STUDY OF ROBUSTNESS AND SOME RELATED PROBLEMS - A NUMERICAL APPROACH

PIJUSH DASGUPTA

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree of the Indian Statistical Institute

ACKNOWLEDGEMENT

I wish to espress my deep gratitude to Professor Jogabrata Roy for his constant help and encouraging guidance during the preparation of this thesis.

I am also grateful to the Indian Statistical Institute for providing me with all facilities of research.

Finally, I wish to thank Shri Arun Das for his careful typing of the manuscript, and my friends at the Computer Science Unit, Indian Statistical Institute, Calcutta for the help received from them at various stages of the work.

Pijush Dasgupta
Pijush Dasgupta

CONTENTS

CHAPTER 1

INTRODUCTION

1.0 General

All commonly used parametric test procedures are derived under certain restrictive assumptions regarding the form of the parent population. These assumptions are usually made to simplify the problem and to arrive at an elegant solution. Some assumptions usually made are normality of the parent distribution of the basic variables and simple randomness of the nature of sampling. Another assumption often made - as in tests of linear hypotheses in a Gauss-Markoff model - is homoscedasticity of the variables. In practice, in most cases, there is no way of knowing a priori whether all the assumptions are valid, nor is it practicable in many situations to test on the basis of a few sample observations, the validity of the assumptions. Hence it is important to know to what extent a procedure is insensitive to variations from underlying assumptions. This insensitivity of a procedure to an assumption is called its 'robustness' [Box (1953)] with respect to that assumption.

Obviously, a study of robustness cannot be exhaustive for the simple reason : 'assumptions can be violated in many more ways than they can be satisfied' [Scheffe (1959)]. Conclusions that are derivable are of a very general nature and 'standards of rigour possible in deducing a mathematical theory from certain assumptions can not be maintained in deriving the consequences of departures from assumptions' [Scheffe (1959)].

In order to study the robustness of a test procedure it is necessary to obtain the distribution function of related test statistic after relaxing the assumptions. Since assumptions can be relaxed in different directions it is often impossible to obtain an exact expression for the distribution function in this relaxed condition. Even in cases where exact expression is obtainable, it is sometimes too complicated to be amenable to numerical computation. Thus it seems that the only tools that are available to us in this situation are heuristic approximations and the technique of simulation. Both these methods have their limitations. A heuristic approximation has the disadvantage that useful error bounds are not available in general and the only way of judging the efficiency of an approximation of this type is by numerical comparison with known results in specific cases. On the other hand, for any effective conclusion the application of the simulation technique requires a large number of so called random numbers, the generation and testing of which itself gives rise to many difficulties. But the problem of studying robustness of the well known statistical procedures is important, and, in the absence of any rigorous mathematical tool for handling this problem, we have no other alternative than to depend on heuristic approximations and simulation inspite of their limitations.

One of the purposes of this thesis is to develop some tools for the study of the robustness properties of a number of well known test procedures. The test procedures studied include student's t test for examing whether the population mean has a specified value, Fisher's t test for equality of two means, analysis of variance for linear hypotheses and the test of independence using the sample correlation coefficient - all based on the primary assumption of normality. A number of heuristic approximations are obtained for the distribution functions of these statistics, by using the first few terms of formal orthogonal polynomial series approximations. We have attempted to assess, in some cases, the accuracy of these approximations by numerical comparisons either with known exact results or with results obtained by simulation.

The other problem investigated in this thesis is concerned with the use of non-optimum statistical procedures, which may be necessary because of cost and other complications. For example one may choose to 'count' instead of 'measure' and test independence of two statistical variables by using a contingency chi-square statistic and not the correlation coefficient even when measurements are known to be normally distributed. In such situations, it is important to know how much information is lost because of the use of the non-optimum procedure. We present here also binomial test procedures for independence based on a simple dichotomy which are likely to be convenient in certain situations.

1.1 A brief review of the previous related work

1.1.1 Approximations to distribution functions using formal series expansions

In order to study the property of robustness of a statistical procedure, it is necessary to know the distribution of the relevant statistic under different types of departures from the basic assumptions. When the distribution can not be obtained exactly, one seeks an approximation. If we assume that characteristics of a probability distribution are displayed, at least approximately, by the first few moments of the corresponding random variable then an attempt can be made to approximate the distribution by retaining a finite number of terms of a formal series expansion where coefficients or parameters of the expansion are obtained in terms of the moments of the distribution.

It has been observed by Cramer (1962) about the Gram-Charlier expansion that in real life validity of this expansion from the consideration of convergence can be justified only for a small class of distributions. In fact, many of the important distributions occurring in statistics are not included in this class. This observation seems to be true in general, about all formal series expansions of density functions. However, as has been pointed by Cramer (1962) 'in practical applications it is in most cases of little value to know the convergence properties of our expansion.

What we really want to know is whether a small number of terms - usually not more than two or three - suffice to give a good approximation to density and distribution functions. If we know this to be the case, it does not concern us much whether the infinite series is convergent or divergent. And conversely, if we know that a series is convergent, this knowledge is of little practical value if it will be necessary to calculate a large number of terms in order to have the series determined to a reasonable approximation.'

Finite term series expansions have been used by Gayen (1949, 1950a, 1950b, 1951), Tiku (1963a, 1963b, 1964), Srivastava (1958, 1959), Roy and Tiku (1962), Roy and Mohammed (1964), Durbin and Watson (1951), Khamis (1960), Roy (1965) and others including the author (1968b, 1968c). The approach due to Gayen and also used by Srivastava is essentially this. They start with a Gram-Charlier expansion for the population density function. The distribution of a statistic is then obtained by integrating the expansion for the approximate density over the relevant region. This is possible when population moments are known. The approach presented in this thesis and also that of Tiku, Roy and Tiku and Roy and Mohammed is different in the sense that here approximations are sought directly for the density function of the statistic. This requires the knowledge of first few moments (or in the case of joint distributions, first few joint moments) of the statistic. It may be pointed here that though the exact distribution of a statistic is difficult to obtain, in many cases it is still possible to compute the moments of the statistic without much difficulty.

Our contribution in approximating distribution function is summarised in section 1.2.1 and presented along with already known approximations (for completeness) in chapter two.

1.1.2 Studies on the robustness of Analysis of Variance

The analysis of variance (ANOVA) is a powerful procedure for testing the significance of difference among means under certain assumptions. Studies on the robustness properties of this procedure have been made by Cochran (1947), Tukey (1956, 1957), Welch (1938), Box (1954a, 1954b), Hammersley (1949), Zackrisson (1959), Hooke (1956), Chakravarti (1967), Box and Anderson (1954, 1955), Pearson (1931), Pitman (1937), Bartlett (1935), Hsu (1938), Geary (1936), Gayen (1949, 1950a, 1950b), David and Johnson (1951a, 1951b, 1952), Horsnell (1953), Collier and Baker (1966, 1963), Donaldson (1967), Srivastava (1958, 1959), Bannerjee (1962), Atiqullah (1962, 1964), Tiku (1963a, 1963b, 1964), Gurland and McCullough (1962) and others. Much of the work prior to 1959 has been reviewed in Scheffé (1959). The general conclusion is that ANOVA procedure, if normality assumption is violated, is more or less unaffected in inferring about means. The same is true concerning inequality of variances when normality assumption is true and samples from different groups are of same size. The studies that have been made mainly concern themselves with the first kind of error or the level of significance. Notable exceptions are studies by Tiku (1964) and Srivastava (1958, 1959). Srivastava (1958, 1959) has investigated the power of t and F tests under deviations of normality, and Tiku (1964) has considered the power of one-way classification F test. Horsnell (1953) has determined the power of ANOVA procedure when there are four groups, and the group variances are not equal. Donaldson (1967) has studied the effects of extreme departures from two underlying assumptions namely, normality and homoscedasticity - on the power of ANOVA test by model sampling experiments. David and Johnson (1951a, 1951b, 1952) have studied the robustness of ANOVA F test in non-normal samples, and have suggested a procedure for investigating the effect of non-normality and heterogeneity of variances on tests of general linear hypothesis, and obtained certain general formuale to facilitate the application of the method to random and mixed models. Empirical investigation into the distribution of F in non-normal samples has also been done by Hack (1958).

If the equality of variance assumption is violated the distribution of the F statistic depends on the variances of the basic variables as nuisance parameters. In the case of one way classification ANOVA, various approximations and exact expressions for the distribution of the F statistic have been obtained by Quensel (1947), David and Johnson (1951b), Horsnell (1953), Box (1954a, 1954b), Bannerjee (1962) and others.

Chakravorti (1967) has considered the effect of violation of the assumption of equality of variances by introducing a-priori distribution of the population variances.

Studies on the robustness properties of ANOVA have been mostly confined to the violation of assumptions of normality and homoscedasticity. The effect of error in the expectation model on the procedure for testing a set of linear hypotheses in a Gauss-Markoff set-up seems to be still largely unexplored. Rao and Mitra (1968) have obtained a set of necessary and sufficient conditions on the model so that the best linear unbiased estimates (BLUEs) of the estimable parametric functions under the old expectation model remain so under the changed one.

It is known that [Tang (1938)] usual variance ratio statistic F employed in ANOVA follows a doubly noncentral F distribution when the true expectation model is different from the assumed one. Madow (1948), Weibull (1953) and Scheffe (1959) have considered the application of this distribution in finding the power functions for ANOVA tests in which interaction or bias effects occur. The difficulty in handling a doubly infinite series is obvious. Price (1964) has given an explicit and exact finite term formula for the doubly noncentral F distribution where the numbers of degrees of freedom are either both odd or both even.

Box (1954a, 1954b) has obtained distributions of quadratic forms useful in the study of robustness of ANOVA. Other useful representations for the distribution of quadratic forms have been obtained by

Robbins and Pitman (1949), Gurland (1953, 1956) and Ruben (1962). Extending and modifying their work Johnson, Kotz and Boyd (1967) and Johnson and Kotz (1968) have developed several expansions which appear well suited for computer applications. Bounds of error have been given for these expansions and uniform convergence of the expansions have been established. Subrahmaniam (1967, 1968a, 1968b) has obtained distributions of some standard statistics starting from a Gram-Charlier type of parent population.

Not in references

Our contribution in studying robustness of ANOVA procedures is summarised in section 1.2.2 and presented in detail in chapter three.

1.1.3 Performance of some tests of independence

Approximations to the sampling distribution of the coefficient of correlati (r) in non-normal samples when the population coefficient of correlation (ρ) is zero has been given by Quensel (1938). Gayen (1951) has obtained a more general result when ρ is not necessarily zero. He has obtained this approximation by starting with a bivariate Gram-Charlier expansion of the joint probability density function of the population and ignoring all joint cumulants of the population above the fourth. He has studied the robustness of the distributions of the sample correlation coefficient r and Fisher's transformation $z = \tanh^{-1} r$ with respect to departures from normality.

It has been observed that when $\rho = 0$ and, in particular, when the variables are independent, the sampling distribution of r is more or less robust, even for small samples. But if the value of ρ is large, the effect of departures from normality is quite appreciable. About Fisher's transformation, the distribution of z remains asymptotically normal, but the convergence of the distribution to normality is slower. The mean and variance of z are, to order n^{-1}, unaffected by skewness in the parental marginal distribution. But the effect of departure from normal value of kurtosis may be considerable.

When the form of the parent distribution is not known and a large number of observations is available, independence between two variables may be tested by forming a two-way contingency chi-square statistic. Properties of chi-square statistics either for goodness of fit or for testing independence have been studied by Eisenhart (1938), Mann and Wald (1942), Cochran (1952), Williams (1950), Hamdan (1963, 1968), Mitra (1955, 1958) and others. Eisenhart and Cochran have shown that the limiting distribution, in the sense of Pitman (1948), of the goodness of fit chi-square in the non-null case follows a noncentral chi-square distribution. Mitra (1955, 1958) has shown that the non-null limiting distribution of the contingency chi-square statistic too follows a non-central chi-square distribution.

Mann and Wald (1942) and Williams (1950) have dealt with optimum number of classes, in the case of goodness of fit testing. Hamdan (1968) has studied the problem of determining the optimum number of classes in a two-way contingency table. Chapman and Meng (1966) have obtained the power function of chi-square test for contingency tables. Harkness and Katz (1964) have given the exact power of the randomised chi-square test in a (2×2) contingency table. They also have given a normal approximation to this power function. The tables prepared by them in a sense, supplement the tables prepared by Bennett and Hsu (1960).

Our contribution in studying the performance of some tests of independence is detailed in chapter four and a summary is given in section 1.2.3.

1.1.4 Effect of deviations from simple sampling on some statistical procedures.

Robustness of t tests and ANOVA procedures has been studied mainly with respect to departure of the populations from normality and homoscedasticity. Not much seems to have been done to study the effect of deviations from simple sampling on these procedures.

Anderson (1942) has obtained the distribution of means for stratified samples for the rectangular population which is symmetric and also for a J-shaped skewed triangular population. The means computed from stratified samples show less variability than the means of simple random samples. It has been observed by him that the stratified sample means from a skewed population are less skewed than the random sample means. He has also shown, for the three populations considered by him, that the effect of stratification in sampling upon the distribution of the standard deviation is to make the distribution more symmetric. It has also been demonstrated that for stratified samples from the rectangular population student's t is much more stable than it is for random samples of the same size.

Our contributions in studying the effect of deviations from simple sampling on t-tests and one-way classification F are presented in chapter five and summarised in section 1.2.4.

1.2 A summary of the present contribution

1.2.1 Approximations to distribution functions using formal series expansions

In chapter two we seek approximations to sampling distributions which are used later to study robustness of some statistical procedures. The density function of a statistic is approximated in terms of a standard density function and orthogonal polynomials associated with it. These approximations can be used when exact moments of the statistics are known. We consider two standard densities: (1) gamma-when the range of the statistic is the non-negative part of the real line, and (2) beta - when the range is the interval 0 to 1. Methods for obtaining similar approximations for joint distributions of two or more statistics, when the exact joint moments of these statistics are known, are also discussed. An expansion for the joint distribution is then used to derive an approximation to the sampling distribution of the statistic $U = X_2/(X_1 + X_2)$ where X_1 and X_2 are non-negative random variables and the joint moments of X_1 and X_2 are known.

The approximations to density functions used in this chapter may turn out to be negative for certain values of the variables; the practical consequences of this are examined in section 2.5.

1.2.2 Studies on robustness of ANOVA

It is known [Tang (1938)] that when the expectation model in a Gauss-Markoff linear set up is wrong, the distribution of ANOVA F follows a doubly non-central F distribution. In chapter three we give a computational procedure for exact evaluation of this distribution function, and also suggest two simple approximations. One of the approximations uses a beta density function and Jacobi polynomials, and in the other we use gamma density functions and Laguerre Polynomials. These results have been published [Dasgupta (1968a, 1968b)]. No attempt is made to obtain error bounds for these approximations, but numerical comparison with the exact results show that generally these approximations do fairly well.

Some authors (for example, Chakravorti (1967)) have suggested the use of the average power of the ANOVA procedure, by assuming a-priori distributions of the unknown parameters. A computational algorithm is derived here for evaluating this kind of an average power assuming that the non-centrality parameters are independently distributed as Pearson's type III variables.

In a different direction we show that the level of significance of the ANOVA procedure and its unbiasedness as a test of significance are not affected by certain types of deviations from the assumed linear model. We also obtain a set of necessary and sufficient conditions on such deviations under which best linear unbiased estimates (BLUEs) of parametric functions retain this property of best linear unbiasedness and simultaneously the ANOVA procedure retains its size and unbiasedness. These are extensions of results due to Mitra and Rao (1968).

An exact expression of the non-null distribution of the F statistic, when the homoscedasticity assumption about the basic observations is violated, is obtained by an extension of the results of Pitman and Robbins (1949) in the restricted case when the 'Within' and 'Between' sums of squares are independently distributed. An approximation is also suggested in this case using a method developed in chapter two. Accuracy of this approximation is judged by observing its performance in two cases where exact distributions are derivable using David and Fix (1960).

Finally we suggest a general approximation to the distribution of F when one or more assumptions are violated. This approximation is based on an expansion for the joint density function of the numer^a^tor and denominator sums of squares. This approach is similar to that of

Tiku (1964). The essential difference between Tiku's method and the one presented here is that unlike Tiku we suggest using scaled sums of squares rather than the sums of squares themselves in obtaining the approximation. This is done in the hope of minimising the number of terms in the approximation without affecting accuracy.

1.2.3 Studies on the performance of some tests of independence

In chapter four we consider the performance of some tests of independence.

It is known that in normal samples the sample coefficient of correlation r provides a uniformly most powerful unbiased test for independence for both one and two sided alternatives. Gayen (1951) has obtained an approximation to the distribution of the sample correlation coefficient when population is non-normal by starting from a bivariate Gram-Charlier expansion of the population density function. An alternative approach is presented in this thesis. The probability density function of $X = \frac{1}{2}(1 + r)$ is approximated using a beta density function and the associated orthogonal Jacobi polynomials. The accuracy of this approximation is judged numerically. A comparison of some approximate values of the distribution function with the exact values, when the population is bivariate normal, at a few selected points show that the approximation is not too bad. The efficiency of the approximation is also studied by two model sampling experiments. Some of these results have already appeared in Dasgupta (1968c). Next, we consider the performance of the contingency chi-square as a statistic for testing independence. An expression for the Pitman power [Pitman (1948)] of the contingency chi-square procedure in normal samples is obtained. This is then compared with the Pitman power of the uniformly most powerful procedure based on r. It is proved that for a (2 × 2) contingency table, when observations are drawn independently from a bivariate normal population, the Pitman power is maximised if the divisions are made for both the variables at their respective means.

Using a result due to Roy (1966) we obtain a class of test procedures based on counting for testing independence quickly in bivariate normal samples. This class contains the most powerful binomial test based on counting for a fixed alternative. The performance of the most powerful test as obtained from the class is studied with that of the binomial test based on median dichotomy and also the test based on the coefficient of correlation,

1.2.4 Effect of deviations from simple sampling on procedures for testing means

In chapter five we study the effect of deviation from simple sampling on some standard statistical tests.

Stratification is usually resorted to for improving the efficiency of the estimates obtained from samples or for convenience where the population physically exists in a stratified form. Sometimes standard statistical test procedures, which are strictly valid for simple random samples only, are used even when observations have been drawn using stratified sampling. This is done, perhaps, with the implicit belief that the result will not be too much vitiated due to deviation from simple sampling. That, this is sometimes dangerous is shown here by obtaining the limiting [Pitman (1948)] power of the student's t procedure for testing the significance of the mean in stratified samples from normal populations when deviations of the alternatives from the null hypothesis are of order $n^{-1/2}$, where n is the sample size. It is demonstrated that this power decreases if stratification is increased by splitting up any one stratum.

We, next, consider the power function of student's t in general non-normal samples when stratification is used. An approximation to the power function is suggested using a method discussed in chapter two, after the joint moments of $\bar{x}$ and s have been obtained using a technique due to David and Johnson (1951a). Use of this approach for obtaining approximations to the power functions in the case of the t test for equality of means, and the one-way classification F procedure is also discussed.

Finally, we consider a generalised 'probability proportional to size' sampling scheme which provides an unbiased estimate of the population mean, and obtain an approximation to the power function of students' t procedure in this type of sampling.

In short, apart from suggesting heuristic approximations in studying robustness properties of some well known statistical tests, some interesting results are obtained in this thesis. These include a set of sufficient conditions under which the ANOVA procedure retains its assumed size and the property of unbiasedness even when the true expectation model is different from the assumed one, a necessary and sufficient condition that (i) the best linear unbiased estimates of some estimable parametric functions under the assumed expectation model remain so under the true model and (ii) the ANOVA procedures pertaining to these parametric functions derived under the assumed model retain their size and the property of unbiasedness in the true model.

It is noted here that the limiting [Pitman (1948)] power of the contingency chi-square test of independence in multinormal sample does not increase with the increase in the number of classes and that the limiting power of the procedure in a (2×2) table is maximum when the table is formed by dividing the marginal distributions at their respective medians. Most powerful binomial tests for independence in bivariate normal samples are obtained for fixed alternatives. The performance of these tests are quite encouraging as compared to the test based on the coefficient of correlation. It is also proved that in normal samples the limiting power [Pitman (1948)] of the standard student's t procedure for testing the significance of the mean value under the assumption of simple random sampling decreases with more and more stratification.

Computations for this dissertation have been carried out using machine systems IBM 1401-8k, Honeywell-400-2k and IBM 360/44-128k systems with programmes written mostly in FORTRAN.

1.3 Notation

Unless otherwise stated the following symbols will be used throughout this thesis :

$A : n \times m$ represents a matrix with n rows and m columns

A' represents the transpose of A

$\mu(A)$ represents the linear manifold generated by columns of A

A^{-} represents a generalised inverse of the matrix A as defined by Rao (1965).

$P_A = A(A'A)^{-}A'$ is the projection operator projecting arbitrary n-vectors onto $\mu(A)$

$A^{\perp}$ denotes a matrix of maximum rank such that $A'A^{\perp} = 0$

$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2), \; -\infty < x < \infty$ represents the density function of a standardised normal variate $(N(0, 1))$

$p_j(\lambda) = \exp(-\lambda)\lambda^j / j!, \; j = 0, 1, \ldots$ represents the probability that a Poisson variable with mean $\lambda(> 0)$ assumes the value j.

$\Phi(x) = \int_{-\infty}^{x} \phi(t)\, dt$ represents the distribution function of $N(0, 1)$

$$\beta(x;\, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\ \Gamma(b)}\, x^{a-1} (1-x)^{b-1},\ 0 \leq x \leq 1,$$

represents the density function of a beta variable with parameters a, b $(a, b > 0)$

$$B(x;\, a, b) = \int_{0}^{x} \beta(t\,;\, a, b)\, dt$$

$g(x;\, m) = \frac{1}{\Gamma(m)} \exp(-x)\, x^{m-1},\ x \geq 0$ represents the density function of a gamma variate with parameter $m (m > 0)$

$$G(x;\, m) = \int_{0}^{x} g(t;\, m)\, dt$$

$f_n(x) = \frac{1}{2^{\frac{n}{2}}\ \Gamma(\frac{n}{2})} \exp(-\frac{1}{2}x)\, x^{\frac{1}{2}n-1}$ represents the density function of a $\chi^2(n)$ variable $(n > 0)$

$\chi^2(n, \lambda)$ represents a non-central chi-square variable with n degrees of freedom and non-centrality parameter λ, i.e. a random variable with density function

$$f_n(x;\, \lambda) = \sum_{j=0}^{\infty} p_j(\tfrac{\lambda}{2})\, f_{n+2j}(x),\ x \geq 0$$

$$F(x, n, \lambda) = \int_0^x f_n(t ; \lambda)\, dt$$

$N(0, 0, 1, 1, \rho)$ represents a bivariate normal variable with density function

$$f(x, y, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]$$

$$-\infty < x, y < \infty, \quad 0 \leq \rho < 1.$$

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ denotes the mean of n real quantities $x_1, x_2, \ldots, x_n$.

CHAPTER 2

APPROXIMATIONS TO DISTRIBUTION FUNCTIONS USING ORTHOGONAL POLYNOMIALS

2.0 Introduction and summary

In the following chapters in this thesis we seek approximations to sampling distributions of various statistics in terms of a standard density function and orthogonal polynomials associated with it. This requires knowledge of the first few moments of the statistic. It is difficult to get any useful estimate of the error in such approximations, but comparisons can be made when exact sampling distributions are known to get an idea of the magnitude of error. Theoretical results on the convergence of such expansions using infinitely many terms have been studied, in some cases, by Szego (1939), Cram (1925, 1928) but are not very useful in our context.

Let $f(x)$ be the exact density function of a statistic X whose r-th moment

$$\Theta(r) = E(X^r)$$

exists for $r = 1,2,...,t$. Let $w(x)$ be a standard density function with associated orthogonal polynomials

$$\pi(j, x) = \sum_{r=0}^{j} c(j, r)x^r$$

satisfying

$$\int_{-\infty}^{\infty} \pi(j, x)\ \pi(k, x)\ w(x)dx = \begin{cases} 0 & \text{if } j \neq k \\ c(j)>0 & \text{if } j = k \end{cases}$$

We shall denote the t-th order approximation to f(x) in terms of w(x) by

$$f_t \equiv f_t(x|w) = w(x) \sum_{j=0}^{t} a(j)\ \pi(j, x)$$

where

$$a(j) = \sum_{r=0}^{j} c(j, r)\ \Theta'(r)/c(j)$$

It is to be noted that f_t itself is not necessarily a density function.

In like manner, we shall call

$$F_t \equiv F_t(x|w) = \int_{-\infty}^{x} f_t(u|w)du$$

the t-th order approximation to the cumulative distribution function

$$F(x) = \int_{-\infty}^{x} f(u)du$$

in terms of w.

We shall be concerned in particular with two different standard distributions namely

(1) gamma - when the range is the non-negative part of the real line, and

(2) beta - when the range is the unit interval from 0 to 1.

Results relating these two standard distributions are discussed in sections 2.1 and 2.2.

Methods for obtaining similar expansions for joint distributions of two or three statistics are described in section 2.3. These results are utilised in section 2.4 to derive an approximation to the sampling distribution of $U = X_2/(X_1 + X_2)$ where X_1 and X_2 are non-negative random variables whose joint moments are known.

The approximations may turn out to be negative for certain values of x; the practical consequences of this are examined in section 2.5.

2.1 Laguerre series expansion

If the statistic X is non-negative, we may like to take as standard the gamma density function

$$g(x;m) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}, \quad 0 \leq x < \infty \tag{2.1.1}$$

where $m > 0$ is a parameter. Associated orthogonal polynomials are called Laguerre polynomials $L(j,x;m)$, $j = 0,1,2\ldots$ and these are defined by

$$j!\, L(j,x;m)\, e^{-x} x^{m-1} = \left(\frac{d}{dx}\right)^j e^{-x} x^{m+j-1}$$

or

$$L(j,x;m) = \sum_{r=0}^{j} c(j,r;m)\, (-x)^r/r! \tag{2.1.2}$$

where

$$c(j,r;m) = \begin{cases} (m+r)\ (m+r+1)\ \ldots\ (m+j-1)/(j-r)!, & \text{for } r=0,1,\ldots,j-1 \\ 1 & \text{for } r=j \end{cases} \tag{2.1.3}$$

for $j = 0,1,2, \ldots$.

Also,

i) $C(j;m) = \int_0^\infty L^2(j,x;m)\ g(x;m)dx = c(j,o,m)$

ii) $\int_0^\infty L(j,x;m)L(k,x;m)\ g(x;m)dx = 0, \quad j \neq k$

and

iii) $\int_0^x g(t\ ;\ m)\ L(r,t;m)dt$

$$= \frac{m}{r}\ g(x;m+1)\ L(r-1,\ x;\ m+1)$$

$$= \frac{x}{r}\ g(x;m)\ L(r-1,\ x;\ m+1) \tag{2.1.4}$$

In particular,

$L(0,\ x;\ m) = 1$

$L(1,\ x;\ m) = m - x$

$L(2,\ x;\ m) = \dfrac{m(m+1)}{2!} - (m+1)x + \dfrac{x^2}{2!}$

$L(3,\ x;\ m) = \dfrac{m(m+1)(m+2)}{3!} - \dfrac{(m+1)(m+2)}{2!}\ x + \dfrac{(m+2)}{2!}\ x^2 - \dfrac{x^3}{3!}$

$$L(4,\ x;\ m) = \frac{m(m+1)(m+2)(m+3)}{4!} - \frac{(m+1)(m+2)(m+3)x}{3!}$$

$$+ \frac{(m+2)\ (m+3)}{2!}\ \frac{x^2}{2!} - (m+3)\ \frac{x^3}{3!} + \frac{x^4}{4!}$$

and

$$a(0) = 1$$

$$a(1) = [m - \theta(1)]/m$$

$$a(2) = \left[\frac{m(m+1)}{2!} - (m+1)\theta(1) + \frac{\theta(2)}{2!}\right] / \frac{m(m+1)}{2!}$$

$$a(3) = \left[\frac{m(m+1)(m+2)}{3!} - \frac{(m+1)(m+2)\theta(1)}{2!} + \frac{(m+2)\theta(2)}{2!} - \frac{\theta(3)}{3!}\right] / \frac{m(m+1)(m+2)}{3!}$$

$$a(4) = \left[\frac{m(m+1)(m+2)(m+3)}{4!} - \frac{(m+1)(m+2)(m+3)\theta(1)}{3!} + \frac{(m+2)(m+3)}{2!}\frac{\theta(2)}{2!} - \frac{(m+3)\theta(3)}{3!} + \frac{\theta(4)}{4!}\right] / \frac{m(m+1)(m+2)(m+3)}{4!} \qquad (2.1.5)$$

If we consider the transformed statistic

$$Y = cX$$

and choose $c = \theta(1)/[\theta(2) - \theta^2(1)]$

$$m = \theta^2(1)/[\theta(2) - \theta^2(1)] = v^{-2}$$

where v is the coefficient of variation (that is, the ratio of the standard deviation to mean) of X, it can be easily shown that the fourth order approximation to the density function of Y at the point x is

$$f_4 = g(x;m)\ \{1 + \alpha(3)\ L(3,x;m) + \alpha(4)\ L(4,x;m)\} \qquad (2.1.6)$$

where $\alpha(3)$ and $\alpha(4)$ are obtained from a(3) and a(4) in (2.1.5) by replacing $\theta(r)$ by $c^r\,\theta(r)$.

The corresponding approximation to the cumulative distribution function of Y is

$$F_4 = G(x;m) + g(x;m)\ \{\frac{1}{3}\ \alpha(3).L(2,x;m+1) + \frac{1}{4}\ \alpha(4).L(3,x;m+1)\} \qquad (2.1.7)$$

where $G(x;m) = \int_0^x g(u;m)du$

Approximations of this type have been used by Roy and Tiku (1962), Roy and Mohammed (1964), Khamis (1960), Tiku (1963a, 1963b, 1964, 1965) and others.

2.2 Jacobi series expansion

If X is a random variable distributed over the interval (0,1), it is natural for us to seek an approximation for its density function in terms of the beta density function

$$\beta(x;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\ \Gamma(b)}\ x^{a-1}\ (1 - x)^{b-1},\ 0 \leq x \leq 1 \qquad (2.2.1)$$

where $a > 0$ and $b > 0$ are two parameters.

Orthogonal polynomials associated with the beta density function are called Jacobi polynomials, and defined as

$$J(j,x;a,b) = \sum_{r=0}^{j} (-1)^r\ c(j,r;\ a,b)x^r \qquad (2.2.2)$$

where

$$c(j;0;\ a,b) = a(a+1)\ \ldots\ (a+j-1)/j!$$

$$c(j,r;a,b) = [(j+a+b-1)(j+a+b)\ \ldots\ (j+a+b+r-2)/r!]$$

$$\times [a+r)\ (a+r+1)\ \ldots\ (a+j-1)/(j-r)!]$$

$$\text{for} \quad r = 1,2,\ldots,\ j-1$$

$$c(j,j;a,b) = (j+a+b-1)\ (j+a+b)\ \ldots\ (2j+a+b-2)/j!$$

Also

i) $$\int_0^1 J^2(j,a,b)\ \beta(x,a,b)dx = c(j;a,b)$$

where $c(o;\ a,b) = 1,$

$$c(j;a,b) = \frac{a(a+1)\ldots(a+j-1)b(b+1)\ \ldots\ (b+j-1)}{j!(2j+a+b-1)(a+b)(a+b+1)\ \ldots\ (a+b+j-2)}$$

$$\text{for } j = 1,2,\ \ldots,$$

and

ii) $$\int_0^x J(r,t;\ a,b)\ \beta(t;\ a,b)dt$$

$$= -\ \frac{ab\ \beta(x;\ a+1,b+1)\ J(r-1,x;\ a+1,\ b+1)}{r(a+b)\ (a+b+1)} \qquad (2.2.3)$$

Here, if we choose

$$a = \frac{\Theta(1)\ (\Theta(1) - \Theta(2))}{\Theta(2) - \Theta^2(1)}$$

$$b = \frac{(\Theta(2) - \Theta(1))(\Theta(1)-1)}{\Theta(2) - \Theta^2(1)}, \qquad (2.2.4)$$

then the fourth order approximation to f(x), the density function of X, in terms of $\beta(x;\ a,b)$ is

$$f_4 = \beta(x;\ a,b)\ [1 + a(3)\ J(3,x;\ a,b) + a(4)\ J(4,x;\ a,b)] \qquad (2.2.5)$$

where $a(j) = \int_0^1 J(j,x;\ a,b)\ f(x)dx/c(j;\ a,b)$

$$= \sum_{r=0}^{j} (-1)^r\ c(j,r;a,b)\ \theta(r)/c(j;a,b),\ j=0,1,\ldots \qquad (2.2.6)$$

Consequently, the fourth order approximation to the distribution function of x is

$$F_4 = B(x;\ a, b) - \beta(x;\ a+1, b+1)[d(3)J(2, x; a+1, b+1) + d(4)\, J(3, x;\ a+1, b+1)] \qquad (2.2.7)$$

where $B(x;\ a, b) = \int_0^x \beta(t;\ a, b)\, dt$,

and, writing $s = a + b$,

$$d(3) = \frac{s+5}{(b+1)(b+2)}\left[\frac{a}{3} - (s+2)\theta(1) + \frac{(s+2)(s+3)\theta(2)}{a+1} - \frac{(s+2)(s+3)(s+4)}{3(a+1)(a+2)}\theta(3)\right]$$

$$d(4) = \frac{(s+7)(s+2)}{(b+1)(b+2)(b+3)}\left[\frac{a}{4} - (s+3)\theta(1) + \frac{3(s+3)(s+4)\theta(2)}{2(a+1)} - \frac{(s+3)(s+4)(s+5)}{(a+1)(a+2)}\theta(3) + \frac{(s+3)(s+4)(s+5)(s+6)}{4(a+1)(a+2)(a+3)}\theta(4)\right] \qquad (2.2.8)$$

2.3 Expansion of joint density function of several variables

We shall illustrate the method by considering the joint frequency density function $f(x_1, x_2)$ of two statistics X_1, X_2 whose joint moments

$$\theta(r_1, r_2) = E(X_1^{r_1}\ X_2^{r_2})$$

are known. Let $w_i(x)$ be the standard density function, considered appropriate for expansion of the marginal frequency density function of X_i and let $\pi_i(j, x)$, $j = 0, 1, 2, \ldots$, denote the orthogonal polynomials associated with w_i, $i = 1,2$. We expand $f(x_1, x_2)$ formally in a power series form

$$f(x_1,x_2) = \sum_{k_1} \sum_{k_2} a(k_1,k_2) \prod_{i=1}^{2} \pi_i(k_i,x_i) w_i(x_i) \qquad (2.3.1)$$

Multiplying both sides by $\prod_{i=1}^{2} \pi_i(j_i,x_i)$ and integrating over the entire two dimensional Euclidean space, we get formally

$$\iint \left[\prod_{i=1}^{2} \pi_i(j_i,x_i) \right] f(x_1,x_2)\, dx_1\, dx_2$$

$$= \sum_{k_1} \sum_{k_2} a(k_1,k_2) \prod_{i=1}^{2} \int \pi_i(k_i,x_i)\, \pi_i(j_i,x_i) w_i(x_i) dx_i \qquad (2.3.2)$$

if the order of summation and integration on the right hand side can be validly interchanged.

From the orthogonality property, all terms of the right hand side vanish except the single term for which $k_i = j_i$, $i = 1,2$.

Consequently

$$a(j_1,j_2) \prod_{i=1}^{2} \int \pi_i^2 (j_i,x_i)\, w_i(x_i) dx_i$$

$$= \iint \left[\prod_{i=1}^{2} \pi_i(j_i,x_i) \right] f(x_1,x_2)\, dx_1\, dx_2 \qquad (2.3.3)$$

Let us write

$$\int \pi_i^2 (j, x)\, w_i dx = c_i(j) \qquad (2.3.4)$$

Further, let

$$\pi_i(j, x) = \sum_{r=0}^{j} c_i(j, r)\, x^r \qquad (2.3.5)$$

so that

$$\prod_{i=1}^{2} \pi_i (j_i, x_i) = \sum_{r_1=o}^{j_1} \sum_{r_2=o}^{j_2} \prod_{i=1}^{2} c_i(j_i, r_i)\ x_i^{r_i}$$

We thus get

$$a(j_1, j_2) = \sum_{r_1=o}^{j_1} \sum_{r_2=o}^{j_2} \left[\prod_{i=1} \frac{c_i(j_i, r_i)}{c_i(j_i)} \right] \theta(r_1, r_2) \qquad (2.3.6)$$

However, we are not interested in the doubly infinite series and the question of its convergence. We are concerned with the truncated series obtained from (2.3.1) by retaining only the terms for which $j_1+j_2 \leq t$. We shall call this truncated series the t-th order approximation to $f(x_1, x_2)$ with respect to the standard frequency density functions w_1, w_2, and denote it by $f_t(x_1, x_2, | w_1, w_2)$ or simply by f_t when there is no possibility of ambiguity. It is a matter of simple extension when more than two statistics are involved.

To estimate the error in replacing f by f_t in general is beyond the scope of ourf investigation. We shall however compute in certain specific cases the numerical magnitude of the error to satisfy ourselves that this approximation does provide usable results in certain situations.

When the standard marginal density functions are fixed, there is one to one correspondence between the t-th order approximation of the type (2.3.1) for a multivariate distribution and its joint moments upto order t. This becomes obvious if we note that (2.3.6) can be represented as

$$G \; \Theta = A \tag{2.3.7}$$

where

$$\Theta = (\theta(0,0),\ \theta(0,1),\ ..,\ \theta(0,t),\ \theta(1,0),..,\ \theta(1,t-1),...,\theta(t,0))'.$$

$$A = (a(0,0),\ a(0,1),\ ..\ a(0,t),\ a(1,0),\ ..\ a(1,t-1),\ ..\ a(t,0))'$$

and

> g_{ij} -- the j-th element of the i-th row of the matrix G -- is the coefficient of $x_1^k\, x_2^l$ in $\pi_1(u,\ x_1)\ \pi_2(v,\ x_2)$ (assuming that $\int \pi_1(u,x)w_1(x)dx = \int \pi_2(v,x)w_2(x)dx = 1$) where (k,l) is the suffix for the j-th component of Θ and (u,v) is the suffix for the i-th component of A,

and the non-singularity of G can be established.

2.4 Approximation to the distribution of $X_2/(X_1+X_2)$ where X_1, X_2 are non-negative statistics.

Let us write

$$m_1 = E(X_1) = \theta(1,0),$$

$$m_2 = E(X_2) = \theta(0,1),$$

$$\mu_{rs} = E[(X_1 - m_1)^r\ (X_2 - m_2)^s],$$

$$Y_1 = c_1\, X_1,\quad c_1 > 0$$

and

$$Y_2 = c_2\, X_2,\quad c_2 > 0. \tag{2.4.1}$$

Since Y_1 and Y_2 are non-negative we may like to take as standard the gamma density functions $g(y_1;\, m)$ and $g(y_2;\, n)$. Then using (2.3.1) the t-th order approximation to the density function $f(y_1, y_2)$ of (Y_1, Y_2) is

$$f_t = \sum_{r+s \leq t} a(r,s)\, L(r,y_1;m)\, L(s,y_2;n)\, g(y_1;m)\, g(y_2;n) \tag{2.4.2}$$

where

$$a(r, s) = \frac{E[L(r,y_1;m)\, L(s,y_2;n)\,|f]}{c(r,0,m)c(s,0,m)}$$

and, L's - the Laguerre polynomials - and $c(r, 0, m)$ are as defined in (2.1.2) and (2.1.3) respectively.

In particular, the first few coefficients are

$$a(0, 0) = 1$$

$$a(1, 0) = (m - c_1 m_1)/m$$

$$a(0, 1) = (n - c_2 m_2)/n$$

$$a(2, 0) = \frac{1}{m(m+1)} (c_1^2\, \mu_{20} - m)$$

$$a(0, 2) = \frac{1}{n(n+1)} (c_2^2\, \mu_{02} - n)$$

$$a(1, 1) = c_1 c_2\, \mu_{11}/mn. \tag{2.4.3}$$

The above relations show that if we choose

$$c_1 = m_1/\mu_{20},$$

$$c_2 = m_2/\mu_{02},$$

$$m = \frac{m_1^2}{\mu_{20}}$$

and $$n = \frac{m_2^2}{\mu_{02}}$$

then $a(1,0) = a(0,1) = a(2,0) = a(0,2) = 0$

and $$a(1,1) = \frac{c_1 c_2\ \rho(\mu_{20}\ \ \mu_{02})^{\frac{1}{2}}}{mn} = \rho(mn)^{-\frac{1}{2}}$$

where ρ is the coefficient of correlation between X_1 and X_2.

From (2.4.2) the second order approximation to f, with the above choice of (c_1, c_2, m, n), is

$$f_2 = g(y_1;m)\ g(y_2;n)\ [1+a(1,1)\ (y_1-m)\ (y_2-n)] \qquad (2.4.4)$$

Consequently, the corresponding approximation to the distribution of $Z = Y_2/(Y_1 + Y_2)$ obtained by integrating f_2 over the region $y_2/(y_1 + y_2) \leq z$ turns out to be

$$F_2(z) = B(z;n,m) + \rho(mn)^{\frac{1}{2}}\ [B(z;n,m)$$

$$-B(z;n+1,m) - B(z;n,m+1) + B(z;n+1,m+1)]$$

$$= B(z;n,m) + \rho(\frac{n}{m})^{\frac{1}{2}}\ \frac{\Gamma(n+m)}{\Gamma(n)\ \Gamma(m)}\ z^n(1-z)^m[-1 + \frac{m+n}{n}\ z] \qquad (2.4.5)$$

Hence, an approximation to the cumulative distribution function of $V = X_2/(X_1 + X_2)$ at the point x is $F_2(x^*)$ where

$$x^* = [1 + \frac{c_1(1 - x)}{c_2 x}]^{-1}$$

$$= [1 + \frac{m_1\ \ \mu_{02}}{m_2\ \ \mu_{20}} \cdot \frac{(1-x)}{x}]^{-1}$$

and n, m, ρ are as defined before.

In chapter 3, this approximation is used in studying the distribution of the variance-rate statistic under departure from usual assumptions. The accuracy of this approximation is examined in particular situations where

$X_1 = U_1 + U_2$ and $X_2 = U_1 + U_3$ and U_1, U_2, U_3 are independent gamma variables.

2.5 Some remarks on situations where the approximation to the density function is negative.

The t-th order approximation f_t is not positive definite. For certain values of x, f_t may assume negative values. This by itself is not a serious defect so long as the probabilities computed from the approximation agree reasonably with the true probabilities in the range in which the statistician is interested. It is however of some importance to examine this point.

Since

$$f_t(x) = w(x) \sum_{j=0}^{t} a(j)\ \pi(j,x)$$

it is clear that $f_t(x)$ is negative in the region in which the t-th degree polynomial

$$P_t(x) = \sum_{j=0}^{t} a(j)\ \pi(j,x)$$

is negative. This can therefore be studied by examining the nature of the roots of $P_t(x) = 0$, which depends on the coefficients $a(j)$, $j = 0,1,2,\ldots t$ once the density function $w(x)$ is fixed. This is taken up in this section for the gamma and beta densities when $t=3$.

As a measure of seriousness of negativity of f_t, we compute the total negative area

$$F_{t-} = 1 - \int |f_t|\, dx$$

the integral being over the whole range of the statistic. The smaller the value of $|F_{t-}|$ the less serious is the problem.

2.5.1 Third order approximation in terms of $g(x;m)$

If we consider a transformed statistic of the type $Y = CX$ where X is a non-normal statistic, and C and m are as defined in section 2.1, then the third order approximation in terms of $g(x;m)$ to the density function of Y at the point x is

$$f_3 = g(x;m)\ \{1 + \alpha(3)\ L(3,x;m)\}$$

We notice that the cubic $L(3, x; m)$ has the value $C_o = \frac{m(m+1)(m+2)}{6}$ at the point $x = o$, has extrema at the points $x_1 = (m+2) - (m+2)^{\frac{1}{2}}$ and $x_2 = (m+2) + (m+2)^{\frac{1}{2}}$ and that

$$C_1 = L(3,\ x_1;\ m) = \tfrac{1}{3}(m+2)\ \{1 - (m+2)^{\frac{1}{2}}\}$$

and

$$C_2 = L(3,\ x_2;\ m) = \tfrac{1}{3}(m+2)\ \{1 + (m+2)^{\frac{1}{2}}\}.$$

The point at which f_3 is negative and the total negative area F_{3-} will depend on the positive real roots of

$$L(3,\ x;\ m) = T \qquad (2.5.1)$$

where $T = -1/\ \alpha(3)$

At most three such roots are possible and when they exist they will lie in regions :

$$0 < r_0 < x_1$$

$$x_1 < r_1 < x_2$$

$$x_2 < r_2 < \infty \qquad (2.5.2)$$

The existence of all these roots, however, depends on the value of T and we have the following possibilities :

(i) No solution exists

(ii) Only one solution r_0 or r_2 exists

(iii) Two solutions r_1 and r_2 exist

(iv) All the three solutions exist.

Enumerating the different values $\alpha(3)$ can assume we have the following situations :

Case I : when $\alpha(3) > 0$

(a) If $|T| \leq |C_1|$, then all the three roots exist,

$f_3 < 0$ for $r_0 < x < r_1$, $x > r_2$ and

$$F_{3-} = \int_{r_0}^{r_1} f_3 \, dx + \int_{r_2}^{\infty} f_3 \, dx = 1 - F_3(r_2) + F_3(r_1) - F_3(r_0)$$

(b) If $|T| > |C_1|$, then only r_2 exists, $f_3 < 0$ for $x > r_2$ and $F_{3-} = 1 - F_3(r_2)$.

Case II : when $\alpha(3) < 0$

(a) If $T \geq max(C_0, C_2)$ then f_3 is non-negative in the entire range and is itself a density function.

(b) If $T < C_0, C_2$, then all roots exist. $f_3 < 0$ for $x < r_0$, $r_1 < x < r_2$ and

$$F_{3-} = F_3(r_0) + F_3(r_2) - F_3(r_1)$$

(c) If $C_0 \leq T < C_2$ — a situation possible only when $m < 2$ — then two roots r_1 and r_2 exist, $f_3 < 0$ for $r_1 < x < r_2$ and $F_{3-} = F_3(r_2) - F_3(r_1)$.

(d) If $C_2 \leq T < C_0$ – a situation possible only when $m > 2$ – then only r_0 exists, $f_3 < 0$ for $x < r_0$ and $F_{3-} = F_3(r_0)$.

Case III : $\alpha(3) = 0$

In this case $f_3(x) = g(x; m)$.

In general for a fixed **m**, the smaller the value of $\alpha(3)$, the smaller will be F_{3-}. If we simplify the expression for $\alpha(3)$ in terms of the moments of the basic variable X, we find that

$$\alpha(3) = 1 - \theta(1)\ \theta(3)/\{\theta(2)(2\theta(2) - \theta^2(1)\} \qquad (2.5.2)$$

or in terms of $\theta(1)$ and central moments $\mu_r = E(x - \theta(1))^r$, $r = 2, 3$

$$\alpha(3) = 1 - \frac{\theta(1)\ \{\mu_3 + 3\mu_2\ \theta(1) + \theta^3(1)\}}{\{\mu_2 + \theta^2(1)\}\ \{2\mu_2 + \theta^2(1)\}} \qquad (2.5.3)$$

If X has a Pearson type III distribution, then $\alpha(3)$ will vanish and the approximation f_3 will coincide with the exact density function of Y.

If X is a statistic based on a sample of size n and if

$$\underset{n \to \infty}{Lt}\ E(X) = \lambda_1$$

and $$\underset{n \to \infty}{Lt} \; n^{\frac{1}{2}r} E(X - \theta(1))^r = \lambda_r , \quad r = 2 , 3$$

exist, then for large n

$$\alpha(3) = -\frac{\lambda_3}{\lambda_1^3} n^{-\frac{3}{2}} + O(n^{-2})$$

This shows that asymptotically at least in a situation like this f_3 will become a density function.

2.5.2 Graph

In figure 2.1 we present the [redacted] area $|F_3|$ as a function of $\alpha(3)$ for $m = 4, 5, 6$ and 7.

2.5.3 Third order approximation in terms of $\beta(x; a, b)$

In the case of third order approximation f_3 in terms of $\beta(x; a, b)$ to the density function of a statistic X whose range is the unit interval 0 to 1, the roots to be studied is of the cubic

$$1 + \alpha(3)\, J(3, x; a, b).$$

About the behaviour of the cubic $J(3, x; a, b)$ we note that

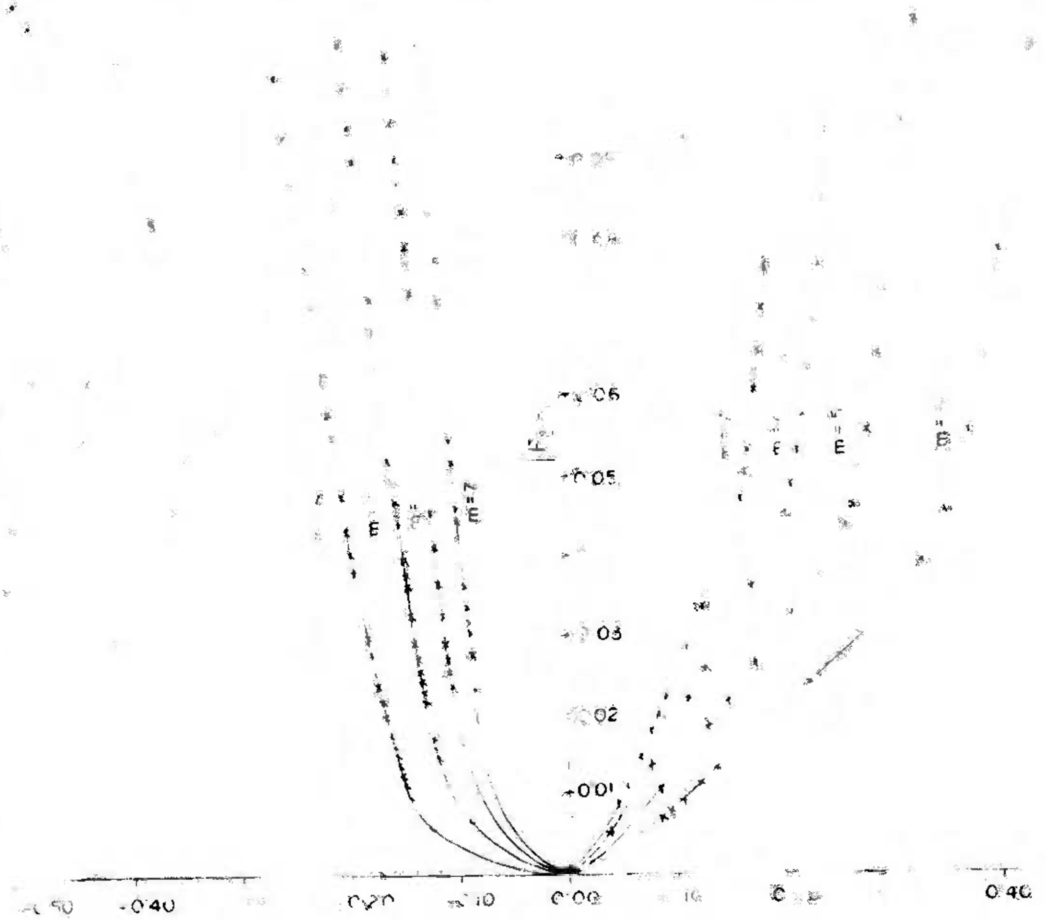
$$C_0 = J(3, 0; a, b) = a(a+1)(a+2)/6$$

and that $J(3, x; a, b)$ has stationary values at the points

$$x_1 = \{(a+2) - ((a+2)(b+2))^{\frac{1}{2}} (a+b+3)^{-\frac{1}{2}}\} / \{a+b+4\}$$

and $$x_2 = \{(a+2) + ((a+2)(b+2))^{\frac{1}{2}} (a+b+3)^{-\frac{1}{2}}\} / \{a+b+4\} ,$$

OF α (3)



Let

$$C_1 = J(\beta_3, x_1; a, b)$$

$$C_2 = J(\beta_3, x_2; a, b)$$

and $$C_3 = J(\beta_3, 1; a, b) = -b(b+1)(b+2)/6$$

We also note that at most three real roots between 0 and 1 of

$$J(\beta_3, x; a, b) = T \tag{2.5.4}$$

where $T = -1/\alpha(\beta_3)$

are possible, these roots, when they exist, lie in regions

$$0 < r_o < x_1$$

$$x_1 < r_1 < x_2$$

$$x_2 < r_2 < 1 \tag{2.5.5}$$

Using the method of analysis used in section 2.5.1 we have the following situations.

Case I : when $\alpha(\beta_3) > 0$

(a) If $|T| \geq max(|C_1|, |C_3|)$, then f_3 is non-negative throughot the range and $F_{3-} = 0$. f_3 is a density function in this case.

(b) If $|C_1| \leq |T| < |C_3|$, then only the root r_2 exists, $f_3 < 0$ for $x > r_2$ and $F_{3-} = 1 - F_3(r_2)$ where F_3 is the approximation to the distribution function corresponding to f_3.

(c) If $|C_3| \leq |T| < |C_1|$, then two roots r_o and r_1 exist, $f_3 < 0$ for $r_o < x < r_1$ and

$$F_{3-} = F_3(r_1) - F_3(r_o).$$

(d) If $|T| < |C_1|, |C_3|$, then all the three roots exist,

$f_3 < 0$ for $r_o < x < r_1$, $x > r_2$ and

$$F_{3-} = F_3(r_1) - F_3(r_o) + 1 - F_3(r_2)$$

Case II : when $\alpha(3) < 0$

(a) If $T \geq max(C_o, C_2)$, then f_3 is non-negative throughout the range and is a density function.

(b) If $T < C_o, C_2$, then all the three roots exist, $f_3 < 0$ for $0 < x < r_o$, $r_1 < x < r_2$ and $F_{3-} = F_3(r_o) + F_3(r_2) - F_3(r_1)$.

(c) If $C_o \leq T < C_2$, then two roots r_1 and r_2 exist and $F_{3-} = F_3(r_2) - F_3(r_1)$.

(d) If $C_2 \leq T < C_o$, then only the root r_o exists, $f_3 < 0$ for $0 < x < r_o$ and $F_{3-} = F_3(r_o)$

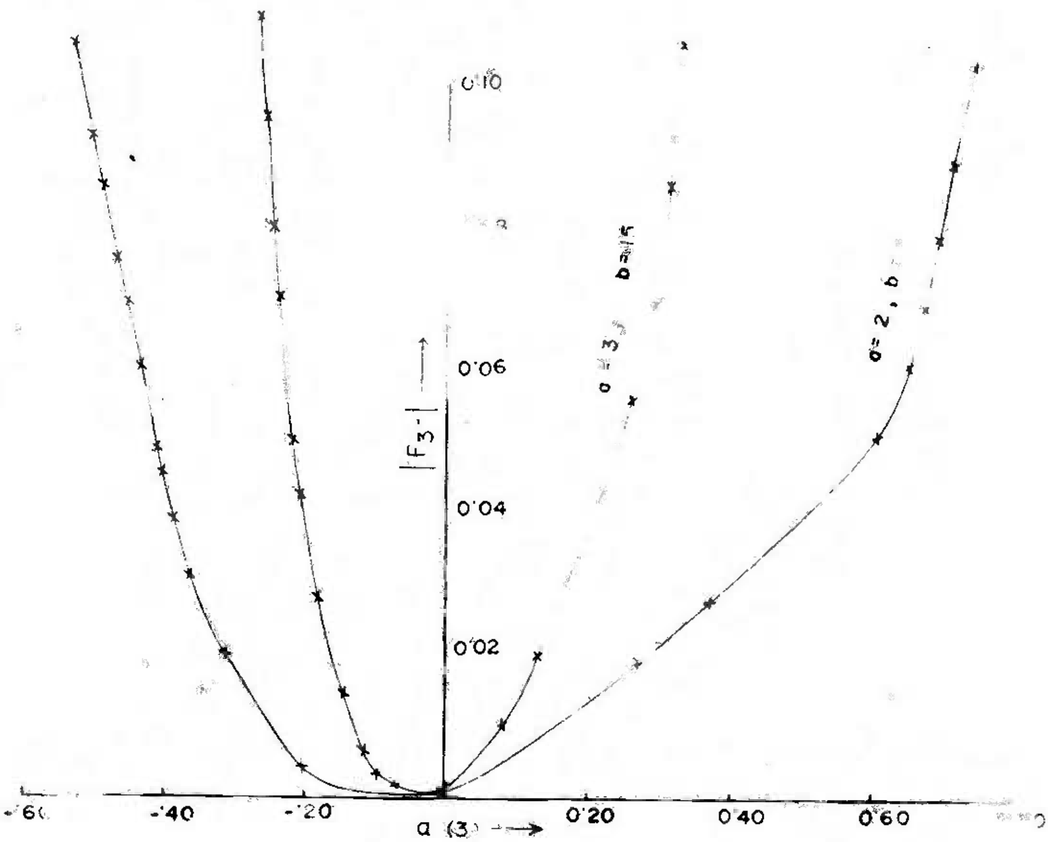
Case III : when $\alpha(3) = 0$

Here $f_3 = \beta(x; a, b)$

2.5.4 Graph

As illustrations we have plotted $|F_{3-}|$ as a function of $\alpha(3)$ for (i) $a = 2$, $b = 5$ and (ii) $a = 3$, $b = 15$.

FIGURE 2 ... A FUNCTION OF
... ACOBI APPROXIMATION



CHAPTER 3

ROBUSTNESS OF ANALYSIS OF VARIANCE

3.0 Introduction and Summary

In this chapter we study the effects of departure from the standard assumptions on the performance of the Analysis of Variance (ANOVA) procedure for testing a set of linear hypotheses under the Gauss-Markoff linear model.

If the other assumptions (normality, independence and homoscedasticity) are valid but the expectations do not follow the linear model assumed, the ANOVA test statistic F follows a doubly non-central F distribution. The transform $T = (1 + \frac{n_1}{n_2} F)^{-1}$, where n_1 and n_2 are the numbers of degrees of freedom for the 'hypothesis sum of squares' and 'error sum of squares' respectively, follows a doubly non-central beta distribution. The distribution function of doubly non-central F has been obtained by Price (1964) in a closed form with finite number of terms when n_1 and n_2 are either both odd or both even. In this chapter we suggest two approximations for the distribution function of T using Laguerre and Jacobi series expansions. These approximations do not require any restrictions on the numbers of degrees of freedom. A computational method is also given for evaluating numerically the exact distribution, to any predetermined order of accuracy, by expanding the distribution in a doubly infinite series.

Next we show that the level of significance (size) of the ANOVA procedure and its unbiasedness as a test of significance are not affected by certain types of deviations from the assumed linear method. We also obtain a set of necessary and sufficient conditions on such deviations under which best linear unbiased estimates (BLUE's) of parametric functions retain this property of best linear unbiasedness and simultaneously the ANOVA procedure retains its size and unbiasedness. These are extensions of results due to Mitra and Rao (1968).

We also suggest a procedure for computing the average power of the ANOVA procedure based on the doubly non-central beta distribution, when the two non-centrality parameters are assumed to be independent Pearson type III random variables.

Finally, we obtain approximations for the distribution of T when one or more of the assumptions of normality, homoscedasticity and independence are violated.

3.1 General linear hypotheses :

We consider a random vector $\eta = (\eta_1, \eta_2, \ldots, \eta_n)'$ which is distributed with

expectation $E(\eta) = A\theta,\ A : n \times p,\ \theta : p \times 1,$

and dispersion $D(\eta) = \sigma^2 I.$ (3.1.1)

This model will be referred to hereafter as Gauss-Markoff model $(A, \sigma^2 I)$ or simply as $(A, \sigma^2 I)$. In addition, it is assumed that each element of η follows the normal distribution. With this so called fixed effects normal model, the basic problem of ANOVA is to test a set of linear hypotheses of the form :

$$H_o : L\theta = K,\quad L : r_1 \times p,\quad K : r_1 \times 1 \qquad (3.1.2)$$

where we assume that all parametric functions in $L\theta$ are 'estimable' in the sense that there exists a constant matrix M such that $E(M\eta) = L\theta$. We shall denote the rank of the matrix A by Rank $(A) = r$ and that of L, by Rank $(L) = r_1$, where of course $r_1 \leq r$.

We make an orthogonal transformation of the n random variables η to n new random variables consisting of $(n - r)$ random variables Y_1, r_1 random variables Y_2 and $(r - r_1)$ random variables Y_3 such that the expectation vector of Y_1 is identically zero, the expectation vector of Y_2 is of the form $CL\theta$ where C is non-singular, and the expectation vector of Y_3 is linearly independent of $L\theta$. To see how this can be done we proceed as follows.

We call a matrix $G : g \times h$ semi-orthogonal if $g < h$ and $GG' = I$. Since A is of rank r, there exists a semi-orthogonal matrix $T_1 : (n - r) \times n$ such that $T_1 A = 0$. From the general theory of linear estimation it is known that the BLUE of $L\theta$ is provided by $L\hat{\theta}$ where $\hat{\theta} = (A'A)^{-} A'\eta$, $(A'A)^{-}$ being any generalised inverse, as defined by Rao(1965), of $(A'A)$. It is easy to check that the row space of the matrix $L(A'A)^{-}A'$ is a sub-space of $\mu(A)$ -- the linear manifold generated by the columns of A, and is of rank r_1. Hence it is possible to find a semi-orthogonal matrix $T_2 : r_1 \times n$ and a non-singular matrix $C : r_1 \times r_1$ such that $CL(A'A)^{-}A' = T_2$. Since $\mu(T_2')$ is a sub-space of $\mu(A)$ and $T_1 A = 0$, we have $T_1 T_2' = 0$. Next we choose a semi-orthogonal matrix $T_3 : (r - r_1) \times n$ with the property $T_3 T_1' = 0$, $T_3 T_2' = 0$.

Now, we submit the random vector η to the following transformation

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} \eta \qquad (3.1.3)$$

The vector Y will have the following expectation and dispersion

$$E \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} T_1 A\theta \\ T_2 A\theta \\ T_3 A\theta \end{bmatrix} = \begin{bmatrix} 0 \\ CL\theta \\ T_3 A\theta \end{bmatrix} = A^*\theta \text{ (say)}, \qquad (3.1.4)$$

$$D(Y) = \sigma^2 I.$$

Thus we see that the hypothesis specified in (3.1.2) is equivalent to

$$H_o : CL\theta = CK \qquad (3.1.5)$$

under the Gauss-Markoff model $(A^*, \sigma^2 I)$. The ANOVA procedure for testing H_o is :

$$\text{Reject } H_o \text{ if } F = \frac{(Y_2 - CK)'(Y_2 - CK)/r_1}{Y_1'Y_1/(n-r)} \geq F_\alpha ,$$

$$\text{accept } H_o \text{ otherwise,} \qquad (3.1.6)$$

where F_α is the upper 100α per cent point of the F distribution with r_1 and $(n - r)$ degrees of freedom and α $(0 < \alpha < 1)$ is a given level of significance.

The procedure (3.1.6) can alternatively be written as :

Reject H_o if $T = \dfrac{Y_1' Y_1}{(Y_2-CK)' (Y_2-CK) + Y_1'Y_1} < x_\alpha$,

accept H_o otherwise (3.1.7)

where x_α is the lower 100α percent point of the beta distribution with parameters $\frac{1}{2}(n - r)$, $\frac{1}{2}r_1$.

It is easily seen that the cumulative distribution function of the statistic T, when the assumed normal fixed effects model is true, is

$$F(x) = \sum_{j=0}^{\infty} p_j(\tfrac{1}{2}\lambda)\ B(x;\ \tfrac{1}{2}(n-r),\ \tfrac{1}{2}r_1 + j) \qquad (3.1.6)$$

where

$$\lambda = (CL\theta - CK)' (CL\theta - CK)/ \sigma^2.$$

The procedure (3.1.7) has the optimum property that among all test procedures of size α with the property that the power depends on the parameters only through λ, it is uniformly most powerful (UMP). For details of other optimum properties see Scheffe (1959). Assumption (3.1.1) with the added condition of normality of the basic variables may be explicitly split into the following components.

1° The random variables $\eta_1, \eta_2, \ldots, \eta_n$ are independent.

2° They have a common, but possibly unknown, variance

$$V(\eta_i) = \sigma^2,\ i = 1, 2, \ldots, n.$$

3° They are distributed normally.

4° $E(\eta_i) = a_{i1}\theta_1 + a_{i2}\theta_2 + \ldots + a_{ip}\theta_p,\ i = 1, 2, \ldots, n$

(3.1.9)

We shall study separately the effects of the breakdown of these components.

3.2 Effect of the wrong expectation model

3.2.1 The general case

Suppose the true model is $(B, \sigma^2 I)$ and not $(A, \sigma^2 I)$ as assumed by us. Then

$$E(Y) = E\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} T_1 B\theta \\ T_2 B\theta \\ T_3 B\theta \end{bmatrix} \tag{3.2.1}$$

If other assumption (1° to 3°) about η are valid, the sampling distribution of the statistic T is then the same as that of

$$\frac{\chi_2^2(n-r, \lambda_2)}{\chi_1^2(r_1, \lambda_1) + \chi_2^2(n-r, \lambda_2)} \tag{3.2.2}$$

where the non-centrality parameters of the independent non-central chi-squares are

$$\lambda_1 = (T_2 B\theta - CK)'(T_2 B\theta - CK)/\sigma^2$$

and

$$\lambda_2 = (T_1 B\theta)'(T_1 B\theta)/\sigma^2 \tag{3.2.3}$$

respectively.

Hence the distribution function of the test statistic T is

$$P(x, \lambda_1, \lambda_2, r_1, n - r)$$

$$= \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} p_i(\tfrac{1}{2}\lambda_1)\, p_j(\tfrac{1}{2}\lambda_2) B(x; \tfrac{1}{2}(n - r) + j, \tfrac{1}{2}r_1 + i)$$

$$= \sum_{k=o}^{\infty} \sum_{i=o}^{k} p_{k-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2)\, B(x; \tfrac{1}{2}(n - r) + i, \tfrac{1}{2}r_1 + k - i) \tag{3.2.4}$$

This can be easily seen by integrating the joint density function of the variables χ_1^2 and χ_2^2 over the region $T \leq x$. The expression (3.2.4) may be called the distribution function of a doubly non-central beta variable with parameters $\frac{1}{2}(n-r)$, $\frac{1}{2}r_1$ and non-centrality parameters λ_2, λ_1. If assumptions 1° - 4° hold, then $\lambda_2 = 0$, and $P \equiv P(x_\alpha, \lambda_1, \lambda_2, r_1, n - r)$, where x_α is the lower 100α percent point of the beta distribution with parameters $\frac{1}{2}(n - r)$, $\frac{1}{2}r$, is the power function of the size - αANOVA test. A non-zero value of λ_2 indicates that the assumption 4° is not-valid, and P, the probability of rejection of H_o, gives us an idea of the robustness of the ANOVA procedure.

About the behaviour of the function $P(x, \lambda_1, \lambda_2, r_1, n-r)$ we have the following result :

Result 3.1

$$P(x, \lambda_1, \lambda_2, r_1, n-r) > P(x, \lambda_1^*, \lambda_2, r_1, n-r)$$

where $\lambda_1^* < \lambda_1$

and

$$< P(x, \lambda_1, \lambda_2^*, r_1, n-r)$$

where $\lambda_2^* < \lambda_2$

which follows from the fact that $F(x, m, \lambda)$, the distribution function of a non-central chi-square variable with m degrees of freedom and non-centrality parameter λ is a decreasing function of λ.

Given $0 < \alpha, \beta < 1$, and positive integers n_1, n_2, let us denote by $S(\alpha, \beta, n_1, n_2)$ the set of points (λ_1, λ_2) for which

$$P(x_\alpha, \lambda_1, \lambda_2, n_1, n_2) = \beta$$

where x_α is defined by

$$P(x_\alpha, 0, 0, n_1, n_2) = \alpha$$

As an illustration, figure 3.1 gives the graphs of the sets $S(\alpha, \beta, n_1, n_2)$ for $\alpha = \beta = 0.05$, $n_2 = 20$ and $n_1 = 3, 4, 5$ and 10 where as figure 3.2 gives the graphs of the sets $S(\alpha, \beta, n_1, n_2)$ for $\alpha = 0.05$, $n_1 = 10$, $n_2 = 20$ and $\beta = 0.05, 0.10, 0.30$ and 0.50. Considerations of space do not permit presentation of such graphs for all useful ranges of parameters, but the author has developed computer programmes in FORTRAN which can be used in such cases.

These graphs are useful in studying robustness of the ANOVA procedure in the context of likely deviations from the assumed linear model as measured by the parameters λ_1, λ_2.

The traditional analysis of randomised blocks experiment makes use of the fundamental assumption that block and treatment effects are additive and there is no interaction between them. If there are t treatments and b blocks then, writing x_{ij} for the response of the j-th treatment of the i-th block, the assumed model is that the x_{ij}'s are independent normal variables with a common variance σ^2 and expectations given by

$$A : E(x_{ij}) = \alpha + \beta_i + \tau_j \; ; \; i = 1, 2, \ldots, b, \quad j = 1, 2, \ldots t$$

where α is the general mean, β_i the effect of the i-th block and τ_j the effect of the j-th treatment. Now if interactions do exist, the expectations would be given by

$$B : E(x_{ij}) = \alpha + \beta_i + \tau_j + \gamma_{ij} \; ; \; i = 1, 2, \ldots b \; ; \; j = 1, 2, \ldots t,$$

$$\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$$

where γ_{ij}'s are interaction parameters. Here the non-centrality parameters of F for testing the null hypothesis $H_0 : (\tau_1 = \tau_2 = \ldots = \tau_t = 0)$

are $\quad \lambda_1 = b \sum \tau_j^2 / \sigma^2$

$$\lambda_2 = \sum \sum \gamma_{ij}^2 / \sigma^2$$

Here λ_1 measures the deviation from the null hypothesis H_0 and λ_2 the deviation from the fundamental assumption about expectations (A), the effect of interactions. For any fixed value of the interaction effect λ_2, the test procedure remains unbiased if the treatment effect λ_1 is at least as large as λ_1^* where

$$P(x_\alpha \, , \lambda_1^*, \lambda_2, t-1, (t-1)(b-1)) = \alpha \; .$$

For example, in a randomised blocks experiment with five treatments and six blocks if $\lambda_2 = 8.6$ then, using the graph for $n_1 = 4$, $n_2 = 20$ in figure 3.1, the ANOVA procedure of size = .05 for testing the equality of treatment effects will remain unbiased for all deviations λ_1 of magnitude at least as large as 2.0.

FIGURE 3·1 (λ_1, λ_2) CURVE FOR $\alpha = 0{\cdot}05$, $\beta = 0{\cdot}05$

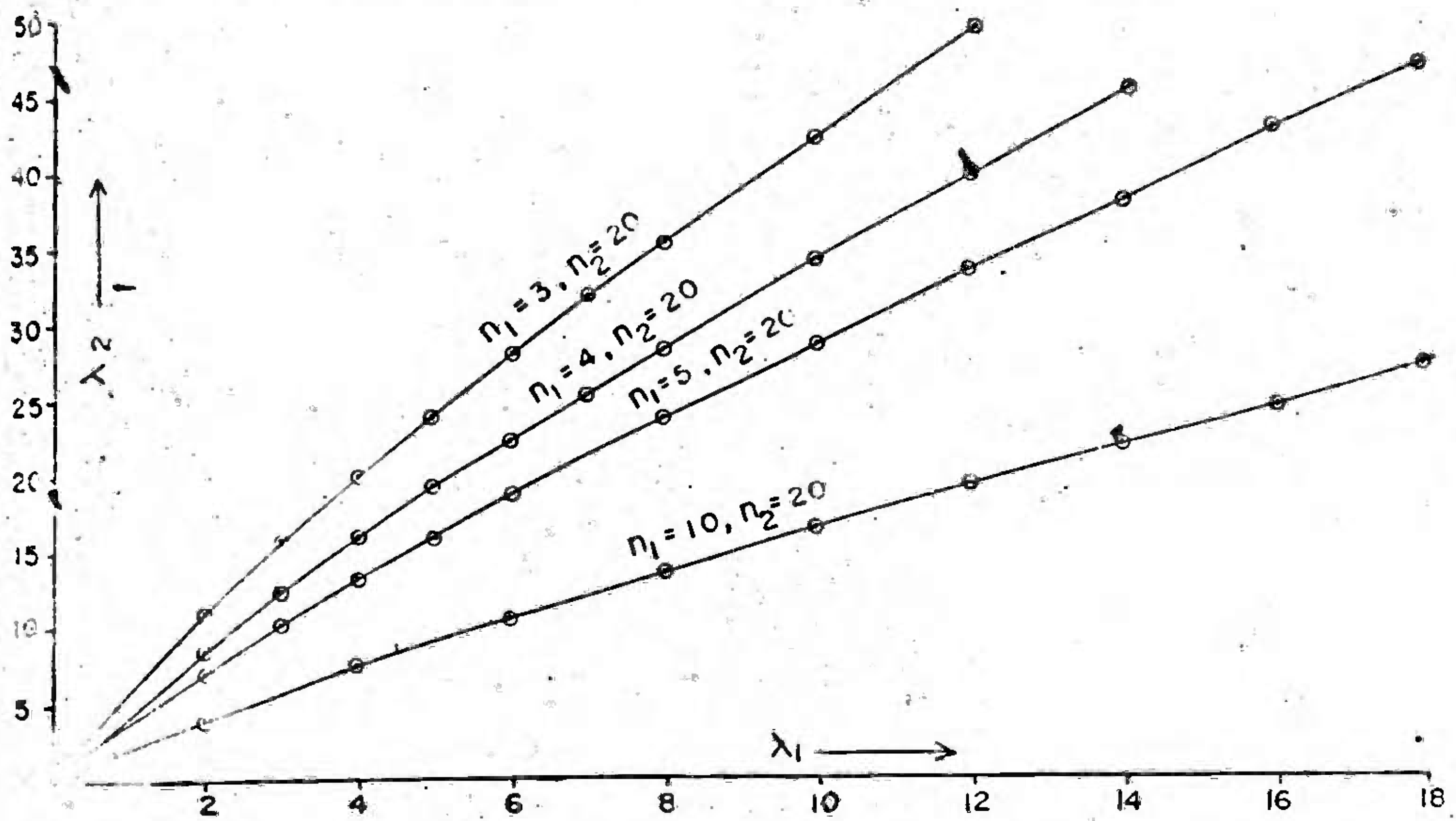
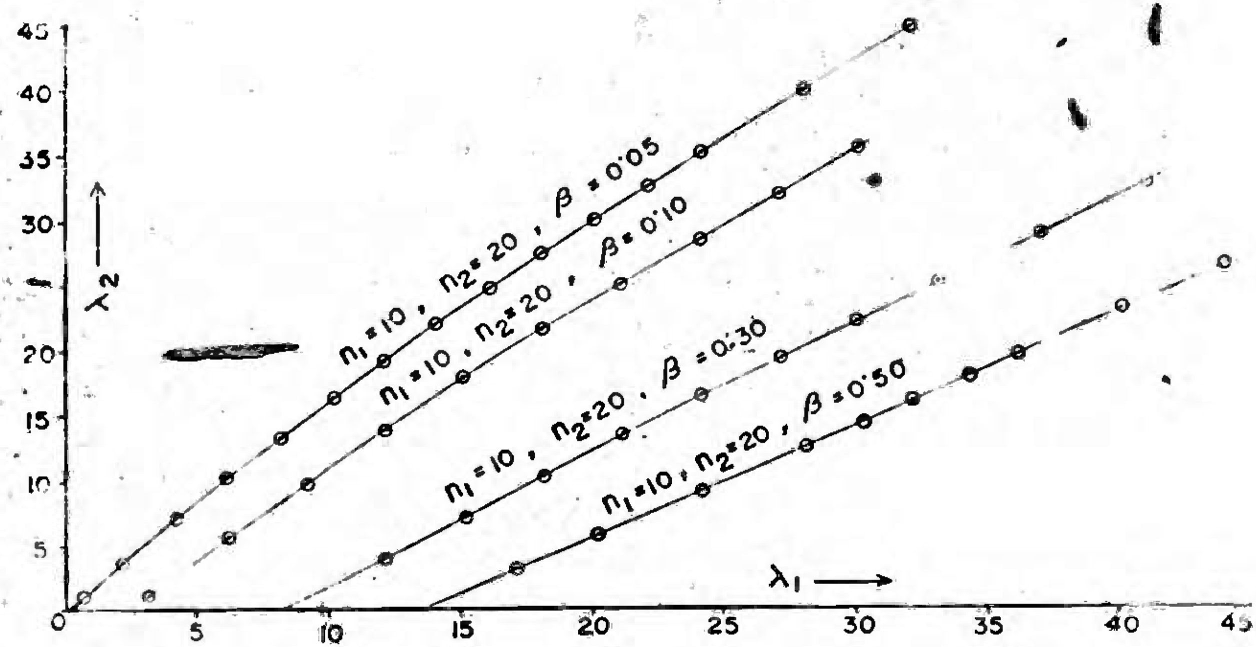


FIGURE 3·2 (λ_1, λ_2) GRAPHS FOR $\alpha = 0{\cdot}05$



Apart from being useful in the study of robustness of the ANOVA procedure, the function $P(x, \lambda_1, \lambda_2, m, n)$ has other applications too. Price (1964) has shown that P gives the probability of error for a particular binary signalling system in which the receiver tries to 'learn' the state of a multiple-parallel-link noise-perturbed channel. It has also been noticed by Turin (1959), Kailath (1961), Wishner (1962), Sebestyen (1961) and Braverman (1962) that the doubly non-central beta distribution has applications in problems of communications, radar and pattern recognition where quadratic form operationson normal data are involved.

3.2.2 A computational procedure for evaluating $P(x, \lambda_1, \lambda_2, r_1, n-r)$.

Let us choose the smallest c such that for a $k_1 \le c$

$$\sum_{k=o}^{c-1} \sum_{i=o}^{k} p_{k-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) + \sum_{i=o}^{k_1} p_{c-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) > 1-\varepsilon, 0 < \varepsilon < 1 \qquad (3.2.5)$$

We then define

$$\bar{P}(x, \lambda_1, \lambda_2, r_1, n-r)$$

$$= \sum_{k=o}^{c-1} \sum_{i=o}^{k} p_{k-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) B(x; \tfrac{1}{2}(n-r) + i, \tfrac{1}{2}r_1 + k - i)$$

$$+ \sum_{i=o}^{k_1} p_{c-i}(\tfrac{1}{2}\lambda_1) p_i(\tfrac{1}{2}\lambda_2)\, B(x; \tfrac{1}{2}(n-r) + i, \tfrac{1}{2}r_1 + c - i) \qquad (3.2.6)$$

Then, since i) $p_{k-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) > 0$ for all i, k

ii) $$\sum_{k=o}^{\infty} \sum_{i=o}^{k} p_{k-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) = 1$$

and

$$iii) \quad 0 < B(x ; \tfrac{1}{2}(n-r) + i , \tfrac{1}{2}r_1 + k - i) < 1$$

for all i, k,

we have

$$0 < P - \bar{P} < \varepsilon$$

The first incomplete beta ratio occuring in (3.2.6) namely, $B(x ; \frac{1}{2}(n-r) , \frac{1}{2}r_1)$ can be computed using the following algorithm. Successive beta ratios are obtained using the following recurrence relations :

$$i) \quad B(x ; m , n+1) = B(x ; m , n) + \frac{x^m (1-x)^n \; \Gamma(m+n)}{n \; \Gamma(m) \; \Gamma(n)}$$

$$ii) \quad B(x ; m+1 , n-1) = B(x ; m , n) - \frac{x^m (1-x)^{n-1}}{m} \; \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}$$

<u>Algorithm 3.1</u>

Let us write $\quad I(x ; p , q) \; = \; \int_0^x t^{p-1} (1-t)^{q-1} \, dt$

$$\text{Then} \quad B(x ; p , q) = \begin{cases} I(x ; p , q)/I(1 ; p , q) & \text{if} \quad x \geq \frac{1}{2} \\ 1 - I(1-x ; q , p)/I(1 ; p , q) , & \text{if} \quad x > \frac{1}{2} \end{cases}$$

We have, integrating $I(x ; p , q)$ by parts,

$$I(x ; p , q) = \frac{x^p (1 - x)^{q-1}}{p} + \frac{(q - 1)}{p(p + 1)} x^{p+1} (1 - x)^{q-2} + \dots$$

$$+ \frac{(q-1)(q-2)\dots(q-[q] + 1)}{p(p+1) \dots (p + [q] - 1)} x^{p+[q]+1} (1-x)^{q-[q]}$$

$$+ \frac{(q-1)(q-2)\dots(q -[q])}{p(p+1) \dots (p+[q] - 1)} I(x ; p +[q] , q - [q])$$

where $[q]$ is the greatest integer not exceeding q. Let us write $u = p + [q]$ and $1 - f = q - [q]$. Then expanding the integrand and integrating term by term we get

$$I(x;u,1-f) = \frac{x^u}{u} + \frac{f\,x^{u+1}}{u+1} + \frac{f(f+1)}{2!}\,\frac{x^{u+2}}{u+2} + \ldots$$

$$+ \frac{f(f+1)\ldots(f+r-1)}{r!}\,\frac{x^{u+r}}{u+r} + R_r$$

where

$$R_r = \frac{f(f+1)\ldots(f+r-1)}{r!}\,\frac{x^{u+r}}{u+r}\left[\frac{f+r}{r+1}\cdot\frac{u+r}{u+r+1}\,x\right.$$

$$\left. + \frac{(f+r)(f+r+1)}{(r+1)(r+2)}\cdot\frac{u+r}{u+r+2}\,x^2 + \ldots\right]$$

$$< \frac{f(f+1)\ldots(f+r-1)}{r!}\,\frac{x^{u+r}}{u+r}\,[x + x^2 + \ldots]$$

$$\leq \frac{f(f+1)\ldots(f+r-1)}{r!}\,\frac{x^{u+r}}{u+r} \qquad \text{if } x \leq \frac{1}{2}$$

Let us choose the smallest r satisfying

$$\frac{(q-1)(q-2)\ldots(q-[q])}{p(p+1)\ldots(p+[q]-1)}\;\frac{f(f+1)\ldots(f+r-1)}{r!}\;\frac{x^{u+r}}{u+r} < \delta$$

With this choice of r, if we replace $I(x;u,f)$ by

$$\sum_{j=0}^{r} \frac{f(f+1)\ldots(f+j-1)}{j!}\,\frac{x^{u+j}}{(u+j)},$$

error will not exceed δ in the computation of $I(x; p, q)$ for $x \leq \frac{1}{2}$. The evaluation of $\bar{P}$ using the above procedure often involves laborious computation. Since P is useful in studying the robustness of the ANOVA procedure when the assumed expectation model is not true it is desirable to have simple numerical approximations to P. In the following section we suggest two heuristic approximations. No attempt is made to estimate the errors of approximation in these cases, but numerical computation shows that the agreement between the exact value and the approximation is, in many cases, not too bad.

3.2.3 Approximation-I P using Jacobi polynomials

Since T is defined over the range $(0, 1)$, one may use methods described in section 2.2, to derive approximations for $P(x, \lambda_1, \lambda_2, r_1, n - r)$ the distribution function of T. Using the first term of (2.2.7) one gets

$$F_1 = B(x;\ a, b) \tag{3.2.7}$$

and using the full expression,

$$F_4 = B(x; a, b) - \beta(x; a, b)\{\ d(3) J(2, x; a+1, b+1) + d(4)\ J(3, x; a+1, b+1)\} \tag{3.2.8}$$

where a, b, d(3), d(4) and the polynomials J are defined in section 2.2 and $\theta(g)$ the g-th moment of T is

$$\theta(g) = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \int_0^1 T^g\, p_{k-i}\left(\tfrac{1}{2}\lambda_1\right) p_i\left(\tfrac{1}{2}\lambda_2\right)$$

$$\times\ \beta\left(T; \tfrac{1}{2}(n-r) + i, \tfrac{1}{2}r_1 + k - i\right) dT$$

Let us write $n_1 = \frac{1}{2}(n-r)$, $n_2 = \frac{1}{2}(n-r+r_1)$.

Then writing

$$\bar{\theta}(g) = \sum_{k=0}^{c-1} \sum_{i=0}^{k} p_{k-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) \frac{\Gamma(n_1+i+g)\,\Gamma(n_2+k)}{\Gamma(n_2+k+g)\,\Gamma(n_1+i)}$$

$$+ \sum_{i=0}^{k_1} p_{c-i}(\tfrac{1}{2}\lambda_1)\, p_i(\tfrac{1}{2}\lambda_2) \frac{\Gamma(n_1+i+g)\,\Gamma(n_2+c)}{\Gamma(n_2+c+g)\,\Gamma(n_1+i)}$$

where c and k_1 are as defined in (3.2.5), we have

$$0 < \theta(g) - \bar{\theta}(g) < \varepsilon$$

In actual computation we use $\bar{\theta}(g)$, instead of $\theta(g)$, $g = 1, 2, 3, 4$, for evaluating (3.2.8).

3.2.4 Approximation-II for P using Laguerre polynomials

The components $\chi_1^2(r_1, \lambda_1)$ and $\chi_2^2(n-r, \lambda_2)$ of T in (3.2.2) are independently distributed. Using the fourth order approximation for the density function described in section 2.1 and also the method given by Roy and Mohammed (1964) for approximating the density function of a non-central chi-square variable, we get,

writing

$$X_i = \chi_i^2 / 2\rho_i, \qquad i = 1, 2,$$

$$\rho_i = \frac{n_i + 2\lambda_i}{n_i + \lambda_i}, \qquad i = 1, 2,$$

$$n_1 = r_1$$

and $$n_2 = n - r,$$

and further noting that χ_1^2 and χ_2^2 are independent, as an approximation to the joint density function of X_1 and X_2

$$q(x_1, x_2) = \sum_{j=0}^{4} \sum_{i=0}^{4} A_i(m_1) A_j(m_2) g(x_1; m_1+i)\ g(x_2; m_2+j) \quad (3.2.8)$$

where

$$A_0(m) = 1 + b(3, m) + b(4, m)$$

$$A_1(m) = -3b(3, m) - 4b(4, m)$$

$$A_2(m) = 3b(3, m) + 6b(4, m)$$

$$A_3(m) = -b(3, m) - 4b(4, m)$$

$$A_4(m) = b(4, m)$$

$$m_i = \frac{(n_i + \lambda_i)^2}{2(n_i + 2\lambda_i)}, \qquad i = 1, 2$$

$$b(3, m_i) = \lambda_i^2\, m_i / 3(n_i + 2\lambda_i)^2, \quad i = 1, 2$$

and $$b(4, m_i) = \lambda_i^2\, m_i (n_i + 4\lambda_i) / 4(n_i + 2\lambda_i)^3, \; i = 1, 2.$$

From (3.2.8) we get, integrating q over $\frac{x_2}{x_1 + x_2} \leq x$, an approximation for $Prob\left[\frac{X_2}{X_1 + X_2} \leq x\right]$ as

$$\bar{Q}(x) = \sum_{j=0}^{4} \sum_{i=0}^{4} A_i(m_1)\ A_j(m_2)\ B(x; m_2 + j, m_1 + i) \quad (3.2.9)$$

The corresponding approximation for the distribution function of T at the point x is

$$\bar{F}_4 = \bar{Q}(y) \qquad (3.2.10)$$

where $$y = \frac{\rho_1 x}{\rho_1 x + \rho_2(1 - x)}$$

which reduces to

$$y = \frac{x(r_1 + 2\lambda_1)/(r_1 + \lambda_1)}{x(r_1+2\lambda_1)/(r_1+\lambda_1) + (1-x)(n-r+2\lambda_2)/(n-r+\lambda_2)} \qquad (3.2.11)$$

It is to be noted that if one uses Patnaik's (1949) approximation for the non-central chi-square distribution both for χ_1^2 and χ_2^2, one obtains the single term approximation

$$\bar{F}_1 = B(y; m_2, m_1) \qquad (3.2.12)$$

when y, m_1 and m_2 are as defined above. This is the first incomplete beta ratio occurring in (3.2.9).

3.2.5 Table for comparing the approximations

Table 3.1 gives the exact value of P and the values obtained from the approximations F_1, F_4, $\bar{F}_1$ and $\bar{F}_4$ at some selected points. It appears from a study of this table that :

(i) in most cases F_1 and $\bar{F}_1$ give fairly good results

(ii) generally $\bar{F}_1$ is better than F_1 and F_4

(iii) however, $\bar{F}_4$ is the best approximation in most cases.

TABLE 3.1 COMPARISON OF THE APPROXIMATIONS TO DOUBLY NON-CENTRAL BETA DISTRIBUTION.

					Approximation-I (using Jacobi polynomials)		Approximation-II (using Laguerre polynomials)		Exact
r_1	$(n-r)$	λ_1	λ_2	x	F_1	F_4	$\bar{F}_1$	$\bar{F}_4$	
5.0	20.0	2.0	7.104	.6	.0553	.0536	.0526	.0531	.0528
5.0	20.0	2.0	7.104	.8	.4636	.4597	.4640	.4650	.4433
4.0	20.0	3.0	12.3734	.6	.0299	.0308	.0286	.0287	.0358
4.0	20.0	8.0	28.46	.8	.4375	.4439	.4452	.4602	.4782
10.0	20.0	24.0	2.8464	.4	.4781	.4864	.4925	.4925	.4948
10.0	20.0	24.0	2.8464	.6	.9648	.9606	.9602	.9573	.9571
10.0	20.0	8.2552	0.0	.4	.1577	.1574	.1555	.1565	.1563
10.0	20.0	8.2552	0.0	.6	.7031	.7072	.7115	.7124	.7123
10.0	20.0	8.2552	0.0	.8	.9900	.9924	.9889	.9877	.9870
10.0	20.0	12.0	3.9714	.4	.1545	.1532	.1509	.1521	.1519
10.0	20.0	12.0	3.9714	.6	.7198	.7241	.7287	.7275	.7284
10.0	20.0	12.0	3.9714	.8	.9950	.9930	.9924	.9913	.9906

3.2.6 A sufficient condition for the ANOVA procedure procedure to retain its size and unbiasedness

In section 3.2.1, we have derived the sampling distribution of T defined in (3.1.7) under the assumed model. It has been found to have a doubly non-central beta distribution with non-centrality parameters λ_2 and λ_1 defined in (3.2.3).

If the null hypothesis and the assumed model are true then the two non-centrality parameters vanish simultaneously. There is, however, a wider class of models under which λ_2 is always zero and the non-centrality parameter λ_1 vanishes only when the null hypothesis is true. Whenever the assumed model belongs to this class, the analysis of variance procedure of course retains its property of unbiasedness and preassigned size. This class of models is characterised by our theorem 3.1.

Definition 3.1

The analysis of variance test procedure for testing $H_o(L\theta = K)$ **obtained under** the Gauss-Markoff model $(A, \sigma^2 I)$ is F-valid under the model $(B, \sigma^2 I)$, if under $(B, \sigma^2 I)$, the test statistic T is distributed as a doubly non-central beta variable with parameters $\frac{1}{2}(n - r)$, $\frac{1}{2}r_1$ and non-centrality parameters $\lambda_2 = 0$, $\lambda_1 = (L\theta - K)'D(L\theta-K)$ where D is a non-negative definite matrix, Or, equivalently, in the notation adopted in section 3.1, if $E(Y_1|B\theta) = 0$ and $E(Y_2|B\theta) = CL\theta$.

Definition 3.2

$\mu(A)$ is the estimation space under the Gauss-Markoff model $(A, \sigma^2 I)$.

Definition 3.3

The set of all n-vectors orthogonal to $\mu(A)$ is the error space under the model $(A, \sigma^2 I)$.

Theorem 3.1

A necessary and sufficient condition that the ANOVA procedure under the model $(A, \sigma^2 I)$ for testing $H_o : (L\theta = K)$ is F-valid under $(B, \sigma^2 I)$ is that

$$B = AW \tag{3.2.12}$$

where W is a $(p \times p)$ matrix such that

$$L(I - HW) = 0 \tag{3.2.13}$$

and $H = (A'A)^- (A'A)$, an idempotent matrix.

Proof :

We use the notation adopted in section 3.1.

Necessity

Since $E(Y_1 \mid B\theta) = 0$, the error space under $(A, \sigma^2 I)$ is a sub-space of the error space under $(B, \sigma^2 I)$, and this implies that there exists a matrix $W : (p \times p)$ such that

$$B = AW$$

we have $Y_2 = CL(A'A)^- A'\eta$ and so

$$E(Y_2 \mid B\theta) = CL\theta$$

$$\Rightarrow CL(A'A)^- A'AW\theta = CL\theta$$

or $$CLHW\theta = CL\theta$$

i.e. $$CL(I - HW)\theta = 0, \text{ for all } \theta$$

which means $L(I - HW) = 0$, C being non-singular.

Sufficiency

If $B = AW$ then the error space under $(A, \sigma^2 I)$ is a sub-space of the error space under $(B, \sigma^2 I)$ and hence all components of Y_1 have expectation zero also under $(B, \sigma^2 I)$. This means

$$E(Y_1 \mid B\theta) = 0.$$

Moreover,

$$E(Y_2 \mid B\theta) = CL(A'A)^- A'B\theta$$

$$= CL(A'A)^- A'AW\theta = CLHW\theta$$

$$= CLHW\theta + CL(I - HW)\theta \quad \text{because of (3.2.13)}$$

$$= CL\theta$$

It is to be noted that when the conditions of the above theorem are satisfied though the test procedure constructed under $(A, \sigma^2 I)$ retains its size and the property of unbiasedness in $(B, \sigma^2 I)$, it no longer has the optimum properties mentioned in section 3.1. We note in particular the following deficiencies :

i) Y_2 is not necessarily the BLUE for $CL\theta$ under $(B, \sigma^2 I)$

ii) Y_1 does not necessarily contain all linear combinations of η belonging to the error space under the model $(B, \sigma^2 I)$.

If in addition to a test procedure's being F-valid, we want the estimates of the parametric functions involved in the null hypothesis to remain best linear unbiased estimates in the changed model, additional restrictions will be required.. A result in this connection is stated in the following theorem.

Theorem 3.2

A necessary and sufficient condition that for all parametric functions which are estimable under the model $(B, \sigma^2 I)$,

a) the BLUEs under $(A, \sigma^2 I)$ will remain BLUEs under $(B, \sigma^2 I)$

b) the test procedures constructed under $(A, \sigma^2 I)$ will remain F-valid under $(B, \sigma^2 I)$

is
$$A = B + (I - P_B)X \tag{3.2.14}$$

where X is any matrix such that

$$\mu(X') \cap \mu(B') = 0 \tag{3.2.15}$$

or, equivalently,

$$A = B + (I - P_B)\, F[I + (I - P_{B'})D]^{\perp'} \tag{3.2.16}$$

where D and F are arbitrary.

For the proof of this theorem we shall use the following due to Mitra and Rao (1968).

Theorem 3.3 (Mitra and Rao)

If for every estimable parametric function the BLUE under $(X_o, \sigma^2 I)$ is also BLUE under $(X, \sigma^2 I)$, it is necessary and sufficient that X is of the form

$$X = X_o + (I - P_{X_o})W \tag{3.2.17}$$

where W is any matrix such that

$$\mu(W') \cap \mu(X'_o) = 0 \tag{3.2.18}$$

or equivalently

$$X = X_o + (I - P_{X_o}) F[I + (I - P_{X'_o})]D^{\perp'} \tag{3.2.19}$$

where D and F are arbitrary.

Proof of the theorem 3.2

Using the theorem 3.3 and further noting that, in the notation adopted in section 3.1, the condition (3.2.14) or its equivalent (3.2.16) implies that $E(Y_1 \mid B\theta) = 0$ and $E(Y_2 \mid B\theta) = CL\theta$, the proof of the theorem immediately follows.

Remark

The condition (3.2.14) is necessary and sufficient for (a) and implies (b).

3.2.7 An expression for the average power of the ANOVA procedure assuming a priori distribution of the non-centrality parameters

Here our object is to obtain a computational procedure for evaluating an expression for the average power of the ANOVA test, by introducing some a priori distributions of the non-centrality parameters. This average power may be used to study the effect of the deviation of the assumed expectation model from the true expectation.

For fixed λ_1 and λ_2, the power of the ANOVA test is given by $P(x, \lambda_1, \lambda_2, r_1, n - r)$ defined in (3.2.4) where x is the lower 100α percent point of the beta distribution with parameters $\frac{1}{2}(n - r), \frac{1}{2}r_1$.

Writing $\delta_i = \frac{1}{2}\lambda_i$, $i = 1, 2$, we consider the following a priori marginal distributions of δ_i, $i = 1, 2$, $(0 \leq \delta_i < \infty)$

$$f_i(\delta_i)\, d\delta_i = \frac{\exp[-\delta_i / a_i b_i]\; \delta_i^{\frac{1}{b_i} - 1}\; d\delta_i}{(a_i b_i)^{1/b_i}\; \Gamma(\frac{1}{b_i})}, \quad i = 1, 2 \qquad (3.2.20)$$

where

$$a_i = E(\delta_i)$$

$$b_i = \frac{V(\delta_i)}{a_i^2} \qquad (3.2.21)$$

We further assume, for simplicity, that the variables δ_1 and δ_2 are independently distributed. Then the average power, averaging over δ_1 and δ_2, is

$$\bar{P} = \int_0^\infty \int_0^\infty P(x, 2\delta_1, 2\delta_2, r_1, n-r) f_1(\delta_1) f_2(\delta_2) d\delta_1 \, d\delta_2$$

$$= \sum_{i=0}^\infty \sum_{j=0}^\infty B(x; \tfrac{1}{2}(n-r) + j, \tfrac{1}{2}r_1 + i) \int_0^\infty \int_0^\infty p_j(\delta_2) p_i(\delta_1) f_1(\delta_1)$$

$$\times f_2(\delta_2) \, d\delta_1 \, d\delta_2. \qquad (3.2.22)$$

Now $\int_0^\infty \dfrac{\exp(-\delta_1) \, \delta_1^i}{i!} f_1(\delta_1) \, d\delta_1$

$$= \frac{\Gamma(i + \frac{1}{b_1}) \, (a_1 b_1)^i}{\Gamma(i+1) \, \Gamma(\frac{1}{b_1}) (1 + a_1 b_1)^{i+(1/b_1)}}$$

$$= c\,(a_1, b_1, i) \qquad \text{(say)}.$$

Hence from (3.2.21)

$$\bar{P} = \sum_{i=0}^\infty \sum_{j=0}^\infty c\,(a_1, b_1, i) \, c\,(a_2, b_2, j) B(x; \tfrac{1}{2}(n-r)+j, \tfrac{1}{2}r_1 + i)$$

$$= \sum_{k=0}^\infty \sum_{j=0}^k c\,(a_1, b_1, k-j) \, c\,(a_2, b_2, j) B(x; \tfrac{1}{2}(n-r)+j, \tfrac{1}{2}r_1 + k - j)$$

or, writing $T(j, k) = c\,(a_1, b_1, k-j) \, c\,(a_2, b_2, j)$,

$$\bar{P} = \sum_{k=0}^\infty \sum_{j=0}^k T(j,k) B(x; \tfrac{1}{2}(n-r)+j, \tfrac{1}{2}r_1 + k - j)$$

For numerical evaluation of this average power we use the following procedure.

3.2.8 Computational procedure for evaluating (3.2.22)

We define

$$\bar{P}^* = \sum_{k=0}^{c-1} \sum_{j=0}^{k} T(j,k)\, B(x; \tfrac{1}{2}(n-r) + j, \tfrac{1}{2}r_1 + k - j) + \sum_{j=0}^{k_1} T(j, c)\, B(x; \tfrac{1}{2}(n-r) + j, \tfrac{1}{2}r_1 + c - j) \qquad (3.2.23)$$

where c and $k_1 (\leq c)$ are chosen in such a way that for a given $\varepsilon (0 < \varepsilon < 1)$ we have

$$\sum_{k=0}^{c-1} \sum_{j=0}^{k} T(j,k) + \sum_{j=0}^{k_1} T(j, c) > 1 - \varepsilon .$$

Since $\sum_{k=0}^{\infty} \sum_{j=0}^{k} T(j, k) = 1$ and all $T(j. k)$ are positive and

$0 < B < 1$, we have $0 < \bar{P} - \bar{P}^* < \varepsilon$ (3.2.24)

As an illustrative example consider the evaluation of the average power when

$a_1 = 6.0$, $a_2 = 1.986$, $b_1 = 0.001$, $b_2 = 0.001$,

$\alpha = .05$, $r_1 = 10$ and $n - r = 20.$

Here $x = .45999$ and $\bar{P}^* = 0.30438$

correct to five places of decimals. It is to be noted that

$$P(.45999, 12.0, 3.872, 10, 20) = 0.30.$$

3.3. Effect on the power function when the assumed dispersion is wrong

3.3.0 Suppose we compute the test statistic for the $H_0(L\,\theta = K)$

under the model $(A, \sigma^2 I)$ whereas the true model is (A, Σ), the normality condition remaining unaffected. Then what will be the effect of this deviation on the power function of the test? In the notation adopted in section 3.1, our primary variables, in the canonical form, are

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} \eta \qquad (3.3.1)$$

and the statistic obtained for testing H_0 is

$$T = \frac{Y_1' Y_1}{Y_1' Y_1 + (Y_2 - CK)' (Y_2 - CK)} \qquad (3.3.2)$$

It is easy to check that $Y_1' Y_1$ is distributed as $\sum_{i=1}^{n-r} \lambda_i \chi_i^2(1)$ where $\lambda_1, \lambda_2, \ldots, \lambda_{n-r}$ are the (n-r) characteristic roots of the matrix $\Sigma_{11} = T_1 \Sigma T_1'$ and $\chi_i^2(1)$, $i = 1, 2, \ldots, n-r$, are independently distributed chi-square variables with one degree of freedom; and $(Y_2 - CK)' (Y_2 - CK)$ is distributed as $\sum_{i=1}^{r_1} \mu_i \chi_i^2(1, \delta_i^2)$ where $\mu_1, \mu_2, \ldots, \mu_{r_1}$ are the r_1 characteristic roots of the matrix $\Sigma_{22} = T_2 \Sigma T_2'$, $\chi_i^2(1, \delta_i^2)$, $i = 1, 2, \ldots, r_1$, are independently distributed non-central chi-square variables, and δ_i is the i-th component of the vector

$DC(L\theta - K)$ where $D : (r_1 \times r_1)$ satisfies $D \; \Sigma_{22} D' = I$.

In general, $Y_1' \, Y_1$ and $(Y_2 - CK)' \, (Y_2 - CK)$ will not be independent. They will, of course, be so if Y_1 and Y_2 are uncorrelated, that is, if $T_1 \, \Sigma \, T_2' = 0$, or equivalently if the row space of $T_2 \, \Sigma = CL(A'A)^{-} A' \, \Sigma$ is a sub-space of $\mu(A)$. We consider below an example where Y_1 and Y_2 are non-correlated even though $D(\eta)$ is different from $\sigma^2 I$.

Example : Consider the model

$$E \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \theta ,$$ and the true dispersion

$$D \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

The statistic for testing $H_0 : (\theta = 0)$ under the assumption of normality, independence and homoscedasticity of η_1 and η_2 is

$$T = \frac{u_2^2}{u_1^2 + u_2^2}$$

where $$u_1 = \frac{1}{\sqrt{2}} (\eta_1 + \eta_2)$$

$$u_2 = \frac{1}{\sqrt{2}} (\eta_1 - \eta_2).$$

Here, under the true model

$$Cov\ (u_1, u_2) = [\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}] \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = 0$$

We shall now obtain an expression for the exact distribution of T when Y_1 and Y_2 are independent.

3.3.1 Distribution of T when Y_1 and Y_2 are independently distributed.

About the distribution of a linear function of independent chi-square variables we have the following theorem due to Robbins and Pitman (1949).

Theorem 3.4 (Robbins and Pitman)

$$\text{Let} \quad X = \chi^2(m) + a_1\, \chi^2(m_1) + \ldots + a_r\, \chi^2(m_r) \qquad (3.3.3)$$

where i) $\chi^2 s$ are independently distributed

and ii) $a_i \geq 1, \quad i = 1, 2, \ldots, r$

Define c_k by the identity

$$\prod_{i=1}^{r} [a_i^{-\frac{1}{2}m_i} [1 - (1 - \frac{1}{a_i}) u]^{-\frac{1}{2}m_i}] = \sum_{k=0}^{\infty} c_k u^k, \ (|u| \leq 1) \qquad (3.3.4)$$

Then obviously

$$c_k \geq 0 \ (k = 0, 1, 2, \ldots), \ \Sigma\, c_k = 1.$$

Let $M = m + m_1 + m_2 + \ldots + m_r$ (3.3.5)

Then for every x, the density function of $\frac{1}{2}X$ can be written as

$$f(x) = \sum_{k=0}^{\infty} c_k \, g(x ; \frac{M}{2} + k) \qquad (3.3.6)$$

Using the method adopted in Robbins and Pitman (1949) for proving the above theorem, the following result is easily obtained about the distribution of a linear function of independent non-central chi-square variables.

Theorem 3.5

Let $X = \chi^2(m_1, \alpha_1) + a_2 \chi^2(m_2, \alpha_2) + \ldots + a_r \chi^2(m_r, \alpha_r)$

(3.3.7)

where i) the non-central chi-square variables are independent

and ii) $a_i \geq 1$, $i = 2, 3, \ldots, r$

Define constants c_k by the identity

$$\prod_{i=1}^{r} [a_i^{-\frac{1}{2} m_i} (1 - [1 - \frac{1}{a_i}] u)^{-\frac{1}{2} m_i}]$$

$$\times \exp [- \frac{1}{2} \sum_{i=1}^{r} \alpha_i (1 - a_i^{-1} u [1 - (1 - \frac{1}{a_i}) u]^{-1})]$$

$$= \sum_{k=0}^{\alpha} c_k u^k, \ |u| \leq 1 \ , \quad a_1 = 1 \qquad (3.3.8)$$

Obviously $c_k \geq 0 \quad (k = 0, 1, 2, \ldots), \quad \Sigma\, c_k = 1.$

Let us further write

$$M = m_1 + m_2 + \ldots + m_r$$

Then for every x, the density function of $\frac{1}{2}X$ can be written as

$$f(x) = \sum_{k=0}^{\infty} c_k\, g(x; \tfrac{1}{2}M + k) \qquad (3.3.10)$$

Using the above two theorems, the distribution function of T, when

$$Y_1'\, Y_1 = \Sigma\, \lambda_i\, \chi_i^2\,(1) \text{ and } (Y_2 - CK)'(Y_2 - CK) = \Sigma\, \mu_j\, \chi_j^2(1, \delta_j^2)$$

are independent, is easily seen to be

$$\text{Prob}\,[T \leq x]$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i\, d_j\, B(z; \tfrac{1}{2}(n-r) + i, \tfrac{1}{2}r_1 + j) \qquad (3.3.11)$$

where, writing

i) $\alpha_i = \lambda_i \,/ \min_i \lambda_i, \qquad i = 1, 2, \ldots, n-r,$

ii) $\beta_j = \mu_j \,/ \min_j \mu_j, \qquad j = 1, 2, \ldots, r_1,$

and iii)

$$z = \left[1 + \frac{\min_i \lambda_i\,(1 - x)}{\min_j \mu_j\, x}\right]^{-1} \qquad (3.3.12)$$

the constants c_i and d_j are defined by identities

$$\sum_{i=0}^{\infty} c_i u^i = \prod_{i=1}^{n-r} [\alpha_i^{-\frac{1}{2}} (1 - [1 - \frac{1}{\alpha_i}] u)^{-\frac{1}{2}}] \qquad (3.3.13)$$

and

$$\sum_{j=0}^{\infty} d_j u^j = \prod_{j=1}^{r_1} [\beta_j^{-\frac{1}{2}} (1 - [1 - \frac{1}{\beta_j}] u)^{-\frac{1}{2}}]$$

$$\times \exp [- \frac{1}{2} \sum_{j=1}^{r_1} \delta_j^2 (1 - \beta_j^{-1} u [1 - (1 - \frac{1}{\beta_j}) u]^{-1})] \qquad (3.3.14)$$

where $|u| \leq 1$.

As regards computation, we observe that if we choose, for a given ε $(0 < \varepsilon < 1)$, integers p and k_1 such that

$$\sum_{k=0}^{p-1} c_i d_{k-i} + \sum_{i=0}^{k_1} c_i d_{p-i} > 1 - \varepsilon \qquad (3.3.15)$$

we have

$$\text{Prob}(T \leq x) - \sum_{k=0}^{p-1} \sum_{i=0}^{k} c_i d_{k-i} B(z; \tfrac{1}{2}(n-r) + i, \tfrac{1}{2} r_1 + k - i)$$

$$- \sum_{i=0}^{k_1} c_i d_{p-i} B(z; \tfrac{1}{2}(n-r) + i, \tfrac{1}{2} r_1 + p - i) < \varepsilon \qquad (3.3.16)$$

When the null hypothesis is true $\delta_j^2 = 0$, $j = 1, 2, \ldots, r_1$, and hence in that case the d's of (3.3.10) can be obtained from the identity

$$\sum_{j=0}^{\infty} d_j u^j = \prod_{j=1}^{r_1} [\beta_j^{-\frac{1}{2}} (1 - [1 - \frac{1}{\beta_j}] u)^{-\frac{1}{2}}] \qquad (3.3.17)$$

The expression (3.3.10) is useful in finding the power function of the test procedure (3.1.7) based on T only when the 'between' and 'within' sums of squares are independent.

In many situations this is not case and we require an expression for the distribution of T when Y_1 and Y_2 are correlated. Since an exact expression is difficult to obtain we suggest the following simple approximation to the distribution of T.

3.3.2 An approximation to the distribution function of T when Y_1 and Y_2 are correlated

Let us write

$$U_1 = (Y_2 - CK)' (Y_2 - CK)$$

$$U_2 = Y_1' Y_1 \tag{3.3.18}$$

Then we have

$$m_1 = E(U_1) = \sum_{i=1}^{r_1} \mu_i (1 + \delta_i^2)$$

$$m_2 = E(U_2) = \sum_{i=1}^{n-r} \lambda_i$$

$$\mu_{20} = V(U_1) = 2 \sum_{i=1}^{r_1} \mu_i^2 (1 + 2\delta_i^2)$$

$$\mu_{02} = V(U_2) = 2 \sum_{i=1}^{n-r} \lambda_i^2$$

$$\mu_{11} = Cov\ (U_1, U_2) = 2 \sum_{i=1}^{n-r} \sum_{j=1}^{r_1} \lambda_i\ \mu_j\ . \rho_{ij}^2 ,$$

where ρ_{ij} is the coefficient of correlation between the i-th component of Y_1 and the j-th component of Y_2.

Now, if we use the method developed in section 2.4, we get the following approximation to the distribution function

$$Prob \quad (T \leq x) \quad \text{of} \quad T = \frac{U_2}{U_1 + U_2}$$

$$P(x) = B(z;\, q,\, p) + R\left(\frac{q}{p}\right)^{\frac{1}{2}} \frac{\Gamma(q+p)}{\Gamma(q)\,\Gamma(p)}\, z^{q}(1-z)^{p}\left[-1 + \frac{p+q}{q} z\right] \tag{3.3.19}$$

where p, q, z and R are as defined below

$$p = \left[\sum_{i=1}^{r_1} \mu_i(1+\delta_i^2)\right]^2 \Big/ 2 \sum_{i=1}^{r_1} \mu_i^2(1+2\delta_i^2) \tag{3.3.20}$$

$$q = \left[\sum_{i=1}^{n-r} \lambda_i\right]^2 \Big/ 2 \sum_{i=1}^{n-r} \lambda_i^2 \tag{3.3.21}$$

$$R = \sum \lambda_i \mu_j \rho_{ij}^2 \Big/ \left[\left(\sum \lambda_i^2\right) \times \left(\sum \mu_j^2(1+2\delta_j^2)\right)\right]^{\frac{1}{2}} \tag{3.3.22}$$

and

$$z = \left[1 + \frac{\sum_j \mu_j(1+\delta_j^2) \times \sum_i \lambda_i^2}{\sum_i \lambda_i \times \sum_j \mu_j^2 (1+2\delta_j^2)} \cdot \frac{1-x}{x}\right]^{-1} \tag{3.3.23}$$

If U_1 and U_2 are independently distributed then $R = 0$, and the approximation (3.3.19) reduces to

$$F(x) = B(z\,;\, n,\, m).$$

This in fact is the approximation one would obtain by assuming independence of U_1 and U_2 and then approximating each of the distributions of the variables U_1 and U_2 by the distribution of a variable of the form $a\chi^2(b)$ where a and b are obtained by equating the first two moments of the approximator to the corresponding moments of the approximant. This has been the approach of Patnaik (1949). The term

$$R\left(\frac{q}{p}\right)^{\frac{1}{2}} \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\, z^q\, (1-z)^p \left(-1+\frac{q+p}{q}z\right) \tag{3.3.24}$$

in (3.3.19) may be thought of as the approximate correction needed because of dependence of U_1 and U_2. Other quantities remaining constant, the correction term is directly proportional to R, the coefficient of correlation between U_1 and U_2. For a large R the magnitude of the correction term will be large and may even exceed $B(z;\, q,\, p)$ - specially when we are trying to use (3.3.19) for approximately finding the power of an ANOVA procedure, the size of a test usually being small - making the approximation highly unreliable. We, therefore, do not recommend the use of the approximation except when it is known that R is small.

3.3.3 Accuracy of the approximation (3.3.19)

In order to have an idea about the accuracy of this approximation we study its performence in two particular cases when exact results are computable.

Let X_1, X_2 and X_3 be three independent random variables with density functions $g(x_1;\, a)$, $g(x_2;\, b)$ and $g(x_3;\, c)$ respectively.

We define

$$U_1 = X_1 + X_2$$

$$U_2 = X_1 + X_3$$

Then $\rho(U_1, U_2)$ = the coefficient of correlation between

$$U_1 \text{ and } U_2 = \frac{a}{[(a+b)\ (a+c)]^{\frac{1}{2}}}$$

Using (3.3.19), we get the following approximation to the distribution Prob $(V \leq x)$ of $V = U_1/U_2$

$$P^*(x) = 1 - F(z) \tag{3.3.25}$$

when

$$p = a + b$$

$$q = a + c$$

$$z = \frac{1}{1+x}$$

$$R = \rho(U_1, U_2)$$

and $F(z)$ is as in (3.3.19).

Let us now consider the exact expression for the distribution of V in the following cases.

<u>Case I</u> $a = b = c = 1$, $\rho = \frac{1}{2}$

David and Fix (1960) have shown that the joint density function of U_1 and U_2 is given by

$$f(U_1, U_2) = e^{-U_1}(1 - e^{-U_2}) \quad \text{if} \quad U_1 > U_2$$

$$= e^{-U_2}(1 - e^{-U_1}) \quad \text{if} \quad U_1 < U_2$$

Therefore, when x is less than 1,

$\text{Prob}(V \leq x) = \text{Prob}(U_1 \leq U_2 x)$

$$= \int_0^\infty e^{-U_2} \int_0^{U_2 x} (1 - e^{-U_1})\, dU_1\, dU_2 = \frac{x^2}{1+x}.$$

Similarly when x is greater than 1

$$\text{Prob}(V \leq x) = \text{Prob}\left(\frac{U_2}{U_1} \geq \frac{1}{x}\right) = 1 - \text{Prob}\left(\frac{U_2}{U_1} \leq \frac{1}{x}\right)$$

$$= 1 - \frac{1}{x(1+x)}.$$

In the following table (table 3.2) we tabulate the exact distribution function of V and also its approximation using (3.3.25).

TABLE 3.2 THE DISTRIBUTION FUNCTION OF V (WHEN $a = b = c = 1$) AND ITS APPROXIMATION

x	Prob$(V \leq x)$ Approximation	Exact	x	Prob$(V \leq x)$ Approximation	Exact
0.1	.0065	.0091	1.3	.6209	.6656
0.2	.0355	.0333	1.4	.6534	.7024
0.3	.0843	.0692	1.5	.6826	.7333
0.4	.1447	.1143	2.0	.7901	.8333
0.5	.2099	.1667	2.5	.8553	.8857
0.6	.2752	.2252	3.0	.8965	.9167
0.7	.3380	.2882	3.5	.9236	.9365
0.8	.3967	.3556	4.0	.9421	.9500
0.9	.4507	.4263	5.0	.9645	.9667
1.0	.5000	.5000	6.0	.9767	.9762
1.1	.5445	.5671	8.0	.9885	.9861
1.2	.5846	.6212	10.0	.9934	.9909

Case II $\quad a = b = 1, \quad c = 2, \quad \rho = (\frac{1}{6})^{\frac{1}{2}}$

The joint density function of U_1 and U_2 in this case has been given by David and Fix (1960) as

$$f(U_1, U_2) = e^{-U_1} [1 - (U_2 + 1) e^{-U_2}] \quad \text{if} \quad U_2 < U_1$$

$$= e^{-U_2} [(U_2 - U_1 + 1) - (U_2 + 1) e^{-U_1}] \quad \text{if} \quad U_1 < U_2$$

The distribution function of V, therefore, may be written as

$$\text{Prob}(V \leq x) = (x - 2)(1 - x) + \frac{1}{1+x} + \frac{1}{(1+x)^2} \quad \text{if} \quad x < 1$$

$$= 1 - \frac{1}{x(1+x)^2} \quad \text{if} \quad x > 1$$

In table 3.3 we tabulate the distribution function of V and present along with it the approximation (3.3.25).

TABLE 3.3 THE DISTRIBUTION FUNCTION OF V (WHEN $a = b = 1$, $c = 2$) AND ITS APPROXIMATION

x	Prob $(V \leq x)$ Approimation	Exact	x	Prob $(V \leq x)$ Approximation	Exact
0.1	.0246	.0255	1.1	.7590	.7939
0.2	.0944	.0878	1.2	.7930	.8278
0.3	.1886	.1707	1.3	.8218	.8546
0.4	.2892	.2645	1.4	.8461	.8760
0.5	.3854	.3611	1.5	.8669	.8933
0.6	.4726	.4556	2.0	.9328	.9444
0.7	.5491	.5443	2.5	.9641	.9673
0.8	.6149	.6242	3.0	.9800	.9792
0.9	.6711	6933	4.0	.9932	.9900
1.0	.7187	.7500	5.0	.9977	.9944

If we look at the tables 3.2 and 3.3 we notice that the performance of the approximation (3.3.19) is not too bad at least in these two particular cases. We may also notice that the performance of the approximation is better in the second case where ρ is only about 0.40.

3.3.4 Applications of the approximation (3.3.19)

Let us consider the problem of testing equality of means of s groups under the assumed model

$$A : E(x_{ij}) = \mu_i \ , V(x_{ij}) = \sigma^2 , i = 1 , 2 ,\dots s; j = 1 , 2 , n_i ,$$

where the true expectations and variances are

$$B : E(x_{ij}) = \mu_i , V(x_{ij}) = \sigma_i^2 , i = 1 ,2 ,\dots s; j = 1 , 2 ,\dots n_i ,$$

x_{ij} being the j-th observation from the i-th group.

Let us write

$$n = \sum_{i=1}^{s} n_i \ , \quad \bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i \ , \quad \bar{\bar{x}} = \sum_{i=1}^{s} n_i \bar{x}_i / n \ ,$$

$$\mu = \frac{1}{n}\sum_{i=1}^{s} n_i \mu_i \ , \quad \sigma_\mu^2 = \sum_{i=1}^{s} n_i (\mu_i - \mu)^2 / n \qquad \text{and}$$

k_{ri} = r-th cumulant of the i-th group, r = 2, 3, 4; i = 1,2,...s.

In the general case where the observations may not be normally distributed the moments of the 'between' sum of squares

$U_1 = \sum_{i=1}^{s} n_i \, (\bar{x}_i - \bar{\bar{x}})^2$ and 'within' sum of squares $U_2 = \sum_{i=1}^{s} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$

are given by (see David and Johnson (1951a))

$$E(U_1) = \sum_{i=1}^{s} k_{2i} \, (1 - \frac{n_i}{n}) + n \, \sigma_\mu^2$$

$$E(U_2) = \sum_{i=1}^{s} (n_i - 1) \, k_{2i}$$

$$V(U_1) = \sum_{i=1}^{s} \frac{k_{4i}}{n_i} (1 - \frac{n_i}{n})^2 + 2 \sum_{i=1}^{s} k_{2i}^2 \, (1 - \frac{2n_i}{n})$$

$$+ \frac{2}{n^2} \Big(\sum_{i=1}^{s} n_i \, k_{2i} \Big)^2 + 4 \sum_{i=1}^{s} n_i \, k_{2i} \, (\mu_i - \mu)^2 + 4 \sum_{i=1}^{s} k_{3t} (\mu_i - \mu)(1 - \frac{n_i}{n})$$

$$V(U_2) = \sum_{i=1}^{s} (n_i - 1)^2 \Big(\frac{k_{4i}}{n_i} + \frac{2k_{2i}^2}{n_i - 1} \Big)$$

$$\mathrm{Cov}(U_1, U_2) = \sum_{i=1}^{s} \frac{n_i - 1}{n_i} (1 - \frac{n_i}{n}) \, k_{4i} + 2 \sum_{i=1}^{s} (n_i - 1)(\mu_i - \mu) \, k_{3i}$$

When the populations are normal $k_{3i} = k_{4i} = 0, \; i = 1, 2, \ldots s,$ $E(U_1), E(U_2)$ remain unaffected, $\mathrm{Cov}(U_1, U_2) = 0$ and the expressions for variances reduce to

$$V(U_1) = 2 \, \Sigma \, k_{2i}^2 \, (1 - \frac{2n_i}{n}) + \frac{2}{n^2} (\Sigma \, n_i \, k_{2i})^2 + 4 \, \Sigma \, n_i \, k_{2i} (\mu_i - \mu)^2$$

$$V(U_2) = 2\Sigma (n_i - 1) \, k_{2i}^2$$

As an illustration, if in a situation like this there are three groups and thirteen observations and if $E(U_1) = 3$, $V(U_1) = 4$, $E(U_2) = 11$, $V(U_2) = 20$ then the approximate value of the power of the usual ANOVA test with assumed size = 0.05 is given by

$$B(0.46644;\ 6.05,\ 2.25) = 0.0577$$

where 0.46644 is our Z as in (3.3.23), 0.54928 being the lower 5 percent point of the beta distribution with parameters 5 and 1.

If, however, the true variances are different even for the observations belonging to the same group, then $\mathrm{Cov}(U_1, U_2)$ will be non-zero and the full expression (3.3.19) is to be used instead of the first term alone.

The following table presents the numerical results of a study when the sums of squares U_1 and U_2 are correlated using the approximation (3.3.19). The table gives for some selected values of r_1, $n-r$, p, q and R, the approximate power using (3.3.19) when the assumed level of significance is 0.05. For computation of this table we have assumed $V(U_1) = 2r_1$ and $V(U_2) = 2(n-r)$. Because of the form of the correction term (3.3.24) we consider for the table only the negative values of R, it being fairly simple to get the value of power for the corresponding positive values of R from the table.

A computer programme for computing $F(x)$ for given $E(U_1)$, $E(U_2)$, $V(U_1)$, $\mathrm{Cov}(U_1, U_2)$, $V(U_2)$ and x is given in the appendix.

TABLE 3.4 APPROXIMATE POWER OF THE ANOVA PROCEDURE WHEN 'BETWEEN' AND 'WITHIN' SUMS OF SQUARES ARE CORRELATED

SIZE=0.05

r_1	$n-r$	p	q	R	Power
2	10	1.000	5.000	-0.3	0.0758
2	10	1.000	5.000	-0.2	0.0672
2	10	1.000	5.000	-0.1	0.0586
2	10	1.000	5.000	0.0	0.0500
2	10	2.250	6.050	-0.3	0.0852
2	10	2.250	6.050	-0.2	0.0760
2	10	2.250	6.050	-0.1	0.0668
2	10	2.250	6.050	0.0	0.0577
2	10	4.000	5.000	-0.3	0.1771
2	10	4.000	5.000	-0.2	0.1655
2	10	4.000	5.000	-0.1	0.1539
2	10	4.000	5.000	0.0	0.1424
2	10	9.000	5.000	-0.3	0.3384
2	10	9.000	5.000	-0.2	0.3298
2	10	9.000	5.000	-0.1	0.3212
2	10	9.000	5.000	0.0	0.3126
2	10	16.000	5.000	-0.3	0.5190
2	10	16.000	5.000	-0.2	0.5182
2	10	16.000	5.000	-0.1	0.5174
2	10	16.000	5.000	0.0	0.5165
4	10	2.000	5.000	-0.3	0.0756
4	10	2.000	5.000	-0.2	0.0670
4	10	2.000	5.000	-0.1	0.0585
4	10	2.000	5.000	0.0	0.0500
4	10	2.000	6.050	-0.3	0.0515
4	10	2.000	6.050	-0.2	0.0447
4	10	2.000	6.050	-0.1	0.0379
4	10	2.000	6.050	0.0	0.0311
4	10	4.500	5.000	-0.3	0.1350
4	10	4.500	5.000	-0.2	0.1244
4	10	4.500	5.000	-0.1	0.1138
4	10	4.500	5.000	0.0	0.1032
4	10	8.000	5.000	-0.3	0.2212
4	10	8.000	5.000	-0.2	0.2104
4	10	8.000	5.000	-0.1	0.1996
4	10	8.000	5.000	0.0	0.1887
4	10	12.500	5.000	-0.3	0.3260
4	10	12.500	5.000	-0.2	0.3175
4	10	12.500	5.000	-0.1	0.3090
4	10	12.500	5.000	0.0	0.3005

TABLE 3.4 APPROXIMATE POWER OF THE ANOVA PROCEDURE WHEN 'BETWEEN' AND 'WITHIN' SUMS OF SQUARES ARE CORRELATED

SIZE=0.05

r_1	$(n-r)$	p	q	R	Power
6	10	3.000	5.000	-0.3	0.0744
6	10	3.000	5.000	-0.2	0.0663
6	10	3.000	5.000	-0.1	0.0581
6	10	3.000	5.000	0.0	0.0500
6	10	3.000	6.050	-0.3	0.0482
6	10	3.000	6.050	-0.2	0.0419
6	10	3.000	6.050	-0.1	0.0356
6	10	3.000	6.050	0.0	0.0294
6	10	8.333	5.000	-0.3	0.1773
6	10	8.333	5.000	-0.2	0.1668
6	10	8.333	5.000	-0.1	0.1563
6	10	8.333	5.000	0.0	0.1459
2	20	1.000	10.000	-0.3	0.0727
2	20	1.000	10.000	-0.2	0.0651
2	20	1.000	10.000	-0.1	0.0576
2	20	1.000	10.000	0.0	0.0500
2	20	1.000	11.025	-0.3	0.0632
2	20	1.000	11.025	-0.2	0.0562
2	20	1.000	11.025	-0.1	0.0492
2	20	1.000	11.025	0.0	0.0422
2	20	4.000	10.000	-0.3	0.1828
2	20	4.000	10.000	-0.2	0.1714
2	20	4.000	10.000	-0.1	0.1601
2	20	4.000	10.000	0.0	0.1487
2	20	9.000	10.000	-0.3	0.3873
2	20	9.000	10.000	-0.2	0.3811
2	20	9.000	10.000	-0.1	0.3750
2	20	9.000	10.000	0.0	0.3688
4	20	2.000	10.000	-0.3	0.0750
4	20	2.000	10.000	-0.2	0.0667
4	20	2.000	10.000	-0.1	0.0583
4	20	2.000	10.000	0.0	0.0500
4	20	2.000	11.025	-0.3	0.0627
4	20	2.000	11.025	-0.2	0.0552
4	20	2.000	11.025	-0.1	0.0477
4	20	2.000	11.025	0.0	0.0402
4	20	8.000	10.000	-0.3	0.2546
4	20	8.000	10.000	-0.2	0.2434
4	20	8.000	10.000	-0.1	0.2322
4	20	8.000	10.000	0.0	0.2210

3.4 The distribution of the ANOVA statistic when one or more of the standard assumptions are violated - an approximation

We have seen that the statistic T defined in (3.1.7) for testing a set of linear hypotheses is of the form

$$T = \frac{Q_2}{Q_1 + Q_2} \qquad (3.4.1)$$

where Q_1 and Q_2 are two quadratic forms of the basic variables $\eta = (\eta_1, \eta_2, \ldots, \eta_n)'$. If the moments of η exist and are known, moments of Q_1 and Q_2 can be computed. If all moments of η upto order $2t$ are available, we can obtain

$$\mu_{rs} = E(Q_1^r Q_2^s), \quad \text{for} \quad r + s \leq t. \qquad (3.4.2)$$

Since Q_1 and Q_2 are both positive, one may use the t-th order approximation to the joint density function of $Y_1 = c_1 Q_1$ and $Y_2 = c_2 Q_2$ in terms of the standard density functions $g(y_1; m)$ and $g(y_2; n)$ as given by (2.4.2). Using the notation adopted in Chapter 2,

$$L(j, x; m) = \sum_{r=0}^{j} \alpha(j, m, r)\, x^r \qquad (3.4.3)$$

where $\alpha(j, m, r) = c(j, r; m)\, (-1)^r / r!$.

Now, rearranging the terms of (2.4.2) we get, as an approximation to the density function of Y_1 and Y_2,

$$f_t = \sum_{r+s \leq t} a(r,s) \left[\sum_{i=0}^{r} \alpha(r, m, i)\, g(y_1 ; m)\, y_1^i \right.$$

$$\left. \times \sum_{j=0}^{s} \alpha(s, n, j)\, g(y_2 ; n) \right] y_2^j$$

$$= \sum_{r+s \leq t} a(r,s) \left[\sum_{i=0}^{r} \alpha(r, m, i) \frac{\Gamma(m+i)}{\Gamma(m)} g(y ; m + i) \right.$$

$$\left. \times \sum_{j=0}^{s} \alpha(s, n, j) \frac{\Gamma(n+j)}{\Gamma(n)} g(y_2 ; n + j) \right]$$

$$= \sum_{r+s \leq t} \sum_{i=0}^{r} \sum_{j=0}^{s} \beta(r, s, m, n, i, j)$$

$$\times\, g(y_1, m + i)\, g(y_2 ; n + j) \qquad (3.4.4)$$

where

$$\beta\,(r, s, m, n, i, j)$$

$$= \frac{a(r,s)\ \alpha(r, m, i)\ \alpha(s, n, j)\ \Gamma(m+i)\ \Gamma(n+j)}{\Gamma(m)\ \Gamma(n)} \qquad (3.4.5)$$

If we choose

$$c_1 = \frac{\mu_{10}}{\mu_{20} - \mu_{10}^2} ,$$

$$c_2 = \frac{\mu_{01}}{\mu_{02} - \mu_{01}^2}$$

$$m = \frac{\mu_{10}^2}{\mu_{20} - \mu_{10}^2}$$

and $$n = \frac{\mu_{01}^2}{\mu_{02} - \mu_{01}^2}$$

then obviously,

$$\beta(0, 0, m, n, 0, 0) = 1$$

and $$\beta(1, 0, m, n, i, j) = \beta(0, 1, m, n, i, j)$$

$$= \beta(2, 0, m, n, i, j) = \beta(0, 2, m, n, i, j)$$

$$= 0 \tag{3.4.6}$$

With these choices of the constants c_1, c_2, m and n

$$f_t = [1 + (y_1 - m)(y_2 - n)]\; g(y_1 ; m)\, g(y_2 ; n)$$

$$+ \sum_{3 \le r+s \le t} \sum_{i=0}^{r} \sum_{j=0}^{s} \beta(r, s, m, n, i, j) g(y_1 ; m+i)$$

$$\times\; g(y_2 ; n + j) \tag{3.4.7}$$

The corresponding approximation to the distribution function Prob $(T \le x)$ of T is given by

$$F(x) = \int_{\frac{Q_2}{Q_1 + Q_2} \le x} f_t \, dQ_1 \, dQ_2$$

$$= B(z ; n, m) + \rho\left(\frac{n}{m}\right)^{\frac{1}{2}} \frac{\Gamma(n+m)}{\Gamma(m)\;\Gamma(n)}\; z^n (1-z)^m \left(-1 + \frac{m+n}{n} z\right)$$

$$+ \sum_{p=3}^{t} \sum_{i=0}^{r} \sum_{j=0}^{p-r} \beta(r, p-r, m, n, i, j) B(z, n+j, m+i) \tag{3.4.8}$$

where $$z = \left[1 + \frac{c_1(1-x)}{c_2 x} \right]^{-1}$$

$$= \left[1 + \frac{\mu_{10}\,(\mu_{02} - \mu_{01}^2)}{\mu_{01}\,(\mu_{20} - \mu_{10}^2)} \cdot \frac{1-x}{x} \right]^{-1} \qquad (3.4.9)$$

The approximation (3.4.8) to the distribution function of T may be used in studying the effects of different types of deviations from the standard assumption on the ANOVA procedure. The essential difference between Tiku's (1963, 1964) approach and ours is that unlike Tiku we expand not the joint density function of Q_1 and Q_2 in terms of Laguerre polynomials as such, but that of the scaled variables $Y_1 = c_1 Q_1$ and $Y_2 = c_2 Q_2$ and then choosing the parameters m, n, c_1 and c_2 in a suitable way, we make four of the terms in the approximation vanish. The use of scaled variables has another advantage. It has been noticed by Tiku (1965) himself in obtaining approximations for non-central chi square and F distributions that the approximations may be quite far from the true result if the basic variables (in his case no-central chi squares) are not suitably scaled and only a few terms of the orthogonal series expansion is used. For example, if one tries to approximate the distribution function of a $\frac{1}{2}\chi^2(n, \lambda)$ variable by a fourth order approximation in terms of the standard density function $g(x; \frac{1}{2}n)$ the result is often disastrous specially when λ is not small, whereas even the first term of the corresponding approximation using $g(x; m)$ -- where $m = \frac{1}{2}\frac{(n+\lambda)^2}{(n+2\lambda)}$ -- usually gives a good result. This has also been noticed by Patnaik (1949).

CHAPTER 4

STUDIES ON SOME TESTS OF INDEPENDENCE

4.0 Introduction and Summary

It is known that tests based on sample correlation coefficient for testing independence in normal samples is uniformly most powerful unbiased for one sided or two sided alternatives. When the parent population is non-normal and we have a large number of observations, independence between two variables is often tested by forming a two-way contingency table and using the contingency chi-square statistic. Properties of chi-square procedures have been studied by Mann and Wald (1942), Cochran (1952), Eisenhart (1938), Mitra (1955, 1958), Williams (1950), Hamdan (1963, 1968) and others. In this chapter we compare the performance of the contingency chi-square procedure with that of the test based on sample correlation coefficient in normal samples, by computing the Pitman (1948) asymptotic powers of the two procedures. Mitra (1955) has shown that in the non-null situation the contingency chi-square statistic follows a non-central chi-square distribution asymptotically when alternatives are of certain forms. Using this result we obtain an explicit expression for the non-centrality parameter of the asymptotic chi-square distribution when basic observations are from a bivariate normal population. The asymptotic power of the contingency chi-square procedure is then compared numerically with the asymptotic power function, again in Pitman's sense, of the uniformly most powerful unbiased (UMPU) procedure based on the sample coefficient of correlation. It is found that increasing the number of classes in each of the two ways of classification does not necessarily increase the power, a fact that has been noticed by many for the case of tests of goodness of fit in the univariate case. Though increasing the number of classes generally increases the non-centrality parameter, the gain in power due to this increase is often offset by an associated increase in the number of degrees of freedom. If a (2 × 2) contingency table is used for testing independence in a bivariate normal population how does one choose the class-limits? It is proved that in this case, power, in Pitman's sense, is maximised when the two means are chosen as the division points. For this case, the ratio of the non-centrality parameter of the contingency chi-square test to that of the correlation coefficient test turns out to be only 0.4.

A statistical procedure optimum in some classical sense may not be economic when cost of measuring the observations and of computation are taken into account. Methods based on counting rather than measurements have sometimes been conveniently used. For example, to study the correlation between the co-ordinates of error of a gun, it may be convenient to count, for a round fired, the number of shots falling within a specified region of a two-dimensional plane. Roy (1956) has given a general procedure for obtaining the class of binomial tests, based on counting, which contains the most powerful test. In this chapter we show how to obtain the most powerful binomial test of independence based on counting against a specified alternative. It is shown that these tests are much better than the binomial test based on median dichotomy and their performance is quite satisfactory even when we compare with the UMPU test based on the coefficient of correlation.

In order to study the effect of non-normality on the sample correlation test we require the distribution of the statistic in non-normal samples. Gayen (1951), by starting with a Gram-Charlier expansion of the joint probability density function of the population, has obtained an approximation to the distribution of r, the sample correlation coefficient, when ρ - the population coefficient of correlation is not necessarily zero. An alternative approach is presented in this thesis. The probability density function of $X = \frac{1}{2}(r + 1)$ is approximated here in terms of a beta density function and Jacobi polynomials. The suitability of this approximation is then studied numerically in the case of bivariate normal populations with non-zero coefficients of correlation. In two other situations the performance of this approximation is studied by model sampling experiments. It is observed in all these three cases the performance of our approximation is not bad.

4.1 Pitman powers of the correlation test and the contingency chi-square test for independence

4.1.1 Contingency chi-square

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be n observations drawn at random from a bivariate population with density function $f(x, y)$. Let $D_1 = (x_o = -\infty, x_1, x_2, \ldots, x_l = \infty)$ and $D_2 = (y_o = -\infty, y_1, y_2, \ldots, y_m = \infty)$ be divisions of the ranges of x and y respectively. If the observations are grouped in a $(l \times m)$ table where f_{ij} = number of observations satisfying

$$x_{i-1} < x \leq x_i$$

$$y_{j-1} < y \leq y_j$$

we get what is called a contingency table.

Let

$$p_{io} = \text{Prob}\ (x_{i-1} < x \leq x_i) = \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} f(x, y)\, dx\, dy, \quad i=1, 2, \ldots, l$$

$$p_{oj} = \text{Prob}\ (y_{j-1} < y \leq y_j) = \int_{-\infty}^{\infty} \int_{y_{j-1}}^{y_j} f(x, y)\, dy\, dx, \quad j=1, 2, \ldots, m$$

$$p_{ij} = \text{Prob}\ (x_{i-1} < x \leq x_i,\ y_{j-1} < y \leq y_j)$$

$$= \int_{x_{i-1}}^{x_i} \int_{y_{j-1}}^{y_j} f(x, y)\, dy\, dx, \quad i = 1, 2, \ldots, l;\ j = 1, 2, \ldots, m$$

Here independence of x and y implies the hypothesis

$$H_o\ (p_{ij} = p_{io} \cdot p_{oj},\quad i = 1, 2, \ldots, l,\quad j = 1, 2, \ldots, m).$$

A procedure for testing the above hypothesis in large samples is :

$$\text{Reject } H_o \text{ if } \chi^2 = n \sum_{i,j} \frac{\left(f_{ij} - \frac{f_{io} f_{oj}}{n}\right)^2}{f_{io} \times f_{oj}} \geq \chi_o^2 ,$$

accept H_o otherwise (4.1.1)

where $f_{io} = \sum_{j=1}^{m} f_{ij}$, $f_{oj} = \sum_{i=o}^{l} f_{ij}$, $n = \sum_{i,j} f_{ij}$

and χ_o^2 satisfies the size restriction

$$\text{Prob}(\chi^2 \geq \chi_o^2 \mid H_o) = \alpha$$

For large n, in the null case, this statistic χ^2 asymptotically follows the chi-square distribution with $(l - 1)(m - 1)$ degrees of freedom and therefore χ_o^2 can be taken as the upper 100α percent point of this distribution.

We want to study the relative efficiency of this procedure with respect to the UMPU procedure based on r - the coefficient of correlation, when the population is known to be bivariate normal. In order to compare their performances, a simple procedure seems to be the comparison of their Pitman asymptotic powers. [see Pitman (1948)].

4.1.2 The asymptotic power function of the contingency chi-square test of independence

Suppose H_o $(p_{ij} = p^o_{ij} = p_{io} \cdot p_{oj}$; for all $i, j)$ is the hypothesis we are testing. Denote by H_{on} the sequence of alternatives

$$H_{on} \; (p_{ij} = p^o_{ij} + \frac{c_{ij}}{\sqrt{n}}), \text{ where } \sum_i c_{ij} = \sum_j c_{ij} = 0 \quad \text{for all } i, j.$$

The limiting power function of the χ^2 test of independence may now be defined as [Pitman (1948)]

$$G = \underset{n \to \infty}{Lt} \; \text{Prob}(\chi^2 \geq \chi^2_o | H_{on})$$ which of course depends on the c_{ij}'s.

This limiting power is called the 'Pitman power' of the chi-square procedure. Mitra (1955) has proved that

$$G = 1 - F(\chi^2_o \, , (l - 1) \, (m - 1) \, , D) \qquad (4.1.2)$$

where F is as defined in section 1.3 and

$$D = \sum_{i,j} \frac{c^2_{ij}}{p^o_{ij}} = \sum_{i,j} \frac{c^2_{ij}}{p_{io} \cdot p_{oj}}$$

4.1.3 Pitman power of the chi-square test when the population is bivariate normal

In the case of a bivariate population where the density function is

$$f(x, y, \rho) = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right],$$

the independence of x and y is equivalent to H_o ($\rho = 0$). The Pitman power of the chi-square procedure in this case may be defined as the limit of the power as $n \to \infty$ when the sequence of alternatives is of the form

$$H_{on} \left(\rho = \frac{\theta}{\sqrt{n}} \right)$$

We prove the following theorem on the Pitman power of the contingency chi-square test of independence in normal samples.

Theorem 4.1

If we neglect terms of order n^{-1} in $(p_{ij} - p_{io}\, p_{oj})$, the Pitman power of chi-square test of independence when the population is bivariate normal and the alternatives are of the form H_{on} ($\rho = \frac{\theta}{\sqrt{n}}$), is given by

$$1 - F(\chi_o^2, (\ell - 1) \times (m - 1), D^*) \qquad (4.1.3)$$

where

$$D^* = \theta^2 \times \sum_i \frac{[\phi(x_i) - \phi(x_{i-1})]^2}{\Phi(x_i) - \Phi(x_{i-1})} \times \sum_j \frac{[\phi(y_j) - \phi(y_{j-1})]^2}{\Phi(y_j) - \Phi(y_{j-1})}$$

and ϕ and Φ are, as defined in section 1.3, the density function and the distribution function respectively of the standard normal variable.

<u>Proof</u>

Expanding $f(x, y, \frac{\theta}{\sqrt{n}})$ about $\theta = 0$,

$$f(x, y, \frac{\theta}{\sqrt{n}}) = f(x, y, 0) + \frac{\theta}{\sqrt{n}} f'(x, y, 0) + O(n^{-1})$$

$$= \phi(x)\ \phi(y) + \frac{\theta}{\sqrt{n}} [x\phi(x) \times\ y\phi(y)] + O(n^{-1}).$$

Therefore

$$p_{ij} = \int_{x_{i-1}}^{x_i} \int_{y_{j-1}}^{y_j} \phi(x)\phi(y)\ dy\ dx + \frac{\theta}{\sqrt{n}} \int_{x_{i-1}}^{x_i} \int_{y_{j-1}}^{y_j} xy\phi(y)\phi(x) dy\ dx + O(n^{-1})$$

$$= p_{io} \cdot p_{oj} + \frac{\theta}{\sqrt{n}} [\phi(x_{i-1}) - \phi(x_i)]\ [\phi(y_{j-1}) - \phi(y_j)] + O(n^{-1})$$

Neglecting terms of order n^{-1}

$$p_{ij} = p_{io} \cdot p_{oj} + \frac{\theta}{\sqrt{n}}\ [\phi(x_{i-1}) - \phi(x_i)]\ [\phi(y_{j-1}) - \phi(y_j)].$$

Also

$$\theta \sum_i [\phi(x_{i-1}) - \phi(x_i)]\ [\phi(y_{j-1}) - \phi(y_j)]$$

$$= \theta \sum_j [\phi(x_{i-1}) - \phi(x_i)] \times [\phi(y_{j-1}) - \phi(y_j)]$$

$$= 0.$$

Hence using (4.1.2) the Pitman power of the contingency chi-square test, defined in (4.1.1), of independence is

$$1 - F(\chi_o^2, (l - 1)(m - 1), D^*)$$

where

$$D^* = \theta^2 \times \sum_i \frac{[\phi(x_i) - \phi(x_{i-1})]^2}{p_{io}}$$

$$\times \sum_j \frac{[\phi(y_j) - \phi(y_{j-1})]^2}{p_{oj}}$$

$$= \theta^2 \sum_{i=1}^{l} \frac{[\phi(x_i) - \phi(x_{i-1})]^2}{\Phi(x_i) - \Phi(x_{i-1})} \times \sum_{j=1}^{m} \frac{[\phi(y_j) - \phi(y_{j-1})]^2}{\Phi(y_j) - \Phi(y_{j-1})} .$$

4.1.4 Pitman power of the UMPU procedure based on the sample correlation coefficient

The statistic $\sqrt{n}\, r$, where r is the coefficient of correlation computed from n independent observations drawn from a bivariate normal population $N(0, 0, 1, 1, \rho)$ is asymptotically distributed as $N(0, 1)$ when $H_o(\rho = 0)$ is true. Hence in large samples the following procedure for testing independence may be used for two-sided alternatives.

Reject H_o if $(\sqrt{n}\, r)^2 > \chi_o^2$,

accept H_o otherwise (4.1.5)

where χ_o^2 is the upper 100α percent point of the chi-square distribution with 1 degree of freedom. Our result about the Pitman power of the procedure (4.1.5) when alternatives are of the form H_{on} ($\rho = \frac{\theta}{\sqrt{n}}$) is stated in theorem 4.2.

Lemma 4.1

If $x_1, x_2, \ldots, x_n$ are n independent but identically distributed variables with $E(x) = m_n$ and $V(x) = \sigma_n^2$, and if the limits

$$\underset{n\to\infty}{\text{Lt}} \sqrt{n}\, m_n = m$$

and

$$\underset{n\to\infty}{\text{Lt}} \sigma_n = \sigma$$

exist, then $\sqrt{n}\,\bar{x}$, where $\bar{x} = \frac{1}{n} \Sigma\, x_i$, asymptotically follows a normal distribution with mean m and standard deviation σ.

Proof :

$$\sqrt{n}\,\bar{x} = \frac{\sqrt{n}(\bar{x} - m_n)}{\sigma_n} \cdot \sigma_n + \sqrt{n}\, m_n$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i \times \sigma_n + \sqrt{n}\, m_n,$$

where $y_i = \dfrac{x_i - m_n}{\sigma_n}, \; i = 1, 2, \ldots, n$

are independently identically distributed with mean zero and standard deviation unity. Hence using the Lindeberg-Levy central limit theorem [Cramer (1962)], $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i$ asymptotically follows a $N(0, 1)$ distribution. Moreover, $\underset{n\to\infty}{\text{Lt}} \sigma_n = \sigma$ and $\underset{n\to\infty}{\text{Lt}} \sqrt{n}\, m_n = m$ and therefore $\sqrt{n}\,\bar{x}$ asymptotically follows a normal distribution with mean m and standard deviation σ.

Using lemma 4.1 it immediately follows that $\sqrt{n}\, r$ when $\rho = \frac{\theta}{\sqrt{n}}$ asymptotically follows a normal distribution with mean θ and standard deviation unity. Hence we have

Theorem 4.2

Pitman power of a size $-\alpha$ procedure based on sample coefficient of correlation for testing independence when the population is bivariate normal and alternatives are of the form $H_{on}(\rho = \frac{\theta}{\sqrt{n}})$ is given by

$$\underset{n\to\infty}{\text{Lt}} \text{ Prob } (nr^2 > \chi_o^2 \mid H_{on}) = 1 - F(\chi_o^2, 1, \theta^2) \qquad (4.1.6)$$

where χ_o^2 is the upper 100α percent point of the chi-square distribution with one degree of freedom.

4,1.5 Numerical comparison of the Pitman powers of contingency chi-square and sample correlation coefficient tests

The following table gives the Pitman power of the chi-square test as compared with the Pitman power of the test based on the coefficient of correlation. The computations have been done, in the case of chi-square tests, by forming classes so that

$$\text{Prob } (x \leq x_i) = i/l, \quad i = 1, 2, \ldots, l$$

and

$$\text{Prob } (y \leq y_j) = j/m, \quad j = 1, 2, \ldots, m.$$

The table shows that the increase in the number of classes does not increase the asymptotic power of the chi-square tests. This happens because though in many cases an increase in the number of classes increases the non-centrality parameter D^* in (4.1.3), the number of degrees of freedom is also increased in the process.

TABLE 4.1 COMPARISON OF THE PITMAN POWERS OF CHI-SQUARE AND CORRELATION COEFFICIENT PROCEDURES

Size = .05

	Pitman Power			
θ.	2 × 2 table	4 × 4 table	6 × 6 table	Test based on r
0.1	.05047	.05023	.05015	.05114
0.2	.05187	.05090	.05057	.05459
0.3	.05419	.05204	.05130	.06037
0.4	.05746	.05367	.05235	.06852
0.5	.06169	.05575	.05364	.07909
0.6	.06688	.05838	.05530	.09215
0.7	.07304	.06155	.05727	.10774
0.8	.08021	.06527	.05961	.12591
0.9	.08838	.06963	.06234	.14669
1.0	.09757	.07466	.06545	.17008
2.0	.24674	.17190	.12611	.51601
3.0	.48007	.37944	.26798	.85084
4.0	.72123	.65726	.50419	.97933
5.0	.88936	.87650	.76037	.99882

In table 4.1 we notice that for all θ the power for the (2×2) table is maximum. In general, however, the Pitman power will depend on the manner in which classes are formed and also on the number of classes. In the next section we show that the effect of dividing any existing class in any of the two ways of classification is to increase the non-centrality parameter D^*. But since the number of degrees of freedom is also increased in the process this may not result in an increase in the power.

4.1.6 Effect of dividing an existing class interval on the non-centrality parameter D^*

Lemma 4.2

If $\omega > 0$,

$x = x_1 + x_2$,

and $0 < \lambda < 1$,

then, $\dfrac{x_1^2}{\lambda\omega} + \dfrac{x_2^2}{(1-\lambda)\omega} \geq \dfrac{x^2}{\omega}$, equality holding when

$$\frac{x_1}{\lambda} = \frac{x_2}{1-\lambda}$$

Proof :

We have

$$x_1^2(1-\lambda)^2 + x_2^2\ \lambda^2 \geq 2x_1x_2\ \lambda(1-\lambda)$$

Dividing both sides by $\lambda(1-\lambda)$

$$x_1^2\left(\frac{1}{\lambda}-1\right)+x_2^2\left(\frac{1}{1-\lambda}-1\right) \geq 2x_1x_2$$

or
$$\frac{x_1^2}{\lambda}+\frac{x_2^2}{1-\lambda} \geq x_1^2+x_2^2+2x_1x_2 = x^2 .$$

or
$$\frac{x_1^2}{\lambda w}+\frac{x_2^2}{(1-\lambda)w} \geq \frac{x^2}{w} .$$

Theorem 4.3

If an already existing class interval, in any of the two ways of classification, is divided into two non-trivial class intervals then the non-centrality parameter D^* in (4.1.3) increases.

Proof :

Suppose we divide the existing class interval $x_{i-1} < x \leq x_i$ into two class intervals $x_{i-1} < x \leq c$ and $c < x \leq x_i$ where $x_{i-1} < c < x_i$. From (4.1.3) the contribution of the class $x_{i-1} < x < x_i$ to the non-centrality parameter D^* is through

$$\frac{[\phi(x_i)-\phi(x_{i-1})]^2}{\Phi(x_i)-\Phi(x_{i-1})} .$$

After the division the contribution would be through

$$\frac{[\phi(c)-\phi(x_{i-1})]^2}{\Phi(c)-\Phi(x_{i-1})}+\frac{[\phi(x_i)-\phi(c)]^2}{\Phi(x_i)-\Phi(c)} .$$

From the lemma 4.2

$$\frac{[\phi(c) - \phi(x_{i-1})]^2}{\Phi(c) - \Phi(x_{i-1})} + \frac{[\phi(x_i) - \phi(c)]^2}{\Phi(x_i) - \Phi(c)} .$$

$$> \frac{[\phi(x_i) - \phi(x_{i-1})]^2}{\Phi(x_i) - \Phi(x_{i-1})} .$$

Here the condition for equality is

$$\frac{\phi(x_{i-1}) - \phi(c)}{\Phi(c) - \Phi(x_{i-1})} = \frac{\phi(c) - \phi(x_i)}{\Phi(x_i) - \Phi(c)}$$

which is clearly impossiple. Hence the theorem.

4.1.7 Optimum points of division in a (2×2) table

When the population is bivariate normal the Pitman power of the contingency chi-square procedure is an increasing function of D^* defined in (4.1.3). The division points at which the power will have a stationary value are given by the solutions of the equations

$$\frac{\delta D^*}{\delta x_i} = 0 \qquad i = 1, 2, \ldots, l - 1$$

$$\frac{\delta D^*}{\delta y_j} = 0 \qquad j = 1, 2, \ldots, m - 1$$

These equations are rather difficult to solve except by trial and error when l and m are large. We shall consider the case when $l = m = 2$.

Let $D_1 = (-\infty, X, \infty)$ and $D_2 = (-\infty, Y, \infty)$ be the divisions of the ranges of x and y respectively. Here the non-centrality parameter in the power function is

$$D^* = \frac{\theta^2 \phi^2(X) \phi^2(Y)}{\Phi(X)(1 - \Phi(X)) \Phi(Y)(1 - \Phi(Y))} \tag{4.1.7}$$

consider the term

$$W = \frac{\phi^2(X)}{\Phi(X) \; (1 - \Phi(X))}$$

Differentiating W twice with respect to X, we get

$$W' = (u_1 - u_2)/v$$

where $u_1 = 2\Phi(1 - \Phi) \phi \phi'$,

$u_2 = \phi^3 (1 - 2\Phi)$,

$v = \Phi^2 (1 - \Phi)^2$,

and

$$W'' = \frac{v(u_1' - u_2') - (u_1 - u_2)v'}{v^2}$$

where $v' = (2\Phi + 4\Phi^3 - 6\Phi^2)\phi$,

$u_1' = 2\Phi(1 - \Phi)\phi \phi'' + 2\phi^2 \phi'(1 - 2\Phi) + 2\Phi (1 - \Phi) (\phi')^2$, and $u_2'' = 3\,\phi^2\,\phi'\,(1 - 2\Phi) - 2\phi^4.$

At the point $X = 0$, $\Phi(X) = \frac{1}{2}$ and $W' = 0$. Thus W has a stationary value at $X = 0$. We further note that at this point $u_1 - u_2 = 0$, v is positive and

$$(u_1'' - u_2'') = -\frac{1}{4\pi} + \frac{1}{2\pi^2}$$

$$= \frac{-1}{4\pi^2}(2 - \pi) < 0, \qquad \text{and}$$

hence $W'' < 0$.

This shows that W has a maximum at $X = 0$. Similarly

$$\frac{\phi^2(Y)}{\Phi(Y)\,(1 - \Phi(Y))}$$

also has a maximum at $Y = 0$. Hence we get the following theorem

Theorem 4.4

Pitman power of a (2×2) contingency chi-square test of independence in the normal case has a maximum when the division points are $X = 0$, $Y = 0$.

4.1.8 Asymptotic efficiency of the (2 × 2) contingency chi-square procedure

We have seen that both the (2×2) contingency chi-square statistic and the statistic $(\sqrt{n}\,r)^2$ are asymptotically distributed as non-central chi-square each one degree of freedom. A measure of relative efficiency of the (2×2) chi-square statistic may be defined [Hannan (1956)], with respect to the UMPU test based on r, as

$$E = \frac{\text{non-centrality parameter of chi-square distribution}}{\text{non-centrality parameter of the distribution of } nr^2}$$

$$= \frac{\theta^2\ \phi^2(X)\ \ \phi^2(Y)}{\theta^2\ \Phi(X)(1 - \Phi(X))\ \Phi(Y)(1 - \Phi(Y))}$$

$$= \frac{\phi^2(X)\ \ \phi^2(Y)}{\Phi(X)(1 - \Phi(X))\ \Phi(Y)(1 - \Phi(Y))}$$

for the division points $x = X,\ \ y = Y.$

From theorem 4.4 the relative efficiency is maximum when $X = 0,$ $Y = 0,$ and this maximum value is

$$E_{max} = \left[\frac{\phi^2(0)}{\Phi(0)\ (1 - \Phi(0))} \right]^2 = \frac{4}{\pi^2} = 0.4 \quad \text{(approx.)}$$

4.2 Binomial procedures for testing independence

Methods based on counting rather than measurements have sometimes been used conveniently in industrial problems when counting is cheaper than measurement. In this chapter we shall discuss such procedures for testing independence when the population is bivariate normal.

4.2.1 Binomial test procedures based on counting

Let x be a p-dimensional random variable with a continuous probability density function f. The problem is to test $H_o : f = f_o$ against the simple alternative $H_1 : f = f_1$ on the basis of a random sample $x_1, x_2, \ldots, x_n$ of size n. Suppose ω is a sub-space of the p-dimensional Euclidean space such that

$$\pi_1 > \pi_0 > 0$$

where π_i = Prob $(x \in \omega | H_i)$, $i = 0, 1$ (4.2.1)

Let us define a variable

$$y_i = 1 \quad \text{if} \quad x_i \in \omega$$
$$= 0 \quad \text{otherwise} \tag{4.2.2}$$

Further, let

$$d = \sum_{i=1}^{n} y_i \tag{4.2.3}$$

Then $\text{Prob}(d = m | H_i) = \binom{n}{m} \pi_i^m (1 - \pi_i)^{n-m}, \; i = 0, 1$ (4.2.4)

The statistic d can be used to test the hypothesis H_0 against the alternative H_1 in the following manner :

Let c be the smallest integer satisfying

$$\sum_{i=c+1}^{n} \binom{n}{i} \pi_0^i (1 - \pi_0)^{n-i} \leq \alpha \tag{4.2.5}$$

Then the test procedure is :

Reject H_0 if $d > c$,

accept H_0 otherwise (4.2.6)

The procedure (4.2.6) for testing H_0 may be called the binomial test procedure for H_0 based on the sub-set ω.

Obviously, the power of this test is given by

$$\beta = \text{Prob}\,(d > c \mid H_1) = \sum_{i=c+1}^{n} \binom{n}{i} \pi_1^i (1 - \pi_1)^{n-i} \qquad (4.2.7)$$

It can be easily checked that β is increasing in π_1 and hence the test procedure (4.2.6) is unbiased, and uniformly so for alternatives H for which

$$\text{Prob}\,(X \in \omega \mid H) > \pi_o$$

The power of the procedure (4.2.6), however, depends on the choice of ω. The question 'what is the best choice of ω ?' naturally arises. Roy (1956) in this connection has proved the following theorem.

Theorem 4.5 (Roy)

Let ω be a given sub-set of the p-dimension Euclidean space satisfying (4.2.1). Then under certain simple condition, it is possible to find a sub-set ω_o belonging to the class :

$$\text{inside } \omega_o : f_o(x) < k f_1(x) \qquad (4.2.8)$$

such that for the same sample size the binomial test based on ω_o is at least as powerful as that based on ω.

Because of this theorem the most powerful binomial test for H_o against a fixed alternative H_1 is based on a subset of the type (4.2.8) and therefore our search for the most powerful binomial procedure reduces to that of finding an optimum value for k. No general method, however, is available for determining this optimum value of k and the best region ω_o may be found only by numerical methods.

4.2.2 Most powerful binomial procedures for testing independence

Let (x_{1i}, x_{2i}) $i = 1, 2, \ldots, n$ be n pairs of random observations from the bivariate normal population $N(0, 0, 1, 1, \rho)$. Theorem 4.5 says that the region ω_o on which the most powerful binomial test (MPBT) procedure, for testing $H_o : (\rho = 0)$ against the alternative $H_1 : (\rho = \rho_1)$, is based satisfies

$$\frac{1}{2\pi} \exp\left[-\frac{1}{2}(x_1^2 + x_2^2)\right] < \frac{k'}{2\pi\sqrt{(1-\rho_1^2)}} \times \exp\left[-\frac{1}{2(1-\rho_1^2)}(x_1^2 + x_2^2 - 2\rho_1 x_1 x_2)\right]$$

inside ω_o

or $$V = \rho_1^2 x_1^2 + \rho_1^2 x_2^2 - 2\rho_1 x_1 x_2 < k, \text{ inside } \omega_o \tag{4.2.9}$$

For a given value of k to find out the probabilities π_o and π_1 we proceed in the following manner.

We have $\pi_o(k) = \text{Prob}(V < k | \rho = 0)$

Since the characteristic roots of the associated matrix of the quadratic form V are given by

$$\begin{vmatrix} \rho_1^2 - \lambda & -\rho_1 \\ -\rho_1 & \rho_1^2 - \lambda \end{vmatrix} = 0 ,$$

therefore, the distribution of V is same as that of

$$Z = \lambda_1 y_1^2 - \lambda_2 y_2^2 \tag{4.2.10}$$

where $\lambda_1 = \rho_1^2 + \rho_1$

$\lambda_2 = -(\rho_1^2 - \rho_1)$ if ρ_1 is positive

or $\lambda_1 = \rho_1^2 - \rho_1$

$\lambda_2 = -(\rho_1^2 + \rho_1)$ if ρ_1 is negative

Chosen so That $\lambda_1, \lambda_2 > 0$

and y_1 and y_2 are independently distributed as $N(0, 1)$. X cf p 112
Following Bhattacharya (1943), it can be easily shown that the density function of Z at any point z is given by

$$f(z) = Z_o \exp[- z(\lambda_2 - \lambda_1)/ 4\lambda_1 \lambda_2] K_o| z(\lambda_1 + \lambda_2)/ 4 \lambda_1 \lambda_2| \quad (4.2.11)$$

where K_m is the Bessel function of order m of the second kind as defined by Watson (1922).

For determining $\pi_1(k) = \text{Prob}(V < k | \rho = \rho_1)$ we note that the coefficient matrix of the quadratic form V is

$$A = \begin{bmatrix} \rho_1^2 & -\rho_1 \\ -\rho_1 & \rho_1^2 \end{bmatrix} \quad (4.2.12)$$

and the dispersion matrix of $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ under H_1 is

$$B = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}$$

Since A is symmetric and B is positive definite by using a suitable transformation V reduces to the form

$$V = \lambda_1 y_1^2 - \lambda_2 y_2^2 \tag{4.2.13}$$

where y_1 and y_2 are independently distributed as $N(0, 1)$, λ_1 and λ_2 are both positive and λ_1 and $-\lambda_2$ are the roots of the determinantal equation

$$|A - \lambda B^{-1}| = 0 \tag{4.2.14}$$

Therefore $\lambda_1 = \lambda_2 = |\rho_1|(1- \rho_1^2)$. Hence

$$\pi_1(k) = \text{Prob}\left(y_1^2 - y_2^2 < \frac{k}{|\rho_1|(1 - \rho_1^2)}\right) \tag{4.2.15}$$

The power of the binomial procedure based on ω_0 is given by

$$P = \sum_{i=c+1}^{n} \binom{n}{i} \pi_1^i(k) (1 - \pi_1(k))^{n-i} \tag{4.2.16}$$

where c is the smallest integer satisfying

$$\sum_{i=c+1}^{n} \binom{n}{i} \pi_0^i(k) (1 - \pi_0(k))^{n-i} \leq \alpha,$$

α being the level of significance,

<u>or</u> in large samples, using a normal approximation for the arcsin transformation of the square root of the proportion,

$$P^* = 1 - \Phi(u) \tag{4.2.17}$$

where $u = t_\alpha - \sqrt{(4n)}\,(\arcsin(\sqrt{\pi_1(k)}) - \arcsin(\sqrt{\pi_0(k)}))$,

t_α being the upper 100α percent point of the standardised normal distribution.

The optimum choice of ω_o is provided by that value of k which maximises the power. In general, the value of k will be a function of the sample size n. The problem becomes somewhat simpler when sample is large so that the power is given, at least approximately, by (4.2.17). In this case the optimum value of k is independent of n and is that value of k which maximises the difference

$$\Delta(k) = \arcsin(\sqrt{\pi_1(k)}) - \arcsin(\sqrt{\pi_o(k)}).$$

After evaluating $\Delta(k)$ for different values of k for different alternative values of ρ, we determine, at least approximately the optimum values of k. These are presented, along with the corresponding values of π_o, π_1 and the difference $\Delta(k)$, in table-4.2.

TABLE 4.2 OPTIMUM VALUES OF k FOR THE MPBT FOR INDEPENDENCE IN NORMAL SAMPLES

Alternative value of ρ	k	π_0	π_1	Δ
0.1	-0.05	0.2697	0.2985	.03200
0.2	0.00	0.4358	0.5000	.06426
0.3	0.00	0.4029	0.5000	.09760
0.4	0.25	0.6144	0.7525	.14924
0.5	0.50	0.6637	0.8377	.20392
0.6	0.80	0.6919	0.9009	.26808
0.7	0.95	0.6703	0.9309	.34570
0.8	0.95	0.6164	0.9531	.44961
0.9	0.95	0.5633	0.9873	.60908

4.2.3 Performance of the most powerful binomial test for independence

We now compare the performance of MPBT for independence as obtained in the previous section for a $N(0, 0, 1, 1, \rho)$ population with that of the binomial test procedure based on median dichotomy i.e. on the subset

$$\omega = (-\infty < x_1 \leq 0, -\infty < x_2 \leq 0) \cup (0 \leq x_1 < \infty, 0 \leq x_2 < \infty)$$

The binomial procedure for $H_0(\rho = 0)$ against all $H_1(\rho > 0)$ based on ω will have the power

$$P_m = \sum_{x=c_\alpha+1}^{n} \binom{n}{x} \pi_1^x (1 - \pi_1)^{n-x} \qquad (4.2.18)$$

where $$\pi_1 = 1 - 2 \int_{-\infty}^{0} \int_{0}^{\infty} f(x, y, \rho)\, dx\, dy \qquad (4.2.19)$$

f being the density function of $N(0, 0, 1, 1, \rho)$ variable and c_α satisfies the size restriction

$$\sum_{c_\alpha+1}^{n} \binom{n}{x} \left(\frac{1}{2}\right)^n \leq \alpha \qquad (4.2.20)$$

This test procedure will be referred to as median dichotomy binomial test (MDBT).

In table 4.3 we compare the performance of the MPBT for specified alternative and MDPT for $n = 25$. The levels of significance exactly attained differ for two tests because of the discreteness of the test statistics. We have taken in either case the size which is nearest to 0.05.

TABLE 4.3 COMPARISON OF THE POWER OF THE MPBT FOR INDEPENDENCE WITH THAT OF MDBT

$n = 25$

Alternative ρ	MPBT for the fixed alternative		MDBT	
	size	power	size	power
0.1	0.0638	0.1146	0.0539	0.0987
0.2	0.0736	0.2120	"	0.1670
0.3	0.0368	0.2120	"	0.2634
0.4	0.0524	0.4345	"	0.3887
0.5	0.0505	0.6390	"	0.5376
0.6	0.0455	0.8383	"	0.6961
0.7	0.0499	0.9739	"	0.8415
0.8	0.0421	0.9991	"	0.9465
0.9	0.0347	1.0000	"	0.9940

From this table though it is difficult to judge how good is the MPBT as compared to the MDBT for all ρ, the former is definitely much better than the later for $\rho \geq 0.4$.

Power of the size-α UMPU procedure for the $H_0(\rho = 0)$ against the class of alternative $H_1(\rho > 0)$ based on the sample coefficient of correlation r is given by

$$\text{Prob}[\, r \geq r_\alpha \mid n, \rho \,] \tag{4.2.21}$$

where r_α is defined by

$$\text{Prob}[\, r \geq r_\alpha \mid n, 0 \,] = \alpha \tag{4.2.22}$$

In order to compare the performance of the three tests namely, MPBT, MDBT and the test based on r, we present in table 4.4, for $n = 25$, the exact powers of the r-test and the MDBT for $\alpha = 0.0539$ and the large sample expression for the power of the MPBT for the same α using (4.2.17).

For given α, (4.2.21) is obtained for different ρ from David (1938) and π_1 (4.2.19) is computed using the table of the Department of Commerce, US National Bureau of standards.

TABLE 4.4 COMPARISON OF THE POWER OF THE MPBT FOR INDEPENDENCE WITH THAT OF THE BINOMIAL TEST BASED ON MEDIAN DICHOTOMY AND ALSO THE UMPU TEST BASED ON r.

$n = 25$, size = .0539

ρ	Power: MPBT for fixed alternative (ρ)	Power: Test based on median dichotomy	r-test
0.1	.0988	.0987	.1280
0.2	.1671	.1670	.2584
0.3	.2636	.2634	.4457
0.4	.4456	.3887	.6596
0.5	.6667	.5376	.8442
0.6	.8583	.6961	.9545
0.7	.9676	.8415	.9936
0.8	.9981	.9465	.9998
0.9	1.0000	.9940	1.0000

It may be noted that whereas in the case of MDBT and the test based on r we use the same procedure for all alternatives of the form $H_1(\rho > 0)$, the MPBT procedure is obtained for a fixed alternative hypothesis and the subset ω changes with change in the alternative hypothesis. This, however, is not a serious drawback. When our objective is to test $H_o(\rho = 0)$ against the class of alternatives $H_1(\rho > 0)$, even if we use the MPBT constructed for $H_1(\rho = 0.3)$ we donot lose much as can be seen by looking at its power function.

The following table is constructed using the large sample expression (4.2.17).

TABLE 4.5 POWER FUNCTION OF MPBT FOR INDEPENDENCE AGAINST $H_1(\rho = 0.3)$

ρ	Power	ρ	Power
0.0	.0539	0.5	.5504
0.1	.0976	0.6	.7162
0.2	.1656	0.7	.8625
0.3	.2636	0.8	.9592
0.4	.3937	0.9	.9963

One method of determining the efficiency of the size-α MPBT will be to find out the increase in the sample size required so that it will have the same power as the size-α UMPU test based on r.

Let β be the power of the size-α test of independence based on r when sample size is n. Suppose the MPBT will require a sample of size n_1 to attain the power. Then, using normal approximation for arcsin transformation of the square root ofa proportion,

$$1 - \Phi\,(t_\alpha - \sqrt{(4\,n_1)}(\arcsin\sqrt{\pi_1} - \arcsin\sqrt{\pi_0})) = \beta$$

or

$$n_1 = \left[\frac{1}{2}\,\frac{t_\alpha - t_\beta}{(\arcsin\sqrt{\pi_1} - \arcsin\sqrt{\pi_0})}\right]^2 \qquad (4.2.23)$$

where t_α is the upper 100α percent point of the standardised normal distribution.

If the ratio of the cost per item sampled for the classical test to that for the binomial test is greater than n_1/n the MPBT should prove more economic.

4.2.4 The following table gives values of n_1 so that MPBT would be as powerful as the correlation test with a sample size(n) 25.

TABLE 4.6 SAMPLE SIZE REQUIRED (n_1) FOR MPBT SO THAT IT IS AS POWERFUL AS r-TEST WITH SAMPLE SIZE = 25

Level of significance = 0.0539

Specified alternative ρ	Power	n_1	ρ	Power	n_1
0.1	.1280	55	0.6	.9545	38
0.2	.2584	56	0.7	.9936	36
0.3	.4457	57	0.8	.9998	33
0.4	.6596	46	0.9	1.0000	27
0.5	.8442	42			

4.2.5 It may be noted that the MPBT for independence as derived in section 4.2.2 can be used only if the means and standard deviations of x_1 and x_2 are known. One method of using this procedure when these are unknown will be to estimate them from independent large samples and use these as the true values.

4.3 An approximation to the distribution of the sample coefficient of correlation when the population is non-normal

It is known that when the population is bivariate normal, r- the sample coefficient of correlation provides an UMPU procedure for testing independence for both one-sided and two-sided alternatives. In order to study the performance of this procedure in non-normal cases, we require the distribution of r when the population is non-normal. In this section we discuss an approximation to the distribution of r in the general non-normal situation and study its accuracy by numerical methods and model sampling experiments. Some of the results discussed here have already been published [Dasgupta (1968c)].

4.3.1 An approximation to the distribution of r

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be n pairs of observations drawn at random from a bivariate population. Let us denote by $f(x)$ yhe probability density function of $X = \frac{1}{2}(r + 1)$ where

$$r = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) / \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \times \sum_{i=1}^{n} (y_i - \bar{y})^2\right]^{\frac{1}{2}}$$

Using the method developed in section 2.2, we have the following fourth order approximation to the cumulative distribution function of r in terms of $\beta(x;\ a, b)$ at the point x

$$F_4 = B(x;\ a, b) - \beta(x;\ a+1, b+1)$$
$$\times\ [\ d(3)\ J(2, x;\ a+1, b+1) + d(4) J(3, x;\ a+1, b+1) \qquad (4.3.1)$$

where

μ_g = g-th moment of r, $g = 1, 2, \ldots$

$$a = \frac{2(1+\mu_1)^2 - (1+\mu_1)(1+2\mu_1+\mu_2)}{2(1+2\mu_1+\mu_2) - 2(1+\mu_1)^2}$$

$$b = \frac{(1-\mu_1)\ (1-\mu_2)}{2(1+2\mu_1+\mu_2) - 2(1+\mu_1)^2} \qquad (4.3.2)$$

and, $d(3)$, $d(4)$ and the Jacobi polynomials $J(r, x;\ a, b)$ are as defined in section 2.2. $\theta(g)$'s occurring in the expressions for $d(3)$ and $d(4)$ are given by

$$\theta(g) = g\text{-th moment of } X = \tfrac{1}{2}(1+r)$$

$$= \frac{1}{2^g} \sum_{j=o}^{g} \binom{g}{j} \mu_j .$$

The expressions for μ_k $(k = 1, 2, 3, 4)$ have been obtained in terms of the population cumulants to order n^{-2} by Cook (1951) and are not reproduced here. Thus knowing the first four moments of

r the expression (4.3.1) can be evaluated numerically for any x. It is to be noted that in order to use Cook's (1951) results we require the knowledge of the population cumulants.

4.3.2 Performance of the approximation (4.3.1) when the population is bivariate normal

When the population is bivariate normal $N(0, 0, 1, 1, \rho)$, the expressions for the first four moments of r, to order n^{-2}, reduce to (see Cook (1951)

$$\mu_1 = \rho\,[1 - \frac{1}{2n} - \frac{3}{8n^2} + \rho^2(\frac{1}{2n} - \frac{3}{4n^2}) + \frac{\rho^4 9}{8n^2}]$$

$$\mu_2 = \rho^2[\frac{1}{\rho^2}(\frac{1}{n} + \frac{1}{n^2}) + 1 - \frac{3}{n} + \frac{3}{n^2} + \rho^2(\frac{2}{n} - \frac{12}{n^2}) + \rho^4\frac{8}{n^2}]$$

$$\mu_3 = \rho^3\,[\frac{1}{\rho^2}(\frac{3}{n} - \frac{9}{2n^2}) + 1 - \frac{15}{2n} + \frac{261}{8n^2} + \rho^2(\frac{9}{2n} - \frac{225}{4n^2}) + \frac{225}{8n^2}\rho^4]$$

$$\mu_4 = \rho^4\,[\frac{1}{\rho^4}\frac{3}{n^2} + \frac{1}{\rho^2}(\frac{6}{n} - \frac{36}{n^2}) + 1 - \frac{14}{n} + \frac{129}{n^2} + \rho^4(\frac{8}{n} - \frac{168}{n^2}) + \frac{72}{n^2}\rho^4]$$

We present below a few values of the distribution function of r at the point u as given in David's (1938) table and the corresponding approximate values using (4.3.1), when the parent population is bivariate normal and also the approximation one would obtain using Fisher's transformation $z = \tanh^{-1} r$. and assuming normality for it.

TABLE 4.7 COMPARISON OF THE APPROXIMATION WITH THE EXACT DISTRIBUTION OF r WHEN THE PARENT POPULATION IS BIVARIATE NORMAL

			Prob$(r \leq u \mid \rho)$		
n	ρ	u	exact	approximation (4.3.1)	Fisher's z approximation
10	0.2	0.2	.4859	.5091	.5000
10	0.2	0.6	.9011	.8905	.9028
10	0.4	0.2	.2494	.2550	.2796
10	0.4	0.6	.7480	.7607	.7611
25	0.2	0.2	.4917	.5040	.5000
25	0.2	0.6	.9882	.9826	.9893
25	0.4	0.2	.1386	.1205	.1501
25	0.4	0.6	.8910	.8823	.8969
50	0.2	0.2	.4942	.5025	.5000
50	0.2	0.6	.9995	.9989	.9996
100	0.2	0.2	.4959	.5017	.5000

Table 4.7 shows that the agreement between the approximation and the exact distribution is not too bad. Fisher's z turns out to be better in most cases considered here. This of course is expected because Fisher's z is specifically meant for bivariate normal populations whereas our approximation is a general one and can be used in the case of any bivariate population for which cumulants are known.

We may note in passing that the approximation (4.3.1) coincides with the exact distribution if the population is bivariate normal with zero correlation coefficient.

4.3.3 Performance of the approximation in two non-normal situations

In order to judge the accuracy of Jacobi approximation (4.3.1) in non-normal situations we consider two cases - one with population correlation coefficient equal to zero and the other with a non-zero coefficient of correlation.

Case 1 : x and y are independently and uniformly distributed over the unit interval 0 to 1.

Here, for a sample of size n, the moments of r to order n^{-2} are, ignoring parent cumulants of order 4 or more,

$$\mu_1(r) = 0$$

$$\mu_2(r) = \frac{1}{n} + \frac{1}{n^2}$$

$$\mu_3(r) = 0$$

$$\mu_4(r) = \frac{3}{n^2} .$$

When $n = 100$, the expressions are

$$\mu_1(r) = 0, \quad \mu_2(r) = .0101, \quad \mu_3(r) = 0 \quad \mu_4(r) = .0003 .$$

Using the above moments of r and the approximation (4.3.1), third column of the table 4.8 is obtained. In order to see the suitability of the approximation in this case the following model sampling experiment is performed.

1000 sets of 100 pairs of random numbers are chosen. For each set a correlation between the random numbers of a pair is computed. The second column of the following table gives the cumulative frequency ratio as obtained from this experiment.

TABLE 4.8 RESULTS OF A MODEL SAMPLING EXPERIMENT AND THE JACOBI APPROXIMATION FOR THE DISTRIBUTION OF r IN THE CASE OF INDEPENDENTLY DISTRIBUTED VARIABLES

u	Prob$(r \leq u)$	
	model sampling result	Jacobi approximation
-0.2	.020	.023
-0.1	.154	.161
0.0	.488	.500
0.1	.846	.839
0.2	.974	.977
0.3	.998	.999
0.4	1.000	1.000

In evaluating the Jacobi approximation in the above case, it is noticed that the contribution due to the third and fourth moments of r are negligible. In a situation like this a simple beta approximation where the parameters of the distribution are chosen to agree with the first two moments of the variable $\frac{1}{2}(1 + r)$ is good enough in large samples.

Case 2 : Let us define

$$x = z_1$$

$$y = 0.2\, z_1 + z_2$$

where z_1 and z_2 are independently and uniformly distributed over the range 0 to 1.

Then the first two moments of r_{xy}, to order n^{-1}, taking $n = 100$ are (using Cook's (1951) result)

$$\mu_1(r) = 0.19467636$$

$$\mu_2(r) = 0.04739163$$

Hence from (4.3.2)

$$a = 59.9435$$

$$b = 40.4076$$

Ignoring the other terms, an approximation to the distribution function of $X = \frac{1}{2}(1 + r)$ at the point x is obtained as

$$P(x) = B(x\ ;\ 59.9435,\ \ 40.4076) \qquad (4.3.3)$$

column (3) of the following table gives the cumulative distribution function of r at different points using the approximation (4.3.3).

In order to have an idea of the exact distribution of r in this case the following model sampling experiment is performed.

1000 sets of 100 pairs of 4 digit random numers are chosen. From each pair (z_1, z_2) we construct a pair (x, y) by

$$x = z_1$$

$$y = 0.2\, z_1 + z_2$$

For each set of 100 pairs, the coefficient of correlation between the variables (x, y) is computed. The results of this experiment is presented in second column of the following table.

TABLE 4.9 BETA APPROXIMATION AND THE RESULTS OF A MODEL SAMPLING EXPERIMENT FOR THE DISTRIBUTION OF $r(z_1, 0.2\, z_1 + z_2)$ WHERE z_1 AND z_2 ARE INDEPENDENTLY UNIFORMLY DISTRIBUTED OVER (0.1).

u	Prob$(r \leq u)$	
	model sampling result	beta approximation
- .15	.000	.000
- .10	.001	.002
- .05	.003	.007
.00	.018	.024
.05	.054	.071
.10	.165	.167
.15	.321	.321
.20	.550	.516
.25	.731	.710
.30	.871	.860
.35	.959	.947
.40	.988	.985
.45	.998	.997
.50	1.000	1.000

CHAPTER 5

EFFECT OF DEVIATIONS FROM SIMPLE SAMPLING ON SOME STATISTICAL TESTS

5.0 Introduction and summary

Most of the studies on robustness of test procedures have been made with respect to departures of parent populations from normality and/or homoscedasticity of the basic variables. Not much seems to have been done to study the effect of deviations from simple random sampling. A notable exception is Anderson's (1940) paper which deals with the distribution of some statistics when stratified sampling scheme is used for different types of populations.

Situations are not infrequent when a statistician has no control over the collection of data, but has to analyse a set of data already collected. Under such situation, he is often forced to make a plausible assumption about the chance mechanism through which the data were generated. An assumption very often made is that the data arose out of a process of simple random sampling from the population under study. If this assumption is not valid, the statistical inferences drawn may or may not remain valid, depending whether the inference procedure is or is not robust in respect of deviations from the sampling mechanism assumed.

In section 5.1 we examine this issue in the context of the t-test of significance of the mean value based on a sample from a normal population, on the assumption that sampling is simple random.

If instead the sample arose out of a process of stratified simple random sampling with proportionate allocation, how might it have affected the operating characteristics or the power function of the test? This is answered by calculating the asymptotic Pitman power of the test. It turns out that this power decreases if any stratum is split up.

A similar question for non-normal populations is considered in sections 5.2 and 5.3.

Finally, we discuss an approximation to the distribution of student's t in a 'probability proportional to size' sampling scheme. This scheme, similar to the one discussed in Lahiri(1951), is defined for drawing a sample from a population with continuous density function, and the fact that the sample mean in this case provides an unbiased estimate of the population mean is proved.

All summations in this chapter unless otherwise indicated are over distinct terms only.

5.1 Pitman power of the student's t test when sampling is stratified

5.1.1 Let us consider a normal population with the density function

$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(x - m)^2\right] \qquad (5.1.1)$$

Suppose this population is divided into k strata, the i-th stratum consisting of the set of values

$$S_i : [\, x \mid x_{i-1} \leq x < x_i \,] \quad , \quad i = 1, 2, \ldots, k$$

where

$$x_o = -\infty$$

$$x_k = +\infty$$

and $\quad x_{i-1} < x_i \quad$ for all i

Suppose n observations have been drawn from this stratified population. Let n_i be the number of observations from the i-th stratum using proportional allocation under the assumption that $m = 0$ i.e.

$$n_i = n\,\theta_i \ , \qquad i = 1, 2, \ldots, k$$

where $\quad \theta_i = \int_{x_{i-1}}^{x_i} \phi(x)\, dx$

An user not knowing the way the sample has been chosen will obviously use for testing $H_o(m = o)$, when he does not have any knowledge about the population variance, the student's statistic

$$t = \frac{\sqrt{n}\,\bar{\bar{x}}}{s} \qquad (5.1.2)$$

where $\bar{\bar{x}}$ = the mean of all the observations

$$= \Sigma\, \theta_i\, \bar{x}_i$$

$\bar{x}_i = \dfrac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$, x_{ij} being the j-th observation from the i-th stratum

$$s^2 = \frac{1}{n-1} \sum_{i,\,j} (x_{ij} - \bar{\bar{x}})^2$$

In order to find out the effect of stratification on the procedure using student's t we compute its Pitman power for alternatives of the form H_1 $(m = \frac{\mu}{\sqrt{n}} > 0)$. The result in this connection is stated in the following theorem.

Theorem 5.1

The Pitman power of the student's t test, when the sample has been drawn using the stratified sampling scheme discussed above, for the null hypothesis H_o $(m = 0)$ against the alternatives H_1 $(m = \frac{\mu}{\sqrt{n}} > 0)$ is given by

$$1 - \Phi \left(\frac{t_\alpha}{\sigma} - \mu\sigma \right) \qquad (5.1.3)$$

where t_α = upper 100α percent point of the standardised normal distribution

$\sigma^2 = 1 - \Sigma\, \theta_i\, m_{io}^2$

and $m_{io} = E(x \mid S_i, H_o) = [\phi(x_{i-1}) - \phi(x_i)] / [\Phi(x_i) - \Phi(x_{i-1})]$

Proof :

Let $m_{ij} = E(x \mid S_i, H_j)$, $j = 0, 1$

Then

$$E(\bar{\bar{x}} \mid H_o) = \sum_{i=1}^{k} \theta_i\, m_{io} = 0 .$$

Therefore ,

$$\sqrt{n}\,\bar{\bar{x}} = \sqrt{n} \sum_i \theta_i \bar{x}_i = \sqrt{n} \sum_i \theta_i (\bar{x}_i - m_{io})$$

$$= \sum_i \sqrt{\theta_i} \sqrt{n_i}\, \bar{y}_i$$

where $y_{ij} = x_{ij} - m_{io}$ and $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$

We have

$$E(\bar{y}_i | H_1) = E(\bar{x}_i | H_1) - m_{io}$$

$$= \frac{\int_{x_{i-1}}^{x_i} x\phi(x - m)\, d\,x}{\int_{x_{i-1}}^{x_i} \phi(x - m)\, d\,x} - m_{io}$$

$$= m + \frac{\phi(x_{i-1} - m) - \phi(x_i - m)}{\Phi(x_i - m) - \Phi(x_{i-1} - m)} - m_{io}$$

Expanding $\phi(x_{i-1} - m) - \phi(x_i - m)$ and $\Phi(x_i - m) - \Phi(x_{i-1} - m)$

about $m = \frac{\mu}{\sqrt{n}} = 0$ and neglecting terms of order $n^{-r}, r \geq 1$,

we get

$$E(\bar{y}_i | H_1) = \frac{\mu}{\sqrt{n}} [1 - \frac{\phi'(x_{i-1}) - \phi'(x_i)}{\phi(x_{i-1}) - \phi(x_i)} m_{io} - m_{io}^2]$$

$$= \frac{\mu \sqrt{\theta_i}}{\sqrt{n_i}} [1 - m_{io} \frac{\phi'(x_{i-1}) - \phi'(x_i)}{\phi(x_{i-1}) - \phi(x_i)} - m_{io}^2]$$

$$\therefore \quad \delta_i = \underset{n_i \to \infty}{\mathrm{Lt}} \sqrt{n_i}\, E(\bar{y}_i)$$

$$= \sqrt{\theta_i}\, \mu[1 - m_{io} \frac{\phi'(x_{i-1}) - \phi'(x_i)}{\phi(x_{i-1}) - \phi(x_i)} - m_{io}^2] \qquad (5.1.5)$$

$$\sigma_{i1}^2(n) = V(y_{ij}|H_1) = V(x_{ij}|H_1)$$

$$= \frac{\int_{x_{i-1}-m}^{x_i-m} (y+m)^2 \phi(y)\, dy}{\Phi(x_i - m) - \Phi(x_{i-1} - m)} - (\frac{\delta_i}{\sqrt{n}} + m_{io})^2 .$$

Therefore, $\underset{n_i \to \infty}{\mathrm{Lt}} \sigma_{i1}^2 (n) = \dfrac{\int_{x_{i-1}}^{x_i} y^2 \phi(y)\, dy}{\theta_i} - m_{io}^2$

$$= V(x_{ij}|H_o) = \sigma_{io}^2 \quad (\text{say}) \qquad (5.1.6)$$

Hence, using lemma 4.1 , $\sqrt{n_i}\, \bar{y}_{ij}$ asymptotically follows a normal distribution with mean δ_i and variance σ_{io}^2 where δ_i and σ_{io}^2 are as defined in (5.1.5) and (5.1.6) respectively. This shows that

$\sqrt{n}\, \bar{\bar{x}} = \sum_{i=1}^{k} \sqrt{\theta_i} \sqrt{n_i}\, \bar{y}_i$ asymptotically follows a normal distribution with mean $D = \Sigma \sqrt{\theta_i}\, \delta_i$ and variance $\sigma^2 = \Sigma\, \theta_i\, \sigma_{io}^2$. Again, noting that $E(s^2|\, H_1)$ converges in probability to the constant $\Sigma\, \theta_i\, \sigma_{io}^2 + \Sigma\, \theta_i (m_{io} - \bar{\bar{m}})^2 = 1$, where $\bar{\bar{m}} = \Sigma\, \theta_i\, m_{io}$, the student's ratio asymptotically follows a normal distribution with mean D and standard deviation σ. After simplification,

$$D = \mu\, [1 - \Sigma\, \theta_i\, m_{io}^2] \tag{5.1.7}$$

and

$$\sigma^2 = 1 - \Sigma\, \theta_i\, m_{io}^2 \tag{5.1.8}$$

Hence, the Pitman power of the students' t test in stratified samples for the null hypothesis $H_o\ (m = o)$ against the alternatives $H_1\ (m = \frac{\mu}{\sqrt{n}} > 0)$ is given by

$$\underset{n \to \infty}{\text{Lt}}\ \text{Prob}\ (t \geq t_\alpha \mid H_1)$$

$$= 1 - \Phi\left(\frac{t_\alpha - D}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{t_\alpha}{\sigma} - \mu\, \sigma\right) \tag{5.1.9}$$

As we have noted in theorem 4.3 the quantity $\Sigma\, \theta_i\, m_{io}^2$ increases if we split up any stratum. This shows that the effect of increasing stratification by splitting up any stratum is to increase the quantity $\frac{t_\alpha}{\sigma} - \mu\, \sigma$ and hence to reduce the Pitman power (5.1.9). It may be noted that the maximum value of σ is 1 which is attained when there is no stratification, and the entire population consists of only one stratum. In this case the Pitman power will be

$$1 - \Phi(t_\alpha - \mu) \tag{5.1.10}$$

The comparison of (5.1.9) with (5.1.10) will give us an idea about the loss of power due to stratification.

5.1.2 Table

The following table gives the Pitman power (5.1.9) when there are two strata. The division point x_1 is chosen in such a way that

$$\theta = \int_{-\infty}^{x_1} \phi(x)\, dx$$

The column which gives the Pitman power for $\theta = 0$ presents the values obtained using (5.1.10).

TABLE 5.1 PITMAN POWER OF THE STUDENT'S t TEST IN STRATIFIED SAMPLES

size = .05

μ \ θ	.00	.005	.01	.025	.05
0.0	.050	.046	.044	.038	.031
0.1	.061	.057	.053	.046	.037
0.2	.074	.069	.065	.056	.045
0.3	.089	.083	.078	.067	.054
0.4	.107	.099	.093	.080	.065
0.5	.126	.117	.110	.095	.077
0.6	.149	.137	.130	.112	.090
0.7	.172	.160	.151	.131	.106
0.8	.199	.185	.175	.151	.123
0.9	.228	.212	.200	.174	.142
1.0	.259	.241	.228	.199	.162
1.2	.328	.306	.291	.254	.209
1.4	.403	.378	.360	.317	.263
1.6	.482	.455	.434	.386	.323
1.8	.561	.532	.511	.459	.389
2.0	.639	.609	.587	.532	.458
2.2	.711	.682	.660	.605	.528
2.4	.775	.748	.728	.674	.598
2.6	.830	.806	.787	.738	.664
2.8	.876	.855	.839	.794	.726
3.0	.912	.895	.881	.843	.781
3.4	.961	.950	.942	.916	.870
3.8	.984	.979	.974	.960	.931
4.2	.995	.992	.990	.983	.967
4.6	.998	.997	.997	.994	.986
5.0	.999	.999	.999	.998	.994

contd.

TABLE 5.1 (Contd.)

θ / μ	.00	.10	.15	.20	.25
0.0	.050	.021	.015	.011	.008
0.1	.061	.026	.018	.013	.009
0.2	.074	.031	.021	.015	.011
0.3	.089	.037	.026	.018	.013
0.4	.107	.044	.031	.022	.016
0.5	.126	.052	.036	.025	.019
0.6	.149	.061	.043	.030	.022
0.7	.172	.072	.050	.035	.026
0.8	.199	.084	.059	.041	.030
0.9	.228	.097	.068	.048	.035
1.0	.259	.112	.079	.056	.041
1.2	.328	.146	.103	.074	.054
1.4	.403	.186	.133	.096	.071
1.6	.482	.232	.169	.123	.091
1.8	.561	.285	.209	.154	.116
2.0	.639	.342	.255	.191	.144
2.2	.711	.404	.306	.232	.177
2.4	.775	.468	.361	.278	.214
2.6	.830	.532	.420	.328	.256
2.8	.876	.596	.480	.381	.301
3.0	.912	.657	.540	.436	.350
3.4	.961	.767	.657	.550	.456
3.8	.984	.854	.760	.660	.563
4.2	.995	.916	.943	.757	.667
4.6	.998	.956	.905	.837	.759
5.0	.999	.979	.947	.897	.835

Contd.

TABLE 5.1 (Contd.)

μ \ θ	0.0	.30	.35	.40	.50
0.0	.050	.006	.005	.004	.003
0.1	.061	.007	.006	.005	.004
0.2	.074	.008	.007	.006	.005
0.3	.089	.010	.008	.007	.006
0.4	.107	.012	.009	.008	.007
0.5	.126	.014	.011	.009	.008
0.6	.149	.016	.013	.011	.009
0.7	.172	.019	.015	.012	.011
0.8	.199	.022	.017	.014	.012
0.9	.228	.026	.020	.017	.014
1.0	.259	.030	.024	.020	.017
1.2	.328	.040	.032	.026	.022
1.4	.403	.053	.042	.035	.030
1.6	.482	.069	.054	.045	.039
1.8	.561	.088	.070	.058	.050
2.0	.639	.111	.088	.074	.064
2.2	.711	.137	.111	.093	.080
2.4	.775	.168	.136	.115	.100
2.6	.830	.203	.165	.141	.123
2.8	.876	.242	.199	.170	.149
3.0	.912	.284	.235	.203	.179
3.4	.961	.378	.319	.279	.248
3.8	.984	.480	.414	.367	.330
4.2	.995	.583	.514	.463	.422
4.6	.998	.682	.613	.561	.517
5.0	.999	.768	.705	.655	.612

5.2 Effect of stratification on the power function of Student's t test - general non-normal situation

Let $f(x)$ be the density function of a population with mean m. We assume that the population is divided into p strata. Suppose $x_{i1}, x_{i2}, \ldots, x_{in_i}$ are n_i observations drawn at random from the i-th stratum. We want to study the effect of stratification on the power function of the student's t test for testing $H_0(m = m_0)$ which one uses when the assumptions of normality of the population and simple randomness of the nature of sampling are valid.

The square of the student's statistic is

$$t^2 = \frac{n(\bar{\bar{x}} - m_0)^2}{s^2} \tag{5.2.1}$$

where $\bar{x}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} x_{ti}$, $\bar{\bar{x}} = \frac{1}{n} \sum_{t=1}^{p} n_t \bar{x}_t$

$$n = \sum_{t=1}^{p} n_t \quad \text{and} \quad s^2 = (n-1)^{-1} \sum_{t,i} (x_{ti} - \bar{\bar{x}})^2$$

Let $K_{1t} = \bar{x}_t$

and $K_{2t} = (n_t - 1)^{-1} \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2$

be the first two K-statistics computed from the sample observations from the t-th stratum.

Then

$$\bar{\bar{x}} = \frac{1}{n} \quad \Sigma \quad n_t \; K_{1t} \tag{5.2.2}$$

Let us write

$$X = n(\bar{\bar{x}} - m_o)^2.$$

Then

$$X = n[\frac{1}{n^2} (\Sigma \; n_t^2 \; K_{1t}^2 + 2\Sigma \; n_t n_{t'} K_{1t} K_{1t'}) - \frac{2m_o}{n} (\Sigma \; n_t K_{1t}) + m_o^2] \tag{5.2.3}$$

and

$$X^2 = \frac{1}{n^2} (\Sigma \; n_t K_{1t})^4 - \frac{4}{n} (\Sigma \; n_t K_{1t})^3 \; m_o$$

$$+ 6(\Sigma \; n_t K_{1t})^2 \; m_o^2 - 4n m_o^3 (\Sigma \; n_t K_{1t}) + n^2 m_o^4$$

$$= \frac{1}{n^2} [\Sigma \; n_t^4 \; K_{1t}^4 + 4 \; \Sigma \; n_t^3 n_{t'} K_{1t}^3 K_{1t'} + 6 \; \Sigma \; n_t^2 n_{t'}^2 K_{1t}^2 K_{1t'}^2$$

$$+ 12\Sigma \; n_t^2 n_{t'} n_{t''} K_{1t}^2 K_{1t'} K_{1t''} + 24 \; \Sigma \; n_t n_{t'} n_{t''} n_{t'''} K_{1t} K_{1t'} K_{1t''} K_{it'''}]$$

$$- \frac{4m_o}{n} [\Sigma \; n_t^3 K_{1t}^3 + 3 \; \Sigma \; n_t^2 n_{t'} K_{1t}^2 K_{1t'} + 6\Sigma \; n_t n_{t'} n_{t''} K_{1t} K_{1t'} K_{1t''}]$$

$$+ 6m_o^2 [\; \Sigma \; n_t^2 K_{1t}^2 + 2\Sigma \; n_t n_{t'} K_{1t} K_{1t'} \;]$$

$$- 4m_o^3 n \; \Sigma \; n_t K_{1t} + n^2 m_o^4 \tag{5.2.4}$$

Let us write

$$S = \sum_{t=1}^{p} \; \sum_{i=1}^{n_t} (x_{ti} - \bar{\bar{x}})^2$$

$$S_1 = \sum_{t=1}^{p} n_t (\bar{x}_t - \bar{\bar{x}})^2$$

$$S_2 = \sum_{t=1}^{p} \; \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2.$$

Then $S = S_1 + S_2$ and $s^2 = (n - 1)^{-1}S$

Let

$$k_1 = \frac{1}{n} \Sigma\, n_t k_{1t}, \quad \text{where} \quad k_{rt},\ r = 1, 2, \ldots,$$

is the r-th cumulant of the t-th stratum,

$$C_t = K_{1t} - k_1 \ ,$$

and

$$\sigma_c^2 = \frac{1}{n} \Sigma\, n_t\, C_t^2 \ .$$

Then following David and Johnson (1951a) we write

$$S_1 = \Sigma\, n_t K_{1t}^2 - \frac{1}{n} \Sigma\, n_t^2 K_{1t}^2 - \frac{2}{n}\Sigma\, n_t n_{t'} K_{1t} K_{1t'} + 2\Sigma\, n_t C_t K_{1t} + n\sigma_c^2 \qquad (5.2.5)$$

and $S_2 = \Sigma(n_t - 1)K_{2t}$ (5.2.6)

For the product XS we have

$$XS = n[\, \bar{\bar{x}}^2\, S_1 + \bar{\bar{x}}^2\, S_2 - 2m_o\, (\, \bar{\bar{x}}\, S_1 + \bar{\bar{x}}\, S_2) + m_o^2\, (S_1 + S_2)\} \qquad (5.2.7)$$

Let us define

$$\alpha_r\, (s\, ,\, t) = E(K_{st}^r) \qquad (5.2.8)$$

and

$$\alpha_{r,s}\, (u\, , v\, ;\, t) = E(K_{u\,t}^r \quad K_{v\,t}^s) \qquad (5.2.9)$$

It is easy to see that [David, Kendall and Barton (1966)]

$$\alpha_1(s, t) = k_{st}$$

$$\alpha_2(1, t) = \frac{k_{2t}}{n_t} + k_{1t}^2$$

$$\alpha_3(1, t) = \frac{k_{3t}}{n_t^2} + \frac{3k_{2t}k_{1t}}{n_t} + k_{1t}^3$$

$$\alpha_4(1, t) = \frac{k_{4t}}{n_t^3} + \frac{3k_{2t}^2 + 4k_{1t}k_{3t}}{n_t^2} + \frac{6k_{2t}k_{1t}^2}{n_t} + k_{1t}^4$$

$$\alpha_{1,1}(1, 2, t) = \frac{k_{3t}}{n} + k_{1t}k_{2t}$$

and

$$\alpha_{2,1}(1, 2, t) = \frac{k_{4t}}{n^2} + k_{2t}\left(\frac{k_{2t}}{n} + k_{1t}^2\right)$$

Then

$$\mu'_{10} = E(X) = \frac{1}{n}\left[\Sigma n_t^2 \alpha_2(1,t) + 2\Sigma n_t n_{t'} k_{1t}k_{1t'} - 2m_o n \Sigma n_t k_{1t} + n^2 m_o^2\right] \qquad (5.2.10)$$

$$\mu'_{20} = E(X^2)$$

$$= \frac{1}{n^2}\left[\Sigma n_t^4 \alpha_4(1, t) + 4\Sigma n_t^3 n_{t'} \alpha_3(1, t)\alpha_1(1, t')\right.$$

$$+ 6\Sigma n_t^2 n_{t'}^2 \alpha_2(1, t)\alpha_2(1, t')$$

$$+ 12\Sigma n_t^2 n_{t'} n_{t''} \alpha_2(1, t)\alpha_1(1, t')\alpha_1(1, t'')$$

$$\left.+ 24\Sigma n_t n_{t'} n_{t''} n_{t'''} \alpha_1(1,t)\alpha_1(1, t')\alpha_1(1, t'')\alpha_1(1, t''')\right]$$

$$- \frac{4m_o}{n} [\Sigma\, n_t^3\, \alpha_3(1, t) + 3\Sigma\, n_t^2 n_{t'}\, \alpha_2(1, t)\, \alpha_1(1, t')$$

$$+ 6\Sigma\, n_t n_{t'} n_{t''}\, \alpha_1(1, t)\, \alpha_1(1, t')\, \alpha_1(1, t'')]$$

$$- 6m_o^2 [\Sigma\, n_t^2\, \alpha_2(1, t) + 2\Sigma\, n_t n_{t'}\, \alpha_1(1, t)\, \alpha_1(1, t')]$$

$$- 4m_o^3\, n\, [\Sigma\, n_t\, \alpha_1(1, t)] + n^2 m_o^4$$

$$\mu_{20} = Var\,(X) = \mu'_{20} - \mu'^2_{10} \qquad (5.2.11)$$

$$\mu'_{01} = E(s^2) = \frac{1}{n-1}\, E(S)$$

$$= [\Sigma\, k_{2t}\, n_t(1 - \frac{1}{n}) + n\sigma_c^2\,] / (n - 1) \qquad (5.2.12)$$

David and Johnson (1951a) have computed $V(S_1)$, $V(S_2)$ and Cov (S_1, S_2). Using these

$$\mu_{02} = V(S) / (n-1)^2$$

$$= \frac{1}{(n-1)^2} [V(S_1) + V(S_2) + 2\ \mathrm{Cov}\ (S_1, S_2)]$$

$$= \frac{1}{(n-1)^2} [[\ \Sigma\, \frac{k_{4t}}{n_t} (1 - \frac{n_t}{n})^2 + 2\Sigma\, k_{2t}^2\, (1 - \frac{2n_t}{n})$$

$$+ \frac{2}{n^2} (\Sigma\, n_t\, k_{2t})^2 + 4\Sigma\, k_{3t}\, C_t (1 - \frac{n_t}{n})$$

$$+ 4\Sigma\, n_t\, k_{2t}\, C_t^2] + [\Sigma (n_t - 1)^2\, (\frac{k_{4t}}{n_t} + \frac{2k_{2t}^2}{n_t - 1})]$$

$$+ 2[\Sigma \frac{n_t - 1}{n_t} (1 - \frac{n_t}{n})\, k_{4t} + 2\Sigma (n_t - 1)\, C_t\, k_{3t}]] \qquad (5.2.13)$$

$$\mu'_{11} = E(X\, s^2) = \frac{n}{n-1}\,[A_1 + A_2 - 2m_o(B_1 + B_2) + (n-1)m_o^2\,\mu'_{01}] \qquad (5.2.14)$$

where

$$A_1 = E(\bar{\bar{x}}^2\, S_1)$$

$$= \frac{1}{n^2}\,\Sigma\, n_t^2(1 - \frac{n_t}{n})\alpha_4(1,t) + \frac{1}{n^2}\Sigma\, n_t n_{t'}(1-\frac{n'_t}{n})\alpha_2(1,\, t)\ \alpha_2(1,\, t')$$

$$- \frac{2}{n^3}\,\Sigma\, n_t^2 n_{t'}\alpha_3\,(1,\, t)\ \alpha_1(1,\, t') - \frac{2}{n^3}\Sigma\, n_t n_{t'} n_{t''}\,\alpha_2(1,t)\alpha_1(1,t')\alpha_1(1,t'')$$

$$+ \frac{2}{n^2}\,\Sigma\, n_t^2\,\alpha_3\,(1,\, t)\ C_t + \frac{2}{n^2}\,\Sigma\, n_t n_{t'}\,\alpha_2(1,\, t)\ \alpha_1(1,\, t')\ C_{t'}$$

$$+ \frac{\alpha_c^2}{n}\,\Sigma\, n_t\ \alpha_2(1,\, t) + \frac{2}{n^2}\,\Sigma\, n_t^2(1 - \frac{n_t}{n})\ \alpha_3\,(1,\, t)\ \alpha_1(1,\, t')$$

$$+ \frac{2}{n^2}\,\Sigma\, n_t n_{t'} n_{t''}\,(1 - \frac{n_{t''}}{n}\,)\ \alpha_1(1,\, t)\ \alpha_1(1,\, t')\quad \alpha_2(1,\, t'')$$

$$- \frac{4}{n^3}\,[\Sigma\, n_t^2 n_{t'}^2\,\alpha_2(1,t)\alpha_2(1,t') + 2\Sigma\, n_t^2 n_{t'} n_{t''}\alpha_2(1,t)\alpha_1(1,t')\ \alpha_1(1,t'')$$

$$+ 2\Sigma\, n_t n_{t'} n_{t''} n_{t'''}\,\alpha_1(1,\, t)\ \alpha_1(1,\, t')\ \alpha_1(1,\, t'')\ \alpha_1(1,\, t''')]$$

$$+ \frac{4}{n^2}\,\Sigma\, n_t^2 n_{t'}\,\alpha_2(1,\, t)\quad \alpha_1(1,\, t')\ C_t + \frac{4}{n^2}\,\Sigma\, n_t n_{t'} n_{t''}\,C_t\ \alpha_1(1,\, t)$$

$$\times\ \alpha_1(1,\, t')\ \alpha_1(1,\, t'')$$

$$+\ \frac{2}{n}\ \sigma_c^2\ \Sigma\, n_t n_{t'}\,\alpha_1(1,\, t')\quad \alpha_1(1,\, t),$$

$$A_2 = E[\,\bar{\bar{x}}^2\ S_2]$$

$$= \frac{1}{n^2}\,\Sigma\, n_t^2(n_t-1)\ \alpha_{2,1}\ (1,\, 2,\, t) + \frac{1}{n^2}\,\Sigma\, n_t^2(n_{t'}-1)\ \alpha_2(1,\, t)\ \alpha_1(2,\, t')$$

$$+ \frac{2}{n^2} \Sigma\, n_t(n_t - 1)\, n_{t'}\, \alpha_{1,1}\,(1, 2, t)\, \alpha_1(1, t')$$

$$+ \frac{2}{n^2} \Sigma\, n_t n_{t'}(n_{t''} - 1)\, \alpha_1(1, t)\, \alpha_1(1, t')\, \alpha_1(2, t''),$$

$$B_1 = E[\,\bar{\bar{x}}\, S_1]$$

$$= \frac{1}{n} \Sigma\, n_t^2 (1 - \frac{n_t}{n})\, \alpha_3\,(1, t) + \frac{1}{n} \Sigma\, n_t n_{t'} (1 - \frac{n_{t'}}{n})\, \alpha_1(1, t)\, \alpha_2(1, t')$$

$$- \frac{2}{n^2} \Sigma\, n_t^2 n_{t'}\, \alpha_2(1, t)\, \alpha_1(1, t') - \frac{2}{n^2} \Sigma\, n_t n_{t'} n_{t''}\, \alpha_1(1, t) \alpha_1(1, t') \alpha_1(1, t'')$$

$$+ \frac{2}{n} \Sigma\, n_t^2\, \alpha_2(1, t) C_t + \frac{2}{n} \Sigma\, n_t n_{t'} \alpha_1(1, t)\, \alpha_1(1, t') + \sigma_e^2\, \Sigma\, n_t \alpha_1(1, t),$$

$$B_2 = E[\,\bar{\bar{x}}\;\; S_2]$$

$$= \frac{1}{n} \Sigma\, n_t(n_t - 1)\, \alpha_{1,1}\,(1, 2, t) + \frac{1}{n} \Sigma\, n_t (n_{t'} - 1)\, \alpha_1(1, t)\, \alpha_1(2, t').$$

Let us write

$$c_1 = \mu'_{10} / \mu_{20}$$

$$c_2 = \mu'_{01} / \mu_{02}$$

$$p = \mu'^2_{10} / \mu_{20}$$

$$q = \mu'^2_{01} / \mu_{02}$$

$$\rho = (\mu'_{11} - \mu'_{01}\;\; \mu'_{10}) / (\,\mu_{02}\;\; \mu_{20})^{\frac{1}{2}}$$

Then using (2.4.6), an approximation to the power function of the size-α student's t test for $H_o(m = m_o)$ for two sided alternatives is given by

$$P(\alpha) = B(z;q,p) + \rho\left(\frac{q}{p}\right)^{\frac{1}{2}} \frac{\Gamma(q+p)}{\Gamma(q)\,\Gamma(p)} z^q(1-z)^p\left[-1+\frac{p+q}{q}z\right] \tag{5.2.15}$$

where
$$z = \left[1 + \frac{c_1(1-x)}{c_2\, x}\right]^{-1},$$

$$x = \frac{1}{1 + t^2(\frac{1}{2}\alpha,\, n-1)},$$

and $t(\alpha, n)$ is the upper 100α percent point of the t distribution with n degrees of freedom.

5.3 Effect of stratification on Fisher's t and one-way classification F tests

5.3.0 In this section we give an outline of the procedure which may be useful in studying the effects of stratification on Fisher's t and one-way classification F tests.

5.3.1 Effect on Fisher's t test

Under assumptions of normality, homoscedasticity and simple random sampling, the statistic known as Fisher's t used for testing the equality of means of two populations P_1 and P_2 when samples from the populations have been drawn independently is

$$t = \frac{\sqrt{(n_1 + n_2)}\ (\bar{\bar{x}}(1) - \bar{\bar{x}}(2))}{\sqrt{\frac{S(1) + S(2)}{n_1 + n_2 - 2}}} \qquad (5.3.1)$$

where n_i, $\bar{\bar{x}}(i)$ and $S(i)$ are the sample size, sample mean and total sum of squares for the sample drawn from P_i, $i = 1, 2$. Let us suppose that the populations are stratified - P_1 is divided into q_1 strata and P_2 into q_2 strata, and we draw a sample of size n_{ij} from the j-th stratum of the population P_i, $i = 1, 2$, $j = 1, 2, \ldots, q_i$.

Then

$$n_i = \sum_{j=1}^{q_i} n_{ij}, \qquad i = 1, 2. \qquad (5.3.2)$$

Let us write $x_{jl}(i)$ = l-th observation of the j-th stratum of the population

$$P_i, \qquad i = 1, 2. \qquad (5.3.3)$$

Then

$$\bar{\bar{x}}(i) = \frac{1}{n_i} \sum_{j=1}^{q_i} \sum_{l=1}^{n_{ij}} x_{jl}(i) \qquad (5.3.4)$$

$$S(i) = \sum_{j=1}^{q_i} \sum_{l=1}^{n_{ij}} (x_{jl} - \bar{\bar{x}}(i))^2 \qquad (5.3.5)$$

We define $n = n_1 + n_2$,

$$X = n(\bar{\bar{x}}(1) - \bar{\bar{x}}(2))^2$$

$$= n(\bar{\bar{x}}^2(1) + \bar{\bar{x}}^2(2) - 2\bar{\bar{x}}(1)\,\bar{\bar{x}}(2)), \qquad (5.3.6)$$

$$Y = \frac{1}{n-2}\,[S(1) + S(2)], \qquad (5.3.7)$$

$$A(i, j) = E(\bar{\bar{x}}^j(i)), \qquad (5.3.8)$$

and

$$B(i, j) = E(S(i)\,\bar{\bar{x}}^j(i)) \qquad (5.3.9)$$

Then

$$E(X) = n\,[A(1, 2) + A(2, 2) - 2A(1, 1)\,A(2, 1)] \qquad (5.3.10)$$

$$E(X^2) = n^2[A(1, 4) + A(2, 4) - 4[(1, 3)A(2, 1) + A(2, 3)A(1, 1)] + 6A(1, 2)\,A(2, 2)] \qquad (5.3.11)$$

$$E(Y) = \frac{1}{n-2}\,[B(1, 0) + B(2, 0)] \qquad (5.3.12)$$

$$V(Y) = \frac{1}{(n-2)^2}\,[V(S(1)) + V(S(2))] \qquad (5.3.13)$$

$$E(XY) = \frac{n}{n-2}\,[B(1, 2) + A(1, 2)\,B(2, 0) + A(2, 2)\,B(1, 0) + B(2, 2) - 2B(1, 1)\,A(2, 1) - 2B(2, 1)\,A(1, 1)] \qquad (5.3.14)$$

The expectations A and B and the variances of $S(1)$ and $S(2)$ can be obtained as in section 5.2. Knowing (5.3.10) - (5.3.14), an approximation of the type (5.2.15) may be used as an approximation to the power function of Fisher's t test for two sided alternatives.

5.3.2 Effect on one-way classification F tests

In a standard situation, the statistic used for testing the equality of means for l different populations is

$$F = \frac{\sum_i n_i(\bar{x}_i - \bar{\bar{x}})^2/(l-1)}{\sum_{i,l} (x_{ij} - \bar{x}_i)^2/(n-l)} \qquad (5.3.16)$$

where

x_{ij} = j-th observation of the i-th population

n_i = size of the sample drawn from i-th population

$\bar{x}_i$ = sample mean for the i-th population

$$= \frac{1}{n_i} \sum_j x_{ij}$$

$$n = \sum_i n_i$$

$$\bar{\bar{x}} = \sum n_i \bar{x}_i / n$$

Let us denote the populations by $P_1, P_2, \ldots, P_l$. Suppose the population P_i is divided into q_i strata, and a sample n_{ij} is drawn from the j-th stratum of P_i.

Let us further write

x_{iju} = u-th observation from j-th stratum of of the population P_i

$$\bar{x}_{ij} = \frac{1}{n_{ij}} \sum_{u=1}^{n_{ij}} x_{iju}, \text{ and } \bar{x}_i = \frac{1}{n_i} \sum_j n_{ij} \bar{x}_{ij}$$

In terms of the observations x_{iju}, when the observations are drawn from the populations after stratification, the statistic F in (5.3.16) reduces to

$$F = \frac{\sum_i n_i(\bar{x}_i - \bar{\bar{x}})^2/(l-1)}{\sum_{i,j,u} (x_{iju} - \bar{x}_i)^2/(n-l)} \qquad (5.3.17)$$

Let us define

$$X = \sum n_i(\bar{x}_i - \bar{\bar{x}})^2$$

$$Y = \sum_{i,j,u} (x_{iju} - \bar{x}_i)^2$$

Then writing

$K_r(i, j)$ for the r-th K-statistic

computed from the sample chosen from the j-th stratum of P_i,

we get

$$X = \sum_i \frac{1}{n_i} [\sum_j n_{ij}^2 K_1^2(i, j) + 2 \sum_{j,j'} n_{ij} n_{ij'} K_1(i, j) K_1(i, j')]$$

$$- \frac{1}{n} \sum_{i,j} n_{ij}^2 K_1^2(i, j) - \frac{2}{n} \sum_{(i,j)\neq(i',j')} n_{ij} n_{i'j'} K_1(i,j) K_1(i',j') \qquad (5.3.18)$$

$$Y = \sum_i \sum_j \sum_u (x_{iju} - \bar{x}_i)^2$$

$$= \sum_i [\sum_j n_{ij}(\bar{x}_{ij} - \bar{x}_i)^2 + \sum_{j,u} (x_{iju} - \bar{x}_{ij})^2]$$

Now if we write

$k_r(i, j)$ = r-th cumulant of the j-th stratum of the population P_i

$$k_1(i) = \frac{1}{n_i} \sum n_{ij} k_1(i, j)$$

$$C_{ij} = k_1(i,\ j) - k_1(i)$$

$$\sigma^2_{c(i)} = \frac{1}{n_i} \Sigma\, n_{ij}\, c^2_{ij}$$

then

$$Y = \sum_i [\, \Sigma\, n^2_{ij}\, K^2_1(i,\ j) - \frac{1}{n_i} (\Sigma\, n_{ij}\, K_1(i,\ j))^2$$

$$+2\Sigma\, n_{ij}\, K_1(i,j)\, C_{ij} + n_i\, \sigma^2_{c\,(i)}] + \sum_i \sum_j (n_{ij} - 1)\, K_2(i,j) \qquad (5.3.19)$$

The moments $E(X)$, $E(Y)$, $V(X)$, $V(Y)$ and $Cov(X,Y)$ can be computed using (5.3.18) and (5.3.19) and the tables of symmetric functions [David, Kendall and Burton (1966)]. Since X and Y are both positive, an approximation to the distribution function of $Z = \frac{Y}{X+Y}$ is obtained using (2.4.6). This may then be used for approximating the power function of the F test.

5.4 Effect on the student's t when 'probability proportional to size' sampling scheme is used

5.4.1 The scheme

Suppose $f(x,\ y)$ is the density function of two random variables x and y. We draw a random observation $(x,\ y)$ from the population. This observation is selected if $r \leq g(x)$ where r has uniform distribution in the two sided closed interval 0 to 1, and $g(x)$ is some function of x satisfying the condition $0 < g(x) \leq 1$. If a drawn observation fails to be selected, we make a second attempt and go on attempting until an observation is selected. This is an extension in the case of continuous populations of Lahiri's (1951) scheme.

Result 5.1

If the above scheme is used to include n independent observations $(x_1, y_1), \ldots, (x_n, y_n)$ in the sample, and if $Z_i = \dfrac{y_i \, E[g(x)]}{g(x_i)}$, $i = 1, 2, \ldots, n$, then $\bar{Z} = \dfrac{1}{n} \sum_{i=1}^{n} Z_i$ is an unbiased estimate of $E(y) = \mu_y$.

Proof :

The probability that a drawn observation will be selected is

$$\int \int g(x) f(x, y) \, dy \, dx = E[g(x)] \qquad (5.4.1)$$

Therefore, the density function of the selected observations is

$$P(x,y) = g(x) f(x,y) + [1 - E(g(x))] \, g(x) f(x,y) + [1 - E(g(x))]^2 g(x) f(x,y) + \ldots$$

Since $0 < g(x) \leq 1$, so is $E[g(x)]$ and hence

$$P(x, y) = \frac{g(x) f(x,y)}{E[g(x)]} \qquad (5.4.2)$$

Therefore

$$E(\bar{Z}) = E(Z_i) = \int \int y \frac{E[g(x)]}{g(x)} \frac{g(x)}{E[g(x)]} f(x, y) \, dx \, dy$$

$$= \int \int y f(x, y) \, dx \, dy = \mu_y .$$

If the observations $y_1, y_2, \ldots, y_n$ are used as basic observations and if the student's t is used for testing $H_o(\mu_y = m_o)$, then

$$t = \frac{\sqrt{n}(\bar{y} - m_o)}{s} \qquad (5.4.3)$$

where $s^2 = (n - 1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$.

Noting that the density function of the selected observation is

$P(x, y) = \frac{g(x) f(x, y)}{E[g(x)]}$, it is possible to compute the expectations

$$\mu'_{10} = E[n(\bar{y} - m_o)^2]$$

$$\mu'_{20} = E[n^2(\bar{y} - m_o)^4]$$

$$\mu'_{01} = E(s^2)$$

$$\mu'_{02} = E(s^4)$$

and $\mu'_{11} = E[s^2 n(\bar{y} - m_o)^2]$.

Then using these moments a simple approximation to the distribution function of $\frac{1}{1 + t^2}$ can be obtained using (2.4.6).

APPENDIX

```
C      PROGRAM FOR COMPUTING APPROXIMATE POWER OF ANOVA PROCEDURE
C      USING APPROXIMATION (3.3.19).
C      INPUT ---  PUNCH  E(U1),E(U2),V(U1),V(U2),COV(U1,U2) AND X
C      ON A CARD WITH FORMAT 6F12.4
C      OUTPUT ---- P,Q,R AND F(X) AS IN (3.3.19)
1      READ 100, X1,X2,X11,X22,X12,Z
100    FORMAT(6F12,4)
       ROW =X12/SQRT(X11)/SQRT(X22)
       C1=X1/X11
       C2=X2/X22
       A=X2*C2
       B=X1*C1
       X=1.0 +C1*(1.0-Z)/(C2*Z)
       X=1.0/X
       CALL BETAB(A,B,X,R,C)
       PROB=SQRT(A)*(-1.0 +(A+B)*X/A)/SQRT(B)
       PROB= PROB*X**A*(1,0-X)**B/C
       COR=ROW*PROB
       RESLT=R+COR
       PRINT 300, A,B,ROW,RESLT
300    FORMAT(5X,4F20.8)
       GO TO 1
       END
       SUBROUTINE BETAB(A,B,X,P,CB)
       I=1
       U=A
       V=B
     5 Y=0.5
     6 R=Y/(1.0-Y)
       S=0.
       W=V-1.
       IF(W)22,22,10
    10 F=1.0/U
       T=F*(Y**U)*((1.0-Y)**(V-1.0))
    15 S=S+T
       W=W-1.0
       IF(W)25,25,20
    20 G=(W+1,0)/(U+V-W-1,0)
       F=F*G
       T=T*G*R
```

```
   GO TO 15
22 F=1.0
   GO TO 27
25 F=F*(W+1.0)
27 U=U+V-W-1.0
   W=-W
   C=0.0
   T=F*(Y**U)/U
30 S=S+T
   C=C+1.0
   T=T*(W+C-1.0)*(U+C-1.0)/(C*(U+C))*Y
   IF(T-.00000005)35,35,30
35 GO TO (40,45,55,65),I
40 U=B
   V=A
   CB=S
   I=2
   GO TO 5
45 CB=CB+S
47 IF(X-0.5)50,50,60
50 U=A
   V=B
   Y=X
   I=3
   GO TO 6
55 BR=S/CB
   GO TO 70
60 U=B
   V=A
   Y=1.0-X
   I=4
   GO TO 6
65 BR=1.0-S/CB
70 P=BR
   RETURN
```

BIBLIOGRAPHY

1. Anderson, P. H. (1942) : Distributions in stratified sampling. Ann. of Math. Stat., 13, 42.

2. Atiqullah, M. (1962) : The estimation of residual variance in quadratically balanced least square problems and the robustness of the F test. Biometrika, 49, 83.

3. -- (1964) : The robustness of the covariance analysis of a one-way classification. Biometrika, 51, 365.

4. Banerjee, S. (1962) : The problem of testing linear hypotheses about population means when the population variances are not equal. Sankhyā, Sr. A, 24, 363.

5. Bennet, B.M. and Hsu, P. (1960) : On the power function of the exact test for the 2 x 2 contingency table. Biometrika, 47, 383.

6. Bartlett, M.S. (1935) : The effect of non-normality on the t distribution. Proc. Camb. Phill. Soc., 31, 223.

7. Bhattacharya, B.C. (1943) : On an aspect of Pearsonian system curves and analogies. Sankhyā, 6, 415.

8. Box, G.E.P. (1953) : Non-normality and tests on variance. Biometrika, 40, 318.

9. -- (1954a): Some theorems on Q-forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in one way classification. Ann. of Math. Stat., 25, 290.

10. --- (1954b): Some theorems on Q-forms applied in the study of analysis of variance problems:
II. Effect of inequality of variance and of correlation between errors in the two way classification.
Ann. of Math. Stat., 25, 484.

11. ------ and Watson, G.S. (1962) : Robustness and non-normality of regression tests.
Biometrika, 49, 93.

12. ------- and Tiao, G. C. (1964) : A note on criterion robustness and inference robustness.
Biometrika, 51, 169.

13. ------- and Anderson, S.L. (1955) : Permutation theory in the derivation of Robust criteria and the study of departures from assumptions.
Jour. of Roy. Stat. Soc., Sr. B, 17,1.

14. Braverman, D. (1962) : Learning filters for optimum pattern recognition.
IRE Tran. on Inf. Th., IT-8, 280.

15. Chakravorti, S.R. (1967) : Effect of the inequality of variances in Analysis of Variance F-test.
Cal. Stat. Assocn. Bull., 16, 103.

16. Chapman, D.G. and Meng, R.C. (1966) : The power of chi-square tests for contingency tables.
Jour. of Amer. Stat. Assocon., 61, 965.

17. Cochran, W.G. (1947) : Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics, 3, 22.

18. ------ (1952) : The test of goodness of fit. Ann. of Math. Stat., 23, 315.

19. Collier, R.O. and Baker, F.B. (1963) : The randomisation distributions of F ratios for the split plot design - an emprical investigation. Biometrika, 50, 431.

20. ------ (1966) : Some Monte Carlo results on the power of the F test under permutation in the simple randomised block design Bilmetrika, 53, 199.

21. Cook, M.B. (1951) : Two applications of bivariate k-statistics. Biometrika, 38, 368.

22. Cramer, H. (1925) : On some class of series used in mathematical statistics. Sixth Scandinovain Cong. of Math., KØbenhavn.

23. Cramer, H. (1928) : On the composition of elementary errors. Scand. Aktdr., 13, 141.

24. ------ (1962) : Mathematical methods of statistics. Asia Publishing House, Bombay.

25. Dasgupta, P. (1968a): Tables of the non-centrality parameter of F test as a function of power. Sankhyā, Sr. B, 30, 73.

26. ------ (1968b): Two approximations for the distribution of double non-central beta. Sankhyā, Sr. B, 30, 83.

27. ------ (1968c): An approximation to the distribution of sample correlation coefficient when population is non-normal. Sankhyā, Sr. B, 30, 425.

28. David, F.N. (1938) : Tables of the ordinates and Probability Integral of the distribution of the correlation coefficient in small samples. Cambridge University Press.

29. ------, Kendall, M.G. and Barton, D.E. (1966) : Symmetric functions and allied tables. Cambridge University Press.

30. David, F.N. and Johnson, N.L. (1951a): The effect of non-normality on the power function of the F test in the analysis of variance. Biometrika, 38, 43.

31. David, F.N. and Johnson, N.L. (1951b): A method of investigating the effect of non-normality and heterogeneity of variance on tests of the general linear hypothesis. Ann. of Math. Stat., 22, 282.

32. ------ (1952) : Extension of a method of investigating the properties of Analysis of Variance tests to the case of random and mixed models. Ann. of Math. Stat. 23, 594.

33. David, F.N. and Fix, E. (1960) : Rank correlation and regression in a non-normal surface. Proc. of 4th Berkeley Symp., Vol-I, p-17.

34. Department of Commerce (US National Bureau of Standards) (1959) : Tables of the bivariate normal distribution function and related functions.

35. Donaldson, T.S. (1967) : Power of the F-test for non-normal distributions and unequal error variances.
Memorandom, RM-5072-PR,
The Rand Corporation.

36. Durbin, J. and Watson, G.S. (1951) : Testing for serial correlation in least squares regression II.
Biometrika, 38, 159.

37. Eisenhart, C. (1938) : The power function of the χ^2 -test.
Bull. of Amer. Math. Soc., 44, 32.

38. Gayen, A.K. (1949) : The distribution of students t in random samples of any size drawn from non-normal universes.
Biometrika, 36, 353.

39. Gayen, A.K. (1950a): The distribution of the variance ratio in random samples of any size drawn from non-normal universes.
Biometrika, 37, 236.

40. ------ (1950b): Significance of difference between the means of two non-normal samples.
Biometrika, 37, 399.

41. ------ (1951) : The frequency distribution of the product moment correlation coefficient in random samples of any size drawn from non-normal universes.
Biometrika, 38, 219.

42. ------ (1971) : Robust tests and estimation in statistics - Presidential address at the section of Statistics. 58th session of the Indian Science Congress, Bangalore.

43. Geary, R.C. (1936) : The distribution of students ratio for non-normal samples. Jour. of Roy. Stat. Soc., Supp. 3, 178.

44. Gurland, J. (1953) : Distribbutions of quadratic forms and ratios of quadratic forms. Ann. of Math. Stat., 24, 416.

45. ------ (1956) : Quadratic forms in normally distributed random variables. Sankhya, 17, 37.

46. Gurland, J. and McCullough, R.S. (1962) : Testing of equality of means after a preliminary test of equality of variances. Biometrika, 49, 403.

47. Hack, H.R.B. (1958) : An emprical investigation into the distribution of the F-ratio in samples from two non-normal populations. Biometrika, 45, 266.

48. Hamdan, M.A. (1963) : The number and width of classes in the chi-square test. Jour. of Amer. Stat. Assoc., 58, 678.

49. ------ (1968) : Optimum choice of classes for contingency tables. Jour. of Amer. Stat. Assoc., 63, 291.

50. Hammersley, J.M. (1949) : The unbiased estimate and standard error of the interclass variance. Metron, 15, 173.

51. Hannan, E.J. (1956) : The assymptotic powers of certain tests based on multiple correlations. Jour. of Roy. Stat. Soc., B, 18, 227.

52. Harkness, W.L. and Katz Leo (1964) : Comparison of the power function for the test of independence in (2 x 2) contingency tables. Ann. of Math. Stat., 35, 115.

53. Horsnell, G. (1953) : The effect of unequal group variances on the F-test for the homogeneity of group means. Biometrika, 40, 128.

54. Hooke, R. (1956) : Some applications of bipolykays to the estimation of variance components and their moments. Ann. of Math. Stat., 27, 80.

55. Hyrenius, H. (1950) : Distribution of Student - Fisher's t in samples of compound normal functions. Biometrika, 37, 429.

56. Johnson, N.L. and Kotz, S. (1968) : Tables of distributions of positive definite quadratic forms in central normal variables. Sankhyā, Sr. B, 30, 303.

57. Kailath, T. (1961) : Optimum receivers for randomly varying channels. Information Theory, p. 109.

58. Kempthorne, O et.al (1961) : Analysis of variance procedures. ARL report 149, Aeronautical Research Lab., United States Air Force.

59. Kendall, M.G. and Stuart, A. (1961) : Advanced Theory of Statistics Vol.2, Charles Griffin and Co., Ltd., London.

60. ------ (1966) : Advanced Theory of Statistics Vol.3, Charles Griffin and Co. Ltd., London.

61. Khamis, S.H. (1960) : Incomplete Gamma Function Expansions of Statistical Distribtbtion Functions. Bull. Inst. Int. Stat., 37(3).

62. Kruskal, W. (1968) : When are Gauss Markoff and least squares estimates identical? A co-ordinate free approach. Ann. of Math. Stat., 39, 70.

63. Kotz, S., Johnson, N.L. and Boyd, D.W. (1967) : Series representation of Quadratic forms in normal variables. I. Central case. Ann. of Math. Stat. 38, 823.

64. Lahiri, D.B. (1951) : A method of sample selection providing unbiased ratio estimators. Bull. Inter. Stat. Inst., 33(2), 133.

65. Madow, W.G. (1940) : Note on tests of departure from normality. Jour. of Math. Stat. Aseoc., 35, 515.

66. Mann, H.B. and Wald, A. (1942) : On the choice of the number of class intervals in the application of χ^2 tests. Ann. of Math. Stat. 13, 306.

67. Mitra, S.K. (1955) : Contributions to the statistical analysis of categorical data. North Carolina Inst. of Stat., Mimeo Sr. No. 142.

68. ------ (1958) : On the limiting power function of the frequency chi-square test. Ann. of Math. Stat., 29, 1221.

69. ------ and Rao, C.R. (1968) : Contributions for optimality and validity of simple least squares theory. Tech. report No. 30/68, RTS., Indian Statistical Institute.

70. Noether, G.E. (1955) : On a theorem of Pitman. Ann. of Math. Stat., 26, 64.

71. Patnaik, P.B. (1949) : The non-central χ and F distribution and their applications. Biometrika, 26, 202.

72. Pearson, E.S. (1931) : The analysis of variance in case of non-normal variation. Biometrika, 23, 114.

73. Pitman, E.J.G. (1937) : Significance tests which may be applied to samples from any population. Suppl. Jour. of Roy. Stat. Soc., 4, 119.

74. ------ (1948) : Non-parametric Statistical Inference North Carolina Inst. of Stat. (Mimeo Lecture Notes).

75. Price, R. (1962a): Error probabilities for adoptive multichannel reception of binary signals.
IRE Trans. on Information Theory IT-8, 305.

76. ------ (1962b): Error probability for adaptive multi-channel reception of binary signals. Addendum,
IRE Trans. on Information Theory, IT-8, 387.

77. ------ (1964) : Some non-central F distributions expressed in closed form.
Biometrika, 51, 107.

78. Quensel, C.E. (1938) : The distribution of the second moment and of the correlation coefficient in samples from populations of type - A.
/ Cited in Gayen (1951) /

79. Rao, C.R. (1951) : An assumptotic expansion of the distribution of Wilk's criterion.
Bull. Inst. Int. Stat., 33(2), 177.

80. ------ (1965) : Linear statistical inference and its applications.
John Wiley and Sons., New York.

81. Robbins, H. (1948) : The distribution of student's t when the population means are unequal.
Ann. of Math. Stat., 19, 406.

82. ------ and Pitman, E.J.G. (1949) : Application of the method of mixtures to quadratic forms in normal variates.
Ann. of Math. Stat., 20, 552.

83. Roy, J. (1956) : On some quick decision methods in multivariate and univariate analysis. Sankhya, 17, 77.

84. ------ (1965) : Power of the likelihood ratio test used in analysis of dispersion. Multivariate Analysis. Ed. by P.R. Krishnaiah.

85. ------ and Tiku, M.L. (1962) : A Laguerre series approximation to the sampling distribution of the variance. Sankhyā, Series A, 24, 181.

86. Roy, J. and Mohammed, J. (1964) : An approximation to the non-central chi-square distribution. Sankhyā, Sr. A, 26, 81.

87. Ruben, H. (1962) : Probability content of regions under spherical normal distributions. IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables. Ann. of Math. Stat., 33, 543.

88. Sansone, G. (1959) : Orthogonal functions. Interscience Publishers, Ins., New York.

89. Sheffe, H. (1959) : The Analysis of Variance. John Wiley and Sons, New York.

90. Sebestyen, G.S. (1961) : Recognition of membership in classes. IRE Trans. on Information Theory, IT-7, 44.

91. Siddiqui, M.M. (1965) : Approximations to the distribution of quadratic forms.
Ann. of Math. Stat., 36, 677.

92. Srivastava, A.B.L. (1958) : Effect of non-normality on the power function of t test.
Biometrika, 45, 421.

93. ------ (1959) : Effect of non-normality on the power of the Analysis of Variance test.
Biometrika, 46, 114.

94. Stevens, J. (1948) : Quality Control by Gauging.
Jour. of Roy. Stat. Soc., Sr. B.

95. Subramaniam, K. (1966) : Some contributions to the theory of non-normality I. Univariate case.
Sankhyā, Sr. A, 28, 389.

96. ------ (1968a): Some contributions to the theory of non-normality-II.
Sankhyā, Sr. A, 30, 411.

97. ------ (1968b): Some contributions to the theory of non-normality-III.
Sankhyā, Sr. B, 30, 383.

98. Szego, G. (1939) : Orthogonal polynomials.
Amer. Math. Soc. Colloquiem Publication, 33.

99. Tiku, M.L. (1963a): Student's t distribution in terms of Hermite and Laguerre polynomials.
Indian Math. Soc. Jour., 27, 91.

100. ------ (1963b): A Laguerre series approximation to one way classification variance ratio distribution.
Jour. of Ind. Soc. of Agri. Stat.

101. ------ (1964) : Approximating the general non-normal variance ratio sampling distributions.
Biometrika, 51, 83.

102. ------ (1965) : Laguerre series forms of non-central χ^2 and F distributions. Biometrika, 52, 415.

103. Tukey, J.W. (1956) : Variances of Variance components I: Balanced designs. Ann. of Math. Stat. 27, 722.

104. Tukey, J.W. (1957) : Variances of variance components II: The unbalanced single classification. Ann. of Math. Stat., 28, 43.

105. Turin, G.L. (1959) : Some computations of error rates for selectively fading multipath channels. Proc. Nat. Electron. Conf., 15, 431.

106. Watson, G.N. (1922) : A treatise on the theory of Bessel functions. Cambridge University Press.

107. Watson, G.S. (1967) : Linear least square regression. Ann. of Math. Stat., 38, 1679.

108. Welch, B.L. (1938a): The significance of the difference between two means when the population variances are unequal. Biometrika, 29, 350.

109. ------ (1938b): On tests of homogeneity. Biometrika, 30, 158.

110. Williams, C.A. Jr. (1950) : On the choice of the number and width of classes for the chi-square test of goodness of fit. Jour. Amer. Stat. Soc., 45, 77.

111. Wishner, R.P. (1952) : Distribution of the normalised periodogram detector.
IRE Trans. on Information Theory, IT - 8, 3429.

112. Zackrisson, U. (1959) : The distribution of student's t in samples from individual non-normal populations.
Publ. Statist. Inst., Univ. of Gothenburg, No. 6.

113. Zyskind, G. (1957) : On canonical forms, negative covariance matrices and best and simple least square linear estimators in linear models.
Ann. of Math. Stat., 38, 1092.

APPENDIX

```
C       PROGRAM FOR COMPUTING APPROXIMATE POWER OF ANOVA PROCEDURE
C       USING APPROXIMATION (3.3.19).
C       INPUT ---  PUNCH  E(U1),E(U2),V(U1),V(U2),COV(U1,U2) AND X
C       ON A CARD WITH FORMAT 6F12.4
C       OUTPUT ---- P,Q,R AND F(X) AS IN (3.3.19)
1       READ 100, X1,X2,X11,X22,X12,Z
100     FORMAT(6F12,4)
        ROW =X12/SQRT(X11)/SQRT(X22)
        C1=X1/X11
        C2=X2/X22
        A=X2*C2
        B=X1*C1
        X=1.0 +C1*(1.0-Z)/(C2*Z)
        X=1.0/X
        CALL BETAB(A,B,X,R,C)
        PROB=SQRT(A)*(-1.0 +(A+B)*X/A)/SQRT(B)
        PROB= PROB*X**A*(1,0-X)**B/C
        COR=ROW*PROB
        RESLT=R+COR
        PRINT 300, A,B,ROW,RESLT
300     FORMAT(5X,4F20.8)
        GO TO 1
        END
        SUBROUTINE BETAB(A,B,X,P,CB)
        I=1
        U=A
        V=B
     5  Y=0.5
     6  R=Y/(1.0-Y)
        S=0.
        W=V-1.
        IF(W)22,22,10
    10  F=1.0/U
        T=F*(Y**U)*((1.0-Y)**(V-1.0))
    15  S=S+T
        W=W-1.0
        IF(W)25,25,20
    20  G=(W+1,0)/(U+V-W-1,0)
        F=F*G
        T=T*G*R
```

```
      GO TO 15
   22 F=1.0
      GO TO 27
   25 F=F*(W+1.0)
   27 U=U+V-W-1.0
      W=-W
      C=0.0
      T=F*(Y**U)/U
   30 S=S+T
      C=C+1.0
      T=T*(W+C-1.0)*(U+C-1.0)/(C*(U+C))*Y
      IF(T-.00000005)35,35,30
   35 GO TO (40,45,55,65),I
   40 U=B
      V=A
      CB=S
      I=2
      GO TO 5
   45 CB=CB+S
   47 IF(X-0.5)50,50,60
   50 U=A
      V=B
      Y=X
      I=3
      GO TO 6
   55 BR=S/CB
      GO TO 70
   60 U=B
      V=A
      Y=1.0-X
      I=4
      GO TO 6
   65 BR=1.0-S/CB
   70 P=BR
      RETURN
```