CHARACTERIZATIONS OF THE NORMAL DISTRIBUTION

by

GOH JEOK-ENG

M.Sc. MATHEMATICS

ABSTRACT

The problem of determining a statistical population belonging to a certain class of distributions is widely investigated in Mathematical Statistics. Of special interest is the characterization of the Normal distribution.

In this thesis, characterizations of the Normal distribution through different considerations are treated in great detail . Chapter II is concerned with the characterization of the Normal distribution by using specified or unspecified distributions of suitable statistics. Chapter III deals mainly with the property of independence of suitable statistics, such as linear statistics, linear and quadratic statistics, linear and polynomial statistics, by which the Normal distribution is characterized. Chapter IV gives some generalizations of some results in Chapter III by replacing the property of independence by a weaker condition of regression such

as constant regression and polynomial regression. Finally, Chapter V discusses the characterization of the Normal distribution through linear structural relations and through properties of sample estimators. Other characterizations which do not fall into the preceding categories are mentioned at the end of Chapter V.

CHARACTERIZATIONS OF THE NORMAL DISTRIBUTION

by

GOH JEOK-ENG

A thesis being submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science.

Department of Mathematics

McGill University

Montreal 110, Quebec

Canada

October 1969

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to Professor A.M. Mathai for his valuable guidance, helpful comments and constant encouragement in the preparation of this thesis.

TABLE OF CONTENTS

CHAPTER I

Characteristic Functions and Conditional Expectations

Introduction 1.0. The characterizations of populations have appeared in the literature from time to time, especially in the last two decades. Different methods, setups and techniques have been developed, among which characterizations through independence of statistics, distributions of statistics and regression properties are the main ones. In short, all these can be said characterizations through properties of statistics. In the monograph by Laha and Lukacs [48] characterizations of populations using the properties mentioned above are studied in great detail , and it also contains a complete bibliography of the work up to 1963 and its historical background. A great portion of research work on characterizations of population is mainly on the normal distribution.

In this thesis, we shall only deal with the characterizations of Normality. The motivation of compiling the work on the characterization of the normal distribution is apparent from the text of the thesis. Further discussion is given at the end of the thesis in Chapter V. Apart from most of the results contained in [48], we also discuss some other results which are not treated in [48] and some recent developments. Our aim is to give an up-to-date complete survey on this subject, and hence it is of expository nature and no original result is obtained in the thesis.

One important technique of the characterizations of Normality is to obtain a functional equation in the characteristic function and obtain a unique solution of this functional equation a functional form of the form exp[p(t)], where p(t) is a polynomial in t say. If the random variable X involved is not degenerate, i.e., X equals some constant with probability one, then exp[p(t)] can only be the characteristic function of some normal distribution in view of a theorem of Marcinkiewicz which states that if $f(t) = \exp[p(t)]$ is a characteristic function, and if p(t) is a polynomial in t, then the degree of the polynomial p(t) cannot be greater than two. Perhaps this is a great achievement in the history of the characterization of Normality. Needless to say, the importance of the characteristic functions to the studies of probability measures or random variables is always emphasized. In the monograph [66] by Lukacs, characteristic functions are studied from the mathematical point of view in great detail.

For our purpose, we introduce here the characteristic function of a distribution function or a random variable and some general properties of conditional expectation in order to provide necessary terminology and notations for the subsequent chapters.

1.1. Characteristic functions.

By a distribution function F(x) of a random variable x, we mean that F(x) is a real-valued function which is non-decreasing and right-continuous such that $F(+\infty) = 1$ and $F(-\infty) = 0$, and the mathematical expectation, denoted by $E(X)$ of the random variable X, we mean the integral $E(X) = \int_{R^1} x dF(x)$ taken in Lebesgue-Stieltjes sense.

The characteristic function $f(t)$ is defined as follows:

$$f(t) = E\left[\exp(itX)\right] = \int_{R^1} e^{itX} dF(x)$$

which is known in analysis as the Fourier-Stieltjes transform of F(x).

An important class of characteristic functions is the class of analytic characteristic functions. This class contains the characteristic functions of the well-known distributions such as the Normal distribution, the Gamma distribution and the Poisson distribution; etc. One of the most important properties of analytic functions is the uniqueness theorem: If the function f(z) is analytic (regular) in the domain D, and if there exists a sequence of points $z_1, z_2, \ldots$ in D having a limit point in D such that $f(z_n) = 0$, $n = 1,2,\ldots$, then the function f(z) vanishes on the entire domain D. The concept of analytic continuation plays an important role in the application of the uniqueness theorem. By an analytic continuation of a function f(z) on a set E, we mean a function F(z) which is analytic in some domain D containing E and coincides with f(z) in the set E. As a consequence of the uniqueness theorem of analytic functions, it is found that if the set E has at least one limit point contained within the domain D, then the function $f(z)$ has at most one analytic continuation to the domain D. This result is very useful for problems of characterizing populations. Instead of considering the characteristic function on the whole real line one needs only to determine the functional form of the characteristic function in some neighbourhood of the origin in such cases. We list some results of the analytic characteristic functions which will be frequently used in our subsequent work. The most important results concerning criteria for analytic characteristic functions refer to a class of entire functions. The proofs of them can be found in [66] and are therefore omitted.

Theorem 1. (Marcinkiewicz) If $f(t) = \exp[p_n(t)]$ is a characteristic function, where $p_n(t)$ is a polynomial of degree n, then n cannot be greater than two.

Proof. See pp. 146 [66].

Theorem 2. Every factor $f_1(z)$ of an entire characteristic function $f(z)$ is an entire characteristic function. The order of the factors of an entire characteristic function cannot be greater than that of $f(z)$.

Proof. See pp. 170 [66].

Theorem 3. (Cramer). The characteristic function $f(t)$ of a normal distribution has only factors which are characteristic functions of some normal distribution.

Proof. See pp. 174[66].

Theorem 4. (Linnik) Let $f_1(t)$, $f_2(t),\dots,f_n(t)$ be arbitrary characteristic functions, and let $a_1,a_2,\dots,a_n$ be positive real numbers. Assume that $f(t)$ is an analytic characteristic function and the relation

$$\prod_{j=1}^{n} [f_j(t)]^{a_j} = \exp\left[i\mu t - \frac{1}{2}\sigma^2 t^2\right]$$

holds in a neighbourhood of the origin, where μ and $\sigma^2 > 0$ are real, and $i^2 = -1$. Then the functions $f_j(t)$, $(j = 1,2,\dots,n)$ are characteristic functions belonging to the normal distribution.

Proof. See pp. 190-196 [66].

1.2. Conditional Expectations

There are several ways of introducting conditional expectation of a random variable Y given another random variable X, and is usually denoted by $E(Y|X)$. In statistical terminology, the conditional expectation X is referred as the regression of Y on X. One way of introducing the $E(Y|X)$ is by means of Radon-Nikodyon theorem. i.e., the conditional expectation $E(Y|X)$ is defined as a $\mathcal{F}_X$-measurable ($\mathcal{F}_X$ = σ-field generated by the random variable X) function up to a set of measure zero such that

$$\int_B E(Y|X)dp = \int_B Ydp \quad \text{for every } B \in \mathcal{F}_X \text{ where } p \text{ is}$$

the probability measure associated with the random variables X and Y.

We see that if B is the whole space, say Ω, then we have the well-known relation

$$E[E(Y|X)] = E(Y) .$$

1.3. General Properties of the Conditional Expectation

1. $E(X|X) = E(X)$

2. $E(X_1 + X_2|X) = E(X_1|X) + E(X_2|X)$.

3. Let X and Y be random variables such that $E(Y)$ exists and let f be a Borel measurable function. If $E(f(X)Y)$ exists, then $E[f(X)Y|X]=f(X)\ E(Y|X)$.

4. If X and Y are independent, then $E(Y|X) = E(Y)$.

All these proofs are straightforward, and hence are omitted. Generalization to a finite number of random variables is also straightforward. One version of $E(Y|X_1,\ldots,X_n)$ is a function $f\,(X_1,\ldots,X_n)$, where f is a Borel function of n variables $x_1,\ldots,x_n$ such that for every $B \in \mathcal{F}_{x_1 \ldots x_n}$ (σ-field generated by $X_1,\ldots,X_n$), we have

$$\int_B f(x_1,\ldots,x_n)dp = \int_B Ydp.$$

Also we have

$$E(E(Y|X_1,\ldots,X_n)) = E(Y).$$

For a detailed treatment of conditional expectation one may refer to [10] and [61].

CHAPTER II

Characterization of Normality by using known distribution of some Statistics.

2.1. Specified distributions of some statistics.

The problem of determining a theoretical distribution belonging to a given class is widely investigated in mathematical statistics. As mentioned in the previous chapter, we shall only confine our attention to the characterization of Normality. By a statistic $S(X_1,\ldots,X_n)$ of a random sample $X_1,X_2,\ldots,X_n$ from a random variable X with d.f. $F(x)$ ($X_1,X_2,\ldots,X_n$ are n i.i.d. r.v.'s having the same d.f. $F(x)$ as X), we shall understand a measurable and single-valued function of $X_1,X_2,..,X_n$, more precisely, $S(X_1,X_2,\ldots,X_n)$ is itself a random variable. If assumptions are imposed on the properties of some specific statistics based on a given random sample, then they will, in general, restrict or determine the distribution of the population under consideration. For instance, assumptions that give explicitly the distribution of S or relate it in some specified manner to the d.f. $F(x)$ can be used to characterize various populations. In this chapter, we make assumptons that (i) $S(X_1,X_2,\ldots,X_n)$ has specified distribution, (ii) $S(X_1,X_2,\ldots,X_n)$ has the same distribution as $F(x)$ by which the normality is characterized.

We first discuss the well-known Cramer's theorem. The theorem was conjectured by P. Levy and was proved by Cramer in 1936 [11] .

Theorem 1. (Cramer) Let X_1 and X_2 be two independent r.v.'s. If X_1+X_2 has a univariate normal distribution, then X_1 and X_2 are normal. (here "identically distributed" is not needed).

This theorem actually is a restatement of Theorem 3, in Chapter One. For a proof, see p. 272 [61]

In the light of this theorem, we have the following as a corollary.

Corollary 1.1. Let $X_1, X_2, \ldots, X_n$ be $n(n \geq 2)$ independent r.v.'s and let $L = \sum_{i=1}^{n} a_i X_i$, a_i, $(i = 1,2,\ldots,n)$ are real. If L is normal, then each X_i, $(i = 1,2,\ldots,)$ is normal.

Based on this corollary, Linnik [56] gave a very elegant proof of the so-called "Skitovich-Darmois" theorem which states, in short, that in two independent linear functions of independent r.v.'s, the components having non-zero coefficients in both forms are normally distributed. He also showed that Cramer's theorem can be deduced from Skitovich-Darmois' theorem. Incidently this result reveals the generality of 'Skitovich-Darmois' theorem. We shall discuss the Skitovich-Darmois' theorem in Chapter III.

One of the most important univariate distributions used in theoretical or applied statistics is no doubt the normal distribution. It is also true that the multivariate normal distribution plays an important role in statistical inference in multivariate analysis.

Just as the univariate normal density function $\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$

where $\sigma > 0$ and μ are real, the density of a multivariate normal distribution has an analogous form, which is defined to be

$$f(X_1,X_2,\ldots,X_p) = (2\pi)^{-p/2}|A|^{1/2}\exp\left[-\frac{1}{2}(X-\underset{\sim}{\mu})'A(X-\underset{\sim}{\mu})\right]$$

where $X = (X_1,\ldots,X_p)'$ is a p-dimensional random vector, $\underset{\sim}{\mu}$ is a p-dimensional scalar column vector, (vector with real components), A is a positive definite p x p matrix and $(X-\underset{\sim}{\mu})'$ denotes the transpose of $X-\underset{\sim}{\mu}$. It can be shown by integration that $A^{-1} = \text{cov}(X)$(the covariance matrix, defined as $(\text{cov}(X_i,X_j))$, $i,j = 1,2,\ldots,p$, see [81], and $E(X) = \underset{\sim}{\mu}$)

It is interesting to note that the marginal distributions, the conditional distributions derived from a multivariate normal distribution are also normal distributions. This is one of the characteristics of the multivariate normal distribution.

The study of multivariate normal distributions is always not as simple as the study of the univariate case. To a great extend, this difficulty is overcome by the result due to Cramer and Wold which states that the distribution of a normally-distributed vector is completely characterized by the one-dimensional normal distribution of the linear function X'L for every fixed scalar vector L. This result enables one to bring over the study of a multivariate normal distribution to that of linear statistics which in most of the cases is found convenient. The study of the multivariate normal distribution adopting this line of approach is revealed in the book [81] by Rao, where a series of results are obtained by using the known properties in the univariate case.

We now give the result by Cramer and Wold mentioned above, and by using it, we obtain an analog of Theorem 1, in multivariate case.

Theorem 2. (Cramer-Wold). Let $X = (X_1, X_2, \ldots, X_p)'$ be a p-dimensional random vector. Then the distribution of $X = (X_1, X_2, \ldots, X_p)'$ has a p-variate normal distribution iff for every p-dimensional non-zero scalar vector $L = (\ell_1\ \ell_2, \ldots, \ell_p)'$, $X'L = \sum_{i=1}^{p} \ell_i X_i$ has a univariate normal distribution.

Proof. It is easy to show by integration that the marginal distribution of any component in the p-dimensional random vector having p-variate normal distribution is normal. We shall only show the sufficiency. Assume that $X'L$ is univariate normal for every non-zero scalar vector $L = (\ell_1, \ldots, \ell_p)'$. Since $X'L$ is normal with the parameters μ and σ^2 say, we may write

$$E(X'L) = \mu \quad \text{and} \quad Var(X'L) = \sigma^2$$

But $$E(X'L) = U'L \quad \text{and} \quad Var(X'L) = L'ML \qquad (2.1)$$

where $E(X) = U$ and $cov(X) = M$ (covariance matrix of X).

Consider the ch.f. of the r.v. $X'L$

$$E(\exp(i\ t X'L)] = \exp\ [it\mu - \frac{1}{2}\sigma^2 t^2] \quad \text{since } X'L \text{ is}$$

normal with mean μ and variance σ^2.

Also, in view of the equation (2.1) we have

$$E[\exp(it\ X'L)] = \exp[it\mu - \frac{1}{2}\sigma^2 t^2]$$

$$= \exp[it\ U'L - \frac{1}{2} L'ML\ t^2] \quad .$$

Let $T = tL$. Then we have

$$E[\exp(i X'T)] = \exp[i U'T - \frac{1}{2} T'MT]$$

This is the ch.f. of a multivariate normal distribution. By the uniqueness theorem, X has a p-variate normal distribution.

In virtue of Theorem 2, we have the following corollary.

Corollary 2.1. Let $X = (X_1, \ldots, X_p)'$ and $Y = (Y_1, \ldots, Y_p)'$ be two independent p-dimensional random vectors. If $Z = X+Y$ is p-variate normal then both X and Y are p-variate normal.

By considering any linear function $Z'L$, using Theorem 2 and Theorem 1. Taking into account of the independence of $X'L$ and $Y'L$ the result easily follows.

2.2. Identically distributed linear statistics.

Let $X_1, \ldots, X_n$ be a random sample from X with distribution function $F(x)$. Consider two different statistics $S_1 = S_1(X_1, \ldots, X_n)$. $S_2 = S_2(X_1, \ldots, X_n)$. In general, the properties of S_1 and S_2 may have a great deal of difference. But for certain statistics S_1 and S_2 it might happen that S_1 and S_2 are identically distributed, or it might happen that S_1 and S_2 are stochastically independent. We shall discuss the characteristics of the independence of statistics in the next chapter. One may think that the properties of two statistics having the same distribution is a characteristic property of some populations.

Indeed, it has long been known that if X_1 and X_2 are independent with common d.f. F(x) with mean zero and variance one, and if $\frac{X_1+X_2}{\sqrt{2}}$ also has the same d.f. F(x), then F(x) is the standard normal distribution function. This result has already been generalized to some linear functions with suitable coefficients of a finite number of i.i.d. r.v.'s. Perhaps the most remarkable results in connection with this aspect are revealed in several papers [50], [51], [52] by Linnik, in which the relation between "independently distributed " and "identically distributed" statistics is investigated in great detail , and a necessary and sufficient condition for the equivalence of the statement that the population is normal with the assertion that two linear statistics are identically distributed is obtained. He also characterized a class of symmetrical distribution which contains the convolutions of symmetric stable laws. Several principal results of Linnik are treated in great detail in [48]. A theorem of Marcinkiewicz on identically distributed linear functions of infinitely (or finitely) many i.i.d.r.v's is also discussed in [48]. We shall only present the statements of these results here.

Theorem 3. (Shimizu) Let $X_1, X_2, \ldots, X_n$ be a random sample from X with d.f. F(x) with mean μ and finite variance σ^2. If there exist non-zero constants $a_1, \ldots, a_n$ such that $L = \sum_{i=1}^{n} a_i X_i$ and X are identically distributed, the F(x) is normal.

Proof. We note that the result can be established by direct applications of some properties of analytic ch.f.'s (see pp. 182 [66]). We give another proof due to Shimizu.

For simplicity, we prove the theorem for $n = 2$ and with $a_1 = a$ and $a_2 = b$. Consider $E\left\{\exp[it(aX_1 + bX_2)]\right\}$ and take into account of the independence of X_1 and X_2. We have

$$E\left\{\exp[it(aX_1 + bX_2)]\right\} = E[\exp(itaX_1)]E[\exp(itbX_2)]$$

By assumption, $f(t) = f(at)\ f(bt)$ (3.1)

where $f(t)$ is the ch.f. of $F(x)$. In a neighbourhood of the origin, we can introduce $\log f(t) = \phi(t)$ (here after $\phi(t)$ will be referred as the cumulant generating function) so that (3.1) beomes

$$\phi(t) = \phi(at) + \phi(bt) \quad . \tag{3.2}$$

Since F has a finite variance, $f(t)$ has continuous second derivative at the origin. We can differentiate (3.2) twice and obtain

$$\phi'(t) = a\,\phi'(at) + b\phi'(bt)$$

$$\phi''(t) = a^2\phi''(at) + b^2\phi''(bt) \tag{3.3}$$

From (3.3) we have

$$\phi''(at) = a^2\,\phi''(a^2t) + b^2\phi''(abt)$$
$$\phi''(bt) = a^2\phi''(abt) + b^2\phi''(b^2t) \tag{3.4}$$

Substitute (3.4) into (3.3), and we get

$$\phi''(t) = (a^2)^2\phi''(a^2t) + a^2b^2\phi''(abt) + a^2b^2\phi''(abt) + (b^2)^2\,\phi''(b^2t) .$$

By induction, we can show for any n (positive integer)

$$\phi''(t) = \sum_{k=o}^{n} \binom{n}{k} (a^2)^k (b^2)^{n-k}\, \phi''(a^k b^{n-k}\, t) \tag{3.5}$$

By letting $t = 0$ in (3.5), we have

$$-\sigma^2 = \sum_{k=o}^{n} \binom{n}{k} (a^2)^k (b^2)^{n-k}(-\sigma^2) \tag{3.6}$$

since $\phi''(0) = -\sigma^2$. It follows that

$$\sum_{k=o}^{n} \binom{n}{k} (a^2)^k (b^2)^{n-k} = 1 \tag{3.7}$$

which in turns implies that

$$0 < |a| < 1 \ , \ 0 < |b| < 1. \tag{3.8}$$

Taking into account of equations (3.5) and (3.6), we have

$$\left| \phi''(t) - \sigma^2 \right| = \left| \sum_{k=o}^{n} \binom{n}{k} (a^2)^k (b^2)^{n-k} [\phi''(a^k b^{n-k}\, t) - \sigma^2] \right|$$

$$\leq \sum_{k=o}^{n} \binom{n}{k} (a^2)^k (b^2)^{n-k} \left| \phi''(a^k b^{n-k}\, t) - \sigma^2 \right| \tag{3.9}$$

Since $\phi''(t)$ is continuous at $t = 0$, and since $0 < |a| < 1$ $0 < |b| < 1$, by continuity, we can make

$$|\phi''(a^k b^{n-k} t) - \sigma^2|$$

so small as we please, say $\epsilon > 0$.

i.e. (3.9) becomes

$$|\phi''(t) - \sigma^2| \leq \epsilon \sum_{k=o}^{b} \binom{n}{k} (a^2)^k (b^2)^{n-k} = \epsilon$$

in virtue of (3.7). Since ϵ is arbitrary, it follows that

$$\phi''(t) = \sigma^2 .$$

Hence $f(t) = \exp\left\{i\mu t - \frac{1}{2}\sigma^2 t^2\right\}$ holds in a neighbourhood of the origin. By analytic continuation $f(t) = \exp\left\{i\mu t - \frac{1}{2}\sigma^2 t^2\right\}$ is true for all t. This completes the proof.

Corollary 3.1. Let X and Y be independent with common d.f. F(x) with mean zero and variance one. Suppose that $\frac{X+Y}{\sqrt{2}}$ also has the d.f. F(x). Then F(x) is the standard normal distribution function.

The foregoing result indicates the possibility that two different linear statsitics might be identically distributed. This problem was first investigated by J. Marcinkiewcz [69] who obtained the following result.

Theorem 4. Let $X_1, X_2, \ldots, X_n, \ldots$ be a finite or infinite sequence of i.i.d. r.v.'s with common d.f. F(x). Assume that the two (finite or infinite) sums $\Sigma\, a_j X_j$ and $\Sigma\, b_j X_j$ exist, and F(x) has moments of any order.

If $\Sigma a_j X_j$ and $\Sigma b_j X_j$ are identically distributed, then either the sequence $\{|a_j|\}$ and $\{|b_j|\}$ are identical, except for the order of the terms, or $F(x)$ is normal (possibly degenerate).

The proof of this result is clearly presented in [48] . This result gives us a sufficient condition for the realization of the normality of $F(x)$. But in statistical analysis, only finite samples are used, and hence it would be much more interesting to formulate the previous result for finite sums. We give another analogous result of the previous one as a characterization of the normality for the finite case.

Theorem 5. Let $X_1, X_2, \ldots, X_n$ be a sample from a r. v. with d.f. $F(x)$ having all finite absolute moments $\beta_k = \int_{-\infty}^{\infty} |x|^k d\,F(x)$, $k = 1,2,\ldots,\infty$.

Let $L_1 = \sum_1^n a_k X_k$ and $L_2 = \sum_1^n b_k X_k$ be two linear functions of $X_1, X_2, \ldots, X_n$ with real coefficients. Suppose that the numbers $|a_1|, |a_2|, \ldots, |a_n|$ are not a permutation of the numbers $|b_1|, |b_2|, \ldots, |b_n|$ and that $\sum_{k=1}^n a_k = \sum_{k=1}^n b_k$, $\sum_{k=1}^n a_k^2 = \sum_{k=1}^n b_k^2$. Then L_1 and L_2 are identically distributed iff $F(x)$ is normal.

A proof of the result is given in [65] . We present the proof in details for it may be of theoretical interest. By considering the characteristic functions of L_1 and L_2 assuming $F(x)$ is normal, it can be easily deduced that the ch. f 's of L_1 and L_2 are identical. By the uniqueness theorem, L_1 and L_2 must be identically distributed.

Necessity. There is no loss of generality in assuming that $F(x)$ is symmetrical for we may consider $L_1' = \sum_1^n a_i(X_i - Y_i)$ and $L_2' = \sum_1^n b_i(X_i - Y_i)$, where Y_i, $i = 1,2,\ldots,n$ are i.i.d. r.v.'s as X_i, $i = 1,2,\ldots,n$ respectively, in view of Cramer's theorem. Clearly, if L_1 and L_2 are identically distributed so are L_1' and L'_2.

Let $f(t)$ be the ch.f. of $F(x)$, and let

$$\phi(t) = \ln f(t) \quad (\text{c.g.f. of } F(x))$$

in a neighbourhood of the origin. By the assumption, we have

$$\sum_1^n \phi(a_k t) = \sum_1^n \phi(b_k t) \tag{5.1}$$

in a certain neighbourhood of the origin.

Since all the moments of $F(x)$ exists we may differentiate the last equation (5.1) any number of times . Let m be a positive integer. Differentiating $2m$ times, and setting $t = 0$, we obtain

$$\left[\sum_1^n (a_k)^{2m} - \sum_1^n (b_k)^{2m}\right] \phi^{2m}(0) = 0 \quad m = 1,2,\ldots .$$

Suppose $\sum_1^n (a_k)^{2m} = \sum_1^n (b_k)^{2m}$

holds for infinitely many times. But this can be true only if $|a_1|, |a_2|, \ldots, |a_n|$ are permutations of $|b_1|, \ldots, |b_n|$ which contradicts our assumption. Hence we must have

$$\phi^{2m}(0) = 0 \quad \text{for all} \quad m.$$

Since F(x) is a symmetric distribution, we have also

$$\phi^{2m-1}(0) \;=\; 0 \quad \text{for all } m.$$

Thus there exists an integer p such that

$\phi^{k}(0) = 0$ for $m > p$. This means the c.g.f. of F(x) is a polynomial of degree not exceeding p. Hence by Marcinkiewciz, theorem in Chapter I. F(x) is normal (possibly degenerate).

Linnik proves several important results related to the identically distributed linear statistics of i.i.d. r.v's. in his paper [51] 1953. He also generalized Marcinkiewicz's result in some sense by introducing the "determining function" $\sigma(z)$ which is an entire function of the complex variable Z. We present here three main results of Linnik taken from [48].

Theorem 6A. (Linnik). Let $X_1,\ldots,X_n$ be n i.i.d.r.v's with common d.f. F(x) . Consider two linear statistics $L_1 = \sum_1^n a_iX_i$ and $L_2 = \sum_1^n b_iX_i$ with the condition $\max(|a_1|,|a_2|,\ldots,|a_n|) \neq \max(|b_1|,|b_2|,\ldots,|b_n|)$. Let r be the greatest real zero of the determining function $\sigma(z) = |a_1|^z + |a_2|^z + \ldots + |a_n|^z - |b_1|^z - \ldots - |b_n|^z$. Suppose F(x) has moments up to order 2m, where $m = [\frac{r}{2} + 1]$ (greatest integer less than $\frac{r}{2} + 1$). Then F(x) is normal if L_1 and L_2 are identically distributed.

In the same paper, Linnik indicates some of the modifications regarding the condition $\max(a_1, \ldots, a_n) \neq \max(b_1, \ldots, b_n)$ without giving a detailed proof the following result.

Theorem 6B. Let $X_1, X_2, \ldots, X_n$ be n i.i.d. r.v.'s with common d.f. F(x). Let $L_1 = \sum_1^n a_i X_i$ and $L_2 = \sum_1^n b_i X_i$ be two linear statistics of $X_1, X_2, \ldots, X_n$. Let r be the exact upper bound of the real parts of the zeros of $\sigma(z)$ and $m = [\frac{r}{2} + 1]$. Suppose F(x) has a finite moment of order 2m. Then F(x) is normal (possibly degenerate) if L_1 and L_2 are identically distributed.

It should be remarked here if $\sigma(z) = 0$, then L_1 and L_2 are identically distributed for any arbitrary d.f. F(x), and if $\sigma(z) \neq 0$ and F(x) has moments of all orders, the conditions of Theorem 6.B are therefore satisfied. Thus Theorem 6.B contains Theorem 4 for the case of linear forms in finitely many variables. The preceding result gives a necessary condition for the normality of the common d.f. of the components in two linear statistics of finitely many i.i. d. r.v.s. Linnik also obtained a necessary and sufficient condition for a population to be normal and two linear statistics to be identically distributed in his paper [51].

Theorem 7. (Linnik) Let $X_1, X_2, \ldots, X_n$ be n i.i.d.r.v.'s with common d.f. F(x). Consider two linear statistics $L_1 = \sum_1^n a_k X_k$, $L_2 = \sum_1^n b_k X_k$ with $\max(|a_1|, |a_2|, \ldots, |a_n|) \neq \max(|b_1|, |b_2|, \ldots, |b_n|)$. Then the following two assertions

(A) $F(x)$ is a normal distribution

(B) L_1 and L_2 are identically distributed, are equivalent iff the the following five conditionsare satisfied

(i) $a_1 + a_2 + \ldots + a_n = b_1 + b_2 + \ldots + b_n$,

(ii) $\sigma(2) = 0$

(iii) all zeros of $\sigma(2)$ which are integers and are divisible by 4 are simple roots.

(iv) all positive roots of $\sigma(2)$ which are even integers of the form $4k + 2$(k integer) have a multiplicty not exceeding 2. If there exists such a double root, then it is unique and is the greatest positive root of $\sigma(2)$.

(v) the determining function $\sigma(2)$ can have at most one odd integer, positive, real root. If such a root exists, then it is simple and $[\frac{r}{2}]$ is odd.

2.3. Student distributions and normal distributions

It is well-known that the student distribution is closely related to the normal distribution and that its applications are always found in statistical literature. Let $X_o, X_1, \ldots, X_n$ $(n \geq 1)$ be i.i.d. normal r.v.'s with mean zero. It is known that the r.v.'s

$$Y_1 = \frac{X_1\sqrt{1}}{|X_o|}, \quad Y_2 = \frac{X_2\sqrt{2}}{\sqrt{X_o^2 + X_1^2}}, \ldots, \quad Y_n = \frac{X_n\sqrt{n}}{\sqrt{X_o^2 + \ldots + X_{n-1}^2}},$$

are r.v.'s distributed according to Student's law with $1,2,\ldots,n$ degrees of freedom respectively, and for $n \geq 2$ by carrying out the transformation techniques, it can be shown that they are independent. It is natural to ask whether the converse of the statement holds. More precisely, does this property characterize the normal distribution uniquely? This problem was investigated by Mauldon [74] (1956) who showed that the answer is negative for $n = 1$ (i.e., only two r.v.'s X_o, X_1). Recently, I.Kotlarski [37] (1966) successfully proved that the answer is in the affirmative for $n \geq 2$ under some conditions. The following theorem is due to Kotlarski (1966).

Theorem 8. (Kotlarski) Let $X_o, X_1, \ldots, X_n$ be $n+1$ $(n \geq 2)$ independent r.v.'s satisfying the conditions that $p(X_k = 0) = 0$ $(k = 0,1,2,\ldots,n)$, and each r.v.'s $X_o,\ldots,X_n$ has a symmetric distribution about the origin. Then the necessary and sufficient condition for X_k to be identically normally distributed with mean zero and common standard deviation σ is that $Y_1, Y_2, \ldots, Y_n$, where Y_k, $k = 1,2,\ldots,n$ are defined as above, are independently distributed according to student's law with $1,2,\ldots,n$ degrees of freedom respectively.

It is well known that for continuous random variables a one-to-one transformation, say $y = h(x)$, with domain S and range space T transforms the probability density $f(x)$ of a continuous r. v. X, say, to the probability density given by

$$g(y) = f(h^{-1}(y))\left|\frac{\partial x}{\partial y}\right| \text{ where } \left|\frac{\partial x}{\partial y}\right| \text{ is the Jacobian of the}$$

transformation. But this inverse result is seldom recognized and seems to have escaped attention. That is, if a r.v. Y has probability density $g(y)$ $(y \in T)$, then $x = h^{-1}(y)$ has probability density $f(x)$, $(x \in S)$ provided the transformation involved is one-to-one. This result is readily seen by applying the transformation techniques usually employed in Statistics in finding out the probability density of a continuous r.v. which is transformed.

The necessity of the theorem can be established easily by means of the transformation techniques. We need only to prove the sufficiency. Since the random variables X_k in Theorem 8 are symmetrical about the origin, their distributions are uniquely characterized by the distribution of U_k defined by

$$U_k = X_k^2 \qquad k = 0,1,2,\ldots,n \tag{8.1}$$

and the characterization is one-to-one since the X_k $(k = 0,1,2,\ldots,n)$ are independent. If X_k $(k = 0,1,2,\ldots,n)$ are normal with mean zero and common standard deviation σ, then the distribution of $U_k (k = 0,1,2,\ldots,n)$ is given by the common density

$$f(u) = \begin{cases} 0 & u \leq 0 \\ \dfrac{1}{\sigma\sqrt{2\pi u}} \exp\left(-\dfrac{u}{2\sigma^2}\right) & u > 0 \end{cases} \tag{8.2}$$

The same arguments apply to

$$V_k = \frac{Y_k^2}{K}, \quad k = 1,2,\ldots, n.$$

When Y_k, $k = 1,2,\ldots, n$ are distributed according to Student's law with k degrees of freedom respectively, then $V_K(k = 1,2,\ldots,n)$ are distributed according to the densities

$$g_k(v) = \begin{cases} 0 & v \leq 0 \\ \dfrac{1}{\beta(\frac{1}{2}, \frac{1}{2} k) v^{1/2}(1+v)^{1/2(k+1)}} & v > 0 \end{cases} \tag{8.3}$$

known as the beta-distribution of the second kind. Hence if we can prove that if each of the r.v.'s $V_1 = \dfrac{U_1}{U_o}$, $V_2 = \dfrac{U_2}{U_o+U_1}$, $V_3 = \dfrac{U_3}{U_o+U_1+U_2}, \ldots, V_n = \dfrac{U_n}{U_o+U_1+\ldots+U_{n-1}}$ is distributed according to (8.3), it implies that each U_k, $k = 1,2,\ldots,n$ is distributed according to the density (8.2), where of course $U_o, U_1, \ldots, U_n$ are $n+1$ independent positive r.v.'s $(n \geq 2)$ then the sufficiency follows.

To do this, we need the following lemmas.

Lemma 8.1. Let $U_o, U_1, \ldots, U_n$ be $n+1$ $(n \geq 2)$ independent positive r.v.'s. Let $Z_j = \dfrac{U_1}{U_o}$, $Z_2 = \dfrac{U_2}{U_o}, \ldots, Z_n = \dfrac{U_n}{U_o}$. If the joint ch.f. of $(\ell n\, z_1, \ell n\, z_2, \ldots, \ell n\, z_n)$ does not vanish, then the joint distribution of $(z_1, \ldots, z_n)$ determines all the distributions of $U_o, U_1, \ldots, U_n$ up to a change of the scale.

Proof. Let $f_k(t)$ be the ch.f. of $\ell n\ U_k$, $k = 0,1,2,\ldots,n$ and let

$$\psi\ (t_1,t_2,\ldots,t_n) = E[\ \exp\ \ i(t_1\ \ell n\ z_1 + \ldots + t_n\ \ell n\ z_n)\ ],$$

$$t_i,\ i = 1,2,\ldots,n$$

are real. By independence of $U_o,U_1,\ldots,U_n$, we have

$$\psi(t_1,\ldots,t_n) = f_1(t_1)f_2(t_2)\ldots f_n(t_n)f_o(-t_1-t_2-,\ldots-t_n).$$

By hypothesis, $\psi(t_1,\ldots,t_n)$ is nonvanishing, we conclude that so are $f_k(t_k)$, $k = 0,1,2,\ldots,n$.

Now if U_o', $U_1',\ldots,U_k'$ and Z_1', $Z_2',\ldots,Z_n'$ also satisfying the conditions of the lemma, and if $\ell n\ U_k'$ has ch.f. $f_k'(t)$ and $(\ell n\ z_1', \ell n\ z_2',\ldots \ell n\ z_n')$ has ch.f. $\psi'(t_1,t_2,\ldots,t_n)$, then by previous argument, we have

$$\psi'(t_1,\ldots,t_n) = f_1'(t_1)f_2'(t_2),\ldots,f_n'(t_n)\ f_o'(-t_1-t_2-\ldots-t_n)$$

$$t_k,\ \ k = 1,2,\ldots, \text{ are real.}$$

If $(Z_1,\ldots,Z_n)$ and $(Z_1',Z_2',\ldots,Z_n')$ have the same joint distribution, then

$$f_1'(t_1)f_2'(t_2),\ldots,f_n'(t_n)f_o'(-t_1-t_2-\ldots-t_n)$$

$$= f_1(t_1)\ldots f_n(t_n)f_o(-t_1-t_2-\ldots-t_n) \tag{8.4}$$

Let

$$f_k'(t) = f_k(t)\ p_k(t),\ \ k = 0,1,2,\ldots,n \tag{8.5}$$

where $p_k(t)$ are complex-valued functions, continuous on the whole line non-vanishing and satisfying the conditions $p_k(0) = 1$. Substituting (8.5) into (8.4), we get

$$p_1(t_1)\ p_2(t_2),\ldots,p_n(t_n)p_o(-t_1-t_2-\ldots-t_n) = 1 \qquad (8.6)$$

Let $p_o(-t) = p(t)$ where t is real, and put $t_k = t$, $t_j = 0$ for $j \neq k$. We obtain from (8.6)

$$p_1(t) = p_2(t) = \ldots = p_n(t) = \frac{1}{p(t)} \quad . \qquad (8.7)$$

Substituting (8.7) into (8.6), we obtain

$$p(t_1+t_2+\ldots+t_n) = p(t_1)p(t_2)\ \ldots\ p(t_n) \ .$$

This is a Cauchy equation. The only complex solution p(t) continuous on the whole real line, nonvanishing and satisfying the condition p(0) = 1 is the exponential function $p(0) = e^{ct}$ (t is real and c is a complex number). Hence

$$p_o(t) = p_1(t) = \ldots = p_n(t) = e^{-ct}$$

and hence $f'_k(t) = f_k(t)\ e^{-ct}$.

But we have $f(-t) = f(t)$, c must be pure imaginary

i.e., $c = -ia$. This implies

$$f'_k(t) = f_k(t)\ e^{iat} \ .$$

The proof is complete.

Lemma 8.2. Let $U_o, U_1, \ldots, U_n$ be n+1 $(n \geq 2)$ independent positive r.v.'s. Let $z_k = \frac{U_k}{U_o}$, $k = 1,2,\ldots,n$. The necessary and sufficient condition U_k to be identically distributed according to the density

$$f(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ \frac{1}{\sigma\sqrt{2\pi u}} \exp\left(-\frac{u}{2\sigma^2}\right) & \text{if } u > 0 \end{cases}$$

is that the n-dimensional r.v. $(z_1,\ldots,z_n)$ is distributed as the density

$$h(z_1,\ldots,z_n) = \begin{cases} \frac{\Gamma(\frac{1}{2}n+\frac{1}{2})}{\pi^{(\frac{1}{2}n+\frac{1}{2})}} \frac{1}{\sqrt{z_1 z_2,\ldots,z_n}} \frac{1}{(1+z_1+z_2+\ldots+z_n)^{n/2+1/2}} & z_k > 0 \\ 0 \qquad \text{otherwise,} & k = 1,2,\ldots,n. \end{cases} \tag{8.8}$$

Proof. A direct calculation shows that the ch.f. $f_k(t)$ of $\ell n\, u_k$, $k = 0,1,2,\ldots,n$ and the ch.f. $\psi(t_1,\ldots,t_n)$ of $(\ell n\, z_1,\ldots, \ell n\, z_k)$ where $Z_1,\ldots,Z_k$ is distributed as (8.8) are given respectively by

$$f_k(t) = \frac{(2\sigma^2)^{it}}{\sqrt{\pi}} \Gamma(\tfrac{1}{2} + it), \quad t \text{ real}$$

$$\psi(t_1,t_2,\ldots,t_n) = \frac{1}{\pi^{(\frac{1}{2}n+\frac{1}{2})}} \Gamma(\tfrac{1}{2} + it_1)\, \Gamma(\tfrac{1}{2} + it_2) \ldots$$

$$\ldots \Gamma(\tfrac{1}{2} + it_n)\, \Gamma(\tfrac{1}{2} - it - \ldots - it_n). \tag{8.9}$$

where t_k, $k = 1,2,\ldots,n$ are real. It is obvious that $\psi(t_1,\ldots,t_n)$ is nonvanishing for all real $t_1,t_2,\ldots,t_n$, and

$$\psi(t_1,\ldots,t_n) = f_1(t_1)f_2(t_2)\ldots f_n(t_n)f_o(-t_1-t_2-\ldots-t_n) \tag{8.10}$$

t_k, $k = 1,2,\ldots,n$ are real. By substituting $f_k(t) = \frac{(2\sigma^2)^{it}}{\sqrt{\pi}} \Gamma(\frac{1}{2} + it)$ in (8.10) and by applying Lemma 8.1 the lemma is established.

We are now in a position to prove the sufficiency of Theorem 8.

It follows from Lemma 8.2 that if $(Z_1,\ldots,Z_n)$ distributed according to (8.8), the $U_k(k = 0,1,2,\ldots, n)$ are identically distributed according to (8.2). Now it suffices to show that if $V_1,V_2,\ldots,V_n$ are independently distributed as (8.3), then $(Z_1,\ldots,Z_n)$ is distributed as (8.8). Since $Z_k = \frac{U_k}{U_o}$, $k = 1,2,\ldots,n$, we see that

$$V_1 = Z_1, \quad V_2 = \frac{Z_2}{1+Z_1}, \quad V_3 = \frac{Z_3}{1+Z_1 + Z_2}, \ldots, V_n = \frac{Z_n}{1+Z_1+\ldots+ Z_{n-1}}, \tag{8.11}$$

The Jacobian of this transformation is

$$J = \frac{\partial(v_1,\ldots,v_n)}{\partial(z_1,\ldots,z_n)} = \frac{1}{(1+z_1)(1+z_1+z_2),\ldots,(1+z_1+\ldots+z_{n-1})} \tag{8.12}$$

The density ot $(V_1,\ldots,V_n)$ is given by

$$g(v_1,\ldots,v_n) = g_1(v_1),\ldots,g(v_n)$$

$$= \prod_{k=1}^{n} \frac{1}{B(\frac{1}{2}, \frac{1}{2} k)\, v_k^{1/2}(1+v_k)^{k+1/2}} \tag{8.13}$$

all $v_k > 0$.

Hence the density of $(z_1,\dots,z_n)$ is given by $h(z_1,\dots,z_n)$

$$= \prod_{k=1}^{n} \frac{}{B(\frac{1}{2},\frac{1}{2}k)\left(\frac{z_k}{1+z_1+z_2+\dots+z_{k-1}}\right)^{1/2}\left(1+\frac{z_k}{1+z_1+z_2+\dots+z_{k-1}}\right)^{(k+1)/2}} |J|$$

where $z_o = 0$

$$= \prod_{k=1}^{n} \frac{(1+z_1+\dots+z_{k-1})^{1/2}(1+z_1+\dots+z_{k-1})^{k+1/2}}{B(\frac{1}{2},\frac{1}{2}k)\; z_k^{1/2}(1+z_1+\dots+z_k)^{k+1/2}} |J|$$

$$= \frac{\Gamma(\frac{1}{2}+\frac{1}{2})\Gamma(\frac{1}{2}+1)\dots\Gamma(\frac{1}{2}+\frac{n}{2})\; 1\cdot(1+z_1)^{(2+2)/2}(1+z_1+z_2)^{(3+2)/2}\dots(1+z_1\dots+z_{n-1})^{\frac{1}{2}(n}}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})\Gamma(1)\dots\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})\; z_1^{1/2}\dots z_n^{1/2}(1+z_1)^{1+1/2}\dots(1+z_1+\dots+z_n)^{(n+1)/2}} |J$$

$$= \frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\pi^{(\frac{n}{2}+\frac{1}{2})}\sqrt{z_1 z_2 \dots z_n}\,(1+z_1+\dots+z_n)^{\frac{n}{2}+\frac{1}{2}}} \qquad z_k > 0,\ k = 1,2,\dots,n\ .$$

This completes the proof.

CHAPTER III

Characterizations of Normality by Independence of suitable Statistics

3.1. Independence of linear statistics.

The appearance of Geary's paper in 1936 proving that the stochastic independence of the sample mean and sample variance implies the Normality of the population under consideration suggests a general method of finding statistical populations by using this property. Geary proved the result under the superflous assumption on the existence of all moments. Soon after Geary's theorem, a series of papers have appeared, generalizing Geary's theorem in various directions. The main generalizations are given by Lukacs [62] Laha [38], Kawata and Sakamoto [33] and Zinger [99]. Lukacs (1942) proved Geary's theorems assuming only the existence of the second moments. Later Kawata and Sakamoto (1949) and Zinger (1951) showed that Geary's theorem is true without the assumption on the existence of moments. There are a number of papers on generalization of Geary's theorem by constructing other statistics instead of the sample mean and sample variance such as Laha [38] and Geisser [25]. Recently the property of independence of statistics is replaced by the regression properties so as to drop the condition of independence to a weaker condition of regression. We shall see how the independence of statistics can be replaced by regression properties in the next chapter. In this chapter, we shall deal with the characterizations through independence of suitable statistics, for instance linear and linear statistics, linear and quadratic statistics, and linear and polynomial statistics. We first consider the independence of two linear statistics which has a rather interesting history and involves a good number of authors (Kac [29], Bernstein [7], Gnedenko [27], Darmois [17] and Skitovich [88]). The problem was discussed in its full generality independently by Darmois (1953) and Skitovich (1954). The following theorem was established.

Theorem 1. (Darmois-Skitovich). Let $L_1 = \sum_1^n a_i X_i$ and $L_2 = \sum_1^n b_i X_i$ be two independent linear statistics, where $X_1, X_2, \ldots, X_n$ are n independent (but not necessarily identically distributed) r.v.'s then each r.v. X_j, $(j = 1,2,\ldots,n)$ with nonzero coefficients in both forms is normally distributed.

As a detailed proof of the theorem is available (see [48] pp. 75-78), we shall only outline the proof. We first note that if any r.v. has zero coefficients in one of the two forms. Then the corresponding r.v. can be arbitary. In view of this, we can omit all those r.v.'s which have zero coefficients in one of the two forms and are left with considering the r.v.'s having nonzero coefficients in both forms. Further, it is obvious that the resulting forms so obtained are also independent. Hence, there is no loss of generality in assuming that

$$a_j b_j \neq 0, \quad (j = 1,2,\ldots,n) \quad \text{and} \quad a_1 = b_1 = 1.$$

We may also group all those r.v.'s such that the ratio of their respective coefficients is equal to some fixed constant. Let the coefficients of the linear forms L_1 and L_2 satisfy the following conditions

$$a_1 = b_1 = 1$$

$$a_j b_j \neq 0, \quad j = 1,2,\ldots,n \qquad (1.1)$$

$$a_j b_k - a_k b_j \neq 0 \quad j \neq k, \quad j,k,\ldots,n \,.$$

Now, taking into account of the independence of L_1 and L_2 and writing it down in terms of the characteristic functions, $f_1, f_2, \ldots, f_n$, of $X_1, \ldots, X_n$ we obtain a relation

$$f(t,s) = \prod_{j=1}^{n} f_j(a_j t + b_j s) = \prod_{j=1}^{n} f_j(a_j t) f_j(b_j s) \qquad (1.2)$$

By continuity of ch.f. at the origin and $f_j(0) = 1$, $j = 1,2,\ldots,n$ there exist a neighbourhood of the origin such that every factor of (1.2) is nonzero, and by taking the logarithm of both sides of (1.2), we arrive at the following equation

$$\sum_{j=1}^{n} \phi_j(a_j t + b_j s) = A_o(t) + B_o(s) \qquad (1.3)$$

where

$$A_o(t) = \sum_{j=1}^{n} \phi_j(a_j t), \quad B_o(t) = \sum_{j=1}^{n} \phi_j(b_j t)$$

In view of (1.1), it is possible to select h_1 and k_1 in such a way that $h_1 + k_1 = s_1$ and $a_n h_1 + b_n k_1 = 0$ hold. Choose a real number s_1 so small that (1.3) is satisfied if t is replaced by $t+h_1$ and s by $s+k_1$. Substituting the quantities and using the method of finite difference (see [48] pp.77), we can eliminate the function ϕ_n by this procedure. Proceed in the same manner, we can finally eliminate the functions $\phi_2,\ldots,\phi_n$, and obtain a difference equation of the function $\phi_1(t)$ of order n. It then follows that $\phi_1(t)$ is a polynomial of degree not exceeding n, and hence by Marcinkiewicz's theorem and the properties of analytic functions, $f_1(t)$ is the ch.f. of a normal distribution. Similarly, $X_2,\ldots,X_n$ can be shown to be normal.

Having established "Darmois-Skitovich's" theorem, we can deduce a good number results from it.

Corollary 1.1. (King and Lukacs 1954) Let $X_1,X_2,\ldots,X_n$ be n-independently (but not necessarily identically) distributed r.v.'s and assume that the n^{th} moment of each X_i $(i = 1,2,\ldots,n)$ exists. The necessary and sufficient conditions for the existence of two independent linear statistics

$$L_1 = \sum_{k=1}^{n} a_k X_k \quad \text{and} \quad L_2 = \sum_{k=1}^{n} b_k X_k \qquad \text{are}$$

(i) Each r.v. with nonzero coefficient in both forms is normally distributed.

(ii) $\sum_{k=1}^{n} a_k b_k \sigma_k^2 = 0$, where $\sigma_k^2 = \text{var}(X_k)$, $k = 1,2,\ldots,n$.

For $n = 2$, and $a_1 = b_1 = a_2 = 1$, $b_2 = 1$, this reduces to a theorem of Bernstein.

The necessity follows immediately from Theorem 1. To prove the sufficiency, assume that (i) and (ii) hold, then it can be shown that L_1

and L_2 are uncorrelated. Since L_1 and L_2 are normal and uncorrelated, they are independent.

Corollary 1.2. (Kac) Let X_1 and X_2 be two independent r.v.'s. If for every real number α

$$Y_1 = X_1\cos\alpha + X_2\sin\alpha, \quad Y_2 = X_1\sin\alpha - X_2\cos\alpha$$

are independent, then X_1 and X_2 are normal.

Instead of considering two independent linear forms, one may consider m linear forms of the independent r.v.'s $X_1, X_2, \ldots, X_n$

$$L_k = \sum_{j=1}^{n} a_j X_j, \quad k = 1,2,\ldots,m, \quad m \leq n.$$

By Theorem 1, one can easily see that all r.v.'s $X_1,\ldots,X_n$ are normally distributed if (i) the linear forms $L_1, L_2, \ldots, L_n$ are mutually independent, (ii) each of the column of the matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

contains at least two nonzero constants. Also, using Cramer-Wold's theorem one can formulate an analogous result of the above result in multivariate case, i.e., if there exist linear forms

$$L_1 = \sum_{i=1}^{n} a_{k_i} X_i, \quad k = 1,2,\ldots,m, \quad 2 \leq m \leq n,$$

of the independent p-dimensional $(p \geq 2)$ random vectors (but not necessarily identically distributed) $X_1,\ldots,X_n$ such that they are independent, then each random vector X_j, $j = 1,2,\ldots,n$ has a p-variate normal distribution if each of the column of the matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{mn} & \cdots & a_{mn} \end{pmatrix}$$

contains at least two nonzero elements.

As seen in the previous discussion, it is worthwhile to investigate the above result in multivariate case for $m = 2$ and with $a_{i1},\ldots,a_{in}$, $i = 1,2$.. replaced by $p \times p$ scalar matrices. However, when $a_{i1},\ldots,a_{in}$, $i = 1,2$, are replaced by $p \times p$ scalar matrices $A_{i1},\ldots,A_{in}$, $i = 1,2$. the reduction to univariate case by using Cramer-Wold's theorem no longer holds. In univariate case, Darmois-Skitovich's theorem tells us that the distribution of the r.v. with zero coefficient can be arbitrary. The same is also true in the matrix case if one of the matrix A_{ik}, $i = 1,2$ for some $k = 1,2,\ldots,n$ is zero (null). However, if a matrix A_{ik} for some $k = 1,2,\ldots,n$ (say) is singular but not null, then some linear combinations of elements of the corresponding random vector X_k are normally distributed, but the distribution of X_k is partly arbitrary. An example is constructed in [26] to illustrate this fact. This means generalization cannot be made by using matrices other than nonsingular matrices. The following result is due to Ghurge and Olkin (1962).

Theorem 2. (Ghurye and Olkin) Let $X_1,\ldots,X_n$ be n independent p-dimensional (column) random vectors, and let $A_1,\ldots,A_n$, $B_1,\ldots,B_n$ be nonsingular p x p matrices. If ΣA_iX_i and ΣB_iX_i are independent, then X_i, $i = 1,2,\ldots,n$ are p-variate normal.

Since the proof of this theorem is every similar to theorem 1 and is lengtly, we shall only sketch the proof.

First considering the ch.f. of $(\sum_1^n A_iX_i, \sum_1^n B_iX_i)$ and taking into account the independence of $\sum_1^n A_iX_i$ and $\sum_1^n B_iX_i$ we obtain a relation of the form

$$\prod_{j=1}^{n} \varphi_j(T'A_j + U'B_j) = F(T')\ G(U') \quad \text{where}$$

$\varphi_j(T) = E \exp [iT'X_j]$ and T,U are p-dimensional (column) scalar vectors. It can be shown that φ_j $(j = 1,2,\ldots,n)$ have no zeros so that $\log \varphi_j(T')$ is defined. Also, by similar arguments of Theorem 1, we can show $\sum_1^n \log \varphi_j(T')$ is a polynomial in T'. By letting T = kV where k is real, and V is a fixed p-dimensional (column) scalar vector, and using the univariate theorem of Marcinkiewicz, it follows that $\sum_1^n \log \varphi_j(kV')$ is a quadratic polynomial in k for each fixed vector V. This implies that $\sum_1^n \log \varphi_j(T')$ is a quadratic polynomial in the vector T' which means that X_i, $i = 1,2,\ldots,n$ are p-variate normal.

3.2. Independence of linear and quadratic Statistics.

As mentioned in the beginning of this chapter, the independence of the sample mean and sample variance implies the Normality of the population. Instead of sample variance, one may use some other quadratic forms. We present here a result due to Rao [84] and obtain the results of Lukacs (1942) and Geisser (1956) as corollaries.

Theorem 3. (Rao, 1958) Let $X_1,\ldots,X_n$ be a sample from X with d.f. $F(x)$. Assume that $E(X) = \mu$ and $\text{var}(X) = \sigma^2$ exist. Let

$$Q = \left(\sum_{k=1}^{m} \sum_{j=1}^{n} a_{kj}^2\right)^{-1} \sum_{k=1}^{n} (a_{k1}X_1 + \ldots + a_{kn}X_n)^2, \quad m \geq 1$$

where $\sum_{j=1}^{n} a_{kj} = 0$ for $k = 1,2,..,m$. The necessary and sufficient condition that F be normal is that $\bar{X} = \sum_{i=1}^{n} X_i/n$ and Q are independent.

Proof. $$E(Q) = \left(\sum_{k=1}^{m} \sum_{j=1}^{n} a_{kj}^2\right)^{-1} \left\{\sum_{k=1}^{m} \sum_{j=1}^{n} a_{kj}^2\, E(X_j^2) + \sum_{k=1}^{m} \sum_{j\neq i}^{n} a_{kj}a_{ki}E(X_jX_i)\right\}$$

$$= \left(\sum_{k=1}^{m} \sum_{j=1}^{n} a_{kj}^2\right)^{-1} \left\{\sum_{k=1}^{m} \sum_{j=1}^{n} a_{kj}^2\, E(X_j^2) - \sum_{k=1}^{m} \sum_{j=1}^{n} a_{kj}^2\, \mu^2\right\}$$

$$= \sigma^2$$

The joint ch.f. of $\bar{X}$ and Q is given by

$$f(t_1,t_2) = E(e^{it_1\bar{X}+it_2Q}) = E(e^{it_1\bar{X}})\, E(e^{it_2Q}) = \varphi_1(t_1)\varphi_2(t_2)$$

where $\varphi_1(t_1) = E(e^{it_1\bar{X}})$ and $\varphi_2(t_2) = E(e^{it_2Q})$.

It follows that

$$\frac{\partial}{\partial t_2} f(t_1,t_2)\Big|_{t_2=0} = \varphi_1(t_1) \frac{\partial}{\partial t_2} \varphi_2(t_2)\Big|_{t_2=0} \tag{3.1}$$

But $\varphi_1(t) = f({}^{t_1}/_{n})^n$ where $f(t) = E(e^{itX})$,

and so

$$\frac{\partial}{\partial t_2} f(t_1,t_2)\Big|_{t_2=0} = i\,(\sum_k \sum_j a_{kj}^2)^{-1}\Big\{ \sum_k \sum_j a_{kj}^2 [f({}^{t_1}/n)]^{n-1} E(X^2 e^{it_1X/n})$$

$$+ (\sum_k \sum_{j\neq i} a_{kj}\, a_{ki})\, f({}^{t_1}/n)^{n-2} [E(X\, e^{it_1X/n})]^2 \Big\} \tag{3.2}$$

Using (3.2) and $\frac{\partial}{\partial t_2} \varphi_2(t_2)\Big|_{t_2=0} = i\sigma^2$, equation (3.1) reduces to

$$- f(t_1) \frac{d^2 f(t_1)}{dt_1^2} + [\frac{df(t_1)}{dt_1}]^2 = \sigma^2 [f(t_1)]^2 \tag{3.3}$$

The solution of (3.3) is the ch.f. of the normal distribution with mean μ and variance σ^2.

To prove the necessity, we prove a more general result as follows:

Let $X_1,\ldots,X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. If the statistic $S = S(X_1,\ldots,X_n)$ is translation-invariant (i.e. $S(X_1,\ldots,X_n) = S(X_1 + a_1,\ldots,X_n + a)$ for any real a), then S is independent of the sample mean $X = \sum_1^n X_i/n$.

Let $f(t_1,t_2)$ denote the ch.f.of the random vector $(\bar{X},S)$. Since $X_1,\ldots,X_n$ are normal with mean μ and variance σ^2, we have

$$f(t_1,t_2) = \left(\frac{1}{\sigma\ 2\pi}\right)^n \int_{R_n} \exp\ i\ \frac{t_1}{n} \sum_{j=1}^{n} X_j + iS_2(X_1,\ldots,X_n)$$

$$- \frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2\ dx_1\ dx_2 \ldots dx_n$$

or

$$f(t_1,t_2) = g(t_1,t_2)\ \exp\left\{ it_1\mu - \frac{\sigma^2}{2n}\ t_1^2 \right\} \tag{3.4}$$

where

$$g(t_1,t_2) = \left(\frac{1}{\sigma\ 2\pi}\right)^n \int_{R_n} \exp\left\{- \frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu - \frac{i\sigma^2}{n} t_1)^2 + \right.$$

$$\left. + it_2\ S(x_1,\ldots,x_n) \right\} dx_1 \ldots dx_n$$

The function $g(t_1,t_2)$ is an entire function of t_1 . Letting $z_j = x_j - \mu - \sigma^2 y/n$, where $-t_1 = iy$ (y real) and taking into consideration that S is translation-invariant, we obtain

$$g(t_1,t_2) = \left(\frac{1}{\sigma\ 2\pi}\right)^n \int \exp\left\{- \frac{1}{2\sigma^2} \sum_{j=1}^{n} z_j^2 + it_2 S(z_1,\ldots,z_n)\right\} dz_1 \ldots dz_n$$

That is, $g(t_1,t_2)$ is a constant for purely imaginary values of t_1. Since it is an entire function of t_1, it must be independent of t_1. We thus obtain

$$g_1(t_2) = g(t_1,t_2)$$

From (3.4), we get

$$f(t_1,t_2) = g_1(t_2) \exp [i\mu t_1 - \frac{\sigma^2 t^2}{2n}]$$

This shows that $\bar{X}$ and S are independent. Finally, we conclude by the above result that the necessity of the theorem holds since Q is invariant under a translation. This completes our proof.

Corollary 3.1. (Lukacs 1942). Let $X_1,\ldots,X_n$ be a random sample from X with d.f. F(x). Assume that E(X) and var(X) exist. Then F(x) is normal iff $\bar{X}$ and $S^2 = \frac{1}{n-1} \sum_{n-1}^{n} (X_i-\bar{X})^2$ are independent

To obtain the result from Theorem 3, let

$$a_{kj} = \begin{cases} 1-\frac{1}{n} & \text{for } k = j \\ -\frac{1}{n} & \text{for } k \neq j \text{ and } m = n \end{cases} .$$

Corollary 3.2. (Geisser 1956) Let $X_1,\ldots,X_n$ be a random sample from X with d.f. F(x). Assume that $E(X) = \mu$ and $\text{var}(X) = \sigma^2$ exist.

Then $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $\delta_p^2 = \sum_{j=1}^{n-p} (X_{j+p} - X_j)^2$, $p = 1,2,\ldots,n-1$

are independent iff F is normal.

To obtain the result from Theorem 3, let

$$a_{kj} = \begin{cases} 1 & \text{for } j = k+p \\ -1 & \text{for } j = k \\ 0 & \text{for other values of } j \text{ and } m = n-p . \end{cases}$$

It should be noted here that a further generalization of the previous theorem has been established by Shimizu (1961).

Following the line of the above approach, one can easily establish the following theorem due to Laha (1956) concerning the independence of the mean and a homogeneous quadratic statistic.

Theorem 4. (Laha 1956) Let $X_1,\ldots,X_n$ be a random sample from a population with d.f. $F(x)$ having finite variance σ^2. Let $Q = \sum_{j=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j$ be a quadratic statistic with the coefficients satisfying the conditions:

(i) $\sum_{i=1}^{n} a_{ii} \neq 0$, $\sum_{j=1}^{n} a_{ij} = 0$ for $i = 1,2,\ldots n$.

Then $\bar{X}$ and Q are independent iff $F(x)$ is normal.

We shall only outline the proof. Since $\overline{X}$ and Q are independent, in terms of ch.f. one can easily obtain a relation of the forms

$$\frac{d^2}{dt^2}\left(\log f(t)\right) = -\sigma^2$$

in a certain neighbourhood of the origin. We conclude from the properties of analytic functions that the solution of the above equation is the ch.f of a normal distribution.

To prove the sufficiency we need only show that $\overline{X}$ and Q are uncorrelated of order (2,2). That is,

$$E(\overline{X}^i Q^j) = E(\overline{X}^i)\, E(Q^j) \quad i = 1,2, \quad j = 1,2, \quad \text{(see pp.72 48).}$$

Direct verifications show that the above relations hold, which completes the proof.

It should be remarked here that the assumption of the existence of finite variance is in fact, superflous in all the results just considered due to a theorem of Zinger (1958) concerning admissible polynomial statistics. A polynomial $p(x_1,\ldots,x_n)$ of degree m is said to be admissible if the coefficients of the terms x_j^m, $j = 1,2,\ldots,n$ are nonzero. The important result obtained by Zinger (1958) concerning admissible polynomial statistics is that if two admissible polynomial statistics $p_1(X_1,\ldots,X_n)$, $p_2(X_1,\ldots,X_n)$, where $X_1,\ldots,X_n$ are n independent (but not necessarily identically distributed) r.v.'s, are independent then each r.v. X_j, $j = 1,2,\ldots,n$ has finite moments of all order. The proof of this result can be found in [48].

3.3. Independence of polynomial Statistics

If the preceding sections, we have seen that the independence of two linear statistics and the independence of a linear and a quadratic statistic characterize Normality. In what follows, we shall consider the independence of two polynomial statistics, or more generally, the independence of certain functions of independent r.v.'s.

We first introduce a special class of polynomials. Let $p(x_1,\ldots,x_n) = \sum_{j_1+j_2+\ldots+j_n \leq r} A_{j_1 j_2,\ldots,j_n} x_1^{j_1} \ldots x_n^{j_n}$ be a polynomial of degree r.

We may write $p(x_1,\ldots,x_n)$ as a sum of a homogeneous polynomial of degree r and another polynomial of degree less than r. i.e.

$$p(x_1,x_2,\ldots,x_n) = p_o(x_1,\ldots,x_n) + p_1(x_1,\ldots,x_n)$$

where $p_o(x_1,\ldots,x_n) = \sum_{j_1+\ldots+j_n = r} A_{j_1,\ldots,j_n} x_1^{j_1} \ldots x_n^{j_n}$ is a homogeneous polynomial of degree r, while $p_1(x_1,\ldots,x_n)$ is a polynomial of degree less than r. The polynomial $p(x_1,\ldots,x_n)$ is said to be non-singular if $p_o(x_1,\ldots,x_n)$ contains the r-th power of at least one variable and $\pi_o(k) \neq 0$ for all integers $k > 0$, where $\pi_o(k)$ is the polynomial obtained by replacing each positive power X_s^j by

$K^{(j)} = k(k-1) \ldots (k-j+1)$ in $p_o(x_1,\ldots,x_n)$. We shall see that the k-statistic of order q is a nonsingular polynomial. By a k-statistic of order q we mean a symmetric homogeneous polynomial stastistic of degree q such that its expectation is equal to the k^{th} cumulant of the population. It is interesting to note that the independence of the sample sum $X_1 +\ldots+X_n$ and a nonsingular polynomial statistic of degree r implies that the ch.f. of the distribution function under consideration is an entire function of finite order with no zeros in the complex plane (see pp. 96[48]). Hence according to Hadamard's factorization theorem, the ch.f. is of the form $f(z) = \exp[p_n(z)]$ and hence we conclude from Marcinkiewicz's theorem that $f(t)$ is the ch.f. of some normal distribution. This readily gives us a characterization of the normal distribution through the independence of sample sum and a nonsingular polynomial statistic. Thus we have

Theorem 5. Let $X_1,\ldots,X_n$ be a random sample from X with distribution function $F(x)$. Let $p(X_1,\ldots,X_n)$ be a nonsingular polynomial statistic. Then $F(x)$ is normal provided that $p(X_1,\ldots,X_n)$ and the sample sum $X_1 + X_2 +\ldots+ X_n$ are independent.

As a direct application of Theorem 5, we have the following four corollaries.

Corollary 5.1. Let $X_1,\ldots,X_n$ be a random sample from X with d.f. $F(x)$. Let $p(X_1,\ldots,X_n)$ be an admissible homogeneous polynomial statistic of degree r such that the expected value of $p(X_1,\ldots,X_n)$ is equal to r^{th} cumulant k_r. Then $F(x)$ is normal provided that $p(X_1,\ldots,X_n)$ and the sample sum are independent.

It can be shown that $p(X_1,\ldots,X_n)$ is a non-singular polynomial statistic. In virtue of Theorem 5, the corollary follows immediately. We note that in particular if $p(X_1,\ldots,X_n)$ is the k-statistic of order r, then conditions of Theorem 5 are automatically satisfied. That is to say the following holds

Corollary 5.2. (Basu and Laha 1954) Let $X_1,\ldots,X_n$ be a random sample from X with d.f. F(x). Let k_p be the k-statistic of order r. The F(x) is normal provided that k_p and the sample sum $X_1+\ldots+X_n$ are independent.

Corollary 5.3. Let $X_1,\ldots,X_n$ be a random sample from X with distribution function F(x). Let

$$A = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji}X_iX_j + \sum_{j=1}^{n} b_jX_j$$

such that

$$B_1 = \sum_{j=1}^{n} a_{jj} \neq 0 \quad \text{and} \quad B_2 = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji} = 0.$$

Then Q and the sample sum are independent iff

(i) F(x) is normal

(ii) $\beta_k = \sum_{j=1}^{n} a_{jk} = 0$ for $k = 1,2,\ldots,n$.

(iii) $\beta = \sum_{j=1}^{n} b_j = 0$.

To prove the necessity, one needs only to form the polynomial $\pi_o(k)$ corresponding to the polynomial Q. It is easy to find that

$$\pi_o(k) = \sum_{j=1}^{n} a_{jj}\, k(k-1) + \sum_{j=1}^{n} \sum_{i \neq j}^{n} a_{ji} kk$$

$$= B_2 k^2 - B_1 k .$$

This shows that Q is indeed a non-singular polynomial of degree two and hence F(x) is normal by Theorem 5.

The conditions (ii) and (iii), and the sufficiency of the corollary follows from a result proven by Laha (1956a). The result states that the necessary and sufficient condition that two real polynomial statistics of the second degree denoted by

$$p_1(X_1, \ldots, X_n) = X'AX + L'X \quad \text{and} \quad p_2(X_1, \ldots, X_n) = X'BX + M'X$$

are independent is that

(a) $AB = 0$, (b) $L'B = 0$, (c) $M'A = 0$, (d) $L'M = 0$.

Here $X = (X_1, \ldots, X_n)'$ $(X_1, \ldots, X_n$ is a random sample) is a random (column) vector, L', M' are column scalar vector and A, B are both $n \times n$ real symmetric matrices.

The following result is another application of Theorem 5.

Corollary 5.4 (Laha, Lukacs and Newman 1960). Let $X_1,\ldots,X_n$ be a random sample from X with d.f. F(x). Let p be a positive integer such that $(p-1)!$ is not divisible by $(n-1)$. Let $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $m_p = \sum_{i=1}^{n} (X_i - \bar{X})^p/n$ be the sample mean and sample central moment respectively. The F(x) is normal iff $\bar{X}$ and m_p are independent.

Proof. The necessity follows from a previous result that in a normal population any translation-invariant statistic is independent of the sample mean. To prove the sufficiency, we first note that the polynomial $\pi_o(k)$ formed by substituting each power X_s^j by $k^{(j)} = k(k-1)\ldots(k-j+1)$ has no non-zero integer roots if $(p-1)!$ is not divisible by $(n-1)$ (see [46]). This means that $\pi_o(k)$ corresponding to m_p does not vanish for any positive integer k, that is, m_p is a non-singular polynomial statistic. The sufficiency follows immediately with the application of Theorem 5.

CHAPTER IV

Characterizations of Normality by means of Regression Properties

4.1. Constant regression and polynomial regression.

In the previous chapter, we have seen that several properties of certain statistics have been used to characterize the normal distribution, namely, the distribution of a certain statistic , and the independence of linear and polynomial statistics. If we examine carefully the proof of Theorem 3 in Chapter III, we see that the assumption that the independence of sample mean and the quadratic statistic Q can be replaced by a condition as $E(Qe^{it\overline{X}}) = E(Q)E(e^{it\overline{X}})$ which also enables us to arrive at the same differential equation. It is well known that the independence of Q and $\overline{X}$ implies $E(Qe^{it\overline{X}}) = E(Q)E(e^{it\overline{X}})$ but not conversely. Hence this is a slight generalization of Theorem 3 in Chapter III. We shall call this kind of property the constant regression of Q on $\overline{X}$ for it is easy to see that the condition $E(Qe^{it\overline{X}}) = E(Q)E(e^{it\overline{X}})$ is equivalent to $E(Q|e^{it\overline{X}}) = E(Q)$.

As mentioned before, the conditional expectation $E(Y|X)$ is called the regression of Y on X, and is a Borel measurable function of X. It is interesting to find out under what condition the regression of Y on X is constant, or a linear function, or a polynomial in X. We first introduce the following definition.

Definition 4.1. Let X and Y be two r.v.'s such that $E(Y|X)$ exists. Let K be a non-negative integer. Then the r.v. Y is said to have a polynomial regression of order k on X if

$$E(Y|X) = \sum_{j=o}^{k} \beta_j X^j \quad \text{holds almost everywhere.}$$

If $k = 0$, we say that Y has a constant regression on X. In such a case, we have $E(Y|X) = \beta_o = E(Y)$, provided that the expected value of Y exists. If $k = 1$ and $\beta_1 \neq 0$ ($k = 2$ and $\beta_2 \neq 0$), then we speak of linear (quadratic) regression.

We first establish a necessary and sufficient condition under which the random variable Y has a polynomial regression on X.

Lemma 1. Let X and Y be two r.v.'s such that $E(X^k)$ (k = a non-negative integer) and $E(Y)$ exist. Then the r.v. Y has a polynomial regression of order k on X iff the relation

$$\text{(A)} \quad E(Ye^{itX}) = \sum_{j=o}^{k} \beta_j E(X^j e^{itX}), \quad \beta_j \text{ is real } (j = 1,2,\ldots,k)$$

holds for all real t.

Proof. Necessity.

$$E(Y|X)\, e^{itX} = \left(\sum_{j=o}^{k} \beta_j X^j \right) e^{itX}$$

$$E(E(Y|X)e^{itX}) = E\left[\left(\sum_{j=o}^{k} \beta_j X^j \right) e^{itX}\right]$$

$$E(E(Ye^{itX}|X)) = E\left[\left(\sum_{j=o}^{k} \beta_j X^j \right) e^{itX}\right]$$

$$E(Ye^{itX}) = \sum_{j=o}^{k} \beta_j E(X^j e^{itX}).$$

To prove the converse, suppose that the relation (A) holds for all real t. Then

$$E(e^{itX}[Y - \sum_{j=o}^{k} \beta_j X^j]) = 0$$

or

$$\int_{R^1} e^{itX} E[Y - \sum_{j=o}^{k} \beta_j X^j | X] \, d\mu_x = 0$$ where μ_x is

the probability measure associated with the r.v. X (see pp.33, [10]). Let

$$\mu(B) = \int_B E(Y - \sum_{j=o}^{k} \beta_j X^j | X) d\mu_x ,$$ where B is a borel

set of R^1 . This is a function of bounded variation defined on all Borel sets of R^1. But then we have

$$\int_{R^1} e^{itX} d\mu = 0 .$$

It is well-known that a function of bounded variation is uniquely determined by its Fourier transform. Hence

$$\mu(B) = \mu(R^1) = 0$$ for all Borel sets B.

But this can happen only if

$$E(Y - \sum_{j=1}^{k} \beta_j X^j | X) = 0 \quad \text{a.e.}$$

This completes the proof.

In particular, for $k = 0$, we see that Y has constant regression on X iff the relation $E(Ye^{itX}) = E(Y)E(e^{itX})$ holds for all real t.

4.2. Constant regression of linear Statistics on another Statistic

Let X_1 $X_2,\ldots,X_n$ be a random sample from a normal distribution. It is easily shown that two uncorrelated linear statistics $L_1 = \sum_1^n a_jX_j$, $L_2 = \sum_1^n b_jX_j$ of the normal variates $X_1,\ldots,X_n$ are independent. We also know from Darmois-Skitovich's theorem that the independence of two linear forms implies that the random variables with nonzero coefficients in both forms are normal. It is, therefore, natural to ask whether the independence of the linear forms can be relaxed to a weaker condition of constant regression; i.e., whether normality can be characterized by the property of constant regression of a linear statistic on another. However, this is, in general, not true. Any linear statistic $L = \sum_1^n a_jX_j$ always has a constant regression or linear regression on the sample sum $X_1+X_2+\ldots+X_n$ by suitably choosing the coefficients a_j $(j = 1,2,\ldots,n)$. This points out that we shall only deal with constant regression of a linear statistic on some linear statistic not proportional to the sample sum. Indeed, under some restrictions on the coefficients of the linear statistics L_1 and L_2, Rao (1967) gives a slight generalization of Darmois-Skitovich's theorem. The following theorems are due to Rao (1967).

Theorem 1. (Rao 1967) Let X_1, X_2 be two i.i.d. r.v.'s with $E(X_1) = 0$. Let $a_1X_1 + a_2X_2$ and $b_1X_1 + b_2X_2$ be two linear functions of X_1 and X_2 with $a_i, b_i \neq 0$ $i = 1,2$ such that

$$E(a_1X_1 + a_2X_2 | b_1X_1 + b_2X_2) = 0 \qquad (1.1)$$

and $|b_2/b_1| \leq 1$ (without loss of generality).

Then we have

(i) X_1, X_2 are degenerate if $|a_2/a_1| \leq 1$ and $|b_2/b_1| < 1$.

(ii) X_1 and X_2 are normally distributed provided that $E(X_1^2) < \infty$

$a_1b_1 + a_2b_2 = 0$, and $|b_2/b_1| < 1$.

Proof of (i). By Lemma 1, we have

$$E[(a_1X_1 + a_2X_2)e^{it(b_1X_1+b_2X_2)}] = 0 \qquad (1.2)$$

holds for all real t.

Let $f(t)$ be their common ch.f. Since $E(X_1)$ exists, the first derivative $f'(t)$ exists and is continuous everywhere (1.2) can thus be written as

$$a_1f'(b_1t)\, f(b_2t) + a_2f(b_1t)\, f'(b_2t) = 0 \qquad (1.3)$$

There exists a neighbourhood of the origin, say $I = (-\delta, \delta)$, such that both $f(b_1 t)$ and $f(b_2 t)$ do not vanish in this neighbourhood.

Dividing (1.3) by $f(b_1 t)\, f(b_2 t)$ and writing $\psi(t) = \frac{f'(t)}{f(t)}$, we obtain

$$a_1\, \psi(b_1 t) + a_2\, \psi(b_2 t) = 0 \quad \text{for every } t \in I = (-\delta,\delta) \ldots \qquad (1.4)$$

This implies

$$\psi(t) = -\frac{a_2}{a_1}\, \psi\left(\frac{b_1 t}{b_1}\right) = \alpha\psi(\beta t) \qquad (1.5)$$

where $\alpha = -\frac{a_2}{a_1}$, $\beta = \frac{b_2}{b_1}$. Now if $|a_2/a_1| \leq 1$ and $|^{b}2/b_1| < 1$,

it follows from (1.5) that

$$\psi(t) = \alpha\psi(\beta t) = \alpha^2\, \psi(\beta^2 t) = \ldots = \alpha^n\, \psi(\beta^n t)$$

$$\psi(t) = \lim_{n \to \infty} \alpha^n\, \psi(\beta^n t) = 0$$

i.e. $\ln f(t) = c$ holds in $I = (-\delta,\delta)$.

By analytic continuation of ch.f., we have

$f(t) = c$ for all real t. That is to say, X_1 and X_2 are degenerate.

Proof of (ii). If $E(X^2) < \infty$, then the second derivative $f''(t)$ of $f(t)$ exists and is continuous everywhere. In such a case the first derivative of $\psi(t)$ also exists in $I = (-\delta,\delta)$ and is continuous at $t = 0$. Therefore $\psi(t)$ is of the form $t\phi(t)$ where $\phi(t)$ is continuous at the origin. Replacing $\psi(t)$ by $t\,\phi(t)$ in (1.4), we obtain for $t \neq 0$

$$\phi(t) = \alpha\,\beta\phi(\beta t) = \alpha^2\beta^2\phi(\beta^2 t) = \ldots = (\alpha\beta)^{2n}\,\phi(\beta^{2n} t)$$

i.e.,
$$\phi(t) = \lim_{n\to\infty} (\alpha\beta)^{2n}\,\phi(\beta^{2n} t) = \phi(0)$$

$$= \begin{cases} \phi(0) = k & \text{when } |\alpha\beta| = 1 \\ 0 & \text{when } |\alpha\beta| < 1. \end{cases}$$

But since $a_1b_1 = -a_2b_2$, we have $|\alpha\beta| = 1$

Hence
$$\psi(t) = kt \quad \text{for} \quad t \in I = (-\delta,\delta)$$

or
$$\log f(t) = \frac{k}{2}t^2 \quad \text{for} \quad t \in I = (-\delta,\delta) \quad .$$

By analytic continuation, $\log f(t) = \frac{k}{2}t^2$ holds for all t.
This completes the proof.

The following theorem is also established by Rao in the same paper [83].

Theorem 2. (Rao 1967) Let $X_1,\ldots,X_n$ be a random sample from X with d.f. $F(x)$ having mean zero. Let there exist two linear functions

$$L_1 = \sum_{i=1}^{n} X_i \quad \text{and} \quad L_2 = \sum_{i=1}^{n} b_iX_i \quad \text{with } |b_n| > \max(|b_1|,\ldots,|b_{n-1}|)$$

and $a_n \neq 0$, such that $E(L_1|L_2) = 0$. Then $F(x)$ is normal provided that $E(X^2) < \infty$, $\sum_{i=1}^{n} a_ib_i = 0$ and $a_ib_i/a_nb_n < 0$ for $i = 1,2,\ldots,n-1$.

Proof. By Lemma 1 one obtains a relation of the form

$$a_1 \psi(b_1 t) + \ldots + a_n \psi(b_n t) = 0 \quad \text{where} \quad \psi(t) = \frac{f'(t)}{f(t)} \tag{2.1}$$

and f(t) is the ch.f. of F(x). As seen in the previous proof $\psi'(t)$ exists and is continuous at the origin, and $\psi(t)$ can be expressed in the form $t\,\emptyset(t)$ where $\emptyset(t)$ is continuous at the origin.

From (2.1), by substituting $\psi(t)$ by $t\,\emptyset(t)$, one has,

$$\emptyset(t) = -\,[\alpha_1\beta_1\,\emptyset(\beta_1 t) + \ldots + \alpha_{n-1}\beta_{n-1}\,\emptyset(\beta_{n-1}t)\,]$$

where $\alpha_i = \dfrac{a_i}{a_n}$, $\beta_i = \dfrac{b_i}{b_n}$, $i = 1,2,\ldots,n-1$

or $\emptyset(t) = \gamma_1\,\emptyset(\beta_1 t) + \ldots + \gamma_{n-1}\emptyset(\beta_{n-1}t)$, $-\gamma_i = \alpha_i\beta_i$, $i = 1,2,\ldots,n-1$

(2.2)

In viewof $\sum_{i=1}^{n} a_i b_i = 0$ and $\dfrac{a_i b_i}{a_n b_n} < 0$ $(i = 1,2,\ldots,n-1)$, we see that $\sum_{i=1}^{n-1} \gamma_i = 1$ and all γ_i, $i = 1,2,\ldots,n-1$ are positive.

We have, by replacing t by $\beta_i t$ in (2.2),

$$\emptyset(\beta_i t) = \gamma_1\emptyset(\beta_1\beta_i t) + \ldots + \gamma_{n-1}\emptyset(\beta_{n-1}\beta_i t),\ i = 1,2,\ldots,n-1$$

so that $\emptyset(t) = \sum_{i=1}^{n-1} \gamma_i\emptyset(\beta_i t) = \sum_{j}^{n-1}\sum_{i}^{n-1} \gamma_i\gamma_j\ \emptyset(\beta_i\beta_j t)$

or $\emptyset(t) = \sum_j \sum_i \delta_{ij}\ \emptyset(\beta_i\beta_j)$ where $\delta_{ij} = \gamma_i\gamma_j$,

$$\text{and} \quad \sum_{ij} \delta_{ij} = 1 \tag{2.3}$$

Hence $\phi(t) - \phi(0) = \sum_j \sum_i \delta_{ij}[\phi(\beta_i\beta_j t) - \phi(0)]$ (2.4)

Proceeding in this manner, one obtains

$$\phi(t) - \phi(0) = \Sigma\, \delta_{i_1,\ldots,i_m} [\phi(\beta_{i_1} \cdots \beta_{i_m} t) - \phi(0)] \quad (2.5)$$

where $\Sigma\, \delta_{i_1 \ldots i_m} = 1$ and summation is taken over all integers i_k, $i \leq i_k \leq m$, $k = 1,2,\ldots,n-1$. Now for any fixed $t \neq 0$, $t < 0$, we can choose m so large that $[\max_i |\beta_i|]^m < \eta/|t|$ where η is such that $|\phi(\mu) - \phi(0)| < \epsilon$ for $|\mu| < \eta$. But then the modulus of the right-hand side of (2.5) is less than ϵ so that for any $\epsilon > 0$, we have $|\phi(t)-\phi(0)| < \epsilon$.

i.e. $\phi(t) = \phi(0) = c$ (say)

$\Rightarrow$ $\phi(t) = ct$

$\log f(t) = \frac{c}{2} t^2$ which completes the proof in view of the analytic continuation of the characteristic function.

As an application of Theorem 2, we have

Corollary 2.1. Let $X_1,\ldots,X_n$ be given as in Theorem 2. If $E(\bar{X}|X_j-\bar{X}) = 0$ for any $j = 1,2,\ldots,n$, where $\bar{X} = \sum_1^n X_i/n$ and if $E(X_i^2) < \infty$, then $F(x)$ is normal.

Instead of considering the conditional expectation of one linear function of the random sample on another linear function, one may consider the conditional expectation of one linear function given several linear functions. Hence an extension of Theorem 2 is possible in the following direction, in which "identically distributed" is not assumed as in the preceding theorems.

Theorem 3. (Rao 1967). Let $X_1,\dots,X_n$ $(n \geq 3)$ be n independently (not necessarily identically) distributed r.v.'s. Let the three linear functions $L_1 = \sum_{i=1}^{n} a_i X_i$, $L_2 = \sum_{i=1}^{n} b_i X_i$, $L_3 = \sum_{i=1}^{n} c_i X_i$ be such that

i) $a_i \neq 0$, $i = 1,2,\dots,n$

ii) for each i, b_i and c_i are not both zero.

iii) $\beta_i \neq \beta_j$ for any $i \neq j$ where $c_i/b_i = \beta_i$ if both c_i and b_i are not zero.

iv) all α_i, where $\alpha_i = \dfrac{c_i}{a_i}$ when $c_i \neq 0$ and $b_i = 0$ are of the same sign and all $\delta_j = \dfrac{b_i}{a_i}$ when $b_i \neq 0$, $c_i = 0$ are of the same sign.

Then $X_1,X_2,\dots,X_n$ are all normally distributed provided that $E(L_1|L_2,L_3) = 0$.

We first establish a useful lemma by Linnik (1964) and Rao (1966).

Lemma 2. (Linnik, 1964 and Rao 1966). Let b_j, $(j = 1,2,\dots,n)$ be real numbers such that $b_j \neq b_k$, $j \neq k$, $j,k = 1,2,\dots,n$. Let

$\phi_1, \phi_2, \ldots, \phi_n$, A, B be continuous functions. Assume that the equation

(*) $\phi_1(u+b_1v) + \phi_2(u+b_2v) + \ldots + \phi_n(u+b_nv) = A(u) + B(v) + Q(u,v)$

holds in $|u| < \delta_o, |v| < \delta_o$, where u and v are real, and **Q** is a quadratic function. Then $\phi_1, \ldots, \phi_n$ A and B are all polynomial functions of degree max(2,n) at most in a neighbourhood of the orign.

Proof. Multiplying both sides of the equation (*) by (x-u) and integrating with respect to u from 0 to x, where $|x| < \delta_o$, we get

$$\sum_{j=1}^{n} \int_o^x (x-u)\phi_j(u+b_jv)du = \int_o^x (x-u)\ A(u)du+B(v) \int_o^x (x-u)du +$$

$$+ \int_o^x (x-u)\ Q(u,v)du = c(x) + x^2B_1(v) + x^3D(v) \qquad \text{(i)}$$

where D(v) is linear in v and $B_1(v)$ is a continuous function of v. Letting $u+b_jv = t$, $|t| < \delta_1 < \delta_o$ in (i), we get

$$\sum_{j=1}^{n} \int_{b_jv}^{x+b_jv} (x+b_jv-t)\ \phi_j(t)dt = x^3D(v) + x^2B_1(v)+c(x) \qquad \text{(ii)}$$

The equality (ii) is true when $|x| < \delta_2$, $|v| < \delta_2$, $0 < \delta_2 < \delta_1$. Since the left-hand side of (ii) is differentiable with respect to v for each fixed value of x, $|x| < \delta$, it follows that $B_1(v)$ is also differentiable with respect to v. Writing (ii) in the form

$$\sum_{j=1}^{n} b_j v \int_{b_j v}^{x+b_j v} \phi_j(t)dt = x^3 D(v) + x^2 B_1(v) + c(x) -$$

$$- \sum_{j=1}^{n} x \int_{b_j v}^{x+b_j v} \phi_j(t)dt + \sum_{j=1}^{n} \int_{b_j v}^{x+b_j v} t\, \phi_j(t)dt$$

and then differentiating it with respect to v, we get

$$\sum_{j=1}^{n} b_j \int_{o}^{x+b_j v} \phi_j(t)dt = hx^3 + B_1'(v)x^2 + B_2(v)x + B_3(v) \qquad \text{(iii)}$$

in which $\sum_{j=1}^{n} b_j \int_{o}^{b_j v} \phi_j(t)dt$ has been added to both sides.

Here h is a constant, $B_2(v)$ and $B_3(v)$ are functions of v, and the equality (iii) holds for a certain domain of v and x. Now differentiating both sides of (iii) with respect to x, we get

$$\sum_{j=1}^{n} b_j \phi_j(x+b_j v) = 3hx^2 + 2B_1'(v)x + B_2(v) \qquad \text{(iv)}$$

Setting $v = 0$ in (iv) we have

$$\sum_{j=1}^{n} b_j \phi_j(x) = f_{21}(x) \qquad \text{(v)}$$

where $f_{21}(x)$ is a polynomial in x of degree two at most.

Starting from the equation (iv), and repeating the procedure as before, we get

$$\sum_{j=1}^{n} b_j^2 \, \phi_j(x) = f_{22}(x)$$

where $f_{22}(x)$ is a polynomial in x of degree three at most. Hence by repeating the procedure n times, we obtain the equations

$$\begin{aligned} b_1\phi_1(x) + \ldots + b_n\phi_n(x) &= f_{21}(x) \\ b_1^2\phi_1(x) + \ldots + b_n^2\phi_n(x) &= f_{22}(x) \\ b_1^n\phi_1(x) + \ldots + b_n^n\phi_n(x) &= f_{2n}(x) \end{aligned} \qquad \text{(vii)}$$

where $f_{2n}(x)$ is a polynomial in x of degree n at most. Since $b_1,\ldots,b_n$ are all different, ϕ_j, $j = 1,2,\ldots,n$ are uniquely determined by solving the equations (vii) and are linear combinations of $f_{21}(x),\ldots,f_{2n}(x)$. This means that each ϕ_j, $j = 1,2,\ldots,n$ is a polynomial of degree $\max(2,n)$ at most in a neightbourhood of the origin.

We now proceed to the proof of the theorem. There is no loss of any generality in assuming that

$b_1 = \ldots = b_s = 0$, $c_{s+1} = \ldots = c_{s+k} = 0$ and $b_i \neq 0$, $c_i \neq 0$ for $i = s+k+1,\ldots,n$.

Denote $Y_i = a_iX_i$, $i = 1,2,\ldots, s+k$

$Y_i = b_iX_i$, $i \geq s+k+1$ so that we have

$$L_1 = Y_1+Y_2 + \ldots + Y_{s+k} + \gamma_{s+k+1}Y_{s+k+1} + \ldots + \gamma_nY_n \quad \text{where } \gamma_j = \frac{a_j}{b_j} \qquad (3.1)$$

$$L_2 = \delta_{s+1}Y_{s+1} + \ldots + \delta_{s+k}Y_{s+k} + Y_{s+k+1} + \ldots + Y_n \quad \text{where} \quad \delta_{s+j} = \frac{b_{s+j}}{a_{s+j}} \ .$$

$$L_3 = \alpha_1 Y_1 + \ldots + \alpha_s Y_s + \beta_{s+k+1}Y_{s+k+1} + \ldots + \beta_n Y_n \text{ where } \alpha_i = \frac{c_i}{a_i}$$

and $\beta_{s+k+j} = \dfrac{c_{s+k+j}}{b_{s+k+j}}$.

If $E(L_1|L_2,L_3) = 0$, then we have

$$E(L_1 e^{it_2L_2 + it_3L_3}) = 0 \quad . \tag{3.2}$$

In terms of ch.f.'s $f_1,\ldots,f_n$ of Y_i, $i = 1,2,\ldots,n$, and writing

$\dfrac{f_i'(t)}{f(t)} = \phi_i(t)$, $i = 1,2,\ldots,n$, the equation (3.2) becomes

$$\sum_1^s \phi_i(\alpha_i t_3) + \sum_{s+1}^{s+k} \phi_i(\delta_i t_2) + \sum_{s+k+1}^{n} \gamma_i \phi_i(t_2 + \beta_i t_3) = 0 \tag{3.3}$$

in a neighbourhood $|t_2| < \delta_1$ and $|t_3| < \delta_3$.

We may write (3.3) in the form

$$\sum_{s+k+1}^{n} \gamma_i \phi_i(t_2 + \beta_i t_3) = - \sum_1^s \phi_i(\alpha_i t_3) - \sum_{s+1}^{s+k} \phi_i(\delta_i t_2) \ .$$

Since all $\beta_{s+k+1},\ldots,\beta_n$ are different, and $-\sum_1^n \phi_i(\alpha_i t_3)$, $-\sum_{s+1}^{s+k} \phi_i(\delta_i t_2)$

are continuous, by Lemma 2, $\phi_i(t_3)$ (i = s+k+1,...,n) $\sum_1^s \phi_i(\alpha_3 t_3)$, and $\sum_{s+1}^{s+k} \phi_i(\delta_i t_3)$ are all polynomials in t. This implies that $\ln f_{s+k+j}$, $j = 1,\ldots,n-s-k$ are also polynomials in t. By Marcinkiewicz's theorem, $Y_{s+k+1},\ldots,Y_n$ are normally distributed.

Furthermore, it is obvious that $\sum_1^s \alpha_1^{-1} \ln f_i(\alpha_i t)$ is a polynomial in t. Since α_i are all of the same sign, by a result listed in Chapter I on analytic ch.f.'s, $f_i(t)$, $i = 1,2,\ldots,s$ are ch.f's of normal distributions. Similarly Y_i, $i = s+1,\ldots,s+k$ are all normally distributed.

It should be noted that if $\beta_{i1} = \beta_{i2} = \ldots = \beta_{ik}$ for a set of indices $i_1,\ldots,i_k$, and if the corresponding $\gamma_{i1},\ldots,\gamma_{ik}$ are of the same sign, then Theorem 3 is still true. This is because the functional equation (3.3) ensures that $\sum_{j=1}^{k} \gamma_{ij}\phi_{ij}(t)$ is a polynomial in t in a certain neighbourhood of the origin. With the same argument as before, the r.v.'s Y_{ij}, $j = 1,2,\ldots,k$ can be seen to be normal.

As a corollary, we have

Corollary 3.1. Let $X_1,\ldots,X_n$ $(n \geq 3)$ be independent (but not necessarily indentically distributed) r.v.'s such that $E(X_i) < \infty$, $i = 1,2,\ldots,n$. Then $E(X|X_1 - X, X_2 - X) = 0$ implies that $X_1,\ldots,X_n$ are all normally distributed.

It was shown in a paper by Rao (1952) that the mean square error of $\bar{X}$ is not smaller than

$$\mathrm{Var}[X-E_o(\bar{X}|X_2-X_1, \ldots, X_n-X_1)]$$

where $X_1,\ldots,X_n$ are i.i.d r.v.'s from a location parameter family, where Var denotes the variance and E_o denotes the conditional expectation when the location parameter vanishes. Also it was shown that $\bar{X}-E_o(\bar{X}|X_2-X_1,\ldots,X_n-X_1)$ is an unbiased estimation of the location parameter, and that $\bar{X}$ is the minimum variance unbiased estimator of the location parameter when the distribution is normal. This means that $E_o(X|X_2-X_1,\ldots,X_n-X_1) = 0$. It is interesting to note that this proposition characterizes a normal distribution without however assuming that the underlying distribution belongs to the location parameter family. It was shown by Kagan, Linnik and Rao (1965) that $E(X|X_2-X_1,\ldots,X_n-X_1) = 0$ implies that X_i is normal, where X_i, $i = 1,2,\ldots,n$ $(n \geq 3)$ are i.i.d. r.v.'s with mean zero. This result turns out to be a special case of the following theorem which is also due to Rao (1967).

Theorem 4. (Rao 1967) Let $X_1,\ldots,X_n$ $(n \geq 3)$ be independent (but not necessarily identically distributed) nondegenerate r.v.'s with $E(X_i) = 0$ $i = 1,2,\ldots,n$. Consider the n linear functions

$$L_1 = a_{i1}X_1 +\ldots+ a_{in}X_n, \quad i = 1,2,\ldots,n$$

with $a_{ij} \neq 0$, $j = 1,2,\ldots,n$ such that the determinant $|(a_{ij})| \neq 0$. Then $X_1,\ldots,X_n$ are all normally distributed if $E(L_1|L_2,\ldots,L_n) = 0$.

It is easy to show from the given condition that

$$E(L_1 e^{it_1L_2+\ldots+it_{n-1}L_n}) = 0 \quad .$$

We may assume $a_{11} = a_{12} =\ldots= a_{1n} = 1$ (without loss of generality) In terms of $\phi_j(t) = \dfrac{f_i'(t)}{f_i(t)}$, where $f_i(t)$ is the ch.f. of X_i, $i = 1,2,\ldots,n$, we obtain a functional equation of the form

$$\phi_1(a_{21}t_1+\ldots+a_{n1}t_{n-1}) +\ldots+ a_n(a_{2n}t_1 +\ldots+ a_{nn}t_{n-1}) = 0 \; .$$

By letting $a_{2j}t_1 +\ldots+ a_{nj}t_{n-1} = T_j$, $j = 1,2,\ldots,n-1$, the above equation can be expressed as

$$\phi_1(T_1) +\ldots+ \phi_{n-1}(T_n) + \phi_n(k_1T_1 +\ldots+ k_{n-1}T_{n-1}) = 0$$

where $k_1,\ldots,k_{n-1}$ are suitable constants. Putting $T_i = 0$ except for $i = r$ and $i = s$ and in view of $\phi_j(0) = 0$, $j = 1,2,\ldots,n$ we obtain

$$\phi_r(T_r) + \phi_s(T_s) = - \phi_n(k_rT_r + k_sT_s) \quad .$$

Hence by Lemma 2 $\phi_r(t)$ and $\phi_s(t)$ are polynomials in t. Since r and s are arbitrary, we conclude that $\phi_1(t)$,..., $\phi_n(t)$ are all polynomials in t. With the same reasoning as before $X_1,\dots,X_n$ are all normally distributed.

Let $X_1,\dots,X_n$ be a random sample from X with d.f. F(x) having mean zero and finite k-th absolute moment. Consider the linear functions $L_1 = \sum_1^n a_iX_i$ and $L_2 = \sum_1^n b_iX_i$ with $\sum_1^n a_ib_i^r \neq 0$, $r = 2,\dots,k-1$. If $E(L_1|L_2) = 0$, then it can be shown that $\sum_1^n a_ib_i = 0$ and that the j^{th} cumulant of F(x) is equal to zero for $j = 3,\dots,k$. The by assuming that F(x) has moments of every order, we immediately get a characterization of the normal distribution (see 83). The converse of the proposition is also true sicne $E(L_1|L_2) = 0$ implies $\sum_1^n a_ib_i = 0$ which in turns implies that L_1 and L_2 are uncorrelated. This means that L_1 and L_2 are independent, and hence the assertion follows at once.

In Theorem 4, we have seen that Normality is characterized by the constant regression of a linear function given several linear functions of the nondegenerate independent r.v.'s. It is interesting to note that the assertion of Theorem 4 is still true if the condition $E(L_1|L_2,\dots,L_n)$ is replaced by an analogous condition as $E(L_i|L_{p+1},\dots,L_n) = 0$ for $i = 1,\dots,p$. To see this, we first transform the L_j to L'_j which takes the form

$$L'_j = X_j + b_{jp+1}X_{p+1} + \dots + b_{jn}X_n \qquad j = 1,2,\dots,p.$$

This is possible since the determinant $|(a_{ij})| \neq 0$. Similarly transform $L_{p+1},\dots,L_n$ to $L'_{p+1},\dots,L'_n$ which take the forms

$$L'_{p+k} = c_{p+k,1}X_1+\dots+c_{p+k,p}X_p + X_{p+k}, \quad k = 1,\dots, n-p.$$

The condition $E(L_i \mid L_{p+1},\dots, L_n) = 0$ is equivalent to $E(L'_i \mid L'_{p+1},\dots,L_n) = 0$, $i = 1,..,p$. From this condition, we get as before a functional equation of the form

$$\phi_i(c_{p+1,i}t_1 +\dots+ c_{ni}t_{n-p})+b_{i,p+1}\phi_{p+1}(t_1)+\dots+b_{in}\phi_n(t_{n-p}) = 0$$

$i = 1,2,\dots,p$. Applying the same arguments as in Theorem 4, and using Lemma 2, we conclude that the assertion is true.

4.3. Constant regression and polynomial regression of a polynomial Statistic on a linear Statistic.

We have seen that the complete independence of two linear statistics can be replaced by a weaker condition, namely the constant regression of one linear statistic on another. The same is true for a quadratic statistic on a linear statistic as mentioned in the beginning of this chapter. A thorough and detailed analysis of the proofs such as that of Theorem 3 in Chapter III readily gives us some modifications of the previous results. Therefore we shall only state the following results without proofs.

Theorem 5. (Laha) Let $X_1,\dots,X_n$ be a random sample from X with d.f. F(x) having finite variance σ^2. Consider the polynomial statistic

$$Q = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk}X_jX_k + \sum_{j=1}^{n} b_jX_j$$

with coefficients satisfying

(i) $B_1 = \sum_{j=1}^{n} a_{jj} \neq 0, \quad B_2 = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} = 0 \quad \text{and} \quad B_3 = \sum_{j=1}^{n} b_j = 0$.

Then F(x) is normal iff Q has constant regression on the sample sum $X_1 + X_2 + \dots + X_n$.

Corollary 5.1. (Laha 1953) Let $X_1,\dots,X_n$ be a random sample from X with d.f. F(x) having a finite variance σ^2. If the regression of any unbiased quadratic statistic $Q = \sum_{ij}^{n} a_{ij}X_iX_j$ of $c\sigma^2 (c \neq 0)$ on $X_1 + X_2 + \dots + X_n$ is constant, then F(x) is normal.

From the definition of unbiasedness, we have

$$E(Q) = c\sigma^2 .$$

The above equation implies that $\sum_{j=1}^{n} a_{jj} = c$ and $\sum_{ij}^{n} a_{ij} = 0$. By Theorem 5, the result follows.

Instead of constant regression, we may as well consider polynomial regression which also characterizes Normality. The following theorem is a generalization of Theorem 5.

Theorem 6. (Laha and Lukacs 1960) Let $X_1, X_2, \ldots, X_n$ be a random sample from X with d.f. $F(x)$ having a finite variance σ^2. Consider a polynomial statistic $Q = \sum_{ij=1}^{n} a_{ij} X_i X_j + \sum_{j=1}^{n} b_j X_j$ with coefficients satisfying $A = A_1(n-1) - A_2 \neq 0$, where $A_1 = \sum_{i=1}^{n} a_{ii}$ and $A_2 = \sum_{i \neq j}^{n} a_{ij}$. Let β_1 and β_2 be two real constants such that $\beta_1 = \frac{\beta}{n}$ and $\beta_2 = \frac{1}{n^2}(A_1 + A_2)$ where $\beta = \sum_{j=1}^{n} b_j$. Then the relation $E(Q|\Lambda) = \beta_o + \beta_1 \Lambda + \beta_2 \Lambda^2$ holds iff (i) $\beta_o = A\frac{\sigma^2}{n}$, (ii) $F(x)$ is normal, where $\Lambda = X_1 + X_2 + \ldots + X_1$.

We first note that if $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} = 0$, $\beta = \sum_{j=1}^{n} b_j = 0$, then $\beta_1 = 0$, and $\beta_2 = 0$. This reduces to the case of Theorem 5. We now proceed to the proof of the Theorem 6. By Lemma 1 we have

$$E(Qe^{it\Lambda}) = \beta_o E(e^{it\Lambda}) + \beta_1 E(\quad e^{it\Lambda}) + \beta_2 E(\Lambda^2 e^{it\Lambda}) \tag{6.1}$$

holds for all real t. Let $f(t)$ be the ch.f. of $F(x)$ and $\emptyset(t) = \ell n\, f(t)$ be its cummulat generating function (valid in a neighbourhood of the origin). We have

$$E(Qe^{it\Lambda}) = -[f(t)]^n \left\{ A_1 \emptyset''(t) + (A_1 + A_2)[\emptyset'(t)]^2 - \beta i \emptyset'(t) \right\}$$

(6.2)

and $$\beta_o E(e^{it\Lambda}) + \beta_1 E(\quad e^{it\Lambda}) + \beta_2 E(\quad ^2 e^{it\Lambda})$$

$$= - [f(t)]^2 \left\{ - \beta_o - n\beta_1 i\emptyset'(t) + n\beta_2\emptyset''(t) + n^2\beta_2[\emptyset'(t)]^2 \right\} \quad (6.3)$$

From (6.1) and using (6.2) and (6.3), we obtain

$$\beta_O = (n\beta_2 - A_1)\ \emptyset''(t) + [n^2\beta_2 - (A_1+A_2)][\emptyset'(t)]^2 +$$

$$+ (\beta - n\beta_1)i\emptyset'(t) \quad (6.4)$$

But since $n^2\beta_2 - (A_1+A_2) = 0$, $\beta - n\beta_1 = 0$, we have from (6.4)

$$\beta_o = (n\beta_2 - A_1)\ \emptyset''(t) .$$

By letting $t = 0$ and using $\emptyset''(0) = - \sigma^2$, we obtain $\beta_o = \frac{\sigma^2}{n} A$ and hence $\emptyset''(t) = - \sigma^2$.

Hence $f(t) = \exp\left\{ - \frac{1}{2}\sigma^2 t^2 + i\alpha t \right\}$ holds in a neighbourhood of the origin. By analytic continuation, we conclude that $F(x)$ is normal.

To prove the sufficiency, write

$$\emptyset(t) = \log f(t) = - \frac{1}{2}\sigma^2 t^2 + i\alpha t \quad .$$

Uisng (6.2) and (6.3) and taking into account $\beta_1 = \frac{\beta}{n}$, $\beta_2 = \frac{A_1+A_2}{n^2}$ and $\beta_2 = A\frac{\sigma^2}{n}$, we have

$$E(Qe^{it\Lambda}) = \beta_o E(e^{it\Lambda}) + \beta_1 E(\Lambda e^{it\Lambda}) + \beta_2 E(\Lambda^2 e^{it\Lambda})$$

for all real t. By Lemma 1, we have

$$E(Qe^{it\Lambda}) = \beta_o + \beta_1\Lambda + \beta_2 \Lambda^2 \quad \text{a.e.}$$

This completes the proof.

By appropriately choosing the coefficients of the polynomial statistic Q and the regression coefficients β_o, β_1 and β_2 we have seen that the regression properties of a polynomial statistic of second degree on a linear statistic characterize the normal distribution. Applying the similar technique as in Theorem 6, Gordon and Mathai (1968) generalize the above result by constructing a polynomial statistic of third degree which has polynomial or constant regression on the sample sum $X_1 + \ldots + X_n$ and obtain a series of characterization theorems for various populations such as Normal, Gamma and Poisson etc. Moreover, a technique that can be used to study any general $r^{th} \leq m^{th}$ order polynomial regression of any m^{th} degree polynomial statistic on a linear one is also revealed in [28] . We present here a few results proven in [28]. The proofs of these are similar to Theorem 6, and hence will be omitted.

Theorem 7. Let $X_1, \ldots, X_n$ be a random sample from X with d.f. F(x) having finite third moment. Let

$$S = \sum_{j,k,m} a_{jkm} X_j X_k X_m + \sum_{j,k} b_{jk} X_j X_k + \sum_j c_j X_j$$

be a polynomial statistic of third order, where a_{jk}, b_{jk} and c_j (for all $j,k,m = 1,\ldots,n$) are real constants.

Assume that the following relations hold

$$\beta_o = 0$$

$$c - n\beta_1 = 0$$

$$B_1 - n\beta_2 = 0$$

$$B_1 + B_2 - n^2\beta_2 = 0$$

$$n(n+2)\beta_3 - 3A_1 - A_2 = 0$$

$$n(n^2-2n+2)\beta_3 - A_1 - A_2 - A_3 = 0$$

$$n\beta_3 - A \neq 0 ,$$

where $A_1 = \sum_j a_{jjj}$, $A_2 = \sum_{j\neq k} (a_{jjk} + a_{jkj} + a_{kjj})$,

$A_3 = \sum_{j\neq k\neq m} a_{jkm}$, $B_1 = \sum_j b_{jj}$, $B_2 = \sum_{j\neq k} b_{jk}$, $C = \sum_j C_j$, and

$\beta_o, \beta_1, \beta_2, \beta_3$ are real constants. Then $F(x)$ is normal iff

$E(S|\Lambda) = \beta_o + \beta_1\Lambda + \beta_2\Lambda^2 + \beta_3\Lambda^3$ a.e., where $\Lambda = X_1 + X_2 + \ldots + X_n$.

The condition $E(S|\Lambda) = \beta_o + \beta_1\Lambda + \beta_2\Lambda^2 + \beta_3\Lambda^3$ yields a fundamental differential equation of third order in the ch.f. $f(t)$ of $F(x)$. The assumptions of the coefficients of the statistic S and the regression coefficients permit us to reduce the fundamental equation to the simple form

$$i(n\beta_3 - A_1) \frac{d^3}{dt^3} \log f(t) = 0$$

which readily gives us a solution of a ch.f. of a normal distribution.

By imposing another set of conditions of the coefficients of the statistic S and the regression coefficients, for example

$$c - n\beta = 0$$

$$n\beta_3 - A = 0$$

$$B_1+B_2-n^2\beta_2 = 0$$

$$n(n+2)\beta_3 - 3A_1-A_2 = 0$$

$$n(n^2 - 2n+2)\beta_3 - A_1-A_2-A_3 = 0$$

$$\beta_o = \sigma^2(B_1-n\beta_2) \neq 0$$

where σ^2 is the variance of $F(x)$, we can also obtain another characterization of the normal distribution. A glance at the conditions of the coefficients of the statistic S and the regression coefficients will tell us the fundamental differential equation resulting from the condition $E(S|\Lambda) = \beta_o + \beta_1\Lambda + \beta_2\Lambda^2 + \beta_3\Lambda^3$ is rather complicated and cannot be solved readily. Hence if we go on considering the regression properties of a m^{th} $(m \geq 3)$ degree polynomial statistic on a linear one, we shall meet with a series of conditions on the coefficients of the m^{th} degree polynomial statistic and the regression coefficients and also have to deal with a very complicated differential equation of m^{th} order. In such cases, it is worthwhile to investigate the analytical properties of those solutions which are characteristic functions.

Linnik and Zinger (1957) discussed a special class of polynomial of degree r that has a constant regression on the sample sum . This class of polynomial is called regular polynomial (see pp.110 [48]) They obtained two important results, one regarding the existence of the moments of the underlying distribution function, the other concerning the analytical properties of the ch.f. of the underlying distribution function from the assumption that a regular polynomial statistic of degree r has a constant regression on the sample sum $X_1 + X_2 + \ldots + X_n$. We shall only state these two results, and for the proofs, we refer the reader to [48] pp.110.

Theorem 8. (Linnik and Zinger 1957). Let $X_1, \ldots, X_n$ be a random sample from X with d.f. $F(x)$ having finite moments up to order m. Let $p(X_1, \ldots, X_n) = \Sigma A_{j_1, \ldots, j_n} X_1^{j_1}, \ldots, X^{j_n}$ be a regular polynomial statistic [see pp. 110] of degree r and order $m (m \leq r)$. If $p(X_1, \ldots, X_n)$ has constant regression on the sample sum $X_1 + \ldots + X_n$, then $F(x)$ has moments of all orders.

Theorem 9. Under the conditions of Theorem 8, if (i) $p(X_1, \ldots, X_n)$ has constant regression on the sample sum $X_1 + \ldots + X_n$ (ii) $m \geq n-1$ then the ch.f. of $f(t)$ of $F(x)$ is an entire function.

It should be noted here that the assertion of Theorem 8 still holds if the sample sum is replaced by any linear statistic $a_1X_1 + \ldots + a_nX_n$.

As an application of Theorem 9, we have the following corollary:

Corollary 9.1. Let $X_1, \ldots, X_n$ be a random sample from X with d.f. $F(x)$ having finite second moment . Let $L_1 = a_1X_1 + \ldots + a_nX_n$ and

$L_2 = b_1X_1 + \ldots + b_nX_n$ be two linear statistics. If L_2^2 has constant regression on L_1, then $F(x)$ has moments of every order.

We note here that the assertion follows by applying Theorem 9 with $P = L_2^2$ with $m = r = 2$. We now give a characterization of the normal distribution.

Theorem 10. (Cacoullas 1967). Let $X_1, \ldots, X_n$ be a random sample from X with d.f. $F(x)$. Let $L_1 = a_1X_1 + \ldots + a_nX_n$ and $L_2 = b_1X_1 + \ldots + b_nX_n$ be two linear statistics with $a_1b_1 + a_2b_2 + \ldots + a_nb_n = 0$ and $a_ja_k > 0$ for all $j, k = 1,2,\ldots,n$. Then L_2^2 has constant regression on L_1 iff $F(x)$ is normal.

As seen before, the sufficiency is obvious. To prove the necessity, consider

$$E(L_2^2 \, e^{itL_1}) = E(L_2^2) \, E(e^{itL_1}) \; . \qquad (10.1)$$

In terms of the ch.f. $f(t)$ of $F(x)$, we have from (10.1).

$$\sum_1^n c_j^2 g_j''(t) \prod_{s\neq j} g_s(t) + \sum_{j\neq k} c_j c_k g_j'(t) g_k'(t) \prod_{r\neq j,k} g_r(t) = - E(L_2^2) \prod_{j=1}^{n} g_j(t) \qquad (10.2)$$

where $c_j = \frac{b_j}{a_j}$ and $g_j(t) = f(a_jt)$, $j = 1,2,\ldots,n$. By corollary 9.1 all moments of $F(x)$ exist, and hence the k^{th} derivatives $g_j^{(k)}(t)$ of $g_j(t)$ exists for every k. Let $\phi_j(t) = \ln g_j(t)$, $j = 1,2,\ldots,n$. (valid in a neighbourhood of the origin). Dividing both sides of (10.2) (permissible in a neighbourhood of the origin) by $\prod_{j=1}^{n} g_j(t)$ we have

$$\Sigma\, c_j^2\, \emptyset''_j(t) + [\Sigma\, c_j \emptyset'_j(t)]^2 = -\, E(L^2_2) \qquad (10.3)$$

We shall show that the cumulants k_r of F(x) are zero for $r > 2$.

Differentiating (10.3) once and setting $t = 0$, we get

$$k_3 \sum_1^n c^2_j\, a^3_j + 2k_1k_2\left(\sum_1^n c_j a_j\right)\left[\sum_j^n c_j a_j^2\right] = 0 \qquad (10.4)$$

by taking into account the relations $\emptyset_j^{(r)}(0) = a^r_j\, \emptyset^{(r)}(0) = i^r a^r_j\, k_r$.

But $\sum_i^n c_j a^2_j = \sum_1^n a_j b_j = 0$ and $\sum_1^n a_j^3\, c^2_j \neq 0$, we must have $k_3 = 0$.

Similarly, we can show $k_r = 0$ for $r > 2$. This completes the proof.

If we are given a sample $X_1,\ldots,X_n$ from X with a symmetrical distribution, then following the proof of Theorem 10, we can show that the distribution is normal iff L^2_2 has a constant regression on L_1 where $L_2 = b_1X_1 +\ldots+ b_nX_n$ and $L_1 = a_1X_1 +\ldots+ a_nX_n$ with $a_j \neq 0$, $j = 1,2,\ldots,n$ and $\sum_1^n a_jb_j = 0$.

4.4. Multivariate Case. Having seen a number of results in univariate case, we now come to consider the multivariate case. Analogous results can be formulated without much difficulties.

Theorem 11. (Kagan, Linnik and Rao 1965). Let $X_1,\dots,X_n$ be a sample from a multivariate distribution with $E(X_i) = 0$. Then $E(\overline{X}|X_2-X_1,\dots, X_n-X_1) = 0$ implies that the common distribution of $X_1,\dots, X_n$ is multivariate normal .

The result follows from Cramer-World theorem and Kagan-Linnik and Rao's theorem in univariate case.

Theorem 12. Let $X_1,\dots,X_n$ be a random sample from a nondegenerate p-dimensional distribution with covariance matrix M. Let

$$Q = \sum_{j=1}^{n} \sum_{i}^{n} A_{ji} X_j X_i'$$

where X_j is a (column) random vector, A_{jj} are $p \times p$ matrices satisfying (i) $\sum_{j=1}^{n} A_{jj} = - \sum_{j \neq k} A_{jk} = A$; (ii) A is nonsingular. Then the distribution is p-variate normal iff Q has constant regression on $L = X_1 + X_2 + \dots + X_n$.

This result is a generalization of Lukacs' result (1942) in multivariate case as well as an analogous form of Laha's result in univariate case. By putting

$$a_{jj} = \frac{1}{n}, \quad j = 1,\dots,n$$

$$a_{jk} = - \frac{1}{n(n-1)}, \quad j \neq k, \quad j,k = 1,\dots, n ,$$

then $$Q = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - X)(X_j - X)'$$ which is the case considered in [62] by Lukacs.

The proof of this result is nothing new, except changes of notations, and hence is omitted.

Similarly, Theorem 10 has an analogous form in multivariate case in the light of Cramer-Wold's theorem.

4.5. Linearity of regression and homoscedasticity

In the conclusion of this chapter, we consider some properties of linearity of regression and homoscedasticity and its applications in characterization problems. It is a remarkable fact that the regression of a component of a two-dimensional normal random vector on another is linear and the conditional variance of a component of a two-dimensional normal random vector does not depend on another. A question arises whether this property is only enjoyed by the normal distribution. In the following we shall establish the normality of the r.v.'s under consideration from the property of the linearity of regression and homoscedasticity of the conditional distribution. For this purpose, we introduce the following concepts.

Let X and Y be two r.v.'s and assume that the second moment of Y exists. The expression

$$E(Y^2|X) - [E(Y|X)]^2$$

is called the conditional variance of Y given X and will be denoted by var $(Y|X)$. We say that the conditional distribution of Y given X is homoscedastic if the conditional variance of Y given X is a constant. i.e., $\mathrm{var}(Y|X) = \sigma_o^2$ a.e., $\sigma_o^2 > 0$. We first establish the following necessary and sufficient conditions for the existence of linearity of regression and homoscedasticity of the conditional distribution of a r.v. Y given X.

Lemma 3. (Rao, Mourier and Rothschild). Let X and Y be two r.v.'s having finite second moments. Then the r.v. Y has linear regression on X and the conditional variance of Y given X is constant, i.e., $E(Y|X) = \alpha + \beta X$, $\mathrm{var}(Y|X) = \sigma_o^2$ a.e. iff the relations

$$\left.\frac{\partial f(u,v)}{\partial v}\right|_{v=o} = -\alpha\, f(u,o) + \beta \frac{d}{du} f(u,o)$$

$$\left.\frac{\partial^2 f(u,v)}{\partial v^2}\right|_{v=0} = -(\sigma_o^2 + \alpha^2) f(u,o) + 2i\alpha\beta \frac{d}{du} f(u,o) + \beta^2 \frac{d^2}{du^2} f(u,o)$$

holds for all real u, where $f(u,v)$ is the ch.f. of the random vector (X,Y).

By multiplying $E(Y|X) = \alpha + \beta X$ and $E(Y^2|X) = \sigma_o^2 + (\alpha + \beta X)^2$ by e^{itX} and taking expectation on both sides with respect to X, the necessity follows. The proof of the sufficiency of the lemma is very similar to Lemma 1.

If the r.v.'s Y and X have mean zero, it is easy to see that $E(Y|X) = \beta X$ in virtue of $E[E(Y|X] = E(Y)$.

Lemma 4. Let (X_i, Y_i), $i = 1,2,\ldots,n$ be n nondegenerate independent (but not necessarily identically distributed) random vectors such that every component of the random vector (X_i, Y_i), $i = 1,2,\ldots,n$, has mean zero and a finite variance and

$$E(Y_i|X_i) = \beta_i X_i \quad Var(Y_i|X_i) = \sigma_{io}^2, \quad i = 1,2,\ldots,n.$$

Let $L_1 = \sum_{i=1}^{n} a_i X_i$ and $L_2 = \sum_{i=1}^{n} b_i Y_i$ be two linear functions with $a_j b_j \neq 0$ $(j = 1,2,\ldots,n)$. Then $E(L_2|L_1) = \beta L_1$ and $Var(L_2|L_1) = \sigma_o^2$ provided that $\frac{b_1\beta_1}{a_1} = \frac{b_2\beta_2}{a_2} = \ldots = \frac{b_n\beta_n}{a_n} = \beta$ is satisfied.

Proof. Let $W_j = a_j X_j$ and $Z_j = b_j Y_j$ for $j = 1,2,\ldots,n$. Then

$$L_1 = \sum_{j=1}^{n} W_j \quad \text{and} \quad L_2 = \sum_{j=1}^{n} Z_j \quad .$$

By Lemma 3, it can be shown that,

$$E(Z_j|W_j) = \beta_j' = \frac{b_j\beta_j}{a_j}$$

$$Var\ (Z_j|W_j) = \sigma_{jo}'^2 = b_j^2\,\sigma_{jo}^2 \quad , \quad i = 1,2,\ldots,n.$$

Let $f_j(u,v)$ and $f_j(u,o)$ denote respectively the ch.f.s of (W_j, Z_j) and W_j, $j = 1,2,\ldots,n$, and similarly $g(u,v)$ and $g(u,o)$ denote that of (L_1, L_2) and L_1 respectively. Then

$$g(u,v) = E\,[\exp(iuL_1 + ivL_2)]$$

$$= E\,[\exp\,(iu \sum_{j=1}^{n} W_j + iv \sum_{j=1}^{n} Z_j)\,] \qquad (i)$$

$$= \prod_{j=1}^{n} f_j\,(u,v)$$

Since $E(Z_j|W_j) = \beta_j' W_j$, $var(Z_j|W_j) = \sigma_{jo}^2$ for $j = 1,2,\ldots,n$, by

Lemma 3, we have

$$\left.\frac{\partial f_j(u,v)}{\partial v}\right|_{v=o} = \beta_j' \frac{df_j(u,o)}{du} \qquad (ii)$$

$$\left.\frac{\partial^2 f_j(u,v)}{\partial v^2}\right|_{v=o} = -\sigma_{jo}'^2 f_j(u,o) + \beta_j'^2 \frac{d^2 f_j(u,o)}{du^2}, \quad j=1,2,\ldots,n\,.$$

Differentiating both sides of (i) with respect to v m times ($m = 1,2$), and then putting $v = o$ and using the relations (ii), we obtain

$$\left.\frac{\partial g(u,v)}{\partial v}\right|_{v=o} = \sum_{j=1}^{n} \beta_j' \frac{df_j(u,o)}{du} \prod_{k\neq j} f_k(u,o) \qquad (iii)$$

$$\left.\frac{\partial^2 g(u,v)}{\partial v^2}\right|_{v=o} = -\prod_{j=1}^{n} f_j(u,o) \sum_{j=1}^{n} \sigma_{jo}'^2 + \sum_{j=1}^{n} \beta_j'^2 \frac{d^2 f_j(u,o)}{du^2} \prod_{k\neq j} f_j(u,o)$$

$$+ \sum_{j\neq k} \beta_j' \beta_k' \frac{df_j(u,o)}{du} \frac{df_k(u,o)}{du} \prod_{\ell\neq j,k} f_\ell(u,o)$$

Letting $v = o$ on both sides of (i), and then differentiating with respect to u m times $(m = 1,2)$, we obtain

$$\frac{dg(u,o)}{du} = \sum_{j=1}^{n} \frac{df_j(u,o)}{du} \prod_{k \neq j} f_k(u,o) \qquad \text{(iv)}$$

$$\frac{d^2 g(u,o)}{du^2} = \sum_{j=1}^{n} \frac{d^2 f_j(u,o)}{du^2} \prod_{k \neq j} f_k(u,o) + \sum_{j \neq k} \frac{df_j(u,o)}{du} \times \frac{d\, f_k(u,o)}{du} \prod_{\ell \neq j,k} f_\ell(u,o).$$

Since $\beta_1' = \beta_2' = \ldots = \beta_n' = \beta$, we have from (iii) and (iv)

$$\left.\frac{\partial\, g(u,v)}{\partial v}\right|_{v=o} = \beta \frac{dg(u,o)}{du}$$

$$\left.\frac{\partial g(u,v)}{\partial v^2}\right|_{v=o} = -\, g(u,o) \sum_{j=1}^{n} \sigma_{o}^{2} + \beta^2 \frac{d^2 g(u,o)}{du^2}$$

By Lemma 3, the proof is complete.

We see from the above result that if there exist two linear functions $L_1 = \sum_{j=1}^{n} a_j X_j$ and $L_2 = \sum_{j=1}^{n} b_j X_j$ with $a_j b_j \neq 0$ $(j = 1,2,\ldots,n)$

where $X_1,\ldots,X_n$ are independent (but not necessarily identically distributed) r.v.'s each having mean zero and finite variance, then $E(L_2|L_1) = \beta$ and $\operatorname{var}(L_2|L_1) = \sigma_o^2$ whenever the relation $b_1/a_1 = b_2/a_2 = \ldots = b_n/a_n = \beta$ is satisfied.

We are now in a position to prove the following theorem .

Theorem 13. (Laha 1957a) Let (X_i,Y_i), $i = 1,2,\ldots,n$ be n non-degenerate independent (but not necessarily identically distributed) random vectors such that $E(X_i) = E(Y_i) = 0$ $E(Y_i|X_i) = \beta_i X_i$ and $\operatorname{var}(Y_i|X_i) = \sigma_{io}^2$ for $i = 1,2,\ldots,n$. Let $L_1 = \sum_{j=1}^{n} a_j X_j$, $L_2 = \sum_{j=1}^{n} b_j X_j$ be two linear functions with $a_j b_j \neq 0$, $(j = 1,2,\ldots,n)$. Then

$$E(L_2|L_1) = \beta \quad \text{and} \quad \operatorname{Var}(L_2|L_1) = \sigma_o^2 \quad \text{iff}$$

(i) each X_j for which $\frac{b_j\beta_j}{a_j} \neq \beta$ is normally distributed while each Y_j and the other X_j's can be arbitrary.

(ii) $$\beta = \Sigma' a_j b_j \beta_j \sigma_j^2 \,/\, \Sigma' a_j^2 \sigma_j^2 \quad \text{and}$$

$$\sigma_o^2 = \sum_{j=1}^{n} b_j^2 \sigma_{jo}^2 + \Sigma' \left(\frac{b_j\beta_j}{a_j} - \beta\right)^2 a_j^2 \sigma_j^2 ,$$

where $\sigma_j^2 = \operatorname{Var}(X_j)$, $(j = 1,2,\ldots,n)$ and Σ' stands for the summation over all indices j for which $\frac{b_j\beta_j}{a_j} \neq \beta$.

Proof: Necessity, With exactly the same notations as those used in Lemma 4, and with the aid of Lemma 3, the following relations can be easily verified.

$$\sum_{j=1}^{n} \beta'_j \frac{df_j(u,o)}{du} \prod_{j\neq k} f_k(u,o) = \beta \sum_{j=1}^{n} \frac{df_j(u,o)}{du} \prod_{k\neq j} f_k(u,o)$$

(13.1)

$$- \prod_{j=1}^{n} f_j(u,o) \sum_{j=1}^{n} \sigma'^2_{jo} + \sum_{j=1}^{n} \beta'^2_j \frac{d^2 f_j(u,o)}{d u^2} \prod_{k\neq j} f_k(u,o) + \sum_{j\neq k} \beta'_j \beta'_k \times$$

$$\frac{df_j(u,o)}{du} \frac{df_k(u,o)}{d u} \prod_{\ell\neq j,k} f_\ell(u,o)$$

$$= - \sigma_j^2 \prod_{j=1}^{n} f_j(u,o) + \beta^2 \sum_{j=1}^{n} \frac{d^2 f_j(u,o)}{du^2} \prod_{k\neq j} f_k(u,o) +$$

$$\sum_{j\neq k} \frac{df_j(u,o)}{du} \frac{d f_k(u,o)}{du} \prod_{\ell\neq j,k} f_\ell (u,o)$$

(13.2)

As before, there exists a neighbourhood in which none of the function $f_j(u,o)$ $(j = 1,2,...,n)$ vanishes. Dividing both sides of (13.1) and (13.2) by $\prod_{j=1}^{n} f_j(u,o)$, we obtain

$$\sum_{j=1}^{n} \beta'_j \frac{df_j(u,o)}{du} \Big/ f_j(u,o) = \beta \sum_{j=1}^{n} \frac{df_j(u,o)}{du} \Big/ f_j(u,o) \qquad (13.3)$$

$$- \sum_{j=1}^{n} \sigma'^2_{jo} + \sum_{j=1}^{n} \beta'^2_j \frac{d^2 f_j(u,o)}{du^2} \Big/ f_j(u,o) + \beta \sum_{j \neq k} \beta'_j \beta'_k \left\{ \frac{df_j(u,o)}{du} \Big/ f_j(u,o) \right\} \times \left\{ \frac{df_k(u,o)}{du} \Big/ f_k(u,o) \right\}$$

$$= - \sigma_o^2 + \beta^2 \left[\sum_{j=1}^{n} \frac{d^2 f_j(u,o)}{du^2} \Big/ f_j(u,o) + \sum_{j \neq k} \left\{ \frac{df_j(u,o)}{du} \Big/ f_j(u,o) \right\} \left\{ \frac{df_k(u,o)}{du} \Big/ f_k(u,o) \right\} \right]$$

(13.4)

(valid in a certain neighbourhood of the origin).

Let $\emptyset_j(u) = \ell n\ f_j(u,o)$ for $j = 1,2,\ldots,n$. Writing (13.3) and (13.4) in terms of the derivatives of $\emptyset_j(u)$, we have

$$\sum_{j=1}^{n} \beta'_j \frac{d\emptyset_j}{du} = \beta \sum_{j=1}^{n} \frac{d\emptyset_j}{du} \qquad (13.5)$$

$$- \sum_{j=1}^{n} \sigma'^2_{jo} + \sum_{j=1}^{n} \beta'^2_j \frac{d^2\phi_j}{du^2} + \left(\sum_{j=1}^{n} \beta'_j \frac{d\phi_j}{du} \right)^2$$

$$= - \sigma_o^2 + \beta^2 \left[\sum_{j=1}^{n} \frac{d^2\phi_j}{du^2} + \left(\sum_{j=1}^{n} \frac{d\phi_j}{du} \right)^2 \right] \tag{13.6}$$

Making use of (13.5), (13.6) further reduces to

$$- \sum_{j=1}^{n} \sigma^2_{jo} + \sum_{j=1}^{n} \beta'^2_j \frac{d^2\phi_j}{du^2} = - \sigma_o^2 + \beta^2 \sum_{j=1}^{n} \frac{d^2\phi_j}{du^2} \tag{13.7}$$

Now, differentiating both sides of (13.5) with respect to u, we have

$$\sum_{j=1}^{n} \beta'_j \frac{d^2\phi_j}{du^2} = \beta \sum_{j=1}^{n} \frac{d^2\phi_j}{du^2} \tag{13.8}$$

Combining (13.7) and (13.8), we have

$$\sum_{j=1}^{n} (\beta'_j - \beta)^2 \frac{d^2\phi_j}{du^2} = \sum_{j=1}^{n} \beta'^2_j \frac{d^2\phi_j}{du^2} - 2\beta \sum_{j=1}^{n} \beta'_j \frac{d^2\phi_j}{du^2} + \beta^2 \sum_{j=1}^{n} \frac{d^2\phi_j}{du^2}$$

$$= \sum_{j=1}^{n} \beta'^2_j \frac{d^2\phi_j}{du^2} - \beta^2 \sum_{j=1}^{n} \frac{d^2\phi_j}{du^2}$$

$$= - \left(\sigma_o^2 - \sum_{j=1}^{n} \sigma'^2_{jo} \right) = c \quad \text{(say)} \tag{13.9}$$

Finally, integrating (13.9) with respect to u, we obtain

$$\sum_{j=1}^{n} \left\{ f_j(u,o) \right\}^{(\beta_j'-\beta)^2} = e^{P(u)}$$

which holds in some neighbourhood of the origin, where P(u) is a quadratic polynomial in u. By Theorem 4 in Chapter I, we conclude that each W_j for which $\beta_j' \neq \beta$ is normally distributed.

Sufficiency. There is no loss of generality in assuming that

$$b_i\beta_i/a_i \neq \beta \text{ for the first } r \text{ pairs } (r \leq n)$$

and $$b_i\beta_i/a_i = \beta \text{ for the remaining } n-r \text{ pairs.}$$

By assumption, the first r r.v'.s , $X_1,\ldots,X_r$ are normally distributed while others are arbitrary. (Using the notations as before, we denote

$$L_1 = W + W_{r+1} + \ldots + W_n$$

$$L_2 = Z + Z_{r+1} + \ldots + Z_n \quad \text{where } W = \sum_{j=1}^{r} W_j,\ Z = \sum_{j=1}^{r} Z_j$$

Now let $f_{wz}(u,v)$ and $f_w(u,o)$ be the ch.f.'s of the distribution of (W ,Z) and the distribution of W respectively, and let $\sigma_j'^2$ denote the variance of W_j, j = 1,2,...,n. Then

$$f_{wz}(u,v) = E\left[e^{iWu+iZv} \right]$$

$$= \prod_{j=1}^{n} f_j(u,v) \tag{13.10}$$

$$\left.\frac{\partial f_{wz}(u,v)}{\partial v}\right|_{v=o} = \sum_{j=1}^{r} \frac{\partial f_j(u,v)}{\partial v} \prod_{j\neq k} f_k(u,v) \Big|_{v=o} \tag{13.11}$$

$$= \sum_{j=1}^{r} \frac{\partial f_j(u,o)}{\partial v} \prod_{j\neq k} f_k(u,o)$$

$$\left.\frac{\partial^2 f_{wz}(u,v)}{\partial v^2}\right|_{v=o} = \sum_{j=1}^{r} \frac{\partial^2 f_j(u,o)}{\partial v^2} \prod_{j\neq k} f_k(u,o)$$

$$+ \sum_{j\neq k}^{r} \frac{\partial f_j(u,o)}{\partial v} \frac{\partial f_k(u,o)}{\partial v} \prod_{\ell\neq j,k}^{r} f_\ell(u,o)$$

By using the fact $E(Y_i|X_i) = \beta_i X_i$, $var(Y_i|X_i) = \sigma_{io}^2$ $(i = 1,2,\ldots,r)$ and Lemma 3, it can be shown as in Lemma 4 that

$$\left.\frac{\partial f_{wz}(u,v)}{\partial v}\right|_{v=o} \Big/ f_w(u,o) = \sum_{j=1}^{r} \beta_j' \frac{df_j(u,o)}{du} \Big/ f_j(u,o) = -\sum_{j=1}^{r} \beta_j' \sigma_j'^2 u \tag{13.12}$$

$$\left.\frac{\partial^2 f_{wz}(u,v)}{\partial v^2}\right|_{v=o} \Big/ f_w(u,o) = -\sum_{j=1}^{r} \sigma_{jo}'^2 - \sum_{j=1}^{r} \beta_j'^2 \sigma_j'^2 + u^2 \left(\sum_{j=1}^{r} \beta_j' \sigma_j'^2 \right)^2 \tag{13.13}$$

since $$f_w(u,o) = \exp\left\{ -\frac{1}{2} u^2 \sum_{j=1}^{r} \sigma_j'^2 \right\}$$

Since W is normal with mean zero and variance $\sum_{j=1}^{r} \sigma_j'^2$, we have

$$\frac{df_{zw}(u,o)}{du} \Big/ f_w(u,o) = -u \sum_{j=1}^{r} \sigma_j'^2 \qquad (13.14)$$

$$\frac{d^2 f_w(u,o)}{du^2} \Big/ f_w(u,o) = -\sum_{j=1}^{r} \sigma_j'^2 + u^2 \left(\sum_{j=1}^{r} \sigma_j'^2 \right)^2 \qquad (13.15)$$

We see from (13.12), (13.13), (13,14) and (13.15) that

$$\left.\frac{\partial f_{wz}(u,o)}{\partial v}\right|_{v=o} = \beta \frac{df_w(u,o)}{du} \quad \text{where}$$

$$\beta = \sum_{j=1}^{r} \beta_j' \sigma_j'^2 \Big/ \sum_{j=1}^{r} \sigma_j'^2$$

and $$\left.\frac{\partial^2 f_{wz}(u,v)}{\partial v^2}\right|_{v=o} \Big/ f_w(u,o)$$

$$= -\sum_{j=1}^{r} \sigma_{jo}'^2 - \sum_{j=1}^{r} \beta_j'^2 \sigma_j'^2 + \beta^2 \left\{ \frac{\partial^2 f_{wz}(u,o)}{\partial u^2} \Big/ f_w(u,o) + \sum_{j=1}^{r} \sigma_j'^2 \right\}$$

$$\Rightarrow \left.\frac{\partial^2 f_{wz}(u,v)}{\partial v^2}\right|_{v=o} = -\sigma_o'^2 f_w(u,o) + \beta^2 \frac{\partial^2 f_w(u,o)}{\partial u^2} \quad \text{where}$$

$$\sigma_o'^2 = \sum_{j=1}^{n} \sigma_{jo}'^2 + \sum_{j=1}^{r} (\beta_j - \beta)^2 \sigma_j'^2 .$$

By Lemma 3, we have

$$E(Z|W) = \beta W, \quad \text{and} \quad \text{var}\,(Z|W) = \sigma_o'^2 .$$

But since $\beta'_{r+1} = \beta'_{r+1} = \ldots = \beta'_n = \beta$, by Lemma 4, we conclude that

$$E(L_2|L_1) = \beta L_1,\ \text{var}(L_2|L_1) = \sigma_o^2 ,$$

where $$\sigma_o^2 = \sum_{j=1}^{n} \sigma_{jo}'^2 + \sum_{j=1}^{r} (\beta'_j - \beta)^2 \sigma_j'^2 .$$

Hence the proof is complete.

Since $E(X|X) = E(X)$ and $\text{var}(X|X) = \text{var}(X)$ for any random variable X we have as a corollary the following.

Corollary 13.1. Let $X_1,\ldots,X_n$ be n independent (but not necessarily identically distributed) nondegenerate r.v.'s each having mean zero and a finite variance $\sigma_j^2 > 0$. Then the necessary and sufficient condition for $E(L_2|L_1) = \beta L_1$ and $\text{var}(L_2|L_1) = \sigma_o^2$ where $L_1 = \sum_1^n a_1 X_i$ and $L_2 = \sum_1^n b_i X_i$ with $a_j b_j \neq 0$ $(j = 1,2,\ldots,n)$, is that

i) each X_j for which $b_j/a_j \neq \beta$ is normally distributed while the remaining X_j's have arbitrary distributions.

ii) $\beta = \Sigma' a_j b_j \sigma^2_j / \Sigma' a^2_j \sigma^2_j$ and $\sigma^2_o = \Sigma' (b_j/a_j - \beta)^2 a^2_j \sigma^2_j$

where the summation runs through all indices j such that $b_j/a_j \neq \beta$.

Multivariate cases of the above results is considered in [73] by Mathai (1967) in which some necessary and sufficient conditions for the existence of rational regression of one stochastic matrix on a number of stochastic matrices are established, and also a series of characterization theorems for the multivariate normal distribution are obtained. We present here a few results proven in [73]. For the proof we refer the reader to [73].

Theorem 14. Let $X_1, \ldots, X_n$ be n independently but not necessary identically distributed stochastic column vectors of order k each with finite covariance matrices. Let $U = a_1X_1 + \ldots + a_nX_n$ and $V = b_1X_1 + \ldots + b_nX_n$ where a's and b's are scalars and $a_jb_j \neq 0$ for $j = 1,2,\ldots,n$. If $E(V|U) = A + bU$ and $var(V|U) = B$ is positive definite and independent of U, then each x_j for which $a_j \neq b\, b_j$ has a multivariate normal distribution.

We note that if the above $X_1, \ldots, X_n$ are replaced by n independently distributed symmetric stochastic matrices (matrix stochastic variables) of order k with each row having finite covariance matrices, then the assertion of the theorem still holds with A representing a matrix of constants, b a scalar and var(V|U) replaced by cov(V|U) = B where B is positive definite and independent of U. Furthermore replacing the scalars a's and b's by nonsingular matrices A's and B's, similar assertion of Theorem 14 can be formulated (see [73]).

CHAPTER V

Characterizations of the Normal Distribution in a Linear Structural Relations and by Properties of Sample Estimations.

The methods of characterizing the normal distribution that have been discussed so far are based on some properties of suitable statistics. In this chapter we shall see that characterizations of the normal distribution can also be done through other considerations. We shall only discuss a few of them, while others not considered here can be found in the bibliography.

5.1. Linear structural relations

To being with, we introduce some model of random (stochastic) variables which is known as linear structural relations or in general linear model. Suppose that we have a set of r.v.'s $X_1,\ldots,X_n$; $Y_1,\ldots,Y_m$; $\epsilon_1,\ldots,\epsilon_n$ such that the following relations are satisfied:

$$X_j = \sum_{j=1}^{m} a_{ji} Y_i + \epsilon_j, \; j = 1,2,\ldots,n,$$

where the $a_{ji}(j = 1,2,\ldots,n, \; i = 1,2,\ldots,m)$ are some real constants. i.e. X_j is a linear function of Y_j's and ϵ_j. The r.v.'s X's,Y's and ϵ's altogether are said to form a stochastic linear structure. The a_{ji} $(j = 1,2,\ldots,n, \; i = 1,2,\ldots,m)$ are always referred as the parameters of the structure. Such model is often encountered in factor analysis mainly used in psychological statistics. We may express the above relation in terms of vector and matrix notations. Accordingly we have

$$X = AY + \epsilon \quad .$$

We shall call the r.v.'s $X_1,\ldots,X_n$ the observable r.v.'s, while the $Y_1,\ldots,Y_m$ the latent variables and $\epsilon_1,\ldots,\epsilon_n$ the error variables . As can be seen from the structure, the distribution of the observable variables is closely related to that of the latent variables, the error variables and the parameters of the structure. A question arises, whether the representation of X's in terms of Y's and ϵ's and $a_{jk}(j = 1,2,\ldots,n,\ \ k= 1,2,\ldots,m)$ is uniquely determined by the latent variables, the error variables and the parameters. i.e., is it possible to have two different structures which have the same distribution for the observable variables? The answer is in the affirmative, if some random variables are normally distributed. Accordingly we define that two structures are equivalent if the distribution of the observable variables is the same in both structures. Linear structural relations also lead to a number of interesting problems such as the investigation of the latent variables or the estimation of the parameters a_{jk} or the identification problem, that is the problem of finding conditions which assure that a parameter is identifiable (A parameter is said to be identifiable if it has the same value in all equivalent structure.)

In this section, we shall only be concerned with the problems of characterizing the latent and the error variables. It appears under the independence of the latent variables and error variables and some restrictions on the parameters that the equivalence of two structures is meaningful only if some latent variables are normally distributed.

Theorem 1. (Rao 1966) Let $X = (X_1,X_2)'$ be a two dimensional random vector such that

$$X_1 = a_{11}Y_1 + \dots + a_{1k}Y_k \quad \text{and} \quad X_1 = b_{11}Z_1 + \dots + b_{1m}Z_m$$
$$X_2 = a_{21}Y_1 + \dots + a_{2k}Y_k \qquad\qquad X_2 = b_{21}Z_1 + \dots + b_{2m}Z_m \tag{1.1}$$

where $Y_1,Y_2,\dots,Y_k$ are independent r.v.'s, $Z_1,Z_2,\dots,Z_m$ are independent r.v.'s and $a_{ij},b_{i\ell}$ $(i = 1,2,\ j = 1,2,\dots,k,\ \ell = 1,2,\dots,m)$ are real constants. Assume that $\binom{a_{1r}}{a_{2r}}$ is not a multiple of any column of the type $\binom{a_{1j}}{a_{2j}}$, $j \neq r$ or of any column of the type $\binom{b_{1j}}{b_{2j}}$, $j = 1,2,\dots,m$. Then the r.v. Y_r is normally distributed.

Proof. Considering the joint characteristic function of X_1,X_2 from the two representations (1.1) taking logarithms and equating them, in terms of the cumulant generating functions $\psi_j(j = 1,2,\dots,k)$, ϕ_j $(j = 1,2,\dots,m)$ of Y_j and Z_j respectively, we have

$$\psi_1(a_{11}u+a_{21}v) + \dots + \psi_k(a_{1k}u+a_{2k}v) = \phi_1(b_{11}u+b_{21}v) + \dots + \phi_m(b_{1m}u+b_{21}v) \tag{1.2}$$

which is valid in a neighbourhood of the origin. There is no loss of generality in assuming that $a_{1r} \neq 0$ and in replacing nonzero a_{1i} and b_{1j} by unity, which only means taking $a_{1i}Y_i$ and $b_{1j}Z_j$ into consideration instead of Y_i and Z_j. Then, by the condition of Theorem 1, the equation (1.2) reduces to

$$\psi_r(u+a_{2r}v) + \eta_1(u+c_{21}v) +\ldots+ \eta_s(u+c_{2s}v) = A(u) + B(v) \quad (1.3)$$

if $a_{2r} \neq 0$, and to

$$\eta_1(u+c_{21}v) +\ldots+ \eta_s(u+c_{2s}v) = \psi_r(v) + B(v)$$

if $a_{2r} = 0$. Here every η is the function obtained adding the ψ-functions and substracting the ϕ functions having a common coefficients for v in (1.2), and $a_{2r}, c_{21}, \ldots, c_{2s}$ can all be taken to be different. By Lemma 3 in Chapter IV, ϕ_r is a polynomial of degree at most S in a neighbourhood of the origin. Hence, in view of Marcinkiewicz's theorem, if Y_r is nondegenerate, it must be normally distributed.

The following result Rao [82] is noteworthy. Suppose that a p-dimensional random vector X takes the two representations X = AY and X = BZ, where A and B are matrices such that no two columns of either A or B are equivalent in the sense that one is a multiple of the other, and Y and Z are random vectors of independent nondegenerate r.v.'s. It may be assumed that A is a p x r matrix and B is a p x s matrix without any restrictions on ranks of A and B. Then the following are true: i) The ranks of A and B are the same. ii) If the i^{th} column of A is not a multiple of any column of B, then i^{th} component of Y is normally distributed. iii) If the i^{th} column of A is a multiple of the j^{th} column of B, then the cumulant generating function of the i^{th} component of Y differs from that of j^{th} component of Z by a polynomial in a neighbourhood of the origin. Therefore if none of the columns of A is a multiple of any columns of B, then X has a p-variate normal distribution.

The next theorem is related with factor analysis models.

Theorem 2. (Rao 1966). Let $X = AY + \epsilon_1$ and $X = BZ + \epsilon_2$ be two representations where A is a p x r matrix of rank m, B is a p x s matrix of rank n, both having no equivalent column, ϵ_1 and ϵ_2 are vectors of error variables, and Y and Z denote vectors of latent variables. Then if $m < n$ we have

i) at least $(n-m)$ of the latent variables in the representation $X = BZ + \epsilon_2$ are normally distributed, ii) there are linear functions $L'\epsilon_1$, $L_2'\epsilon_2$ of the error variables which differ by a nondegenerate normal component, (a variable is said to have a normal component if its distribution can be expressed as the convolution of two distributions of which one is normal. iii) The cumulant generating functions of at least one of the pairs (ℓ_i, ℓ_i') (ℓ_i and ℓ_i' are respectively the i^{th} component of ϵ_1 and ϵ_2) $i = 1,2,\ldots,p$ differ by a second degree polynomial in a neighbourhood of the origin.

Since the condition $m < n$ implies that at least (n-m) columns of B are not multiples of columns of A, as seen above, the corresponding r.v.'s are then normally distributed which proves (i). The proofs of (ii) and (iii) are omitted here. For the proofs we refer the reader to the original paper [82].

5.2. Characterization of the normal distribution by properties of sample estimations.

By a Gauss-Markoff Model we mean a linear model

$$Y_i = a_{i1}\theta_1 + \ldots + a_{im}\theta_m + \epsilon_i, \quad i = 1,2,\ldots,n, \qquad (3.1)$$

such that $E(Y_i) = a_{i1}\theta_1 + \ldots + a_{im}\theta_m$, $\mathrm{var}(Y_i) = \sigma^2 (i = 1,2,\ldots,n)$ and $E(\epsilon_i) = 0$ $(i = 1,2,\ldots,n)$, where $Y_1,\ldots,Y_n$ are independent (sometimes "uncorrelated" may be preferable), $\theta_1,\ldots,\theta_m$ are real unknown parameters and $a_{i1},\ldots,a_{im}$ $(i = 1,2,\ldots,n)$ are known constants i.e., unlike in linear structural relations where $\theta_1,\ldots,\theta_m$ are random variables. The theory of least squares is concerned with the problem of estimating unknown parameters $\theta_1,\ldots,\theta_m$ in a linear model. The essentials of the theory are found in the works of Gauss (1809) and Markoff (1900). A unified approach to the least square theory covering all the practical situations has been suggested, using the concept of a generalized inverse (g-inverse) of a singular matrix in [81]. The least squares estimator of an estimable parametric function $q_1\theta_1 + \ldots + q_m\theta_m$ (see pp 182 [81]) is the linear function of $Y_1,\ldots,Y_n$ which has the minimum variance in the class of all linear unbiased estimators. This means that the least squares estimator depends only on the coefficients a_{ij} $(i = 1,2,\ldots,n,\ \ j = 1,2,\ldots,m)$ and not on the exact distribution of $Y_1,\ldots,Y_n$. However, it is shown in [79] that if $Y_1,\ldots,Y_n$ have normal distributions with a common variance independent of $\theta_1,\ldots,\theta_m$, then the least square estimator of $q_1\theta_1 +\ldots+ q_m\theta_m$ has minimum variance in the class of all unbiased estimators. In this section, we shall see that the converse of this proposition also holds under certain conditions.

Let us assume that $(Y_i-E(Y_i))$, $i = 1,2,\ldots,n$ have the same distribution $F_{\theta\phi}$ which may depend on $\theta = (\theta_1,\ldots,\theta_m)$ and certain other unknown parameter ϕ. We first notice that the mean of the distribution $F_{\theta\phi}$ is zero whatever θ and ϕ may be.

Theorem 3. Let $Y_i = a_{i1}\theta_1 +\ldots+ a_{im}\theta_m + \epsilon_i$, $i = 1,2,\ldots,n$ be a Gauss-Markoff Model. Assume that the rank of (a_{ij}) is unity and that $Z = b_1Y_1 +\ldots+ b_nY_n$ is the least squares estimator of the essentially unique estimable linear parametric function $q_1\theta_1 +\ldots+ q_m\theta_m$. Further assume that $F_{\theta\phi}$ has finite moments up to order $2s$ for each θ,ϕ and that, $b_1,\ldots,b_n$ are all different from zero (without loss of generality). If $Z = b_1Y_1 +\ldots+ b_nY_n$ has minimum variance in the class of all unbiased estimators which are polynomials of order s or less and the vector $(b_1,\ldots,b_n)$ is not a multiple of a vector with entries only ± 1 as its elements, then Y_i agrees with a normal distribution up to moments of order $(s+1)$.

We first note that since the rank of (a_{ij}) $i = 1,2,\ldots,n$, $j = 1,2,\ldots,m$ is unity, there is only one independent estimable linear parametric function, and there exist $(n-1)$ linear functions

$$Z_j = c_{j1}Y_1 + \ldots + c_{jn}Y_n \quad j = 1,2,\ldots,n-1 \quad \text{such}$$

that

$$E(Z_j) = 0, \; E(Z_j Z) = 0, \quad E(Z_j Z_i) = 0, \quad i \neq j.$$

Since $E(Z_j Z) = 0$ and $\mathrm{Var}(Z_j Z) < \infty$ if $s \geq 2$, it follows from a result by Rao (see [81] pp.257, (i)) that if Z has minimum variance in the class of all unbiased estimators of the second degree, then

$$E[Z(Z_j Z)] = E(Z_j Z^2) = 0 \quad .$$

Similarly, we have

$$E(Z_j Z^r) = 0, \quad r = 1,2,\ldots,s \; , \quad j = 1,\ldots,n-1 \; .$$

Consider

$$\psi(t) = E\left\{ iZ_j \, e^{it[Z-E(Z)]} \right\}$$

$$= \; [c_{j1}\phi(b_1 t) + \ldots + c_{jn}\phi(b_n t)] \; f(b_1 t)\ldots f(b_n t) \quad .$$

Here $f(t)$ denotes the ch.f. of $F_{\theta\phi}$ and $\phi(t) = \dfrac{f'(t)}{f(t)}$ (in a neighbourhood of the origin). Since the moments of $F_{\theta\phi}$ exist up to order $2s$, $\phi(t)$ is differentiable s times. Thus differentiating $\phi(t)$ r times ($r \leq s$) and then putting $t = 0$, in view of $E(Z_j Z^r) = 0$ $r = 1,2,\ldots,s$, we obtain

$$(\sum_i c_{ji} b_i^{\,r}) k_{r+1}(\theta,\emptyset) = 0, \quad r \leq s \quad \text{and} \quad j = 1,2,\ldots,n-1,$$

where $k_{r+1}(\theta,\emptyset)$ is the $(r+1)^{th}$ cumulant of $F_{\theta\emptyset}$. From the above equation, we have either

$$\sum c_{ji} b_i^{\,r} = 0$$

or

$$k_{r+1}(\theta,\emptyset) = 0 \quad \text{or both.}$$

Since $E(Z_j Z) = 0$, we have $\sum_i c_{ji} b_i = 0$, $j = 1,2,\ldots,n-1$.

Then $\sum_i c_{ji} b_i^{\,r} = 0 \quad j = 1,2,\ldots,n-1$ implies

$$b_j^{\,r} = \lambda_r b_j\ , \quad j = 1,2,\ldots,n,$$

which holds only if $r = 1$ or $b_j = 0$, or proportional to ± 1, $j = 1,2,\ldots,n$. Hence we must have

$$k_{r+1}(\theta,\emptyset) = 0 \quad \text{for} \quad r = 1,2,\ldots,s.$$

This completes the proof.

From the above theorem, we see that if all the moments of $F_{\theta\emptyset}$ exist, then under the conditions of theorem 3 on the coefficients $b_1,\ldots,b_n$ a necessary condition for the least squares estimator to be the minimum variance unbiased estimator is that the variables are normally distributed. This is the main theorem proven in the earlier paper by Rao (1959).

The case that the rank of (a_{ij}) is greater than one is also considered in [83] in which a similar result is obtained under some sightly different condition on the matrix (a_{ij}).

In conclusion, we mention some other types of characterization of the normal distribution that have been done so far.

Ferguson (1962) has studied the families of distributions involving location and scale parameters and he obtained several characterizations of the normal distribution, one of which extends the result by Teicher (1961) on the characterization of the normal distribution by maximum likelihood estimate of the location parameter in a family of distributions involving location parameter. Actually the above mentioned result by Teicher has long been known in literature and can be found in the work of Gauss. Patil and Seshadri (1964) considered the problem of characterizing the Binomial, Exponential, Normal and Power series from the conditional distribution of X given $X+Y$ in a bivariate case. Mathai (1967) dealt with the problem of the structural properties of the conditional distributions and obtained several characterization theorems by assigning specific form for the conditional distribution of a random variable given a set of random variables.

We have seen that characterizations of the normal distribution can be achieved from various aspects and through various methods, among which those considered in the second to the fourth chapters seem more important and have been investigated in great detail . Others are far from complete. All these works may have thrown some lights in the determination of an underlying theoretical distribution from the knowledge of either properties of statistics or of some other information regarding the experimental or other situations. These results or the characteristic properties enhances the status of the Normal distribution as a very important distribution in statistical literature.

BIBLIOGRAPHY

1. Aitchison, J. (1963), Inverse distributions and independent Gamma-distributed products of random variables, Biometrika 50, 505-508

2. Basu, D. (1951), On the independence of linear functions of independent chance variables, Bull. Int. Statist. Inst. 33,83-96.

3. Basu, D. and Laha, R.G. (1954), On some characterizations of the normal distribution, Sankhya ser.A 13, 359-362.

4. Basu, D. (1955), On statistics independent of a complete sufficient statistic, Sankhya ser.A 15, 377-380.

5. Basu, D. (1956), A note on the multivariate extension of some theorems related to the univariate normal distribution, Sankhya ser.A 17, 221-224.

6. Basu, D. (1958), On statistics independent of sufficient statistics, Sankhya ser.A 20, 223-226.

7. Bernstein, S. (1941), Sur une propriete caracteristique de la loi de Gauss, Trans. Leningrad. Polytech. Inst. 3, 21-22.

8. Cacoullos, T. (1967), Characterizations of normality by constant regression of linear statistics on another linear statistic, Ann. Math. Statist. 38, 1894-1898.

9. Cacoullos, T. (1967), Some characterizations of normality, Sankhya ser. A 30, 399-404.

10. Chung, K.L. (1968), A course in Probability Theory, Harcourt Brace and World Inc., New York.

11. Cramer, H. (1936), Über eine Eigenschaft der Normalen Verteilungs-funktion, Math. Zeitschrift, 41, 405-414.

12. Cramer, H. (1937), Random variables and Probability Distributions, Cambridge Tracts in Mathematics and Math. Phy. 36, Cambridge University Press.

13. Cramer, H. (1946), Mathematical methods of Statistics, Princeton Univ. Press.

14. Daly, J.F. (1946), On the use of the sample range in an analogue of student's t-test, Ann. Math. Statist. 17, 71-72.

15. Darmois, G. (1951), Sur une propriete caracteristique de la loi de probabilite de Laplace, C.R. Acad. Sci. Paris, 232, 1999-2000.

16. Darmois, G. (1951), Sur diverses propriete's characteristiques de la loi de probabilite de Laplace-Gauss, Pro. Int. Stat. Conf. India.

17. Darmois, G. (1953), Analysis generale des liaisons stochastiques etade particuliere de l'analyse factorielle lineaire, Rev. Inst. Internat. Stat. 21, 2-8.

18. Dynkin, E.B. (1951), Necessary and sufficient statistics for a family of probability distributions. Uspehi, Mat. Nauk (N.S.) 6 No. 1 (41) 68-90. (Eng. transl. in Selected Trans. in Math. Statist. and Probability 1, 17-40).

19. Evgrafov, M.A. (1966), Analytic functions, W.B. Saunders Co., Philadelphia.

20. Ferguson, F.S. (1962), Location and scale parameters in exponential families of distributions, Ann. Math. Statist. 33, 986-1001.

21. Fix, E. (1946), Distributions which leads to linear regressions First Berkely Sym. on Math. Statist. and Probability, 79-91.

22. Fisher, R.A. (1925), Applications of student's distribution, Metron 5, No. 3, 90-104.

23. Frechet, M. (1951), Generalizations de la loi de probabilite de Laplace, Annals de L'institute Henri Poincare 12, 1-29.

24. Geary, R.A. (1936), Distribution of student's ratio for non-normal samples. Jour. Roy. Statist. Soc., Suppl. Ser. B. 3, 178-184.

25. Geisser, S. (1956), A note on the normal distribution, Ann. Math. Statist. 27, 858-859.

26. Ghurge, S.G. and Olkin,I. (1962), A characterization of the multivariate normal distribution, Ann. Math. Statist. 33, 533-541.

27. Gnedenko, B.V. (1948), On a theorem of S.N. Bernstein, Izvestiya Akad. Nauk. SSSR (ser. Mat) 12, 97-100.

28. Gordon, F.S. (1968), Characterizations of populations using regression properties,Ph.D. Thesis, McGill Univ. Montreal, Canada.

29. Kac, M. (1939), On a characterization of the Normal distribution, Amer. Jour. Math. 61, 726-728.

30. Kagan, A.M. (1966), On the estimation theory of location parameter, Sankhya ser. A. 28, 335-352.

31. Kagan, A.M., Linnik, Yu, V. and Rao, R.C. (1966), On a characterization of the normal law based on a property of the sample average, Sankhya ser. A. 28, 405-406.

32. Kagan, A.M. and Shalaevskii, O.V. (1967), A characterization of the normal law by the property of partial sufficiency, Theory of probability and its applications 12, 512-514.

33. Kawata, T. and Sakamoto, H. (1949), On the characterization of the normal population by the independence of the sample mean and the sample variance, Jour. Math. Soc. Japan, 1,111-115.

34. King, E.P. and Lukacs, E. (1954), A property of the normal distribution, Ann. Math. Statist. 25, 389-394.

35. Koopmans, T.C. and Reier sØl , o. (1950), Identification of structural characteristics, Ann. Math. Statist. 21, 165-187.

36. Kotlarski, I. (1960), On random variables whose quotient follows the Cauchy law, Colloquium Math. 7, 277-284.

37. Kotlarski, I. (1966), On characterizing the normal distribution by student's law, Biometrika 53, 603-606.

38. Laha, R.G. (1953), On an extension of Geary's theorem, Biometrika 40, 228-229.

39. Laha, R.G. (1955) On characterizations of probability distributions and statistics, Ph.D. Thesis submitted to Calcutta University.

40. Laha, R.G. (1956a), On the stochastic independence of two second degree polynomial in normally distributed variates, Ann. Math. Statist. 27, 790-796.

41 Laha, R.G. (1956b), On the stochastic independence of a homogeneous quadratic statistic and the mean, Vestnik Leningrad, Univ. 11, 25-32.

42. Laha, R.G. (1957a), On a characterization of the normal distribution from properties of suitable linear statistics, Ann. Math. Statist. 28, 127-139.

43. Laha, R.G. (1957b), On some characterization problems connected with linear structural relations, Ann. Math. Statist. 28, 405-414.

44. Laha, R.G. (1959a), On the laws of Cauchy and Gauss, Ann. Math. Statist. 30, 1165-1174.

45. Laha, R.G. (1959b), On a class of distribution functions where the quotient follows the Cauchy law. Trans. Amer. Math. Soc. 93,205-215.

46. Laha, R.G. Lukacs, E. and Newman, M. (1960), On the independence of a sample central moment and the sample mean. Ann. Math. Statist. 31, 1028-1033.

47. Laha, R.G. and Lukacs, E. (1960), On a problem connected with quadratic regression, Biometrika 47, 335-339.

48. Laha, R.G. and Lukacs, E. (1964), Applications of characteristic functions, Charles Griffin and Co., London.

49. Laha, R.G. and Lukacs, E. (1964), On identically distributed stochastic integrals. Trans. 3rd Prague conference on Information Theory, Statistical Decision Functions, Random processes, June 1962. Publ. House Czechoslovak Acad. Sci.

50. Linnik, Yu. V. (1952), Linear statistics and the normal distribution law, Doklady Adad. Nauk SSSR (N.S) 83, 353-355. (Eng. Transl.in Selected Transl. in Math. Statist. and Probability).

51. Linnik, Yu, V. (1953), On some identically distributed statistics, Doklady Akad. Nauk. SSSR (N.S) 89, 9-11. (Eng.Transl.in Selected Transl. in Math. Statist. and Probability).

52. Linnik, Yu. V. (1954), Independently and equally distributed statistics and some statistical and analytical problems connected with them. Mimeographed notes, Calcutta.

53. Linnik, Yu. V. and Zinger, A.A. (1955), On an analytic extension of a theorem of Cramer and its application, Vestnik Leningrad Univ. 10, 51-56.

54. Linnik, Yu. V. (1956a), On polynomial statistics in connection with the analytical theory of differential equation, Vestnik Leningrad, Univ. 11, 35-48. (Eng. Transl. in Selected translation in Math. Statist. and Probability, 1,191-206.)

55. Linnik, Yu, V. (1956b), On the question of finding a general distribution from the distribution of a statistic, Teorige veroyata i primen, 1, 466-478. (Eng. Transl. in Theory of probability and its applications 1, 422-434).

56. Linnik, Yu, V. (1956c), A remark on Cramer's theorem on the decomposition of the normal law, Teoriya veroyata i. primen 1, 479-480. (Eng. Transl. in Theory of probability and its applications 1. 435-436.

57. Linnik, Yu, V. and Zinger, A.A. (1957), On a class of differential equattions with an application to certain questions of regression theory, Vestnik Leningrad Univ. Ser. Mat. Mech. Astr. 12, 121-130. (Eng. Transl. in Selected trans. in Math. Statist. and Probability, 3.)

58. Linnik, Yu, V. (1960). Decomposition of probability distributions. Eng. Transl., 1964, Oliver and Boyd.

59. Linnik, Yu, V. and Zinger, A.A. (1964), A characteristic property of the normal distribution, Theory of probability and its applications 9, 624-626.

60. Linnik, Yu, V. and Zinger, A.A. (1964), On polynomial statistics for the normal and related laws, Theory of probability and its applications 9, 494-497.

61. Loève, M. (1963), Probability Theory (3rd Ed.) D. Van Nostrand Co., Inc. New York.

62. Lukacs, E. (1942), A characterization of the normal distribution, Ann. Math. Statist. 13, 91-93.

63. Lukacs, E. (1952), The stochastic independence of symmetric and homogeneous linear and quadratic statistics, Ann. Math. Statist. 23, 442-449.

64. Lukacs, E. (1955), Applications of Faa di Bruno's formula in mathematical statistics, Ann. Math. Monthly 62, 340-348.

65. Lukacs, E. (1956), Characterization of populations by properties of suitable statistics, Proc. 3rd Berkeley Symp. on Math. Statist. and Prob. Theory, 2, 195-214.

66. Lukcas, E. (1960), Characteristic functions, Chalres Griffin and Co., London.

67 Lukacs, E. (1962), On the characterization of a family of populations which includes the Poisson population, Ann. Univ. Sci. Budapest, Sect. Mat. 3-4, 159-175.

68 Mamay, L.V. (1960), On the theory of characteristic functions, Vestnik Leningrad Univ. 15, 85-89, (Eng. Transl. in Selected Transl. in Math. Statist. and Probability 4, 153-170).

69 Marcinkiewicz, J. (1939), Sur une propriete de la loi de Gauss Math. Zeitschrift 44, 612-618.

70. Mathai, A.M. (1966), Some characterizations of the one parameter family of distributions, Canad. Math. Bull. 9, 95-102.

71 Mathai, A.M. (1966), On multivariate exponential type distributions, Jour. Indian Statist. Assoc. 4, 143-154.

72. Mathai, A.M. (1967), On the structural properties of the conditional distributions, Canad. Math. Bull. 10, 239-245.

73. Mathai, A.M. (1967), On a problem of regression, South African Statist. Jour. 1, 43-48.

74. Mauldon, J.G. (1956), Characterizing properties of statistical distributions, Quar. Jour. Math. Oxford. sec. 2,7, 155-160.

75. Neyman, J. (1951), Existence of consistent estimates of the directional parameter in a linear structural relation between two variables, Ann. Math. Statist. 22, 497-512.

76 Patil, G.P. and Seshadri, V. (1963), A characterization of a bivariate distribution by the marginal and the conditional distributions of the sample component, Ann. Inst. Statist. Math., Tokyo, 15, 215-221.

77 Patil G.P. and Seshadri, V. (1964), Characterization theorems for some univariate probability distributions, Roy. Statist. Soc. Ser. B. 26, 286-292.

78 Prokhorov, Yu, V. (1965), On a characterization of a class of probability distributions by distributions of some statistics, Theory of probability and its applications 10, 438-445.

79 Rao, C.R. (1952), Some theorems on minimum variance estimation, Sankhya ser. A, 12, 27-42.

80 Rao, C.R. (1959), Sur une caracterisation de la distribution normal etablie d'apress une propriete optimum des estimations linearies le Calcul des Prob. et ses Application LXXXVII, 165-172, Coll. Internat. Du. Cent. Nat. De la Recherche Sci.

81. Rao, C.R. (1965), Linear statistical inference and its applications, John Wiley and Sons Inc., New York.

82 Rao, C.R. (1966), Characterization of the distribution of random varaibles in linear structural relations. Sankhya ser. A, 28,251-260.

83 Rao, C.R. (1967), On some characterizations of the normal law, Sankhya ser. A, 29, 1-14.

84 Rao, J.N.R. (1958), A characterization of the normal distribution, Ann. Math. Statist. 29, 914-919.

85 Shimizu, R. (1961), A characterization of the normal distribution, Ann. Inst. Statist. Math., Tokyo, 13, 53-56.

86 Shimizu, R. (1962), Characterization of the normal distribution Ann. Inst. Statist. Math., Tokyo, 14, 175-178.

87 Skitovich, V.P. (1953), On a property of the normal distribution Doklady Akad. Nauk SSSR, 89, 217-219.

88 Skitovich, V.P. (1954), Linear combinations of independent random variables and the normal distribution law, Izu. Adad. Nauk SSSR (ser. Mat) 18, 185-200 (Eng. Transl. in Selected Transl. in Math. Statist. and Probability 2, 211-228).

89 Skitovich, V.P. (1956), On a characterization of Brownian motion, Teoriya veroyant i. primen. 1, 361-364. (Eng. Transl. in Theory of probability and its applications 1, 326-328).

90 Steck, G.P. (1958), A uniqueness property not enjoyed by the normal distribution, Ann. Math. Statisti. 29, 604-606.

91 Steyn, H.S. (1957), On regression properties of discrete system of probability functions, Indag. Math. 19, 119-127.

92 Steyn, H.S. (1960), On regression properties of multivariate probability functions of Pearson types, Indag. Math. 27, 303-311.

93 Tambankar, M.V. (1967), A characterization of normality, Ann. Math. Statist. 38, 1924-1927.

94 Teicher, H. (1956), Identification of a certain stochastic structure, Econometrica, 24, 172-177.

95 Teicher, H. (1961), Maximum likelihood characterization of distribution, Ann. Math. Statist. 32, 1214-1222.

96 Tweedie, M.C.K. (1946), The regression of the sample variance on the sample mean, Jour. London Math. Soc. 21, 22-28.

97 Vershik, A.M. (1964), Some characteristic properties of Gaussian stochastic processes, Theory of probability and its applications, 9, 353-356.

98 Wolfowitz, J. (1952), Consistent estimates of the parameters of a linear structural relation, Skaud Aktuarietidskrift, 35, 132-151.

99 Zinger, A.A. (1951), On independent samples from normal populations, Uspeli Matem. Nauk (N.S.) 6 No.5, 45, 172-175.

100 Zinger, A.A. (1956), On a problem of A.N. Kolmogorov, Vestnik Leningrad, G.V. 53-56.

101 Zinger, A.A. (1958), Independence of quasi-polynomial statistics and analytic properties of distributions, Teoriya veroyata i. primen,3, 265-284. (Eng. Transl.in Theory of probability and its applications 3, 247-265).

102 Zinger, A.A. On the distribution of polynomial statistics in samples from a normal population, Doklady Akad. Nauk SSSR, 149, 1, 20-21 (In Russian).

Index of Symbols

ch.f.'s	=	characteriztic functions
c.g.f's	=	cumulant generating functions
d.f.	=	distribution function
i.i.d.	=	independently and identically distributed
r.v.'s	=	random variables
$a \in I$	=	a is an element of the set I
$\Rightarrow$	=	logical implication
Σ	=	summation sign
$\int$	=	integral sign
Var(X)	=	$E[(X-E(X))^2]$ = variance of X
iff	=	if and only if
R^1	=	real line
R_n	=	n-dimensional Euclidean space
pp	=	page
a.e.	=	almost everywhere
ℓn	=	log = logarithm