormous growth in the methods of non-parametric analysis. The bibliography on non-parametric statistics and other related topics by Savage (1953) runs to 63 pages in the Journal of the American Statistical Association, and this gives a fair idea of the recent growth in the subject.

It is well nigh impossible to examine and deal with this huge literature in the course of our lectures. The subject of non-parametric methods can be broadly divided into two parts - its 'test of significance' aspect and its 'analytical' aspect. We will mainly deal with the 'analytical' (in parts I and II) and briefly touch upon the recent discoveries made at the I.S.I. (in part III), which normally belong to the 'test of significance' domain.

The 'test of significance' approach may also be called the 'descriptive' approach. In this approach, we have a null hypothesis in mind and set of alternatives, at times only vaguely defined, against which we wish to propose some tests. The tests are usually based on 'measures of divergence' computed from the sample, which have the property that, measured from the population they assume a value a_0 for the distributions of the null hypothesis and a value $\neq a_0$ for

distributions in the alternative. Large deviations of the 'measure of divergence' from a_0 form the critical region. Such tests are usually easy to propose. These tests become 'non-parametric' or 'distribution free' if the exact or at least the approximate sampling distributions of the test function is independent of the parent distribution. The evaluation of the sampling distribution of these tests is the more difficult problem in many instances.

One of the first instances of this approach is the famous discrepancy χ^2 of Karl Pearson (1900). The discrepancy χ^2 statistic is a 'measure of divergence' between the sample and the distribution of the null hypothesis; and the limiting distribution of this statistic under the null hypothesis was established to be a χ^2 distribution. Other great pioneers of this approach are Fisher (1935), Pitman (1937), (1937a), (1937b), Mosteller (1948) and others. An excellent account of the works of these and other pioneers is given in Siegal's (1956) book.

In Part III of our lectures we shall deal with some new methods of estimation of the density function by Parthasarathy (1961) and the methods of fixed interval analysis and fractile analysis first formulated by Mahalanobis (1958) and (1961). Work on the latter topic has been done by Takeuchi (1961) and the author (1961), (1961a). The results in this connection are still in their infancy and these actually belong to the domain of the 'tests of significance'.

It is to be admitted that the 'tests of significance approach is the earliest phase of all non-parametric procedures. The success of this phase depends on the penetrating intuition that should be exercised in order that the tests proposed actually possess certain optimum properties. Thus it is very necessary that we have a general theory of non-parametric methods which starts with a general set up, defines some optimum properties of tests, examines whether the tests based on intuition possess the optimum properties and develops general techniques of obtaining tests with these optimality criteria. Such an approach could be termed as the 'analytical' approach. For developing this approach it is necessary, at the outset, to review the techniques of the 'analytic' approach to parametric problems, preferably in a more general set up, and then, to proceed to apply them in non-parametric problems. Such a review can be found in the notes by Lehmann (1949), (1950) and Lehmann (1950a). The special techniques that are suited to non-parametric problems were developed by the pioneers Lehmann and Stein (1949), Lehmann (1950a), (1951), (1953) etc. An excellent and satisfying account of these topics is given in Fraser (1957). We hope to follow this book in our lectures. This comprises part I of our lectures.

The theory of limiting distributions plays a central role in probability theory. Under the name of asymptotic distribution theory it is very useful in Statistics. The parameter, n , that tends to

infinity in these limit theorems is usually the sample size and in those instances where small sample distributions are difficult to obtain or compute, the asymptotic distributions which are easier to obtain, serve as convenient approximations. Thus the general theory of limit distributions can be very helpful in non-parametric problems also. There are also some limit theorems that have been developed specially for non-parametric problems. The results of Hoeffding and Robbins (1948), Hoeffding (1948), (1951), Wald and Wolfowitz (1944), Noether (1949), Mood (1940), etc., can be quoted in this context. The results of the present author (1961b) are also useful in many situations. We shall deal with these results in Part II of our lectures.

PART I

METHODS OF NON-PARAMETRIC ANALYSIS

CHAPTER II

GENERAL THEORIES OF INFERENCE

2.1. The decision theoretic approach.

The general problem of statistical inference can be mathematically summarised as follows $(\mathcal{X}, \mathcal{S})$ is the sample space together with an associated σ -field of events. P_θ for each $\theta \in \Theta$ is a probability distribution. (We shall adhere to this notation throughout this book.) On the basis of a random observation x some inference has to be drawn about θ , the classifying index of the true probability distribution on $(\mathcal{X}, \mathcal{S})$. For this purpose we can define \mathcal{A} the class of all inferences, a , which we would like to make. As soon as x is observed we make an inference or a decision, $d(x)$, which is a member of \mathcal{A} . Thus the procedure of inference that we adopt can be viewed as a function d from \mathcal{X} into \mathcal{A} . d is called a decision function. It is also possible that the statistician chooses one of the inferences from \mathcal{A} at random when x is observed instead of choosing a fixed member $d(x)$. This may be called a randomised decision function, denoted by $\delta_x(\cdot)$, which for each x , is a measure on $(\mathcal{A}, \mathcal{J})$ where \mathcal{J} is some suitable σ -field of subsets on \mathcal{A} .

The statistician will then take into account the loss involved in making an inference a when the true situation is that denoted by θ . For this we have the loss function, denoted by $W(a, \theta)$ a real valued non-negative function on $\mathcal{A} \times \Theta$. The average loss or the

risk, $R(d, \theta)$, in using a decision function when θ is true is

$$R(d, \theta) = E_{\theta}(W(d(X), \theta)) = \int_{\mathcal{X}} W(d(x), \theta) dP_{\theta},$$

and $R(\delta, \theta)$, the risk in using $\delta_x(\cdot)$ is given by

$$R(\delta, \theta) = \int_{\mathcal{X}} \int_{\mathcal{A}} W(a, \theta) \delta_x(da) P_{\theta}(dx)$$

where it should be assumed that $\delta_x(A)$ for each A and $d(x)$ are measurable functions.

The performance of a decision function is judged by its risk function.

Definition: A decision function δ is said to be a uniformly minimum risk decision function (UMRDF) if $R(\delta, \theta) \leq R(\delta', \theta)$ for any decision function δ' and any θ .

Such UMRDF do not exist in general. There are certain problems, where we can produce UMRDF within a suitably restricted class of decision functions.

Definition: Let $M(\delta) = \sup_{\theta} R(\delta, \theta)$. δ is said to be a minimax decision function (MDF) if $M(\delta) \leq M(\delta')$ for any other decision function δ' .

Quite often $M(\delta) = \infty$ for all decision functions δ so that the above definition does not provide an optimal decision function. There are other problems where non-trivial MDF's exist. The minimax property stated here is related to the most stringent property defined in section 2.3.

Definition: A decision function δ is said to be inadmissible if

there exists another decision function δ' such that

$$R(\delta, \theta) \geq R(\delta', \theta) \quad \text{for all } \theta$$

and $R(\delta, \theta) > R(\delta', \theta_0)$ for some θ_0

A decision function is said to be admissible if it is not inadmissible.

Admissibility should be a property possessed by any decision function that is to be used in practice. There are several techniques of obtaining admissible decision functions, but we shall not relate them here, for admissibility alone is not an optimum property.

Definition: A class D of decision functions is said to be a complete class if for every $\delta \notin D$ there exists a $\delta' \in D$ such that

$$R_{\delta}(\theta) \geq R_{\delta'}(\theta) \quad \text{for all } \theta.$$

One of the important problems of decision theory is to obtain a complete class. One can refer to some of Wald's results in this connection, for example see Wald (1950).

All the main problems of inference, namely point estimation, testing of hypothesis, confidence intervals and tolerance regions can be treated by the general approach outlined above by suitably choosing \mathcal{Q} and $W(a, \theta)$. As an illustration, \mathcal{Q} can be taken to be the range of $g(\theta)$ in the point estimation of a real valued function $g(\theta)$. $W(a, \theta)$ can then be chosen to be $(a - g(\theta))^2$ if the mean square deviation is to be the criterion of the performance of our decision procedure.

The following lemma shows that randomised decision procedures can be ignored in the above point estimation problem and other similar problems. Let \mathcal{Q} be an interval in R_k , the Euclidean space of k -dimensions. Let the loss function $W(a, \theta)$ be a convex (downwards) function of a for each θ .

Lemma 2.1. In this case corresponding to each randomised decision function δ there exists a non-randomised decision function d such that

$$R(d, \theta) \leq R(\delta, \theta) \quad \text{for all } \theta.$$

Proof: Define $d(x) = \int_D a \delta_x(da)$.

$$\begin{aligned} \text{Now } R(d, \theta) &= \int_{\mathcal{X}} W(d(X), \theta) dP_\theta = \int_{\mathcal{X}} [W(\int_D a \delta_x(da), \theta)] dP_\theta \\ &\leq \int_{\mathcal{X}} [\int_D W(a, \theta) \delta_x(da)] dP_\theta \\ &= R(\delta, \theta). \end{aligned}$$

2.2. The estimation of real valued functions of the parameter.

The problem here is the estimation of a real valued function $g(\theta)$ of θ on the basis of an observation x . \mathcal{Q} is the range of $g(\theta)$. Let the loss function $W(a, \theta)$ be convex (downwards) for each θ . This is satisfied by the conventional choices of the loss function namely, $W(a, \theta) = |a - g(\theta)|^p$, $p \geq 1$. Theorem 2.4, otherwise called the Rao-Blackwell theorem with the Lehmann-Scheffe extension is the fundamental result in the above situation.

Definition: $t(x)$ is said to be a sufficient statistic for the family of distributions $\{P_\theta\}$ if the conditional distribution of x given $t(x) = t$ is independent of θ (a.e. P_θ).

Definition: A statistic $t(x)$ is said to be complete with respect to a family of distributions $\{P_\theta\}$ on (X, S) if for any function f of t ,

$$E_\theta[f(t(X))] = \int_X f(t(x)) dP_\theta = 0 \text{ for all } \theta$$

implies that $f(t(x)) = 0$ a.e. $\{P_\theta\}$.

Definition: A statistic $s(x)$ is said to be an unbiased estimate of $g(\theta)$ if

$$E_\theta(s(X)) = g(\theta) \text{ for all } \theta.$$

Lemma 2.2. If $s(x)$ is unbiased for $g(\theta)$ and $t(x)$ is sufficient for $\{P_\theta\}$ then there exists a function u of t such that

$$R(s, \theta) \geq R(u, \theta) \text{ for all } \theta.$$

Proof:

$$\text{Let } u(t) = E_\theta(s(X) | t(x) = t). \quad \dots \quad (1)$$

It is to be noted that the right hand side does not depend on θ . It is obvious that u is unbiased for $g(\theta)$. Now from (1) we have

$$W(u(t), \theta) \leq E(W(s(X), \theta) | t(x) = t)$$

Taking the expectations over the distribution of t we have

$$R(U, \theta) \leq R(s, \theta).$$

Corollary 2.3. The class of estimates depending only on t forms

a complete class among the class of all estimates.

Theorem 2.4. Let $t(x)$ be a statistic complete and sufficient with respect to $\{P_\theta\}$. If $s(x)$ is unbiased for $g(\theta)$ there exists a statistic u depending only on t which is the minimum risk unbiased estimate (MRUE) of $g(\theta)$.

Proof:- Let u be defined as in (1). Now let $s'(x)$ be any unbiased estimate of $g(\theta)$. Define $u'(t) = E(s'(X) | t(x) = t)$. Then $R(s', \theta) \geq R(u', \theta)$ and u' is unbiased for $g(\theta)$. Now $E_\theta(u(T) - u'(T)) = g(\theta) - g(\theta) = 0$ for all θ . Hence $u(t) = u'(t)$ a.e. Thus $R(u, \theta) = R(u', \theta) \leq R(s', \theta)$. This completes the proof.

Corollary 2.5. Every function of t is the MRUE of its expectation.

Corollary 2.6. Every function of θ which is unbiasedly estimable possesses a MRUE.

2.3. The theory of testing hypothesis.

In the theory of testing hypothesis there are only two elements, a_1 and a_2 , in the space of actions. a_1 corresponds to the statement ' θ is in w ' and a_2 to the statement ' θ is in \bar{w} - w ,' where w is some fixed subset in $\bar{\Omega}$. One of the actions (say a_1) is usually called the null hypothesis, H , and the other is called the alternative, A . In practice the null hypothesis corresponds to the situation sanctioned by previous experience and the alternative takes into account the possible changes that might have set in the old situation. The problem is to

find out whether such a change has occurred. Thus the hypothesis and alternative are not set on a common footing. We would like to seek some protection against making the decision a_2 too often when actually $\theta \in W$. This is made by fixing 'The size' of our decision procedure as defined by (3).

The special nature of \mathcal{Q} makes the loss assume a simple form:

$$W(a_1, \theta) = \begin{cases} 0 & \text{if } \theta \in W \\ W_1(\theta) & \text{if } \theta \in \bar{W} \end{cases}$$

$$W(a_2, \theta) = \begin{cases} W_2(\theta) & \text{if } \theta \in W \\ 0 & \text{if } \theta \in \bar{W} \end{cases}$$

Any non-randomised decision procedure $d(x)$ is of the form $d(x) = a_1$ if $x \in X - W$ and $d(x) = a_2$ if $x \in W$. Such a procedure involves the acceptance of H when $x \in X - W$ and rejection of H when $x \in W$. W is called the critical region. A randomised decision function δ_x is for each x a distribution assigning probabilities $1 - \phi(x)$ and $\phi(x)$ to the points a_1 and a_2 respectively. $\phi(x)$ is the probability with which H is rejected when x is observed and is called a test function. We can easily see

$$R(\delta, \theta) = R(\phi, \theta) = \begin{cases} \beta(\phi, \theta) W_2(\theta) & \text{if } \theta \in W \\ (1 - \beta(\phi, \theta)) W_1(\theta) & \text{if } \theta \in \bar{W} \end{cases} \dots(2)$$

where $\beta(\phi, \theta) = E_{\theta}(\phi(X)) = \int \phi(x) dP_{\theta}$.

$\beta(\phi, \theta)$ is called the power of the test ϕ and it is the probability of rejecting H under the distribution P_θ . From (2) we see that we should minimise $\beta(\phi, \theta)$ for each $\theta \in W$ and maximise $\beta(\phi, \theta)$ for each $\theta \in \bar{W}$ to obtain tests with smaller average loss.

Definition: The size, α , of a test ϕ is defined by the relation

$$\alpha = \sup_{\theta \in W} \beta(\phi, \theta) \quad \dots \quad (3)$$

Definition: If C is a certain class of tests, then $\phi^* \in C$ is said to be most powerful in C if

$$\beta(\phi^*, \theta) \geq \beta(\phi, \theta) \text{ for each } \theta \in \bar{W} \text{ and each } \phi \text{ in } C.$$

The following theorem called the Neyman-Pearson Lemma is of fundamental importance in the theory of testing hypothesis. All the techniques in this theory are consequences of this lemma.

Let X be a random variable with a density function $f(x, \theta)$, $\theta \in \bar{W} = \{0, 1\}$. $H: \theta = 0$ $A: \theta = 1$. (This notation, to be used repeatedly hereafter, means that $H = \{\theta = 0\}$ and $A = \{\theta = 1\}$.).

Theorem 2.7. There exists a size α test ϕ^* which is most powerful in the class, C_α , of all tests of size α . Constants a and c , $0 \leq a < 1$, $0 \leq c \leq \infty$ can be found so that ϕ^* can be written in the form

$$\phi^*(x) = \begin{cases} 1 & \text{if } q(x) > c \\ a & \text{if } q(x) = c \\ 0 & \text{if } q(x) < c \end{cases} \quad \dots(4)$$

where $g(x) = \frac{f(x, 1)}{f(x, 0)}$.

Proof: The proof of this important lemma is elementary and is found in many text books; for instance in Fraser (1957), Lehmann (1949), and so is omitted.

When the hypothesis or alternative contains only a single point it is said to be simple. Otherwise it is said to be composite. The Neyman -Pearson lemma finds the most powerful (m.p.) size α test when both the hypothesis and alternative are simple. We shall see below how we can adapt this lemma to yield m.p. size α tests in other situations.

Let the hypothesis, $H : \theta \in w$, be composite and the alternative, $A : \theta \in \bar{w} = \{\theta_1\}$, be simple. C_α is the class of all tests of size α for this problem. In such situations we can sometimes use the technique of least favourable distributions to obtain the m.p. test in C_α . Let us take a point θ_0 in w so that P_{θ_0} resembles P_{θ_1} very closely. Let $C_\alpha(\theta_0)$ denote the class of all tests of size α for the problem $H : \theta = \theta_0, A : \theta = \theta_1$. Let ϕ^* be the m.p. test in $C_\alpha(\theta_0)$. If ϕ^* can be shown to be in C_α then ϕ^* is m.p. in C_α . In such a case θ_0 is called a least favourable distribution in w . In certain other situations we may use a mixture of the distributions in w , obtained from some a priori distribution, to play the role of θ_0 . In such instances the a priori distribution is defined on a suitable σ -field

of sets in w which makes $P_{\theta}(B)$ a measurable function of θ for each $B \in \mathcal{S}$.

Definition: A test ϕ is said to be unbiased if $\beta(\phi, \theta) \geq \alpha$ for all $\theta \in \Omega - w$ where α is the size of the test.

Unbiasedness of a test is a desirable property in as much as the probability of rejection of the null hypothesis at a distribution in the alternative should not be smaller than at a distribution in the hypothesis. A m.p. size α test, whenever it exists, is always unbiased since it is more powerful than the test $\phi(x)$ which is identically equal to α . It often happens that in situations with a composite alternative there is no m.p. test in C_{α} , the class of tests of size α , but there exists a m.p. test in C_{α}^U , the class of unbiased size α tests. Thus unbiasedness may also be helpful to restrict the class of tests in the search of a m.p. test.

The condition of unbiasedness is not easily algebraically expressed. There is another condition, which may not be a very desirable property, but is easily amenable to algebraic methods, and this is the similarity of a test.

Definition: A test ϕ is said to be similar of size α if

$$\beta(\phi, \theta) = \alpha \quad \text{for all } \theta \in w \quad \dots \quad (5)$$

Consider the situation where the distributions $\{P_{\theta}\}$, $\theta \in w$ admit of a sufficient statistic $t(x) = t$.

Definition: A test $\phi(x)$ is said to be of Neyman structure, if

$$E_{\theta}(\phi(X) | t(x) = t) = \alpha \text{ for all } t \text{ and for all } \theta \in w \dots \quad (6)$$

We note that for each t , the conditional distributions $P_{\theta}(\cdot | t(x)=t) = Q_{\theta}^t$ can be indexed by θ varying in $\{0, \bigcup_{\theta \in w} \theta\}$, where $Q_0^t = Q_{\theta}^t$ for some θ (and therefore for all θ) in w . Thus if we use tests of Neyman structure, the problem of testing would, for each t , correspond to the problem $H : \theta = 0, A : \theta \in \bigcup_{\theta \in w} \theta$, that is a problem with a simple hypothesis instead of a composite hypothesis. The following theorem (Theorem 2.8) shows that under certain conditions the restriction to tests of Neyman structure is equivalent to the restriction to similar tests.

Definition: A statistic $t(x)$ is boundedly complete if $|t(x)| < K$ and $E_{\theta}(t(X)) = 0$ implies $t(x) = 0$ a.e. $\{P_{\theta}\}$.

Theorem 2.8. A necessary and sufficient condition that all similar tests are of Neyman structure (with regard to the sufficient statistic $t(x)$) is that $t(x)$ is boundedly complete.

Proof: Let $t(x)$ be boundedly complete and $\phi(x)$ be a similar test of size α . Let $\psi(t) = E_{\theta}(\phi(X) | t(x) = t), \theta \in w$. Then $0 \leq \psi(t) \leq 1$ and $E_{\theta}(\psi(T)) = \alpha$ for $\theta \in w$. Since $t(x)$ is boundedly complete we must have $\psi(t) - \alpha = 0$ a.e. Q_0^t , and hence $\phi(x)$ is of Neyman structure.

Let $\psi(t)$ be a function of t such that $|\psi(t)| < K$ and $E_{\theta}(\psi(t)) = 0$ for all $\theta \in \omega$. Consider the test $\frac{\psi(t) + K}{2K} = \phi(x)$ which is similar and of size $\frac{1}{2}$. So it should be of Neyman structure, that is, $\frac{\psi(t) + K}{2K} = \frac{1}{2}$. Thus $\psi(t) = 0$ a.e. Q_0^t .

Consider the sample space (X, S) and the family of measures $\{P_{\theta}\}$ $\theta \in \Omega$. Let G be a class of 1-1 transformations of X onto itself. The product $g_1 g_2$ of two transformations g_1 and g_2 is defined as $(g_1 g_2)(x) = g_1(g_2 x)$. Let G be a group of transformations under this operation. If X has the distribution P_{θ} then gX has the distribution $P(gX \in B) = P_{\theta}(g^{-1}B) = P_{\theta} g^{-1}$. For each θ , let this distribution be $P_{g\theta}$ and belong to the distributions indexed by Ω . Thus corresponding to each g , there is a transformation \bar{g} of Ω into Ω that takes θ into $\bar{g}\theta$. Let us assume that this transformation is 1-1 and $\bar{G} = \{\bar{g}\}$ be a group of transformations. Let the testing problem be $H: \theta \in \omega, A: \theta \in \Omega - \omega$. Further let $\bar{g}\omega = \omega$ for each \bar{g} in \bar{G} . Then G is said to constitute an invariant group of transformations for the above testing problem.

The principle of invariance states that in such a situation we should restrict ourselves to invariant tests, that is to tests $\phi(x)$ with the property, $\phi(x) = \phi(gx)$ for all $g \in G$.

An example will illustrate the idea of invariance. If X has a normal distribution with mean θ and variance 1 then to test $H: \theta = 0, A: \theta > 0$, one should get the same results whether we measure X or aX

for any a , $0 < a < \infty$. This natural condition is generalised as the addition of invariance.

We shall see how the principle of invariance effects a reduction in the sample space, the space of the parameter, the hypothesis and the alternative. After this reduction is carried out we should apply other methods to obtain tests with desired optimum properties. An optimum property possessed by invariant tests forms the content of the famous Hunt and Stein Lemma (Theorem 2.16 and Corollary 2.17).

We now partition \mathcal{X} into a union of disjoint invariant sets. Thus starting from a point x we form the set $G_x = \{gx\}$, $g \in G$. \mathcal{X} can be written as $\bigcup_{x \in \mathcal{X}} G_x$. We can pick out x'_t to form $\{x_t\}$ indexed by $t \in T$ such that $\bigcup_{t \in T} G_{x_t} = \mathcal{X}$ and $G_{x_t} \cap G_{x_{t'}} = \emptyset$ if $t \neq t'$. We can have a similar partitioning of Ω by means of the transformation group \bar{G} . Let $\Omega = \bigcup_{\mu \in M} \bar{G}_{\theta_\mu}$, $\mu \in M$, where $\bar{G}_\theta = \{\bar{g}\theta\}$, $\bar{g} \in \bar{G}$; $\bar{G}_{\theta_\mu} \cap \bar{G}_{\theta_{\mu'}} = \emptyset$ if $\mu \neq \mu'$. Since w is invariant under G , $w = \bigcup_{\mu \in K} \bar{G}_{\theta_\mu}$ where K is some subset of M . A function $f(x)$ on \mathcal{X} is said to be invariant if $f(x) = f(gx)$ for each $g \in G$. The following two lemmas concern invariant functions.

Lemma 2.9. Any invariant function on \mathcal{X} is constant on each G_{x_t} , that is, depends on x only through t . The proof is elementary.

Lemma 2.10. The distribution of an invariant function on \mathcal{X} depends on θ only through μ .

Proof: Let $f(x)$ be an invariant function. Then $f(x) = f(gx)$

for each x .

$$\begin{aligned} P_{\theta}(f(x) \in B) &= Q_{\theta}^f(B) \\ &= P_{\theta}(f(gx) \in B) = P_{\theta}(f(y) \in g^{-1}B) \\ &= P_{g\theta}(f(y) \in B) = Q_{g\theta}^f \end{aligned}$$

Thus $Q_{\theta}^f = Q_{g\theta}^f$. Hence the lemma.

The principle of invariance, the invariance of H and the above two lemmas show clearly how the testing problem can be reduced to the following. Let \mathcal{U} be the σ -field induced by the invariant function $t(x)$. We have the new sample space (T, \mathcal{U}) and probability measures $Q_{\mu}, \mu \in M$, where $Q_{\mu}(B)$ is $P_{\theta}\{t(x) \in B\}$ for some θ , and therefore for any, θ in \bar{G}_{μ} . The hypothesis and alternative are now reduced to K and $M-K$ respectively.

Definition: The power envelope of tests of size α is defined as

$$Y(\theta) = \sup_{\phi \in C_{\alpha}} \beta(\phi, \theta) \text{ where } C_{\alpha} \text{ is the class of all tests of size } \alpha.$$

Definition: The short-coming of any size α test ϕ is defined as

$$S(\phi) = \sup_{\theta \in \bar{G}_{\mu} - w} (Y(\theta) - \beta(\phi, \theta)).$$

Definition: ϕ^* is said to be more stringent than ϕ if

$$S(\phi^*) \leq S(\phi).$$

The following theorems, (stated without proof for they are plainly obvious, outline the general methods of obtaining most stringent tests.

Theorem 2.11. If $Y(\theta)$ is constant over $\bar{G}_{\mu} - w$ and ϕ^* is such that it maximises $\inf_{\theta \in \bar{G}_{\mu} - w} \beta(\phi, \theta)$ among all tests ϕ of size α then ϕ

is the most stringent test of size α .

Theorem 2.12: Let $\gamma(\theta)$ be constant over $\bar{\Omega}_d$, for each d in D , where $\bigcap_{d \in \bar{D}} \bar{\Omega}_d = \bar{\Omega} - w$. Let ϕ_d^* be the test that maximises $\inf_{\theta \in \bar{\Omega}_d} \beta(\phi, \theta)$ among all tests of size α and let ϕ_d^* not depend on d . Then $\phi^* (= \phi_d^*)$ is the most stringent test of size α .

Theorem 2.13. Let ϕ^* be of size α and the most powerful for the problem $H: \theta \in w, A: \theta = \theta_1$ where θ_1 is some fixed point in $\bar{\Omega} - w$. Let $\beta(\phi^*, \theta) \geq \beta(\phi^*, \theta_1)$ for all $\theta \in \bar{\Omega} - w$. Then ϕ^* maximises $\inf_{\theta \in \bar{\Omega} - w} \beta(\phi, \theta)$ among all size α tests.

Let the group of transformations G be a compact topological group. Any finite group and the group of orthogonal matrices as examples of such groups. We have the following theorems.

Theorem 2.14: Let $\phi(x)$ be any test. There exists an invariant test $\psi(x)$ such that

$$\inf_{g \in G} \int \phi(gx) dP_{\theta}(x) \leq \int \psi(x) dP_{\theta}(x) \leq \sup_{g \in G} \int \phi(gx) dP_{\theta}(x) \quad \dots \quad (7)$$

Proof: Let μ be the left invariant Haar measure on the group G $\mu(B) = \mu(gB)$ for all sets B of the σ -field in G .

Let $\psi(x) = \int_G \phi(gx) d\mu(g) / \mu(G)$. $\psi(x)$ is invariant since,

$$\begin{aligned} \psi(g_1 x) &= \int_G \phi(g_1 g x) d\mu(g) / \mu(G) \\ &= \int_G \phi(g' x) d\mu(g') / \mu(G) \\ &= \psi(x) \end{aligned}$$

Now $\int_G \int \phi(gx) dP_\theta(x) d\mu(g) = \int dP_\theta(x) \int_G \phi(gx) d\mu(g) = \int \psi(x) dP_\theta(x)$. We find that $\int \psi(x) dP_\theta(x)$ is an average of the quantities $\int \phi(gx) dP_\theta(x)$. This proves (7).

Lemma 2.15. The power envelope $\gamma(\theta)$ is constant over any $\bar{G}_{\theta, \mu}$.

This lemma is plain.

Theorem 2.16. Given any non-invariant test ϕ there exists a more stringent invariant test

Proof:- Take any $\bar{G}_{\theta, \mu}$. $\gamma(\theta)$ is constant over $\bar{G}_{\theta, \mu}$. Let $\psi(x) = \int_G \phi(gx) d\mu(g) / \mu(G)$. Since $\psi(x)$ is invariant $\beta(\psi, \theta) = \int \psi(x) dP_\theta(x)$ is constant for θ in $\bar{G}_{\theta, \mu}$. We shall show that

$$\beta(\psi, \theta) \geq \inf_{\theta \in \bar{G}_{\theta, \mu}} \int \psi(x) dP_\theta(x) = \inf_{\bar{g} \in \bar{G}} \int \phi(x) dP_{\bar{g}\theta, \mu}(x) \dots (8)$$

This is an immediate consequence of theorem 2.15.

Thus (8) and theorems 2.12 and 2.13 show that ψ is more stringent than ϕ .

Theorem 2.14 and 2.16 can be shown to be true in the following more general situations also. X is the euclidean space of k dimensions and the group G is of the following type: $gx = Ax + B$, where A is a positive definite metric, etc., for other examples see Lehmann (1959) pp.335.

2.4. The theorem of confidence regions.

The space of actions, A , is the class of all subsets w of $(\bar{\cdot})$. The element w of A corresponds to the statement that θ is in w . The natural loss function that one shall assume in this case is

$$W(w, \theta) = \begin{cases} 0 & \text{if } \theta \in w \\ 1 & \text{if } \theta \notin w \end{cases}$$

For each x , the decision procedure, $S(x)$, is a subset of $\bar{\Omega}$ and is called a confidence region.

$$R(S(x), \theta) = P_{\theta}(S(X) \text{ does not cover } \theta).$$

If $R(S(x), \theta)$ is a constant ($=\alpha$) for all θ then $S(x)$ is called a $100(1 - \alpha)\%$ confidence region for θ .

The construction of a $100(1 - \alpha)\%$ confidence region is very simple if we note the very close relationship between critical region tests and confidence regions. Let W_{θ_0} be a critical region of size α for the testing problem $H : \theta = \theta_0, A : \theta \in \bar{\Omega} - \theta_0$. There is a one to one correspondence between a family of critical regions $\{W_{\theta}\}$, $\theta \in \bar{\Omega}$ and a $100(1 - \alpha)\%$ confidence region $S(x)$. Thus given $\{W_{\theta}\}$ define

$$T(x) = \{\theta : x \notin W_{\theta}\}.$$

It is easy to verify that $T(x)$ is a $100(1 - \alpha)\%$ confidence region.

Conversely if $S(x)$ is a 100% confidence region then define

$W_{\theta} = \{x : \theta \notin S(x)\}$. Then it is easy to verify that W_{θ_0} is a size α critical region for the testing problem, $H : \theta = \theta_0, A : \theta \in \bar{\Omega} - \theta_0$.

This one to one correspondence can be extended to randomised tests also. Given any randomised test $\phi(x)$, ^{it} can be viewed as a critical region test on the sample space $X \times R$ of the random variable (X, T) where X has the distribution P_{θ} and T has a rectangular distribution on $R = [0, 1]$ and X and T are independent. Let $W(x, r) = \{(x, r) : r \leq \phi(x)\}$. Then the test that rejects if $(x, r) \in W(x, r)$ is a

critical region test. Since $E(\phi(X)) = P(W(X,r))$ the critical region test $W(x,r)$ and the randomised test $\phi(x)$ are equivalent.

Optimal properties of a confidence region, $S(x)$ are defined in terms of the probabilities of $S(x)$ covering θ' under the distribution of P_{θ} .

Definition: A confidence region $S(x)$ is said to be shorter than another confidence region $S'(x)$ if

$$P_{\theta} \{S(X) \text{ covers } \theta'\} \leq P_{\theta} \{S'(X) \text{ covers } \theta'\} \text{ for all } \theta, \theta', \theta \neq \theta'.$$

Definition: $S(x)$ is the shortest $100(1 - \alpha)\%$ confidence if $S(x)$ is shorter than every other $100(1 - \alpha)\%$ confidence region.

Definition: A $100(1 - \alpha)\%$ confidence region $S(x)$ is said to be unbiased if

$$P_{\theta} \{S(X) \text{ covers } \theta'\} \leq P_{\theta} \{S(X) \text{ covers } \theta\} \text{ for all } \theta \neq \theta'.$$

It is easily proved that in the one to one correspondence between confidence regions and test functions that shortest unbiased confidence region correspond to most powerful unbiased test functions. This fact reduces the problem of constructing optimum confidence regions to a problem in tests of hypothesis.

Other optimality conditions on confidence are known (see for instance Kendall (1959) p. 72) but the construction of such optimal regions becomes difficult.

2.5. The theory of tolerance regions.

Let (\mathcal{X}, S) be the sample space and the associated σ -field of events of a random variable X with a distribution P_θ , $\theta \in \bar{\Omega}$. (\mathcal{X}^n, S^n) is the n -dimensional product space and the associated measure is $P_\theta \times P_\theta \times \dots \times P_\theta$. A tolerance region is a mapping $T(x_1, \dots, x_n)$ from \mathcal{X}^n to S . $P_\theta(T(x_1, \dots, x_n))$, the probability of the Borel set $T(x_1, \dots, x_n)$ is a real valued statistic on \mathcal{X}^n .

Definition: $T(x_1, \dots, x_n)$ is said to be a $\alpha - \beta$ tolerance region if

$$\text{Prob} \left\{ P_\theta(T(x_1, \dots, x_n)) \geq \beta \mid P_\theta \times \dots \times P_\theta \right\} = \alpha \quad \dots \quad (9)$$

Definition: $T(x_1, \dots, x_n)$ is said to be a distribution free $\alpha - \beta$ tolerance region if the distribution of $P_\theta(T(x_1, \dots, x_n))$ under $P_\theta \times \dots \times P_\theta$ is the same for all θ and (9) holds.

Definition: $T(x_1, \dots, x_n)$ is said to be a β -expectation tolerance region if $E_\theta(P_\theta(T(x_1, \dots, x_n))) = \beta$ for all θ .

One notes atonce that in practice we would like to choose α and β to be large. Again, even though $T(x_1, \dots, x_n) = \mathcal{X}$ is a 1-1 tolerance region it is hardly useful. A tolerance region should seek to limit a very large probability under P_θ to as small a set as possible. This condition of optimality can be formulated as follows. Let μ be a measure on (\mathcal{X}, S) .

Definition: A tolerance region T is better than T' if

$$E_{\theta} (\mu(T')) \leq E_{\theta} (\mu(T)).$$

The determination of a best $\alpha - \beta$ tolerance region is a difficult problem. In case μ is a probability measure, the determination of a best β - expectation tolerance region can be reduced to a problem in testing of hypothesis. Let the indicator function of $T(x_1, \dots, x_n)$ be $\phi(x; x_1, \dots, x_n)$. The condition that T is a β -tolerance region is

$$\int \dots \int \phi(x; x_1, \dots, x_n) dP_{\theta}(x) dP_{\theta}(x_1) \dots dP_{\theta}(x_n) = \beta \text{ for all } \theta \dots (10)$$

The condition of optimality of T is that

$$\int \dots \int \phi(x; x_1, \dots, x_n) d\mu(x) dP_{\theta}(x_1) \dots dP_{\theta}(x_n) \text{ a minimum } \dots (11)$$

The problem of choosing ϕ to satisfy (10) with the condition (11) is equivalent to finding a most powerful similar region test for the problem

$H: (X, X_1, \dots, X_n)$ distributed independently and identically as $P_{\theta}, \theta \in \Omega$

$A: (X_1, \dots, X_n)$ distributed independently and identically as $P_{\theta}, \theta \in \Omega$ and X independent of (X_1, \dots, X_n) and distributed as μ .

For further material on these lines see Fraser and Guttman (1956).

CHAPTER III

APPLICATIONS TO NONPARAMETRIC METHODS

3.1. The estimation of real parameters.

Let (\mathcal{X}, S) be the real line with its usual Borel σ -field of subsets. The admissible class of distributions for the random variable X is $\{P_\theta\}$, $\theta \in \Omega$, the class of all distributions on (\mathcal{X}, S) . X_1, \dots, X_n are independently and identically distributed as P_θ . This distribution of (X_1, \dots, X_n) on (\mathcal{X}^n, S^n) will be denoted by P_θ^n . $g(\theta)$ is a real valued functional of θ defined possibly only on a subset w of Ω . One of the problems in nonparametric estimation is to estimate $g(\theta)$ on the basis of (X_1, \dots, X_n) .

Definition. A real valued functional $f(x_1, \dots, x_n)$ on \mathcal{X}^n is said to be an unbiased estimator of $g(\theta)$ if

$$E_\theta(f(X_1, \dots, X_n)) = \int f(x_1, \dots, x_n) dP_\theta(x_1) \dots dP_\theta(x_n) = g(\theta)$$

for all $\theta \in w$.

Definition. $g(\theta)$ is said to be unbiasedly estimable if there exists an integer n and a function $f(x_1, \dots, x_n)$ such that $f(x_1, \dots, x_n)$ is unbiased for $g(\theta)$. The smallest such integer n is called the degree of $g(\theta)$.

As an example if $g(\theta)$ is the mean (whenever it exists) of the distribution P_θ , then $g(\theta)$ is unbiasedly estimable by x_1 and so is of degree 1. Similarly the k^{th} raw moment of P_θ is unbiasedly estimable

and of degree 1. We can also show that the k^{th} central moment of P_{θ} is unbiasedly estimable and of degree k .

If $g_1(\theta)$ and $g_2(\theta)$ are unbiasedly estimable and of degree k_1 and k_2 respectively, then we can easily see that $g_1(\theta) + g_2(\theta)$ is unbiasedly estimable and of degree less than or equal to $\min(k_1, k_2)$ and that $g_1(\theta)g_2(\theta)$ is unbiasedly estimable and of degree less than or equal to k_1k_2 .

Let

$$h(x_1, \dots, x_n) = \frac{1}{n!} \sum f(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \quad \dots \quad (1)$$

where the summation is over all permutations of the n integers (i_1, i_2, \dots, i_n) of $(1, 2, \dots, n)$. $h(x_1, \dots, x_n)$ is now symmetric in x_1, \dots, x_n . $f(x_1, \dots, x_n)$ and $h(x_1, \dots, x_n)$ are unbiased for the same parametric function, $g(\theta)$ say. Thus we can always find a symmetric and unbiased estimate for any unbiasedly estimable parametric function.

Consider the function $t(x_1, \dots, x_n) = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ where $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ is a permutation of (x_1, x_2, \dots, x_n) such that $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. This function is called the order statistic. The set of all (x_1, \dots, x_n) with $t(x_1, \dots, x_n) \in B$ ($B \in \mathcal{S}^n$) is a set, symmetric in the x 's. The induced distribution of $t(x_1, \dots, x_n)$ under P_{θ}^n is defined over the σ -field of such symmetric sets and is denoted by $P_{\theta, t}^n$.

Theorem 3.1. $t(x_1, \dots, x_n)$ is a sufficient statistic for the class of measures P_θ^n , $\theta \in \bar{\Omega}$.

Proof. Let $A \in \mathcal{S}^n$, B be a set of the σ -field of symmetric sets $\phi(A; x_1, \dots, x_n)$ denotes the indicator function of A , that is

$$\begin{aligned} \phi(A; x_1, \dots, x_n) &= 1 && \text{if } (x_1, \dots, x_n) \in A \\ &= 0 && \text{if } (x_1, \dots, x_n) \notin A \end{aligned}$$

$$\begin{aligned} P_\theta^n(A \cap B) &= \int_B \phi(A; x_1, \dots, x_n) dP_\theta(x_1) \dots dP_\theta(x_n) \\ &= \int_B \phi(A; x_{i_1}, \dots, x_{i_n}) dP_\theta(x_1) \dots dP_\theta(x_n) \end{aligned}$$

(because B is symmetric)

where (i_1, \dots, i_n) is a permutation of the integers $(1, \dots, n)$

$$\begin{aligned} \text{Thus } P_\theta^n(A \cap B) &= \int_B \frac{i(A; x_1, \dots, x_n)}{n!} dP_\theta(x_1) \dots dP_\theta(x_n) \\ &= \int_B \frac{i(A; x_1, \dots, x_n)}{n!} dP_{\theta, t}^n, \end{aligned}$$

where $i(A; x_1, \dots, x_n)$ is the number of permutations of (x_1, \dots, x_n) occurring in A . Thus, by definition, the conditional probability $P_\theta^n(A \mid t = (x_1, \dots, x_n)) = i(A; x_1, \dots, x_n)/n!$. This being independent of θ , t is sufficient.

Lemma 3.2. Let $Q(p_1, \dots, p_n)$ be a homogeneous polynomial of degree greater than 0. Let $Q(p_1, \dots, p_n) = 0$ whenever $0 \leq p_i \leq 1$, $i = 1, \dots, n$ and $p_1 + \dots + p_n = 1$. Then $Q(p_1, \dots, p_n) = 0$ for all p_1, \dots, p_n .

Proof: The lemma is trivial when $n = 1$. Now let the lemma be true for $n = m - 1$. We shall prove that the lemma is true for $n = m$.

We note that $Q(cp_1, \dots, cp_n) = c^k Q(p_1, \dots, p_n)$ when Q is of degree $k (\neq 0)$. Then $Q = 0$ for $0 \leq p_i \leq 1, \sum p_i = 1$ implies $Q = 0$ for all p_i with $0 \leq p_i (i = 1, \dots, n)$.

Writing out $Q(p_1, \dots, p_n)$ as a polynomial in p_n we note that for a fixed p_1, \dots, p_{n-1} , Q is zero for all $p_n \geq 0$. Thus each coefficient of p_n^r is zero. But since these coefficients are homogeneous polynomials of degree less than $m-1$, the lemma is true for $n = m$.

The lemma now follows from induction.

Lemma 3.3. Any function of $t(x_1, \dots, x_n)$ is symmetric in x_1, \dots, x_n and conversely.

Proof: This lemma is obvious.

Definition A distribution that has a constant density function on disjoint intervals I_1, \dots, I_k and a zero density elsewhere is called a distribution uniform over intervals.

Theorem 3.4. Let $\bar{\Omega}_1$ be the class of all θ such that the distributions P_θ is uniform over a collection of n disjoint intervals. Then $t(x_1, \dots, x_n)$ is complete for the distributions $P_\theta^n, \theta \in \bar{\Omega}_1$.

Proof: Let $h(x_1, \dots, x_n)$ be symmetric in x_1, \dots, x_n . It is a function of $t(x_1, \dots, x_n)$. Taking the expectation under a distribution uniform over I_1, \dots, I_n we have

$$0 = \sum_{i_1=1}^n \dots \sum_{i_n=1}^n p_{i_1} \dots p_{i_n} J(i_1, \dots, i_n) \dots \quad (2)$$

where $J(i_1, \dots, i_n) = \frac{1}{\ell(i_1) \dots \ell(i_n)} \int_{I_{i_1}} \dots \int_{I_{i_n}} h(x_1, \dots, x_n) dx_1 \dots dx_n$

and $\ell(1), \dots, \ell(n)$ are the length of the intervals I_1, \dots, I_n respectively. Therefore (2) can be written as

$$0 = \sum p_1^{a_1} \dots p_n^{a_n} c(a_1, \dots, a_n) \dots \quad (3)$$

where the summation is taken over all integers a_1, \dots, a_n such that $\sum a_i = n$ and where $c(a_1, \dots, a_n)$ is an integral multiple of $J(i_1, \dots, i_n)$ having a_1 of the i'_α 's equal to 1, a_2 of the i'_α 's equal to 2, and so on.

The expression on the right hand side of (3) satisfies the conditions of lemma 3.2 and so $c(a_1, \dots, a_n) = 0$ and hence $J(i_1, \dots, i_n) = 0$. It follows that

$$\int_{I_{i_1}} \dots \int_{I_{i_n}} h(x_1, \dots, x_n) dx_1, \dots, dx_n = 0 \text{ for all}$$

$i_1, \dots, i_n = 1, 2, \text{ or } n$ and all disjoint intervals I_1, \dots, I_n

By Lebesgue's theorem $h(x_1, \dots, x_n) = 0$ a.e. This being true for every symmetric function h , t is complete.

Theorem 3.5. The statistic t is complete for the class of all distributions P_θ^n with $\theta \in \bar{I}_2$, where $\theta \in \bar{I}_2$ whenever P_θ is a discrete distribution, with masses at n points on the real line.

Proof: The proof of this theorem is on the same lines as the previous theorem.

Definition: Let $\bar{\Omega}_3$ be the set of all θ such that P_θ has a continuous distribution function. Let $\bar{\Omega}_4$ be the set of all θ such that P_θ is an absolutely continuous distribution (w.r.t. Lebesgue measure). Let $\bar{\Omega}_5$ be the set of all θ for which P_θ is a discrete distribution. We now have the following obvious lemma.

Lemma 3.6. $t(x)$ is complete for the class of all distributions P_θ^n with $\theta \in \bar{\Omega}_3$, or $\bar{\Omega}_4$, or $\bar{\Omega}_5$.

Theorem 3.7. Let $g(\theta)$ be a parametric function defined on w , a subset of $\bar{\Omega}$. Let $f(x_1, \dots, x_n)$ be unbiased for $g(\theta)$. We have (i) $h(x_1, \dots, x_n)$ defined by (1) is unbiased for $g(\theta)$, (ii) $R(h, \theta) \leq R(f, \theta)$ whenever the loss function is convex and (iii) if w contains $\bar{\Omega}_1$ or $\bar{\Omega}_2$ then h is the unique unbiased minimum risk estimator.

Proof: These conclusions follow immediately from theorems 2.4, 3.1, 3.4 and 3.5.

Theorem 3.7 is the fundamental theorem in the theory of nonparametric estimation. It states that we should always restrict our estimates to symmetric functions and in many situations this is the best solution.

As examples we note that the raw moments and the k -statistics are unique unbiased minimum variance estimates of their expectations.

3.2. The theory of testing hypothesis applied to nonparametric problems.

We start the discussion with a very interesting non-parametric problem admitting of a uniformly most powerful test. Let (X^n, S^n) be the sample space and P_θ^n with $\theta \in \bar{\Omega}_4$ be the class of admissible distributions. The density function of P_θ is denoted by $f_\theta(x)$. The p^{th} quantile of the distribution P_θ is denoted by $\xi_{\theta,p}$. The hypothesis H is the set of all θ with $\xi_{\theta,p} = 0$, the alternative A is the set of all θ with $\xi_{\theta,p} > 0$.

The density function $f_\theta(x)$ of a distribution P_θ can be described by three quantities p^* , $f_+(x)$ and $f_-(x)$, where $0 \leq p^* \leq 1$, $f_+(x)$ is a density function on $(-\infty, 0)$ and $f_-(x)$ is a density function on $(0, +\infty)$, as follows. Given any such triplet, $p^*f_-(x) + (1-p^*)f_+(x) = f(x)$ is a density function of some P_θ with θ in $\bar{\Omega}_4$. Conversely let $p^* = P_\theta \{ X < 0 \}$,

$$f_-(x) = \frac{f_\theta(x)}{p^*} \text{ if } x < 0 \quad \text{and } = 0 \text{ otherwise,}$$

$$f_+(x) = \frac{f_\theta(x)}{1-p^*} \text{ if } x > 0 \quad \text{and } = 0 \text{ otherwise.}$$

Then $f_\theta(x) = p^*f_-(x) + (1 - p^*)f_+(x) \quad \dots \quad (4)$

Describing⁺ distributions in $\bar{\Omega}_4$ by p^* , $f_-(x)$ and $f_+(x)$ we note that H and A reduce to

$$H : p^* = p, \quad A : p^* < p.$$

+ In the problem of testing for the probabilities in k cells, a general distribution can be described in terms of the cell probabilities p_1, \dots, p_k and density functions $f_1(x), \dots, f_k(x)$ with $\sum p_i = 1$

The Karl Pearson's χ^2 test can be then obtained as a good large sample test.

Choose and fix a distribution $(p, f_-(x), f_+(x))$ in the hypothesis. Let us take the distribution $(p^*, f_-(x), f_+(x))$ with a $p^* < p$ as our alternative. Let us determine the most powerful size α test for this problem. Using the Neyman-Pearson lemma 2.7. we find the test to be

$$\begin{aligned} \phi(x_1, \dots, x_n) &= 1 && \text{if } k(x_1, \dots, x_n) > c \\ &= a && \text{if } k(x_1, \dots, x_n) = c \\ &= 0 && \text{if } k(x_1, \dots, x_n) < c \end{aligned}$$

$$\begin{aligned} \text{where } k(x_1, \dots, x_n) &= \frac{\prod_{i=1}^n [p^* f_-(x_i) + (1 - p^*) f_+(x_i)]}{\prod_{i=1}^n [p f_-(x_i) + (1 - p) f_+(x_i)]} \\ &= \left(\frac{p^*}{p}\right)^{n-s(x_1, \dots, x_n)} \left(\frac{1-p^*}{1-p}\right)^{s(x_1, \dots, x_n)} \\ &= \left(\frac{p(1-p^*)}{p^*(1-p)}\right)^{s(x_1, \dots, x_n)} \left(\frac{p^*}{p}\right)^n \end{aligned}$$

where $s(x_1, \dots, x_n)$ is the number of the x_i 's greater than 0. Thus the most powerful test can be written as

$$\begin{aligned} \phi(x_1, \dots, x_n) &= 1 && \text{if } s(x_1, \dots, x_n) > c' \\ &= a && \text{if } s(x_1, \dots, x_n) = c' \\ &= 0 && \text{if } s(x_1, \dots, x_n) < c' \end{aligned}$$

where a and c' have to be chosen so that the test is of size α . We note the test is similar with regard to the original hypothesis. Thus it is most powerful similar against the alternative $(p^*, f_-(x), f_+(x))$. Since the test does not depend on the alternative it is uniformly most

powerful against the original set of alternatives.

The above test is called the sign test. There are a few other problems where most powerful size α tests exist and they are just minor variations of the above problem. Thus to test $\xi_{\theta,p} = \xi_0$ against $\xi_{\theta,p} > \xi_0$ the sign test based on $x_1 - \xi_0, \dots, x_n - \xi_0$ is most powerful similar. The matched pair sign test which tests that the median of the difference of two random variables is zero is uniformly most powerful. The hypothesis can be extended in all these cases to say, $\xi_{\theta,p} \leq 0$ instead of $\xi_{\theta,p} = 0$, and the sign test would remain most powerful and of size α .

For the problem of two sided alternatives, e.g. $H : \xi_{\theta,p} = 0$, $A : \xi_{\theta,p} \neq 0$, we can show the existence of uniformly most powerful unbiased similar size α tests based on $s(x_1, \dots, x_n)$. This test will correspond to the test in the Binomial distribution for $p^* = p$ against $p^* \neq p$. For detailed proofs refer to Fraser (1953).

We now pass on to certain problems where we can construct most powerful similar tests against some specific alternatives. As an example let Ω_6 be the set of all θ such that the distribution P_θ is absolutely continuous, symmetric and has median equal to 0. Let Ω_7 be the set of θ such that the distribution P_θ has the density function $\frac{1}{\sqrt{2\pi}\sigma} \exp(-(x-\mu)^2/2\sigma^2)$; $\mu > 0$; $\sigma > 0$. Set the problem to be $H : \theta \in \Omega_6$, $A : \theta \in \Omega_7$. We will now determine the most powerful similar size α test. Consider the statistic $u(x_1, \dots, x_n) = (|x_{(1)}|, |x_{(2)}|, \dots, |x_{(n)}|)$ which is the unsigned order statistic.

We first note that $u(x_1, \dots, x_n)$ as a sufficient and complete statistic for the distributions in $(\bigcap)_6$. This follows from simple extensions in Theorems 3.1 and 3.4. When x_1, \dots, x_n are distinct (this happens with probability one) the conditional distribution of X_1, \dots, X_n given $u(X_1, \dots, X_n) = (d_1, \dots, d_n)$ is the probability distribution that assigns the mass $(\frac{1}{2})^n$ to each of the possibilities $(\pm d_1, \pm d_2, \dots, \pm d_n)$. This distribution will be called the randomization distribution.

Applying theorem 2.8 we note that it is enough to construct the most powerful test on this conditional sample space for each u . The resulting test will be most powerful similar size α test. Let the alternative be restricted to a single distribution with parameters μ and σ . Using the Neyman-Pearson lemma, we find that the most powerful size α test on the conditional space is given by

$$\begin{aligned} \phi(x_1, \dots, x_n) &= 1 && \text{if } h(x_1, \dots, x_n) > c \\ &= a && \text{if } h(x_1, \dots, x_n) = c \\ &= 0 && \text{if } h(x_1, \dots, x_n) < c \end{aligned}$$

where $h(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left(-\frac{1}{2} \sum (x_i - \mu)^2 \right)$.

Now with $\mu > 0, \sigma > 0$, $h(x_1, \dots, x_n)$ is a monotonic function of

$\sum x_i / \sqrt{\sum x_i^2} = v(x_1, \dots, x_n)$. Thus the most powerful similar size α test against $(\bigcap)_7$ is given by

$$\begin{aligned} \phi(x_1, \dots, x_n) &= 1 && \text{if } v(x_1, \dots, x_n) > c(u) \\ &= a(u) && \text{if } v(x_1, \dots, x_n) = c(u) \\ &= 0 && \text{if } v(x_1, \dots, x_n) < c(u) \end{aligned}$$

where a and c are so determined that the size is α under randomisation on each of the conditional sample spaces.

For testing $H : \theta \in \bar{\Omega}_6$ against $A : \theta$ such that P_θ has density function $(2\pi\sigma)^{-\frac{1}{2}} \exp(-(x-\mu)^2/2)$ $\mu \neq 0, \sigma > 0$, we can show that the two sided equal tail area test based on the ^{distribution} randomisation \int of v is the most stringent similar. To prove this we first note that the power envelope is constant at the two distributions with density functions $(2\pi\sigma)^{-1} \exp(-(x-\mu)^2/2\sigma^2)$ and $(2\pi\sigma)^{-1} \exp(-(x+\mu)^2/2\sigma^2)$. Next choosing an apriori distribution with masses $\frac{1}{2}$ and $\frac{1}{2}$ at these two distributions we find the most powerful size α test against this mixture of distributions of the alternative. This test reduces to the two sided equal tail area test mentioned earlier and is most stringent similar in virtue of theorem 2.13. The above two tests are known as Fisher's randomisation tests.

The above technique of constructing tests with such optimal properties can be carried out whenever complete and sufficient statistics are available. As a further illustration we present the following example without all the details.

The sample space is $(\mathcal{X}^{n_1+n_2}, S^{n_1+n_2})$. The null hypothesis consists of all distributions $P_\theta^{n_1+n_2}, \theta \in \bar{\Omega}_4$. Under the alternative

X_1, \dots, X_{n_1} are independent and identical normal variables with mean μ_1 and variance σ^2 and $X_{n_1+1}, \dots, X_{n_1+n_2}$ are independent and identical normal variables with mean μ_2 and variance σ^2 , with $\mu_1 > \mu_2$. Consider the test

$$\begin{aligned} \phi(x_1, \dots, x_n) &= 1 && \text{if } \sum_{i=1}^{n_1} x_i > c(t) \\ &= a(t) && \text{if } \sum_{i=1}^{n_1} x_i = c(t) \\ &= 0 && \text{if } \sum_{i=1}^{n_1} x_i < c(t) \end{aligned}$$

where a and c are determined so that the size of the test on each conditional space, with fixed $t(x_1, \dots, x_{n_1+n_2})$, is α . This test is most powerful similar. Now let the alternative be extended to include the class of all normal distributions with $\mu_1 \neq \mu_2$. Consider the test

$$\begin{aligned} \phi(x_1, \dots, x_n) &= 1 && \text{if } |\bar{x}_1 - \bar{x}_2| > c(t) \\ &= a(t) && \text{if } |\bar{x}_1 - \bar{x}_2| = c(t) \\ &= 0 && \text{if } |\bar{x}_1 - \bar{x}_2| < c(t) \end{aligned}$$

$$\bar{x}_1 = \sum_{i=1}^{n_1} x_i / n_1, \quad \bar{x}_2 = \sum_{i=n_1+1}^{n_1+n_2} x_i / n_2$$

where a and c are chosen to make the test of size α on each conditional sample space with fixed $t(x_1, \dots, x_{n_1+n_2})$. This test is most stringent similar. These tests are called the Pitman randomisation tests.

Other applications of these techniques can be found in Lehmann and Stein (1949).

Another large class of tests can be generated by the use of the principle of invariance. Let (X^n, S^n) be the sample space. Under the hypothesis (X_1, \dots, X_n) has the distribution P_θ^n with $\theta \in \bar{\Omega}_3$. Under the alternative X_1, \dots, X_n are independent and have distributions $P_{\theta_1}, \dots, P_{\theta_n}$ respectively $\theta_1, \dots, \theta_n \in \bar{\Omega}_3$ and for some $i, j, \theta_i \neq \theta_j$. We thus parametrise the admissible distributions by $\underline{\theta} = (\theta_1, \dots, \theta_n)$ (corresponding to the distribution $P_{\theta_1} \times \dots \times P_{\theta_n}$) and $\underline{\theta}$ varies in $\bar{\Omega}_3^n$. The Hypothesis corresponds to $\underline{\theta}$ such that $\theta_1 = \dots = \theta_n$ and the alternative corresponds to $\underline{\theta}$ such that not all θ_i 's are equal.

Consider the class G of transformations on the real line that are strictly increasing and continuous. This class of transformations is a group. Let X be a random variable with distribution P_θ with $\theta \in \bar{\Omega}$. The random variable gX has a distribution which corresponds to some θ in $\bar{\Omega}$. Let this parameter be denoted by $\bar{g}\theta$. Thus corresponding to each transformation g there is a transformation \bar{g} on $\bar{\Omega}_3$. The class of \bar{g} such transformations \bar{g} forms a group of transformations on $\bar{\Omega}_3$. The transformation g can be visualised as a transformation on X^n as follows, $g(x_1, \dots, x_n) = (gx_1, \dots, gx_n)$. Correspondingly if (X_1, \dots, X_n) has a distribution specified by $(\theta_1, \dots, \theta_n)$, $g(X_1, \dots, X_n)$ will have a distribution specified by $\bar{g}(\theta_1, \dots, \theta_n) = (\bar{g}\theta_1, \dots, \bar{g}\theta_n)$. Further the null hypothesis and alternative are left invariant by the

transformations in \bar{G} . The principle of invariance demands that we use only those tests that are invariant under G .

Let T be set of points (x_1, \dots, x_n) in \mathcal{X}^n with $x_i = x_j$ for some i and j . The probability of T is zero for all the admissible distributions. Consider the function $r(x_1, \dots, x_n) = (r_1, \dots, r_n)$ on $\mathcal{X}^n - T$ that gives the ranks (in the increasing order of magnitude) of x_1, \dots, x_n . It is easy to see that $r(x_1, \dots, x_n)$ is the maximal invariant function with respect to the group of transformations G . Let $h_1(u), \dots, h_n(u)$ be n strictly increasing continuous functions from $[0, 1]$ onto $[0, 1]$. For each such (h_1, \dots, h_n) consider the collection of distribution functions $(h_1(F_\theta(x)), \dots, h_n(F_\theta(x)))$ with $\theta \in \bar{\Omega}_3$.

Lemma 3.8. The collection of such distributions is invariant under \bar{G} . The distribution of any invariant statistic is the same under the distributions of each such collection.

Proof: The first part of the lemma is obvious. To prove the second part we show that it is true for the function $r(x_1, \dots, x_n)$.

$$\begin{aligned} \text{Let } \theta_2 = \bar{g}\theta_1. \text{ Then Prob. } & \left\{ r(X_1, \dots, X_n) = r \mid h_1^{F_{\theta_1}}, \dots, h_n^{F_{\theta_1}} \right\} \\ &= \text{Prob. } \left\{ r(gX_1, \dots, gX_n) = r \mid h_1^{F_{\bar{g}\theta_1}}, \dots, h_n^{F_{\bar{g}\theta_1}} \right\} \\ &= \text{Prob. } \left\{ r(X_1, \dots, X_n) = r \mid h_1^{F_{\theta_2}}, \dots, h_n^{F_{\theta_2}} \right\} \\ &= \text{Prob. } \left\{ r(X_1, \dots, X_n) = r \mid h_1^{F_{\theta_2}}, \dots, h_n^{F_{\theta_2}} \right\}. \end{aligned}$$

This proves the lemma.

We note that though the collection of distributions defined through (h_1, \dots, h_n) does not yield a maximal invariant set, the distribution of every invariant function is constant over such a set. We also note that the distributions of the hypothesis correspond to the functions (e, \dots, e) where $e(u) = u$. Now let the alternatives be restricted to the set of distributions corresponding to (h_1, \dots, h_n) . Restricting ourselves to invariant tests we note that the hypothesis and alternative become simple. If $\text{Prob.} \{r(X_1, \dots, X_n) = \underline{r} \mid h_1, \dots, h_n\}$ were known then an application of Neyman - Pearson lemma will yield the most powerful invariant test.

The following theorem due to Hoeffding (1951a) enables us to calculate $\text{Prob.} \{r(X_1, \dots, X_n) = \underline{r}\}$ under any distribution. Let P and P^* be two distributions on \mathcal{X}^n . Their densities are $f(\underline{x})$ and $f^*(\underline{x})$ respectively. Let us assume that $f(\underline{x})$ is symmetric in x_1, \dots, x_n . Given any permutation $\underline{r} = (r_1, \dots, r_n)$ of the integers $(1, \dots, n)$ let $S_{\underline{r}}$ denote the set of all (x_1, \dots, x_n) with $r(x_1, \dots, x_n) = \underline{r}$. Let \underline{r}' denote the permutation which when applied to \underline{r} takes it to $\underline{1} = (1, \dots, n)$. We also note that the density function of the order statistics $X_{(1)}, \dots, X_{(n)}$ under P is $n! f(x_1, \dots, x_n)$ on $S_{\underline{1}}$ and ^{is} zero elsewhere.

Theorem 3.9.

$$P \left\{ r(X_1, \dots, X_n) = \underline{r} \mid P^* \right\} = \frac{1}{n!} E_P \left\{ \frac{f^*(X_{(r'_1)}, \dots, X_{(r'_n)})}{f(X_{(r'_1)}, \dots, X_{(r'_n)})} \right\}$$

Proof:

$$\begin{aligned}
 P \{ r(X_1, \dots, X_n) = \underline{r} \mid P^* \} &= \int_{S_{\underline{r}}} f^*(x_1, \dots, x_n) dx_1 \dots dx_n \\
 &= \int_{S_I} f^*(x_{r'_1}, \dots, x_{r'_n}) dx_1 \dots dx_n \\
 &= \frac{1}{n!} \int_{S_I} \frac{f^*(x_{r'_1}, \dots, x_{r'_n})}{f(x_{r'_1}, \dots, x_{r'_n})} \cdot n! f(x_1, \dots, x_n) dx_1, \dots, dx_n \\
 &= \frac{1}{n!} E_P \left\{ \frac{f^*(X_{(r'_1)}, \dots, X_{(r'_n)})}{f(X_{(r'_1)}, \dots, X_{(r'_n)})} \right\} .
 \end{aligned}$$

As an example let us take $n = n_1 + n_2$, $(h_1, \dots, h_{n_1+n_2}) = (e, \dots, e, h, \dots, h)$ where $h(u)$ is differentiable. $(h_1, \dots, h_{n_1+n_2}) \dots n_1 \dots n_2 \dots$ defines a class of alternatives in the two sample problem.

$$\begin{aligned}
 \frac{f^*(x_1, \dots, x_n)}{f(x_1, \dots, x_n)} &= \frac{\frac{d}{dx_1} F_{\theta}(x) \dots \frac{d}{dx_{n_1}} F_{\theta}(x) \frac{d}{dx_{n_1+1}} h(F_{\theta}(x)) \dots \frac{d}{dx_n} h(F_{\theta}(x))}{\frac{d}{dx_1} F_{\theta}(x) \dots \frac{d}{dx_{n_1}} F_{\theta}(x) \frac{d}{dx_{n_1+1}} F_{\theta}(x) \dots \frac{d}{dx_n} F_{\theta}(x)} \\
 &= h'(u_{n_1+1}) \dots h'(u_n)
 \end{aligned}$$

where $u = F_{\theta}(x)$.

Thus

$$\begin{aligned}
 \text{Prob} \{ r(X_1, \dots, X_n) = \underline{r} \mid (e, \dots, e, h, \dots, h) \} \\
 &= \frac{1}{n!} E(e, \dots, e) \left\{ \frac{f^*(X_{(r'_1)}, \dots, X_{(r'_n)})}{f(X_{(r'_1)}, \dots, X_{(r'_n)})} \right\} \\
 &= \frac{1}{n!} E \left\{ h'(U_{r'_{n_1+1}}) \dots h'(U_{r'_n}) \right\}
 \end{aligned}$$

where U_1, \dots, U_n are the order statistics in a sample of n independent observations on the rectangular distribution on $(0, 1)$.

Specialising further let us take $h(u) = u^k, k > 0$. After some tedious algebra (for instance see Lehmann(1953)) we can obtain

$$P\{r(X_1, \dots, X_n) = r \mid (e, \dots, e, u^k, \dots, u^k)\}$$

$$= \text{const.} \times \prod_{j=1}^{n_2} \frac{\Gamma(s_j + jk - j) \Gamma(s_{j+1})}{\Gamma(s_j) \Gamma(s_{j+1} + jk - j)}$$

where s_1, \dots, s_{n_2} are the ranks $r'_{n_1+1}, \dots, r'_{n_1+n_2}$ arranged in increasing order of magnitude and $s_{n_2+1} = n_1+n_2+1$.

Thus if $k = 2$, the most powerful rank invariant test of size α would be

$$\phi(s_1, \dots, s_{n_2}) = 1 \text{ if } s_1(s_2+1) \dots (s_{n_2}+n_2-1) > c$$

$$= a \text{ if } s_1(s_2+1) \dots (s_{n_2}+n_2-1) = c$$

$$= 0 \text{ if } s_1(s_2+1) \dots (s_{n_2}+n_2-1) < c$$

when a and c are chosen to make test of size α .

As another example let $h(u) = h_p(u) = qu + pu^2, q = 1-p, 0 < p \leq 1$. We will find out the locally most powerful invariant test against the alternatives $0 < p \leq 1$.

$$\begin{aligned}
 P \{ r(X_1, \dots, X_n) = r | (e, \dots, e, h_p, \dots, h_p) \} \\
 = \frac{1}{n!} E \{ (q + 2pU_{s_1}) \dots (q + 2pU_{s_2}) \} \\
 \left[\frac{d}{dp} P(r(X_1, \dots, X_n) = r | (e, \dots, e, h_p, \dots, h_p)) \right]_{p=0} \\
 = \frac{1}{n!} E \left\{ \sum_{i=1}^{n_2} (2U_{s_i} - 1) \right\} = k \sum_{i=1}^{n_2} s_i + \ell
 \end{aligned}$$

where $k > 0$ and ℓ are two constants. From this it follows that the locally most powerful rank test of size α is

$$\begin{aligned}
 \phi(s_1, \dots, s_2) &= 1 \text{ if } \sum_{i=1}^{n_2} s_i > c \\
 &= a \text{ if } \sum_{i=1}^{n_2} s_i = c \\
 &= 0 \text{ if } \sum_{i=1}^{n_2} s_i < c
 \end{aligned}$$

where a and c are chosen to give the test size α . This is the one sided Mann-Whitney test.

For other interesting examples of rank tests with optimum properties we can refer to Lehman (1953) ^{and} Terry (1952).

3.3. Confidence regions in non-parametric problems.

Confidence regions in non-parametric problems are arrived at in the same way as in parametric problems.

In this section we give just one example, namely the confidence interval for the population p th quantile. If X_1, \dots, X_n are n

independent and identical random variables then the interval $(X_{(k)}, X_{(l)})$ is a shortest unbiased $100(1 - \alpha)\%$ confidence region

if
$$\sum_{r=k}^{n-l} \binom{n}{r} p^r (1-p)^{n-r} = \alpha \quad \dots \quad (5)$$

and
$$\sum_{r=k}^{n-l} r \binom{n}{r} p^r (1-p)^{n-r} = np\alpha \quad \dots \quad (6)$$

This follows easily by an inversion of the sign test for the p th quantile. In many situations the test would be randomised and this should be reduced to a critical region test before the inversion is carried out. The details of such methods are explained in section 2.4.

3.4. Tolerance regions in non-parametric problems.

Distribution free tolerance regions are very easy to construct as will be obvious from theorem 3.10. There is an important outstanding problem of defining optimality criteria and determining such optimal tolerance regions. This was mentioned in section 2.5. Till this problem is solved, the material in this section will remain to be of purely academic interest.

Firstly we discuss some sampling properties of the order statistics of a rectangular population which play an important role in the construction of tolerance regions. Let Z_1, \dots, Z_n be n independent random variables with a uniform distribution on $(0, 1)$. The order statistics

are $Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}$. Let $C_1 = Z_{(1)}, C_2 = Z_{(2)} - Z_{(1)}, \dots$
 $C_n = Z_{(n)} - Z_{(n-1)}, C_{n+1} = 1 - Z_{(n)}$. Then C_1, C_2, \dots, C_{n+1} have a
 symmetric distribution with density function

$$n! \text{ if } 0 \leq c_i \leq 1, \sum_{i=1}^{n+1} c_i = 1$$

$$0 \text{ otherwise.}$$

If $S_k = C_{i_1} + \dots + C_{i_k}$ ($i_1 \neq i_2 \neq \dots \neq i_k$)

then

$$P\{S_k \geq \beta\} = \frac{1}{B(k, n-k+1)} \int_{\beta}^1 y^{k-1} (1-y)^{n-k} dy$$

$$= \alpha(n, k, \beta).$$

Thus S_k is a $\alpha(n, k, \beta) - \beta$ tolerance region for the rectangular
 population. Given n and β , we can choose k so that $\alpha(n, k, \beta) \geq \alpha$.

Let x_1, x_2, \dots, x_n be n independent observations on a random
 variable X with a probability distribution P . Let its distribution
 function $F(x)$ be continuous. Let $\phi_1(x), \dots, \phi_n(x)$ be n functions
 on \mathcal{X} with continuous induced distributions.

Let (p_1, \dots, p_n) be a permutation of the n
 integers $(1, \dots, n)$. Arrange the values $\phi_1(x_1), \dots, \phi_n(x_n)$ in their
 decreasing order of magnitude. Let p_1 th largest value, denoted by
 $\max_{i=1}^n (p_i) \phi_1(x_i)$, be r_1 . We then divide \mathcal{X} into two regions

$$S_{1, \dots, p_1} = \{x : \phi_1(x) > r_1\}$$

$$S_{p_1+1, \dots, n} = \{x : \phi_1(x) < r_1\}$$

by means of the cut $T_{p_1} = \{x : \phi_1(x) = r_1\}$. Since $\phi_1(x)$ has a continuous distribution function we can assume that there will be only one x_i with $\phi_1(x_i) = r_1$, that is, only one x_i in T_{p_1} . There are $(p_1 - 1)$ x_i 's in S_1, \dots, p_1 and $(n - p_1)$ x_i 's in S_{p_1+1}, \dots, n .

The second cut is made by using the function $\phi_2(x)$ and the integer p_2 . If p_2 is one of the integers $1, \dots, p_1 - 1$ then the region S_1, \dots, p_1 is cut into two regions

$$S_{1, \dots, p_2} = S_{1, \dots, p_1} \cap \{x : \phi_2(x) > r_2\}$$

$$S_{p_2+1, \dots, p_1} = S_{1, \dots, p_1} \cap \{x : \phi_2(x) < r_2\}$$

by means of the cut $T_{p_2} = S_{1, \dots, p_1} \cap \{x : \phi_2(x) = r_2\}$ where

$r_2 = \max_{i=1, \dots, p_1} \phi_2(x_i)$. If p_2 is one of the integers $p_1 + 1, \dots, n$

then the region S_{p_1+1, \dots, p_n} is cut into two regions

$$S_{p_1+1, \dots, p_2} = S_{p_1+1, \dots, n} \cap \{x : \phi_2(x) > r_2\}$$

$$S_{p_2+1, \dots, n} = S_{p_1+1, \dots, n} \cap \{x : \phi_2(x) < r_2\}$$

by means of the cut $T_{p_2} = S_{p_1+1, \dots, n} \cap \{x : \phi_2(x) = r_2\}$ where

$r_2 = \max_{i=p_1+1, \dots, n} \phi_2(x_i)$. We continue this construction to the next stage in a similar fashion. In the end we will have n cuts

T_1, \dots, T_n and $(n+1)$ regions S_1, S_2, \dots, S_{n+1} that will cover X .

Theorem 3.10. $P(S_1), P(S_2), \dots, P(S_{n+1})$ have a joint distribution which is the same as the joint distribution of C_1, C_2, \dots, C_{n+1} .

Proof: The proof is done stepwise, corresponding to the stages in the constructional procedure, and a correspondence is set up with the simple case of sampling from the uniform distribution.

Let $G(y)$ be the distribution function $\phi_1(X)$. Since it is assumed to be continuous $G(\phi_1(X))$ has a uniform distribution. Therefore the joint distribution $P(S_{1, \dots, p_1}, S_{p_1+1, \dots, n+1})$ is the same as the joint distribution of $1 - Z_{(n-p_1+1)}$ and $Z_{(n-p_1+1)}$, ^{is the same as} in other words, the joint distribution of $C_{n+1} + \dots + C_{n-p_2}$ and $C_1 + \dots + C_{n-p_1}$.

Given $\max_1^n (p_1) \phi_1(X_1) = r_1$ the conditional distribution of X_1, \dots, X_n is the distribution of $p_1 - 1$ independent observations on X restricted to S_{1, \dots, p_1} and $n - p_1$ independent observations on X restricted to $S_{p_1+1, \dots, n}$ even as the conditional distribution of Z_1, \dots, Z_n given $Z_{(n-p_1+1)} = G(r_1)$ is the distribution of $p_1 - 1$ independent uniform variables on $(Z_{(n-p_1+1)}, 1)$ and $n - p_1$ independent uniform variables on $(0, Z_{(n-p_1+1)})$. Thus the further division of \mathcal{X} into regions $S_{1, \dots, p_1}, S_{p_1+1, \dots, p_2}, S_{p_2+1, \dots, n+1}$ (in case $p_1 < p_2$) can be likened to corresponding division of $(0, 1)$ into the regions

$$(Z_{(n-p_1+1)}, 1), (Z_{(n-p_2+1)}, Z_{(n-p_1+1)}), (0, Z_{(n-p_2+1)}),$$

and the joint distributions of the probabilities of these regions under P and the uniform distribution, respectively are the same. Applying

we show that
this procedure repeatedly $\wedge P(S_1), \dots, P(S_{n+1})$ have the same joint distribution as C_1, \dots, C_{n+1} . Hence the theorem.

Thus to choose an $\alpha - \beta$ tolerance region we first find k such that $\alpha(n, k, \beta) = \alpha$. Then the union of any k regions from S_1, \dots, S_{n+1} forms a $\alpha - \beta$ distribution free tolerance region.

The functions $\phi_1(x), \phi_2(x), \dots, \phi_n(x)$ and the permutation (p_1, \dots, p_n) can be chosen in a more general fashion by allowing ϕ_i and p_i to depend on r_1, \dots, r_{i-1} , $i = 2, \dots, n$, without altering the construction of the tolerance region and the truth of Theorem 3.10. Theorem 3.10 can be generalised to include discontinuous distributions also. For these generalisations one can consult the series of papers by Sheffe and Tukey (1945), Tukey (1947), (1948), Fraser and Formleighton (1951), Fraser (1951), (1953a) etc.

3.5. Efficiency and consistency of estimates.

The performance of an unbiased estimate t of $g(\theta)$ is measured by its variance $E(t - g(\theta))^2$. Under certain assumptions in parametric analysis, the well known theorem of Cramer-Rao (See Cramer (1946) pp.480 or Rao (1952) pp. 131) gives a lower bound $b(\theta)$ to this variance. Thus we have the following definition of efficiency.

Definition : The efficiency $e(t)$ of an estimate t of $g(\theta)$ is defined by the relation

$$e(t) = \frac{E(t - g(\theta))^2}{b(\theta)} \quad \dots \quad \dots \quad (7)$$

When t_n is an estimate based on n independent observations on a variable X , the asymptotic efficiency of $\{t_n\}$ can be defined to be the limit of the $e(t_n)$.

The conditions under which the Cramer-Rao lower bound $b(\theta)$ holds do not obtain in the problems of non-parametric estimation. In fact even a formal evaluation of $b(\theta)$ is not possible. Thus $e(t)$ cannot serve as a measure of efficiency in non-parametric problems. Bahadur (1960) has adopted another definition of efficiency on the basis of the concentration of estimates. It can be shown that for a large class of estimates that

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\log P(|t_n - g(\theta)| > \epsilon)}{\epsilon^2} = \log \varrho(t) \quad \dots \quad (8)$$

with $0 < \varrho(t) < 1$. $\varrho(t)$ is a measure of the efficiency of t_n , and smaller values of $\varrho(t)$ indicate higher efficiency of $\varrho(t)$. The evaluation $\varrho(t)$, in case it exists, is difficult even in parametric problems which means that this method is now not available for non-parametric problems. Nevertheless $\varrho(t)$ based on considerations of concentration seems to be more satisfying than $e(t)$ based on variance.

The performance of a test is based on its power function. Let $\{\phi_n\}$, $\{\phi_n^*\}$ be two sequences of size α tests based on X^n for the

problem $H : \theta = \theta_0$, $A : \theta \in \bar{I} - \theta_0$. Let θ_i be a sequence of points in A tending to θ_0 in some topology in \bar{I} . Let $\{n_i\}$, $\{n_i^*\}$ be defined so that

$$\lim_{i \rightarrow \infty} \beta(\phi_{n_i}, \theta) = \lim_{i \rightarrow \infty} \beta(\phi_{n_i}^*, \theta) = \beta, \quad 0 < \beta < 1.$$

Then the efficiency $e(\phi, \phi^*)$ of $\{\phi_n\}$ w.r.t. $\{\phi_n^*\}$ is defined by

$$e(\phi, \phi^*) = \lim_{i \rightarrow \infty} \frac{n_i^*}{n_i} \quad \dots \quad (9)$$

whenever it exists. This efficiency, known as Pitman's efficiency, when it exists, depends on β and the sequence $\{\theta_i\}$. In certain parametric problems it becomes independent of β and the sequence $\{\theta_i\}$. For such theorems and other reductions one can consult Fraser (1957) chap. 7. An alternative definition of efficiency of a test due to Rao (1961) (analogous to Bahadur's asymptotic slope c.f. Bahadur (1960)) is the following. It can be shown that for a large class of tests $\{\phi_n\}$ (when \bar{I} is an interval on the real line) that

$$\lim_{\theta \rightarrow \theta_0} \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\log(1 - \beta(\phi_n, \theta))}{(\theta - \theta_0)^2} = \log q(\phi) \quad (10)$$

with $0 < q < 1$. As in the case of estimates the test with the smallest q is most efficient. Certain very interesting theorems concerning this new definition of efficiency can be found in Rao (1961) and Bahadur (1960),

unfortunately for parametric problems only.

The known results concerning the efficiency of a non-parametric test are those that compare a non-parametric test with a parametric test at parametric alternatives. For instance a general two sample non-parametric test is compared with the Fisher's t-test when the alternative is restricted to normal distributions with the same variance but different means. For such examples see Fraser (1957) Chap. 7. Some very interesting limit distributions have been developed by Hoeffding (1952) in this connection. The main criticism against this approach is that the non-parametric test and the parametric test have been developed for entirely different problems and a comparison between them could be misleading. If the efficiency of a non-parametric test is high it means that it is quite good at the parametric alternatives, it throws no light on the performance of the test at non-parametric alternatives. In fact it may happen that a test with low efficiency at parametric alternatives may be good at non-parametric alternatives. Thus a comparison between two non-parametric tests alone and at all alternatives will be useful. Notwithstanding the fact that this might be a difficult problem, we should remark that we rarely come across two non-parametric tests for the same problem and so such comparisons are rarely necessary. And, the previous sections of this chapter were for developing good tests for specific problems.

PART II

LIMIT DISTRIBUTIONS

Part II

Chapter IV

LIMITING DISTRIBUTIONS

1.1. General techniques of limiting distributions.

In the last chapter we introduced several tests all of which had to be adjusted to be of a previously chosen size. This means that the sampling distribution of these test functions would have to be determined. This problem is more complicated in the case of non-parametric tests than in the case of parametric tests. As an alternative, the limiting sampling distributions of these statistics as some constants (usually the sample size, n) in the distributions tends to infinity are determined and then used as approximations to the exact sampling distributions. This would enable us not only to fix the size approximately, but also to compute the power of the test, though only approximately. Thus the theory of limiting distributions is a necessary appendage to the theory of non-parametric inference. It is also interesting to note that the theory of limiting distributions has been enriched by certain special problems peculiar to non-parametric problems. A knowledge of the several techniques of limiting distributions would be very helpful in many situations.

To define convergence and limit distributions of random variables we should first examine how a distribution is defined. If there

are several ways of defining a distribution, a natural mode of convergence can be associated with each definition. The study of the interrelation of these different modes of convergence will be useful in deciding whether a given sequence of distributions converges in an accepted mode.

By a distribution we mean a probability measure λ defined on the measurable subsets of (X, S) . We assume that the space is topological and all continuous functions are measurable. A sequence of distributions λ_n is said to converge to a distribution λ if

$$\int g d\lambda_n \rightarrow \int g d\lambda$$

as $n \rightarrow \infty$ for every bounded continuous function $g(x)$. If X is the Euclidean space of k dimensions, R_k , then a distribution λ can be defined in terms of a distribution function $F(x)$. we then have the following well known result.

Theorem 4.1. A sequence of distributions λ_n converges to λ if and only if.

$$F_n(x) \rightarrow F(x) \text{ for each}$$

continuity point of $F(x)$.

In the case $X = R_k$ a distribution can be defined in terms of the characteristic function $\phi(t) = \int e^{it \cdot x} d\lambda$. with $t \in R_k$. We have the following theorem.

Theorem 4.2. A sequence of distribution λ_n converges to λ if and only if $\phi_n(t) \rightarrow \phi(t)$ for each t and $\phi(t)$ is continuous at 0.

A distribution λ on R_k may sometimes be defined uniquely in terms of its moments $\mu_{\alpha_1, \dots, \alpha_k}$, $\alpha_i = 0, 1, 2, \dots; i = 1, \dots, k$. The following theorem is due to Shohat and Tamarkin.

Theorem 4.3. A sequence of distributions λ_n with moment sequence $\{\mu_{\alpha_1, \dots, \alpha_k}^{(n)}\}$ converges to λ with moment sequence $\mu_{\alpha_1, \dots, \alpha_k}$ if the sequence $\{\mu_{\alpha_1, \dots, \alpha_k}^{(n)}\}$ converges to $\mu_{\alpha_1, \dots, \alpha_k}$ for each $(\alpha_1, \dots, \alpha_k)$ as $n \rightarrow \infty$ and the moment sequence $\mu_{\alpha_1, \dots, \alpha_k}$ defines a distribution uniquely.

To verify this last condition of the theorem we would require another result due to Shohat and Tamarkin that gives the restrictions under which a moment sequence determines a distribution uniquely. It however would suffice to note that the moments of the multivariate normal distribution satisfy those restrictions.

In Cramer (1946) we find that the distribution λ of a random variable X on R_k is uniquely defined if the distribution of every linear function LX of X is available. In this connection we have the following result due to Varadarajan (1958):

Theorem 4.4. If for each linear function L , the distribution of LX_n converges to a distribution λ_L then the distributions λ_n of X_n converge to a distribution λ and λ_L is the distribution of LX computed under λ .

The Euclidean space R_k can be written as the product space $R_p \times R_r$ if $p + r = k$. A random variable X on R_k is of the form (Y, Z) with Y in R_p and Z in R_r . Let μ be a distribution on R_r . Let $\nu(\cdot, z)$ for each $z \in R_r$ be a distribution on R_p . The collection $\mu, \nu(\cdot, z)$ define a distribution λ on R_k as follows. For each Borel set A in R_k define

$$\lambda(A) = \int_{R_r} \nu(A_z, z) d\mu(z) \quad \dots \quad (1)$$

where $A_z = \{y : \underline{x} = (y, z) \in A\}$. This λ is a distribution on R_k and is called the joint distribution of (Y, Z) . μ is called the marginal distribution of Z and $\nu(\cdot, z)$ is called the conditional distribution of Y given $Z = z$. Theorems 4.5 and 4.6 due to Sethuraman (1961), (1) are useful in determining the convergence of λ_n when the associated distributions $\mu_n, \nu_n(\cdot, z)$ converge.

Definition. A sequence of distributions μ_n is said to converge strongly to μ if $\mu_n(B) \rightarrow \mu(B)$ for every Borel set B .

Theorem 4.5. Let μ_n converge strongly to μ . Let $\nu_n(\cdot, z)$ converge to $\nu(\cdot, z)$ for almost all $[u] z$. Then λ_n (associated with $\mu_n, \nu_n(\cdot, z)$) converges to λ where λ is defined by (1).

Let $\{\phi_n(t, \theta)\}$ be a family of sequences of characteristic functions on R_p with θ in an closed bounded interval I in R_r .

Definition. $\{ \phi_n(t, \theta) \}$ is said to converge uniformly to a characteristic function $\phi(t, \theta)$ in the UC* sense relative to θ in I if

- a) $\sup_{\theta \in I} |\phi_n(t, \theta) - \phi(t, \theta)| \rightarrow 0$ for each t
- b) $\phi(t, \theta)$ is equicontinuous in $\theta \in I$ at $t = 0$
- c) $\phi(t, \theta)$ is continuous in θ for each t

Theorem 4.6. Let the sequence of characteristic functions of $\nu_n(\cdot, z)$ converge to the characteristic function of $\nu(\cdot, z)$ in the UC* sense relative to z in any bounded closed interval of R_r . Let μ_n converge to μ . Then λ_n (associated with $\mu_n, \nu_n(\cdot, z)$) converges to λ where λ is defined by (1).

After this brief review of the general techniques we now present some of the fundamental theorems of limiting distributions which will be used in our applications.

Let $(X_1, Y_1), (X_2, Y_2)$ be a sequence of random variables on R_2 . Let the limiting distribution of X_n be $F(x)$. Let the limiting distribution of Y_n be degenerate at c . We have the following theorem.

Theorem 4.7. The limiting distribution function of $X_n + Y_n$ is $F(x - c)$. The limiting distribution function of X_n / Y_n is $F(cx)$ if $c > 0$.

We state ~~three more theorems without proof.~~

Theorem 4.8. If $X_1, X_2, \dots, X_n, \dots$

is a sequence of independent and identical random variables on R_1 , and if $E(X_1) = \mu$ exists then $\frac{X_1 + \dots + X_n}{n}$ tends to μ in probability. In fact $\frac{X_1 + \dots + X_n}{n}$ tends μ with probability one.

Theorem 4.9. Let X_1, X_2, \dots be a sequence of independent and identical random variables with finite mean μ and variance σ^2 . Then the limiting distribution function of $(X_1 + \dots + X_n - n\mu)/n^{1/2} \sigma$ is $(2\pi)^{-1/2} \int_{-\infty}^x \exp(-t^2/2) dt$.

This theorem is one of the forms of the Central Limit Theorem. Another form of the same theorem is cast as Theorem 4.10. Let

$$\begin{array}{cccc} X_{11}, & \dots & X_{1\nu_1} & \\ X_{21}, & \dots & \dots & X_{2\nu_2} \\ \dots & \dots & \dots & \dots \\ X_{n1}, & \dots & \dots & X_{n\nu_n} \\ \dots & \dots & \dots & \dots \end{array}$$

be a triangular scheme of random variables on R_1 which are independent in each row. Let $E(X_{ni}) = 0$ for all n and i .

$$E(X_{ni}^2) = \mu_{2,n,i} \text{ and } E(|X_{ni}|^3) = \mu_{3,n,i}.$$

$$\text{Let } \rho_n^3 = \frac{1}{\nu_n} \sum_{i=1}^{\nu_n} \mu_{3,n,i} \quad \sigma_n^2 = \frac{1}{\nu_n} \sum_{i=1}^{\nu_n} \mu_{2,n,i}$$

Theorem 4.10. The limiting distribution of $\frac{1}{\sqrt{\nu_n}}(X_{n1} + \dots + X_{n\nu_n})$ is

$$(2\pi)^{-\frac{1}{2}} \sigma^{-1} \int_{-\infty}^x \exp(-t^2/2\sigma^2) dt$$

if $\sigma_n^2 \rightarrow \sigma$, $\nu_n \rightarrow \infty$ and $\rho_n/\sigma_n \rightarrow 0$ as $n \rightarrow \infty$.

4.2. Central limit theorems for m -dependent random variables.

Definition. Let X_1, X_2, \dots be a sequence of random variables.

This sequence of random variables is said to be m -dependent if

(X_1, \dots, X_r) is independent of (X_s, X_{s+1}, \dots) whenever $s-r > m$.

Let

$$A_i = 2 \sum_{j=0}^{m-1} \text{cov}(X_{i+j}, X_{i+m}) + \text{var}(X_{i+m}) \quad \dots \quad (2)$$

where cov and var stand for covariance and variance respectively.

The central limit theorem for m -dependent random variables due to

Hoeffding and Robbins (1948) is the following:

Theorem 4.11. Let X_1, X_2, \dots be a m -dependent sequence of variables.

Let $E(X_i) = 0$, $E(|X_i|^3) \leq R^3 < \infty$ for all $i = 1, 2, \dots$. Let the limit

$$\lim_{p \rightarrow \infty} \frac{1}{p} \sum_{h=1}^p A_{i+h} = A \quad \dots \quad (3)$$

exist uniformly in i . Then the limiting distribution distribution

function of $\sum_{i=1}^n X_i / (nA)^{\frac{1}{2}}$ is

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^x \exp(-t^2/2) dt.$$

Proof. The proof of this theorem is a modification of the proof of the central limit theorem for Markov chains due to Bernstein. Let k be some integer less than n to be chosen to satisfy (5) and (6). Integers $\nu \geq 1$ and r , $0 \leq r \leq k-1$ can be found such that $n = \nu k + r$.

$$S = X_1 + \dots + X_n = S' + T$$

where $S' = U_1 + \dots + U_\nu$

and $U_i = X_{(i-1)k+1} + X_{(i-1)k+2} + \dots + X_{ik-m}$

and $T = \sum_{i=1}^{\nu-1} (X_{ik-m+1} + \dots + X_{ik}) + X_{\nu k-m+1} + \dots + X_{\nu k+r}$.

We shall show that $n^{-\frac{1}{2}} S'$ has a limiting normal distribution with mean 0 and variance Λ and that $n^{-\frac{1}{2}} T$ tends to 0 in probability.

The theorem is then immediate from Theorem 4.7 and the relation

$$n^{-\frac{1}{2}} S = n^{-\frac{1}{2}} S' + n^{-\frac{1}{2}} T.$$

Consider $n^{-\frac{1}{2}} S'$

$$n^{-\frac{1}{2}} S' = \left(\frac{\nu k}{n} \right)^{\frac{1}{2}} \nu^{-\frac{1}{2}} \sum_{i=1}^{\nu} k^{-\frac{1}{2}} U_i \dots \quad (4)$$

If k be chosen so that

$$\frac{\nu k}{n} \rightarrow 1 \dots \dots \quad (5)$$

then by an application of Theorem 4.7, it is enough to show that

$\nu^{-\frac{1}{2}} \sum_{i=1}^{\nu} k^{-\frac{1}{2}} U_i$ has a limiting normal distribution. Now for $s > m$ we

have

$$\begin{aligned}
 & E \left\{ (X_{i+1} + \dots + X_{i+s})^2 \right\} \\
 &= E \left\{ (X_{i+1} + \dots + X_{i+s-1})^2 \right\} + 2E \left\{ (X_{i+1} + \dots + X_{i+s-1})X_{i+s} \right\} \\
 &\quad + E \left\{ X_{i+s}^2 \right\} \\
 &= E \left\{ (X_{i+1} + \dots + X_{i+s-1})^2 \right\} + \Lambda_{i+s-m} \\
 &= E \left\{ (X_{i+1} + \dots + X_{i+m})^2 \right\} + \sum_{h=1}^{s-m} \Lambda_{i+h}
 \end{aligned}$$

Thus $E \{ U_i^2 \} = E \left\{ (X_{(i-1)k+1} + \dots + X_{(i-1)k+m})^2 \right\} + \sum_{h=1}^{k-2m} \Lambda_{(i-1)k+h} .$

From some well known inequalities we have

$$E \{ |X_i|^2 \} \leq E^{2/3} \{ |X_i|^3 \} \leq R^2$$

and $E \{ |X_i X_j| \} \leq [E(X_i^2) E(X_j^2)]^{1/2} \leq R^2 .$

Hence

$$E \left\{ (X_{(i-1)k+1} + \dots + X_{(i-1)k+m})^2 \right\} \leq m^2 R^2$$

and $\left| E(U_i^2) - \sum_1^{k-2m} \Lambda_{(i-1)k+h} \right| \leq m^2 R^2$

and $\left| \frac{1}{\nu} \sum_{i=1}^{\nu} E \left\{ (k^{-\frac{1}{2}} U_i)^2 \right\} - \frac{k-2m}{k\nu} \sum_{i=1}^{\nu} \frac{1}{k-2m} \sum_{h=1}^{k-2m} \Lambda_{(i-1)k+h} \right|$
 $\leq m^2 R^2 / k \quad \dots \quad (6)$

As $\nu \rightarrow \infty, k \rightarrow \infty$ from (6) and (3) it follows that

$$\lim_{\nu \rightarrow \infty} \frac{1}{\nu} \sum_{i=1}^{\nu} E \left\{ (k^{-\frac{1}{2}} U_i)^2 \right\} = A \quad \dots \quad (7)$$

Now

$$E \{ |X_i X_j X_k|^3 \} \leq [E \{ |X_i|^3 \} \cdot E \{ |X_j|^3 \} \cdot E \{ |X_k|^3 \}]^{1/3} \leq R^3.$$

Thus

$$E \{ |k^{-\frac{1}{2}} U_i|^3 \} \leq k^{-3/2} (k-m)^3 R^3 \leq k^{3/2} R^3.$$

Thus

$$\frac{1}{\nu} \sum_1^{\nu} E \{ |k^{-\frac{1}{2}} U_i|^3 \} \leq k^{3/2} R^3 \quad \dots \quad (8)$$

$$\text{Let } k^{3/\nu} \rightarrow 0 \text{ as } \nu, k \rightarrow \infty \quad \dots \quad (9)$$

(A suitable choice for k satisfying (5) and (8) is n^α for some α with $0 < \alpha < \frac{1}{4}$). (6), (8), (9) and the fact $\nu \rightarrow \infty$ are enough to verify the conditions of theorem 4.10. Thus $\nu^{-1/2} \sum_1^{\nu} k^{-\frac{1}{2}} U_i$ has an limiting normal distribution.

Since $k \rightarrow \infty$, $k > 2m$ after a certain stage. Then

$$n^{-1} E \{ T^2 \} \leq \frac{1}{n} [(\nu-1)m^2 + (k+m)^2] \rightarrow 0 \text{ as } \nu, k \rightarrow \infty.$$

Thus $n^{-\frac{1}{2}} T$ tends to 0 in probability. Hence the theorem.

Definition. A sequence of random variables X_1, X_2, \dots is said to be stationary if the joint distribution of $X_i, X_{i+1}, \dots, X_{i+r}$ is independent of i for all r .

The following theorem is immediate.

Theorem 4.12. If X_1, X_2, \dots is a stationary m -dependent sequence of with $E(X_1) = 0$ and $E \{ |X_1|^3 \} < \infty$ then the limiting distribution function of $n^{-\frac{1}{2}}(X_1 + \dots + X_n)$ is

$$(2\pi\Lambda)^{-\frac{1}{2}} \int_{-\infty}^x \exp(-t^2/2\Lambda) dt \quad \text{where}$$

$$\Lambda = \text{var}(X_1^2) + 2[\text{cov}(X_1X_2) + \dots + \text{cov}(X_1X_{m+1})]$$

Bivariate extensions and multivariate extensions of this theorem are available. These can be easily established by the use of Theorem 4.4. We shall now give an example as an application of Theorem 4.12.

Z_1, Z_2, \dots is a sequence of independent and identical random variables with $E(Z) = 0$, $E(Z_1^2) = 1$, $E(|Z_1|^3) < \infty$. Let $X_i = Z_i Z_{i+1}$. X_1, X_2, \dots a 1-dependent stationary sequence of random variables. Thus the limiting distribution function of $n^{-\frac{1}{2}}(Z_1Z_2 + \dots + Z_nZ_{n+1})$ is

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^x \exp(-t^2/2) dt.$$

$$\begin{aligned} \text{Let } W_n &= n^{\frac{1}{2}} \frac{Z_1Z_2 + \dots + Z_n Z_{n+1}}{(Z_1^2 + \dots + Z_n^2)} \\ &= \frac{n^{-\frac{1}{2}} (Z_1 Z_2 + \dots + Z_n Z_{n+1})}{(Z_1^2 + \dots + Z_n^2)/n}. \end{aligned}$$

Since $\sum_1^n Z_i^2/n$ tends to 1 in probability the normalised serial correlation coefficient W_n has a limiting normal distribution with mean zero and variance 1.

4.3. The limiting distributions of U-statistics.

We saw in Chapter 3 that if $f(x_1, \dots, x_m)$ is an unbiased estimator of a parametric function $g(\theta)$, then in many situations $h(x_1, \dots, x_m)$ is the minimum variance unbiased estimator of $g(\theta)$

where

$$h(x_1, \dots, x_m) = \frac{1}{n!} \sum_P f(x_{i_1}, \dots, x_{i_m})$$

and the summation is over all permutations (i_1, \dots, i_m) of the integers $(1, \dots, m)$. If X_1, \dots, X_n are n independent and identical random variables (each on R_1) then as an application of theorem 3.7 the minimum variance unbiased estimator of $g(\theta)$ when θ is restricted to a class containing $(\bar{1})_1$ or $(\bar{1})_2$, is $U(x_1, \dots, x_n) = \frac{1}{\binom{n}{m} C} \sum_C h(x_{i_1}, \dots, x_{i_m})$ where the summation is over all combinations of the integers (i_1, \dots, i_m) out of $(1, \dots, n)$. U is called a U-statistic and h is called the kernel of the U-statistic. Theorem 4.13 is concerned with the limiting distribution of U-statistics.

Let X_1, X_2, \dots be a sequence of independent and identical random variables. Let $E \{ h^2(X_1, \dots, X_m) \} < \infty$.

Define

$$h_c(x_1, \dots, x_c) = E \left\{ h(x_1, \dots, x_c, X_{c+1}, \dots, X_m) \right\}$$

$$c = 1, 2, \dots, m-1.$$

Then $E(h_c(X_1, \dots, X_c)) = E(h(X_1, \dots, X_m)) = \gamma$ say, for all c . Let $\text{Var}(h_c(X_1, \dots, X_c)) = \zeta_c$. It is easy to see that

$$\text{cov}(h(X_1, \dots, X_m), h(X_1, \dots, X_c, X_{m+1}, \dots, X_{2m-c})) = \zeta_c.$$

Thus $\text{Var} (U(X_1, \dots, X_n))$

$$\begin{aligned}
 &= \frac{1}{\binom{n}{m}^2} \sum_{c=1}^m \binom{n}{m} \binom{m}{c} \binom{n-m}{m-c} \zeta_c \\
 &= \frac{1}{\binom{n}{m}} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \zeta_c \\
 &\approx \frac{m^2 \zeta_1}{n} + o(1/n) \quad \text{as } n \rightarrow \infty \quad \dots \quad (10)
 \end{aligned}$$

Theorem 4.13.

$$n^{\frac{1}{2}}(U_n(X_1, \dots, X_n) - \gamma)$$

has a limiting normal distribution with mean 0 and variance $m^2 \zeta_1$.

Proof: Let $Y_n = mn^{-\frac{1}{2}} \sum_{i=1}^n (h_1(X_i) - \gamma)$.

From Theorem 4.9, it follows that Y_n has an limiting normal distribution with mean 0 and variance $m^2 \zeta_1$. We note the following

$$\text{Var} (n^{\frac{1}{2}} U_n) = m^2 \zeta_1 + \dots + o(1)$$

$$\text{Var} (Y_n) = m^2 \zeta_1$$

$$\begin{aligned}
 \text{cov} (n^{\frac{1}{2}} U_n, Y_n) &= \frac{m}{\binom{n}{m}} \sum_c \sum_i \text{cov} (h(X_{i_1}, \dots, X_{i_m}), h_1(X_i)) \\
 &= \frac{m}{\binom{n}{m}} n \binom{n-1}{m-1} \zeta_1 \\
 &= m^2 \zeta_1
 \end{aligned}$$

Thus

$$E (n^{\frac{1}{2}} U_n - Y_n)^2 = o(1).$$

Hence $n^{\frac{1}{2}}U_n - Y_n \rightarrow 0$ in probability. The theorem is now an immediate consequence of theorem 4.7.

Hoeffding (1948) who gave this proof has proved a similar theorem when the random variables X_1, X_2, \dots are independent and not identical under some stronger conditions. Lehmann (1951) defined a generalised U-statistic based on two sequences of random variables. Let $h(x_1, \dots, x_{m_1}; y_1, \dots, y_{m_2})$ be a function symmetric in the m_1 x-arguments and m_2 y-arguments separately. Let $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ be $n_1 + n_2$ independent random variables. Let

$$U_{n_1, n_2} = \frac{1}{\binom{n_1}{m_1} \binom{n_2}{m_2} c} \sum h(X_{i_1}, \dots, X_{i_{m_1}}; Y_{j_1}, \dots, Y_{j_{m_2}})$$

where the summation is over all combinations of m_1 indices (i_1, \dots, i_{m_1}) from $(1, \dots, n_1)$ and m_2 indices (j_1, \dots, j_{m_2}) from $(1, \dots, n_2)$.

U_{n_1, n_2} is called a generalised U-statistic.

Theorem 4.14. Let X_1, X_2, \dots be a sequence of independent and identical random variables. Let Y_1, Y_2, \dots be another independent sequence of independent and identical random variables. Let

$$E \{ h(X_1, \dots, X_{m_1}; Y_1, \dots, Y_{m_2}) \} = \gamma \text{ and } E \{ h^2(X_1, \dots, X_{m_1}; Y_1, \dots, Y_{m_2}) \}$$

If $n_1, n_2 \rightarrow \infty$ such that $n_1/n_2 \rightarrow c, 0 < c < \infty$ then $[n_1 n_2 / (n_1 + n_2)]^{\frac{1}{2}}$

$(U_{n_1, n_2} - \gamma)$ has a limiting normal distribution.

Prof: - The proof proceeds on the same lines as that of Theorem 4.13 and so is omitted.

Examples of U-statistics in non-parametric methods are numerous, the minimum variance unbiased estimators of section 3.1 being the first examples. The Mann-Whitney U-test described in 3.2 involves a statistic which is a U-statistic in the generalised sense and theorem 4.14 can be used to obtain its distribution both under the null hypothesis and under the alternatives.

4.4. The limiting distribution of the order statistics and the limiting joint distribution of U-statistics and order statistics.

The results of this section are useful in many non-parametric problems. Since the proofs are involved we shall just state the results, at times hinting the nature of the proof. Important and conclusive work about the limiting distributions of order statistics were done Gnedenko (1943) and Smirnov (1949), (1952) and about the limiting joint distributions of U-statistics and order statistics by Sethuraman (1961).

Let X_1, X_2, \dots be a sequence of independent and identical random variables with a distribution function $F(x)$. Let $X_{1,n}, X_{2,n}, \dots, X_{n,n}$ be the order statistics based on X_1, \dots, X_n . For convenience we quite often drop the suffix n and denote $X_{i,n}$ by $X_{(i)}$. Theorems 4.15, 4.16 and 4.17 deal with the asymptotic distribution of $X_{(a_n)}$ where a_n is a sequence of integers with $1 \leq a_n \leq n$.

Definition. A sequence of random variables Z_n is said to have an asymptotic distribution function $F(x)$ if there exist constants

$a_n > 0, b_n$ such that

$$P \{ a_n Z_n + b_n \leq x \} \rightarrow F(x)$$

at each continuity point of $F(x)$.

Definition. Two distributions with distribution functions $F(x)$ and $G(x)$ are said to be of the same type if there exist numbers $a > 0, b$ such that

$$F(ax + b) = G(x)$$

Theorem 4.15

$$\text{Let } \sqrt{n} \left(\frac{a}{n} - p \right) \rightarrow 0 \text{ as } n \rightarrow \infty; 0 < p < 1 \quad \dots \quad (11)$$

There exist constants $\alpha_n > 0, \beta_n$ such that $(X_{(a_n)} - \beta_n)/\alpha_n$ has a limiting distribution if and only if $u_n(x) = [F(\alpha_n x + \beta_n) - p] \times [n/p(1-p)]^{\frac{1}{2}} \rightarrow u(x)$ as $n \rightarrow \infty$, where $u(x)$ is a non-decreasing function. In this case the limiting distribution function $\Phi_{\Lambda}(x)$ is given by

$$\Phi_{\Lambda}(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{u(x)} \exp(-t^2/2) dt \quad \dots \quad (12)$$

Further $u(x)$ can assume only one of the following forms.

$$u(x) = \begin{cases} -\infty & x < 0 \\ cx^\alpha & x > 0, c > 0, \alpha > 0 \end{cases} \quad \dots \quad (13)$$

$$u(x) = \begin{cases} -c|x|^\alpha & x < 0, c > 0, \alpha > 0 \\ \infty & x > 0 \end{cases} \quad \dots \quad (14)$$

$$u(x) = \begin{cases} -c_1 |x|^\alpha & x < 0 \quad c_1 > 0, \alpha > 0 \\ c_2 x^\alpha & x > 0 \quad c_1 > 0 \end{cases} \quad \dots \quad (15)$$

$$u(x) = \begin{cases} -\infty & x < -1 \\ 0 & -1 < x < 1 \\ \infty & x > 1 \end{cases} \quad \dots \quad (16)$$

Proof: The necessary and sufficient conditions A_1, A_2, A_3 and A_4 on $F(x)$ for $u(x)$ to be of the form (13), (14), (15) and (16) respectively, together with the proof are found in Smirnov (1949).

Theorem 4.16. Let $a_n = k$, a constant for all $n \dots (17)$.

There exist constants $\alpha_n > 0, \beta_n$ such that $(X_{(a_n)} - \beta_n)/\alpha_n$ has a limiting distribution if and only if

$$v_n(x) = nF(\alpha_n x + \beta_n) \rightarrow v(x)$$

as $n \rightarrow \infty$, where $v(x)$ is a nondecreasing function of x . In this case the limiting distribution function $\bar{\Phi}_B(x)$ is given by

$$\bar{\Phi}_B(x) = \sum_{r=k}^{\infty} e^{-v(x)} (v(x))^r / r! \quad \dots \quad (18)$$

Further $v(x)$ can assume only one of the following three forms

$$v(x) = \begin{cases} 0 & x < 0 \\ x^\alpha & x > 0, \alpha > 0 \end{cases} \quad \dots \quad (19)$$

$$v(x) = \begin{cases} |x|^{-\alpha} & x < 0, \alpha > 0 \\ 0 & x > 0 \end{cases} \quad \dots \quad (20)$$

$$v(x) = e^x, \quad -\infty < x < \infty \quad \dots \quad (21)$$

Proof: The necessary and sufficient conditions B_1, B_2 and B_3 on $F(x)$ for $v(x)$ to be of the form (19), (20) and (21) respectively, together with the proof are found in Gnedenko (1943), Smirnov (1949). We may note that the case $a_n = n-k$, k a constant, is treated in the same way as above.

Theorem 4.17. Let $\frac{a_n}{n} \rightarrow 1, n - a_n \rightarrow \infty$ as $n \rightarrow \infty \dots$ (22)

Let $F(x)$ satisfy condition C, namely

$$\frac{1 - F(x(1 + \varepsilon(x)))}{1 - F(x)} = 1 - \alpha \varepsilon(x) + o(\varepsilon(x)) \dots \quad (23)$$

as $x \rightarrow \infty$, where $\varepsilon(x)$ is any function of x which tends to 0 as $x \rightarrow \infty$ and $\alpha > 0$ is a fixed constant. Then there exist constants $\alpha_n > 0, \beta_n$ such that the limiting distribution function of

$$(X_{(a_n)} - \beta_n) / \alpha_n \text{ is } (2\pi)^{-\frac{1}{2}} \int_{-\infty}^x \exp(-t^2/2) dt.$$

Proof: The proof of this theorem can be found in Kawata (1951)

Theorem 4.18. Let X_1, X_2, \dots be a sequence of independent and identical random variables with the same distribution function $F(x)$.

Let U_n be a U-statistic with a kernel $h(x_1, \dots, x_m)$. Let

$$E \{ h^2(X_1, \dots, X_m) \} < \infty$$

A) Let $F(x)$ satisfy condition A_1, A_2, A_3 or A_4 . Let a_n satisfy (11). Then the joint limiting distribution function of

$$\left\{ (X_{(a_n)} - \beta_n) / \alpha_n, n^{\frac{1}{2}} (U_n - \gamma) \right\} \text{ is } \Lambda(u(x), y)$$

where $\Lambda(z, y) = \frac{1}{2\pi} \int_{-\infty}^z \exp(-t^2/2) \int_{-\infty}^{Q(y,t)} \exp(-u^2/2) du dt \dots (24)$

and $Q(y, t) = \frac{y - t\theta}{\delta}$ where θ and δ depend on the distribution of $h_1(X_1)$ and the p th quantile of X_1 .

B) Let $F(x)$ satisfy condition C_1, C_2 or C_z . Let a_n satisfy (17). Then the limiting joint distribution function of

$$\left\{ (X_{(a_n)} - \beta_n)/\alpha_n, n^{\frac{1}{2}}(U_n - \gamma) \right\} \text{ is}$$

$$\sum_{r=k}^{\infty} e^{-v(x)} \frac{v^r(x)}{r!} x \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y/m} \exp(-t^2/2) dt \quad (25)$$

C) Let $F(x)$ satisfy condition C. Let a_n satisfy (22).

Then the limiting joint distribution of $\left\{ (X_{(a_n)} - \beta_n)/\alpha_n, n^{\frac{1}{2}}(U_n - \gamma) \right\}$ is

$$\frac{1}{2\pi} \int_{-\infty}^x \exp(-t^2/2) dt \times \int_{-\infty}^{y/m} \exp(-u^2/2) du \quad \dots \quad (26)$$

Proof: Let k_n the number of random variables among X_1, \dots, X_n that are less than or equal to $\alpha_n x + \beta_n$. The distribution of k_n and $X_{(a_n)}$ are related to one another.

The conditional distribution of $n^{\frac{1}{2}}(U_n - \gamma)$ given $c_n k_n + d_n = z$ ($c_n > 0$ and d_n are suitable constants) is shown to converge in the UC* sense with respect to z in any bounded interval in the cases A and B, and to converge in the ordinary sense for each z in the case C.

The distribution of $c_n k_n + d_n$ converges to a distribution in the ordinary sense in cases A and B and in the strong sense in case C. Applications of theorems 4.5 and 4.6 yield the joint limiting distribution of $n^{1/2}(U_n - \gamma)$ and $c_n k_n + d_n$. The relationship between the distributions of k_n and $X_{(a_n)}$ establish the theorem. The complete details of the proof are found in Sethuraman (1961).

4.5. Combinatorial Central Limit Theorems.

We presented several tests in section 3.2 that were called permutation or randomisation tests since the test depended on a statistic whose distribution was to be determined under equally likely permutations. This class of problems has led to completely new and fascinating problems in limit distributions. The first result in this connection is due to Wald and Wolfowitz (1944). Noether (1949) gave a generalisation of this result and it forms our theorem 4.19.

Let $\mathcal{H}_n = (h_{n1}, \dots, h_{nn})$, for $n = 1, 2, \dots$ be a triangular sequence of real numbers.

Definition: A triangular sequence of real number \mathcal{H}_n is said to satisfy condition W if

$$\frac{\frac{1}{n} \sum_{i=1}^n (h_{in} - \bar{h}_n)^r}{\left[\frac{1}{n} \sum_{i=1}^n (h_{in} - \bar{h}_n)^2 \right]^{r/2}} = o(1)$$

as $n \rightarrow \infty$ for $r = 3, 4, \dots$ where $\bar{h}_n = \sum_{i=1}^n h_{in} / n$.

Definition: A triangular scheme H_n is said to satisfy the condition N if

$$\frac{\sum_{i=1}^n (h_{in} - \bar{h}_n)^r}{\left[\sum_{i=1}^n (h_{in} - \bar{h}_n)^2 \right]^{r/2}} = o(1)$$

as $n \rightarrow \infty$ for $r = 3, 4, \dots$

Condition W is a stronger condition than condition N.

Corresponding to a triangular sequence A_n we define a random variable $X_n = (X_1, \dots, X_n)$ which taken each permutation $(a_{1n}, a_{2n}, \dots, a_{nn})$ with the same probability $1/n!$. Then corresponding to two triangular sequences A_n and C_n we investigate the limiting distribution, as $n \rightarrow \infty$, of

$$L_n = c_{1n} X_1 + \dots + c_{nn} X_n.$$

It is easy to see that $E(L_n) = n \bar{c}_n \bar{a}_n$,

$$\text{Var}(L_n) = \frac{1}{n-1} \sum_{i=1}^n (c_{in} - \bar{c}_n)^2 \sum_{i=1}^n (a_{in} - \bar{a}_n)^2$$

where $\bar{a}_n = \sum a_{in}/n$ and $\bar{c}_n = \sum c_{in}/n$.

Theorem 4.19. If C_n satisfies condition W and A_n satisfies condition N then the limiting distribution of $L_n^0 = (L_n - E(L_n)) / (\text{Var}(L_n))^{1/2}$ is normal with mean 0 and variance 1.

Proof: Let C_{e_1, \dots, e_m} be the symmetric function

$$C_{e_1, \dots, e_m} = \sum_P c_{i_1}^{e_1} \dots c_{i_m}^{e_m}$$

where the summation is over all permutations (i_1, \dots, i_m) of m integers chosen from $(1, \dots, n)$. Let A_{e_1, \dots, e_m} denote the corresponding symmetric function derived from A_n, L_n is unaffected by change of scale and location. Hence we can assume without loss of generality that $A_1 = C_1 = 0, A_2 = C_2 = n$. These relations together with the conditions W and N imply

$$C_r = o(n) \quad r = 3, 4, \dots, \quad A_r = o(n^{r/2}), \quad r = 3, 4, \dots \quad (27)$$

We find that $E(L_n) = 0$ and $\text{Var}(L_n) = \frac{1}{n-1} C_2 A_2$ is $o(n)$ (i.e. is of the order of n). From theorem 4.7 we see that it is enough to prove that $n^{-\frac{1}{2}} L_n$ has a limiting normal distribution. This we proceed to do by an application of theorem 4.3. We shall prove that the r th moment of $n^{-\frac{1}{2}} L_n$ approaches the r th moment of the standard normal distribution.

$$\begin{aligned} \mu_r &= n^{-r/2} E(L_n^r) \\ &= n^{-r/2} \sum_{i_1=1}^n \dots \sum_{i_r=1}^n E(c_{i_1}^{X_{i_1}} \dots c_{i_r}^{X_{i_r}}) \\ &= n^{-r/2} [C_r E(X_1^r) + \dots + c(r, e_1, \dots, e_m) C_{e_1, \dots, e_m} E(X_1^{e_1} \dots X_m^{e_m}) \\ &\quad + \dots + c_{1 \dots 1} E(X_1 \dots X_r)] \dots \quad (28) \end{aligned}$$

where $e_1 + \dots + e_m = r$, e_k for $k = 1, \dots, m$ is a positive integer and $c(r, e_1, \dots, e_m)$ is the number of ways that the r indices e_1, \dots, i_r can be tied into m groups so that the m groups in the order in which their first element occurs in the sequence i_1, \dots, i_r , are, of size e_1, \dots, e_m , respectively. Since $E(X_1^{e_1} \dots X_m^{e_m}) \sim n^{-m} A_{e_1, \dots, e_m}$, we have

$$n^{-r/2} c_{e_1, \dots, e_m} E(X_1^{e_1} \dots X_m^{e_m}) \sim n^{-r/2+m} c_{e_1, \dots, e_m} A_{e_1, \dots, e_m} \dots \quad (29)$$

and we designate by $B(r, e_1, \dots, e_m)$ the right hand side of this relation. To complete the proof of the theorem we need the following lemma which we shall prove later.

Lemma 4.20. $B(r; e_1, \dots, e_m) \sim 0$ unless $m = \frac{r}{2}$, $e_1 = \dots = e_m = 2$ in which case $B(r, 2, \dots, 2) \sim 1$.

By (28) μ_r is the sum of a finite number of expressions $B(r, e_1, \dots, e_m)$. Therefore, if $r = 2s+1$, ($s = 1, 2, \dots$), $\mu_{2r+1} \sim 0$ since at least one of the e 's in each B must be odd. If $r = 2s$, $\mu_{2s} \sim c(2s, 2, \dots, 2)$. Since the first index of the expression being summed in (28) can be tied with any of the $(2s-1)$ others, the next free with any of the $(2s-3)$ others, etc., it is seen that $\mu_{2s} \sim (2s-1) \cdot (2s-3) \cdot \dots \cdot 3$. But these are the moments of the standard normal distribution. Hence the theorem.

Proof of lemma 4.20. Let $\Lambda(j_1, \dots, j_h) = \Lambda_{j_1}, \dots, \Lambda_{j_h}$. Then by the theory of symmetric functions $\Lambda_{e_1, \dots, e_m}$ can be expressed uniquely as a linear combination of a finite number of $\Lambda(j_1, \dots, j_h)$, where $j_1 + \dots + j_h = e_1 + \dots + e_m = r$, and the j 's correspond to sums of e 's. Since $\Lambda_1 = 0$ we need consider only $\Lambda(j_1, \dots, j_h)$ with $j_g \geq 2$ ($g = 1, \dots, h$). If some $j_g > 2$ then $\Lambda(j_1, \dots, j_h) = o(n^{r/2})$. If all $j_g = 2$ then $\Lambda(2, \dots, 2) = \Lambda_2^{r/2} = n^{r/2}$, r is even, and all the e 's must be 1's or 2's; therefore $m > r/2$ unless $m = 2r$, $e_1 = \dots = e_m = 2$. Thus we have $\Lambda_{e_1, \dots, e_m} = o(n^{r/2})$ if $m < r/2$ and certainly $\Lambda_{e_1, \dots, e_m} = o(n^m)$ if $m \geq r/2$ unless $m = r/2$ and $e_1 = \dots = e_m = 2$ in which case $\Lambda_{2, \dots, 2} \sim \Lambda(2, \dots, 2) = n^{r/2}$. Similarly writing C_{e_1, \dots, e_m} as a sum of products of the form $C(j_1, \dots, j_h)$ we obtain the relations $C_{e_1, \dots, e_m} = o(n^m)$ if $m < r/2$ and $= o(n^{r/2})$ if $m > r/2$. Combining these results we have $\Lambda_{e_1, \dots, e_m} C_{e_1, \dots, e_m} = o(n^{m+r/2})$ unless r is even and all the e 's are equal to 2, in which case $\Lambda_{2, \dots, 2} C_{2, \dots, 2} \sim n^r$. This proves the lemma.

The condition N can be given two simpler forms, both of which are more convenient for application, as in the following theorem.

Theorem 4.21. The condition N for \mathcal{H}_n is equivalent to either of the following two conditions

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n |h_{in} - \bar{h}_n|^r}{\left[\sum_{i=1}^n (h_{in} - \bar{h}_n)^2 \right]^{r/2}} = 0 \text{ for some } r > 2 \dots (30)$$

$$\frac{\max_1^n (h_{1n} - \bar{h}_n)^2}{\sum_1^n (h_{in} - \bar{h}_n)^2} \rightarrow 0 \quad \dots \quad (31)$$

Proof: The proof is simple and therefore omitted.

Generalisations of theorem 4.19 of the following type are immediate by an application of theorem 4.4.

Theorem 4.22. Let the triangular sequences A_n, C_n and D_n satisfy conditions N, W and W respectively. Further let

$$\rho_n = \frac{\sum_1^n (c_{in} - \bar{c}_n)(d_{in} - \bar{d}_n)}{[\sum_1^n (c_{in} - \bar{c}_n)^2 \sum_1^n (d_{in} - \bar{d}_n)^2]^{\frac{1}{2}}} \rightarrow \rho \quad \dots \quad (32)$$

as $n \rightarrow \infty$. Then the limiting joint distribution of

$$L_n^o = \frac{L_n - E(L_n)}{(\text{Var}(L_n))^{\frac{1}{2}}}, \quad L_n'^o = \frac{L_n' - E(L_n')}{(\text{Var}(L_n'))^{\frac{1}{2}}}$$

where $L_n = \sum_1^n c_{in} X_i$, $L_n' = \sum_1^n d_{in} X_i$, is bivariate normal with means 0, variance 1 and correlation ρ .

Theorem 4.23, the remarkable generalisation of Theorem 4.19, is due to Hoeffding and is given below. Let $b_n(i, j)$, $i, j = 1, \dots, n$, be n^2 real numbers defined for each n . Let (R_1, \dots, R_n) designate the random variable that takes each permutation of $(1, \dots, n)$ the same probability $1/n!$. Let $L_n = \sum_{i=1}^n b_n(i, R_i)$. Define

$$d_n(i,j) = b_n(i,j) - \frac{1}{n} \sum_{g=1}^n b_n(g,j) - \frac{1}{n} \sum_{h=1}^n b_n(i,h) + \frac{1}{n^2} \sum_{g,h=1}^n b_n(g,h).$$

Then it is easy to see that

$$E(L_n) = \frac{1}{n} \sum_{i,j=1}^n b_n(i,j), \text{ Var}(L_n) = \frac{1}{n-1} \sum_{i,j=1}^n d_n^2(i,j).$$

Theorem 4.23. Let

$$\lim_{n \rightarrow \infty} \frac{\frac{1}{n} \sum_{i,j=1}^n d_n^2(i,j)}{\left[\frac{1}{n} \sum_{i,j=1}^n d_n^2(i,j) \right]^{r/2}} = 0, \quad r = 3, 4 \quad \dots \quad (33)$$

Then $L_n^0 = \frac{L_n - E(L_n)}{(\text{Var}(L_n))^{1/2}}$ has a limiting normal distribution with mean

0 and variance 1.

Proof: The proof of this theorem also makes use of Theorem 4.3. The proof though elementary is very lengthy and can be found in Hoeffding (1951). We may remark that condition (33) is satisfied if

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq i, j \leq n} d_n^2(i,j)}{\frac{1}{n} \sum_{i,j=1}^n d_n^2(j,j)} = 0 \quad \dots \quad (34)$$

Theorem 4.24. X_1, X_2, \dots is a sequence of independent and identical random variables with $\text{Var}(X_1) > 0$ and $E(|X_1|) < \infty$. Then with probability 1, the triangular sequence $\chi_n = (X_1, \dots, X_n)$ satisfies condition N.

Proof:- From the second part of Theorem 4.8 it follows that, with probability 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E(X_1), \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^2 = E(X_1^2) \text{ and}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^3 = E(X_1^3).$$

Thus

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \text{Var}(X_1)$$

and $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}_n|^3 = E(|X_1 - E(X_1)|^3)$ with probability 1.

Since $\text{Var}(X_1) > 0$, we have with probability 1 that

$$\lim_{n \rightarrow \infty} \frac{n^{-1} \sum_{i=1}^n |X_i - \bar{X}_n|^3}{\left[n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]^{3/2}} = \frac{E(|X_1 - E(X_1)|^3)}{[\text{Var}(X_1)]^{3/2}}.$$

This together with Theorem 4.21 proves the theorem.

We give two examples to illustrate the use of these theorems. The rank correlation between a pair of observations $(x_1, y_1) \dots (x_n, y_n)$ is defined to be

$$\frac{\sum (r_i - \frac{n+1}{2})(s_i - \frac{n+1}{2})}{\left(\frac{n^2 - 1}{12} \right)}$$

where (r_1, \dots, r_n) and (s_1, \dots, s_n) are the rank statistics based on (x_1, \dots, x_n) and (y_1, \dots, y_n) respectively. Under the null hypothesis of perfect independence, (r_1, \dots, r_n) and (s_1, \dots, s_n) are random

variables taking all permutations of $(1, \dots, n)$ with the same probability $1/n!$. We now find the asymptotic distribution of the rank correlation coefficient under the null hypothesis. Let in Theorem 4.19 $C_n = (1, \dots, n)$ and $A_n = (1, \dots, n)$. Then the rank correlation coefficient is a linear function of L_n .

Now $\bar{c}_n = \frac{n+1}{2}$; $\frac{1}{n} \sum_1^n (i - \frac{n+1}{2})^2 = \frac{n^2-1}{12}$; $\frac{1}{n} \sum_1^n i^r = O(n^r)$. Therefore

$\frac{1}{n} \sum_1^n (i - \frac{n+1}{2})^r = O(n^r)$. Thus C_n satisfies W. Hence A_n satisfies N.

From theorem 4.19 it follows that L_n and hence the rank correlation coefficient ~~are~~ asymptotically normal.

Another example take the Pitman test criterion (see section 3.) with $n_1 = n_2 = \frac{n}{2}$ $\frac{1}{n} \sum_1^n x_i - \frac{1}{n} \sum_{n+1}^{2n} x_i$. Assuming that X_1, X_2, \dots are

independent and identically distributed with $E(X_1^3) < \infty$ and $\text{Var}(X_1) > 0$ we shall establish that the asymptotic distribution of this statistic, under the randomisation distribution, is normal with probability 1.

Setting up a correspondence with Theorem 4.24 we see that we should show that the triangular scheme $C_{2n} = (\frac{1}{n}, \dots, \frac{1}{n}, -\frac{1}{n}, \dots, -\frac{1}{n})$ satisfies condition W.

Now $\bar{c}_n = 0$, $\frac{1}{2n} \sum c_{in} = \frac{1}{n^2}$ if r is even and $= 0$ if r is odd.

These show that condition W is satisfied. The same asymptotic distribution can be established also when $n_1, n_2 \rightarrow \infty$ so that $n_1/n_2 \rightarrow c$, $0 < c < \infty$.

PART III

RECENT WORK AT THE I.S.I.

Chapter V

ESTIMATES FOR THE DENSITY FUNCTION OF A DISTRIBUTION

5.1. Introduction.

Let X_1, X_2, \dots, X_n be n independent and identical random variables with a distribution function $F(x)$. An estimate of $F(x)$ that suggests itself at once is the empirical distribution function $S_n(x)$. The famous Glivenko-Cantelli lemma asserts that $\sup_x |S_n(x) - F(x)|$ tends to 0 with probability one or that $S_n(x)$ is a strongly uniform consistent estimate of $F(x)$. In this chapter we propose a class of estimates $f_n(x)$ for the density function $f(x)$ of $F(x)$, which are strongly mean consistent for $f(x)$, that is, possess the property $\int |f_n(x) - f(x)| dx \rightarrow 0$ as $n \rightarrow \infty$ with probability one. The method of construction and the proof of mean consistency depend on the theory of integral operators of the Fejer type, a discussion on which may be found in Acheiser (1956). The results described in this chapter are due to K. R. Parthasarathy of the Indian Statistical Institute, and have not yet been published.

5.2. Definition and properties of kernels of the Fejer type.

Definition. A function $K(x)$ is said to be of the Fejer type if it satisfies the following:

- a) $K(x) = K(-x)$, $K(x) \geq 0$
- b) $\int_{-\infty}^{\infty} K(x) dx = 1$
- c) $K(x)$ is bounded in the interval $[-1, 1]$
- d) $x^2 K(x)$ is bounded everywhere
- e) $K(x)$ is of bounded variation.

Examples of such kernels are $\frac{1}{\pi} \frac{a}{a^2 + x^2}$, $a > 0$; $(\frac{a}{\pi})^{1/2} e^{-ax^2}$, $a > 0$;

$\frac{2 \sin^2 x/2}{\pi x^2}$; etc. It is easy to prove the following.

Lemma 5.1. If $K(x)$ is a kernel of the Fejer type then $(\int_{-\infty}^{\infty} K^2(x) dx)^{-1} K^2(x)$ and $\int_{-\infty}^{\infty} K(x-y)K(y)dy$ are kernels of the Fejer type.

The following theorem is important in the proof of the mean consistency of the density estimate in section 5. Let W_1 be the class of functions $f(x)$ such that $\frac{f(x)}{1+x^2}$ is integrable. For any $f \in W_1$ we define the operator K_λ by the relation

$$K_\lambda(f) = \lambda \int_{-\infty}^{\infty} f(u) K(\lambda(x-u)) du, \lambda > 0 \quad \dots \quad (1)$$

K_λ is said to be an operator with a kernel of the Fejer type.

Theorem 5.2. Let $f(x) \in W_1$ and $f(x, \lambda) = K_\lambda(f(x))$ where K_λ is an operator with a kernel of the Fejer type.

i) If $\lim_{h \rightarrow 0} \int_0^h \left| \frac{f(x+t) + f(x-t)}{2} - g(x) \right| dt = 0, h > 0$

then $\lim_{\lambda \rightarrow \infty} f(x, \lambda) = g(x).$

ii) If $f(x)$ is the density function of a probability distribution then so is $f(x, \lambda)$ and

$$\lim_{\lambda \rightarrow \infty} \int |f(x, \lambda) - f(x)| dx = 0$$

Proof:- For the proof of the first part of the theorem we refer to Achieser (1956). If $f(x)$ is a density function, it is obvious that $f(x, \lambda)$ is also a density function. Since $f(x)$ is integrable we have from a theorem of Lebesgue that

$$\lim_{h \rightarrow 0} \int_0^h \left| \frac{f(x+t) + f(x-t)}{2} - f(x) \right| dt = 0$$

almost every where with respect to the Lebesgue measure. Hence by the first part

$$\lim_{\lambda \rightarrow \infty} f(x, \lambda) = f(x) \text{ a.e. (Lebesgue measure).}$$

Since $f(x, \lambda)$ and $f(x)$ are densities the use of Scheffé's Theorem (See Scheffé (1947)) establishes the second part of the theorem.

We now state a theorem due to Prohorov (1959) which is useful in proving some theorems in section 5.4.

Theorem 5.3. If X_1, X_2, \dots, X_n are independent random variables with mean zero, $|X_i| \leq c, i = 1, 2, \dots, n$

and $E(X_1 + \dots + X_n)^2 = \sigma^2$, then

$$P \{ |X_1 + \dots + X_n| > x \} = 2 \exp \left[-\frac{x}{2\sigma} \sinh^{-1} \left\{ \frac{xc}{2\sigma^2} \right\} \right] \dots (2)$$

Proof:- For the proof we refer to Prohorov (1959).

5.3. Estimates of the density function.

Let x_1, x_2, \dots, x_n be n independent observations on a random variable X with a density function $f(x)$. Consider the estimate

$$\hat{f}(x, \lambda) = \lambda \sum_{r=1}^n \frac{K(\lambda(x - x_r))}{n} \dots (3)$$

for $\lambda > 0$. $\hat{f}(x, \lambda)$ is an unbiased estimate of $f(x, \lambda) = K_\lambda(f) = \lambda \int f(u) K(\lambda(x-u)) du$. From theorem 5.2. we know that $f(x, \lambda)$ converges to $f(x)$ in the mean. Thus it could be possible to choose a sequence λ_n with $\lambda_n \rightarrow \infty$ so that $\hat{f}(x, \lambda_n)$ converges to $f(x)$ in the mean with probability one. This we propose to do in the next section.

5.4. The strong mean consistency of $\hat{f}(x, \lambda_n)$.

Lemma 5.4. The functions $\hat{f}(x, \lambda)$ and $f(x, \lambda)$ are square integrable densities.

Proof: In fact the functions $\hat{f}(x, \lambda)$ and $f(x, \lambda)$ are bounded by λ and bounded densities are always square integrable.

Lemma 5.5. If $\lambda = n^\theta$, $0 < \theta < \frac{1}{2}$, then

$$\lim_{n \rightarrow \infty} \int |\hat{f}(x, \lambda) - f(x, \lambda)|^2 dx = 0 \quad (4)$$

with probability 1.

Proof:- Let

$$I = \int |\hat{f}(x, \lambda) - f(x, \lambda)|^2 dx$$

Then

$$I = \frac{1}{2} \sum_r \sum_s \int \lambda^2 K(\lambda(x-x_r)) K(\lambda(x-x_s)) dx \\ - \frac{2}{n} \sum_r \int \lambda K(\lambda(x-x_r)) f(x, \lambda) dx + \int |f(x, \lambda)|^2 dx \dots (5)$$

Let $K^*(x) = \int K(x-y)K(y)dy$

and $\lambda \int K^*(\lambda(x-y))f(y)dy = f^*(x, \lambda)$. . . (6)

Using (6) and (5) we find

$$I = \frac{1}{n} \sum_r \left[\sum_s \frac{\lambda K^*(\lambda(x_r - x_s))}{n} - f^*(x_r, \lambda) \right] \\ - \frac{1}{n} \sum_r [f^*(x_r, \lambda) - E(f^*(x_r, \lambda))] . . . (7)$$

Since $K^*(x)$ is a Fejer kernel by lemma 5.1, there exists a constant c_1 such that

$$K^*(x) < c_1, f^*(x, \lambda) \leq c_1 \lambda \quad (8)$$

Thus by (7) and (8)

$$I \leq \frac{2c_1 \lambda}{n} + \frac{1}{n} \sum_r \left| \sum_{\substack{1 \leq s \leq n, \\ s \neq r}} \frac{[\lambda K^*(\lambda(x_r - x_s)) - f^*(x_r, \lambda)]}{n} \right| \\ + \frac{1}{n} \left| \sum_r [f^*(x_r, \lambda) - E(f^*(x_r, \lambda))] \right| \\ = 2c_1 \frac{\lambda}{n} + Y_n + Z_n, \text{ say} \quad (9)$$

Since

$$|f^*(x_r, \lambda) - E(f^*(x_r, \lambda))| \leq 2c_1 \lambda$$

and

$$E |f^*(x_r, \lambda) - E(f^*(x_r, \lambda))|^2 \leq E(f^*(x_r, \lambda))^2 \leq \lambda^2$$

and $\sinh^{-1} x \leq \log(1+x)$

we have by lemma 5.3 that

$$P[Z_n > \delta] \leq 2 \exp \left[-\frac{n\delta}{4c_1\lambda} \log \left(1 + \frac{\delta}{4c_1\lambda} \right) \right] \quad \dots (10)$$

Further, we have

$$\begin{aligned} P[Y_n > \delta] &\leq nP \left[\left| \frac{\sum_{s=2}^n [\lambda K^*(\lambda(x_1 - x_s)) - f^*(x_1, \lambda)]}{n} \right| > \delta \right] \\ &= nE \left\{ P \left[\left| \frac{\sum_{s=2}^n [\lambda K^*(\lambda(x_1 - x_s)) - f^*(x_1, \lambda)]}{n} \right| > \delta \mid x_1 \right] \right\} \end{aligned}$$

When x_1 is given, $x_1 - x_2, x_1 - x_3, \dots$ are independent random variables. Noting that

$$E[\lambda K^*(\lambda(x_1 - x_s)) - f^*(x_1, \lambda) \mid x_1] = 0$$

and

$$\left| \lambda K^*(\lambda(x_1 - x_s)) - f^*(x_1, \lambda) \right| \leq 2c_1 \lambda$$

and applying lemma 5.3, we have

$$P[Y_n > \delta] \leq 2n \exp \left[-\frac{n\delta}{4c_1\lambda} \log \left(1 + \frac{\delta}{4c_1\lambda} \right) \right] \dots (11)$$

Thus if $\lambda = n^\theta$, $0 < \theta < \frac{1}{2}$, we have

$$\sum_n \left\{ P[Y_n > \delta] + P[Z_n > \delta] \right\} < \infty \quad (12)$$

(9), (12) and an application of the Borel-Cantelli lemma, establish (4) with probability one. Hence the theorem

We now prove the strong mean consistency of $\hat{f}(x, \lambda)$.

Theorem 5.6. If x_1, x_2, \dots is a sequence of independent observations from a population with a density function $f(x)$ and $\lambda_n = n^\theta$, $0 < \theta < \frac{1}{2}$, then

$$\lim_{n \rightarrow \infty} \int | \hat{f}(x, \lambda) - f(x) | dx = 0 \quad (13)$$

with probability 1.

Proof:- Let $S_n(x)$ be the empirical distribution function based on x_1, \dots, x_n . Let

$$G_\lambda(x) = \lambda \int_{-\infty}^x K(\lambda y) dy \quad (14)$$

From the Glivenko-Cantelli lemma it follows that $S_n(x)$ converges to $F(x)$ with probability 1. Further $G_{\lambda_n}(x)$ converges to the distribution which is degenerate at the origin. Hence the convolution

$S_n(x) * G_{\lambda_n}(x)$ of the distributions $S_n(x)$ and $G_{\lambda_n}(x)$ converges to the distribution $F(x)$ with probability 1. $\hat{f}(x, \lambda_n)$ and $f(x, \lambda_n)$ are

the density functions of the distributions $S_n(x) * G_{\lambda_n}(x)$ and $F(x) * G_{\lambda_n}(x)$ respectively. Hence for any given $\varepsilon > 0$, there exists a constant

$A > 0$ depending on the sequence x_1, x_2, \dots of observations and ε but

not on n such that for all $n \geq n_0$,

$$\int_{|x|>A} \hat{f}(x, \lambda_n) dx + \int_{|x|>A} f(x, \lambda_n) dx < \varepsilon \quad (15)$$

$$\begin{aligned} & \int |f(x, \lambda_n) - f(x)| dx \\ & \leq \int_{|x|<A} |\hat{f}(x, \lambda_n) - f(x, \lambda_n)| dx + \int_{|x|>A} |\hat{f}(x, \lambda_n)| dx \\ & \quad + \int_{|x|>A} f(x, \lambda_n) dx + \int |f(x, \lambda_n) - f(x)| dx \quad (16) \end{aligned}$$

Since $\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$, by Theorem 5.2 the last term in (16) tends to zero. For $n \geq n_0$, by (15) the two middle terms are less than ε . By lemma 5.5 the first term tends to zero as $n \rightarrow \infty$ since the integral is over a bounded interval. This completes the proof of the theorem.

Thus there arises a new problem of choosing the kernel $K(x)$ with some optimum properties, since all kernels yield estimates with the strong mean consistency property. Since $E\{f(x, \lambda)\} = f(x, \lambda)$, a measure of the bias of the density estimate would be $\int |f(x, \lambda) - f(x)| dx = e(K, \lambda)$. Imposing certain desirable restrictions on $e(K, \lambda)$, it can be shown that estimates satisfying these restrictions can be obtained by restricting K to the class of all kernels that are linear combinations of $(\frac{\sin x}{x})^{2r}$, $r = 1, 2, \dots$

Chapter VI

FIXED INTERVAL ANALYSIS AND FRACTILE ANALYSIS

6.1. The methods of fixed interval analysis and fractile analysis.

One of the important problems of statistics is the study of the relationship of one variable Y with another variable X and of the comparison of such relationships among different populations. These are usually made through the study of the regression function of Y on X , i.e. through the study of the function $E(Y | X=x)$. A common problem met with in practice is the comparison of the regression functions in two populations (Y, X) and (Y', X') . We can give illustrations for this problem from practically every applied field of Statistics. We will however content ourselves with one example which we will refer to, for purposes of illustration, in the sequel. Data on a pair of random variables (Y, X) and (Y', X') are available. Y and X refer to the consumption of milk (volume) and total expenditure per month, respectively, of an individual in a population P . Y' and X' refer to the same variates in a different population P' . The problem is to compare the patterns of relationship of the consumption of milk with the total expenditure in the two populations.

The problem just stated can be expressed in symbols as follows:

We wish to test whether or not

$$\lambda(x) = \lambda'(x) \text{ for all } x \dots \dots \dots (1)$$

where

$$\lambda(x) = E(Y \mid X = x) \dots \dots \dots (2)$$

$$\lambda'(x) = E(Y' \mid X' = x) \dots \dots \dots (3)$$

This is the sort of regression problem we shall be concerned with in this chapter. The classical method of approaching this problem consists in assuming that the regression functions $\lambda(x)$ and $\lambda'(x)$ are of a certain algebraic form, completely determined except for a finite number of 'parameters', say, a polynomial, a trigonometric series or the like. The problem of testing the equality of the regression functions then reduces to the problem of testing the equality of these parameters. The difficulty with these methods is that the assumption of algebraic model for the two regression functions is too great an over simplification of the actual situations. Further, it is very difficult to test whether any particular model fits in a given practical set up. We will describe in this section two methods, fixed interval analysis and fractile analysis, which do not involve the assumptions of the usual regression models. In both these methods we compare the two regressions functions intervalwise instead of pointwise, where the intervals are largely at our choice.

In fixed interval analysis we fix to start with g fixed intervals

$$(a_0, a_1], (a_1, a_2], \dots, (a_{g-1}, a_g] \quad \dots \quad (4)$$

by means of the $(g+1)$ constants

$$a_0, a_1, \dots, a_g \quad \dots \quad (5)$$

satisfying

$$-\infty = a_0 < a_1 < \dots < a_{g-1} < a_g = +\infty \quad \dots \quad (6)$$

We then define the regression $\bar{y}(a, b)$ of Y on X in an interval $(a, b]$ by the relation

$$\bar{y}(a, b) = E(Y \mid a < X \leq b) \quad \dots \quad (7)$$

Let

$$\bar{y}(a_{i-1}, a_i) = \bar{y}_i, \quad i = 1, \dots, g \quad \dots \quad (8)$$

The quantities \bar{y}'_i are defined in the same way for the random variable (Y', X') . In fixed interval analysis we test the hypothesis that

$$\bar{y} = \bar{y}' \quad \dots \quad (9)$$

where

$$\bar{y} = (\bar{y}_1, \dots, \bar{y}_g), \quad \bar{y}' = (\bar{y}'_1, \dots, \bar{y}'_g) \quad \dots \quad (10)$$

It is obvious that such a hypothesis becomes meaningful only when the intervals $(a_0, a_1], \dots, (a_{g-1}, a_g]$ of (4) are chosen carefully. For instance, in our illustrative example, let the populations P and P' correspond to two constituent states of the Indian Union at the same point in time. Several comparable social strata can be defined in the two populations by a suitable choice of the intervals of (4). Such strata might correspond to expenditures classified as 'lower class',

'lower middle class', etc. In this case the hypothesis (9) is meaningful and fixed interval analysis is appropriate to the situation.

It should be noted that the use of a large number of strata would mean that the hypothesis $\bar{y} = \bar{y}'$ is practically equivalent to the hypothesis $\lambda(x) = \lambda'(x)$.

The practical details of the method are as follows. We estimate the quantities \bar{y} and \bar{y}' suitably and test for their difference. Let

$$(y_1, x_1), \dots (y_n, x_n) \dots \quad (11)$$

and

$$(y'_1, x'_1), \dots (y'_n, x'_n) \dots \quad (12)$$

be two independent samples S and S' from the populations P and P' respectively. Let n_i be the number of observations of the samples S with x-components in the interval $(a_{i-1}, a_i]$, $i = 1, \dots, g$.

Let

$$\bar{v}_i = \frac{\sum_{a_{i-1} < x_r \leq a_i} y_r}{n_i}, \quad i = 1, \dots, g \quad (13)$$

where the summation is made over all y_r 's for which x_r 's lie in $(a_{i-1}, a_i]$. We calculate \bar{v}'_i , $i = 1, \dots, g$ in a similar way from the sample S'. The vectors \bar{v} and \bar{v}' are estimates of \bar{y} and \bar{y}' . We define several measures of divergence between the sample estimates for the two regressions.

$$\bar{D} = \sum_{i=1}^g |\bar{v}_i - \bar{v}'_i| \dots \quad (14)$$

$$\bar{\Delta} = \sum_{i=1}^g [\bar{v}_i - \bar{v}'_i]^2 \quad \dots \quad (15)$$

$$\bar{F} = (\bar{v} - \bar{v}') B (\bar{v} - \bar{v}')' \quad \dots \quad (16)$$

where B is some positive definite matrix. Large values of these statistics will form the corresponding rejection region for testing the hypothesis in (9). It is an important problem to determine the distributions of these statistics, at least in large samples, and set up significance points for these statistics.

We now proceed to remark that there are many problems where fixed interval analysis becomes conceptually meaningless or partially so. As an example let us suppose that in our practical example, the populations P and P' correspond to two different states having different currencies. After setting up interval limits to the total social groups in one population, we may find it very difficult expenditure reflecting different to demarcate comparable limits in the currency of the second population. The official exchange rate cannot be used here since it does not reflect the actual purchasing power of the two currencies. The real exchange rate that does this is not easily available. We have thus presented a typical case of the general situation where X and X' are not comparable and where fixed interval analysis is not useful.

In the above example we can safely assume that the total expenditure is a monotonic function of the socio-economic level of an individual. Thus, though the total expenditures in the two countries are not

directly comparable they are monotonically related. Groupings based on the ranks of X and X' will be comparable in a meaningful way. Professor P. C. Mahalanobis made use of this fact in proposing a new method called fractile analysis for such situations.

We describe this method after defining some quantities. Let $\theta_1, \theta_2, \dots, \theta_{g-1}$ be the $\frac{1}{g}$ th, $\frac{2}{g}$ th, ... $\frac{g-1}{g}$ th quantiles (or fractiles) of the distribution of X . Let $\theta_0 = -\infty$, $\theta_g = +\infty$. Let $\theta'_0, \theta'_1, \dots, \theta'_{g-1}, \theta'_g$ be the corresponding quantiles for the distribution of X' . Let

$$\nu_i = \bar{y}(\theta_{i-1}, \theta_i) \quad i = 1, \dots, g \quad \dots \quad (17)$$

$$\nu'_i = \bar{y}'(\theta'_{i-1}, \theta'_i) \quad i = 1, \dots, g \quad \dots \quad (18)$$

We see that the intervals $(\theta_0, \theta_1), (\theta_1, \theta_2) \dots (\theta_{g-1}, \theta_g)$ of X represent the lowest $100/g$ percent section, the second lowest $100/g$ percent section the highest $100/g$ percent section, respectively, of population P . A similar interpretation can be given for the intervals $(\theta'_0, \theta'_1), (\theta'_1, \theta'_2), \dots (\theta'_{g-1}, \theta'_g)$ of X' in the second population P' . Thus these intervals are comparable in a very important sense, although the X values are different. The method of fractile analysis consists in testing the hypothesis

$$\nu = \nu' \quad \dots \quad (19)$$

where

$$\underline{y} = (y_1, \dots, y_g), \quad \underline{y}' = (y'_1, \dots, y'_g) \dots \quad (20)$$

The practical method adopted is as follows. Let S and S' be the samples from the populations P and P' respectively, as in (11), (12). Rearrange the observations in the sample S so that the x's are in the increasing order of magnitude thus

$$(y_{(1)}, x_{(1)}), (y_{(2)}, x_{(2)}), \dots (y_{(n)}, x_{(n)}) \dots \quad (21)$$

with

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} \dots \dots \dots \quad (22)$$

The notation adopted here being symmetric in x and y may be a bit confusing. We remark that we will never have to order the y's so that $(x_{(r)}, y_{(r)})$ will always represent the observation whose x - component has the rth rank. Let $n = m.g$ where m and g are integers. We define the quantities

$$v_i = \sum_{(i-1)m < r \leq im} y_{(r)}/m, \quad i = 1, \dots, g \dots \quad (23)$$

Quantities $v'_i, i = 1, \dots, g$ are defined in a similar way from S' vectors \underline{y} and \underline{y}' are estimates of \underline{y} and \underline{y}' . Suitable measures of divergence are then defined. Professor P.C.Mahalanobis defined one such measure called separation, as follows. Plot the ordinates v_1, v_2, \dots, v_g corresponding to the equidistant points 1, 2, ... g. Join the successive points by straight lines; the curve C thus obtained is called the fractile graph of S. The fractile graph C' of

S' is obtained in a similar way and drawn on the same paper and to the same scale. The area A between these two graphs, and the ordinates at 1 and g called the separation, is a measure of divergence between the two sample regressions. The algebraic expression for A is

$$A = \sum_{i=1}^{g-1} \left[\frac{1}{2} |v_i - v'_i| + \frac{1}{2} |v_{i+1} - v'_{i+1}| - \delta(v_i - v'_i, v_{i+1} - v'_{i+1}) \frac{|(v_i - v'_i)(v_{i+1} - v'_{i+1})|}{|v_i - v'_i| + |v_{i+1} - v'_{i+1}|} \right] \dots \quad (24)$$

where $\delta(a, b) = \begin{cases} 0 & \text{if } ab > 0 \\ 1 & \text{if } ab < 0 \end{cases} \dots \dots \quad (25)$

Some other measures of divergence which have been considered are

$$D = \sum_{i=1}^g |v_i - v'_i|, \quad \dots \dots \quad (26)$$

$$\Delta = \sum_{i=1}^g (v_i - v'_i)^2, \quad \dots \dots \quad (27)$$

$$\Gamma = (v - v') B (v - v')', \quad \dots \dots \quad (28)$$

where B is some positive definite matrix. Large values of these statistics form the corresponding critical regions for testing the hypothesis in (19). We repeat the remark made earlier, that at least the large sample distributions of these statistics should be determined before using these measures for testing purposes. Till now only descriptive methods are available for the methods of fixed interval

analysis and fractile analysis; for examples see Mahalanobis (1960), Das (1960), Das and Sharma (1960).

We find the requisite limiting distributions in section 6.3 and these can now be used in practice.

Let us add a few points bringing out the special character of the method of fractile analysis in comparison with the method of fixed interval analysis. For applying the method of fixed interval analysis the random variables X and X' must be directly comparable and some intervals resembling strata should be formed. We then compare the regression in these intervals. The method of fractile analysis is applicable to the more difficult situation where X and X' are not directly comparable but are monotonically related to a common character that X and X' are supposed to measure. This, then is the general set up where fractile analysis can be used. Such situations have been encountered in Econometrics, Psychometry, Demography etc., and fractile analysis has been applied, though, as yet only in a descriptive way. See Mahalanobis (1960), Das (1960), Das and Sharma (1960), Som (1960). Fractile analysis is now being utilised in the National Sample Survey of India on a large scale.

Several modifications of the method of fractile analysis can be made. After ranking the individuals we can take groups with varying proportions instead of equal proportions. Thus a group already formed may be split up into several sub-groups for a more detailed analysis

of the regression in that group. This has been employed in Mahalanobis (1960). For calculating the sample estimate of μ , the median, mode or some other suitable characteristic can be taken instead of the mean. These may be easier to compute in practice. This has been stated in Mahalanobis (1958a). We can thus go on multiplying the number of modifications we could make.

In practice, it happens that frequently we take several and independent sub-samples S_1, \dots, S_k from a population P instead of just one sample. These are called interpenetrating sub-samples and their usefulness in a large number of situations has been recognised. See Mahalanobis (1946), Yates (1949), Ghosh (1949), Lahiri (1954), (1957). If interpenetrating sub-samples are taken from both the populations P and P' we can get measures, D, D_1 and D_2 , of divergence, from the samples measuring the divergence between the combined samples of P and P' , within the sub-samples of P and within the sub-samples of P' , respectively. Thus D would be the between divergence whereas D_1 and D_2 would be the within divergences. These terminologies are analogous to those in analysis of variance problems and are self-explanatory. Suggestive as they are they permit us to develop some tests of at least a descriptive nature. See Mahalanobis (1958a), (1960).

Finally we add that the quantities calculated in the methods of fixed interval analysis and fractile analysis can be suitably modified

to estimate relative concentration curves and relative concentration ratios. The utility of these is well known. See for example, Davis (1941), Roy, Chakravarti and Laha (1959), Iyengar (1960).

We now give a brief description of the method of obtaining the concentration curve etc. The following symbols will be used:

$$\bar{v}_i = n_1 \bar{v}_1 + n_2 \bar{v}_2 + \dots + n_i \bar{v}_i; \quad i = 1, \dots, g \quad (29)$$

$$\bar{v}_0 = 0, \bar{v}_g = \bar{v} \quad \dots \quad \dots \quad \dots \quad (30)$$

$$\bar{q}_i = \bar{v}_i / \bar{v}; \quad i = 0, 1, \dots, g \quad \dots \quad \dots \quad (31)$$

$$\bar{p}_i = (n_1 + n_2 + \dots + n_i) / n; \quad i = 1, \dots, g \quad (32)$$

$$\bar{p}_0 = 0 \quad \dots \quad \dots \quad \dots \quad (33)$$

$$V_i = (v_1 + \dots + v_i); \quad i = 1, \dots, g \quad \dots \quad (34)$$

$$V_0 = 0, V_g = V \quad \dots \quad \dots \quad \dots \quad (35)$$

$$q_i = V_i / V; \quad i = 0, 1, \dots, g \quad \dots \quad \dots \quad (36)$$

$$p_i = i / g; \quad i = 0, \dots, g \quad \dots \quad \dots \quad (37)$$

When the method of fixed interval analysis is used an estimate of the relative concentration curve of Y on X is obtained by plotting the points

$$(\bar{p}_0, \bar{q}_0), (\bar{p}_1, \bar{q}_1), \dots, (\bar{p}_g, \bar{q}_g) \quad \dots \quad (38)$$

and joining successive points by straight lines. The relative concentration ratio is estimated by

$$\bar{c} = \frac{1}{\bar{v}} \left[\sum n_i \bar{v}_i (\bar{p}_{i-1} + \bar{p}_i) \right] - 1 \quad \dots \quad (39)$$

When the method of fractile analysis is used an estimate of the relative concentration curve of Y on X is obtained by plotting the points

$$(p_0, q_0), (p_1, q_1), \dots, (p_g, q_g) \quad \dots \quad (40)$$

and joining successive points by straight lines. The relative concentration ratio is estimated by

$$c = \frac{2}{g \bar{v}} \sum i v_i - \frac{g+1}{g} \quad \dots \quad (41)$$

6.2. Notation, definitions etc.

In this section we develop the notations etc., to be used in sections 6.3, 6.4 and 6.5. Since we are developing the theories of fixed interval analysis and fractile analysis side by side and since they are similar to one another in many respects, the notations employed will also be similar. Whenever possible we will distinguish the quantities involved in fixed interval analysis by a bar $\bar{\quad}$.

Let (Y, X) be a random variable on the Euclidean plane with distribution function $G(y, x)$. The distribution function of X will be denoted by $F(x)$.

$$\text{Let } (y_1, x_1), (y_2, x_2), \dots, (y_n, x_n) \quad \dots \quad (42)$$

be a sample S of n independent observations on (Y, X) . Let us arrange the observations according to the increasing order of magnitude

of the x'_s , thus:

$$(y_{(1)}, x_{(1)}), (y_{(2)}, x_{(2)}), \dots, (y_{(n)}, x_{(n)}) \dots \quad (43)$$

The fixed interval method of analysis involves stratifying the population into g strata. Let these strata be formed by the predetermined constants a_0, a_1, \dots, a_g satisfying

$$-\infty = a_0 < a_1 < \dots < a_{g-1} < a_g = +\infty \quad \dots \quad (44)$$

These constants introduce g strata in the domain of (Y, X) as follows: The r th stratum ($r = 1, \dots, g$) contains all (y, x) with $a_{r-1} < x \leq a_r$. We define

$$\pi_r = F(a_r) - F(a_{r-1}); \quad i = 1, \dots, g \quad \dots \quad (45)$$

throughout the discussion on fixed interval analysis we shall assume that

$$\pi_r > 0, \quad r = 1, \dots, g \quad \dots \quad (46)$$

Let us define the means, variances and covariances of Y and X (these are assumed to exist throughout this discussion) in these g strata

$$\begin{aligned} \bar{\mu}_r &= E(X \mid a_{r-1} < X \leq a_r); & \bar{\nu}_r &= E(Y \mid a_{r-1} < X \leq a_r) \\ \bar{\sigma}_r^2 &= V(X \mid a_{r-1} < X \leq a_r); & \bar{\tau}_r^2 &= V(Y \mid a_{r-1} < X \leq a_r) \\ \bar{\varrho}_r \bar{\sigma}_r \bar{\tau}_r &= \text{cov}(Y, X \mid a_{r-1} < X \leq a_r); & & r = 1, \dots, g. \end{aligned} \quad (47)$$

Let n_1, n_2, \dots, n_g be the number of observations in the sample S in the 1st, 2nd, ..., g^{th} stratum, respectively, introduced by the constants in (44). We shall denote the proportion of observations in these strata by p_1, p_2, \dots, p_g , i.e.

$$p_i = n_i / n; \quad i = 1, \dots, g \quad \dots \quad (48)$$

The means, variances and covariances the sample S in these strata are dedined as follows:

$$\begin{aligned} \bar{u}_i &= \sum_{a_{i-1} < x_r \leq a_i} x_r / n_i; & \bar{v}_i &= \sum_{a_{i-1} < x_r \leq a_i} y_r / n_i \\ \bar{s}_i^2 &= \sum_{a_{i-1} < x_r \leq a_i} (x_r - u_i)^2 / n_i; & \bar{t}_i^2 &= \sum_{a_{i-1} < x_r \leq a_i} (y_r - v_i)^2 / n_i \\ \bar{r}_i \bar{s}_i \bar{t}_i &= \sum_{a_{i-1} < x_r \leq a_i} (x_r - u_i)(y_r - v_i) / n_i; \quad i = 1, \dots, g \end{aligned} \quad (49)$$

In the theorems of the next section will be interested in the limiting distribution of the following statistics.

$$\bar{\xi}_i(n) = \sqrt{n}(\bar{u}_i - \mu_i); \quad \bar{\eta}_i(n) = \sqrt{n}(\bar{v}_i - \nu_i); \quad \dots \quad (50)$$

$$\bar{\zeta}_i(n) = \sqrt{n}(p_i - \pi_i); \quad i = 1, \dots, g$$

$$\begin{aligned} \bar{\Xi}(n) &= (\bar{\xi}_1(n), \dots, \bar{\xi}_g(n)); & \bar{\Psi}(n) &= (\bar{\eta}_1(n), \dots, \bar{\eta}_g(n)); \\ \bar{\Sigma}(n) &= (\bar{\zeta}_1(n), \dots, \bar{\zeta}_g(n)). & \dots & \dots \dots \dots \quad (51) \end{aligned}$$

In problems connected with fractile analysis we shall assume the following:

The distribution function $G(y, x)$ admits of a density function $g(y, x)$ which is continuous and which does not vanish. | (52)

The density function of $F(x)$ will be denoted by $f(x)$.

In the fractile method of analysis the population is to be divided into a previously assigned number, g , of strata thus. Let θ_i be defined by

$$F(\theta_i) = 1/g \quad i = 1, \dots, g-1 \quad \dots \quad (53)$$

$$\theta_0 = -\infty, \quad \theta_g = +\infty \quad \dots \quad (54)$$

Then the r^{th} stratum is defined as the region of all points (y, x) with

$$\theta_{r-1} < x \leq \theta_r, \quad r = 1, \dots, g.$$

The means, variances and covariances (these are assumed to exist throughout this discussion) in these strata are defined by

$$\begin{aligned} \mu_r &= E(X \mid \theta_{r-1} < X \leq \theta_r); \quad \nu_r = E(Y \mid \theta_{r-1} < X \leq \theta_r); \\ \sigma_r^2 &= V(X \mid \theta_{r-1} < X \leq \theta_r); \quad \tau_r^2 = V(Y \mid \theta_{r-1} < X \leq \theta_r); \\ \rho_r \sigma_r \tau_r &= \text{cov}(Y, X \mid \theta_{r-1} < X \leq \theta_r); \quad r = 1, \dots, g. \end{aligned} \quad (55)$$

We also require the regression function

$$E(Y \mid X = x) = \lambda(x) \quad \dots \quad (56)$$

$$\lambda(\theta_i) = \lambda_i, \quad i = 1, \dots, g-1 \quad \dots \quad (57)$$

$$\lambda_0 = 0, \quad \lambda_g = 0 \quad \dots \quad (58)$$

The corresponding quantities are defined from the sample S.

It is assumed that $n = m \cdot g$ where m is an integer.

$$\begin{aligned}
 u_i &= \frac{1}{m} \sum_{(i-1)m < r \leq im} x_{(r)}; \\
 v_i &= \frac{1}{m} \sum_{(i-1)m < r \leq im} y_{(r)}; \\
 s_i^2 &= \frac{1}{m} \sum_{(i-1)m < r \leq im} (x_{(r)} - u_i)^2/m; \quad t_i^2 = \frac{1}{m} \sum_{(i-1)m < r \leq im} (y_{(r)} - v_i)^2/m; \\
 r_i s_i t_i &= \frac{1}{m} \sum_{(i-1)m < r \leq im} (x_{(r)} - u_i)(y_{(r)} - v_i)/m; \quad i = 1, \dots, g.
 \end{aligned} \tag{59}$$

In the next section we will be interested in the limiting distribution of the following statistics.

$$\begin{aligned}
 \xi_i(n) &= \sqrt{m}(u_i - \mu_i); \quad i = 1, \dots, g \\
 \eta_i(n) &= \sqrt{m}(v_i - \nu_i); \quad i = 1, \dots, g
 \end{aligned} \tag{60}$$

$$\zeta_i(n) = \sqrt{n}(x_{(im+1)} - \theta_i); \quad i = 1, \dots, g-1 \tag{61}$$

$$\begin{aligned}
 \xi(n) &= (\xi_1(n), \dots, \xi_g(n)); \quad \eta(n) = (\eta_1(n), \dots, \eta_g(n)); \\
 \zeta(n) &= (\zeta_1(n), \dots, \zeta_{g-1}(n)). \quad \dots \tag{62}
 \end{aligned}$$

In the following sections we shall have occasion to consider another random variable (Y', X') constituting a population P' . The constants for this population will be obtained by adding a ' mark to those of P . In the same way the statistics from a sample with some suffix will be obtained by adding that suffix to the statistics of the sample S.

6.3. Limit distributions

In this section we state without proof some results concerning the limiting distributions of the statistics entering (51) and (62). The proofs of all these theorems can be found in Sethuraman (1961a). The results concerning the statistics entering fractile graphical analysis can be found in Sethuraman (1961).

Theorem 6.1. Let condition (45) hold. As $n \rightarrow \infty$, the sequence of random variables $(\bar{X}(n), \bar{Y}(n), \bar{Z}(n))$ converges to a random variable $(\bar{X}, \bar{Y}, \bar{Z})$ with a multivariate normal distribution with mean vector 0 and variance covariance matrix

$$\begin{pmatrix} \bar{\Sigma}\bar{\Lambda}^{-1} & \bar{E}\bar{\Lambda}^{-1} & 0 \\ \bar{E}\bar{\Lambda}^{-1} & \bar{T}\bar{\Lambda}^{-1} & 0 \\ 0 & 0 & \bar{K} \end{pmatrix} \quad \dots \quad (63)$$

where $\bar{\Sigma} = \text{diag.} * (\bar{\sigma}_1^2, \dots, \bar{\sigma}_g^2) \quad \dots \quad (64)$

$$\bar{T} = \text{diag.} (\bar{\tau}_1^2, \dots, \bar{\tau}_g^2) \quad \dots \quad (65)$$

$$\bar{E} = \text{dia.} (\bar{\sigma}_1 \bar{\sigma}_1 \bar{\tau}_1, \dots, \bar{\sigma}_g \bar{\sigma}_g \bar{\tau}_g), \quad \dots \quad (66)$$

$$\bar{\Lambda} = \text{diag.} (\pi_1, \dots, \pi_g) \quad \dots \quad (67)$$

$$\bar{K} = \begin{pmatrix} \pi_1(1 - \pi_1), \pi_1(1 - \pi_2) \dots \pi_1(1 - \pi_g) \\ \pi_1(1 - \pi_2), \pi_2(1 - \pi_2) \dots \pi_2(1 - \pi_g) \\ \dots \\ \pi_1(1 - \pi_g), \pi_2(1 - \pi_g) \dots \pi_g(1 - \pi_g) \end{pmatrix} \quad \dots \quad (68)$$

* $\text{diag.}(b_1, \dots, b_k)$ stands for the diagonal matrix $\begin{pmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_g \end{pmatrix}$

Corollary 6.2 The distribution of $\bar{\eta}$ is multivariate normal with mean vector 0 and variance covariance matrix where

$$\bar{\Lambda} = \bar{\Gamma} \Pi^{-1} \dots \quad (69)$$

Theorem 6.3. Let condition (52) hold. As $m \rightarrow \infty$ (i.e. $n \rightarrow \infty$) the sequence of random variables $(\xi(n), \eta(n), \zeta(n))$ converges weakly to a random variable (ξ, η, ζ) with a multivariate normal distribution.

Let

$$\begin{aligned} M_i &= i(\theta_i - \mu_i) - (i-1)(\theta_{i-1} - \mu_i); \quad i = 2, \dots, g-1 \\ M_1 &= (\theta_1 - \mu_1), \quad M_g = -(g-1)(\theta_{g-1} - \mu_g) \end{aligned} \quad (70)$$

$$\begin{aligned} M_i^0 &= (g-1)(\theta_i - \mu_i) - (g-i+1)(\theta_{i-1} - \mu_i); \quad i = 2, \dots, g-1 \\ M_1^0 &= (g-1)(\theta_1 - \mu_1), \quad M_g^0 = -(\theta_{g-1} - \mu_g) \end{aligned} \quad (71)$$

$$\begin{aligned} N_i &= i(\lambda_i - \nu_i) - (i-1)(\lambda_{i-1} - \nu_i); \quad i = 2, \dots, g-1 \\ N_1 &= (\lambda_1 - \nu_1), \quad N_g = -(g-1)(\lambda_{g-1} - \nu_g) \end{aligned} \quad (72)$$

$$\begin{aligned} N_i^0 &= (g-i)(\lambda_i - \nu_i) - (g-i+1)(\lambda_{i-1} - \nu_i); \quad i = 2, \dots, g-1 \\ N_1^0 &= (g-1)(\lambda_1 - \nu_1), \quad N_g^0 = -(\lambda_{g-1} - \nu_g) \end{aligned} \quad (73)$$

$$\begin{aligned}
 \text{Let } \Sigma_{ij}^* &= \frac{1}{g} M_i M_j^0 & j > i \\
 &= \frac{1}{g} M_j M_i^0 & i > j \\
 &= \sigma_i^2 + \frac{1}{g} M_i M_i^0 + (\theta_i - \mu_i)(\theta_{i-1} - \mu_{i-1}); & i=j; i \neq 1, g & \quad (74) \\
 &= \sigma_1^2 + \frac{1}{g} M_1 M_1^0 & i = j = 1 \\
 &= \sigma_g^2 + \frac{1}{g} M_g M_g^0 & i = j = g
 \end{aligned}$$

$$\begin{aligned}
 \Lambda_{ij} &= \frac{1}{g} N_i N_j^0 & j > i \\
 &= \frac{1}{g} N_j N_i^0 & i > j \\
 &= \tau_i^2 + \frac{1}{g} N_i N_i^0 + (\lambda_i - \nu_i)(\lambda_{i-1} - \nu_{i-1}); & i = j; i \neq 1, g & \quad (75) \\
 &= \tau_1^2 + \frac{1}{g} N_1 N_1^0 & i = j = 1 \\
 &= \tau_g^2 + \frac{1}{g} N_g N_g^0 & i = j = g
 \end{aligned}$$

$$\begin{aligned}
 E_{ij}^* &= \frac{1}{g} M_i N_j^0 & j > i \\
 &= \frac{1}{g} N_i M_j^0 & j < i \\
 &= \sigma_i \sigma_i \tau_i + \frac{1}{g} M_i N_i^0 + (\theta_i - \mu_i)(\lambda_{i-1} - \nu_{i-1}); & i=j; i \neq 1, g & \quad (76) \\
 &= \sigma_1 \sigma_1 \tau_1 + \frac{1}{g} N_1 N_1^0 & i = j = 1 \\
 &= \sigma_g \sigma_g \tau_g + \frac{1}{g} M_g N_g^0 & i = j = g
 \end{aligned}$$

Corollary 6.4. (ξ, η) has a multivariate normal distribution with mean vector 0 and variance covariance matrix

$$\left(\begin{array}{c|c} \Sigma^* & E^* \\ \hline E^* & \Lambda \end{array} \right) \begin{array}{l} g \\ g \end{array} \quad (77)$$

Corollary 6.5. \mathcal{Y} has a multivariate normal distribution with mean vector 0 and variance covariance matrix Λ .

Theorem 6.6. Let $S_{(1)}, \dots, S_{(k)}$ be k independent samples, each of size n , on (Y, X) . Let S be the pooled sample. Let $(\xi_{(i)}(n), \mathcal{Y}_{(i)}(n)), \dots, (\xi_{(k)}(n), \mathcal{Y}_{(k)}(n))$ and $(\xi(n), \mathcal{Y}(n))$ be statistics computed from these samples. Let

$$(\xi^{\circ}(n), \mathcal{Y}^{\circ}(n)) = \frac{1}{\sqrt{k}} \sum_{i=1}^k (\xi_{(i)}(n), \mathcal{Y}_{(i)}(n)) \quad \dots \quad (78)$$

Let condition(52) hold.

Then $(\xi(n), \mathcal{Y}(n)) - (\xi^{\circ}(n), \mathcal{Y}^{\circ}(n)) \rightarrow 0$ in probability.

The proofs of all these theorems depend heavily on theorems 4.5, 4.6.

6.4. Theoretical applications.

In this section we show that the limiting distributions of section 6.3 in effect reduce a large sample S from a population P to just one observation (\underline{y} or \underline{y}) from a multivariate normal distribution. We then derive the limiting distribution of the measures of divergence $\bar{D}, \bar{\Delta}, \bar{\Gamma}, \Lambda, D, \Delta$ and $\bar{\Gamma}$ (see section 2) used in fixed interval analysis and in fractile analysis. We indicate the tests that can be used when several (interpenetrating) samples are available from each population. The asymptotic distributions of the concentration

ratios (39) and (41) are shown to be normal.

Let samples S and S' be drawn from the populations P and P' respectively, and the method of fixed interval analysis be employed.

Corollary 6.2 states that \bar{v} is asymptotically normal with mean and variance covariance matrix $\bar{\Lambda}/n$. We write this in symbols as

$$\bar{v} \sim MN(\bar{y}, \bar{\Lambda}/n) \quad \dots \quad (79)$$

Similarly

$$\bar{v}' \sim MN(\bar{y}', \bar{\Lambda}'/n') \quad \dots \quad (80)$$

(79) and (80) show that the samples S and S' are now reduced to the vectors \bar{v} and \bar{v}' respectively, with asymptotic multivariate normal distributions.

Let $n, n' \rightarrow \infty$ in such a way that $n/n' \rightarrow c$, $0 < c < \infty$.

Then

$$\bar{v} - \bar{v}' \sim MN(\bar{y} - \bar{y}', (\bar{\Lambda} + c\bar{\Lambda}')/n) \quad (81)$$

From (81) we can easily obtain the asymptotic distributions of \bar{D} , $\bar{\Delta}$ and $\bar{\Gamma}$ under the null hypothesis (9). $n\bar{\Delta}$ has a limiting distribution which is the distribution of $\sum \beta_r \chi_r^2$ where $\chi_1^2, \dots, \chi_g^2$ are independent χ^2 with 1 d.f. and

$$\beta_1 = \left(\frac{\bar{\tau}_1^2}{\pi_1} + c \frac{\bar{\tau}'_1{}^2}{\pi'_1} \right), \dots, \beta_g = \left(\frac{\bar{\tau}_g^2}{\pi_g} + c \frac{\bar{\tau}'_g{}^2}{\pi'_g} \right).$$

For a particular choice of B , $n\bar{\Gamma}$ is equal to $n \sum_1^g [(\bar{v}_i - \bar{v}'_i)^2 / (\frac{\bar{\tau}_i^2}{\pi_i} + c \frac{\bar{\tau}'_i{}^2}{\pi'_i})]$ and has a limiting distribution that is a χ^2 with g degrees of freedom. The limiting distribution of $\sqrt{n}\bar{D}$ exists but does not have

a simple algebraic form.

Let $S_{(1)}, \dots, S_{(k)}$ be k independent and equally valid (interpenetrating) sub-samples of the same size n from the population P . Let $S'_{(1)}, \dots, S'_{(k')}$ be k' interpenetrating subsamples of size n' each from P . If n and n' are large these two sets of samples can be reduced to two samples $\bar{v}_{(1)}, \dots, \bar{v}_{(k)}$ and $\bar{v}'_{(1)}, \dots, \bar{v}'_{(k')}$ from multivariate normal distributions with parameters* $(\bar{\mu}, \bar{\Lambda}/n)$ and $(\bar{\mu}', \bar{\Lambda}'/n')$ respectively.

The problem of fixed interval analysis is to test the hypothesis $\bar{\mu} = \bar{\mu}'$. Since $\bar{\Lambda}$ and $\bar{\Lambda}'$ are diagonal the problem can be viewed as the problem of simultaneous independent Fisher-Behren tests.

To pose the problem as one in classical multivariate analysis we should strengthen hypothesis (9) to the hypothesis

$$\bar{\mu} = \bar{\mu}', \quad \bar{\Lambda} = \bar{\Lambda}' \quad \dots \quad (82)$$

Multivariate analysis now yields us two solutions to this problem.

We can use the likelihood ratio criterion

$$\frac{G}{r-1} \left\{ [kcs_r^2 + k's_r'^2 + kk'c(\bar{v}_r - \bar{v}'_r)^2] / [kcs_r^2 + k's_r'^2] \right\} \quad (83)$$

where $\bar{v}_r = \sum_{j=1}^k \bar{v}_{r(j)}/k$

$$s_r^2 = \sum_{j=1}^k (\bar{v}_{r(j)} - \bar{v}_r)^2/k \quad \dots \quad \dots \quad (84)$$

* the parameters are the mean vector and the variance covariance matrix, respectively.

Large values of this criterion will form the region of rejection. The distribution of this criterion has been evaluated by Box (1949). Another method is to use the criterion

$$\sup_{1 \leq r \leq g} \frac{|\bar{v}_r - \bar{v}'_r| [ckk'(k + k' - 2)]^{\frac{1}{2}}}{[(cks_r^2 + k's'_r^2)(k' + kc)]^{\frac{1}{2}}} \dots \quad (35)$$

The distribution of this criterion is the distribution of the absolute maximum of g independent t -distributions each with $k + k' - 2$ degrees of freedom. Significance points of this distribution can easily be obtained from the tables of the t -distribution.

Let samples S and S' be drawn from the populations P and P' respectively, and the method of fractile analysis be used.

Corollary 6.5 states that

$$\underline{y} \sim MN(\underline{\gamma}, \Lambda/m) \quad (86)$$

$$\underline{y}' \sim MN(\underline{\gamma}', \Lambda'/m') \quad (87)$$

(86) and (87) show that the samples S and S' are now reduced to the vectors \underline{y} and \underline{y}' respectively, with asymptotic multivariate normal distributions.

Let $n, n' \rightarrow \infty$ in such a way that $n/n' \rightarrow c, 0 < c < \infty$.

Then

$$\underline{y} - \underline{y}' \sim MN(\underline{\gamma} - \underline{\gamma}', (\Lambda + c \Lambda')/m) \quad (88)$$

From (88) we can obtain the asymptotic distribution of Λ , D , Δ and Γ under the null hypothesis (19). $m \Delta$ has a limiting distribution which is the distribution of $\sum \beta_r \chi_r^2$ where $\chi_1^2, \dots, \chi_g^2$ are independent χ^2 with 1 d.f. and β_1, \dots, β_g are the latent roots of $(\Lambda + c \Lambda')$. Γ has a similar limiting distribution. In particular

$$\Gamma = m(\underline{v} - \underline{v}')(\Lambda + c \Lambda')^{-1}(\underline{v} - \underline{v}')' \quad (89)$$

has a limiting distribution that is a χ^2 with g d.f. The limiting distribution of $\sqrt{m} \Lambda$ and $\sqrt{m} D$ exist but do not have simple algebraic expressions. Crude approximations to the limiting distribution of $\sqrt{m} \Lambda$ can be made by the use of the following easily proved inequality

$$\frac{\sqrt{\Delta}}{6} \leq \Lambda \leq \sqrt{g} \cdot \sqrt{\Delta} \quad (90)$$

Let $S_{(1)}, \dots, S_{(k)}$ ($S'_{(1)}, \dots, S'_{(k')}$) be $k(k')$ interpenetrating subsamples from $P(P')$. If $n(n')$ is large then $v_{(1)}, \dots, v_{(k)}$ ($v'_{(1)}, \dots, v'_{(k')}$) is a sample from a multivariate normal distribution with parameters $(\underline{v}, \Lambda/m)$ ($(\underline{v}', \Lambda'/m')$).

The problem of fractile analysis is to test the hypothesis $\underline{v} = \underline{v}'$. To tackle this problem as one in multivariate analysis we use a restricted hypothesis

$$\underline{v} = \underline{v}' \quad , \quad \Lambda = \Lambda' \quad . \quad . \quad . \quad (91)$$

We now have the familiar problem of testing the equality of the mean when the variance covariance matrix of one multivariate normal population is a constant multiple of that of the other. The Mahalanobis D^2 -statistic will be used. Let

$$v_i^0 = \sum_{\alpha=1}^k v_i(\alpha) / k$$

$$s_{ij} = \sum_{\alpha=1}^k (v_i(\alpha) - v_i^0)(v_j(\alpha) - v_j^0) / k \quad \dots \quad (92)$$

$$s = (s_{ij})$$

$$\tilde{s} = [kcs + k's'] / c \quad \dots \quad \dots \quad (93)$$

Then

$$\frac{k + k' - g - 1}{g} \cdot \frac{kk'}{ck + k'} (v^0 - v^0')(\tilde{s})^{-1} (v^0 - v^0) ' \quad (94)$$

is our test criterion. Its distribution is an F distribution with g and $k + k' - g - 1$ d.f.

Let $\underline{y} (\underline{y}')$ be derived from $S(S')$, the sample obtained by pooling $S_{(1)}, \dots, S_{(k)} (S'_{(1)}, \dots, S'_{(k)})$. As an application of Theorem 6.4 we can substitute \underline{y} and \underline{y}' for \underline{v}^0 and \underline{v}^0' in (94) without changing the limiting distribution.

In the preceding discussion we assumed that the interpenetrating samples from each population are of the same size. We can remove this restriction by simple modifications. The tests mentioned above require several samples from each populations, and make use of

them only through the \bar{y} 's or \bar{y} 's. This involves considerable labour and waste of information. When only one sample is available from each population we cannot make use of the measures of divergence Δ , D , etc., for testing since their limiting distributions involve unknown constants, In 6.5 we suggest methods of over-coming this difficulty.

In section 6.1 we described how concentration curves can be drawn with the help of the data of fixed interval analysis and fractile analysis. We now give the limiting distributions of the concentration ratio, as a direct application of the theorems of section 6.3 and a theorem of Cramer (1946) pp. 366.

$$\text{Let } \bar{y} = \frac{\sum \pi_i \bar{y}_i (\bar{\pi}_{i-1} + \bar{\pi}_i)}{(\sum \pi_i \bar{y}_i)} - 1 \quad \dots \quad (95)$$

$$\bar{\pi}_i = \pi_1 + \dots + \pi_i \quad \dots \quad \dots \quad (96)$$

$$\frac{(\bar{\pi}_{i-1} + \bar{\pi}_i)}{(\sum \pi_r \bar{y}_r)} - \frac{\bar{y}}{(\sum \pi_r \bar{y}_r)} = \bar{d}_i \quad i = 1, \dots, g \quad (97)$$

$$\frac{2 \bar{y}_i \pi_{i-1} + \sum_i^g \pi_r \bar{y}_r}{(\sum \pi_r \bar{y}_r)} - \frac{\bar{y}}{(\sum \pi_r \bar{y}_r)} \cdot \bar{y}_i = e_i$$

$$i = 1, \dots, g \quad (98)$$

Then $\sqrt{n} (\bar{e} - \bar{y})$ has a limiting distribution that is normal with mean 0 and variance

$$\sum_{i=1}^g \bar{d}_i^2 \nu_i^2 + e \bar{K} e' \quad \dots \quad \dots \quad (99)$$

Let
$$\gamma = \frac{2 \sum i \nu_i^2}{g(\nu_1 + \dots + \nu_g)} - \frac{g+1}{g} \quad \dots \quad (100)$$

$$\frac{2i}{g(\nu_1 + \dots + \nu_g)} - \frac{\gamma}{(\nu_1 + \dots + \nu_g)} = d_i; i = 1, \dots, g \quad (101)$$

The $\sqrt{m} (C - \gamma)$ has a limiting distribution that is normal with mean 0 and variance

$$\bar{d} \wedge \bar{d}' \quad \dots \quad \dots \quad (102)$$

Though we have not explicitly mentioned, it should be noted that (46) is assumed when fixed interval analysis is employed and (52) is assumed when fractile analysis is employed, in this section.

6.5. Methods of testing with just one sample.

Among the statistics \bar{F} , $\bar{\Delta}$ and $\bar{\Gamma}$ that can be used to test the null hypothesis (9) of fixed interval analysis, $\bar{\Gamma}$ is most suited for practical use since its limiting distribution is most simple. Further $\bar{\Gamma}$ corresponds to the statistic that is used for testing simultaneous that several means are zero when all the variances are known. When the variables concerned are independently normally distributed this test corresponds to the Hotelling's T. Thus $\bar{\Gamma}$ seems to be the best criterion we can use in this situation.

$$\bar{\Gamma} = \sum [(\bar{v}_i - \bar{v}'_i)^2 / (\frac{\bar{t}_i^2}{\pi_i n} + \frac{\bar{t}'_i^2}{\pi'_i n'})] \dots \quad (103)$$

We note that $\frac{\bar{t}_i^2}{n_i}$ and $\frac{\bar{t}'_i^2}{n'_i}$ are consistent for $\frac{\bar{t}_i^2}{\pi_i n}$ and $\frac{\bar{t}'_i^2}{\pi'_i n'}$, $i = 1, \dots, g$, the unknown constants that enter in $\bar{\Gamma}$. Let

$$\bar{\Gamma}^* = \sum [(\bar{v}_i - \bar{v}'_i)^2 / (\frac{\bar{t}_i^2}{n_i} + \frac{\bar{t}'_i^2}{n'_i})] \dots \quad (104)$$

We note that $\bar{\Gamma}^*$ can be calculated from S and S' and that its limiting distribution is a χ^2 with g degrees of freedom. The critical region for testing the null hypothesis of fixed interval analysis will be the region of large values of $\bar{\Gamma}^*$

Our applications so far (of the Theorems 6.1, 6.2 etc.) depend in essence on the fact that in large samples we can treat v_1, \dots, v_g as independently normally distributed variables. This can be used in many ways. For instance, if we have two samples $S_{(1)}$ and $S_{(2)}$ (pooled, they form S) from one population P and only one, S' , from the second population P' , our test criterion would be

$$\bar{\Gamma}^* = \sum [(\bar{v}_i - \bar{v}'_i)^2 / (\frac{n_{i(1)} \bar{t}_i^2(1) + n_{i(2)} \bar{t}_i^2(2)}{n_{i(1)} + n_{i(2)}} + \frac{\bar{t}'_i^2}{n'_i})] \dots \quad (105)$$

The limiting distribution of this statistic is again χ^2 with g d.f.

We have seen in section 6.4 (e.g. in (86)) that

$$y \sim MN (y, \Lambda/m) \dots \quad (106)$$

Comparing this relation with (79) we note an important difference between fixed interval analysis and fractile analysis. These relations show that \bar{v} and v are asymptotically normal with variance covariance matrices $\bar{\Lambda}/n$ and Λ/m , respectively, $\bar{\Lambda}$ is always diagonal, but Λ is not diagonal in general. $\bar{\Lambda}$ is diagonal only when $\lambda_i = \dots = \lambda_{i-1}$ for all i . This does not hold in general, for instance when $\lambda(x)$ is strictly monotone. Thus whereas $\bar{v}_1, \dots, \bar{v}_g$ can be considered to be independent in large samples, v_1, \dots, v_g are dependent even in the limit.

Among the several measures of divergence introduced in section 6.1, Δ and $\bar{\Gamma}$ can be used since their limiting distributions under the null hypothesis are of a simple form.

$$m\Delta = \sum m(v_i - v'_i)^2$$

has a limiting distribution which is the distribution of $\sum_1^g \beta_i \chi_i^2$ where $\chi_1^2, \dots, \chi_g^2$ are independent chi-squares with 1 d.f. and β_1, \dots, β_g are the latent roots of $\Lambda + c\Lambda'$. This distribution can be approximated by [see Satterthwaite (1946). For other approximations see Robbins and Pitman (1949)] dZ where Z has a χ^2 -distribution with α d.f.; d and α are given by the relations

$$d = \sum \beta_i^2 / \sum \beta_i \tag{107}$$

$$\alpha = (\sum \beta_i)^2 / \sum \beta_i^2 \tag{108}$$

$$\bar{\Gamma} = (v - v') \left(\frac{1}{m} + \frac{1}{m'} \right)^{-1} (v - v')' \quad (109)$$

has a limiting distribution that is a χ^2 with g d.f.

These two statistics can not be used in practice unless Λ and Λ' are known. In trying to estimate Λ from S it is natural to use t_i^2 as an estimate of Λ_{ii} . It can be shown from some generalisations of the results of Hoeffding (1953) that t_i^2 is not even consistent for Λ_{ii} in general. This unfortunate fact that no simple consistent estimate of Λ is available from S , is a great set-back to the construction of nonparametric tests in the method of fractile analysis.

We therefore proceed to construct convenient test procedures when (Y, X) is known to follow some special distribution. Let us assume that the distribution of (Y, X) is bivariate normal with parameters $\nu, \mu; \sigma^2, \rho^2; \rho$. We now evaluate the matrix Λ for this distribution. We note that Λ does not depend on ν, μ and σ^2 . Using (76) we find that Λ can be written down in the form

$$\Lambda = \sigma^2 I + \rho^2 \sigma^2 Q \quad \dots \quad (110)$$

where I is the identity matrix and Q is a $g \times g$ matrix depending only on ρ . It can be easily demonstrated that the matrix Q is doubly symmetric, i.e.

$$Q_{i,j} = Q_{j,i} = Q_{g-i+1, g-j+1} = Q_{g-j+1, g-i+1} \quad (111)$$

The matrix Q has certain interesting properties (not basic to our main work) which are easily derived from the fact that $\frac{1}{g} \sum v_i = \bar{y}$,

the mean of y 's. Thus, Q is negative semi-definite, $\sum_j Q_{ij} = 0$ for each i , one of the latent roots of Q is zero and so on.

In table I, at the end of this chapter, we have given these matrices Q when $g = 2, 3, 4, \dots, 10$. Since the matrix Q is doubly symmetric, we give only Q_{ij} for $j \leq i \leq g-j+1$, $1 \leq j \leq \lfloor \frac{g+1}{2} \rfloor$ for each g .

We have also tabulated the latent roots (q_1, \dots, q_g) of Q in table II for $g = 2, 3, 4, \dots, 10$. Actually Q, q_i, μ_i etc.. all depend on g , so that $Q_g, q_{g,i}, \mu_{g,i}$ etc., would be the more appropriate notations for them. We however drop the suffix g whenever we feel that it will not cause confusion.

Let $\hat{\tau}^2$ and $\hat{\sigma}^2 \tau^2$ be consistent estimates of τ^2 and $\sigma^2 \tau^2$ respectively. $\hat{\Lambda}$ is now consistently estimated by $\hat{\Lambda} = \hat{\tau}^2 I + \hat{\sigma}^2 \tau^2 Q$. (Q can be got from Table I for some values of g .) Then

$$\Gamma^* = (\underline{y} - \underline{y}') \left(\frac{\hat{\Lambda}}{m} + \frac{\hat{\Lambda}'}{m'} \right)^{-1} (\underline{y} - \underline{y}')' \quad \dots \quad (112)$$

will be distributed as a χ^2 with g d.f.

The latent roots of $\hat{\Lambda}$ are given by $\hat{\tau}^2 + \hat{\sigma}^2 \tau^2 q_1, \dots, \hat{\tau}^2 + \hat{\sigma}^2 \tau^2 q_g$ and these are consistent estimates of the latent roots of Λ . Let us denote the consistent estimates of the latent roots of $\Lambda + c \Lambda'$ (which is got in the same way) by $\hat{\beta}_1, \dots, \hat{\beta}_g$. Then the limiting distribution of $m\Delta$ can be approximated by $\sum \hat{\beta}_i \chi_i^2$ where $\chi_1^2, \dots, \chi_g^2$

are independent chisquares with 1 d.f. We should note that large values of the statistics $m \Delta$ and Γ^* form the critical region for testing the null hypothesis (19) of fractile analysis.

We now suggest two methods of estimating τ^2 and $\rho^2 \tau^2$ from the sample S. Let t^2 and r be the sample variance of y and the sample correlation coefficient between y and x . Then t^2 and $r^2 t^2$ are consistent estimates of τ^2 and $\rho^2 \tau^2$ respectively. These can now be used as suggested in the preceding paragraph. Since the sample size of S will be large it will be time consuming to compute t^2 and r . We now suggest another method that will be computationally easier.

This method makes use of the sample S only through u_i 's and v_i 's. Consider

$$\tilde{t}^2 = \frac{1}{g} \sum v_i^2 - \left(\frac{1}{g} \sum v_i \right)^2 \quad \dots \quad (113)$$

$$\tilde{s}^2 = \frac{1}{g} \sum u_i^2 - \left(\frac{1}{g} \sum u_i \right)^2 \quad \dots \quad (114)$$

$$\tilde{r} \tilde{s} \tilde{t} = \frac{1}{g} \sum u_i v_i - \left(\frac{1}{g} \sum u_i \right) \left(\frac{1}{g} \sum v_i \right) \quad \dots \quad (115)$$

We can easily demonstrate (by showing that the expectations converge and variances tend to zero) that $m \tilde{t}^2$, $m \tilde{s}^2$ and $m \tilde{r} \tilde{s} \tilde{t}$ are consistent for

$$\frac{g-1}{g} \tau^2 + \frac{1}{g} \tau^2 \rho^2 L, \quad \frac{g-1}{g} \sigma^2 + \frac{1}{g} \sigma^2 L \quad \text{and} \quad \frac{g-1}{g} \rho \tau \sigma + \frac{1}{g} \rho \sigma \tau L$$

respectively, where

$$L = \text{Tr } Q + \mu_{g,1}^2 + \dots + \mu_{g,g}^2 \dots \quad (116)$$

$$\text{Tr } Q = q_{g,1} + \dots + q_{g,g} \dots \dots \quad (117)$$

In tables III and IV we have given $\text{Tr } Q_g$ and $\sum_1^{gs} \mu_{g,i}^2$ for $g = 2, 3, \dots, 10$. Making use of these tables, we can easily obtain consistent estimates for γ^2 and $q^2 \gamma^2$ based on \tilde{t}, \tilde{s} and \tilde{r} .

The following are the Q matrices Q_g for $g = 2(1)10$, $(Q_g)_{ij}$ has been given only for $j \leq i \leq g-j+1$, $1 \leq j \leq \lceil \frac{g+1}{2} \rceil$, since Q is doubly symmetric. (See (111).)

Table 1.
 $10^5 \times (Q_g)_{ij}$

j	1	2	3
i	(g = 2)		
1	-31831		
2	31831		
i	(g = 3)		
1	-42954		
2	28431	-56862	
3	14523		
i	(g = 4)		
1	-49140		
2	25339	-65199	
3	14903	24958	
4	8899		
i	(g = 5)		
1	-53210		
2	23160	-69826	
3	14141	21961	-72205
4	9677	15028	
5	6231		

Table I(contd.)

 $10^5 \times (q_g)_{ij}$

j	1	2	3	4	5
i	(g = 6)				
1	-56149				
2	21549	-72858			
3	13389	19854	-76069		
4	9513	14107	19207		
5	6987	10361			
6	4712				
i	(g = 7)				
1	-50399				
2	20301	-75034			
3	12735	18296	-78601		
4	9216	13240	17285	-79482	
5	7023	10089	13172		
6	5379	7728			
7	3744				
i	(g = 8)				
1	-60194				
2	19302	-76689			
3	12179	17091	-80419		
4	8907	12500	15866	-81711	
5	6913	9701	12314	15509	
6	5489	7703	9777		
7	4324	6068			
8	3081				
i	(g = 9)				
1	-61670				
2	18477	-78001			
3	11702	16127	-81804		
4	8617	11875	14769	-83312	
5	6761	9317	11588	14202	-83736
6	5466	7532	9368	11482	
7	4460	6146	7644		
8	3586	4941			
9	2602				

Table I (contd.)

$10^5 \times (q_g)_{ij}$

j	1	2	3	4	5
i	(g = 10)				
1	-62913				
2	17783	-79073			
3	11290	15333	-82902		
4	8352	11343	13891	-84531	
5	6600	8963	10976	13193	-85187
6	5394	7326	8971	10782	12982
7	4480	6084	7451	8955	
8	3727	5062	6199		
9	3044	4134			
10	2241				

Table II

Latent roots $(q_{g,1}, \dots, q_{g,g}) = q_g$ of the Q matrices: $g=2(1)10,$

$$10^5 \times q_g$$

$$g = 2$$

(00000, -63662)

$$g = 3$$

(00000, -85294, -57478)

$$g = 4$$

(00000, -93249, -80483, -54946)

$$g = 5$$

(00000, -96323, -77658, -90674, -53621)

$$g = 6$$

(00000, -97716, -52822, -94976, -88817, -75820)

$$g = 7$$

(00000, -96960, -52293, -98448, -93896, -87424, -74527)

$$g = 8$$

(00000, -51920, -98876, -97989, -86339, -96315, -93020, -73569)

$$g = 9$$

(00000, -51644, -99148, -85466, -98579, -97583, -95767, -92295, -72828)

$$g = 10$$

(00000, -51433, -98945, -84746, -99332, -98309, -97226, -95297, -91684, -72239)

Table III

Trace $\text{Tr } Q$, of the Q matrices; $g = 2(1) 10$.

g	$-\text{Tr } Q_g$
2	0.63662
3	1.42770
4	2.28678
5	3.18277
6	4.10152
7	5.30550
8	5.98026
9	6.93310
10	7.82212

Table IV

The following gives $\sum_1^g \mu_{g,i}^2$ for $g=2(1) 10$.

g	$\sum_1^g \mu_{g,i}^2$
2	0.636620
3	2.379689
4	3.442239
5	4.484782
6	5.516166
7	6.540582
8	7.560287
9	8.576615
10	9.590464

REFERENCES

- Achieser, N.I. (1956): Theory of approximation Translated by C.J.Hyman, Fredrick Kugar, New York.
- Bahadur, R.R. (1960) : On the asymptotic efficiency of tests and estimates, Sankhyā, 22, 229-252.
- Box, G.E.P. (1949): A general distribution theory for a class of likelihood criteria, Biometrika, 36, 317-346.
- Cramer, H. (1946): Mathematical methods of statistics, Princeton University press.
- Das, Rhea, S. (1960): Applications of fractile graphical analysis to psychometry: I. Item analysis, Psychological Studies, 5, 11-18.
- Das, Rhea, S. and Sharma, K.N. (1960): Applications of fractile graphical analysis to psychometry: II. Reliability, Psychological Studies, 5, 71-77.
- Davis, H.T. (1941): The theory of econometrics, The principia press, Bloomington, Indiana.
- Fisher, R.A. (1935): The design of experiments, Oliver and Boyd, Edinburgh
- Fraser, D.A.S. (1951): Sequentially determined statistically equivalent blocks, Ann. Math. Stat., 22, 372-381.
- (1953): Non-parametric theory scale and location parameters, Can. Jour. Math., 6, 1-26.
- (1953a): Non-parametric tolerance regions, Ann. Math. Stat., 24, 44-55.
- (1957): Non-parametric methods in statistics, John Wiley and Sons, New York.
- Fraser, D.A.S. and Guttman, I. (1956): Tolerance regions, Ann. Math. Stat., 27, 16-32.

- Fraser, D. A. S. and Wormleighton, R. (1951): Non-parametric estimation, IV Ann. Math. Stat., 22, 294-298.
- Ghosh, B. N. (1949): Interpenetrating (network of) sub-samples, Cal. Stat. Assos. Bull., 2, 108-119.
- Gnedenko, B. V. (1943): Sur la distribution limits du terms maximum d'une serie aleatoire, Ann. Math., 44, 423-453.
- Hoeffding, W. (1948): On a class of statistics with asymptotically normal distributions, Ann. Math. Stat., 19, 293-325.
- (1951): A combinatorial central limit theorem, Ann. Math Stat., 22, 558-566.
- (1951a): Optimum non-parametric tests, Proc. Second Berkeley Symposium 83-92, University of California Press
- (1952): The large sample power of tests based on permutations of observations, Ann. Math. Stat., 23, 169-192.
- (1953): On the distribution of expected values of the order statistics, Ann. Math. Stat., 24, 93-100.
- Hoeffding, W. and Robbins, H. (1948): The central limit theorem for depend ent random variables, Duke Math. Jour., 15, 773-780.
- Iyengar, N. S. (1960): On a method of computing Eugel elasticities from concentration curves, Econometrica, 28, 882-891.
- Kawata, T. (1951): Limit distributions of single order statistics, Rep. Stat. Appl. Res. Union of Japanese Scientists and Engineers, 1, 4-9.
- Kendall, M. G. (1959): The advanced theory of statistics, Vol.2, (third edition), Charles Griffin and Company, London, pp.72.
- Lahiri, D. B. (1954): Technical paper on some aspects of the development of the sample design, National Sample Survey of India, Report No.5, Indian Statistical Institute, Calcutta.
- (1957): Observations on the use of interpenetrating samples in India, Bulletin de l'institut international de statistique, 36, Book 3, 144-152.

- Lehmann, E. (1949): Theory of hypothesis testing, mimeographed notes, University of California.
- (1950): Theory of estimation, mimeographed notes, University of California.
- (1950a): Some principles of the theory of testing hypothesis, Ann. Math. Stat., 21, 1-26.
- (1951): Consistency, unbiasedness of certain non-parametric tests, Ann. Math. Stat., 22, 165-179.
- (1953): The power of rank tests, Ann. Math. Stat., 24, 23-43.
- (1959): Testing statistical hypothesis, John Wiley and Sons, New York.
- Lehmann, E. and Stein, C. (1949): On the theory of some non-parametric hypothesis, Ann. Math. Stat., 20, 28-45.
- Mahalanobis, P. C. (1946): Recent experiments at the Indian Statistical Institute, Jour. Royal Stat. Soc., 109, 325-378.
- (1958): Lectures in Japan - draft proof.
- (1960): A method of fractile graphical analysis, Econometrica, 28, 325-351.
- (1961): A method of fractile graphical analysis, Sankhya, Series A 23, 325-251.
- Mood, A. M. (1940): The distribution theory of runs, Ann. Math. Stat., 11, 367-392.
- Mosteller, F. (1948): A k-sample slippage test for an extremal population, Ann. Math. Stat., 19, 58-65.
- Parthasarathy, K. R. (1961): On a class of estimates of the density function, unpublished.
- Pearson, Karl (1900): On the criterion that a given system of deviations from the probable in the case of a correlated system is such that it can be reasonably supposed to have arisen from random sampling, Phil. Mag. Series 5, 50, 157-175.

- Pitman, E.J.G. (1937): Significance tests which may be applied to samples from any populations, Supplement to Jour. Royal Stat. Soc., 4, 119-130.
- Pitman, E.J.G. (1937a): Significance tests which may be applied to samples from any populations II. The correlation coefficient test, Supplement to Jour. Royal Stat. Soc., 4, 225-232.
- Pitman, E.J.G. (1976): Significance tests which may be applied to samples from any populations III. The analysis of variance test, Biometrika, 29, 322-335.
- Prohorov, Yu. V. (1959): An extremal problem in probability theory. Theoria Veroyot i. ee Primen., 4, 211.
- Rao, C. R. (1952): Advanced statistical methods in biometric research, John Wiley and Sons, New York.
- (1961): Efficient estimates and optimum inference procedures in large samples, read at the Royal Statistical Society - to be published in the Jour. Royal Stat.Soc.
- Robbins, H. and Pitman, E.J.G.(1949): Application of a method of mixtures to quadratic forms in normal variables, Ann. Math. Stat. 20, 552-560.
- Roy, J. Chakravorty, I.M. and Laha, R.G. (1959): A study of concentration curves as description of consumer pattern, Studies relating to planning for national development, 2, 73-82. Indian Statistical Institute, Calcutta.
- Satterthwaite, (1946): An approximate distribution of estimates of variance components, Biometrics, 2, 110-114.
- Savage, I.R. (1953): Bibliography on non-parametric statistics, Jour. Amer. Stat. Assos., 48, 844-906.
- Scheffe, H. (1947): A useful convergence theorem for probability distributions, Ann. Math. Stat., 18, 434-438.
- Scheffe, H. and Tukey, J.W. (1945): Non-parametric estimation, I. Validation of order statistics, Ann. Math. Stat., 16, 187-192.
- Sethuraman, J. (1961): Some results concerning asymptotic distributions and their applications, Unpublished Ph.D. thesis submitted to the Indian Statistical Institute.

- Sethuraman, J. (1961a): Some limit theorems connected with fractile graphical analysis, Sankhya, Series A, 23, 79-60.
- (1961b): Some limit theorems for joint distributions, sankhya Series A, 23,
- Siegal, S. (1956): Non-parametric statistics for the behavioural sciences, McGram-Hill Book Company, New York.
- Smirnov, N. V. (1949): Limit distributions for the terms of a variational series, Trudy. Mat. Inst. A. Steklova, 25, 1-61(Russian).
- (1952): Limit distributions for the terms of a variational series, Ann. Math. Soc. Translation No.67.
- Som, R. K. (1960): A statistical study of the population of India, Unpublished D.Phil. Thesis, submitted to the University of Calcutta.
- Takenochi, K. (1961): On some properties of the error area in fractile graphical analysis, Sankhya Series A, 23, 65-78.
- Terry, M.E. (1952): Some rank order statistics which are most powerful against specific parametric alternatives, Ann. Math. Stat., 23, 346-366.
- Tukey, J.W. (1947): Non-parametric estimation II. Statistically equivalent blocks and tolerance regions - the continuous case, Ann. Math. Stat., 18, 529-539.
- (1948): Non-parametric estimation III. Statistically equivalent blocks and tolerance regions - the discontinuous case, Ann. Math. Stat., 19, 30-39.
- Varadarajan, V. S. (1958): A useful convergence theorem, Sankhya, 20, 221-222.
- Wald, A. (1950): Statistical Decision functions, John Wiley and Sons., New York.
- Yates, F. (1949): Sampling methods for censuses and surveys, Charles Griffin and Company, London.

