# THE FIDUCIAL ARGUMENT IN STATISTICAL INFERENCE

by

Gregory W. BENNETT, B.A. (Hons), B.Sc.

A Thesis submitted for the degree of Doctor of Philosophy

in the University of Adelaide

Department of Mathematics

February, 1965.

CONTENTS

# SUMMARY

Some of the apparent inconsistencies, which have arisen from attempts to extend the fiducial argument to deal with the multivariate normal distribution, are examined against a background of three assumptions which have been derived from an examination of the various examples of the argument given by Fisher. These assumptions are designed to attempt to describe the relationship between sampling and fiducial densities so that the latter may be directly and simply derived from the former. It is shown that some of the difficulties encountered are a direct result of the desire to maintain too close an association between the two types of distribution and that it is generally unreasonable to expect such an analogy.

Multivariate situations demonstrate the need for considerable care in the choice of approach to the problem of forming inferences and the uniqueness property of a fiducial distribution is shown to be dependent not only on the data given but also on the complete specification of the problem to be solved, a requirement which needs more careful consideration than is the case with simple univariate densities to which the fiducial argument was first applied. The type of inconsistency seen by Mauldon in his choice of pivotal quantities, is shown to be due to an inadequate appreciation of this requirement.

The three well-known solutions to the question of finding the fiducial distribution of the parameters of a bivariate normal viz. those due to (i) Segal, (ii) Fisher in the Journal of the Royal Statistical Society and (iii) Fisher in the last section of "Statistical Methods and Scientific Inference", are compared numerically after the implicit factor in the last density is determined. It is proposed that the first is the correct result under symmetric conditions while the second, being based on the existence of a very close relationship between sampling and fiducial results, must be rejected because of the limited actual nature of this relationship. The last solution must be discarded because it is shown to be composed of elements relevant to two different natural specifications. Some suggestions are made about the extent to which the sampling-fiducial analogy is valid, the likelihood function playing an important part in this question.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

G.W. Bennett

## ACKNOWLEDGEMENTS

It is not very often possible for a research student to be able to work with the person directly responsible for the introduction of the basic ideas around which a thesis is to be constructed. I am, therefore, very grateful to be able to acknowledge my debt to the late Professor Sir Ronald Fisher, who was kind enough to discuss with me some of the points raised in the body of this work; at the same time I may express regret that time did not permit the discussion of the modifications to some of the fundamental ideas which have been suggested in the sections dealing with the multivariate normal density. Similarly I must record my thanks to my supervisor, Professor E.A. Cornish, for his assistance and encouragement during the preparation of this thesis, to my colleagues in the C.S.I.R.O. Division of Mathematical Statistics for many helpful discussions and to Professor D.A. Sprott with whom I have conducted a considerable correspondence.

The picture describing the various solutions proposed for the multivariate normal distribution could not have been completed without a great deal of numerical computing; I should like to express my gratitude to Mr. J.G. Sanderson for the use of CIRRUS, a machine developed by the University of Adelaide, and to the Computing Research Section of C.S.I.R.O. for time on its new CDC 3600-series machine. Finally I must thank Mrs. K. Teisseire who performed the unenviable task of typing this thesis.

## CANDIDATE'S NOTE

Throughout this thesis, the following conventions have been adopted

| | | |
|---|---|---|
| $\Gamma(a+b)/2$ | means | $\Gamma\{\frac{a+b}{2}\}$ |
| $x^{(a+b)/2}$ | means | $x^{\{\frac{a+b}{2}\}}$ |
| $x^{a/2-1}$ | means | $x^{\{\frac{a}{2}-1\}}$ |

$\underline{A}$ means that the symbol denotes a matrix of the indicated size (generally (p×p).

$\underline{A}'$ denotes the transpose of $\underline{A}$

i.e. $(\underline{A}')_{ij} = \underline{A}_{ji}$.

No distinction has been made between matrices and vectors since the context always makes clear the size of the matrix in question.

# CHAPTER I

## INTRODUCTION

As a subject for mathematical study, probability theory first came under notice because of its possible application to gambling investments. During a period when both mathematics and gambling were flourishing it was by no means unnatural that those primarily interested in the latter should attempt to persuade those primarily interested in the former to try to predict favourable odds to lay in connection with various games.

The entire framework of modern statistical theory rests on probability theory and a great deal of knowledge has been amassed relating to the correct design and analysis of experiments. However the prime purpose of practical experiments lies in the conclusions which can be drawn from their results. It is then somewhat surprising that the problems involved in forming a theory of statistical inference in particular and scientific inference in general, have received only desultory attention until recent times for in some way these problems may be said to have given rise to the whole subject of statistics.

That a theory of inference was necessary was clear to many people and there is general agreement that the first serious formal attempt at setting up a system was made by Thomas Bayes in the eighteenth century. The fundamental element of Bayes original attempt and the subsequent

Bayesian theories was the postulated existence of an *a priori* distribution, completely known, for the parameters involved in the problem. It is well known that Bayes himself doubted whether this postulate was either axiomatic or obvious and many subsequent writers, notably George Boole and Venn, criticised the eager adoption and generalisation of this postulate by many mathematicians including the Marquis de Laplace who composed a large treatise on probability among his mathematical works.

There is wide agreement that it would be desirable to establish a theory of inference which does not depend on any such assumption and furthermore one which depends only on the experimental data available for the problem under consideration. After the tremendous progress made in other branches of statistics early in this century, this question naturally recurred and almost simultaneously two attempts were begun to produce such a theory, one by Professor J. Neyman in Poland and the other by Professor R.A. Fisher in England. Opinions differ considerably as to how far the former, the Confidence Interval Theory, or the latter, the Fiducial Theory, succeed in their attempts; it is also to be noted that the extensions of Bayes' original theory have many supporters who do not see the necessity for the efforts expendend by the proponents of either of the other theories.

It is proposed to outline the development of the theory developed by Fisher and to show that it is not extendable simply and uniquely to deal with problems involving several variates and several parameters. In an attempt to carry out any extensions it becomes necessary to examine carefully the assumptions underlying the theory and to modify the requirements relating to the complete description of the problem and data. Modified forms of these assumptions are given and examined with relation to specific multivariate problems.

Outline of the Development of Fiducial Probability.

After successfully attacking many of the fundamental distribution problems occurring in statistical analysis as it then was, and outlining some of the problems to be solved in the course of setting statistical theory on a firm basis, Fisher in 1930 began his work on statistical inference. Like many before him he felt dis-satisfied with the Bayesian or Inverse Probability approach which at the time was restricted almost entirely to supposing a uniform distribution of parameter under a Postulate of Ignorance. However he felt this did not mean that probability theory could not be used in the formation of inferences:

> "There are, however, certain cases in which statements in terms of probability can be made with respect to the parameters of the population. One illustration may be given before considering in what ways its logical content differs from the corresponding statement of a probability inferred from known *a priori* probabilities. In many cases the random sampling distribution of a statistic, T, calculable directly from the observations, is expressible solely in terms of a single parameter, of which T is the estimate found by the method of maximum likelihood. If T is a statistic of continuous variation, and P the probability that T should be less than any specified value, we have then a relationship of the form
>
> $$P = F(T, \theta) .$$
>
> If we now give to P any particular value such as 0.95, we have a relationship between the statistic T and the parameter θ, such that T is the 95 per cent value corresponding to a given

θ, and this relationship implies the perfectly objective fact that in 5 per cent of samples T will exceed the 95 per cent value corresponding to the actual value of θ in the population from which it is drawn. To any value of T there will moreover be usually a particular value of θ to which it bears this relationship; we may call this the "fiducial 5 per cent value of θ" corresponding to a given T. If, as usually if not always happens, T increases with θ for all possible values, we may express the relationship by saying that the true value of θ will be less than the fiducial 5 per cent value corresponding to the observed value of T in exactly 5 trials in 100. By constructing a table of corresponding values, we may know as soon as T is calculated, what is the fiducial 5 per cent value of θ, and that the true value of θ will be less than this value in just 5 per cent of trials. This then is a definite probability statement about the unknown parameter θ, which is true _in the absence_ of any assumptions as to its _a priori_ distribution".

Proc. Camb. Phil. Soc. Vol. 26 1930 pp. 530.

Originally _irrespective_ appeared instead of _in the absence_, this latter being a correction due to Fisher himself.

Although this is by no means a formal statement of a basis for a theory, certain points made in the above paragraph are to be noted. Firstly there is the isolation of the statistic T and the parameter θ so that other parameters do not appear; then there is the uniqueness of the correspondence between T and θ i.e. for any pair (P,θ) there is one and only one T and for any pair (P,T) there is one and only one θ; finally there are the monotonicity and inversive properties of P as a function of T and of θ e.g. the original relationship

$$P = F(T,\theta)$$

from which $Pr(T>T_0)$ is calculable, yields $Pr(\theta<\theta_0)$ rather than $Pr(\theta>\theta_0)$.

While comparing methods of estimating sampling variance from normal data, Fisher was struck by a property of the mean square error, viz. if the observations are supposed to have been corrected for their mean, then their distribution conditional on the mean square error is independent of the population parameters, in particular of the population variance. That is, the distribution of the observations can be written

$$dF(x_1....xn;\xi,\sigma^2) = dF_1(\bar{x};\xi,\sigma^2)dF_2(s^2;\sigma^2)dF_3(\theta_1...,\theta_{n-2})$$

where $dF_3$ is independent of ξ and $\sigma^2$. Subsequently Fisher used this property in a generalised form to define the concept of Sufficiency. Suppose a sample $x_1...x_n$ of independent observations from $dF(x;\theta)$ is available; a function $x = x(x_1...x_n)$ is said to be sufficient for θ if the distribution of the observations, conditional on the calculated value of x, does not depend on θ, i.e.

$$\frac{d}{d\theta}[dF(x_1...x_n|x)] \equiv 0.$$

In this sense x can be said to contain all the information about θ which was contained in $x_1...x_n$.

Initially this property was of interest from the point of view of condensing data so that a few quantities could provide an adequate description of the original data. However it was later realised that such a property would be very desirable in any quantity used for drawing inferences from actual experiments.

If T is sufficient for θ and there is only one parameter then

$dF(T,\theta)$ contains all the information relevant to $\theta$ in the sample and the 'isolation' of T and $\theta$ is a natural consequence of the sufficiency. However if there are other parameters as well as $\theta$ then the criterion of sufficiency may not be strong enough to naturally separate T and $\theta$; in fact the problems involving several parameters produce many difficult questions.

The properties of uniqueness and monotonicity are closely linked and are essential for the (algebraic) inversion of the relation:

$$P = F(T,\theta).$$

But to be able to derive a distribution of $\theta$ from the inversion of such relations, a great deal more is required of $F(T,\theta)$. For example there must be values $\theta_M$ and $\theta_m$ in the permissible range of $\theta$ such that

$$F(T,\theta_M) = 1$$

$$F(T,\theta_m) = 0$$

for all possible values of T; also for any T

$$\frac{\partial}{\partial\theta} F(T,\theta)$$

must not change sign as $\theta$ varies. In short F must possess the same properties with respect to $\theta$ as it does with respect to T. Thus for inferences of the type proposed by Fisher in the case of one parameter and one variable, symmetry between the observations and parameter must exist; if sufficient estimation is also required then the results of Koopman and Darmois impose further restrictions on the form of the density function so that the number of cases which can be handled is not large but involves the commonly occurring distributions such as the negative exponential, normal and the Poisson.

This first paper aroused considerable comment, much of it unfavourable, and it was not until some time later that a more formal attempt was made to set up a theory. In 1935 Fisher put forward the solution to four inference problems and began at least one major controversy. The problems all related to normal theory, were:

(i) Given a sample of observations what can be said about a possible future observation?

(ii) Given a sample of observations what can be inferred about $\xi$ and $\sigma$?

(iii) The so-called Behrens-Fisher problem;

(iv) What inferences are available from an analysis of variance under certain rather special conditions?

While the fourth problem is not of very much interest, the Behrens-Fisher problem has been the subject of a very large number of papers in the last thirty years and is in fact much more complex than it at first appears.

Problem (i) was solved using the Student t-distribution in the usual manner and problem (ii) used a limiting process based upon a generalisation of the method used in problem (i). However this provided a simultaneous distribution of $\xi$ and $\sigma$ which, unlike the simultaneous sampling density of the sufficient statistics $\bar{x}$ and s, did not factorise into two independent distributions one for $\xi$ and one for $\sigma$; how then were the individual fiducial distributions to be defined? Fisher took the corresponding marginal distribution as the appropriate quantity and so obtained the usual t and $\chi$ distributions for $\xi$ and $\sigma$. The question of the definition of individual distributions in the case of several parameters leads to many confusing results as will be shown and great care is necessary when dealing with these situations.

The need for sufficient statistics was emphasised for the first time in this paper and a tentative general form of the fiducial argument proposed when a set of jointly sufficient statistics was known to exist. The corresponding marginal distribution was suggested as a natural definition of the fiducial distribution of a single parameter in multi-parameter cases.

For many years after this the positive work in fiducial theory consisted of the examination of various special distributions and attempts to describe the fundamentals of the logical basis of the theory. The critical and destructive work came from many quarters. In the first instance the supporters of *a priori* notions resented the strong criticism they received and there were many bitter exchanges between Fisher and the Bayesians in which it was often hard to distinguish between scientific and personal disagreement. As has often happened in the past, an independent attempt to solve these fundamental scientific questions had been begun by another research worker. Professor Jerzy Neyman had commenced his work on confidence theory in Poland and was unaware of Fisher's ideas until he came to England. After an initial confusion of identity of the two theories, again and unfortunately personal feelings entered into the scientific dispute, this time to such an extent that there was almost zero communication between the two resulting schools of thought.

The cases examined relied almost entirely on normal theory and were little more than extensions of t and $\chi$. However from the logical point of view the basic assumptions underlying the theory were gradually formulated more explicitly. As already mentioned the concept of sufficiency was emphasised as being necessary because of the essential difference between deductive and inductive logic, for in the former a result proved using only some of the basic axioms is still valid and exact when the remaining axioms are taken into account but in the latter, for optimum inferences all the relevant information must be used. The concept of sufficiency provided a criterion for separating the relevant information. When a simultaneous distribution failed to factorise, the appropriate marginal distribution was to be taken as the fiducial distribution of the parameter in question. The distributions determined for the parameters involved sample characteristics which acted as "parameters" in these circumstances and the inferences made were to be referred to a population determined by these sample characteristics. For example the Student t-distribution of $\xi$ from $N(\xi,\sigma^2)$ was to be interpreted as referring to the possible parameter values which could be associated with a sample whose characteristics were $(n,\bar{x},s^2)$ i.e. imagine a large collection of normal samples of size n with mean $\bar{x}$ and variances $s^2$, then if the true population means for these samples were known and plotted out they would follow the distribution of $t_{n-1}$ where

$$t_{n-1} = \sqrt{n}(\bar{x} - \xi)/s.$$

Fisher also pointed out that there was a difference between problems of inference and problems of designing tests of significance although in many cases the quantities used were the same for each situation. In designing a test of significance it is not necessary to use sufficient statistics although it is desirable, but for inference purposes it is important to use all the information supplied by the sample.

This then was the state of affairs existing in 1954 when a Symposium on Interval Estimation was held by the Royal Statistical Society. At this Symposium the Creasy-Fieller problem appeared and in the subsequent discussion the following remarks were made by Fisher:

"We should have been given a normal bivariate sample, of which the statistics used before are now estimates of the variances and covariance of the errors of the co-ordinates of the estimated centre $(\bar{x},\bar{y})$, namely

$$P = S(x - \bar{x})^2/N(N-1)$$

$$Q = S(x - \bar{x})(y - \bar{y})/N(N-1)$$

$$R = S(y - \bar{y})^2/N(N-1),$$

then recognising, as a simple application of "Student's" test that, for any ratio $\lambda:\mu$,

$\lambda(\xi-\bar{x})+\mu(\eta-\bar{y})>t\sqrt{(\lambda^2P+2\lambda\mu Q+\mu^2R)}$

with the one-sided probability appropriate to t for N-1 degrees of freedom, it would seem that the appropriate fiducial probability of the pair $(\xi,\eta)$ lying in any elementary region must be

$$\frac{N-1}{2\pi\sqrt{(PR-Q^2)}} \qquad \frac{d\xi d\eta}{(1+r^2)^{(N+1)/2}}$$

where r is defined by

$$(PR-Q^2)r^2 = R(\xi-\bar{x})^2 - 2Q(\xi-\bar{x})(\eta-\bar{y}) = P(\eta-\bar{y})^2,$$

which may be easily generalised to d dimensions, and be integrated over any restricted region to give the fiducial probability to be associated with it".

With one exception to be noted this was the first attempt to deal with problems involving more than one variate and indeed these problems lead to some real difficulties for fiducial theory. But at the time the difficulty was that Fisher gave no indication as to how he arrived at the result and when hard pressed had to admit he could not give the necessary formal derivation. A derivation of this result was put forward by Cornish in 1961 but this proof was not endorsed by Fisher except when a common variance could be associated with each variate.

The exception mentioned above is due to I.E. Segal. Treating the whole problem as one in pure mathematics, Segal took the known distribution of the vector of sample means $\underline{\bar{x}}$ and the estimate $\underline{S}$ of the variance-covariance matrix and changed variables to

$$\underline{A} = \underline{\Sigma}^{-1/2}(\underline{\bar{x}}-\underline{\xi}), \quad B = \underline{\Sigma}^{-1/2}\underline{S}\ \underline{\Sigma}^{-1/2}$$

where $\underline{\xi}$ and $\underline{\Sigma}$ are the population characteristics. The quantities $\underline{A}$ and $\underline{B}$ then have a distribution independent of $\underline{\xi}$ and $\underline{\Sigma}$ and are thus pivotal quantities. Formally changing variables from $(\underline{A},\underline{B})$ to $(\underline{\xi},\underline{\Sigma}^{-1/2})$ then yields a density for these two quantities which Segal called their fiducial distribution. This density as given is fairly complex but changing variables from $(\underline{\xi},\underline{\Sigma}^{-1/2})$ to $(\underline{\xi},\underline{\Sigma}^{-1})$ shows that the latter pair have a distribution obtainable from that of $\underline{\bar{x}}$ and $\underline{S}$ by interchanging $\underline{\bar{x}}$ and $\underline{\xi}$ in the normal factor and $\underline{S}$ and $\underline{\Sigma}^{-1}$ in the Wishart density.

For some reason this result passed almost unnoticed and later references to the paper containing it were more concerned with the author's comments on pivotal quantities. How this result compares with the above conjecture will be shown later in Section V.

1956 saw the publication of Statistical Methods and Scientific Inference the final section of which contained yet another attempt to derive the fiducial distribution of the parameters of the bivariate normal distribution. This attempt however was incomplete for part of the final density function was missing viz.:

$$-\frac{\partial}{\partial\rho}\int_{-1}^{R} dF(r,\rho)dr.$$

It will be shown that the form of this function leads to still further complications for fiducial theory and that this attempt by Fisher appears to be composed of pieces from two situations.

So much material of a conflicting and apparently contradictory nature has arisen from the consideration of the bi- and multi-variate normal distributions that a complete revision of many of the fundamental ideas has been necessary in order to preserve the theory of fiducial inference from a collapse which some writers feel is still inevitable.

# CHAPTER II

## UNIVARIATE DISTRIBUTIONS

As an example of the fiducial argument and the logic involved in it, consider the simple univariate uniparameter distribution with frequency function given by

$$f(x,\theta) = \theta e^{-\theta x}.$$

It is not difficult to show that

$$T = {}^{n}/\sum x_i = {}^{n}/X$$

is the maximum likelihood estimate of $\theta$ from a sample of n independent observations. Since $2\theta x$ is evidently distributed as $\chi^2_2$, $2\theta X$ is distributed as $\chi^2_{2n}$ by the reproductive property of the parent density and it is then easily shown that T is a sufficient estimate of $\theta$ so that as far as questions of inference about $\theta$ are concerned the distribution of T contains all the relevant information.

For any particular given value of $\theta$, there is a well defined distribution of X and as $\theta$ varies from 0 to $\infty$ the graphs of these distributions generate a surface as in figure 1. Note that the sections parallel to OX are normalised i.e. for any given $\theta$

$$\int_0^\infty f(X,\theta)dX \equiv 1$$

but sections parallel to O$\theta$ are not:

$$\int_0^\infty f(X,\theta)d\theta \not\equiv \text{constant}$$

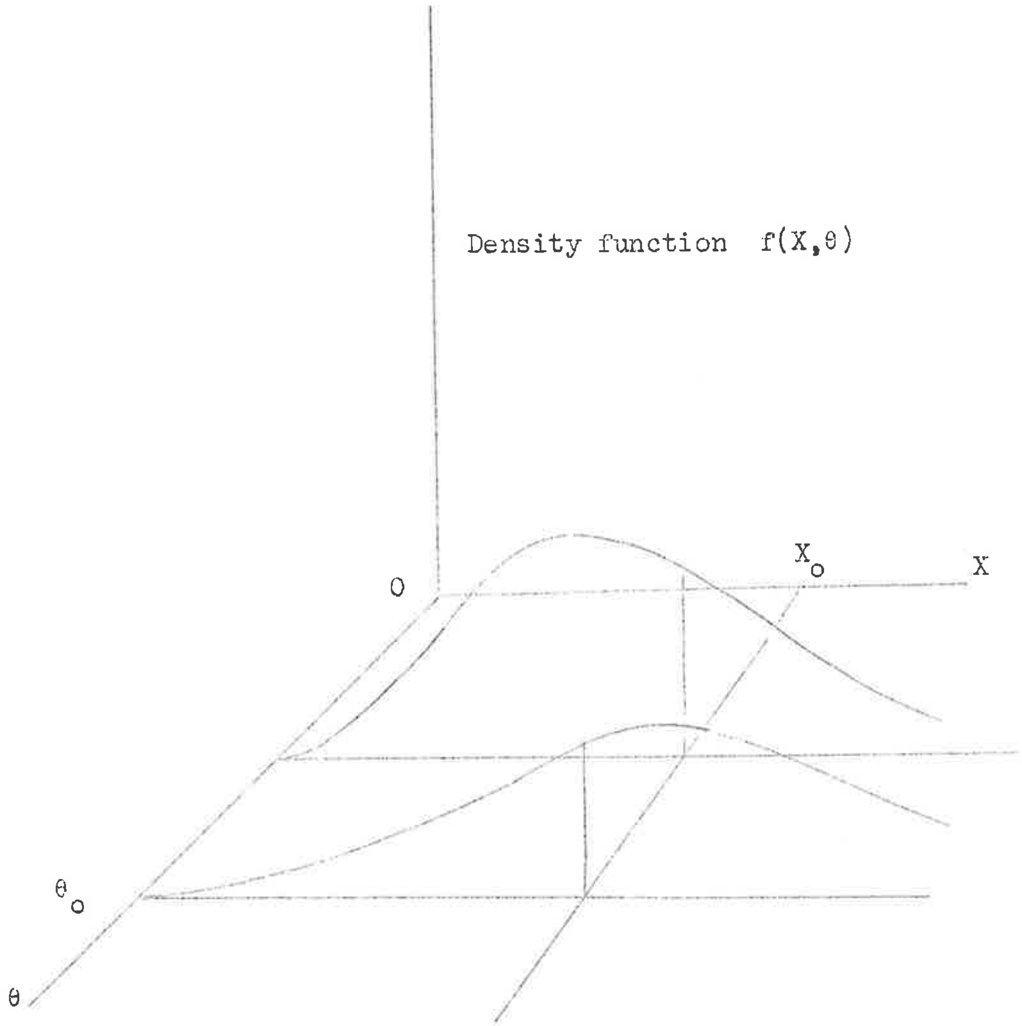


FIGURE 1.

$$f(X,\theta) = \theta^n X^{n-1} e^{-\theta X} / \Gamma(n).$$

Given $\theta$, $\chi^2 = 2\theta X$ is distributed as $\chi^2_{2n}$ so that

$$Pr(\chi^2>\alpha) = P(\alpha,n)$$

can be found exactly and depends only on $\alpha$ and n, furthermore the inequality $\chi^2>\alpha$

yields $\theta>\alpha/2X$

and the probability statement yields

$$Pr(\theta>\alpha/2X) = P(\alpha,n). \qquad (*)$$

It is this last transitional step which usually causes the opponents of fiducial theory to object to the reasoning. In the first place it is a probability statement and as such only has meaning when the population to which it refers is specified. Clearly the population to which (*) refers is not the population from which the observations $x_i$ were drawn. It is a population of samples which resemble the actual data in all essential respects, in this case the values of n and $X = \sum x_i$. Imagine the set of all possible samples derivable from any negative exponential distribution and take the subset of samples of size n with $\sum x_i = X$ as observed, then the value of $\theta$ corresponding to any sample can be anywhere in the range $(o,\infty)$ and some values more likely than others. It is not obvious that the frequency distribution of $\theta$ is described by (*) but in the absence of any _a priori_ information and assuming the sample to be a random one of independent observations, since X is sufficient for $\theta$ and thus contains all the information relative to $\theta$, it is the inference of fiducial theory that (*) does describe the distribution of $\theta$ as far as the observed data allow.

The fiducial distribution so defined then gives inferences about the unknown parameter $\theta$ against the background of the observed sample. Essential elements in its construction are the existence of a sufficient statistic for $\theta$ and the existence of a quantity $\chi^2$ which is a function of $\theta$

and the sufficient statistic with a completely specified distribution. To these must be added the requirements that the sample is a random sample of independent observations and that there is no other information given relevant to $\theta$, such as an a priori distribution. In forming an inference of any kind it is important and necessary to take into account all the information available and to outline the method used to form the inference. The above process supplies a unique set of probability statements for each sample and no two sets of these contradict each other for they refer to different populations if the sample characteristics differ.

As mentioned earlier, the sections parallel to $O\theta$ are not normalised since

$$\int_0^\infty f(X,\theta) \quad = {}^n/X^2$$

but given the a priori distribution of $\theta$ as $f_1(\theta)$, the a posteriori distribution is easily found as

$$f(X,\theta)f_1(\theta) \div \int_0^\infty f(X,\theta)f_1(\theta)d\theta \quad ;$$

simple normalisation of the sections parallel to $O\theta$ amounts to presupposing a uniform a priori density on $(0,\infty)$ for $\theta$.

As far as univariate uniparameter distributions are concerned, the existence of a sufficient statistic restricts the form of the frequency function $f(x,\theta)$ to be exponential for the Koopman-Darmois results imply that there exist real single-valued analytic functions of $\theta$ and $x$; $\theta_1$, $\theta_2$, $X_1$, $X_2$ such that

$$f(x,\theta) = \exp\left[\theta_1 X_1+\theta_2+X_2\right].$$

For a sample of n independent observations,

$$\mathcal{L} = \exp \sum[\theta_1 X_1(x_i) + \theta_2 + X_2(x_i)]$$

and $$L = \theta_1 \sum X_1(x_i) + n\theta_2 + \sum X_2(x_i),$$

thus $$\partial L/\partial\theta = \theta_1' \sum X_1(x_i) + n\theta_2' = 0$$

implies that there is a function $X_1 = X_1(x_i)$ whose sample mean is a sufficient statistic for $\theta$. Suppose now that its distribution has frequency function

$$\phi(y,\theta) \quad -\infty<y, \ \theta<\infty$$

where as usual $\phi \equiv 0$ outside the admissible actual ranges of $y$ and $\theta$. Then

$$\Phi(Y,\theta) = \int_{-\infty}^{Y} \phi(y,\theta)dy$$

is uniformly distributed on $[0,1]$ for all $\theta$ so that a quantity with a known distribution always exists in this case. To see whether corresponding fiducial statements can be derived, it is necessary to consider the properties of $\Phi(Y,\theta)$ as a function of $\theta$.

In the remarks of Fisher already quoted from the first paper on fiducial inference, it was supposed that $\Phi(Y,\theta)$ was monotonic (decreasing) in $\theta$ for all $Y$ and even if this is the case, so that

$$Pr(y<Y) = \Phi(Y,\theta)$$

leads to $\theta>\theta_1$, it is still necessary for $\Phi$ to be a distribution function with respect to $\theta$ if probability statements are to be available. If $t$ is the lower boundary of admissible $\theta$ values and $T$ the upper boundary then it is necessary that

$$\left.\begin{aligned} \Phi(Y,t) &= {1 \atop 0} \\ \Phi(Y,T) &= {0 \atop 1} \end{aligned}\right\} \qquad (1)$$

depending on whether $\Phi$ is in fact decreasing or increasing in $\theta$. Thus $Y$ and $\theta$ are in a sense interchangeable since $\Phi$ has the same properties with respect to each variable. This recipricocity has a habit of recurring in fiducial theory; if it occurs here then the appropriate fiducial frequency function is found from

$$\frac{d}{d\theta}\,(\text{Cum.d.f. of } \theta) = \frac{d}{d\theta}\,(1 - P(\theta > \theta_1))$$
$$= \frac{d}{d\theta}\,(1 - P(y < Y))$$
$$= -\frac{d}{d\theta}\,\Phi(Y, \theta)$$

a formula which Fisher used widely as defining the fiducial distribution in situations similar to those described above and in others where all the restrictions mentioned did not apply. However the monotonicity properties needed are not evident, all that is known is that the frequency element is expressible in the form

$$\exp\,[\theta_1 X_1 + \theta_2 + X_2]$$

which will still be valid if variables are changed to $X_1$ and the differential element is absorbed into $X_2$. Hence symmetry between X and θ exists but while $\theta_1$ and $\theta_2$ are single-valued and analytic, they are not necessarily monotone so that this essential requirement may be missing. Even if it is present, the requirements (1) place further restrictions on the form of function with which the argument in this simple form can deal.

Probably the greatest restriction of all is the requirement of sufficiency itself. Fortunately many of the commonly employed distributions do possess sufficient statistics but if these do not exist then it is necessary to look for ancillary statistics and to consider the distribution of the appropriate estimate conditional on the values of the ancillaries and to examine the behaviour of this function. The subject of ancillary statistics is not very well charted and little is known about the corresponding fiducial problems.

The Univariate Normal.

The concept of sufficiency and its generalisation to joint sufficiency appears to play an important role in the problems associated with multiparameter and multivariate distributions. Some of these points can be demonstrated in connection with the univariate normal distribution. In the usual notation the appropriate parameter estimates are

$$\bar{x} = \frac{1}{n}\sum x_i, \quad s^2 = \frac{1}{n-1}\sum (x_i-\bar{x})^2$$

which form a jointly sufficient set of statistics together with the ancillary quantity n, the sample size. While $\bar{x}$ is individually sufficient for $\xi$, $s^2$ is not sufficient for $\sigma^2$ unless the distribution of the observations is considered conditional on $\bar{x}$, for variance $(\bar{x}) = \frac{\sigma^2}{n}$. The inequalities

$$[-\infty \leqq \bar{x} \leqq X, \quad 0 \leqq s \leqq S]$$

lead to

$$[-\infty \leqq \sqrt{n}\frac{(\bar{x}-\xi)}{\sigma} \leqq \sqrt{n}\frac{(X-\xi)}{\sigma}, \quad 0 \leqq \sqrt{n-1}.\frac{s}{\sigma} \leqq \sqrt{n-1}\frac{S}{\sigma}]$$

or

$$[-\infty \leqq u \leqq U, \quad 0 \leqq v \leqq V]$$

see figures 2 and 3. Similarly

$$[-\infty \leqq \sqrt{n}\frac{(X-\xi)}{\sigma} \leqq U, \quad 0 \leqq \sqrt{n-1}.\frac{S}{\sigma} \leqq V]$$

when considered as functions of $\xi$ and $\sigma$ map figure 3 into figure 4 so that the simple extension

$$-\frac{\partial^2}{\partial\xi\partial\sigma}\Pr(\bar{x} \leqq X, \ s \leqq S)$$

will not lead to the correct density for $\xi$ and $\sigma$ since the above mappings do not lead by a series of probability inequalities from

$$\Pr(\bar{x} \leqq X, \ s \leqq S)$$

to

$$\Pr(\xi \geqq \Xi, \ \sigma \geqq \Sigma)$$

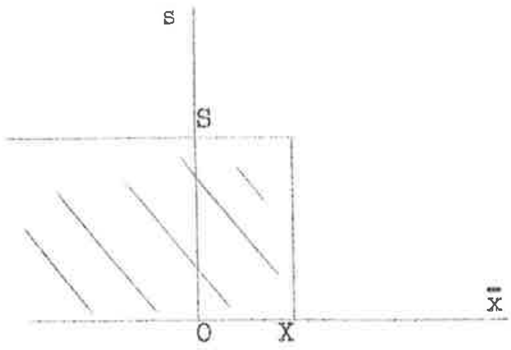


FIGURE 2.

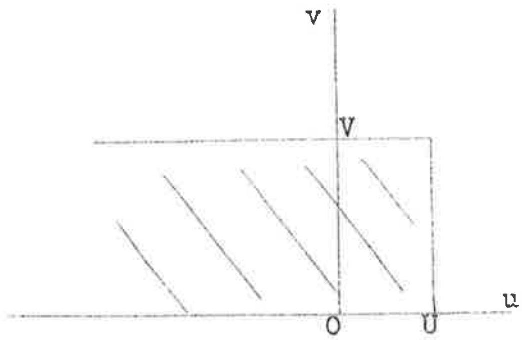


FIGURE 3.

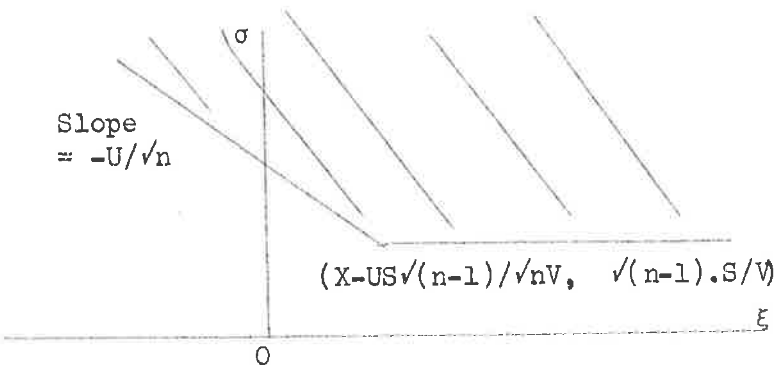


FIGURE 4.

or more particularly

$Pr(\bar{x} \geq X, s \geq S)$ does not lead to $Pr(\xi \leq \Xi, \sigma \leq \Sigma)$.

It is necessary for the pivotal quantities u and v above to be defined over a slice-shaped region similar to figure 4 in order to obtain

$$Pr(\xi \leq \Xi, \sigma \leq \Sigma)$$

upon which $\partial^2/\partial\xi\partial\sigma$ may be used to obtain the required density. This region in turn requires one of the same shape to be considered for $\bar{x}$ and s. This extra consideration is brought about by the fact that u is a function of both $\xi$ and $\sigma$ while v is a function of $\sigma$ above; it can be removed by using $w = u/v$ and v instead but then it reappears because now w is a function of both $\bar{x}$ and s. Such a complication can only be avoided if the parameters in question have a joint set of sufficient statistics each of which is also individually sufficient so that the whole problem reduces to a series of univariate uniparameter cases such as already considered.

Having obtained the simultaneous distribution of $\xi$ and $\sigma$ in the form

$$\frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-n(\xi-\bar{x})^2/2\sigma^2} \, d\xi \, \frac{(n-1)^{(n-1)/2} s^{n-1}}{\Gamma(n-1)/2 . 2^{(n-3)/2} \sigma^n} e^{-(n-1)s^2/2\sigma^2} d\sigma \qquad (2)$$

and having noted that it does not factorise into two pieces, one for $\xi$ and one for $\sigma$, the problem of the definition of the fiducial distribution of $\xi$ and $\sigma$ individually arises. The natural quantities to consider are the corresponding marginal distributions which are described by

$$t_{n-1} = \sqrt{n}(\xi-\bar{x})/s$$

$$\chi_{n-1} = s\sqrt{n-1}/\sigma$$

and to use these as the fiducial distributions, noting that both of these quantities are extensions of the usual significance tests associated with a normal sample.

It is possible to use the ideas of marginal and conditional distributions to build up expression (2). Remembering that when the observations are considered to be conditional on $\bar{x}$, s is sufficient for $\sigma$ and $\sqrt{(n-1)}s/\sigma$ is distributed as $\chi_{n-1}$, the density of $\sigma$ can be directly obtained. For given $\sigma$, $\sqrt{n}(\bar{x}-\xi)/\sigma$ is distributed as N(0,1) and moreover $\bar{x}$ is the estimate of $\xi$ even when $\sigma$ is supposed known. The normal density of $\xi$ given $\sigma$ is then obtained and multiplication of the two factors gives (2). Similarly one can start with

$$t_{n-1} = \sqrt{n}(\bar{x}-\xi)/s$$

to obtain the marginal density for $\xi$; then since for given $\xi$

$$u^2 = \frac{1}{n}\sum(x_i-\xi)^2$$

is the appropriate estimate of $\sigma^2$ and is essentially distributed as $\chi^2_n$, inversion leads to the conditional distribution of $\sigma$ given $\xi$. Multiplication of the two leads to (2) once more. However it is important and necessary to use the fact that when $\xi$ is known, $s^2$ is no longer the best estimate of $\sigma^2$ and must be replaced by the sufficient estimate $u^2$. This is an example of the necessity of always using all the available information in forming inferences. The use of the marginal distribution as the fiducial distribution is a matter of definition and its identity with the extensions of the usual significance tests appears to be a consequence of the form of the normal and $\chi^2$ distributions and of the statistic-parameter symmetry of the pivotal quantities.

In a similar manner to the above, the fiducial distributions of $\alpha$ and $\beta$, the parameters of a simple linear regression model, can be derived as can fiducial limits for ratios of variances determined from analyses of variance.

# CHAPTER III

## CONDITIONS TO BE EXAMINED FOR THE FIDUCIAL ARGUMENT

Since the aim of fiducial theory is to set up a system in which the uncertainty existing about a parameter or a set of parameters is expressible in terms of probability, and in particular in terms of a well defined distribution, it is natural to desire that the concepts of marginal and conditional distribution be meaningful and useful. It is also desirable and essential to determine the mechanisms available for deriving fiducial distributions and in this connection the properties of the densities of sampling statistics is important.

If it turns out that many of these simple distributions and the corresponding significance tests can be directly used then application of the theory of fiducial inference will be easy and wide-spread. For the last twelve months or so it has appeared that any attempt to carry the analogy between sampling and fiducial distributions into multivariate problems leads to inconsistent and confusing results which have induced many people to reject the theory entirely.

It is proposed to show that although such inconsistencies do appear relative to a background based on a Sampling Distribution-Fiducial Distribution analogy, the fault lies in the assumption of this analogy rather than in the concept of a fiducial distribution. For the purpose of providing concrete examples, the bivariate normal distribution will be extensively used.

The background for the results is provided by the three assumptions given below:

A. Joint fiducial distributions may be constructed from marginal and conditional densities which may be derived by inversion of the appropriate sampling densities.

B. If this construction can be performed in several different ways each will lead to the same result.

C. The formula $\pm\partial F(T,\theta)/\partial\theta$, where $F(T,\theta)$ is the cumulative distribution function of T, may be used to carry out the inversions subject to restrictions of monotonicity and differentiability on F considered as a function of $\theta$, and provided

$$F(T,\infty) = \begin{matrix}1\\0\end{matrix}\Big\} \qquad -\infty<T<\infty$$

$$F(T,-\infty) = \begin{matrix}0\\1\end{matrix}\Big\}$$

depending on whether F is an increasing or decreasing function of $\theta$.

The first part of A is necessary if the theory of fiducial inference is to exist for the idea of a fiducial distribution entails the concepts of marginal and conditional distributions and if, as Fisher maintained from the outset, fiducial probability obeys the same rules as ordinary probability, these concepts must play their parts as usual. Since the method of dealing with simple distributions is well known, the second part of A may provide one way of using this knowledge to simplify the construction of joint densities.

Condition B is necessary to maintain internal consistency in the theory for the same set of sufficient statistics for the same problem must lead to the same set of inferences.

Condition C is designed to allow for the evaluation of fiducial frequency functions when simple pivotal quantities do not exist. The use of the formula $\pm\partial F(T,\theta)/\partial\theta$ has already been mentioned and discussed.

# CHAPTER IV

## THE BIVARIATE NORMAL DISTRIBUTION

The bivariate normal distribution provides an example on which these conditions can be tested and the varying results examined and interpreted. The likelihood function for a sample of n independent observations is essentially

$$L = (\sigma_1\sigma_2\sqrt{1-\rho^2})^{-n}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{S(x_1-\xi_1)^2}{\sigma_1^2} - 2\rho\frac{S(x_1-\xi_1)(x_2-\xi_2)}{\sigma_1\sigma_2} + \frac{S(x_2-\xi_2)^2}{\sigma_2^2}]\}$$

whence the five estimation equations can be formed and solved. As is the case with the univariate normal, the estimation of the location parameters is independent of whether the scale parameters are known or not but the estimation of the latter depends on the knowledge available about $\xi_1$ and $\xi_2$; if they are known, the estimates are

$$u_{11} = \frac{1}{n} S(x_1-\xi_1)^2 , \quad u_{22} = \frac{1}{n} S(x_2-\xi_2)^2 , \quad u_{12} = S(x_1-\xi_1)(x_2-\xi_2)/n\sqrt{u_{11}}\sqrt{u_{22}};$$

if they are not known the usual estimates

$$s_1^2, s_2^2, r$$

are obtained. It is also to be noted that $\bar{x}_1$, $\bar{x}_2$ are jointly sufficient for $\xi_1$ and $\xi_2$ while $s_1^2$, $s_2^2$, r are only sufficient if the distribution is considered conditional on the values of $\bar{x}_1$ and $\bar{x}_2$. Furthermore the five estimates are jointly sufficient for the five parameters yet no one of the estimates is individually sufficient for its parameter.

Some well-known facts about the joint distribution of the statistics may be recalled. $\bar{x}_1$ and $\bar{x}_2$ are distributed independently of $s_1$, $s_2$, r and each has a marginal normal distribution with the corresponding ξ and σ as parameters. Similarly $s_i^2$ is distributed as $\sigma_i^2 \chi^2_{n-1}/n-1$. The marginal density of r can be written as

$$dF(r,\rho) = \frac{n-2}{\pi} (1-\rho^2)^{(n-1)/2} (1-r^2)^{(n-4)/2} I_{n-1}(\rho r)\, dr \qquad (3)$$

where

$$I\mu(x) = \int_0^\infty (\text{Cosh } z - x)^{-\mu} dz, \qquad |x|<1.$$

In an attempt to invert the joint sampling distribution, Fisher made use of A as a first step. Since $\bar{x}_1$, $\bar{x}_2$ are the estimates of $\xi_1$ and $\xi_2$ even when $\sigma_1,\sigma_2,\rho$ are known and since, for given $\sigma_1,\sigma_2$ ,ρ, they have a bivariate normal density described by

$$N(\underline{\xi}, \underline{\Sigma}/n)$$

where

$$\underline{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} , \qquad \underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} ,$$

a simple probability inversion for the conditional distribution exists

i.e. $\Pr(\bar{x}_1-\xi_1)\sqrt{n}/\sigma_1>\alpha_1,\ (\bar{x}_2-\xi_2)\sqrt{n}/\sigma_2>\sigma_2|\sigma_1,\sigma_2,\rho)$

leads to $\Pr(\xi_1<\beta_1,\xi_2<\beta_2|\sigma_1,\sigma_2,\rho)$

or the cumulative density of $\xi_1$ and $\xi_2$ since conditions of sufficiency and monotonicity are clearly satisfied. From A, the portion required for the complete specification of the joint density is the marginal density of $\sigma_1,\sigma_2,\rho$ which can be obtained from

$dF(s_1,s_2,r : \sigma_1,\sigma_2,\rho)$ =

$$\frac{(n-1)^{n-1}s_1^{\,n-2}s_2^{\,n-2}(1-r^2)^{(n-4)/2}}{\pi\Gamma(n-2)\sigma_1^{\,n-1}\sigma_2^{\,n-1}(1-\rho^2)^{(n-1)/2}}\,.$$

$$\exp\{-\frac{(n-1)}{2(1-\rho^2)}\left[\frac{s_1^2}{\rho_1^2} - 2r\rho\frac{s_1 s_2}{\rho_1\rho_2} + \frac{s_2^2}{\rho_2^2}\right]\}\, ds_1 ds_2 dr.$$

Unfortunately it is not possible to find 3 simple pivotal quantities for this distribution while at the same time maintaining its symmetry and form because of the awkward quantity r.

Noting that $dF(r,\rho)$ is a function of r and $\rho$ only, Fisher proceeded to apply A and C to obtain

$$-\frac{\partial}{\partial\rho}\int_{-1}^{r} dF(t,\rho)$$

as the marginal density of $\rho$; since

$$dF(s_1,s_2|r)$$

is a symmetric function of $\frac{s_1}{\sigma_1}$, $\frac{s_2}{\sigma_2}$ and possesses, for given $r,\rho$, pivotal quantities

$$\frac{s_1}{\sigma_1}\sqrt{\frac{n-1}{1-\rho^2}}\,,\quad \frac{s_2}{\sigma_2}\sqrt{\frac{n-1}{1-\rho^2}}$$

a simple probability statement inversion analogous to that used for $\bar{x}_1$ and $\bar{x}_2$ yielded the conditional density of $\sigma_1,\sigma_2$ given $\rho$. This when multiplied by the marginal frequency element of $\rho$ as determined above gave a density for $\sigma_1,\sigma_2,\rho$ which completed the joint distribution element. Because of the awkward nature of $dF(r,\rho)$ the result remained formal since

$$-\frac{\partial}{\partial\rho}\int_{-1}^{r} dF(t,\rho)$$

could not be found. Thus a comparison with the other results given by Segal and conjectured by Fisher was not possible. The missing function can be found in the following somewhat round-about way.

Although $dF(s_1,s_2,r;\sigma_1,\sigma_2,\rho)$ does not possess simple pivotals preserving symmetry and form, a change of variables to

$$x_1 = \frac{s_1}{\sigma_1}\sqrt{n-1},\quad x_2 = \frac{s_2\sqrt{n-1}\,\sqrt{1-r^2}}{\sigma_2\sqrt{1-\rho^2}},\quad x_3 = \frac{\sqrt{n-1}}{\sqrt{1-\rho^2}}\left(r\frac{s_2}{\sigma_1} - \rho\frac{s_1}{\sigma_2}\right)$$

provides a complete factorisation into $\chi_{n-1}$, $\chi_{n-2}$ and $N(0,1)$. Moreover

$$x_3 = \frac{r}{\sqrt{1-r^2}}x_2 - \frac{\rho}{\sqrt{1-\rho^2}}x_1.$$

Starting from the joint distribution of an independent z, $\chi_{n-1}$, $\chi_{n-2}$ and transforming to r, $\chi_{n-1}$, $\chi_{n-2}$ where r is defined through the above relation, and integrating out for $\chi_{n-1}$, $\chi_{n-2}$ yields the sampling distribution of r. Similarly changing variables to $\rho$, $\chi_{n-1}$, $\chi_{n-2}$ and integrating yields a density for $\rho$ of the form

$$\frac{n-2}{\pi}(1-\rho^2)^{(n-3)/2}(1-r^2)^{(n-2)/2}[I_{n-2}(\rho r) + \rho r I_{n-1}(\rho r)]. \qquad (4)$$

Differentiation of (4) and $-\frac{\partial}{\partial\rho}\int_{-1}^{r} dF(t,\rho)$ with respect to r and equating leads to the relation

$$r[(n-2)I_{n-2}(\rho r) + (2n-3)\rho r I_{n-1}(\rho r) - (n-1)(1-\rho^2r^2)I_n(\rho r)] \equiv 0$$

identically in $\rho$ and r, so that either $r \equiv o$ or the expression in brackets, must vanish. If $I_\mu(x)$ is evaluated by expanding the integrand in a power series of $^1/\text{Cosh } z$ and integrating termwise, then it is found that

$$I_\mu(x) = \sum_{\nu=o}^{\infty} \frac{\Gamma(\mu+\nu)/2\,\Gamma(\mu+\nu)/2}{\Gamma(\mu)\Gamma(\nu+1)} 2^{\mu+\nu-2} x^\nu \quad , \quad |x|<1$$

and it is then a simple matter to verify that $I\mu$ does satisfy the above Legendre type of recurrence relation. Thus (4) represents the missing factor and Fisher's complete result may be written:

$$\frac{n-2}{\pi}(1-\rho^2)^{(n-3)/2}(1-r^2)^{(n-2)/2}[I_{n-2}(\rho r) + \rho r\, I_{n-1}(\rho r)]$$

$$\times \frac{(n-1)^{n-1}(s_1 s_2)^{n-1}}{(1-\rho^2)^{n-1}(\sigma_1\sigma_2)^n\Gamma(n-1)} \exp\{-\frac{n-1}{2(1-\rho^2)}[\frac{s_1^2}{\sigma_1^2} - 2r\rho\frac{s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}]\} \div I_{n-1}(\rho r)$$

$$\times \frac{n}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-\frac{n}{2(1-\rho^2)}[\frac{(\xi_1-\bar{x}_1)^2}{\sigma_1^2} - 2\rho\frac{(\xi_1-\bar{x}_1)(\xi_2-\bar{x}_2)}{\sigma_1\sigma_2} + \frac{(\xi_2-\bar{x}_2)^2}{\sigma_2^2}]\} \quad . \ (5)$$

This is evidently a very cumbersome expression which is analytically intractable except in the trivial case when r vanishes. Then the marginal density of $\sigma_i$ can be expressed as

$$k(n)\frac{s_i^{n-1}}{\sigma_i^n}\int_0^1 \frac{\exp\{-(n-1)s_i^2/2\sigma_i^2(1-\rho^2)\}}{1-\rho^2}\, d\rho$$

which does not simplify to the $\chi^2$density obtainable by inverting the distribution of $s_i^2$ as would be required by condition A, for equating the two implies that

$$\int_0^1 \frac{\exp\{-(n-1)s_i^2\rho^2/2\sigma_i^2(1-\rho^2)\}}{1-\rho^2}\, d\rho$$

is a function of n alone which is evidently not so.

Similarly writing

$$t_i = \sqrt{n}(\xi_i - \bar{x}_i)/\sqrt{(n-1)}s_i \quad , \quad i = 1,2$$

their joint density when r = 0 is

$$\frac{(n)}{(1+t_1^2)^{n/2}(1+t_2^2)^{n/2}} \sum_{\mu=0}^{\infty} \frac{\Gamma(\mu+n/2)\Gamma(\mu+n/2)}{\Gamma(\mu+1)\Gamma(\mu+(n+1))/2)} \left(\frac{t_1^2}{1+t_1^2}\right)^{\mu} \cdot \left(\frac{t_2^2}{1+t_2^2}\right)^{\mu} \quad (6)$$

which does not have simple Student distributions as marginals for $t_1$ and $t_2$, and in any case differs from the function put forward by Fisher at the R.S.S. symposium. A proof that the distribution (5) leads to results differing from the $\chi^2$ and t results even when $r \neq 0$ will be given in the next section.

# CHAPTER V

## THE MULTIVARIATE NORMAL DISTRIBUTION

### Solutions of Segal, Fisher and Cornish

It is convenient to compare these solutions by considering the multivariate normal distribution in which the p×1 vector $\bar{\underline{x}}$ is supposed to be distributed as $N(\underline{\xi},\underline{\Sigma})$ with mean vector $\underline{\xi}$ and p×p variance-covariance matrix $\underline{\Sigma}$, in the density

$$(2\pi)^{-p/2}|\underline{\Sigma}|^{-1/2} \text{ etr } [-\underline{\Sigma}^{-1}(\underline{x}-\underline{\xi})(\underline{x}-\underline{\xi})'/2] \, d\underline{x}$$

where

$$|\underline{A}| = \text{determinant of } \underline{A}, \text{ etr } [\underline{A}] = \exp [\text{trace } \underline{A}], d\underline{A} = \prod_{i,j=1}^{p} dA_{ij}.$$

The usual terminology of $\bar{\underline{x}}$ for sample mean vector and $\underline{S}$ for sample variance-covariance matrix will be used. These quantities have joint density, for a sample of size n:

$$\left(\frac{n}{2\pi}\right)^{p/2}|\underline{\Sigma}|^{-1/2} \text{ etr } [-n\underline{\Sigma}^{-1} (\bar{\underline{x}}-\underline{\xi})(\bar{\underline{x}}-\underline{\xi})'/2] \, d\bar{\underline{x}}$$

$$\times \frac{(n-1)^{p(n-1)/2}|\underline{S}|^{(n-p-2)/2}}{\pi^{p(p-1)/4}|\underline{\Sigma}|^{(n-1)/2}2^{p(n-1)/2}\Gamma_p(n-1)/2} \text{ etr } [-(n-1)\underline{\Sigma}^{-1}\underline{S}/2]d\underline{S} \qquad (7)$$

where

$$\Gamma_p(d) = \prod_{i=1}^{p}\Gamma(d - (i-1)/2).$$

The standard result

$$\int_{\underline{A}>0} |\underline{A}|^{(\alpha-p-1)/2} \operatorname{etr}[-\underline{B}^{-1}\underline{A}/2]d\underline{A}$$

$$= 2^{p\alpha/2}\pi^{p(p-1)/4}\Gamma_p(\alpha/2)|\underline{B}|^{\alpha/2} \qquad (8)$$

will be often used, the integration being taken over the region where $\underline{A}$ is positive definite.

Segal's process is expressed in the changes of variable

$$(\bar{\underline{x}},\underline{S}) \to (\bar{\underline{x}}-\underline{\xi},\ \underline{\Sigma}^{-1/2}\underline{S}\underline{\Sigma}^{-1/2}) \to (\underline{\xi},\underline{\Sigma}^{-1/2})$$

and leads to the density (7) with $\bar{\underline{x}}$ and $\underline{\xi}$ interchanged in the normal factor and $\underline{\Sigma}^{-1}$ and $\underline{S}$ interchanged in the Wishart factor. To find the marginal distribution of the vector $\underline{\xi}$, write

$$\underline{U} = \underline{\Sigma}^{-1},\ \underline{t} = n^{1/2}\underline{S}^{-1/2}(\bar{\underline{x}}-\underline{\xi}) \ ;\ \text{Jacobian} = n^{-p/2}|\underline{S}|^{1/2}$$

then

$$dF(\underline{t}) = \int_{\underline{U}>0} \left(\frac{n}{2\pi}\right)^{p/2} \frac{(n-1)^{p(n-1)/2}|\underline{S}|^{n/2}|\underline{U}|^{(n-p-1)/2}}{\pi^{p(p-1)/4}2^{p(n-1)/2}\Gamma_p(n-1)/2\ .n^{p/2}}$$

$$\times \ \operatorname{etr}\{-\underline{U}[(n-1)\underline{S} + \underline{S}^{1/2}\underline{t}\ \underline{t}'\underline{S}^{1/2}]\}\ d\underline{U}\ .$$

(8) gives

$$dF(\underline{t}) = \frac{(2\pi)^{-p/2}(n-1)^{p(n-1)/2}|\underline{S}|^{n/2}2^{pn/2}\pi^{p(p-1)/4}\Gamma_p(n/2)}{\pi^{p(p-1)/4}2^{p(n-1)/2}\Gamma_p(n-1)/2}|(n-1)\underline{S}+\underline{S}^{1/2}\underline{t}\underline{t}'\underline{S}^{1/2}|^{-n/2}$$

and since $\Gamma_p(n/2) \div \Gamma_p(n-1)/2 = \Gamma(n/2) \div \Gamma(n-p)/2$,

this reduces to

$$dF(\underline{t}) = \frac{\Gamma(n/2)}{[\pi(n-1)]^{p/2}\Gamma(n-p)/2} \cdot \frac{d\underline{t}}{|\ \underline{I} + \underline{t}\ \underline{t}'/(n-1)|^{n/2}} \qquad (9)$$

Now (9) is a particular form of a general density

$$dF(\underline{x}) = \frac{|\underline{A}|^{1/2}\Gamma(\beta)}{(\pi\alpha)^{p/2}\Gamma(\beta-p/2)} \cdot \frac{d\underline{x}}{|\underline{I} + \underline{A}\ \underline{x}\underline{x}'/\alpha|^{\beta}} \qquad (10)$$

which has the property that the distribution of any $k \leq p$ linear functions of the x's defined by $\underline{y} = \underline{H}\,\underline{x}$ , $\underline{H}$ $(k \times p)$, is given by

$$\frac{|\underline{HA}^{-1}\underline{H}^{1}|^{-1/2}\Gamma(\beta - (p-k)/2)}{(\pi\alpha)^{k/2}\Gamma(\beta-p/2)} \cdot \frac{d\underline{y}}{|\underline{I} + (\underline{HA}^{-1}\underline{H}^{1})^{-1}\ \underline{y}\ \underline{y}'/\alpha|^{\beta-(p-k)/2}} \qquad (11)$$

This result is proved in Appendix I.

Let $\underline{R}$ be the sample correlation matrix and $\underline{s}$ the diagonal matrix whose i-th entry is $S_{ii}^{1/2} = s_i$, i.e. $\underline{S}$ is the matrix of estimates of variance, then

$$\underline{S} = \underline{s}\ \underline{R}\ \underline{s}$$

and changing variables in (9) by

$$\underline{\tau} = \underline{s}^{-1}\underline{S}^{1/2}\underline{t}$$

the density of $\underline{\tau}$ is

$$\frac{\Gamma(n/2)}{((n-1)\pi)^{p/2}\Gamma(n-p)/2.\,|\underline{R}|^{1/2}} \cdot \frac{d\underline{\tau}}{|\underline{I} + \underline{R}^{-1/2}\underline{\tau\tau}'\ \underline{R}^{-1/2}/(n-1)|^{n/2}}$$

$$= \frac{\Gamma(n/2)}{((n-1)\pi)^{p/2}\Gamma(n-p)/2.\,|\underline{R}|^{1/2}} \cdot \frac{d\underline{\tau}}{|\underline{I} + \underline{R}^{-1}\underline{\tau\tau}'/(n-1)|^{n/2}} \qquad (12)$$

But

$$\underline{\tau} = \underline{s}^{-1}\underline{S}^{1/2}\underline{t}$$

$$= \underline{s}^{-1}n^{1/2}(\underline{\bar{x}}-\underline{\xi})$$

so that

$$\tau_i = \sqrt{n}(\xi_i-\bar{x}_i)/s_i$$

and from the form (12) and the reproductive property of (10) given in (11), the density of $\tau_i$ is given by

$$\frac{\Gamma(n-p-1)/2}{\sqrt{\pi(n-1)}.\Gamma(n-p)/2} \cdot \frac{d\tau_i}{[1+\tau_i^2/(n-1)]^{(n-p+1)/2}}$$

i.e. $u_i = \tau_i\sqrt{(n-p)}/\sqrt{(n-1)}$

is distributed as Student's t with n-p degrees of freedom. This is not in accord with A which requires that $\tau_i$ should be distributed as Student's t with n-1 degrees of freedom.

As far as the variances are concerned, the reproductive property of the Wishart distribution implies that the leading k×k minors of $\underline{\Sigma}^{-1}$ have Wishart densities with n-1 degrees of freedom but these are not simply proportional to the inverses of the leading k×k minors of $\underline{\Sigma}$. It is necessary to first transform to the matrix $(\underline{\Sigma}^{-1})^{-1}$ and then integrate out to show that $s_i^2/\sigma_i^2$ is distributed as $\chi_{n-p}^2/n-p$ instead of $\chi_{n-1}^2/n-1$ as obtained from the inversion of the marginal density of $s_i^2$. When there are only two variates, the density of $\rho$ given by Segal's result is identical with that of r with $\rho$ and r interchanged, which is obviously different from expression (4) obtained from C and $F(r,\rho)$.

The conjecture of Fisher at the R.S.S. Symposium (which will be abbreviated by FRSS to distinguish it from the result, (5), given in "Statistical Methods and Scientific Inference" which will be abbreviated by FSI) and its subsequent generalisation by Cornish lead to a form for the simultaneous distribution of the quantities

$$t_i = \sqrt{n}(\xi_i-\bar{x}_i)/s_i \ , \quad i = 1\ldots p$$

which is analogous to (12):

$$dF(\underline{t}) = \frac{\Gamma(n+p-1)/2}{(\pi(n-1))^{p/2}\Gamma(n-1)/2.|\underline{R}|^{1/2}} \cdot \frac{d\underline{t}}{|\underline{I}+\underline{R}^{-1}\underline{t}\,\underline{t}'(n-1)|^{(n+p-1)/2}} \quad (13)$$

whence, from (10) and (11), the density of $t_i$ is

$$\frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}.\Gamma(n-1)/2} \cdot \frac{dt_i}{[1 + t_i^2/(n-1)]^{n/2}} \quad .$$

That is $t_i$ is distributed in a t-density with n-1 degrees of freedom as A requires. This is not surprising for the derivation of (13) given by Cornish relies heavily on a generalisation of A.

No corresponding result was conjectured for the variances and correlations but using the conditions given it is possible to determine this as follows. When the variance-covariance matrix, $\underline{\Sigma}$, is known the vector $\bar{\underline{x}}$ is still the correct estimate of $\underline{\xi}$ and its density is $N(\underline{\xi},\underline{\Sigma}/n)$. This $\underline{z} = \sqrt{n}\underline{\Sigma}^{-1/2}(\bar{\underline{x}}-\underline{\xi})$ has the distribution $N(\underline{0},\underline{I})$ and leads to the conditional fiducial density of $\underline{\xi}$ on $\underline{\Sigma}$ as

$$N(\bar{\underline{x}},\underline{\Sigma}/n).$$

The marginal density of $\underline{\xi}$ is to be given by (13) expressed as a function of $\underline{\xi}$. Suppose now $f(\underline{\Sigma}^{-1},\underline{S})$ provides the density of $\underline{\Sigma}^{-1}$ required - it is more convenient to consider $\underline{\Sigma}^{-1}$ than $\underline{\Sigma}$ - clearly it will depend only on $\underline{S}$ and n since these are the sufficient statistics when the mean has been allowed for. By A, $f(\underline{\Sigma}^{-1},\underline{S})$ must then satisfy the equation

$$\int_{\underline{\Sigma}^{-1}>0} \left(\frac{n}{2\pi}\right)^{p/2} |\underline{\Sigma}^{-1}|^{1/2} \mathrm{etr}[-n\underline{\Sigma}^{-1}(\underline{\xi}-\bar{\underline{x}})(\underline{\xi}-\bar{\underline{x}})'/2] f(\underline{\Sigma}^{-1},\underline{S})\, d\underline{\Sigma}^{-1}$$

$$= \frac{\Gamma(n+p-1)/2.n^{p/2}}{((n-1)\pi)^{p/2}\Gamma(n-1)/2.|\underline{S}|^{1/2}} \cdot \frac{d\underline{\xi}}{|\underline{I} + n\underline{S}^{-1}(\underline{\xi}-\bar{\underline{x}})(\underline{\xi}-\bar{\underline{x}})'/(n-1)|^{(n+p-1)/2}} \quad (15).$$

From the fact that

$$c \int_{\underline{U}>0} \left(\frac{n}{2\pi}\right)^{p/2} |\underline{U}|^{1/2} \mathrm{etr}[-n\underline{U}\underline{y}\underline{y}'/2]\, |\underline{U}|^{(\alpha-p-1)/2} \mathrm{etr}[-(n-1)\underline{U}\underline{S}/2]\, d\underline{U}$$

$$= c\left(\frac{n}{2\pi}\right)^{p/2} \frac{2^{p(\alpha+1)/2}\Gamma_p(\alpha+1)/2\pi^{p(p-1)/4}}{(n-1)^{p(\alpha+1)/2}|\underline{S}|^{(\alpha+1)/2}} \cdot |\underline{I} + n\underline{S}^{-1}\underline{yy}'/(n-1)|^{-(\alpha+1)/2},$$

which is a consequence of (8), it can be seen that

$$f(\underline{\Sigma}^{-1},\underline{S}) = \frac{(n-1)^{p(n+p-2)/2}|\underline{S}|^{(n+p-2)/2}|\underline{\Sigma}^{-1}|^{(n-3)/2}}{\pi^{p(p-1)/4}\Gamma_p(n+p-2)/2.2^{(n+p-2)/2}} \operatorname{etr}[-(n-1)\underline{\Sigma}^{-1}\underline{S}/2] \quad (16)$$

satisfies (15) and what is more its uniqueness is ensured by a multivariate analogue of the uniqueness theorem for the Laplace transform.

This function represents a Wishart distribution with n+p-2 degrees of freedom i.e. p-1 more than ascribed by Segal's solution. As with that function $s_i^2/\sigma_i^2$ is distributed as $\chi^2$ but this time as $\chi^2_{n-1}/n-1$ as required by A. The correlation density for two variates is that of r with r and p interchanged and with (n+1) for n. Clearly these results are analogous to those derivable from (12) but are at variance with those derivable from (5) which appears to hold no place at all in relation to (12) and (16). However (5) is more closely related to (12) than to (16). If the density obtained through Segal's approach is written out completely for the case of two variates and is compared with FSI (i.e. (5)), then it is seen that the only point of difference is the density marginally given to the correlation coefficient. How this latter density fits into the picture will be described in Chapter VI.

For the case of two variates, the results obtained may be surveyed conveniently in tabular form as below:

| | Sample Size n | | | |
|---|---|---|---|---|
| | Marg. Samp. Inverse | Segal | FRSS | FSI |
| $\xi_1$ | $t_{n-1}$ | $t_{n-2}$ | $t_{n-1}$ | Not t |
| $\xi_2$ | $t_{n-1}$ | $t_{n-2}$ | $t_{n-1}$ | Not t |
| $\sigma_1$ | $\chi_{n-1}$ | $\chi_{n-2}$ | $\chi_{n-1}$ | Not $\chi$ |
| $\sigma_2$ | $\chi_{n-1}$ | $\chi_{n-2}$ | $\chi_{n-1}$ | Not $\chi$ |
| $\rho$ | $-\partial F(r,\rho,n)/\partial\rho$ | $F(\rho,r,n)$ | $F(\rho,r,n+1)$ | $-\partial F(r,\rho,n)/\partial\rho$ |

The entries in the last column may be justified by observing that the forms necessary to produce the t and $\chi$ densities, as exemplified by entries in the other columns, fail to arise in FSI. As remarked above FSI and Segal do not differ very much but the exceptionally awkward form of FSI makes it difficult to determine analytically how much difference the respective correlation densities do in fact make. Accordingly numerical calculation of the percentage points of FSI, (5), for various values of n and r have been undertaken and Table I contains the values computed at the 90%, 95% and 99% levels for the distribution of one t-like quantity

$$t = \sqrt{n}(\bar{x} - \xi)/s$$

from (5). The density (5) is a symmetric function of r so that values for $r \geq 0$ only had to be computed. Details of the computation may be found in Appendix II. Values for $t_{n-1}$ and $t_{n-2}$ are given for comparison.

Note that the distribution asymptotes in two directions viz. as $r \to \pm 1$ and as $n \to \infty$ and that in a sense it lies "between" FRSS and Segal's results. As might be expected the small sample results are the most discrepant and from a purely numerical point of view, the higher percentage

TABLE I

Percentage points of the marginal distribution of 1 Student-type T derived from the distribution given in Fisher:Statistical Methods and Scientific Inference.

Sample Number n = 5

| o/o | ±.000 | ±.200 | ±.400 | ±.600 | ±.800 | ±.900 | ±.950 | $t_{n-1}$ | $t_{n-2}$ |
|---|---|---|---|---|---|---|---|---|---|
| 90 | 2.53 | 2.50 | 2.48 | 2.42 | 2.32 | 2.25 | 2.18 | 2.13 | 2.35 |
| 95 | 3.36 | 3.33 | 3.27 | 3.18 | 3.04 | 2.93 | 2.82 | 2.78 | 3.18 |
| 99 | 5.85 | 5.68 | 5.60 | 5.48 | 5.13 | 4.80 | 4.52 | 4.60 | 5.84 |
| | | | | Sample Number n = 10 | | | | | |
| 90 | 1.96 | 1.95 | 1.93 | 1.92 | 1.88 | 1.86 | ur | 1.83 | 1.86 |
| 95 | 2.42 | 2.42 | 2.39 | 2.37 | 2.32 | 2.30 | ur | 2.26 | 2.31 |
| 99 | 3.50 | 3.49 | 3.47 | 3.41 | 3.37 | 3.28 | ur | 3.25 | 3.36 |
| | | | | Sample Number n = 15 | | | | | |
| 90 | 1.84 | 1.82 | 1.82 | 1.81 | 1.78 | 1.77 | ur | 1.76 | 1.77 |
| 95 | 2.24 | 2.23 | 2.22 | 2.20 | 2.17 | 2.16 | ur | 2.15 | 2.16 |
| 99 | 3.14 | 3.09 | 3.08 | 3.08 | 3.03 | 3.00 | ur | 2.98 | 3.01 |
| | | | | Sample Number n = 20 | | | | | |
| 90 | 1.78 | 1.77 | 1.77 | 1.76 | 1.75 | 1.74 | ur | 1.73 | 1.73 |
| 95 | 2.15 | 2.15 | 2.15 | 2.13 | 2.11 | 2.10 | ur | 2.09 | 2.10 |
| 99 | 2.85 | 2.85 | 2.83 | 2.81 | 2.80 | 2.80 | ur | 2.86 | 2.88 |
| | | | | Sample Number n = 40 | | | | | |
| 90 | 1.72 | 1.72 | 1.70 | 1.69 | 1.68 | ur | ur | 1.68 | 1.69 |
| 95 | 2.05 | 2.05 | 2.05 | 2.04 | 2.04 | ur | ur | 2.02 | 2.02 |
| 99 | 2.75 | 2.75 | 2.74 | 2.73 | 2.72 | ur | ur | 2.70 | 2.71 |

ur indicates that the value obtained is unreliable.

points are not as accurate as the lower ones.

From the form (10) it is not easy to determine the integral over regions of the form $x_i \leq X_i$, $i = 1 \ldots p$, but the natural integral to consider is that over regions defined by

$$q = \underline{x}'\underline{Ax} \leq Q$$

since $|\underline{I} + \underline{Axx}'/\alpha| = [1 + \underline{x}'\underline{Ax}/\alpha]$. To determine the distribution of q, change variables by

$$\underline{x} = \underline{Cy}$$

where $\underline{C}'\underline{AC} = \underline{I}$, Jacobian $= |\underline{A}|^{-1/2}$

then

$$dF(Q) = \int_{q = \underline{y}'\underline{y} \leq Q} \frac{\Gamma(\beta)}{(\pi\alpha)^{p/2}\Gamma(\beta - p/2)} \cdot \frac{dy}{[1 + \underline{y}'\underline{y}/\alpha]^{\beta}}$$

Now make a hyperspherical transformation in p dimensions and integrate for the angles giving

$$dF(Q) = \frac{\Gamma(\beta)\ 2\pi^{p/2}}{(\pi\alpha)^{p/2}\Gamma(\beta - p/2).\Gamma(p/2)} \int_0^{\sqrt{Q}} \frac{r^{p-1}dr}{[1 + r^2/\alpha]^{\beta}}$$

so that the frequency element of $q = r^2$ describes an F-distribution for

$$r^2/\alpha = p/(2\beta - p)\ .\ F(p, 2\beta - p).$$

Thus (12) implies that $\underline{\tau}'\underline{R}^{-1}\underline{\tau}/n-1)$ is distributed as $p/(n-p).F(p,n-p)$ and (13) implies that $\underline{t}'\underline{R}^{-1}\underline{t}/n-1$ is distributed as $p/(n-1)\ .\ F(p,n-1)$. However the quadratic forms are the same for both viz.

$$n/n-1\ .\ (\underline{\xi} - \underline{\bar{x}})'\underline{S}^{-1}(\underline{\xi} - \underline{\bar{x}})$$

so that both results cannot be correct. The former is consistent with the use of Hotelling's $T^2$ whereas the latter, as with previous results, requires an increase in the degrees of freedom. Thus for a given contour

$$\underline{t}'\underline{R}^{-1}\underline{t} = \alpha$$

the probability content under (13) is greater than that under (12). Corresponding results for (5) are given in Table II where the probabilities given are those corresponding to the ellipses

$$\underline{t}'\underline{R}^{-1}\underline{t} = 2F(2,n-1)$$

i.e. those associated with (13). It will be seen that the integral again shows asymptoting in two directions but that the variation is much slower for any sample size and percentage ellipse, i.e. the effect of varying r is not very marked. Examination of the values of the ordinates of the bivariate marginal of (5) shows that the curves

$$\underline{t}'\underline{R}^{-1}\underline{t} = 2F(2,n-1)$$

are very nearly curves of constant density of FSI. Again the computional details may be found in Appendix II.

An examination of the integrals associated with the hyper-ellipses provides one way of discriminating between the two expressions (12) and (13). Cornish, in defence of his solution, has remarked that Hotelling's form does not give any results when the number of variates, p, exceeds the degrees of freedom, n-1, available for the estimation of the variance-covariance matrix. This is clear because the Hotelling result is described by F(p,n-p) so that if $p > n-1$, the second set of degrees of freedom is zero or negative which means that no estimate of $\underline{\Sigma}$ is available or rather, the matrix $\underline{S}$ can be formed but it is now singular, since in general $\underline{S}$ is constructed as a product

$$\underline{S} = \underline{X}\,\underline{X}',$$

where $\underline{X}$ is p × (degrees of freedom), and is thus p× p; however the rank of $\underline{S}$ is min (p, degrees of freedom) which in this case is the latter

## TABLE II

### 90 per cent ellipses

r

| n | .000 | .200 | .400 | .600 | .800 | .900 | .950 |
|---|---|---|---|---|---|---|---|
| 5 | .843 | .844 | .842 | .840 | .835 | .830 | .830 |
| 10 | .878 | .875 | .878 | .871 | .873 | .877 | .866 |
| 15 | .884 | .882 | .885 | .884 | .887 | .883 | .884 |
| 20 | .984 | .891 | .892 | .892 | .896 | .887 | .895 |

95 per cent

| n | .000 | .200 | .400 | .600 | .800 | .900 | .950 |
|---|---|---|---|---|---|---|---|
| 5 | .913 | .913 | .912 | .909 | .904 | .904 | .897 |
| 10 | .934 | .934 | .932 | .932 | .932 | .933 | .929 |
| 15 | .941 | .940 | .939 | .940 | .939 | .941 | .943 |
| 20 | .942 | .943 | .942 | .941 | .942 | .938 | .949 |

99 per cent

| n | .000 | .200 | .400 | .600 | .800 | .900 | .950 |
|---|---|---|---|---|---|---|---|
| 5 | .978 | .978 | .977 | .976 | .974 | .974 | .973 |
| 10 | .984 | .984 | .984 | .984 | .984 | .983 | .982 |
| 15 | .987 | .986 | .986 | .987 | .987 | .986 | .987 |
| 20 | .988 | .988 | .988 | .987 | .988 | .987 | .994 |

quantity. Hence

$$\text{rank } (\underline{S}) < \text{order } (\underline{S})$$

i.e. $\underline{S}$ is singular.

The solution FRSS and its generalisation to (13) provide results which do not allow for this indeterminacy, for the correct procedure when $\underline{S}$ is singular is by no means clear. On this account it would seem that (13) must be discarded in favour of another solution.

Turning back to the bivariate case for a moment, suppose the question to be answered is : What inferences can be made about $\xi_1$? There are two possible approaches. First of all since $x_1$ margin ally is distributed as $N(\xi_1, \sigma_1{}^2)$ it is possible to use this fact to derive

$$t_1 = \surd n(\xi_1 - \bar{x}_1)/s_1$$

with the Student t-distribution on n-1 degrees of freedom. This however overlooks the effect of variate $x_2$ by eliminating it before trying to form any inferences. A second approach is to find the joint fiducial distribution of $\xi_1, \xi_2, \sigma_1, \sigma_2, \rho$ and subsequently eliminate the last four variates by integration. Combination of A and B and comparison with (13) requires that the results of these two approaches should be the same i.e. the inferences obtained about $\xi_1$ are independent of the dependence of $x_1$ on $x_2$, or for that matter on any further variates in a multivariate population. It is very difficult to credit this; if it is known that $x_1$ and $x_2$ are independent then clearly the two processes are identical because of the complete factorisation of the bivariate density but the existence of a real dependence must alter the result produced by the second process as outlined above. Moreover the requirement that marginal fiducial and marginal sampling results should match must be regarded with suspicion since the effect of other variates would always be overlooked, yet Fisher repeatedly used this property in an attempt to derive fiducial results.

Suppose $f(\underline{x},\underline{\theta})$ represents the joint density of the sufficient statistics for the parameters $\underline{\theta}$ and that the marginal density of $x_1$ depends only on $\theta_1$, then the above two processes can be represented schematically as follows:

1. $$f(\underline{x},\underline{\theta}) \xrightarrow{\int dx_1 \ldots dx_n} f_1(x_1,\theta_1) \xrightarrow{\text{inversion}} f_2(\theta_1,x_1)$$

2. $$f(\underline{x},\underline{\theta}) \xrightarrow{\text{inversion}} f_3(\underline{\theta},\underline{x}) \xrightarrow{\int d\theta_1 \ldots d\theta_r} f_4(\theta_1,\underline{x})$$

whence the inequality of the final results is clear. If $f(\underline{x},\underline{\theta})$ factorises into a set of univariate-uniparameter densities then the results will be identical; it is also possible to obtain equality in cases where the density is of a particular form as occurs with the univariate normal, but with the multivariate normal it is unreasonable to expect it.

On the other hand the solution obtained through the pivotals

$$\underline{t} = n^{1/2}\underline{\Sigma}^{-1/2}(\underline{\bar{x}} - \underline{\xi}) \ , \ \underline{W} = \underline{\Sigma}^{-1}\underline{S}$$

in the multivariate normal case does allow for the presence of other variates and the fact that the variability associated with any one variate is not simply describable by two parameters, by only permitting less stringent inferences subject, as well, to the requirement that the Matrix $\underline{S}$ shall be non-singular,

The two approaches discussed above really answer two different questions for the first one provides inferences about $\xi_1$ ignoring the parameters relating to $x_2 \ldots x_n$ i.e. about $\xi_1$ ignoring $\xi_2 \ldots \xi_n, \sigma_1 \ldots \sigma_n$, etc., while the second provides inferences about $\xi_1$ after eliminating $\xi_2 \ldots \xi_n$ i.e.

the two match up to analysis of variance entries for

Blocks ignoring Treatments

and Treatments adjusted for Blocks.

Similarly a comparison of the inferences available about the variances leads to the same conclusion : that the two $\chi^2$ densities are really appropriate to two different situations which cannot be considered as being identical. As far as the quantities $(r,\rho)$ are concerned, inferences about $\rho$ can only be considered in the face of a bivariate population so there is no question of a variate being ignored accounting for the various results, in particular those of Segal and FSI. The place of the latter will be more clearly shown in Chapter VI. However it might be noted that although Fisher was aware of the necessity for modifying parameter estimates when other parameters were assumed to be known, he made no comment about this in deriving FSI, for when $\rho$ is supposed to be known, $s_i^2$ is no longer the appropriate estimate of $\sigma_i^2$ but should be replaced by

$$\frac{s_i^2(1-\rho^2)}{(1-\rho r)} , \qquad i = 1,2$$

and moreover the density $dF(s_1,s_2|r)$ is not brought to a pivotal form by changing to these variates since it is still a function of $r$ and $\rho$. This was avoided at the time by claiming that $(\sigma_1,\sigma_2)$ were in a "higher stratum" than $\rho$ but why this should have been so was never made clear. The factorisation of the likelihood function separates $(\xi_1,\xi_2)$ from $(\sigma_1,\sigma_2,\rho)$ but does not sub-divide the latter group. This matter will be taken up again in Chapter VII.

The Limit Process.

Further support for the form derived by Segal can be obtained by considering a generalisation of the limiting process employed by Fisher in dealing with the univariate normal in 1935. To do this, two analogues of the quantities t and z are required and the following simply derivable results will be used:

(i) if $\underline{S}_1$ and $\underline{S}_2$ are estimates of $\underline{\Sigma}$ on $n_1$ and $n_2$ degrees of freedom respectively, then $\underline{S}_2\,\underline{S}_1^{-1} = \underline{W}$ has density

$$\frac{n_1^{n_1 p/2}\; n_2^{n_2 p/2}}{\pi^{p(p-1)/4}\,\beta_p(n_1/2, n_2/2)} \cdot \frac{|\underline{W}|^{(n_2-p-1)/2}\; d\underline{W}}{|n_1\underline{I} + n_2\underline{W}|^{(n_1+n_2)/2}} \qquad (17)$$

where

$$\beta_p(a,b) = \Gamma_p(a)\Gamma_p(b)/\Gamma_p(a+b) \;;$$

(ii) if $\underline{x}$ is distributed as $N(\underline{0},\underline{\Sigma})$ and $\underline{S}$ estimates $\underline{\Sigma}$ on n degrees of freedom, then

$$\underline{t} = \underline{S}^{-1/2}\underline{x}$$

has density

$$\frac{\Gamma(n+1)/2}{(\pi n)^{p/2}\,\Gamma(n-p+1)/2} \cdot \frac{d\underline{t}}{|\underline{I} + \underline{t}\underline{t}'/n|^{(n+1)/2}} \qquad (18).$$

Following Fisher's method, suppose an observed sample of $n_2+1$ observations from $N(\underline{\xi},\underline{\Sigma})$ yields $\bar{\underline{x}}_2, \underline{S}_2$ as the calculated estimates while the corresponding quantities from an unmade sample of $n_1+1$ observations are to be $\bar{\underline{x}}_1, \underline{S}_1$, then

$$\underline{W} = \underline{S}_2\underline{S}_1^{-1}$$

will be distributed in the form (17) and

$$\underline{t} = \left[\frac{n_1\underline{S}_1 + n_2\underline{S}_2}{n_1+n_2}\left(\frac{1}{n_1+1} + \frac{1}{n_2+1}\right)\right]^{-1/2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$$

will be distributed in the form (18) with

$$n = n_1 + n_2$$

and independently of $\underline{W}$. The joint density is then the product of the two and a subsequent change of variables:

$$(\underline{W},\underline{t}) \to (\underline{S}_1^{-1},\bar{\underline{x}}_1)$$

gives a Jacobian of

$$|\underline{S}_2|^{(p+1)/2} \left| \frac{n_1\underline{S}_1+n_2\underline{S}_2}{n_1+n_2} \left(\frac{1}{n_1+1}+\frac{1}{n_2+1}\right) \right|^{-1/2}$$

$n_1$ is now allowed to tend to infinity and because of the properties of $\bar{\underline{x}}_1$ and $\underline{S}_1$,

$$\bar{\underline{x}}_1 \to \underline{\xi}\ , \quad \underline{S}_1^{-1} \to \underline{\Sigma}^{-1}\ .$$

Also

$$\begin{aligned} & \lim_{m\to\infty} |\underline{I} + \underline{A}/m|^m \\ = & \lim_{m\to\infty} \prod_{i=1}^{p} (1 + \lambda i/m)^m \\ = & \exp\ (\lambda_1+\ldots+\lambda_p) \\ = & \operatorname{etr}\ [\underline{A}] \end{aligned}$$

where the $\lambda_i$ are the eigenvalues of $\underline{A}$. The algebra involved in the simplification is somewhat awkward and is not given here but there is no real difficulty. Application of Stirling's formula yields the limit as $n_1 \to \infty$ of the density as

$$\frac{n_2^{n_2p/2}}{2^{n_2p/2}\,\pi^{p(p-1)/4}\,\Gamma_p(n_2/2)} \quad |\underline{S}_2|^{n_2/2}\,|\underline{\Sigma}^{-1}|^{(n_2-p-1)/2}\,\operatorname{etr}[-\underline{\Sigma}^{-1}n_2\underline{S}_2/2]d\underline{\Sigma}^{-1}$$

$$\times \quad \frac{(n_2+1)^{p/2}|\underline{\Sigma}^{-1}|^{1/2}}{(2\pi)^{p/2}} \quad \operatorname{etr}\ [-(n_2+1)\underline{\Sigma}^{-1}(\underline{\xi}-\bar{\underline{x}}_2)(\underline{\xi}-\bar{\underline{x}}_2)'/2]d\underline{\xi} \qquad (19)$$

which is just the general result obtained through using

$$\underline{t} = n^{1/2}\underline{\Sigma}^{-1/2}(\bar{\underline{x}}-\underline{\xi})\ ,\ \underline{W} = \underline{\Sigma}^{-1}\underline{S}$$

as pivotals for the general multivariate density. As yet it has not proved possible to find quantities which will yield (13) and (16), i.e. FRSS, through the limit process and such quantities cannot exist since they would require degrees of freedom beyond those available in the data. Thus the solution formally derived by Segal as a function of $\underline{\xi}$ and $\underline{\Sigma}^{-1/2}$ is the solution appropriate to the multivariate normal distribution, its marginal densities yielding inferences about the appropriate parameters allowing for the effect of the other variates.

# CHAPTER VI

## THE REGRESSION DISTRIBUTIONS

The support for Segal's solution must be qualified a little further. Considerable comment arose in 1955 following a paper by J.G. Mauldon in which an asymmetric fiducial distribution for the parameters $\sigma_1, \sigma_2, \rho$ of a bivariate normal, was derived. The result can be obtained from the following type of consideration.

As already mentioned the Wishart distribution cannot be symmetrically factorised in a simple fashion for the elements of $\underline{\Sigma}^{1/2}$ are complicated functions of $\sigma_1, \sigma_2$ and $\rho$. However $\underline{\Sigma}^{1/2}$ is just one of the solutions of the matrix equation

$$\underline{A}'\underline{A} = \underline{\Sigma} \qquad (20)$$

and corresponds to the requirement that $\underline{A}$ should be square and symmetric; it is then unique. Equation (20) has infinitely many solutions but restricting $\underline{A}$ to be square, each solution is an orthogonal transform of any other solution, for if

$$\underline{A}'\underline{A} = \underline{\Sigma} \text{ and } \underline{B}'\underline{B} = \underline{\Sigma}$$

then since $\underline{\Sigma}$ is supposed non-singular

$$(\underline{B}')^{-1}\underline{A}'\underline{A}\,\underline{B}^{-1} = \underline{I}$$

$$(\underline{A}\underline{B}^{-1})'(\underline{A}\underline{B}^{-1}) = \underline{I}$$

or $$\underline{A}\underline{B}^{-1} = \underline{H},\ \underline{H} \text{ orthogonal}$$

i.e. $$\underline{A} = \underline{H}\,\underline{B}\ .$$

The matrix $\underline{A}$ is also uniquely determined if it is required to be triangular i.e.

$$A_{ij} = 0 \quad \text{if } j > i,$$

and this representation corresponds to the situation considered by Mauldon. Consider the density.

$$K.|\underline{\Sigma}|^{-1/2}\text{etr}[-n\underline{\Sigma}^{-1}(\bar{\underline{x}}-\underline{\xi})(\bar{\underline{x}}-\underline{\xi})'/2]d\bar{\underline{x}}$$

$$\times \quad |\underline{\Sigma}|^{-(n-1)/2}|\underline{S}|^{(n-p-2)/2}\text{etr}[-(n-1)\underline{\Sigma}^{-1}\underline{S}/2]d\underline{S}$$

and write $\underline{\Sigma}^{-1} = \underline{T}'\underline{T}$. Change variables to

$$\underline{\beta} = \underline{T}(\bar{\underline{x}}-\underline{\xi})$$

$$\underline{G} = \underline{T}\,\underline{S}\,\underline{T}'$$

$$\text{Jacobian} = |\underline{T}|^{-(p+2)/2}$$

$$dF(\underline{\beta},\underline{G}) \propto \text{etr}[-n\underline{\beta}\underline{\beta}'/2]\ |\underline{G}|^{(n-p-2)/2}\text{etr}[-(n-1)\underline{G}/2]d\underline{\beta}\ d\ \underline{G}\ .$$

Now transform to $\underline{\xi}$ and $\underline{T}$ and then to $\underline{\xi}$ and $\underline{\Sigma}^{-1} = \underline{T}'\underline{T}$, the Jacobian is (see Deemer and Olkin ref.)

$$|\underline{\Sigma}|^{-1/2} \prod_{i=1}^{p} t_{ii}^{\,p-2i+1} \prod_{i=1}^{p} k_{ii}^{\,2(p-i+1)}$$

where $\underline{S} = \underline{K}\underline{K}'$ and $\underline{K}$ has the same shape as $\underline{T}$. Then

$$dF(\underline{\xi},\underline{\Sigma}^{-1}) \propto |\underline{S}|^{(n-p-2)/2}|\underline{\Sigma}^{-1}|^{(n-p-2)/2}\Pi t_{ii}^{\,p-2i+1}\Pi k_{ii}^{\,2(p-i+1)}$$

$$\times \quad \text{etr}[-(n-1)\underline{\Sigma}^{-1}\underline{S}/2\ d\underline{\Sigma}^{-1}\ .\ |\underline{\Sigma}^{-1}|^{1/2}\text{etr}[-n\underline{\Sigma}^{-1}(\underline{\xi}-\bar{\underline{x}})(\underline{\xi}-\bar{\underline{x}})'/2]d\underline{\xi}\ .$$

Evidently the density now ascribed to $\underline{\Sigma}^{-1}$ has some asymmetric factors.

Suppose there are only two variates, then

$$\underline{\Sigma}^{-1} = \begin{bmatrix} 1/\sigma_1^2(1-\rho^2) & -\rho/\sigma_1\sigma_2(1-\rho^2) \\ -\rho/\sigma_1\sigma_2(1-\rho^2) & 1/\sigma_2^2(1-\rho^2) \end{bmatrix} = \begin{bmatrix} a & b \\ o & c \end{bmatrix}\begin{bmatrix} a & o \\ b & c \end{bmatrix}$$

yields

$$a = 1/\sigma_1 = t_{11},\ b = -\rho/\sigma_1\surd(1-\rho^2) = t_{21},\ c = 1/\rho_2\surd(1-\rho^2) = t_{22}$$

and similarly

$$k_{11} = s_1,\quad k_{22} = s_2\surd(1-r^2).$$

A factor

$$\sigma_1^{-4}\sigma_2^{-4}(1-\rho^2)^{-3}$$

arises from the transformation from $\underline{\Sigma}^{-1}$ to $\sigma_1,\sigma_2,\rho$ and then the density appears as

$$dF(\sigma_1,\sigma_2,\rho\,|s_1,s_2,r) \propto$$

$$\frac{s_1^{\,n}s_2^{\,n-2}(1-r^2)^{(n-2)/2}}{\sigma_1^{\,n+1}\sigma_2^{\,n-1}(1-\rho^2)^{(n+1)/2}}\ \text{etr}\ [-(n-1)\underline{\Sigma}^{-1}\underline{S}/2]\ d\sigma_1 d\sigma_2 d\rho \qquad (21)$$

so that $\sigma_1$ and $\sigma_2$ appear asymmetrically in the fiducial density so ... Mauldon considered that this was sufficient to point out a paradox in fiducial theory, for corresponding to equation (20) there is the alternative relation

$$\underline{\Sigma} = \underline{A}\ \underline{A}'\ ,\ \underline{A}\ \text{triangular}$$

which has the same properties. The above argument can be used to derive (21) but with ($s_1 \leftrightarrow s_2$) and ($\sigma_1 \leftrightarrow \sigma_2$). Hence there would be two asymmetric densities and the uniqueness property insisted upon by Fisher meant that a contradiction had been reached even if no account was taken of the symmetric inversions.

To see what part the form (21) plays in the whole theory, it is necessary to consider the relationship existing between the two variates $x_1$ and $x_2$ of a bivariate normal population. Basically the statement:

$x_1$ and $x_2$ have a bivariate normal distribution

$$N(\xi_1,\xi_2,\sigma_1,\sigma_2,\rho),$$

means that the two variates are dependent and that their simultaneous density function is specified by the five parameter values. However no further information is given about the dependence. There appear to be three possibilities:

(i) $x_1$ and $x_2$ are dependent upon each other without one playing the part of a determining variate;

(ii) $x_1$ depends on $x_2$, rather than vice-versa;

(iii) $x_2$ depends on $x_1$, rather than vice-versa.

The last two situations describe the basic relationship in data analysed by a covariance analysis where often the second measurement is one on the same characteristic as the first but after some time or treatment delay.

Under (i) there is no reason to distinguish between the variates and it is thus appropriate to use a symmetric form of fiducial inversion as is given by the pivotal quantity

$$\underline{W} = \underline{\Sigma}^{-1}\,\underline{S}.$$

However under situation (ii) or (iii) where one variate can be used to describe the pattern of behaviour of the other, it is necessary to use this added datum to form inferences. The factorisations

$$\underline{\Sigma} = \underline{A}'\underline{A} \quad \text{and} \quad \underline{\Sigma} = \underline{AA}'$$

allow for the extraction of a linear relationship between the variates, for $A_{22}$ represents the variance of one variate adjusted for the linear regression on the other, i.e.

$$\sigma^2_{ij} = \sigma^2_i\,(1-\rho^2),$$

while $A_{21}/A_{22}$ is the regression coefficient involved in this relationship.

Properly then, the problem should be regarded as one concerning these parameters rather than $\sigma_1, \sigma_2, \rho$ and it is not surprising that simple mathematical transformations from the regression parameters to $(\sigma_1, \sigma_2, \rho)$ fail to take account of the added information available about the direction of the $x_1 \leftrightarrow x_2$ dependence. Just as the same set of experimental data can be analysed in several different ways without arousing comment provided the reasons for doing so are given, and no contradictions are seen in the different analyses because of the varying basic models, so it must be expected that these differences will be reflected in the inferences produced because of the requirement that <u>all</u> the available information be used.

Even admitting square matrices only, the equation

$$\underline{\Sigma} = \underline{A}'\underline{A}$$

has infinitely many solutions as already mentioned, so that superficially it would seem that infinitely many inversions must exist. However apart from the symmetric and triangular solutions, no factorisations have been found which reduce the Wishart density to pivotal form or which seem to have any physical meaning corresponding to an interpretation of the data.

Admitting the validity of such expressions as (21) under the appropriate experimental conditions implies that in higher dimensions there will be many possible fiducial distributions. For example combinations of the variety:

$x_1$ depends on $x_2$ and $x_3$ but the latter pair cannot be separated, must be allowed for in 3 dimensions. This corresponds to

$$\underline{\Sigma} = \underline{A}'\underline{A}$$

where

$$\underline{A} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

and $a_{32} = a_{23}$; furthermore this representation is only uniquely determined if $a_{32} = a_{23}$. The various results obtainable correspond to different basic models for the data analysis and are not contradictory since they are not directly comparable, and so the inconsistency seen by Mauldon disappears.

By integrating (21) for $\sigma_1$ and $\sigma_2$ using the same type of transformations as are used to derive $F(r,\rho)$ from the bivariate Wishart density, it is easily found that the density of $\rho$ so obtained is identical with that derived from

$$-\frac{\partial}{\partial\rho}\int_{-1}^{r} dF(t,\rho)$$

viz. expression (4). This is not surprising because the derivation of (4) used as pivotal quantities just the elements of the matrix

$$\underset{\sim}{B} = \underset{\sim}{T}\,\underset{\sim}{K}$$

$$= \begin{bmatrix} s_1/\sigma_1 & 0 \\ -\rho s_1/\sigma_1\sqrt{(1-\rho^2)} + rs_2/\sigma_2\sqrt{(1-\rho^2)} & s_2\sqrt{(1-r^2)}/\sigma_2\sqrt{(1-\rho^2)} \end{bmatrix}$$

in the above notation. Clearly the form (21) with $(s_1 \leftrightarrow s_2)$, $(\sigma_1 \leftrightarrow \sigma_2)$ will yield the same density for $\rho$. Thus the marginal density ascribed to $\rho$ by Fisher in FSI comes from a completely different source than the other portions of FSI and it is perhaps a little ironical to note that Fisher warned against the use of such quantities as the elements of $\underset{\sim}{B}$ in an attempt to derive a fiducial density, immediately after he had completed the derivation of FSI which now turns out to be, in part, dependent upon these elements.

The complex construction of (5) = FSI can then be detailed as:

(a) the marginal density of $\rho$ is as obtained from (either of) the regression factorisations;

(b) $dF(\sigma_1, \sigma_2 | \rho)$ is as obtained from the symmetric inversion using $\underline{\Sigma}^{-1}\underline{S}$ as pivotal for the Wishart density:

(c) $dF(\xi_1, \xi_2 | \sigma_1, \sigma_2, \rho)$ is the bivariate normal as expected since the estimates of $\xi_1$ and $\xi_2$ do not depend on the knowledge available about the scale parameters, save that $\underline{\Sigma}$ is non-singular.

Thus (5) describes no situation at all and must be discarded as representing the fiducial distribution of the parameters of a bivariate normal distribution.

# CHAPTER VII

## PIVOTAL QUANTITIES

The process of inverting sampling distributions to generate fiducial distributions is dependent upon the existence of pivotal quantities which are functions of the observations and parameters having the property that their distribution is completely known, i.e. it is independent of observations and parameters. As already pointed out, where a sufficient statistic exists for a single parameter, the marginal cumulative probability function is always a pivotal being uniformly distributed on [0,1]. Generally this function is not available in a convenient form so that simpler quantities such as t and X are sought. It is to be noted however that the c.p.f. is a unique pivotal in the above case in the sense that all pivotal quantities are functions of it. The foregoing discussion of the bivariate normal makes it evident that the requirement of "one distribution - one set of pivotals - one fiducial inversion" cannot be maintained.

During the months before his death, Fisher gave some lectures at the University of Adelaide dealing with various aspects of inference in general and the fiducial theory in particular and he produced two sets of conditions to be laid upon pivotals; these are given below.

Set 1.

(i) Each shall be monotonic in one "parameter", uniformly for all possible sets of statistics.

(ii) Their simultaneous distribution must be independent of all parameters.

(iii) The statistics shall themselves constitute an exhaustive set.

Set 2.

(i) They must have a simultaneous sampling distribution independent of all parameters (ignoring those of lower strata).

(ii) Each must involve only one parameter of the stratum.

(iii) Each shall vary monotonically with that parameter uniformly for variates of the statistics.

(iv) Jointly they must involve a set of statistics exhaustive for these parameters.

Thus the second set is a refinement of the first but one or two points require explanation. The inverted commas around the word parameter in 1(i) are to indicate that the population character involved in the pivotal quantity may in fact be a function of the usual characteristics which describe the model. That this can lead to some confusion has already been demonstrated in connection with the symmetric and regression densities. Secondly the concept of a stratum of parameters is introduced in the second set. This idea was introduced to govern the way in which simultaneous fiducial distributions could be constructed from marginal and conditional densities by dividing the parameters into groups. The parameters in any one group were to be considered as of the same logical type while those in different groups could not be freely mixed. Unfortunately no indication was given as to how this stratification was to be determined; $\xi$ and $\sigma^2$ were considered to lie in different strata ($\sigma^2$ in the lower), as were $(\xi_1,\xi_2)$ and $(\sigma_1{}^2,\sigma_2{}^2,\rho)$ but this latter group was further subdivided into $(\sigma_1{}^2,\sigma_2{}^2)$ and $(\rho)$. In the course of deriving FSI Fisher gave as justification for beginning with the distribution of r, the fact that it depends only on $\rho$; this is clearly insufficient since the density of either estimate of variance depends only on the corresponding population character.

The examples already given of the symmetric and triangular factorisations of the normal variance-covariance matrix show that different sets of pivotals can provide different end results when expressed in terms of the same variates, yet nevertheless they are not inconsistent. Thus the problem of stratification can only be considered when referred to a completely specified situation which may necessitate a transformation of the parameters commonly employed.

It has already been pointed out that the behaviour of the estimates used in forming inferences is very important and since the estimates found by the Method of Maximum Likelihood have, in general, properties of Efficiency and Sufficiency which make them the estimates to be used, it is not unnatural to consider a system of stratification based upon the likelihood function itself.

In dealing with the symmetric case of the bivariate normal it was noted that the likelihood function could be factorised (treating it as a tensor pseudo-density with respect to the observations):

$$L = f_1(\underline{\bar{x}},\underline{\xi},\underline{\Sigma})f_2(\underline{S},\underline{\Sigma})\ f_3(\text{indep. of } \underline{\xi},\underline{\Sigma})$$

and furthermore the equation

$$\partial L/\partial \underline{\xi} = \underline{0}$$

had a solution independent of $\underline{\Sigma}$ (provided $|\underline{\Sigma}|\neq 0$). Suppose in general that the likelihood function $L$ can be transformed to

$$L = f_1(x_1,\theta_1;\theta_2\ldots\theta_p)f_2(x_2,\theta_2;\theta_3\ldots\theta_p)\ldots f_p(x_p,\theta_p)$$

in such a way that $\partial L/\partial\theta_1 = 0$ has a solution independently of the information available about $\theta_2\ldots\theta_p$ while the solution of

$$\partial L/\partial\theta_2 = 0$$

may depend on the solution already obtained for the $\theta_1$ equation etc. Then a natural stratification is set up. It may be that complete factorisation is not possible i.e. $L$ may factorise into functions of several parameters and variates as with the multivariate normal, then a stratification into groups is provided and the parameters within each group are of equal status.

Such a stratification may not be sufficiently general to deal with all desirable situations but in the cases so far analysed it provides a useful method of dealing with the successive inversions and ensures the uniqueness of the result for a given set of data which includes any information relevant to variate inter-relationships. With this method of grouping the problems associated with the Wishart density in particular do not arise, for the symmetric inversion uses $\underline{\Sigma}^{-1}\underline{S}$ as pivotal where

$$\underline{A} > 0$$

is to be understood as meaning that $\underline{A}$ is positive definite

i.e. $$\underline{x}'\underline{A}\,\underline{x} \geq 0$$

with equality if and only if $\underline{x} = \underline{0}$. The regression densities pose less of a problem for the factorisation of $L$ is complete and the successive inversions are exactly analogous to those used in the case of $N(\xi,\sigma^2)$. This also shows that care is needed in using functions of parameters as 1(i) was designed to imply. It is true that the regression 'parameters' are functions of $\sigma_1,\sigma_2$ and $\rho$ but their density is not simply obtainable from the symmetric result because of the difference in model so that functions of parameters cannot be used indiscriminately.

As a matter of interest, it is to be noted that the examples and basis of the fiducial theory require a considerable degree of symmetry to exist between statistics and parameters and indeed in all cases so far examined the pivotal quantities show this symmetry. The normal location parameters have pivotals which change sign only on the interchange of estimate and parameter i.e.

$$(\bar{x} - \xi) \to (\xi - \bar{x}) = -(\bar{x} - \xi);$$

the scale parameter $\sigma$ has a pivotal which inverts i.e.

$$s/\sigma \to \sigma/s = (s/\sigma)^{-1}$$

and so does the symmetric dispersion matrix

$$\underline{\Sigma}^{-1}\underline{S} = \underline{S}\,\underline{\Sigma}^{-1} = (\underline{\Sigma}^{-1}\underline{S})^{-1} .$$

Indeed the correlation coefficient density is the only common density which cannot be brought to symmetric form by a simple pivotal and it is precisely this lack of symmetry which causes

$$-\partial F(r,\rho)/\partial\rho$$

to return a function which is not a trivial transform of

$$F(r,\rho) .$$

The symmetry in univariate situations has already been mentioned in Chapter II.

Chapter VIII will give details of a case in which

(a) the method of ordinary pivotals coupled with the type of build up used by Fisher in forming FSI = (5);

(b) the limit process;

(c) the build up actually used by Fisher in forming FSI;

and for two variates

(d) the method of matrix pivotals analogous to $\underline{W} = \underline{\Sigma}^{-1}\underline{S}$,

all give the same answer.

# CHAPTER VIII

## A PARTICULAR EXAMPLE

The multivariate normal distribution to be considered is that with means $\xi_1,\ldots,\xi_p$ and variance-covariance matrix

$$\underline{\Sigma} = \sigma^2 \begin{matrix} 1 & \rho & \cdots \\ \rho & 1 & \cdots \\ \cdot & \cdot & \cdots \\ \rho & \rho & \cdots 1 \end{matrix} \qquad (p\times p)$$

$$= \sigma^2(1-\rho)\,\underline{I} + \sigma^2\rho\underline{1}\,\underline{1}',$$

where $\underline{1}' = [1\ 1\ \ldots\ 1]$ $(1\times p)$. As is usual with all multivariate (non-singular) normal distributions, the sample means are the appropriate maximum likelihood estimates of the population means, so that only the distribution of the estimates of $\sigma$ and $\rho$ must be found.

If n observations are available, write

$$\sum_{i=1}^{p} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 = S$$

and

$$\sum_{i<j=1}^{p} \sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{ij} - \bar{x}_j) = P,$$

then the likelihood function can be written as

$$e^L = (2\pi)^{-n/2}\{\sigma^{2p}(1-\rho)^{p-1}[1 + (p-1)\rho]\}^{-n/2}$$

$$\times \exp - \frac{1}{2\sigma^2}\left\{ \frac{[1+(p-2)\rho]S}{(1-\rho)\{1+(p-1)\rho\}} - \frac{2\rho P}{(1-\rho)\{1+(p-1)\rho\}} \right\}$$

since it can easily be shown that

$$\underset{\sim}{\Sigma}^{-1} = \frac{1+(p-1)\rho}{\sigma^2(1-\rho)\{1+(p-1)\rho\}} \cdot \underset{\sim}{I} - \frac{\rho}{\sigma^2(1-\rho)\{1+(p-1)\rho\}} \underset{\sim}{1}\underset{\sim}{1}' .$$

The appropriate estimates of $\rho$ and $\sigma^2$ are given by

$$r = 2P/(p-1)S,$$

$$s^2 = S/p(n-1)$$

respectively.

Perhaps the simplest way to find the distribution of these statistics is to make use of the special form of the variance-covariance matrix. Since this is of the form

$$\underset{\sim}{A} = k\,\underset{\sim}{I} + m\,\underset{\sim}{1}\,\underset{\sim}{1}',$$

there exists an orthogonal matrix $\underset{\sim}{H}$ which does not depend on k or m such that $\underset{\sim}{H}'\,\underset{\sim}{A}\,\underset{\sim}{H}$ is diagonal and in fact the matrix $\underset{\sim}{H}$ such that

$$\underset{\sim}{H}'\,\underset{\sim}{1}\,\underset{\sim}{1}'\,\underset{\sim}{H}$$

is diagonal, will be sufficient to accomplish this. Moreover $\underset{\sim}{1}\,\underset{\sim}{1}'$ is a matrix of rank one so that it has only one non-zero eigenvalue $\lambda$. Thus

$$\lambda = \text{tr}\,\underset{\sim}{1}\underset{\sim}{1}' = \text{tr}\,\underset{\sim}{1}'\underset{\sim}{1} = p.$$

Hence

$$\underset{\sim}{H}'\underset{\sim}{A}\underset{\sim}{H} = \begin{bmatrix} k + pm & & \\ & \ddots k & \\ & & \ddots k \end{bmatrix}$$

for a suitable arrangement of the rows of $\underset{\sim}{H}$.

To apply this to the problem in hand, write

$$\underset{\sim}{X} = [x_1,\dots,x_p] = [x_{ij} - \bar{x}_i],\ i=1,\dots p;\ j=1\dots n,$$

then clearly $$S + 2P = \text{tr } \underline{1}\underline{1}'\underline{X}\underline{X}',$$

and $$s^2 = \text{tr } \underline{X}\,\underline{X}'$$

so that writing $\underline{X} = \underline{H}\,\underline{Y}$ where $\underline{H}$ is as above, it follows that

$$S = \text{tr } \underline{HYY'H'} = \text{tr } \underline{H'}\,\underline{HYY'}$$
$$= \text{tr } \underline{YY'}$$
$$= \sum\sum y_{ij}^{2}$$

and $S + 2P = \text{tr } \underline{11'}\,\underline{XX'} = \text{tr } \underline{11'}\,\underline{HYY'}\,\underline{H'} = \text{tr } \underline{H'}\,\underline{11'}\,\underline{HYY'}$

$$= \text{tr} \begin{bmatrix} p & \underline{0}' \\ \underline{0} & \underline{0} \end{bmatrix} \underline{YY'}$$

$$= p \sum y_{1j}^{2}.$$

Hence $s^2$ becomes $\sum\sum y_{ij}^{2}/p(n-1)$ and $r$ becomes

$$\frac{1}{p-1} \cdot \frac{p\sum y_{1j}^{2} - \sum\sum y_{ij}^{2}}{\sum\sum y_{ij}^{2}} \quad .$$

From standard normal theory it is known that the y's are each normally distributed with variances

$$\sigma^2\{1 + (p-1)\rho\} , \sigma^2(1-\rho),\ldots,\sigma^2(1-\rho)$$

respectively and since $\underline{\Sigma}$ has the same structure as $\underline{A}$, the $y_i$ are distributed independently of each other.

Clearly the distribution of $s^2$ and $r$ is obtainable from that of

$$u = \sum y_{ij}^{2} \text{ and } v = \sum_{i=2}^{p} \sum y_{ij}$$

since $s^2 = (u+v)/p(n-1)$ and $r = 1/(p-1).\{(p-1)u-v\}/(u+v)$. However, from the reproductive property of independent $\chi^2$ variates,

$$u/\sigma^2\{1 + (p-1)\rho\}$$

is distributed as $\chi^2_{n-1}$ independently of

$$v/\sigma^2(1-\rho)\ ,$$

which is distributed as $\chi^2_{(p-1)(n-1)}$. A simple change of variables gives the distribution of $s^2$ and $r$ as

$$\frac{p(n-1)^{p(n-1)/2}(p-1)^{(p-1)(n-1)/2}\{1+(p-1)r\}^{(n-1)/2-1}(1-r)^{(p-1)(n-1)/2-1}}{2^{p(n-1)/2}\Gamma(n-1)/2.\Gamma(p-1)(n-1)/2.(1-\rho)^{(p-1)(n-1)/2}\{1+(p-1)\rho\}^{(n-1)/2}} \times$$

$$\frac{(s^2)^{p(n-1)/2-1}}{(\sigma^2)^{p(n-1)/2}} \times \exp\left\{-\frac{p(n-1)s^2[1+(p-2)\rho-(p-1)\rho r]}{2\sigma^2(1-\rho)\{1+(p-1)\rho\}}\right\}\ d(s^2)dr \qquad (22)$$

It is possible to proceed in the way Fisher did in deriving FSI for the marginal frequency function of $r$ is

$$f(r,\rho) = \frac{p(p-1)^{(p-1)(n-1)/2}\Gamma p(n-1)/2.(1-\rho)^{(n-1)/2}}{p^{p(n-1)/2}\Gamma(n-1)/2.\Gamma(p-1)(n-1)/2} \times$$

$$\frac{\{1+(p-1)\rho\}^{(p-1)(n-1)/2}(1-r)^{(p-1)(n-1)/2-1}}{\{1+(p-2)\rho-(p-1)\rho r\}^{p(n-1)/2}} \times$$

$$\{1+(p-1)r\}^{(n-1)/2-1}dr, \qquad (23)$$

and depends only on $\rho$. Since both $r$ and $\rho$ are restricted to the range $[-1/(p-1),1]$ by the requirement that the distribution be non-singular, it is necessary to consider

$$-\frac{\partial}{\partial\rho}\int_{-1/(p-1)}^{r<1} f(t,\rho)dt$$

which after some awkward algebra simplifies to (23) with the interchanges

$$1-p \leftrightarrow 1+(p-1)r\ ;\quad 1+(p-1)\rho \leftrightarrow 1-r$$

and it can be easily verified that this is a probability density with respect to $\rho$ over the range $-1/(p-1)<\rho<1$ provided $-1/(p-1)<r<1$.

In a similar fashion, the distribution of $s^2$ for given $r$ can be found and inverted by forming

$$-\frac{\partial}{\partial\sigma}\int_0^s f(t,\sigma|r,\rho)dt,$$

which yields

$$\frac{2p^{p(n-1)/2}(n-1)^{p(n-1)/2}\{1+(p-2)\rho-(p-1)\rho r\}^{p(n-1)/2}}{2^{p(n-1)/2}\Gamma p(n-1)/2.[(1-\rho)\{1+(p-1)\rho\}]^{p(n-1)/2}}\left\{\frac{s}{\sigma}\right\}^{p(n-1)}\frac{d\sigma}{\sigma}\times$$

$$\exp\left\{-\frac{p(n-1)s^2\{1+(p-2)\rho-(p-1)\rho r\}}{2\sigma^2(1-\rho)\{1+(p-1)\rho\}}\right\} \qquad (24)$$

However there is an alternative method of finding these densities which at the same time shows that the formula $-\partial F(T,\theta)/\partial\theta$ can be used and avoids all the algebra which it entails.

In the first place, it has already been shown (Bennett and Cornish 1963) that for the case p=2, the sampling density of $s^2$ and r can be expressed in terms of the matrices $\underline{\Sigma}$ and $\underline{S}$, but for p>2, (22) cannot be so expressed for there is an odd factor remaining which is essentially the product of the distinct eigenvalues of $\underline{S}$. This suggests examining the distribution of the eigenvalues:

$$a = s^2(1-r)\ ,\quad b = s^2\{1+(p-1)r\}$$

$$\partial(\sigma^2,r)/\partial(a,b) = \{b+(p-1)a\}^{-1}.$$

If at the same time the substitutions

$$\alpha = \sigma^2(1-\rho)\ ,\quad \beta = \sigma^2\{1+(p-1)\rho\}$$

are made, the density of a and b appears as

$$K(n,p).(a/\alpha)^{(p-1)(n-1)/2-1}(b/\beta)^{(n-1)/2-1}$$

$$\exp[-(n-1)\{b/\beta+(p-1)a/\alpha\}]d(a/\alpha)d(b/\beta)\ .$$

That is

$$(p-1)(n-1).(a/\alpha) \text{ and } (n-1).(b/\beta)$$

are independently distributed as $\chi^2_{(p-1)(n-1)}$ and $\chi^2_{n-1}$ respectively.

Returning to the original distribution for a moment, if $\rho$ is supposed to be known then the appropriate estimate of $\sigma^2$ is essentially

$$\frac{\{1 + (p-2)\rho - (p-1)\rho r\}s^2}{(1-\rho)\{1 + (p-1)\rho\}}$$

and, moreover, inspection of the distribution of $s^2$ for given r shows that it can be specified by writing

$$\chi^2 = \frac{p(n-1)\{1 + (p-2)\rho - (p-1)\rho r\}s^2}{\sigma^2(1-\rho)\{1 + (p-1)\rho\}} = \chi^2_{p(n-1)}$$

which quantity is proportional to

$$b/\beta + (p-1)a/\alpha$$

in terms of the new variates introduced above.

Accordingly change variables to

$$u = b/\beta+(p-1)a/\alpha, \quad v = a\beta/b\alpha, \quad \text{Jacobian} = u/\{1 + (p-1)v\}^2,$$

and

$$dF(u,v) = K(n,p)u^{p(n-1)/2-1}e^{-(n-1)u/2}du. \frac{v^{(p-1)(n-1)/2-1}dv}{\{1 + (p-1)v\}^{p(n-1)/2}},$$

which with the value of K(n,p) substituted yields

$$v = \frac{1-r}{1 + (p-1)r} \cdot \frac{1 + (p-1)\rho}{1-\rho}$$

as having the distribution of $e^{2z}$ with $n_1 = (n-1)(p-1)$ and $n_2 = n-1$ degrees of freedom, independently of $(n-1)u$ which is distributed as $\chi^2$ with $p(n-1)$ degrees of freedom.

Hence for the marginal distribution of $r$, $(1-\rho)/\{1 + (p-1)\rho\}$ is a scale parameter for $(1-r)/\{1 + (p-1)r\}$ and since $x < x_1$ corresponds to

$$\frac{1-x}{1 + (p-1)x} > \frac{1-x_1}{1 + (p-1)x_1},$$

the formula $-\partial F(T,\theta)/\partial\theta$ is justified. Also $(n-1)u$ is a pivotal for the conditional distribution of $s^2$ given r and is a function of the conditional estimate of $\sigma^2$ when $\rho$ is supposed known. These facts coupled with the monotonicity properties of u as a function of s and $\sigma$ show that $-\partial F(s_1\sigma|r,\rho)/\partial\sigma$ can be properly used to find the distribution of $\sigma^2$ given $\rho$. Thus the derivation of the correlation density and (24) is reduced to a series of probability inequalities exactly analogous to those used in dealing with the univariate normal.

For the case p=2, it has already been shown that

$$t = (r-\rho)/\sqrt{(1-r^2)}.\sqrt{(1-\rho^2)}$$

has essentially a Student t-distribution with n-1 degrees of freedom (Bennett and Cornish 1963). Since t is an increasing function of r and a decreasing function of $\rho$, it too may be used to invert the distribution of r. Naturally t and the above v must be functions of each other and it can easily be verified that

$$2 + 4t^2 = (v^2+1)/v.$$

The complete fiducial distribution is obtained by multiplying the correlation density by the conditional distribution of the variance and then using the fact that, as already pointed out, the sample means are the appropriate estimates of the population parameters even when $\sigma^2$ and $\rho$ are supposed known. Thus the multivariate normal conditional density of the means is multiplied by the appropriate marginal density to yield the final result. By direct integration the marginal density of the means can be found in the form

$$\frac{n^{p/2}\Gamma n/2.\Gamma n(p-1)/2}{(n-1)^{p/2}s^p(1-r)^{(p-1)/2}\{1 + (p-1)r\}^{1/2}\pi^{p/2}\Gamma(n-1)/2.\Gamma(n-1)(p-1)/2} \times$$

$$\left\{ 1 + \frac{n(S + 2P)}{p(n-1)s^2\{1 + (p-1)r\}} \right\}^{-n/2} \left\{ p-1 + \frac{n\{(p-1)S - 2P\}}{p(n-1)s^2(1-r)} \right\}^{-n(p-1)/2}$$

where S and P are as defined earlier. This may be written somewhat more neatly by introducing the matrices

$$\underline{A} = \frac{n}{p(n-1)s^2\{1 + (p-1)r\}} \underline{1}\,\underline{1}',$$

$$\underline{B} = \frac{n}{p(n-1)s^2(1-r)} [p\,\underline{I} - \underline{1}\,\underline{1}'] ,$$

so that $|\underline{A} + \underline{B}| = n^p/[(n-1)^p s^{2p}(1-r)^{p-1}\{1 + (p-1)r\}]$. Then the above distribution may be written as

$$dF(\underline{\xi}) = K(n,p)$$

$$\frac{|\underline{A} + \underline{B}|^{1/2} d\underline{\xi}}{|\underline{I} + \underline{A}(\underline{\xi}-\underline{\bar{x}})(\underline{\xi}-\underline{\bar{x}})'|^{n/2} |(p-1)\underline{I} + \underline{B}(\underline{\xi}-\underline{\bar{x}})(\underline{\xi}-\underline{\bar{x}})'|^{n(p-1)/2}} \quad . \qquad (25)$$

The above pivotal analysis shows that Fisher's FSI approach gives the same results as the pivotal method in this case. The reasons are fairly clear; in the first place complete factorisation of the 'Wishart' density is possible and each piece yields a simple pivotal which possesses the necessary monotonicity properties and moreover the recipricocity property mentioned in Chapter VII for

$$v(\rho,r) = 1/v(r,\rho)$$

and $$u(\sigma,s) = 1/u(s,\sigma).$$

Secondly the conditional density

$$dF(s^2|r)$$

is a function of the conditional estimate of $\sigma^2$ given $\rho$. These properties are themselves brought about by the relatively simple form of the density and the fact that the 'Wishart' density reduces to a two variate density by transformation of variates to its eigenvalues.

# CHAPTER IX

## CONCLUSION

Fisher at no stage gave any formal set of rules for the derivation of fiducial distributions in any general case but was content to give illustrative examples. The conditions A,B,C were derived by examining these examples and from discussions with Fisher. In the preceding chapters an attempt has been made to show that the simple ideas involved in dealing with such densities as the negative exponential and the univariate normal cannot be taken over to deal with the multivariate normal density. The theory of fiducial inference is designed to supply probability densities to describe the uncertainty about distribution parameters. Therefore the concepts of marginal and conditional densities must play their parts just as they do in sampling theory. In the negative exponential and ordinary normal cases, the marginal and conditional sampling densities re-appear as describing the marginal and conditional fiducial densities but this is due to the rather special factorisation of the likelihood function considered as a function of the parameters, i.e. on the natural stratification of these characters of the parent population.

When there are several variates it has been shown that this stratification is only partial for the parameters are naturally separated into groups and the sampling densities are symmetric with respect to these groups of parameters and their estimates rather than with respect to any one parameter-statistic pair. It would appear that the desire to maintain too close an analogy with sampling densities has been responsible for the derivation of such forms as (13) which return the t and $\chi$ densities for mean and variance which would be obtained from inverting the marginal sampling densities of the sufficient statistics. However the density returned for any one of the correlation coefficients then presents a problem for the 'degrees of freedom' involved are increased beyond those originally available.

This in turn raises another point. The concept of degrees of freedom as applied to sampling statistics estimating variation or association is well understood; it is not obvious, perhaps, that this concept should find itself exactly parallelled in fiducial theory. In fact it must still be taken as referring to the sample characters rather than the parameters which are the new distribution variates. It is thus difficult to imagine this characteristic of the fiducial distribution exceeding the degrees of freedom available for the estimation of the parameter in question.

Relying on the likelihood stratification, it has been shown that the solution first formally obtained by Segal should be used to provide inferences about the parameters of a multivariate normal distribution when there is no outside information regarding the variates themselves. The corresponding loss in 'degrees of freedom' for the densities of means and variances has been accounted for in terms of the presence of extra variates, which increase rather than diminish the uncertainty relating to any one parameter. However the sampling distributions of Hotelling and

Wishart play their parts but only with respect to the sets of parameters produced by the stratification.

The inconsistencies seen by Meuldon as a result of considering the triangular factorisation of the Wishart matrix have been explained in terms of the direction of the dependence between the variates involved. Information of this variety is vital to any attempt to draw inferences from actual data so that it is not only peculiar to fiducial theory, but fiducial theory seems to be the only one in which the natural interpretation of this triangular factorisation simply appears.

There is not sufficient space here to discuss the relationship between fiducial theory and alternate theories of inference. Several facts may be noted however. For adherents to either the fiducial or confidence theory, the Bayesian or inverse probability type of argument possesses one major stumbling block viz. How is the appropriate a priori distribution determined? This stumbling block has been noticed by many Bayesian statisticians themselves but they have either ignored it or preferred not to see it. There are cases, mainly those involving problems of inheritance in genetics, where such distributions can be constructed and indeed it would be foolish to overlook them, but the general assumption of a uniform density or of some specific function, is not easily justified.

At an early date in the 1930's there was some confusion regarding the possible identity of the theories advanced by Fisher and Neyman since the numerical results obtained in many cases were the same while the logical interpretation differed. Admittedly Fisher was more restrictive in his applications while Neyman wanted to create a very general theory. That the two approaches differed was first satisfactorily demonstrated over the Behrens-Fisher problem which has taken up more than its fair share of Journal space in the last thirty years. Perhaps more unfortunate than any other aspect of the differences was the personal atmosphere which crept in and prevented any possibility of the discussion of fundamentals between two men of undoubted genius.

The most essential difference between the two theories seems to be that whereas the confidence theory can provide regions describing the probability of occurrence of a series of events <u>under any given hypothesis or relative to one of a series of alternative hypotheses about the parameters</u>, fiducial theory attempts to remove the necessity of using any hypotheses beyond the random nature of the data. Initially the fiducial argument was applied to fairly special situations, viz. those in which sufficient statistics existed and the appropriate monotonicity conditions were satisfied. At the same time it was designed to go further than the routine tests of significance by providing a method of specifying the uncertainty inherent in any inferences drawn from experimental data. Moreover at no stage was it claimed that the method was universally applicable or would prove to be so when further developed. Essentially it provides a measure on the subsets of the parameter space and as such it suffers the disadvantage of any non-trivial measure, i.e. all subsets cannot be measured let alone all densities so that there will be densities on which the method, in its existing form, will fail to be applicable. Confidence theory also provides such a measure and the non-identity (or rather the non-equivalence) of the two methods means that there will be problems which can be treated uniquely by one method which cannot be treated by the other so that to say that one theory is 'better' than another has little meaning.

There is one point in the early exposition of fiducial theory that requires some qualification. The following quotation from "The Fiducial Argument in Statistical Inference": [Annals of Eugenics Vol. 6, Part 4, pp 301-398, 1935] will serve to illustrate the point.

"If a sample of n observations, $x_1, \ldots, x_n$, has been drawn from a normal population having a mean value $\mu$, and if from the sample we calculate the two statistics

$$\bar{x} = \frac{1}{n} S(x)$$

and

$$s^2 = \frac{1}{n-1} S(x-\bar{x})^2,$$

where S stands for summation over the sample, "Student" has shown 1925 that the quantity t, defined by the equation

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s} \quad ,$$

is distributed in different samples in a distribution dependent only from the size of the sample, n. It is possible, therefore, to calculate, for each value of n, what value of t will be exceeded with any assigned frequency, P, such as 1 per cent. or 5 per cent. These values of t are, in fact, available in existing tables (Fisher, 1925-34).

It must now be noticed that t is a continuous function of the unknown parameter, the mean, together with observable values, $\bar{x}$, s and n, only. Consequently the inequality

$$t > t,$$

is equivalent to the inequality

$$\mu < \bar{x} - st_1/\sqrt{n},$$

<u>so that this last inequality must be satisfied with the same probability as the first."</u> (Writer's underlining).

The underlined portion of the last sentence is the type of statement which caused many people to baulk and ask the questions: Why <u>must</u>? What is obvious about it? Algebraically it is trivial, but logically it is not. Basically there is no necessity for the statement to be logically true it is just that in view of the evidence available and taking regard for the properties of the statistics involved, there is reason for taking the statement as logically rational and measuring the uncertainty about $\mu$. As with all inferences there will be those that disagree with part or all of the procedure but provided the grounds on which such statements are based are made clear, they have a right to consideration and make a good deal more sense than results derived by some other methods.

Criticisms of the nature levelled by Lindley in his two-observation example just cannot be taken seriously for to overlook one observation in favour of another or to fail to combine their information before making any inferences just penalises the precision of the inferences and makes them worthless.

To return to the conditions A,B,C which served as a background for the comparison of the various densities available for consideration, some modifications may be made which seem reasonable. Naturally the sampling density of the set of jointly sufficient estimates must serve as a starting point for the derivation of a fiducial distribution but, as already pointed out, it is not reasonable to suppose that the marginal sampling densities will simply invert to give the appropriate fiducial density even although the sampling distribution may provide an exact significance test. The reason for this is that several parameters of the set specifying the population may be of the same logical type in the sense that the knowledge extractable from the sample about any one is dependent upon the information available about the others. A system of stratification based on the factorisation of the likelihood function itself and the solution of the estimation equations, seems to reveal this grouping which determines the extent to which A is applicable. Once this grouping is established, the sampling densities can be used to invert the necessary probability statements for the groups as a whole. This process is further dependent upon any information relevant to the dependence between variates, which will enable the correct 'parameters' to be used e.g. it will be clear which of

$$\underline{\Sigma} = \underline{A}^2 \quad \text{or} \quad \underline{\Sigma} = \underline{A}'\underline{A}, \quad \underline{A} \text{ triangular}$$

is the appropriate representation to be employed.

This latter point also applies to B for the concept of One distribution-One inversion, must be modified to ensure that the problem to be solved is completely specified in the first place. However allowing for this and the modification of A, in all examples so far examined the consistency required by B has been found to be satisfied.

Mention has already been made of the restrictions implied by the use of the formula of condition C although they are not really more stringent than those inherent in the existence of a fiducial distribution for if

$$P = Pr(\theta_1 < a_1, \ldots, \theta_k < a_k)$$

is found by probability inversion then

$$\frac{\partial^k P}{\partial \theta_1 \ldots \partial \theta_k}$$

represents the frequency element of the fiducial density. The negative sign arises from the fact that in most simple cases estimate and parameter occupy logically reciprocal positions as already noted in section (7). The main danger in the indiscriminate use of $\pm\partial F(T_1\theta)/\partial\theta$ is that it will be mis-applied by splitting up one of the parameter strata.

Although the full extent of the applicability of fiducial theory is not yet known, particularly in relation to problems in which ancillary statistics must be employed, the existing results provide clear concise methods of dealing with a large number of applied problems and at the same time highlight the care necessary in designing, analysing and interpreting experimental data.

## APPENDIX I

Suppose the (p×1) vector $\underline{t}$ is distributed in the density

$$\frac{|\underline{A}|^{1/2}\Gamma(2\beta)/2}{(\pi\alpha)^{p/2}\Gamma(2\beta-p)/2} \qquad \frac{d\underline{t}}{|\underline{I} + \underline{A}\ \underline{t}\underline{t}'/\alpha|^{\beta}}$$

$$= \frac{|\underline{A}|^{1/2}\Gamma(2\beta)/2}{(\pi\alpha)^{p/2}\Gamma(2\beta-p)/2} \qquad \frac{d\underline{t}}{(1 + \underline{t}'\underline{A}\underline{t}/\alpha)^{\beta}}$$

and the marginal density of k<p linear functions $\underline{y} = \underline{H}\underline{t}$ is required, $\underline{H}$(k×p). First changing variables to $\underline{x} = \underline{M}^{-1}\underline{t}$ where

$$\underline{M}'\underline{A}\ \underline{M} = \underline{I}\ ,$$

gives a Jacobian of $|\underline{M}| = |\underline{A}|^{-1/2}$ and the required ensity is

$$\frac{\Gamma(2\beta)/2}{(\pi\alpha)^{p/2}\Gamma(2\beta-p)/2} \cdot \frac{d\underline{x}}{(1 + \underline{x}'\underline{x}/\alpha)^{\beta}}$$

integrated over $\underline{y} = \underline{H}\ \underline{t} = \underline{H}\ \underline{M}\ \underline{x} < \underline{Y}$. Now $\underline{H}\underline{M}$ is (k×p) and can be augumented by a matrix $\underline{N}$(p-k×p) to a (p×p) matrix, where $\underline{N}$ is arbitrary for the moment. Then change variables by

$$\begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix} = \begin{bmatrix} \underline{H}\ \underline{M} \\ \underline{N} \end{bmatrix} \underline{x}$$

$$\underline{x} = \begin{bmatrix} \underline{H}\ \underline{M} \\ \underline{N} \end{bmatrix}^{-1} \begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}, \quad \text{Jacobian} = \begin{vmatrix} \underline{H}\ \underline{M} \\ \underline{N} \end{vmatrix}^{-1} ;$$

$$\underline{x}'\underline{x} = [\underline{y}'\underline{z}'][\underline{M}'\underline{H}'\ \ \underline{N}']^{-1} \begin{bmatrix} \underline{H}\ \underline{M} \\ \underline{N} \end{bmatrix} \begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}$$

$$= [\underline{y}'\underline{z}'] \begin{bmatrix} \underline{H} & \underline{M} \\ & \underline{N} \end{bmatrix} (\underline{M}'\underline{H}' \quad \underline{N}')^{-1} \begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}$$

$$= [\underline{y}' \ \underline{z}'] \ [(\underline{H}\ \underline{M}\ \underline{M}'\underline{H}')^{-1} \dotplus \underline{I}] \begin{bmatrix} \underline{y} \\ \underline{z} \end{bmatrix}$$

if $\underline{N}$ is chosen so that $(\underline{H}\ \underline{M})\underline{N}' = \underline{0}$ and $\underline{N}\ \underline{N}' = \underline{I}_{p-k}$, which is always possible just by choosing the columns of $\underline{N}$ to be an orthogonal basis for the space of solutions of the equation

$$(\underline{H}\ \underline{M})\ \underline{x} = 0$$

Since $|\underline{A}| = |\underline{A}\ \underline{A}'|^{1/2}$, $\begin{vmatrix} \underline{H} & \underline{M} \\ & \underline{N} \end{vmatrix}^2 = |\underline{H}\underline{M}\underline{M}'\underline{H}'| = |\underline{H}\underline{A}^{-1}\underline{H}'|$

so that the density becomes

$$\frac{\Gamma(\ \beta)/2}{(\pi\alpha)^{p/2}\Gamma(2\beta-p)/2.|\underline{H}\underline{A}^{-1}\underline{H}'|^{1/2}} \int_{\underline{z}=-\underline{\infty}}^{\underline{\infty}} \frac{d\underline{z}}{[1 + \underline{y}'(\underline{H}\underline{A}^{-1}\underline{H})^{-1}\underline{y}/2 + \underline{z}'\underline{z}/2]^{\beta}} .$$

The integration can be carried out by making a hyperspherical transformation in p-k dimensions thus reducing the problem to a standard beta-function type and producing the final result as

$$\frac{\Gamma[2\beta-(p-k)]/2}{(\pi\alpha)^{k/2}\Gamma(2\beta-p)/2.|\underline{H}\underline{A}^{-1}\underline{H}'|^{1/2}} \qquad \frac{d\underline{y}}{[1 + \underline{y}'(\underline{H}\underline{A}^{-1}\underline{H})^{-1}\underline{y}/2]^{\beta-(p-k)/2}} .$$

This is equivalent to (11) since

$$(1 + \underline{a}'\underline{B}\underline{a}) = \underline{I} + \underline{B}\ \underline{a}\ \underline{a}'$$

as can be easily shown by examining the eigenvalues of the matrix $\underline{B}\ \underline{a}\ \underline{a}'$ which are $\underline{a}'\underline{B}\ \underline{a}$ and (p-1) zeroes. Thus the proof is completed.

## APPENDIX II

As already indicated, expression (5) is very hard to handle but to find the density of one Student-type t, some explicit reduction is possible. First of all $\xi_1$ was eliminated and then the variable

$$t = \sqrt{n}(\xi_1-\bar{x}_1)/s_1$$

was introduced. By proceeding in the same way as that use to obtain the correlation density from the bivariate Wishart distribution, it was then possible to integrate out for $\sigma_1$ and $\sigma_2$. With the notation

$$R = r/\sqrt{[1 + t^2(1-x^2)/n-1]}$$

$$I_\mu(x) = \int_o^\infty \frac{dy}{(\text{Cosh } y - x)^\mu}$$

$$f(x,r) = (1-x^2)^{(n-2)/2} \left\{ \frac{I_{n-2}(xr) + xrI_{n-1}(xr)}{I_{n-1}(xr)} \right\} ,$$

the density of t can be written as

$$f(t;r,n) = \frac{(n-2)(1-r^2)^{(n-2)/2}}{2\pi\sqrt{n-1}} \int_{-1}^{1} \frac{f(x,r) \qquad dx}{(1-xR)^{n-1}[1 + t^2(1-x^2)/n-1]^{n/2}} .(*)$$

Of the integrand, f(x,r) is independent of t so that for given n and r it only need be evaluated once whereas the remainder must be calculated for each value of t.

The procedure adopted was to fix first n, then r; calculate f(x,r) for x = -0.99(0.01)0.99 so that the range [-1,1] was divided into 199 points, then for each t = 0(0.2)k, calculate the remainder of the integrand for the dissection points of [-1,1] and integrate numerically using Gregory-Newton Formula up to and including 8th differences. k was fixed as approximately the 1% point of the Student distribution on n-1 degrees of freedom so that good approximation to the appropriate percentage points was obtained without undue calculation.

The values of n and r taken were as follows

$$n = 5,10,15,20$$

and for each n,

$$r = 0,0.2,0.4,0.6,0.8,0.9,0.95.$$

It was not necessary to consider negative values of either r or t since

$$f(t;r,n) = f(t;-r,n) = f(t;r,n).$$

Cumulative probabilities as given in Table I were calculated using Gregory-Newton Formula to 3rd differences and interpolating for the percentage points, having first checked the accuracy by using Student-t ordinates as data. Computationally the larger values of r gave more trouble than the larger values of n.

Mr. J.G. Sanderson was kind enough to allow the writer to use the newly completed CIRRUS computer developed by the University of Adelaide to handle the job and a listing of the program is given below.

```
12 1 ..MARGINAL FIDUCIAL ORD (FISHER) OF T. G.W.B.
0-END 12
          .ALLOT(W(1/199),ROI(1/199),ROT(1/199),ROW,DRO,
                 TT,DT,R,PI=3.1415927,AN,AN1,AN2,C,AJ,RO,
                 A1,A2,A3,A4,A5,P1,P2,P3,AI,T,D,B1,B2,.I
                 N,NO,M,N1,I,J)
          .FUNCTION(SQRT(.R),PWR(.R,.R),ARG(.R,.R),.I INT
                   (.R))
          .FOR I=1(1)199,.READ .UNIT 3(W(I))
START:    .READ .UNIT 3(ROW,DRO,TT,DT,R,N,NO)
          .WRITE .UNIT 2,F('R=',R,',  N=',N,',  NO=',NO,',
                          TT=',TT,',  DT=',DT)
F:        .FORMAT(/S,F6.2,S,I3,S,I3,S,F6.2,S,F6.2)
          M=N-1
          N1=199
          AN=N
          AN1=AN-1
          AN2=(AN-2)/2
          C=PWR(1-R'2,AN2)*AN2/(PI*SQRT(AN1))
          .FOR J=1(1)N1,
          1(AJ=J
          RO=ROW+AJ*DRO
          A1=RO*R
          A2=1-A1'2
          A3=SQRT(A2)
          A4=ARG(-A1,A3)
          A5=1-RO'2
          P2=A2*A4/(A3+A1*A4)
```

```
        .FOR I=3(1)M,
        2(AI=I
        P3=(AI-1)*A2/((AI-2)*P2+(2*AI-3)*A1)
        P1=P2
        P2=P3)2
        ROI(J)=PWR(A5,AN2)*(P2+A1))1
        .FOR I=1(1)NO,
        1(AI=I-1
        T=TT+AI*DT
        D=0
        .FOR J=1(1)N1,
        2(AJ=J
        RO=ROW+AJ*DRO
        B1=1-RO'2
        B2=SQRT(1+T'2*B1/AN1)
        ROT(J)=1/(B2*PWR(B2-RO*R,AN1))
        D=D+C*W(J)*ROI(J)*ROT(J))2
        .WRITE .UNIT 2,L8(D,T))1
L8:     .FORMAT(E12.6,F5.1)
        .GO .TO START
        .WAIT
        .END
```

Analogously the expression for $f(t_1,t_2;r,n)$ is

$$\frac{(n-2)(1-r^2)^{(n-2)/2}}{2\pi^2\left(1+\frac{t_1^2}{n-1}\right)^{n/2}\left(1+\frac{t_2^2}{n-1}\right)^{n/2}} \int_{-1}^{1} f(x,r)\, I_n\,(xR)dx$$

where
$$R = \frac{(n-1)r + t_1t_2}{\sqrt{(n-1)+t_1^2}\ \sqrt{(n-1)+t_2^2}} .$$

Now $f(t_1,t_2;r,n)$ possesses the symmetries

$$f(t_1,t_2;r,n) = f(t_2,t_1;r,n)$$

$$f(t_1,t_2;r,n) = f(-t_1,t_2;-r,n)$$

and
$$f(-t_1,-t_2;r,n) = f(t_1,t_2;r,n),$$

so that by a rotation of axes through 45 degrees it can be shown that it is sufficient to evaluate the function over the triangular region specified by

$$t_2^2 - t_1^2 \geq 0.$$

This considerably reduces the amount of work involved. Again for each $(n,r)$ , $f(x,r)$ need only be found once but the rest of the function has to be evaluated for each pair of values $(t_1,t_2)$.

The values of the parameters used were:

$$n = 5,\ .0,\ 15,\ 20$$

and for each n,

$$r = 0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 0.9,\ 0.95.$$

The requirements of subsequent integration over the ellipses defined by Fisher's FRSS density required that $t_1$ be taken steps of 0.2 from 0 to a value corresponding approximately to the axis length of these ellipses, while $t_2$ was taken from $-t_1(\max)$ to $+t_1(\max)$ in steps of 0.2. Integration over the ellipses as given in Table II was accomplished by crude summation after a check on the accuracy of the method. Some idea of the work involved may be gained from the fact that the complete calculation was performed on a Control Data Corporation 3600 machine and took over 100 minutes. A listing of the program is attached below.

FTN4.10

```
      PROGRAM LOT
      DIMENSION W(199),ROI(199)
  100 FORMAT(3F9.5,F10.7)
  101 FORMAT(10F8.6)
  102 FORMAT(I6,F6.3,I6,3F6.2)
      READ100,DT,ROW,DRO,PI
      READ101,W
   12 READ102,N,RR,NO,F90,F95,F99
      IF(N)200,200,300
  200 STOP
  300 CONTINUE
      AN=N
      AN1=AN-1.
      AN2=(AN-2.)/2.
      C=AN2*(1.-RR*RR)**AN2/(PI*PI)
      X90=SQRTF((1.+RR)*2.*F90)
      Y90=SQRTF((1.-RR)*2.*F90)
      X95=SQRTF((1.+RR)*2.*F95)
      Y95=SQRTF((1.-RR)*2.*F95)
      X99=SQRTF((1.+RR)*2.*F99)
      Y99=SQRTF((1.-RR)*2.*F99)
      M=1
      B90=0.
      B95=0.
      B99=0.
      C90=0.
      C95=0.
```

```
  C99=0.
  DO2I=1,NO
  II=2*I-1
  AI=I-1
  T1=AI*DT
  DO2J=1,II
  AJ=J-1
  T2=AJ*DT
  BT=SQRTF((1.+T1*T1/AN1)*(1.+T2*T2/AN1))
  B=BT**(-AN)
  R=(RR+T1*T2/AN1)/BT
  M=M+1
  D=0.
  DO4K=1,199
  AK=K
  RO=ROW+AK*DRO
  A1=RO*R
  A2=1.-A1*A1
  A3=SQRTF(A2)
  A4=A2*A3
  IF(A1)5,6,7
5 A5=ATANF(A3/(-A1))
  GO TO 8
6 A5=PI/2.
  GO TO 8
7 A5=PI+ATANF(A3/(-A1))
8 P1=A5/A3
  P2=1./A2+A1*A5/A4
  DO9L=3,N
```

FTN4.10

```
      AL=L
      P3=((AL-2.)*P1+(2.*AL-3.)*A1*P2)/((AL-
     1.)*A2)
      P0=P1
      P1=P2
    9 P2=P3
      IF(M-2)10,10,4
   10 ROI(K)=(1.-RO*RO)**AN2*(P0/P1+A1)
    4 D=D+C*B*W(K)*ROI(K)*P2
      QF90=(T1*T1-2.*RR*T1*T2+T2*T2)/(1.-RR
     *RR)-2.*F90
      QF95=(T1*T1-2.*RR*T1*T2+T2*T2)/(1.-RR
     *RR)-2.*F95
      QF99=(T1*T1-2.*RR*T1*T2+T2*T2)/(1.-RR
     *RR)-2.*F99
      IF(QF90)24,24,25
   24 IF(T1)26,27,26
   27 A=D
      GO TO 2
   26 Z=T1*T1-T2*T2
      IF(Z)28,29,28
   29 B90=B90+D
      GO TO 2
   28 C90=C90+D
      GO TO 2
   25 IF(QF95)34,34,35
   34 Z1=T1*T1-T2*T2
      IF(Z1)38,39,38
   39 B95=B95+D
      GO TO 2
   38 C95=C95+D
```

```
      GO TO 2
   35 IF(QF99)44,44,2
   44 Z2=(T1*T1-T2*T2)
      IF(Z2)48,49,48
   49 B99=B99+D
      GO TO 2
   48 C99=C99+D
    2 CONTINUE
      D190=(A+2.*B90+4.*C90)*0.04
      D195=D190+(2.*B95+4.*C95)*0.04
      D199=D195+(2.*B99+4.*C99)*0.04
  400 FORMAT(1X,29HELLIPSE T1R(-1)T=2F(2,N-1)=
              2*,F10.2)
  401 FORMAT(1X,6HAXES  ,2F10.4)
  402 FORMAT(1X,11HSAMPLE SIZE,I6,18HSAMPLE CORR-
              ELATION,F10.4)
  403 FORMAT(1X,10HINTEGRAL =,F10.6//)
      PRINT400,F90
      PRINT401,X90,Y90
      PRINT402,N,RR
      PRINT403,D190
      PRINT400,F95
      PRINT401,X95,Y95
      PRINT402,N,RR
      PRINT403,D195
      PRINT400,F99
      PRINT401,X99,Y99
      PRINT402,N,RR
      PRINT403,D199
      GO TO 12
      END
```

# BIBLIOGRAPHY

At one time or another, reference has been made to all of the papers Fisher wrote on the subject of fiducial inference but only the more important publications are given below since a complete list would be unduly lengthy.


Bayes, T. Phil. Trans. Roy. Soc. 53, 370-418 (1763)

Reprinted: Biometrika.

Bennett, G.W. C.S.I.R.O. Div. of Math. Stats. Tech. Paper 18 (1964)

Bennett, G.W. and Cornish, E.A. Bull. I.S.I. 40.2, 902-919 (1964)

Cornish, E.A. Aust. J. Statist. 2.1, 32-40 (1960)

C.S.I.R.O. Div. of Math. Stats. Tech. Paper 6 (1960)

C.S.I.R.O. Div. of Math. Stats. Tech. Paper 8 (1961)

C.S.I.R.O. Div. of Math. Stats. Tech. Paper 11 (1962)

Deemer, W.L. and Olkin, I. Biometrika, 38, 345-367 (1951)

Fisher, R.A. Phil. Trans. Roy. Soc. A, 222, 309-368 (1921)

Proc. Camb. Phil. Soc. 22, 700-725 (1925)

Proc. Camb. Phil. Soc. 26, 528-535 (1930)

Annals of Eugenics 6, 391-398 (1935)

Sankhyā, 7, 129-132 (1945)

Annales de l'Inst. Henri Poincare, 10, 191-213 (1948)

J.R.S.S.B, 17, 69-78 (1955)

"Statistical Methods and Scientific Inference"

Oliver and Boyd (1956,1959)

Sankhyā, 23.1, 3-8 (1961)

Fraser, D.A.S. Biometrika 48.3.4, 261-280 (1961)

Ann. Math. Stats. 32.3, 661-676 (1961)

Lindley, D.V. J.R.S.S.,B, 20, 102-107 (1958)

Mauldon, J.G. J.R.S.S.,B, 17, 79-85 (1955)

Sprott, D.A. J.R.S.S.,B, 22.2, 312-318 (1960)

Ann. Math. Stats. 32.2, 616-618 (1961)

J.R.S.S.,B, 23.2, 460-468 (1961)