SOME PROBLEMS OF INFERENCE IN TWO-DIMENSIONAL DISTRIBUTIONS

by

Dorothy Ann Anderson

A thesis submitted for the degree of

Doctor of Philosophy

of the

Australian National University

June, 1979

SOME PROBLEMS OF INFERENCE IN TWO-DIMENSIONAL DISTRIBUTIONS

by

Dorothy Ann Anderson

A thesis submitted for the degree of

Doctor of Philosophy

of the

Australian National University

June, 1979

To

B.T.

CONTENTS

PREFACE

I would like to thank my parents for their support and encouragement which enabled me to obtain a university education.

I am grateful to the Australian Government for the financial support of a Post-Graduate Research Award, and also to the Australian National University for their financial support.

I wish to thank all members of the Department of Statistics, Institute of Advanced Studies, Australian National University for providing a stimulating and friendly atmosphere in which to work. In particular I thank my supervisor, Professor P.A.P. Moran for his continued advice and interest. He suggested several of the problems and introduced me to a broader outlook on statistics. I would also like to thank Mrs B. Cranston for practical advice and cheerful encouragement. Finally I wish to express my gratitude to Mrs B. Geary for her excellent typing of this thesis.

* * * * * * * * *

Chapter 2 is an expanded and revised version of a paper that has been published (Anderson (1977)). An earlier version of chapter 4 has been submitted for publication and other work in this thesis is in preparation for submission.

Where I have consciously used results from other authors, the appropriate references have been made, except where the material can be regarded as part of the folklore of the subject. Very little research is done entirely in isolation, and I have asked for, and been given advice and suggestions on lines of approach. However none of the work has been of a collaborative nature. With these reservations, and to the best of my knowledge, this thesis is the product of my own research.

ABSTRACT

This thesis is concerned with problems of, and associated with, inference in some particular two-dimensional distributions. This includes obtaining properties of parameter estimators in multiparametric and multi-dimensional distributions which is generally a very messy procedure. Since the theory of how to obtain maximum likelihood estimators and their asymptotic properties is well developed, and because these asymptotic properties are deemed desirable for inferential procedures, maximum likelihood estimators should be examined, if only to enable comparisons to be made with estimators that are easier to calculate. Although the theory of maximum likelihood estimation is fairly well developed, the practical difficulties this can entail tend to be glossed over. Obtaining properties of other estimators can be even more difficult. Often only asymptotic moments, if that, can be obtained and no distributional properties. Here a variety of methods are used to extract as much information as possible about the parameter estimators.

Problems associated with inference are not confined to obtaining properties of parameter estimators. It may be necessary to verify that a particular estimation scheme, such as maximum likelihood, can be used for a particular problem. In addition if such an estimation scheme is inappropriate or unwieldy, some other approach may have to be considered. As an exploratory step it may be desirable to establish distributional or probabilistic properties of the distribution involved, and some work of this kind is done in this thesis.

GLOSSARY

It is difficult if not impossible to find a uniquely defined notation for every parameter, variable and expression, while retaining conventions regarding the use of specific notation for certain parameters. For example, conventionally ρ is used to represent the correlation between two variables but it is also used to represent the radius of a circle. In this thesis there has been notational duplication of this type. However a fairly unified notation has been attempted for the two chapter groups, chapters 2 and 3 and chapters 4, 5 and 6. Generally Greek letters denote parameters, while Latin letters denote variables or sample values. Expressions can be denoted by either. The glossary does not include well understood statistical notational abbreviations, such as $E(\cdot)$ to denote expectation of a variable, $\overset{d}{\sim}$ to denote "distributed as", $R[0, \lambda]$ to represent a random variable whose probability density function is uniform on the interval $[0, \lambda]$, or $N(\theta, \sigma^2)$ to represent a random variable which has a normal distribution with mean θ and variance σ^2.

(i) Chapters 2 and 3

A, B	standard normal variables
ARE(Q, Q^*)	asymptotic relative efficiency of Q relative to Q^*
c_1, c_2, c_3, c_4, c_5	integrals defined by (2.11), (2.12), (2.13), (2.14) and (2.15) respectively
c_6	$[\mathrm{ARE}(w, r)]^{\frac{1}{4}}$
d_1, d_2	defined (2.16)
f	probability distribution function
F	cumulative distribution function
g	defined (2.1)
h	interval width for numerical integration
l	log-likelihood function
$m(\theta)$	marginal information of θ
n	size of sample of (X_i, Y_i) values, or arbitrary positive integer
r	sample correlation coefficient of A and B
r^*	Kowalski and Tarter's (1969) approximation to r

r_F, r_K, r_S	Fisher-Yates, Kendall's and Spearman's rank correlation coefficient respectively
T	defined (2.1)
$V(\theta, \rho)$	$\mathrm{Var}(\hat{\theta})/\mathrm{Var}(\overline{\theta})$
$v(\alpha, \beta)$	$= \lim_{\rho \to 0} \nu(\rho\alpha, \rho\beta)$
w	sample correlation coefficient of X and Y
X, Y	random variables
α	$= \int_0^1 \{T'(s)\}^2 ds$ - chapter 2 only
α, β	shape parameters of X and Y respectively in §3.2 - §3.6
$\Gamma(s)$	gamma function
$\Pi\Gamma(\alpha, \beta; \omega)$	set of bivariate distributions defined in §3.2
δ	position constant for numerical integration
θ	Chapter 2 - parameter of distribution of X Chapter 3 - parameter of association between X and Y
$\overline{\theta}$, $\overline{\lambda}$	marginal maximum likelihood estimates of θ and λ
$\hat{\theta}$, $\hat{\lambda}$	maximum likelihood estimates of θ and λ
λ	parameter of distribution of Y
$\nu(\alpha, \beta)$	maximum possible correlation between X and Y in chapter 3
ρ	correlation coefficient of A and B
$\overline{\rho}$	estimate of ρ obtained by substituting $\overline{\theta}$ and $\overline{\lambda}$ into (2.3)
$\hat{\rho}$	maximum likelihood estimate of ρ
$\Phi(\cdot)$	standard normal cumulative distribution function
$\Phi'(x)$	derivative of $\Phi(x)$ with respect to x
$\chi(s, t)$	characteristic function of X and Y
ψ	defined (2.1)
ω	correlation coefficient of X and Y
$\approx$	approximately
$\simeq$	asymptotically or approximately, in context

(ii) Chapters 4, 5 and 6

A	$= \sigma^2/\rho^2 = \lambda^{-2} = \tfrac{1}{2}\gamma^{-1}$
a, b	major and minor axis of ellipse in §4.14

AML1, AML2	approximate maximum likelihood 1 and 2 respectively
FR1, FR2	approximate functional relationship 1 and 2 respectively
G, $G(\tau)$	cumulative distribution function of τ
G_0	true distribution of τ for §4.3
$\hat{G}$	maximum likelihood estimator of G
$I_\nu(z)$	modified Bessel function of the first kind and order ν
$I_{a,b}(z)$	$I_a(z)/I_b(z)$
k	eccentricity of ellipse in §4.14
$k_j(\tilde{\mu})$	jth cumulant of "sample" distribution of $\tilde{\mu}$, derived from the M replications of the estimation of $\tilde{\mu}$ (from samples of size N) $k_1(\tilde{\mu})$ is the "sample" mean, $k_2(\tilde{\mu})$ is the "sample" variance
L.S.	Least Squares
m	known degrees of freedom of non-central chi distribution
M	number of replications of estimation of an estimator
N	sample size, sample size on which estimator is based
p	radial length of ellipse
P.L.S.	Perturbed Least Squares
R	set of real numbers
$R_1, \ldots, R_4$	rectangles in the (γ_1, γ_2) plane
r	used for r_i where meaning is clear
r_i	$\begin{cases} \left[(X_i-\xi)^2+(Y_i-\eta)^2\right]^{\frac{1}{2}} & \text{- chapters 4 and 6} \\ \left[\sum_{j=1}^{m} (X_{ij}-\xi_j)^2\right]^{\frac{1}{2}} & \text{- chapter 5} \end{cases}$
$\bar{r}$	$N^{-1} \sum_{i=1}^{N} r_i$
r_i^*	$\left[(X_i-\xi^*)^2+(Y_i-\eta^*)^2\right]^{\frac{1}{2}}$

$sd(\tilde{\mu})$	standard deviation of $\tilde{\mu}$
T	$[0, 2\pi)$
$\overline{T}$	$[0, 2\pi]$
t, t_i	angular variate, especially in §4.14
X, Y	random variables
$\overline{X}, \overline{Y}$	$N^{-1} \sum_{i=1}^{N} X_i$ and $N^{-1} \sum_{i=1}^{N} Y_i$ respectively
X_i	$\begin{cases} \text{univariate random variable - chapters 4 and 6} \\ (X_{i1}, X_{i2}, \ldots, X_{im}) \quad \text{- chapter 5} \end{cases}$
$\overline{X}_j$	$N^{-1} \sum_{i=1}^{N} X_{ij}$ - chapter 5
α	probability of a (γ_1, γ_2) point falling outside a specified rectangle
α'	$\frac{1}{2}$ probability that a γ_1 or γ_2 point falls outside a specified interval
α_i	defined by $\cos \alpha_i = (X_i - \xi)/r_i$
β	defined by (4.21) or (5.5), function of A appearing in the variance - covariance matrix of $(\hat{\rho}, \hat{\sigma}^2)$
γ	$= \frac{1}{2}\rho^2/\sigma^2 = \frac{1}{2}\lambda^2 = \frac{1}{2}A^{-1}$
$\gamma_1(\tilde{\mu})$	$= k_3(\tilde{\mu})/k_2(\tilde{\mu})^{\frac{3}{2}}$ measure of skewness
$\gamma_2(\tilde{\mu})$	$= k_4(\tilde{\mu})/k_2(\tilde{\mu})^2$ measure of kurtosis
Γ	$= \{G\}$ a given space of G, for §4.3
δ	defined (4.22) or (5.10), function of A appearing in the variance of $\hat{\xi}$
$\delta(\cdot, \cdot)$	metric defined by (4.5)
Δ	$= (\beta/\rho^2 - 1)(\frac{1}{2}m + A^{-1}) - 1$, $m = 2$ in chapter 4
$\{(\varepsilon_i, \nu_i)\}$	sequence of independent bivariate random variables with zero means and finite variances
η	y-coordinate of the centre of the circle
$\eta^*, \hat{\eta}$	estimate of η defined in context, maximum likelihood estimate of η
η_a, η_c, η_T	estimate of η due to Angell & Barber, Chan and Thom, respectively

θ	$= (\xi, \eta, \rho, \sigma^2)$ vector of structural parameters
θ_0	true value of θ for §4.3
$\hat{\theta}$	$= (\hat{\xi}, \hat{\eta}, \hat{\rho}, \hat{\sigma}^2)$, maximum likelihood estimate of θ
κ	$= (\theta, G)$ is the generic point in the space $\Omega \times \Gamma$ in §4.3; parameter in von Mises distribution in §4.13
κ_0	$= (\theta_0, G_0)$
$\hat{\kappa}$	maximum likelihood estimate of κ
κ_j	jth cumulant of distribution of estimator. κ_1 is the mean, κ_2 the variance
λ^2	$= \rho^2/\sigma^2$, noncentrality parameter of non-central chi distribution
$\hat{\lambda}$	maximum likelihood estimator of λ
μ	parameter, in §4.13 location parameter of a von Mises distribution
$\mu^*, \tilde{\mu}, \hat{\mu}$	estimators of μ defined in context, maximum likelihood estimator of μ
ξ	x-coordinate of centre of circle
ξ_j	jth coordinate of centre of circle $j = 1, 2, \ldots, m$
$\xi^*, \hat{\xi}, \xi_a, \xi_c, \xi_T$	see $\eta^*, \hat{\eta}, \eta_a, \eta_c, \eta_T$
ρ	radius of circle
$\hat{\rho}, \rho_a, \rho_c, \rho_T$	see $\hat{\eta}, \eta_a, \eta_c, \eta_T$
ρ_f	estimator of ρ using functional relationship equation
ρ_a^*	$= N^{-1} \sum_{i=1}^{N} \left[(X_i-\xi_a)^2+(Y_i-\eta_a)^2\right]^{\frac{1}{2}}$
ρ_1, ρ_2	estimators of ρ based on moments of r
ρ_1^*, ρ_2^*	defined using r_i^* instead of r_i in ρ_1 and ρ_2
ρ_3, ρ_4	estimators based on distance between (X_i, Y_i) and (X_j, Y_j)
σ^2	variance of ε_i and ν_i now considered to be normal variables

$\hat{\sigma}^2$, σ^{*2}, σ_a^2, σ_c^2, σ_T^2	see $\hat{\eta}$, η^*, η_a, η_c, η_T
σ_4^2	estimator of σ^2 associated with ρ_4
σ_a^{*2}	$= N^{-1} \sum_{i=1}^{N} \left(r_i^2 - \rho_a^{*2}\right)$
τ_i	angle from x-axis to the expectation vector of (X_i, Y_i)
τ_i^*	functional relationship estimate of τ_i
$\hat{\tau}_i$	maximum likelihood estimate of τ_i
φ_j	$j = 1, 2, \ldots, m$ angular variates
φ	angular displacement of major axis of ellipse, in §4.14
$\chi_{(m)}$	chi random variable with m degrees of freedom
$\chi(s, t)$	characteristic function of X and Y
ψ_j	$j = 1, 2, \ldots, m$ angular variates
Ω	space of possible values of θ
$\bar{\Omega} \times \bar{\Gamma}$	completed space of $\Omega \times \Gamma$: also compact

CHAPTER 1

INTRODUCTION

This thesis is mainly concerned with problems of, and associated with, inference in some particular two-dimensional distributions. Making statistical inferences about parameters of a two-dimensional distribution is generally a very messy procedure. One tends to require simultaneous estimation of a number of parameters, and one needs to know the properties of these estimators. Often comparisons need to be made between sets of alternative estimators. For these comparisons such things as the bias, efficiency of estimation and ease of numerical calculation of the estimators are desired. In addition if the estimation equations are particularly messy the use of approximations may need to be considered. One often wants distributional and probabilistic properties of the two-dimensional random vector to aid estimation of the parameters or obtain a better understanding of the two-dimensional distribution. Also the study of the estimation of the parameters of a two-dimensional random variable can often shed light on parametric estimation in an associated one- or multi-dimensional random variable, and *vice versa*.

Consequently although this thesis is mainly concerned with statistical inference in two-dimensional distributions that is not the only topic considered here. In chapter 3 there is a discussion of infinite divisibility of bivariate gamma distributions; the motivation for which evolved from the work in chapter 2. Chapter 2 is concerned with estimation of the parameters of a particular class of bivariate distributions. One of the marginal distributions of this class is the exponential distribution, but the class does not include the gamma distribution. The work in chapter 3 eliminates a possible line of inquiry as to how the analysis in chapter 2 could be extended to the allied distribution with gamma marginals, and highlights some of the difficulties associated with this distribution. Also in chapter 5 there is a discussion of the estimation of the non-centrality and scale parameters of a non-central chi distribution with known degrees of freedom. This evolved from the work in chapter 4. The circular structural model is considered in chapter 4. Although estimation in the general case is discussed there, due to its intractability more emphasis is placed on a special case. This involves a random variable which has a non-central chi distribution with unknown scale parameter and two degrees of freedom. Chapter 5 also explores the m-dimensional hyper-spherical generalization of this circular model.

As indicated above, the circular structural model is examined in chapter 4. One practical application of the circular structural model is to stone circle data and this is considered in chapter 6. Some of the techniques actually used to estimate circular parameters are briefly mentioned. A circle fitting algorithm suggested for fitting stone circles is examined and compared with other estimation procedures.

When examining properties of estimators it is quite usual to be able to find only asymptotic properties, that is large sample properties. For example given certain regularity conditions a maximum likelihood estimator is asymptotically unbiased and, asymptotically, has a normal distribution with a specified variance. To find this variance we must be able to integrate the second derivative of the log-likelihood function. Techniques of numerical integration need to be employed if these integrals are not available from the literature. Such is the case in chapters 2 and 4. Numerical integration of integrals also occurs in other contexts. For example in chapter 3 numerical integration is used to describe the relationship between two variables.

Further, if other estimators of the parameters are available then by using the known efficiency properties of maximum likelihood estimators it may be possible to put a bound on the variance of the asymptotic distribution. This is explored in chapter 4.

However very often one might wish to make statistical inferences about parameter estimates that have been derived from a small sample. In a finite sample the distribution of the maximum likelihood estimator for example may not even have finite moments. Assuming the existence of moments we would like to know how large the sample size has to be before it can reasonably be assumed that the asymptotic properties of the estimator hold. A partial answer to this question can be obtained from simulation studies. In addition, for some of the estimators used no distributional properties can be derived, and only asymptotic moments can be found. We would like to know whether or not the distributions of any of these estimators can be assumed to be reasonably close to the normal distribution. Simulation studies can provide the answer to this question, as well as indicating whether or not any of the estimators have unexpected small sample properties. Some results of simulation studies carried out can be found in the second half of chapter 6.

Essentially only two different parametric estimation and inferential problems are considered in this thesis; a particular class of bivariate distributions, considered in chapters 2 and 3, and a circular structural model and allied models considered in chapters 4, 5 and 6. The problems associated with statistical inference are similar in both cases, and some of these problems have been described above. This thesis also illustrates the fact that a minor modification of a model can increase the mathematical complexity by an order of magnitude. This can turn a model from which a moderate amount of information can be extracted into one for which practically no information can be obtained.

CHAPTER 2

STATISTICAL INFERENCE FOR A CLASS OF BIVARIATE DISTRIBUTIONS

1. Introduction

One of the difficulties associated with bivariate distributions which adds greatly to their complexity is that they are not completely specified by their marginal distributions. The complete joint correlation structure or the method of construction of the joint distribution of the variates must also be specified. One such method of construction is the version of the translation method suggested by Nataf (1962). Moran (1969) used such a construction to study the statistical analysis of a bivariate gamma distribution, useful for its applications in meteorology and hydrology. It is shown here that if the conditions for maximum likelihood estimation and one further condition on the marginal cumulative distribution functions are satisfied, then quite general results can be obtained for the elements of the asymptotic covariance matrix for the vector parameter estimator found by maximum likelihood. This facilitates hypothesis testing involving the various parameters. The vector parameter estimator is consistent and, if additional conditions proposed by Wald (1949) are satisfied, has an asymptotic multivariate normal distribution. The further condition on the marginal cumulative distribution functions required for this discussion is always satisfied by distributions which have only location and scale parameters unknown.

Define the random variables X and Y to have respective cumulative distribution functions $F(X;\ \theta)$ and $F(Y;\ \lambda)$. For simplicity we restrict X and Y to have the same distributional form each with one unknown parameter, where these unknown parameters are of the same type. For example, they can both be location or scale parameters. Extensions to more parameters and differing marginal distributions involves no new principles only more algebra. Much work has been done in obtaining and studying the distributional properties of constructed bivariate distributions which have prescribed marginals. Johnson and Kotz (1972) relate many of these. Kimeldorf and Sampson (1975a) examined transformations to one parameter bivariate distributions such as the one used here and, (1975b), examined in particular the bivariate uniform distribution and concepts of bivariate dependence. In this chapter we will be more concerned with aspects of statistical analyses than distributional properties.

A description of the method of translation can be found in section 4.3 of Mardia (1970a), but for completeness is included here. Let A and B be a pair of random variables having a standard bivariate normal distribution with correlation ρ. Define random variables U and V by

$$U = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{A} \exp\left(-\tfrac{1}{2}t^2\right) dt = \Phi(A) ,$$

and $V = \Phi(B)$. X and Y can then be defined by

$$U = \int_{-\infty}^{X} f(x;\ \theta) dx = F(X;\ \theta) ,$$

and $V = F(Y;\ \lambda)$. Thus the joint probability density of X and Y is written;

$$f(X,\ Y) = \left(1-\rho^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\rho^2A^2-2\rho AB+\rho^2B^2}{2\left(1-\rho^2\right)}\right\} f(X;\ \theta) f(Y;\ \lambda) ,$$

where $A = \Phi^{-1}[F(X;\ \theta)]$, $B = \Phi^{-1}[F(Y;\ \lambda)]$ and Φ^{-1} is the inverse function of Φ.

For maximum likelihood estimation, $F(X;\ \theta)$ and $F(Y;\ \lambda)$ must satisfy certain conditions. We require that the range of the random variable X does not depend on the unknown parameter θ, and that F be twice differentiable with respect to θ and differentiable with respect to X. We will also require the existence of higher derivatives of f with respect to θ. Similar conditions apply to Y and $F(Y;\ \lambda)$. The condition on the marginal cumulative distribution functions is that

$$\begin{aligned} \partial F(X;\ \theta)/\partial\theta &= T(F) g(\theta) \\ &= \psi(A) g(\theta) , \end{aligned} \tag{2.1}$$

where $T(F)$ is a function of F alone, $\psi(A)$ is a function of A alone and $g(\theta)$ is a function of θ alone. The following extension is required when the F's have more than one unknown parameter. Let $\theta = \left(\theta_1,\ \theta_2,\ \ldots,\ \theta_k\right)$ be the vector of parameters for F, for some k; then

$$\partial F(X;\ \theta)/\partial\theta_i = \psi_i(A) g_i(\theta) , \tag{2.2}$$

must hold for $i = 1,\ 2,\ \ldots,\ k$. Notice that $g_i(\cdot)$ can be a function of

$\boldsymbol{\theta}$ and not just θ_i .

With the extension to more than one parameter given, and the corresponding extension of the conditions to maximum likelihood assumed, we can prove that (2.2) is satisfied for distributions with scale and location parameters as we will see in section 5. Initially it was the analysis of the bivariate gamma distribution of Moran (1969) that was in mind. The logical place to start was to consider an exponential marginal. Although the analysis turned out to have a much wider range of application than the exponential marginal, in its present form it is not immediately applicable to a gamma marginal. However the analysis is a first step towards that goal.

It should be emphasised that the normalizing transformation can be applied even when condition (2.1) does not hold. If (2.1) does hold then the integrals required to evaluate the elements of the asymptotic covariance matrix of the vector of parameter estimators involve only functions of A and B which do not involve the unknown parameter values except perhaps the correlation coefficient ρ . This simplifies the calculations.

2. Parameter Estimation

From (2.1),

$$\partial^2 F/\partial\theta^2 = \psi(A)g'(\theta) + \psi'(A)g(\theta)\,\partial A/\partial\theta ,$$

where

$$\partial A/\partial\theta = (2\pi)^{\frac{1}{2}}\exp\left(\tfrac{1}{2}A^2\right)\,\partial F/\partial\theta .$$

Hence

$$\partial^2 F/\partial\theta^2 = \psi(A)g'(\theta) + (2\pi)^{\frac{1}{2}}\exp\left(\tfrac{1}{2}A^2\right)\psi(A)\psi'(A)g^2(\theta) ,$$

and

$$\partial^2 A/\partial\theta^2 = 2\pi A\,\exp\left(A^2\right)(\partial F/\partial\theta)^2 + (2\pi)^{\frac{1}{2}}\exp\left(\tfrac{1}{2}A^2\right)\,\partial^2 F/\partial\theta^2 .$$

Given a sample of size n , $(X_1, Y_1), \ldots, (X_n, Y_n)$, if

$l = \sum_{i=1}^{n} \log f(X_i, Y_i)$ is the log-likelihood, and subscripts 1, 2 and 3 of l , A and B refer to partial differentiation with respect to ρ, θ and λ respectively, we obtain the following equations:

$$l_1 = n\rho\left(1-\rho^2\right)^{-1} - \left(1-\rho^2\right)^{-2} \sum_{i=1}^{n} \left\{\rho A^2-\left(1+\rho^2\right)AB+\rho B^2\right\} , \tag{2.3}$$

$$l_2 = -\left(1-\rho^2\right)^{-1} \sum_{i=1}^{n} \rho(\rho A-B)A_2 + \sum_{i=1}^{n} \partial \log f\left(X_i;\ \theta\right)/\partial\theta , \tag{2.4}$$

$$l_{11} = n\left(1+\rho^2\right)\left(1-\rho^2\right)^{-2}$$
$$- \left(1-\rho^2\right)^{-3} \sum_{i=1}^{n} \left\{\left(1+3\rho^2\right)A^2-2\rho\left(3+\rho^2\right)AB+\left(1+3\rho^2\right)B^2\right\} ; \tag{2.5}$$

$$l_{12} = - \sum_{i=1}^{n} \left(1-\rho^2\right)^{-2}\left(2\rho A-\left(1+\rho^2\right)B\right)A_2 , \tag{2.6}$$

$$l_{22} = - \sum_{i=1}^{n} \left(1-\rho^2\right)^{-1}\left\{\rho^2 AA_{22}+\rho^2A_2^2-\rho A_{22}B\right\} + \sum_{i=1}^{n} \partial^2 \log f\left(X_i;\ \theta\right)/\partial\theta^2 , \tag{2.7}$$

$$l_{23} = \sum_{i=1}^{n} \left(1-\rho^2\right)^{-1}\rho A_2B_3 , \tag{2.8}$$

and similarly for l_3, l_{13} and l_{33}.

Since $E\left(l_1\right) = E\left(l_2\right) = E\left(l_3\right) = 0$, we easily obtain some identities, for example

$$E\left(\rho^2AA_2\right) = E\left(\rho A_2B\right) . \tag{2.9}$$

Remembering that A_{22} is a function of A and not of B or λ or Y, and that A and B have a bivariate normal distribution, we find that

$$\begin{aligned} E\left(\rho BA_{22}\right) &= E\left[\rho B\left\{2\pi Ae^{A^2}\left(\frac{\partial F}{\partial\theta}\right)^2+(2\pi)^{\frac{1}{2}}e^{\frac{1}{2}A^2}\ \frac{\partial^2F}{\partial\theta^2}\right\}\right] \\ &= E_A\left\{A_{22}E(\rho B\,|A)\right\} . \end{aligned}$$

Since A_{22} does not involve explicit functions of B, this is equal to

$$E_A\left(\rho^2AA_{22}\right) . \tag{2.10}$$

The actual evaluation of the estimates will have to be done iteratively. A reasonable starting point might be $(\bar\rho, \bar\theta, \bar\lambda)$, where $\bar\theta$ and $\bar\lambda$ are the marginal maximum likelihood estimates and $\bar\rho$ the estimate obtained by substituting $\bar\theta$ and $\bar\lambda$ into (2.3), also using $\bar\theta$ and $\bar\lambda$ to get the A's and B's . If the log-likelihood has several local maxima, which could be possible for some marginal distributions, care is needed to ensure proper convergence. $\bar{\theta}$ and $\bar{\lambda}$ should be fairly close to $\hat{\theta}$ and $\hat{\lambda}$, especially if ρ is near zero.

Since A and B have a bivariate normal distribution,

$$E\left(-l_{11}\right) = n\left(1+\rho^2\right)\left(1-\rho^2\right)^{-2} .$$

Now from (2.6) using identity (2.9),

$$E\left(-l_{12}\right) = n\rho\left(1-\rho^2\right)^{-1}E\left(AA_2\right) ,$$

where

$$\begin{aligned} E\left(AA_2\right) &= E\left\{A\psi(A)g(\theta)(2\pi)^{\frac{1}{2}}\exp\left(\tfrac{1}{2}A^2\right)\right\} \\ &= g(\theta)\int_{-\infty}^{\infty} A\psi(A)dA . \end{aligned}$$

Thus

$$E\left(-l_{12}\right) = n\rho\left(1-\rho^2\right)^{-1}g(\theta)c_1 ,$$

where

$$c_1 = \int_{-\infty}^{\infty} A\psi(A)dA .$$

Similarly,

$$E\left(-l_{13}\right) = n\rho\left(1-\rho^2\right)^{-1}g(\lambda)c_1 . \qquad (2.11)$$

c_1 is a univariate integral involving only A and functions of A . It does not involve the parameters of the bivariate distribution except through A , due to the condition imposed on $\partial F/\partial\theta$. From (2.7),

$$E\left(-l_{22}\right) = n\left(1-\rho^2\right)^{-1}E\left\{\rho^2AA_{22}+\rho^2A_2^2-\rho A_{22}B\right\} + nE\left\{-\partial^2 \ln f(X;\ \theta)/\partial\theta^2\right\} .$$

Define $m(\theta) = E\{-\partial^2 \ln f(X;\theta)/\partial\theta^2\}$, the marginal information. Using the assumed existence of the second derivative of $f(X;\ \theta)$ with respect to θ , it can be shown that

$$m(\theta) = \int \frac{1}{f}\left[\frac{d}{dx}\left(\frac{\partial F}{\partial\theta}\right)\right]^2 dx ,$$

$$\frac{d}{dx}\left(\frac{\partial F}{\partial \theta}\right) = \frac{dA}{dx} \cdot \frac{d}{dA}\left(\frac{\partial F}{\partial \theta}\right)$$
$$= g(\theta)\Phi'(A)\,\frac{dA}{dx}\left\{\frac{d}{d\Phi(A)}\,T[\Phi(A)]\right\} ,$$

using (2.1).

$$m(\theta) = g^2(\theta)\int_{-\infty}^{\infty} [T'\{\Phi(A)\}]^2\Phi'(A)dA ,$$

since

$$f(x;\ \theta)dx = \Phi'(A)dA .$$

Thus

$$m(\theta) = g^2(\theta)\int_0^1 \{T'(s)\}^2 ds = \alpha g^2(\theta)$$

say. Then from (2.7), using (2.10),

$$E\left(-l_{22}\right) = n\left(1-\rho^2\right)^{-1}E\left(\rho^2 A_2^2\right) + n\alpha g^2(\theta) ,$$

where we assume that $E\left(AA_{22}\right)$ is finite. As before we can show that

$$E\left(AA_{22}\right) = g^2(\theta)\{c_2+c_3\} + g'(\theta)c_1 ,$$

where

$$c_2 = (2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty} A\psi(A)\psi'(A)\exp\left(\tfrac{1}{2}A^2\right)dA , \tag{2.12}$$

and

$$c_3 = (2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty} A^2\psi^2(A)\exp\left(\tfrac{1}{2}A^2\right)dA , \tag{2.13}$$

so we assume that c_2 and c_3 are finite. We can show that

$$E\left(\rho^2 A_2^2\right) = \rho^2 g^2(\theta)c_4 ,$$

where

$$c_4 = (2\pi)^{\frac{1}{2}}\int_{-\infty}^{\infty} \psi^2(A)\exp\left(\tfrac{1}{2}A^2\right)dA . \tag{2.14}$$

Therefore

$$E\left(-l_{22}\right) = ng^2(\theta)\left\{\rho^2\left(1-\rho^2\right)^{-1}c_4+\alpha\right\} ,$$

and

$$E\left(-l_{33}\right) = ng^2(\lambda)\left\{\rho^2\left(1-\rho^2\right)^{-1}c_4+\alpha\right\} .$$

All the expectations so far involve integrals independent of ρ, θ and λ. Unfortunately although the integral involved in $E\left(A_2B_3\right)$ is independent of θ and λ, it does depend on ρ. We have for (2.8), using (2.1),

$$E\left(A_2B_3\right) = g(\theta)g(\lambda)E\left\{2\pi \exp\left(\tfrac{1}{2}A^2+\tfrac{1}{2}B^2\right)\psi(A)\psi(B)\right\} .$$

Write

$$c_5 = \left(1-\rho^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(A)dA \int_{-\infty}^{\infty} \psi(B)\exp\left\{-\frac{\rho^2A^2-2\rho AB+\rho^2B^2}{2\left(1-\rho^2\right)}\right\}dB , \qquad (2.15)$$

then

$$E\left(-l_{23}\right) = -n\rho\left(1-\rho^2\right)^{-1}g(\theta)g(\lambda)c_5 .$$

We then obtain the matrix

$$D = \begin{bmatrix} \left(1+\rho^2\right)\left(1-\rho^2\right)^{-1} & \rho g(\theta)c_1 & \rho g(\lambda)c_1 \\ \rho g(\theta)c_1 & g^2(\theta)\left\{\rho^2c_4+\left(1-\rho^2\right)\alpha\right\} & -\rho g(\theta)g(\lambda)c_5 \\ \rho g(\lambda)c_1 & -\rho g(\theta)g(\lambda)c_5 & g^2(\lambda)\left\{\rho^2c_4+\left(1-\rho^2\right)\alpha\right\} \end{bmatrix} .$$

Let

$$d_1 = \rho^2c_4 + \left(1-\rho^2\right)\alpha - \rho c_5 \quad \text{and} \quad d_2 = \rho^2c_4 + \left(1-\rho^2\right)\alpha + \rho c_5 . \qquad (2.16)$$

The asymptotic variance-covariance matrix of $(\hat{\rho}, \hat{\theta}, \hat{\lambda})$ is $\left(1-\rho^2\right)D^{-1}$, and this gives

$$\mathrm{Var}(\hat{\rho}) \simeq \frac{\left(1-\rho^2\right)^2}{n\left(1+\rho^2\right)}\left[1 - \frac{2\rho^2\left(1-\rho^2\right)c_1^2}{\left(1+\rho^2\right)d_1}\right]^{-1} .$$

Therefore we have a multiplying factor of

$$\left[1-2\rho^2\left(1-\rho^2\right)c_1^2/\left\{\left(1+\rho^2\right)d_1\right\}\right]^{-1}$$

which multiplies the asymptotic variance that $\hat{\rho}$ would have, were θ and λ known. Similarly

$$\text{Var}(\hat{\theta}) \simeq \frac{\alpha}{nm(\theta)} \left(1-\rho^2\right) \left[\frac{\left\{\left(1+\rho^2\right)\left(d_2-\rho c_5\right)-\rho^2\left(1-\rho^2\right)c_1^2\right\}}{d_2\left\{\left(1+\rho^2\right)d_1-2\rho^2\left(1-\rho^2\right)c_1^2\right\}}\right] ,$$

where $\left(nm(\theta)\right)^{-1}$ is the asymptotic variance of the marginal estimator $\overline{\theta}$ obtained by maximum likelihood using only the X values. The ratio of these asymptotic variances gives an indication of how much information about θ is contained in the marginal sample distribution of X.

Next we have

$$\text{cov}(\hat{\theta}, \hat{\rho}) \simeq \frac{\left(1-\rho^2\right)^2}{ng(\theta)} \left[\frac{\rho c_1}{\left\{\left(1+\rho^2\right)d_1-2\rho^2\left(1-\rho^2\right)c_1^2\right\}}\right]$$

and

$$\text{cov}(\hat{\theta}, \hat{\lambda}) \simeq \frac{\left(1-\rho^2\right)}{ng(\theta)g(\lambda)} \left[\frac{\left\{\rho\left(1+\rho^2\right)c_5+\rho^2\left(1-\rho^2\right)c_1^2\right\}}{d_2\left\{\left(1+\rho^2\right)d_1-2\rho^2\left(1-\rho^2\right)c_1^2\right\}}\right] .$$

3. Results for the Correlation Coefficients

Let ω be the correlation coefficient of X and Y. Fréchet (1951) considered bounds for ω, for given marginals. Let

$$F^+(X, Y) = \min\{F(X; \theta), F(Y; \lambda)\}$$

and

$$F^-(X, Y) = \max\{F(X; \theta)+F(Y; \lambda)-1, 0\} ,$$

with ω^+ and ω^- the respective correlation coefficients, then

$$\omega^- \leq \omega \leq \omega^+ .$$

Whitt (1976) also attributes these bounds to Hoeffding (1940). Mardia (1970b) showed that these bounds are attainable in our case.

To test the hypothesis $\rho = 0$, there are many statistics that can be considered. Let $\hat{\rho}$ be the maximum likelihood estimate of ρ, r the sample correlation coefficient between the A and B values, and w the sample correlation coefficient between the X and Y values. Other estimators of correlation that can be considered are the three rank correlation coefficients r_S, r_K and r_F, that is, Spearman's, Kendall's and the Fisher-Yates correlation coefficient. The definitions of these can be found in Fieller, Hartley and Pearson (1957) for example. These three statistics rely on n associated pairs obtained by ranking the A_i's in ascending order and then considering the pairs (j, v_j) where j is the rank of A_i and v_j the rank of B_i. We lose some information in using ranks so we expect some reduction in the power of such tests. Fieller *et al* (1957) pointed out that since we have applied monotonic transformations from X_i to A_i and from Y_i to B_i, ranking the A and B random variables is equivalent to ranking the X and Y random variables.

In order to get the power of such tests we require the notion of asymptotic relative efficiency (ARE) in the Pitman sense (see for example, Fraser [(1957), p. 273]). If Q and Q^* are two test statistics, to test $H_0 : \rho = \rho_0$, the ARE is defined as the limit of the reciprocal of the sample sizes required for both tests to attain the same power, and this is independent of the power chosen if both test statistics are asymptotically normally distributed for all ρ near ρ_0. If Q_n and Q_n^* are two sequences of test statistics satisfying certain regularity conditions, which ensure the required asymptotic normal distribution, it can be shown that the ARE of Q relative to Q^* is

$$\text{ARE}(Q, Q^*) = \lim_{n\to\infty} \left[\frac{dE(Q_n)/d\rho \mid \rho = \rho_0}{dE(Q_n^*)/d\rho \mid \rho = \rho_0}\right]^2 \frac{\text{Var}(Q_n^*) \mid \rho = \rho_0}{\text{Var}(Q_n) \mid \rho = \rho_0},$$

where Q and Q^* are the limiting test statistics as $n \to \infty$. Stuart (1954) gave the relationship between the ARE of two tests and the asymptotic ratio of the derivatives of the power function. As Fieller *et al* (1957) remarked, the appropriate sample correlation coefficient is r. Lancaster (1957) and Maung (1941) proved that $|\rho| \geq |\omega|$, so the appropriate hypothesis to test independence is $H_0 : \rho = 0$ and not $H_0 : \omega = 0$. For tests of independence between arbitrary random variables one problem that arises is finding a suitable model for the alternative hypothesis. However for the bivariate normal this problem does not arise since we have the natural model

$$A = \left(1-\rho^2\right)^{\frac{1}{2}} z_1 + \rho z_2 \quad \text{and} \quad B = z_2$$

where z_1 and z_2 are independent normal random variables. The ARE's of the rank correlation coefficients relative to r can be found from the literature. That is,

$$\mathrm{ARE}\left(r_S,\ r\right) = \mathrm{ARE}\left(r_K,\ r\right) = 9/\pi^2 \simeq 0.91\ ,\ \text{at}\ \ \rho = 0\ ,$$

and from Bhuchongkul (1964) who stated it explicitly and Konijn (1956),

$$\mathrm{ARE}\left(r_F,\ r\right) = 1\ ,\ \text{at}\ \ \rho = 0\ .$$

The complicating factor is that r is unobtainable unless θ and λ are known, since it requires the actual sample values of A and B, where the means and variances of A and B are not assumed to be known. Kowalski and Tarter (1969) defined a quantity r^* as an approximation to r. They used an approximation for Φ^{-1} and an estimation procedure to find the sample marginal cumulative distribution functions to obtain estimated A and B values. r^* is the correlation between these estimated A and B values. They compared r, r^* and w, and for certain simulated cases the power functions of r and r^* were shown to be almost identical and generally higher than that for w. For the bivariate case considered here, evaluation of r^* for large n would involve a large amount of work and the rank correlation coefficients perform almost as well.

From Ghosh (1966) we obtain

$$E(r|\rho) = b_n\rho\ {}_2F_1\left[\tfrac{1}{2},\ \tfrac{1}{2};\ \tfrac{1}{2}(n+1);\ \rho^2\right]\ ,$$

where

$$b_n = 2(n-1)^{-1}\left\{\Gamma(n/2)/\Gamma\left((n-1)/2\right)\right\}^2\ ,$$

and

$${}_2F_1[a,\ b;\ c;\ z] = \frac{\Gamma(c)}{\Gamma(b)\Gamma(a)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!}$$

is the hypergeometric function. From Slater [(1966), eq. 1.4.1.1],

$$d\left\{{}_2F_1[a,\ b;\ c;\ z]\right\}/dz = abc^{-1}\ {}_2F_1[a+1,\ b+1;\ c+1;\ z]$$

and

$${}_2F_1[\cdot,\ \cdot;\ \cdot;\ 0] = 1\ ,$$

and therefore

$$\partial E(r|\rho)/\partial\rho = b_n \; {}_2F_1\left[\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}(n+1); \rho^2\right] + b_n\rho^2(n+1)^{-1} \; {}_2F_1\left[\tfrac{3}{2}, \tfrac{3}{2}; \tfrac{1}{2}(n+3); \rho^2\right] .$$

Consequently

$$\partial E(r|\rho)/\partial\rho = b_n \quad \text{at} \quad \rho = 0 .$$

Using Stirling's approximation for $\Gamma(n)$ and $\lim_{n\to\infty} \left(1-n^{-1}\right)^n = e^{-1}$, we can show that $\lim_{n\to\infty} b_n = 1$. Also at $\rho = 0$,

$$\mathrm{Var}(r) = (n-1)^{-1} .$$

Consider w , the sample correlation coefficient of the X and Y values. From Pitman (1937), $\mathrm{Var}(w|\rho = 0) = (n-1)^{-1}$, assuming the existence of ω . Cook (1951) and Gayen (1951) showed that $E(w)$ can be expressed in the form

$$E(w) = \omega + n^{-1}\left\{-\tfrac{1}{2}\omega\left(1-\omega^2\right)+b\right\} .$$

We would then expect that

$$\partial E(w)/\partial\rho = \partial\omega/\partial\rho \left\{1+O\left(n^{-1}\right)\right\} .$$

But first we must ensure that $\partial b/\partial\omega = O(1)$. In fact b is a sum of terms which are products and cumulants of the bivariate population (X, Y) , divided by non-negative powers of n . Since these population cumulants do not involve n their derivatives with respect to ω can not introduce any positive powers of n . This does assume that the bivariate population (X, Y) has cumulants of all orders. However the argument could probably be extended as long as w is a consistent estimator of ω .

Thus we show that

$$\mathrm{ARE}(w, r) = (\partial\omega/\partial\rho \mid \rho = 0)^2 = c_6^4 ,$$

say. We need ω as a function of ρ before we can proceed any further.

Consider $\hat{\rho}$, the maximum likelihood estimator of ρ . From Cox and Hinkley [(1974), pg. 310],

$$E(\hat{\rho}) = \rho + b_1 n^{-1} .$$

Here $\hat{\rho}$ is a consistent estimator of ρ as is r , and when evaluating the ARE of two consistent estimators of the parameter it is usual to use

$$\mathrm{ARE}(Q, Q^*) = \lim_{n\to\infty} \frac{\mathrm{Var}\left(Q_n^*\right) \mid \rho = \rho_0}{\mathrm{Var}\left(Q_n\right) \mid \rho = \rho_0} .$$

At $\rho = 0$, $\mathrm{Var}(\hat{\rho}) = n^{-1}$, so we can show that

$$\mathrm{ARE}(\hat{\rho}, r) = 1 .$$

Fieller *et al* (1957) and Fieller and Pearson (1961) also considered Fisher's z-transformation,

$$z_a = \mathrm{artanh}\, r_a = \tfrac{1}{2} \log_e\left\{\left(1+r_a\right)/\left(1-r_a\right)\right\}$$

where r_a is a sample, or sample rank, correlation coefficient. They studied z_F, z_K and z_S empirically and showed that these z-transformations have variances almost independent of ρ . Since the z transformation has a remarkable normalizing property, z could also be used to test hypotheses involving values of ρ other than $\rho = 0$.

Kimeldorf and Sampson (1975b) claim that there are advantages in studying continuous bivariate distributions by means of their uniform translates, since bivariate uniform random variables are independent if and only if their joint probability distribution function is constant, so spikes in the probability distribution function can be interpreted as regions of local dependence. Of course again if θ and λ are unknown, we can not transform the marginal variables exactly and this technique could only be used if we had a large enough sample so that the marginal cumulative distribution functions were sufficiently smooth.

4. Bivariate Exponential

It is easy to show that the exponential distribution satisfies condition (2.1). We have $F(X; \theta) = 1 - \exp(-X/\theta)$, so that

$$\begin{aligned}\partial F(X; \theta)/\partial\theta &= -X\theta^{-2}\exp(-X/\theta)\\ &= +\theta^{-1}\{1-\Phi(A)\}\ln\{1-\Phi(A)\} ,\end{aligned}$$

since $X = -\theta \ln\{1-\Phi(A)\}$. In the notation of (2.1), $g(\theta) = +\theta^{-1}$ and $T\{\Phi(A)\} = \{1-\Phi(A)\}\ln\{1-\Phi(A)\}$. Thus we can show that $T'(x) = -1 - \ln(1-x)$, so that $\alpha = 1$, and writing Φ for $\Phi(A)$ for convenience, we get

$$c_1 = \int_{-\infty}^{\infty} A(1-\Phi)\ln(1-\Phi)dA = -0.2978 ,$$

$$c_2 = \int_{-\infty}^{\infty} A(1-\Phi)\ln(1-\Phi)\{1-\ln(1-\Phi)\}dA = -1.3535 < \infty ,$$

$$c_3 = (2\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} A^2\exp\left(\tfrac{1}{2}A^2\right)\{(1-\Phi)\ln(1-\Phi)\}^2dA = 0.4806 < \infty ,$$

$$c_4 = (2\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(\tfrac{1}{2}A^2\right)\{(1-\Phi)\ln(1-\Phi)\}^2dA = 0.9067 ,$$

$$c_5 = \left(1-\rho^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} T\big(\Phi(A)\big)dA \int_{-\infty}^{\infty} T\big(\Phi(B)\big)\exp\left(-\frac{\rho^2A^2-2\rho AB+\rho^2B^2}{2\left(1-\rho^2\right)}\right)dB ,$$

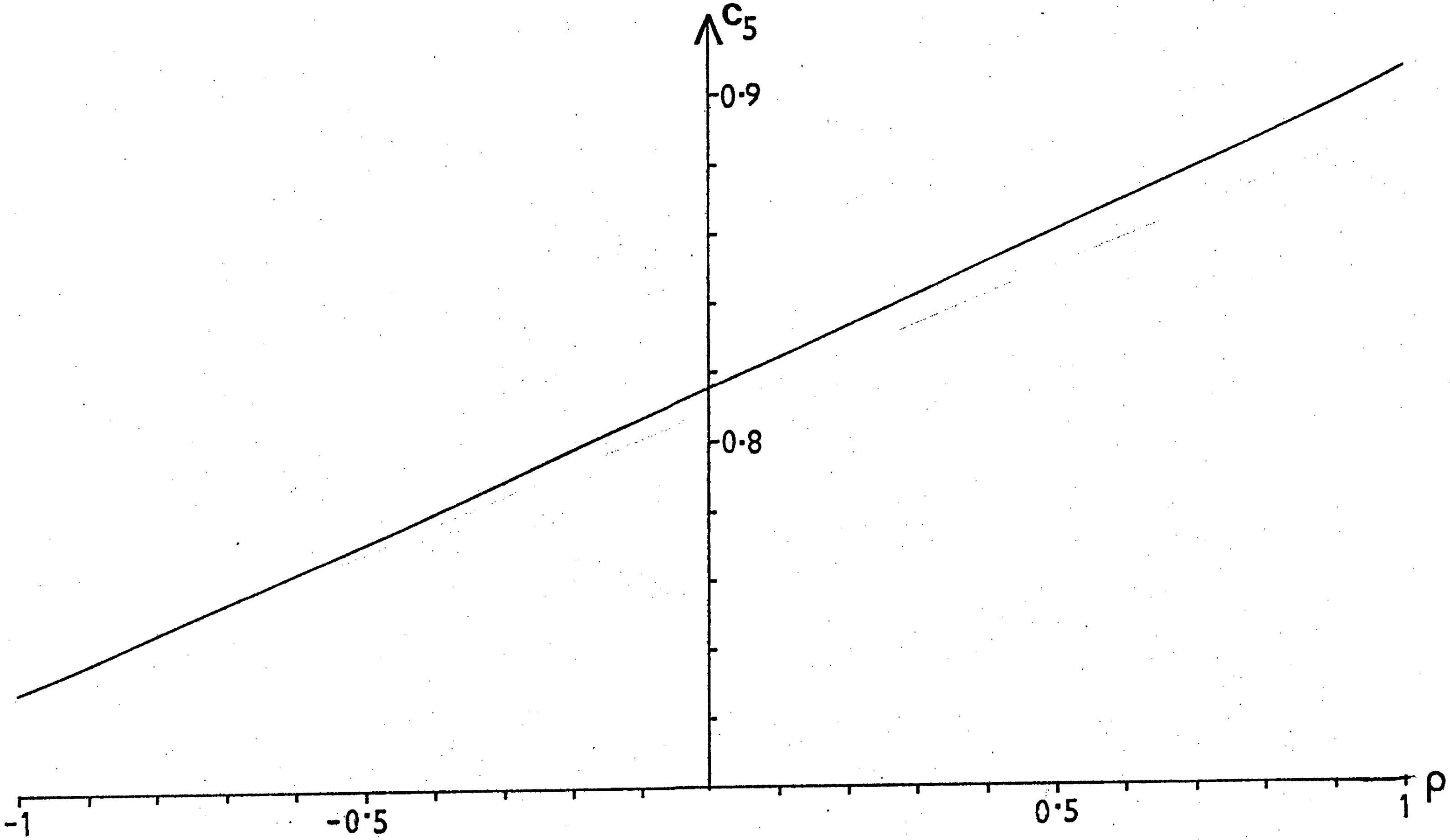
and

$$\omega = (2\pi)^{-1}\left(1-\rho^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \ln(1-\Phi)dA \int_{-\infty}^{\infty} \ln(1-\Phi)\exp\left(-\frac{A^2-2\rho AB+B^2}{2\left(1-\rho^2\right)}\right)dB - 1 .$$

For c_5 and ω as functions of ρ see Tables I and II and Figures 2.1 and 2.2.

ρ	c_5	ρ	c_5	ρ	c_5
-0.9	0.7377	-0.2	0.7981	0.5	0.8607
-0.8	0.7462	-0.1	0.8069	0.6	0.8698
-0.7	0.7548	0.0	0.8158	0.7	0.8790
-0.6	0.7633	0.1	0.8247	0.8	0.8882
-0.5	0.7720	0.2	0.8336	0.9	0.8974
-0.4	0.7806	0.3	0.8426		
-0.3	0.7894	0.4	0.8516		

TABLE I



Fig· 2·1

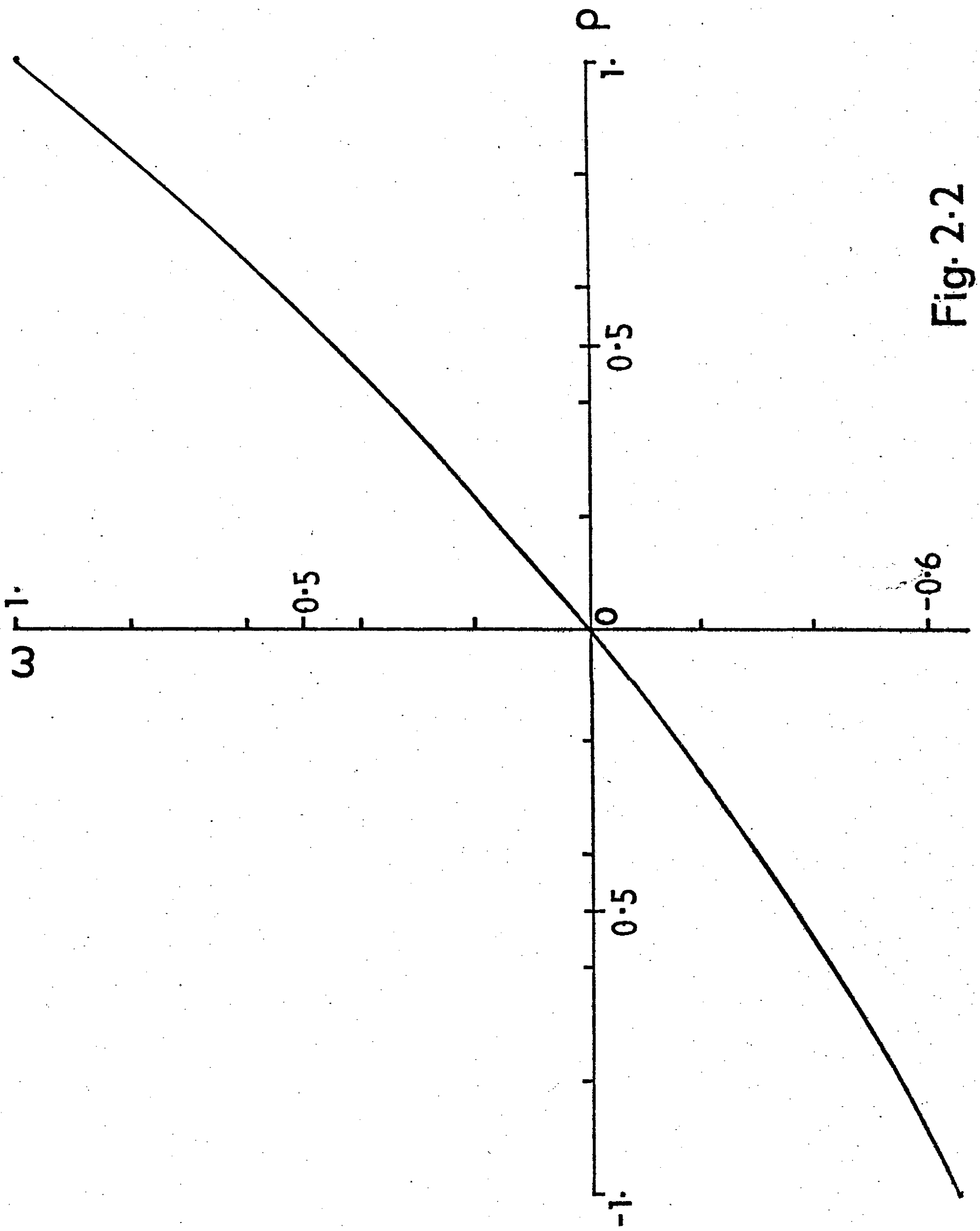


Fig. 2·2

ρ	ω	ρ	ω	ρ	ω
-0.9	-0.5953	-0.2	-0.1561	0.5	0.4531
-0.8	-0.5425	-0.1	-0.0798	0.6	0.5548
-0.7	-0.4864	0.0	0.0000	0.7	0.6603
-0.6	-0.4270	0.1	0.0834	0.8	0.7696
-0.5	-0.3644	0.2	0.1703	0.9	0.8829
-0.4	-0.2983	0.3	0.2608		
-0.3	-0.2289	0.4	0.3551		

TABLE II

Some limits for c_5 and ω can be obtained fairly easily; thus

$\lim_{\rho\to-1} c_5 = 0.7293$, $\lim_{\rho\to+1} c_5 = c_4 = 0.9067$, $\lim_{\rho\to-1} \omega = 1 - \pi^2/6 = -0.6449$

and $\lim_{\rho\to+1} \omega = 1$. An interesting feature of c_5 is that it varies fairly slowly with ρ . If we have a fairly good estimate of ρ , ρ^* say, then the value of c_5 at ρ^* will not vary very much from the value at ρ .

Since we now have ω as a function of ρ , we can calculate ARE(w, r) . At $\rho = 0$, it can be shown that

$$\partial\omega/\partial\rho = \left[\int_{-\infty}^{\infty} A \ln\{1-\Phi(A)\}\Phi'(A)dA\right]^2 = c_6^2 .$$

Now $c_6 = -0.9032$, so that

$$\text{ARE}(w, r) = 0.665 \quad \text{at} \quad \rho = 0 .$$

This is much less than the asymptotic efficiencies of any of the rank tests relative to r .

At this stage we should describe how the numerical integrations of c_i , $i = 1, 2, \ldots, 6$, and ω were performed and discuss the accuracies of the resulting approximations. Unless otherwise stated the normal cumulative distribution function values used were values to ten decimal places from the tables of the U.S. National Bureau of Standards (1948). Some useful techniques used in studying convergence of series to integrals can be found in Moran (1958).

Univariate Integrals: To evaluate numerically an integral of the form

$$I = \int_{-\infty}^{\infty} k(A)dA \ ,$$

we approximate it by the sum

$$S = \sum_{n=-\infty}^{\infty} h \cdot k(nh+\delta) \ ,$$

where h is the interval width and δ is a position constant. In this case we replace S by

$$S_1 = \sum_{n=-z}^{+z} h \cdot k(nh+\delta) \ ,$$

where $z = (6-\delta)h^{-1}$. The value of the integrand outside $[-6-h/2, 6+h/2]$ is negligible, and these terms in the sum S can be omitted. Also end effects were negligible so that no correction for these was needed. It was found that using $\delta = 0$ or $h/2$ resulted in intervals that were symmetrically placed about the origin, with the origin as either an end point or a mid point of an interval. Then the relation $\Phi(-A) = 1 - \Phi(A)$ could be used to evaluate $k(A)$ and $k(-A)$ simultaneously. Therefore $\Phi(A)$ was only needed for A positive. The values of S_1 adopted were those with $h = 0.1$ and $\delta = 0$. To obtain an estimate of the error, S_1 was also calculated for $h = 0.05, 0.2, 0.3, 0.4, 0.5, 0.6$ and 1.0, each with $\delta = 0$, and $h = 0.1$ with $\delta = 0.05$.

In all cases for c_1 the sum S_1 varied by no more than 1×10^{-7}, even for $h = 1.0$. The corresponding variation for c_4 was around 10^{-6}, and for c_6 was less than 1×10^{-6}, leading to a difference in calculated c_6^4 value of around 1×10^{-5}. The differences for c_2 and c_3 were marginally larger. The errors involved in not having the exact values for the normal distribution were also not significant. c_1, c_2, c_3, c_4 and c_6 were also calculated with $h = 0.1$ and $h = 0.01$, and $\delta = 0$, using the normal approximation due to C. Hastings in Abramowitz and Stegun [(1972), eqn. 26.2.17], which had a claimed error in Φ of not more than 7.5×10^{-8}. The difference between the value of the sum S_1 with $h = 0.1$

and $h = 0.01$ was negligible, and there was close agreement with the value of the sum S_1 found by using the tabulated cumulative distribution function values; differences of about 2×10^{-6} for c_1 and 4×10^{-6} for c_4. Since the integrands are all well behaved, approach zero fairly rapidly with increasing A and systematic errors appear unlikely, it seemed reasonable to assume at least four decimal point accuracy for c_1 and c_4.

Bivariate Integrals: Here the integrals are of the form

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(A,\, B)dAdB \ ,$$

and were approximated by the double sum

$$S_2 = \sum_{i=-z}^{z} \ \sum_{j=-M}^{+M} h^2 k\left(ih+\delta_1,\ jh+\delta_2\right)$$

where $z = \left(6-\delta_1\right)h^{-1}$ and $M = \left(6-\delta_2\right)h^{-1}$. Due to the relationship $\Phi(-A) = 1 - \Phi(A)$, and since $k(A,\, B) = k(B,\, A)$, reduction in the computational complexity is attained by letting $\delta_1 = \delta_2 = \delta$, and $\delta = 0$ or $h/2$. As the univariate sum approximations to the integrals were so accurate it was decided to calculate S_2 for $h = 1.0$, 0.5 and 0.2 with $\delta = 0$, initially, to see if smaller square intervals were really necessary. Again the value of the integrand outside $\{(A,\, B) : -6-h/2 \le A,\, B \le 6+h/2\}$ is negligible and edge effects also are negligible. S_2 was also calculated with $h = 0.2$ and $\delta = 0.1$. In addition the normal approximation due to Hastings from Abramowitz and Stegun (1972) was used with $h = 0.25$ and $\delta = 0$, as a rough check.

c_5 was calculated first. Not only is c_5 expressed as a double integral, but it is also a function of ρ and so was evaluated for values of ρ between -1 and $+1$ in steps of 0.1. Comparing the sum approximation with $h = 1.0$ and $h = 0.5$ it was apparent that there was less difference in S_2 if $|\rho|$ was not too large. Except for $|\rho| \ge 0.7$, the variations in the sums were less than 1×10^{-5}. The difference in S_2 for interval widths 0.5 and 0.2 was less than 1×10^{-6} for all ρ.

Again it seemed reasonable to assume four decimal point accuracy. In fact variations in the fifth decimal place in c_1, c_4 or c_5 did not significantly affect the multiplying factor for $\hat{\rho}$, or the ratio $\mathrm{Var}(\hat{\theta})/\mathrm{Var}(\overline{\theta})$. The value of S_2 using the normal approximation due to Hastings agreed with the value of S_2 using the tabulated normal distributional values to at least four figures for all ρ .

Similarly the double sum approximating the double integral for ω was evaluated. Similar results were obtained, except that S_2 using the normal approximation due to Hastings occasionally rounded to a slightly different four figure decimal, than the value obtained using the tabulated normal distributional values. Another interesting point is that when ρ is zero ω is known to be zero, and this provides an additional check against errors in the normal distributional values. For $h = 0.5$, the approximation for ω was less than 2×10^{-8} . The convergence of the double sum to the double integral becomes slower as the absolute value of ρ increases, but not much.

The multiplying factor $\left[1-2\rho^2\left(1-\rho^2\right)c_1^2/\left\{\left(1+\rho^2\right)d_1\right\}\right]^{-1}$ as a function of ρ is given in Table III and Figure 2.3, and the ratio $\mathrm{Var}(\hat{\theta})/\mathrm{Var}(\overline{\theta})$, being for convenience represented by $V(\theta, \rho)$, is given in Table III and Figure 2.4.

ρ	mult. factor	$V(\theta, \rho)$	ρ	mult. factor	$V(\theta, \rho)$
-0.9	1.010	0.425	0.1	1.002	0.999
-0.8	1.016	0.643	0.2	1.008	0.996
-0.7	1.020	0.774	0.3	1.018	0.993
-0.6	1.022	0.859	0.4	1.033	0.990
-0.5	1.020	0.916	0.5	1.051	0.988
-0.4	1.016	0.953	0.6	1.072	0.987
-0.3	1.011	0.977	0.7	1.096	0.987
-0.2	1.006	0.991	0.8	1.122	0.988
-0.1	1.002	0.998	0.9	1.148	0.989
0.0	1.000	1.000	1.0	1.194	0.995

TABLE III

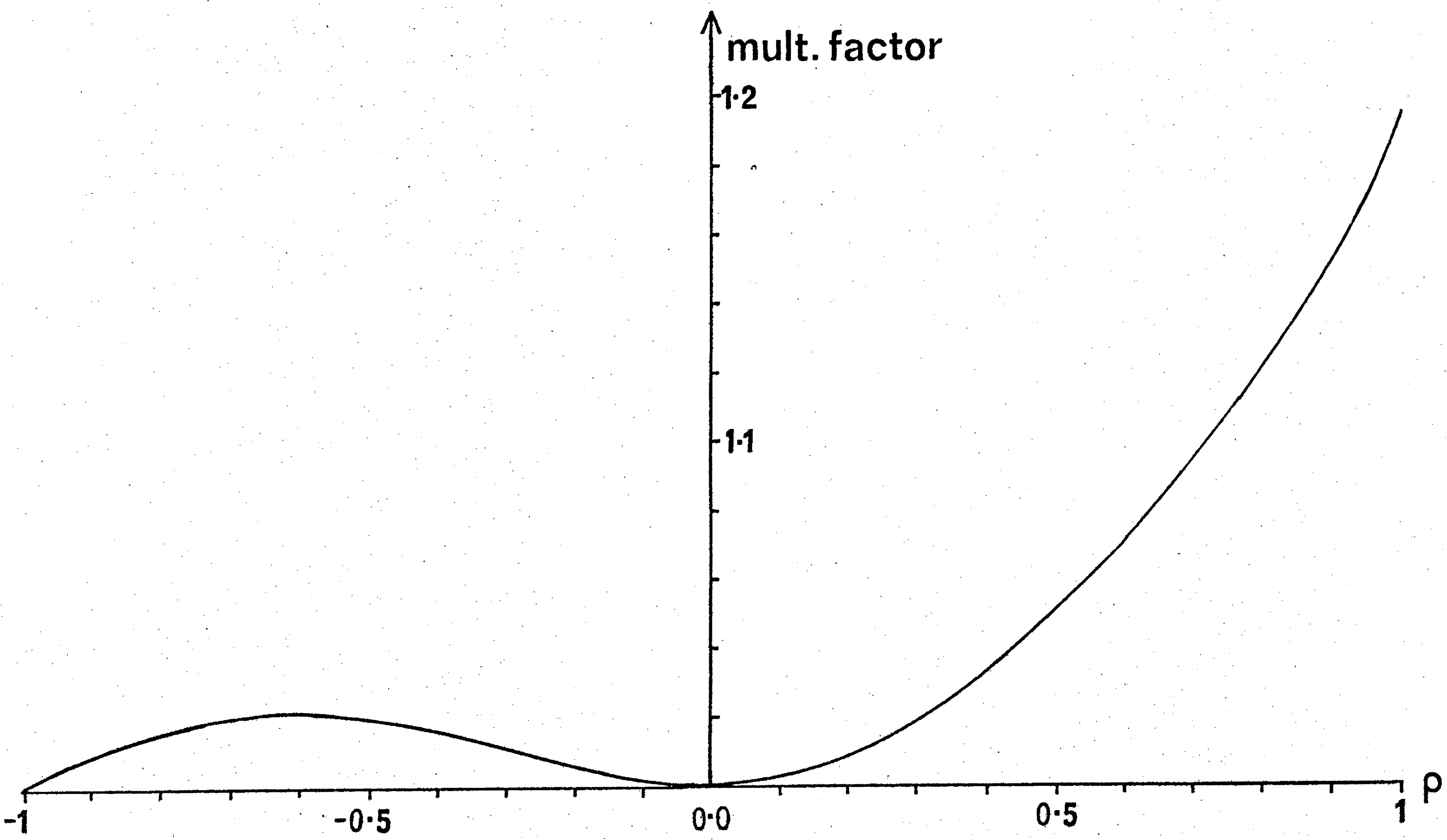
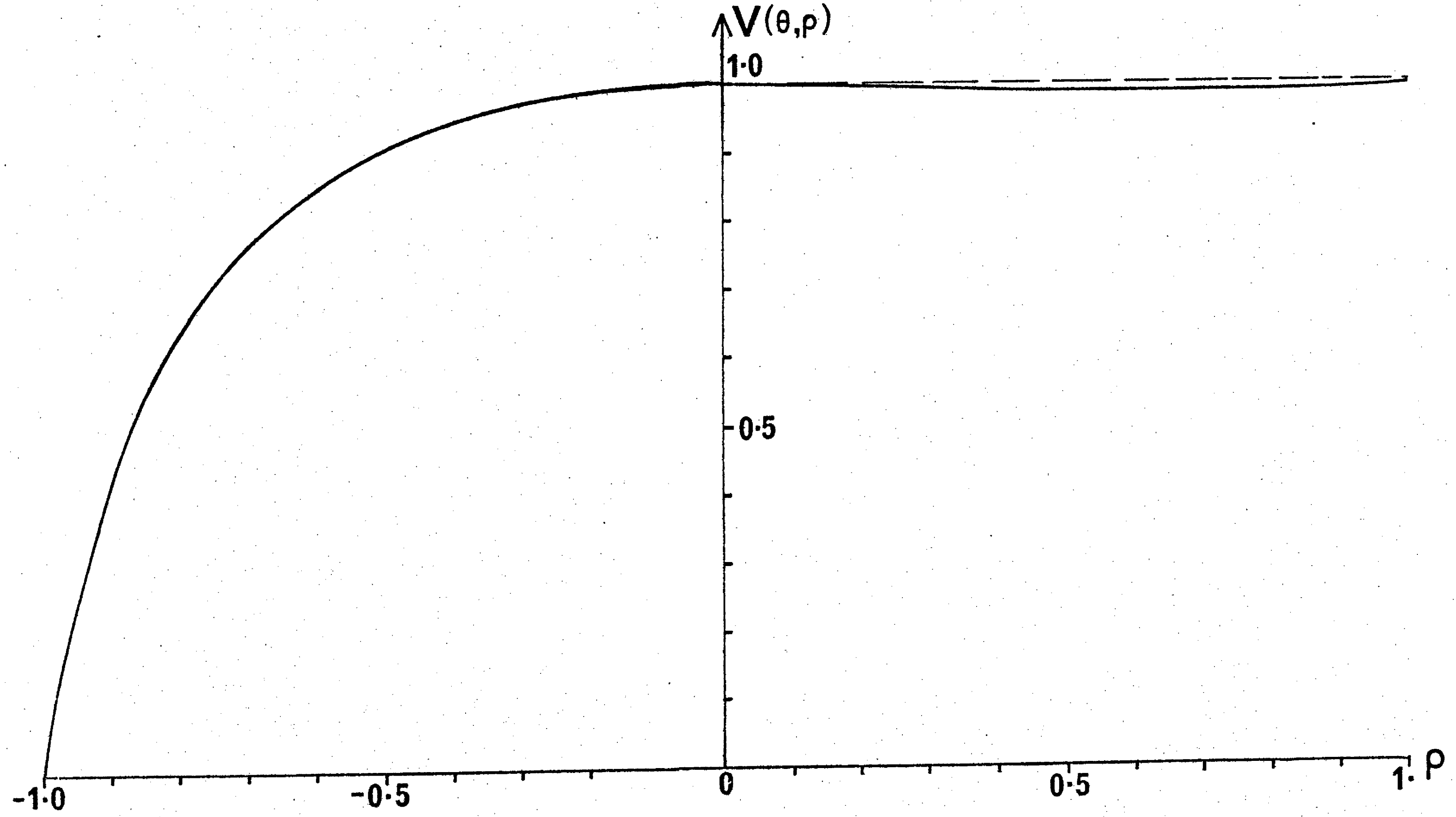


Fig. 2·3



Fig· 2·4

Also $\lim_{\rho\to -1}$ (mult. factor) = 1.0 and $\lim_{\rho\to -1} V(\theta, \rho) = 0.000$.

A few points about $V(\theta, \rho)$ should be noted. As was indicated previously, knowledge of the Y values gives information about the parameters of the X distribution. This is not the case in the bivariate normal distribution, but here we actually have a reduction in the variance of the estimator of θ because information from the Y observations influences estimation of θ .

The multiplying factor shows how much is lost asymptotically by not knowing the marginal parameters θ and λ when estimating ρ . For the bivariate exponential distribution this multiplying factor is not symmetric in ρ , unlike the case for the bivariate normal with unknown marginal scale parameters. This multiplying factor is also less than the corresponding value of $(1+\rho^2)$ for the bivariate normal.

5. Comments and Further Results

It would be interesting to determine the class of distributions which satisfy (2.1) since it does appear to be satisfied by distributions other than the exponential. If F satisfies (2.1) then

$$\partial \log T(F)/\partial\theta = g(\theta) ,$$

so that,

$$\log T(F) = \int^{\theta} g(\theta)d\theta + k(X) ,$$

and therefore

$$F = M\{g_1(\theta)+k(X)\} ,$$

where $k(X)$ is a function of X alone and $g_1(\theta) = \int^{\theta} g(s)ds$. A slight extension is required for marginal distributions with more than one parameter. It is obvious that θ can be a location type parameter, or replacing $M(\cdot)$, by $M_1 \log(\cdot)$, that θ can be a scale type parameter.

Notice that $\hat{\theta}$ and $\hat{\rho}$ are uncorrelated if

$$c_1 = \int_{-\infty}^{\infty} A\psi(A)dA = 0 .$$

When this occurs the multiplying factor for $\mathrm{Var}(\hat{\rho})$ equals one. This will occur if $\psi(A)$ is symmetric in A. If F satisfies (2.1), and θ is a location parameter then $F(X;\, \theta) = F(X-\theta)$ and

$$\partial F(X-\theta)/\partial\theta = -\partial F(X-\theta)/\partial X = -f(X;\, \theta) \;,$$

so that

$$f(X;\, \theta) = \psi(A)g(\theta) \;.$$

If $f(X-\theta)$ is symmetric around $X = \theta$, then it can easily be shown that $\psi(A)$ is symmetric, that is $\psi(-A) = \psi(A)$. So $c_1 = 0$ if X is distributed symmetrically around a location parameter θ. If θ is a scale parameter, then $F(X;\, \theta) = F(X/\theta)$ and

$$\partial F(X;\, \theta)/\partial\theta = -X\theta^{-2}\partial F(X;\, \theta)/\partial X = -X\theta^{-2}f(X/\theta) \;,$$

so that

$$-X\theta^{2}f(X/\theta) = \psi(A)g(\theta) \;.$$

Here even if $f(X/\theta)$ is symmetrically distributed around 0, it is easily shown that $\psi(A)$ is not symmetric. $\psi(A)$ can not be symmetric if X has an infinite range. So $c_1 \neq 0$ in general if θ is a scale parameter. This is obvious when $A = X - \theta$ or $A = X/\theta$. In the first case $\psi(A) = \Phi'(A)$, but in the second $\psi(A) = A\Phi'(A)$.

Condition (2.1) can not be extended to be applicable for variables whose range depends on an unknown parameter. For example if we try to carry out the analysis for X distributed as $R[0,\, \theta)$, c_1 becomes infinite and the analysis breaks down. We can go further and suggest that the normalizing transformation is inapplicable for variables defined on a finite interval of the real line. In these cases transformation to rectangular marginals might be more applicable.

Another question is raised by the form of the maximum likelihood equations for the estimators of the marginal parameters. For which distributions does

$$\sum_{i=1}^{n} (\rho A-B)A_2 = 0 \;, \text{ at } (\bar{\theta},\, \bar{\lambda}) \;?$$

This holds if $A = a + bs(X)$ where $s(X)$ is some function of X. This includes a subset of the bivariate distributions considered by Johnson

(1949) where $s(X) = X$, $\sinh^{-1}X$, $\log X$ or $\log\{X(1-X)^{-1}\}$. Consequently the marginal distributions obeying the above inequality include the two parameter log normal distribution. Otherwise if θ is a location parameter of X write $A = s(X-\theta)$. By definition $\overline{\theta}$ is the solution of

$$\sum_{i=1}^{n} \partial \log f(X-\theta)/\partial\theta = 0 \ .$$

Here $g(\theta) = 1$, and $f(X-\theta) = \psi(A)$, so the above equation becomes

$$\sum_{i=1}^{n} \{\psi'(A)/\psi(A)\} \, \partial A/\partial\theta = 0 \ .$$

Also

$$\rho E(A \ \partial A/\partial\theta) = 0 \ ,$$

so that if $\rho \neq 0$,

$$\sum_{i=1}^{n} \{\psi'(A)/\psi(A)\} \, \partial A/\partial\theta = \sum_{i=1}^{n} A \ \partial A/\partial\theta = 0 \ .$$

But $\partial A/\partial\theta \neq 0$, and the above equality must hold for all n . Hence the only solution of the above equality occurs when A is proportional to $\psi'(A)/\psi(A)$, that is when $\psi(A)$ is proportional to $\exp\left(-\tfrac{1}{2}A^2\right)$, and X is normally distributed.

Now assume θ is a scale parameter of X , then write $A = s(X/\theta)$. $\overline{\theta}$ is still the solution of

$$\sum_{i=1}^{n} \partial \log f(X/\theta)/\partial\theta = 0 \ .$$

Here $g(\theta) = -\theta^{-1}$, so $f(X/\theta) = X^{-1}\theta\psi(A)$, hence the above expression is equivalent to

$$\sum_{i=1}^{n} \partial \log \psi(A)/\partial\theta = 0 \ ,$$

or

$$\sum_{i=1}^{n} \psi'(A)/\Phi'(A) = 0 \ , \qquad (2.17)$$

using

$$\partial A/\partial\theta = [\Phi'(A)]^{-1}\ \partial F/\partial\theta = g(\theta)\psi(A)/\Phi'(A)\ .$$

Here $E(A\ \partial A/\partial\theta) \neq 0$, so that if $\rho \neq 0$,

$$\rho = \left(\sum_{i=1}^{n} B\ \partial A/\partial\theta\right) \Big/ \left(\sum_{i=1}^{n} A\ \partial A/\partial\theta\right)\ .$$

Since

$$-\frac{d}{dA}\left[\frac{\psi(A)}{\Phi'(A)}\right] = A\,\frac{\psi(A)}{\Phi'(A)} + \frac{\psi'(A)}{\Phi'(A)}\ ,$$

we can show that if (2.17) holds then

$$\sum_{i=1}^{n} A\psi(A)/\Phi'(A) = 1\ ,$$

and

$$\rho = -\theta \sum_{i=1}^{n} B\ \partial A/\partial\theta\ .$$

Similarly

$$\rho = -\lambda \sum_{i=1}^{n} A\ \partial B/\partial\lambda\ ,$$

or

$$\rho = \sum_{i=1}^{n} B\psi(A)/\Phi'(A) = \sum_{i=1}^{n} A\psi(B)/\Phi'(B)\ .$$

This must be true for all n , and hence

$$\psi(A) = A\Phi'(A)\ ,$$

that is X is normally distributed.

The problem of what to do when condition (2.1) does not hold still remains. The univariate and bivariate integrals will now be functions of the unknown parameters θ and λ . All these integrals will have to be evaluated for a range of the parameters θ and λ . If the integrals vary slowly with θ and λ , perhaps they could be approximated using $\hat{\theta}$ and $\hat{\lambda}$. An alternative is to see how much information is lost by using the marginal estimates $(\bar{\theta}, \bar{\lambda})$. Some distributions, such as the bivariate gamma distribution, are of enough practical interest to make it worthwhile to consider such a study. In the next chapter we will show that this bivariate gamma distribution can not be infinitely divisible for negative ρ, and some of the positive range of ρ. Also it is shown that it can not even be divisible for some of the range of ρ. This means that we can not consider constructing such a distribution as the sum of other bivariate distributions and so eliminates one possible method of examining the distribution.

CHAPTER 3

BIVARIATE GAMMA DISTRIBUTIONS AND INFINITE DIVISIBILITY

1. Introduction

In the last chapter the bivariate gamma distribution of Moran (1969) was introduced. In his paper Moran remarked that it would be interesting to know whether or not this bivariate gamma distribution is infinitely divisible. In this chapter we study this question, and although a complete solution is not available, the partial solution is certainly suggestive. Although this particular bivariate gamma distribution was initially the object of the investigation, the results have a much wider range of applicability, and highlight some interesting points about bivariate distributions.

The bivariate random vector $V = (X, Y)$ is said to be infinitely divisible if for every positive integer n there exists a set of mutually independent, identically distributed random vectors $V_1, V_2, \ldots, V_n$ such that $V = V_1 + V_2 + \ldots + V_n$. Equivalently if $\chi(s, t)$ is the characteristic function of (X, Y), then $[\chi(s, t)]^{1/n}$ must be a characteristic function for every n.

Infinite divisibility is a theoretical concept; one of its most important applications being in the theory of limit distributions of sums of independent random variables, see for example Lévy (1937) and Gnedenko and Kolmogorov (1954). All infinitely divisible distributions can be generated by stochastic processes, more specifically by processes with stationary independent increments. This provides the motivation for the study of infinite divisibility, but here we will be more concerned with establishing infinite divisibility rather than its applications.

Much of the work on infinitely divisible distributions has been confined to univariate distributions. Extensive bibliographies can be found in Fisz (1962) and Steutel (1973), although many of the problems posed in the latter paper have now been solved. A whole range of techniques is available for establishing infinite divisibility in the univariate case. For example, for positive random variables the Laplace transform and properties associated with it are often examined instead of the characteristic function (see for example Appendix C). When we come to bivariate distributions things change. Very little work has been done, and what has been done generally relies on the characteristic function being relatively easy to find. But bivariate characteristic functions are often difficult if not impossible to find. This difficulty occurs for the bivariate gamma of Moran, and some other line of approach must be tried. Since gamma variables are closed under convolution an examination of the correlation ω seemed indicated.

2. Negative Correlation

Let

$$f_X(x) = \Gamma(\alpha)^{-1}e^{-x}x^{\alpha-1} \text{ , and } f_Y(y) = \Gamma(\beta)^{-1}e^{-y}y^{\beta-1} \text{ ,}$$

then we will denote the set of bivariate distributions with these marginals by $\mathbb{\Gamma}(\alpha, \beta; \omega)$ where ω is the correlation between X and Y. We are interested in specifying the range of ω for which it is possible for an element of $\mathbb{\Gamma}(\alpha, \beta; \omega)$ to be infinitely divisible. Analogously we are interested in the range of ω for which it is impossible for any element of $\mathbb{\Gamma}(\alpha, \beta; \omega)$ to be infinitely divisible.

Suppose there exists an infinitely divisible element of $\mathbb{\Gamma}(\alpha, \beta; \omega)$ where ω is negative. Then consideration of the marginal variates of $[\chi_{X,Y}(s, t)]^{1/n}$ leads us to the conclusion that there must exist at least one element of $\mathbb{\Gamma}(\alpha/n, \beta/n; \omega)$, (U, V) say. But for positive random variables,

$$\omega \geq -E(U)E(V)/\{\text{Var } U \cdot \text{Var } V\}^{\frac{1}{2}} = -(\alpha\beta)^{\frac{1}{2}}/n \text{ ,}$$

which must be true for all n , so that

$$\omega \geq 0 \text{ .}$$

There are no infinitely divisible elements of $\mathbb{\Gamma}(\alpha, \beta; \omega)$ if ω is negative. Notice that theorem 4.3 of Whitt (1976) shows that if $\alpha = \beta$, the minimum correlation between X and Y is non-increasing as α runs through the values $t2^{-n}$, $n \geq 1$. Whitt merely confines himself to that observation and speculates whether this minimum is strictly decreasing from 0 to -1 as α goes from 0 to infinity. He makes no observations about infinite divisibility.

The above result can be extended as follows: Let (X, Y) be a bivariate distribution such that $0 \le X, Y < \infty$, $EX = m_1$, $EY = m_2$, $\mathrm{Var}\, X = V_1$ and $\mathrm{Var}\, Y = V_2$, where all these moments are finite. Assume that both X and Y are infinitely divisible. If there is an infinitely divisible bivariate distribution with marginals X and Y then for each n there must also be a bivariate distribution with marginals S_i and T_i such that $X = \sum_{i=1}^{n} S_i$, $Y = \sum_{i=1}^{n} T_i$, $ES_i = m_1/n$, $ET_i = m_2/n$, $\mathrm{Var}\, S_i = V_1/n$ and $\mathrm{Var}\, T_i = V_2/n$. Also S_i and T_i must be non-negative random variables, otherwise there would be a non-zero probability that either X or Y or both, was negative. Then

$$\begin{aligned} \omega = \mathrm{Corr}(X, Y) &= \mathrm{Corr}\left(S_i, T_i\right) \\ &\ge -ES_iET_i/\{\mathrm{Var}\, S_i \;\mathrm{Var}\, T_i\}^{\frac{1}{2}} \\ &= m_1m_2V_1^{-\frac{1}{2}}V_2^{-\frac{1}{2}}n^{-1} , \end{aligned}$$

which must hold for all n , consequently

$$\omega \ge 0 .$$

This inequality holds for all bivariate distributions with non-negative infinitely divisible marginal variables which have finite first two moments, and is applicable to both discrete and continuous random variables.

If $(-\infty < X, Y \le 0)$ and X and Y are infinitely divisible then,

$$\omega = \mathrm{Corr}(X, Y) = \mathrm{Corr}(-X, -Y) \ge 0 ,$$

by the above. If instead $(-D \le X < \infty, -C \le Y < \infty)$ where $D \ge C \ge 0$, then

$$\omega = \mathrm{Corr}(X, Y) = \mathrm{Corr}(X+D, Y+D) \ge 0 .$$

For Moran's bivariate gamma distribution, when ρ is negative then ω is also negative, hence certainly this distribution is not infinitely divisible for the entire possible negative range of ω .

Non-negative infinitely divisible random variables with finite first and second moments include the negative binomial distribution, Pareto distribution, the log normal distribution, the latter two proved infinitely divisible by Thorin (1977a) and (1977b) respectively, the inverse Gaussian distribution (see Johnson and Kotz [(1970), pg. 139]), and $z \overset{d}{\sim} 1/X$ where X has the inverse Gaussian distribution (see Johnson and Kotz [(1970), pg. 149]). Two other non-negative distributions recently proved to be infinitely divisible are the ratio of two independent gamma variables, see Goovaerts *et al* (1978), and $z \overset{d}{\sim} \left(\chi_k^2\right)^{-1}$ see Kelker (1971), Grosswald (1976a), (1976b), Ismail and Kelker (1976) and Ismail (1977), where infinite divisibility was established as part of the proof that the t distribution is infinitely divisible.

There are many more, including for example the non-central chi-square distribution and the Poisson distribution.

As examples of how useful these observations on correlation are, let us consider two bivariate exponential distributions. The first, attributed to Gumbel (1960), has probability distribution function;

$$f(x, y) = \{(1+\theta x)(1+\theta y)-\theta\}\exp(-x-y-\theta xy) \ , \quad 0 \le \theta \le 1 \ ,$$

with characteristic function

$$\chi_{X,Y}(s, t) = \int_0^\infty (1-\theta+\theta y)(1-it-\theta y)^{-1}\exp\{-(1-is)y\}dy$$
$$+ \theta \int_0^\infty (1+y)(1-it-\theta y)^{-2}\exp\{-(1-is)y\}dy \ .$$

This characteristic function is extremely messy and probably involves exponential integrals. However if we examine the correlation between X and Y, which is somewhat easier to find (see for example Johnson and Kotz [(1972), pg. 262]),

$$\mathrm{Corr}(X, Y) = \theta^{-1}\exp(1/\theta)E_1\left(\theta^{-1}\right) - 1 \ ,$$

where $E_1(t)$ is the exponential integral defined by

$$E_1(t) = \int_t^\infty u^{-1}e^{-u}du \ ,$$

and is well tabulated (see Abramowitz and Stegun [(1972), pg. 238]). It can be shown that

$$-0.40365 \le \omega \le 0 \ .$$

From above, this distribution can not be infinitely divisible unless $\omega = 0$.

The second distribution is attributed to Eyraud (1936), Gumbel (1960)

Morgenstern (1956) and has probability distribution function;

$$f(x,\ y) = \left\{1-\theta\left(2e^{-x}-1\right)\left(2e^{-y}-1\right)\right\}\exp(-x-y)\ ,$$

where $|\theta| \leq 1$. The correlation is easily found to be

$$\omega = \tfrac{1}{4}\theta\ ,$$

thus the possible range for ω is

$$-\tfrac{1}{4} \leq \omega \leq \tfrac{1}{4}\ .$$

The distribution is not infinitely divisible if $\theta < 0$, and it is only necessary to consider $\theta \geq 0$. The characteristic function for this bivariate distribution is easier to find than for the previous example, and is given by

$$\chi_{X,Y}(s,\ t) = (1-is)^{-1}(1-it)^{-1}\{1-4\theta st(1-\tfrac{1}{2}it)^{-1}(1-\tfrac{1}{2}is)^{-1}\}\ .$$

It only remains to examine this for $\theta \geq 0$, and that will not be done here.

3. Classical Bivariate Gamma Distributions

For bivariate gamma distributions we still have to consider whether or not the distribution is infinitely divisible when ω is positive. If the marginal variables are independent and ω is zero, then

$$\chi_{X,Y}(s,\ t) = (1-is)^{-\alpha}(1-it)^{-\beta}\ ,$$

and this is obviously infinitely divisible, being the product of two infinitely divisible characteristic functions. Vere-Jones (1967) proved that if α is equal to β , there exists an infinitely divisible bivariate gamma distribution for all ω between zero and one, and this has characteristic function first given by Kibble (1941);

$$\chi_{X,Y}(s,\ t) = [(1-is)(1-it)+\omega st]^{-\alpha}\ .$$

Griffiths (1969a) and Sarmanov (1970) gave generalizations of the above distribution when α is greater than β ,

$$\chi_{X,Y}(s,\ t) = (1-is)^{-(\alpha-\beta)}[(1-is)(1-it)+\kappa st]^{-\beta}\ , \tag{3.1}$$

which Griffiths (1969a) showed was infinitely divisible. Actually this is the canonical form bivariate gamma with probability density function given by,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)\left[1 + \sum_{n=1}^{\infty} a_n L_n^{\alpha-1}(x)L_n^{\beta-1}(y)\right] ,$$

where $\alpha > \beta$, and for some $0 \leq \kappa \leq 1$,

$$a_n = \kappa^n[\Gamma(\alpha)\Gamma(\beta+n)]^{\frac{1}{2}}/[\Gamma(\beta)\Gamma(\alpha+n)]^{\frac{1}{2}} .$$

$L_n^{\alpha}(x)$ are Laguerre polynomials (see Erdélyi [(1953), p. 188]. Now $\omega = \kappa(\beta/\alpha)^{\frac{1}{2}}$ and thus

$$0 \leq \omega \leq (\beta/\alpha)^{\frac{1}{2}} .$$

Therefore infinitely divisible elements of $\mathbb{\Gamma}(\alpha, \beta; \omega)$ exist for at least $0 \leq \omega \leq (\beta/\alpha)^{\frac{1}{2}}$. Griffiths went further and examined a wider range of bivariate gamma distributions. But for all of these, $0 \leq \omega \leq (\beta/\alpha)^{\frac{1}{2}}$, and Griffiths (1969b) showed that not all of this wider range of bivariate distributions are infinitely divisible. They have characteristic function given by,

$$\chi_{X,Y}(s, t) = (1-is)^{-\alpha}(1-it)^{-\beta} \int_0^1 [1+\kappa st(1+is)^{-1}(1+it)^{-1}]^{-\beta} d\mu(t) ,$$

where $\mu(\cdot)$ is a distribution function on $[0, 1]$.

To see if the interval for ω , $0 \leq \omega \leq (\beta/\alpha)^{\frac{1}{2}}$, can be extended we first look at the other classical bivariate gamma distributions. Cheriyan (1941) and David and Fix (1961) constructed a bivariate distribution as follows: let U_1, U_2 and U_3 be independent gamma variables with shape parameters θ_1, θ_2 and θ_3 respectively. Define

$$X = U_1 + U_3 \quad \text{and} \quad Y = U_2 + U_3 ,$$

then X and Y have gamma distributions with shape parameters $\theta_1 + \theta_3$ and $\theta_2 + \theta_3$ respectively, and a joint distribution with correlation coefficient

$$\omega = \theta_3/\{(\theta_1+\theta_3)(\theta_2+\theta_3)\}^{\frac{1}{2}} .$$

This follows at once using the moments of X and Y . The characteristic function is

$$\chi_{X,Y}(s,\ t) = (1-is)^{-\theta_1}(1-it)^{-\theta_2}[(1-is)(1-it)+st]^{-\theta_3} .$$

This is infinitely divisible by construction, since we have merely added two independent, infinitely divisible bivariate gamma distributions,

$$(X,\ Y) = (\theta_1,\ \theta_2) + (\theta_3,\ \theta_3) .$$

If we let $\alpha = \theta_1 + \theta_3$, $\beta = \theta_3$ and $\kappa = \theta_2$ where $\alpha > \beta$, then

$$\omega = (\beta/\alpha)^{\frac{1}{2}}\{1 + \kappa/\beta\}^{-\frac{1}{2}} < (\beta/\alpha)^{\frac{1}{2}} .$$

McKay (1934) had previously considered the joint distribution of $X = U_1 + U_3$ and $Y = U_3$, which is just a special case (or limiting case) of the above distribution. McKay's bivariate gamma is also infinitely divisible and has joint probability distribution function given by

$$f(x,\ y) = \Gamma(\alpha)^{-1}\Gamma(\alpha-\beta)y^{\alpha-1}(x-y)^{\alpha-\beta-1}e^{-x} ,$$

with characteristic function

$$\chi_{X,Y}(s,\ t) = (1-is)^{-(\alpha-\beta)}[(1-is)(1-it)+st]^{-\beta} ,$$

and correlation $\omega = (\beta/\alpha)^{\frac{1}{2}}$. This characteristic function is very interesting since it corresponds to (3.1) with $\kappa = 1$. In this case, since the characteristic function uniquely defines the distribution, McKay's distribution is just a special case of the canonical correlation model.

Dussauchoy and Berland (1972) gave the distribution for two dependent gamma random variables X and Y with the property that $X - bY$ and Y are independent. They showed that the characteristic function is

$$\chi_{X,Y}(s,\ t) = \left(1 - is/\theta_1\right)^{-\alpha}\left(1 - ibt/\theta_2\right)^{-\beta}\left(1 - i(t+bs)/\theta_2\right)^{-\beta} ,$$

where $E(\theta_1 X) = \alpha$, $E(\theta_2 Y) = \beta$, and $0 < b \le \theta_1/\theta_2$. Here $\mathrm{Corr}(X,\ Y) = b\theta_2\theta_1^{-1}(\beta/\alpha)^{\frac{1}{2}} \le (\beta/\alpha)^{\frac{1}{2}}$. They proved that the characteristic function is infinitely divisible and derived the joint probability density function, which in addition to a term similar to the density of McKay's distribution, also involves a confluent hypergeometric function. Note that the density is only non-zero when $x > by$, similar to the McKay distribution which requires $x > y$.

All the distributions given in this section have marked similarities, and include all of the infinitely divisible bivariate gamma distributions discussed so far. This means that we have not yet found an infinitely divisible bivariate gamma distribution which has correlation outside the range $0 \le \omega \le (\beta/\alpha)^{\frac{1}{2}}$. It is interesting to know whether or not it is possible for a bivariate gamma distribution to be infinitely divisible if $\omega > (\beta/\alpha)^{\frac{1}{2}}$, where $\alpha > \beta$.

4. The Bivariate Gamma Distribution of Moran

Now we return to the bivariate gamma distribution of Moran. We are interested in examining this distribution when ω is positive. Let $A = \Phi^{-1}\left[F_X(x;\ \alpha)\right]$ and $B = \Phi^{-1}\left[F_Y(y;\ \beta)\right]$, where (A, B) has a standard bivariate normal distribution with correlation ρ, as in chapter 2. Lancaster (1957) and Maung (1941) proved that $|\rho| \ge |\omega|$, and when ρ is zero, ω is also zero. Further it is easy to show that when ω is zero, X and Y are independent, so that the distribution of X and Y is infinitely divisible. We have already dealt with negative ω; consequently we will only need to consider positive ω. As ρ varies from 0 to 1, ω varies from 0 to $\nu(\alpha, \beta)$, where $\nu(\alpha, \beta) = 1$ if and only if α is equal to β. When α is equal to β and ω is one, the distribution is degenerate on the line $Y = X$, and has characteristic function

$$\left(1-i(s+t)\right)^{-\alpha},$$

and is infinitely divisible. If α is not equal to β then $\nu(\alpha, \beta)$ is still attained when the probability is concentrated on the line

$$F_X(x;\ \alpha) = F_Y(y;\ \beta),$$

that is,

$$\Gamma(\alpha)^{-1} \int_0^x e^{-u} u^{\alpha-1} du = \Gamma(\beta)^{-1} \int_0^y e^{-v} v^{\beta-1} dv . \tag{3.2}$$

This relationship defines Y uniquely as a function of X. For example if $\alpha = \beta$, $Y = X$, or if $\beta = 1$ and α integral,

$$Y = X - \log_e \left(\sum_{r=0}^{\alpha-1} X^r/r! \right) .$$

Before continuing further we will develop some properties of $\nu(\alpha, \beta)$.

LEMMA 1. $\nu(\alpha, \beta) \geq (\beta/\alpha)^{\frac{1}{2}}$.

Equality holds when $\alpha = \beta$. If $\alpha \neq \beta$, this follows immediately from the canonical correlation bivariate gamma which is defined in a non-degenerate manner for $0 \leq \omega \leq (\beta/\alpha)^{\frac{1}{2}}$, for all $\alpha, \beta > 0$. This means that $\nu(\alpha, \beta)$ must be at least $(\beta/\alpha)^{\frac{1}{2}}$. Further, since at $\nu(\alpha, \beta)$ a bivariate gamma distribution is degenerate on the line

$$F_X(x;\ \alpha) = F_Y(y;\ \beta) ,$$

for α, β positive, we can deduce that strict inequality holds when $\alpha \neq \beta$.

For example, if $\beta = 1$ and $\alpha = 2$, then at $\mathrm{Corr}(X, Y) = \nu(2, 1)$,

$$Y = X - \ln(X+1) ,$$

$$\begin{aligned} \nu(2, 1) &= 2^{-\frac{1}{2}}\{E(XY)-2\} \\ &= 2^{-\frac{1}{2}}\left\{E\left(X^2\right)-E\left(X\ln(X+1)\right)-2\right\} , \end{aligned}$$

and to evaluate this we need,

$$\int_0^\infty t^2 \exp(-t)\ln(1+t)dt .$$

Using Bierens de Haan [(1939), Table 353 (6)], this is

$$2 + eE_1(1) ,$$

where $E_1(t)$ is the exponential integral defined previously. This function is tabulated (see for example Abramowitz and Stegun [(1972), p. 239]). Using the numerical value of $E_1(1)$ we obtain,

$$\nu(2, 1) = 2^{-\frac{1}{2}}\left\{2-eE_1(1)\right\} = 0.9925 ,$$

whereas

$$(\beta/\alpha)^{\frac{1}{2}} = 2^{-\frac{1}{2}} = 0.7071 .$$

Later we will examine an approximation for $\nu(\alpha, \beta)$ valid for α and β large.

LEMMA 2. $\nu(\alpha, \beta)$ *tends to* 1 *as* α *and* β *tend to infinity.*

Define

$$X^* = (X-\alpha)/\alpha^{\frac{1}{2}} \quad \text{and} \quad Y^* = (Y-\beta)/\beta^{\frac{1}{2}} ,$$

then if

$$\nu(\alpha, \beta) = \mathrm{Corr}(X, Y)$$

it follows that

$$\nu(\alpha, \beta) = \mathrm{Corr}(X^*, Y^*) ,$$

and X^* and Y^* are degenerate on the line

$$F_{X^*}(x^*) = F_{Y^*}(y^*) .$$

As α and β tend to infinity X^* tends in distribution to A and Y^* tends in distribution to B. The above relationship tends to

$$\Phi(X^*) = \Phi(Y^*) ,$$

that is the solution tends to a linear relationship between X^* and Y^*, and the correlation ω tends to one.

LEMMA 3. $\nu(q\alpha, q\beta)$ *is a continuous function of* q .

Now

$$\nu(q\alpha, q\beta) = (\beta/\alpha)^{\frac{1}{2}}\Gamma(\beta q+1)^{-1} \int_0^\infty XY^{q\beta}e^{-Y}dY - q(\alpha\beta)^{\frac{1}{2}} ,$$

where X is defined uniquely as a function of Y by

$$\Gamma(\alpha q)^{-1} \int_0^X e^{-u}u^{\alpha q-1}du = \Gamma(q\beta)^{-1} \int_0^Y e^{-v}v^{q\beta-1}dv .$$

$\gamma(\alpha q, X)$ is a continuous monotonically increasing function of X, continuous in q, from Erdélyi [(1953), p. 141] and thus has a well defined inverse, which is also continuous in q. $\gamma(\beta q, Y)/\Gamma(\beta q)$ is also continuous in q, therefore X is defined uniquely as a continuous function of Y and q. Thus $\nu(q\alpha, q\beta)$ can easily be seen to be a continuous function of q.

LEMMA 4. *We would like to prove that* $\nu(q\alpha, q\beta)$ *is a strictly increasing function of* q *. It is only necessary to prove something slightly weaker, and we will do this.*

If (X_1, Y_1) and (X_2, Y_2) are independent elements of $\Gamma(\alpha, \beta; \omega)$ where for some small positive ε, $\omega = \nu(\alpha, \beta) - \varepsilon$, then since $X_1 + X_2$ and $Y_1 + Y_2$ have gamma distributions with respective shape parameters 2α and 2β and $\mathrm{Corr}(X_1+X_2, Y_1+Y_2) = \omega$, we have just constructed an element of $\Gamma(2\alpha, 2\beta; \omega)$. We can let ε be zero and the argument is unaffected. This only proves that $\nu(r\alpha, r\beta)$ is not less than $\nu(\alpha, \beta)$ for r integral, and $\nu(s^n\alpha, s^n\beta)$ is a nondecreasing function of n for n and s integral. We can also show that

$$\nu(2^{n-1}\alpha, 2^{n-1}\beta) = \nu(2^n\alpha, 2^n\beta) ,$$

can not hold for all n. If it did hold for all n integral, then since $\{\nu(2^n\alpha, 2^n\beta)\}$ can be regarded as a Cauchy sequence, $\lim_{n\to\infty} \nu(2^n\alpha, 2^n\beta) = \nu(\alpha, \beta)$. But $\lim_{n\to\infty} \nu(2^n\alpha, 2^n\beta) = 1$, by lemma 2, and since $\nu(\alpha, \beta)$ is strictly less than 1 for $\alpha \neq \beta$, we have a contradiction. Therefore for at least one n,

$$\nu(2^{n-1}\alpha, 2^{n-1}\beta) < \nu(2^n\alpha, 2^n\beta) .$$

For finite n, $\nu(2^n\alpha, 2^n\beta)$ is strictly less than 1 for $\alpha \neq \beta$, so that for each n, there exists n_0 greater than n such that

$$\nu(2^n\alpha, 2^n\beta) < \nu(2^{n_0}\alpha, 2^{n_0}\beta) .$$

However we can show that

$$\nu(2\alpha, 2\beta) > \nu(\alpha, \beta) .$$

When $\omega = \nu(\alpha, \beta)$, $X_1 = h(Y_1)$ and $X_2 = h(Y_2)$ where h is the continuous monotonically increasing function which uniquely determines X_i as a function of Y_i, $i = 1, 2$. Therefore we write

$$X_1 + X_2 = h(Y_1) + h(Y_2) ,$$

and if $\nu(2\alpha, 2\beta) = \nu(\alpha, \beta)$ then $X_1 + X_2 = h^*(Y_1+Y_2)$ where h^* is the continuous monotonically increasing function which uniquely determines $X_1 + X_2$ as a function of $Y_1 + Y_2$. Then it must be true that

$$h^*(Y_1+Y_2) = h(Y_1) + h(Y_2) ,$$

for all Y_1 and Y_2. Taking $\varepsilon > 0$, this equality implies that

$$h^*(Y_1+Y_2) = h(Y_1+\varepsilon) + h(Y_2-\varepsilon) .$$

For simplicity consider the case when $Y_1 = Y_2$,

$$h^*(2Y_1) = h(Y_1+\varepsilon) + h(Y_1-\varepsilon) = h(Y_1) + h(Y_1) .$$

If h is not linear then we can certainly find a Y_1 and ε such that $h(Y_1+\varepsilon) - h(Y_1) \neq h(Y_1) - h(Y_1-\varepsilon)$, and obtain a contradiction. Of course if h is linear then $\alpha = \beta$, $\nu(\alpha, \beta) = 1$ and $\nu(2\alpha, 2\beta) = 1$. If $\alpha \neq \beta$, then $X_1 + X_2 = h^*(Y_1+Y_2)$ can not hold, and $X_1 + X_2$ is not completely determined by $Y_1 + Y_2$. The distribution (X_1+X_2, Y_1+Y_2) is not degenerate and

$$\nu(\alpha, \beta) < \nu(2\alpha, 2\beta) .$$

This argument extends to show that

$$\nu(2^{n-1}\alpha, 2^{n-1}\beta) < \nu(2^n\alpha, 2^n\beta) ,$$

for all integral n, and also

$$\nu(2^{-n}\alpha, 2^{-n}\beta) < \nu(2^{-n+1}\alpha, 2^{-n+1}\beta) .$$

We have proved that the sequence $\{\nu(2^{-n}\alpha, 2^{-n}\beta)\}$ is strictly decreasing in n, for n integral, and is bounded below by $(\beta/\alpha)^{\frac{1}{2}}$, and is therefore convergent. Similarly $\nu(s^{-n}\alpha, s^{-n}\beta)$ is a strictly decreasing sequence for integral n and s.

However we have not shown that $\nu(q\alpha, q\beta)$ is an increasing function of q, although since $\nu(q\alpha, q\beta)$ is continuous we can prove that as n increases the sequences $\{\nu(s^{-n}\alpha, s^{-n}\beta)\}$ all decrease to the same limit. This is all we need to consider infinite divisibility, since if $\nu(\alpha, \beta) > \nu(\alpha/2, \beta/2)$, then for

$$\nu(\alpha, \beta) \geq \omega \geq \nu(\alpha/2, \beta/2) ,$$

(X_1, Y_1) is not only not infinitely divisible, it is not divisible at all. That is, we can not find (U_1, V_1) and (U_2, V_2) independent and identically distributed such that $(X_1, Y_1) = (U_1+U_2, V_1+V_2)$. However it would be nice to be able to complete the proof.

5. $\omega(\alpha, \beta)$ for α and β Large

For fixed ρ we know from Lancaster (1957) and Maung (1941) that $|\omega| \leq |\rho|$, and as α and β tend to infinity, $(X-\alpha)/\alpha^{\frac{1}{2}}$ tends in distribution to A and $(Y-\beta)/\beta^{\frac{1}{2}}$ tends in distribution to B. It can be shown that as α and β tend to infinity, ω tends to ρ. Thus for fixed ρ, ω does not remain constant as α and β increase; ω must be a function of α and β. This is unlike the canonical correlation bivariate gamma where ω remains fixed as long as β/α remains constant. In Lancaster's construction of the proof that $|\omega| \leq |\rho|$, he expressed X in the form

$$X = \sum_{i=1}^{\infty} a_i H_i(A)(i!)^{-\frac{1}{2}} ,$$

where $H_i(A)$ are the Hermite polynomials (see Erdélyi [(1953), p. 192], and a shift in scale and location has been made so that $\sum_{i=1}^{\infty} a_i^2 = 1$. Also

$$Y = \sum_{i=1}^{\infty} b_i H_i(B)(i!)^{-\frac{1}{2}} ,$$

where $\sum_{i=1}^{\infty} b_i^2 = 1$, and $a_i = b_i$ if $\alpha = \beta$. Then using the orthogonal properties of the Hermite polynomials

$$\mathrm{Corr}(X, Y) = \sum_{i=1}^{\infty} a_i b_i \rho^i .$$

Notice that by construction $X(A) = F_X^{-1}[\Phi(A);\ \alpha]$. X is certainly a function of A and is completely determined by A. For each observation we can regard A as a variable on the real line, then we can discuss expressing X as an infinite sum of orthogonal polynomials of A. Here A is a variable on the real line to which we associate the measure $(2\pi)^{-\frac{1}{2}}\exp\{-\frac{1}{2}A^2\}dA$ instead of the measure $1.dA$. We require $X(A)$ to be such that

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} X^2 \exp\{-\tfrac{1}{2}A^2\}dA < \infty .$$

Then "closeness" is defined by

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \left[X - \sum_{i=1}^{\infty} a_i H_i(A)(i!)^{-\frac{1}{2}} \right]^2 \exp\{-\tfrac{1}{2}A^2\} dA \ .$$

The a_i's and b_i's are functions of the cumulants of X and Y. For large α and β we can find these a_i's and b_i's by using the Cornish-Fisher expansion (see for example Abramowitz and Stegun [(1972), p. 935]). Then we can approximate arbitrarily closely to X and Y by including the a_i and b_i terms whose order are greater than a specified power of $\alpha^{-\frac{1}{2}}$ and $\beta^{-\frac{1}{2}}$. Normally the Cornish-Fisher expansion is used for a standardized sum of n independent identically distributed random variables, distributed as X, where the standardized cumulants are increasing powers of $n^{-\frac{1}{2}}$. Here the cumulants of X are increasing powers of $\alpha^{-\frac{1}{2}}$, and each cumulant multiplies a finite linear combination of Hermite polynomials, each of which is merely a polynomial function of A. Thus X can be expanded in terms of powers of A. Similarly Y can be expanded in terms of powers of B, and $\omega(\alpha, \beta)$ evaluated.

For α equal to β,

$$\omega(\alpha, \beta) = \mathrm{Corr}(X, Y) = \rho\left[1 - 2(1-\rho)/9\alpha + O\left(\alpha^{-\frac{3}{2}}\right)\right]$$

and if α is not equal to β,

$$\omega(\alpha, \beta) = \mathrm{Corr}(X, Y) = \rho\left[1-\left\{(\alpha+\beta)-(2\alpha\beta)^{\frac{1}{2}}\rho\right\}/9\alpha \quad \ldots\right] .$$

One interesting point is that at least to the α^{-1} term we obtain the same expression for $\mathrm{Corr}(X, Y)$ if we use the Wilson and Hilferty (1931) approximation to the chi-square distribution for each marginal.

$$F_X(x;\ \alpha) = \Phi\left\{\left[(x/\alpha)^{\frac{1}{3}}-1 + 1/9\alpha\right](3\alpha)^{\frac{1}{2}}\right\} ,$$

that is

$$X \doteqdot \alpha\left(\tfrac{1}{3}A\alpha^{-\frac{1}{2}}+1-(9\alpha)^{-1}\right)^3 .$$

The validity of this approximation has only been examined for univariate gamma distributions although a slightly simplified version was examined by Jensen (1976) as an approximation to the bivariate gamma distribution of

Griffiths (1969a).

Note that the difference between the Cornish-Fisher expansion and the Wilson-Hilferty approximation is that in the former we can approximate arbitrarily closely to $F(X;\ \alpha)$ by including more and more terms of the expansion, whereas for the latter, for fixed X and α, the approximation deviates from $F(X;\ \alpha)$ by a fixed amount. Therefore we must check that it does closely approximate the moments as well as the cumulative distribution function.

When ρ is one, Moran's bivariate gamma distribution attains maximum correlation. Using the expansion for $\omega(\alpha,\ \beta)$, we have an approximation for $\nu(\alpha,\ \beta)$ at least for α and β large,

$$\nu(\alpha,\ \beta) \simeq \left[1 - \tfrac{1}{9}\left(\beta^{-\frac{1}{2}}-\alpha^{-\frac{1}{2}}\right)^2 + \ldots\right] < 1\ .$$

Even for α and β moderate, this approximation is reasonably close to the correct value. For example if $\alpha = 2$, $\beta = 1$, this approximation gives 0.9905 whereas the correct value is 0.9925.

For the bivariate gamma distribution of Moran there is a distribution for every positive ω less than or equal to $\nu(\alpha,\ \beta)$. Therefore there certainly are bivariate gamma distributions which can not be infinitely divisible for certain ranges of positive correlation.

6. Maximum Positive Correlation for Infinite Divisibility to be Possible

Since $\left\{\nu\left(2^{-n}\alpha,\ 2^{-n}\beta\right)\right\}_n$ is a decreasing sequence bounded below, the minimum occurs as n tends to infinity. Since $\nu(q\alpha,\ q\beta)$ is continuous in q this is equivalent to letting q decrease to zero. If

$$\upsilon(\alpha,\ \beta) = \lim_{q\to 0} \nu(q\alpha,\ q\beta)\ ,$$

then $\upsilon(\alpha,\ \beta)$ is the maximum value of ω for which it is possible that an element of $\mathbb{\Gamma}(\alpha,\ \beta;\ \omega)$ is infinitely divisible. We can derive an expression for $\upsilon(\alpha,\ \beta)$ from $\nu(q\alpha,\ q\beta)$ in the following manner:

$$\nu(q\alpha,\ q\beta) = (\beta/\alpha)^{\frac{1}{2}}\Gamma(\beta q+1)^{-1}\int_0^\infty Xe^{-Y}Y^{\beta q}dY - q(\alpha\beta)^{\frac{1}{2}}\ ,$$

where X is the unique function of Y given by

$$\Gamma(\alpha q)^{-1}\int_X^\infty e^{-u}u^{\alpha q-1}du = \Gamma(\beta q)^{-1}\int_Y^\infty e^{-u}u^{\beta q-1}du\ .$$

Letting $q \to 0$ we can see that

$$v(\alpha, \beta) = (\beta/\alpha)^{\frac{1}{2}} \int_0^\infty Xe^{-Y} dY ,$$

where X is the unique function of Y defined by

$$\int_X^\infty e^{-u}u^{-1}du = (\beta/\alpha) \int_Y^\infty e^{-u}u^{-1}du ,$$

that is

$$E_1(X) = (\beta/\alpha)E_1(Y) ,$$

where $E_1(\cdot)$ is the exponential integral defined previously.

If we let $q \to 0$, we are really interested in the correlation between the variables $X/(q\alpha)^{\frac{1}{2}}$ and $Y/(q\beta)^{\frac{1}{2}}$ where

$$E\{X/(q\alpha)^{\frac{1}{2}}\} = (q\alpha)^{\frac{1}{2}} \to 0 \quad \text{as} \quad q \to 0 ;$$

but

$$\mathrm{Var}\{X/(q\alpha)^{\frac{1}{2}}\} = 1 .$$

We have a positive random variable which tends in probability to zero, but leaves a thin smear of probability out at the tail so that the variance remains constant. From Kahaner and Waterman (1978), if we define $Q_\alpha(\xi)$ by

$$F_X(Q_\alpha; \alpha q) = \xi , \quad 0 < \xi < 1 ,$$

then $Q_\alpha(\xi)$ is a percentile of X and

$$Q_\alpha(\xi) \mathrel{\dot{\sim}} \xi^{1/\alpha q}$$

where

$$g(\alpha q) \mathrel{\dot{\sim}} f(\alpha q)$$

if

$$\lim_{\alpha q \to 0^+} f(\alpha q)/g(\alpha q) = 1 .$$

We know that $E\{X/(\alpha q)^{\frac{1}{2}}\} = (\alpha q)^{\frac{1}{2}}$, and this tends to zero fairly slowly compared with $\xi^{1/\alpha q}$. Thus for all $\xi < 1$, as q decreases, eventually

$$Q_\alpha(\xi) < (\alpha q)^{\frac{1}{2}} .$$

For example if $(\alpha q) = 1 \times 10^{-3}$ and $\xi = 0.99$ then

$$Q_\alpha(\xi) \simeq 4 \times 10^{-5} .$$

However if we try using the approximation of Kahaner and Waterman,

$$X \mathrel{\dot{\widetilde{}}} Y^{\alpha/\beta} \tag{3.3}$$

we would obtain

$$v(\alpha,\ \beta) \approx (\beta/\alpha)^{\frac{1}{2}}\Gamma(\alpha/\beta + 1) > 1 \quad \text{for} \quad \alpha = 2\ , \quad \beta = 1\ .$$

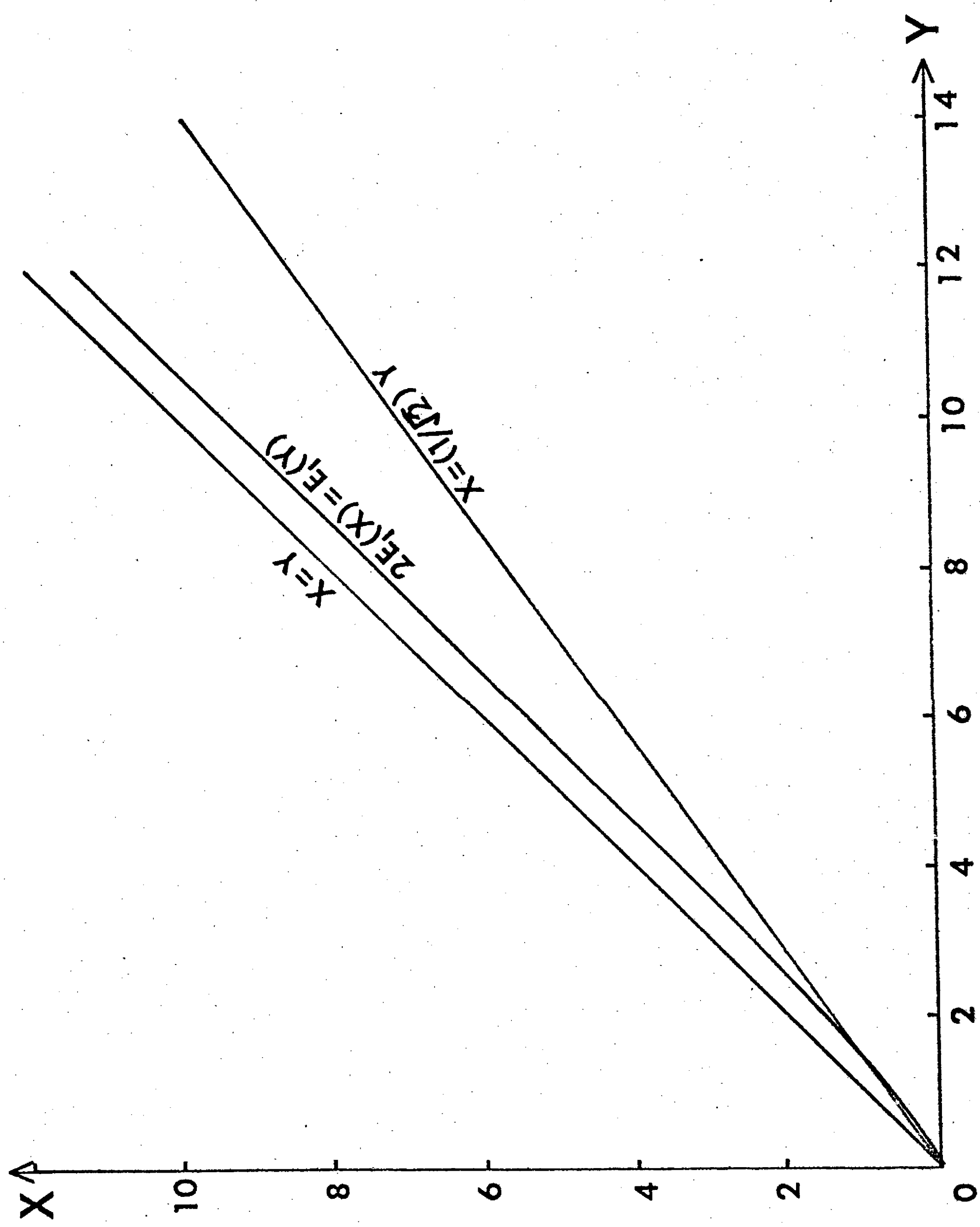
The error occurs because the approximation is only valid when X and Y both tend to zero. There is sufficient probability in the tails of the distributions of X and Y to upset (3.3) as a global approximation. The major difficulty in obtaining algebraic bounds for the integral involved in $v(\alpha,\ \beta)$ is an inability to find a suitable approximation for X as a function of Y, and this is tied up with the inability in this context to use adequate approximations to the function $E_1(t)$.

It is possible to integrate $v(\alpha,\ \beta)$ numerically, since we can numerically find X as a function of Y or *vice-versa*, using Abramowitz and Stegun [(1972), eq. 5.1.53 and 5.1.56]. Equation 5.1.53 gives a polynomial approximation to $E_1(x)$ for $0 \le x \le 1$ with claimed error less than 2×10^{-7} . Equation 5.1.56 gives a rational approximation to $xe^{x}E_1(x)$ for $1 \le x \le \infty$ with claimed error less than 2×10^{-8} .

$v(\alpha,\ \beta)$ can also be written as

$$v(\alpha,\ \beta) = (\alpha/\beta)^{\frac{1}{2}} \int_0^\infty Ye^{-X}dX ,$$

and this latter integral was used to evaluate $v(\alpha,\ \beta)$. The sum replacing this integral converged faster than the analogous sum replacing $\int_0^\infty Xe^{-X}dX$, which in turn converged faster than the sum replacing $\int_0^\infty Xe^{-Y}dY$. $E_1(X)$ is calculated, and $(\alpha/\beta)E_1(X)$ found. Then Y is the solution of



Fig· 3·1

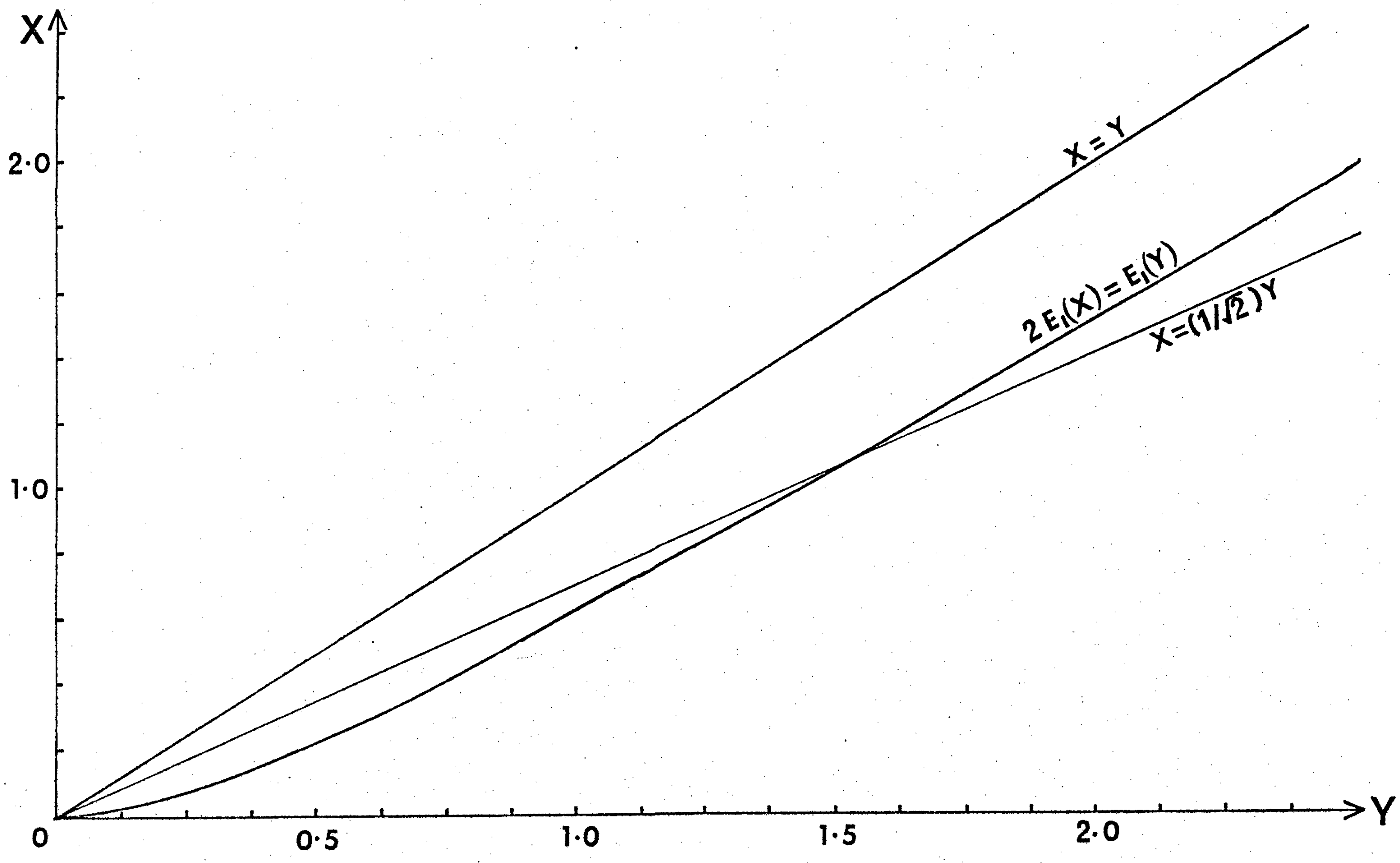


Fig. 3·2

$E_1(Y) - c = 0$, where $c = (\alpha/\beta)E_1(X)$, is found by a Newton-Raphson iterative technique. The only difficulty occurring was that of selecting a suitable initial value for Y for the iterative procedure when X was small. The methods used to approximate the integral were the same as described in the last chapter. An interval width of 0.1 was sufficient to obtain $v(\alpha, \beta)$ to four figures, although this was checked using an interval width of 0.05 .

α/β	$v(\alpha, \beta)$	α/β	$v(\alpha, \beta)$
1	1	6	0.8534
2	0.9753	7	0.8309
3	0.9400	8	0.8108
4	0.9075	9	0.7927
5	0.8787	10	0.7762

TABLE IV

$v(\alpha, \beta)$ is quite high and this raises the interesting question of whether we can find any infinitely divisible bivariate gamma distributions with correlation between $(\beta/\alpha)^{\frac{1}{2}}$ and $\nu(\alpha, \beta)$.

Figures 3.1 and 3.2 give

$$E_1(X) = (\beta/\alpha)E_1(Y)$$

in the case where $\beta/\alpha = 0.5$. As a comparison the lines $X = Y$ and $X = (\beta/\alpha)^{\frac{1}{2}}Y$ are also given.

7. Bivariate Poisson Distribution

Analogous questions do not arise for the bivariate Poisson distribution, that is the class of bivariate distributions which have Poisson marginals. Dwass and Teicher (1957) proved that for a bivariate Poisson distribution to be infinitely divisible it must have characteristic function given by

$$\chi_{X,Y}(s, t) = \exp\{\alpha(e^{is}-1)+\beta(e^{it}-1)+\gamma(e^{i(s+t)}-1)\} ,$$

for some non-negative α, β and γ . This is precisely the characteristic function of the bivariate Poisson distribution given by Campbell (1934) and Holgate (1964). That is, if U_1, U_2 , and U_3 are independent Poisson random variables with respective means α, β and γ then,

$$X = U_1 + U_3 \quad \text{and} \quad Y = U_2 + U_3$$

has the above characteristic function and

$$\mathrm{Corr}(X,\ Y) = \gamma/\left[(\alpha+\gamma)^{\frac{1}{2}}(\beta+\gamma)^{\frac{1}{2}}\right] .$$

The maximum correlation between X and Y is the minimum of the square roots of the ratios of the marginal means, and this was pointed out by Holgate (1964).

8. Other Familes of Distributions Closed Under Convolution

The family of gamma distributions is an example of a family of distributions that is closed under convolution, if U_1 and U_2 have gamma distributions and are independent, then $U_1 + U_2$ has a gamma distribution. Many but not all of these families are also infinitely divisible. The normal, gamma, Poisson and Cauchy distributions are infinitely divisible, but the m-dimensional Wishart distribution for $m > 2$ is not infinitely divisible, as shown by Lévy (1948), and neither is the binomial distribution.

Since gamma distributions are closed under convolution it is possible to make statements about the maximum correlation possible between marginal variates in an infinitely divisible bivariate gamma distribution. This is not needed in considering the bivariate Poisson distribution since there is only one infinitely divisible bivariate distribution with Poisson marginals. From this construction we can see that we can always find at least one infinitely divisible bivariate distribution with marginals closed under convolution, if their marginals are also infinitely divisible.

Let U_1, U_2 and U_3 be independent elements of a family of distributions which is closed under convolution and each element of which is infinitely divisible. Then

$$X = U_1 + U_3 \quad \text{and} \quad Y = U_2 + U_3 ,$$

belong to the same family of random variables. $(X,\ Y)$ has characteristic function

$$\chi_{X,Y}(s,\ t) = \chi_{U_1}(s)\chi_{U_2}(t)\chi_{U_3}(s+t) ,$$

and this is obviously infinitely divisible. If in addition the first two moments of U_i , $i = 1, 2, 3$ are finite and v_i is the variance of U_i ,

$i = 1, 2, 3$, then

$$\omega = \mathrm{Corr}(X, Y) = v_3/\{(v_1+v_3)(v_2+v_3)\}^{\frac{1}{2}} ,$$

the extreme case occurring when $X = U_1 + U_3$ and $Y = U_3$. For bivariate distributions with Poisson marginals this construction yields the only infinitely divisible distribution. The construction is valid even for marginal distributions which have no moments such as the Bessel function distribution considered by Feller (1966):

$$f_X(x) = e^{-x}(\theta/x)I_\theta(x) , \quad x \geq 0 , \quad \theta > 0 .$$

However for such distributions consideration of correlation is inappropriate.

In theory if we had a bivariate distribution whose marginals were infinitely divisible, closed under convolution and whose first two moments were finite, we could analyse it in a similar manner to that done for the bivariate distribution with gamma marginals. In practice this may not be possible.

9. Conclusion

We still have not shown whether the bivariate gamma distribution of Moran is infinitely divisible if $0 < \omega \leq v(\alpha, \beta)$. As we can see from Table IV, this includes most of the positive possible range of ω . This problem is extremely difficult to solve and will not be attempted here. Also although we have given $v(\alpha, \beta)$ as the maximum correlation possible if (X, Y) is to be infinitely divisible we have not exhibited any infinitely divisible distributions with

$$(\beta/\alpha)^{\frac{1}{2}} < \omega \leq v(\alpha, \beta) .$$

However we have shown that Moran's bivariate gamma distribution can not be infinitely divisible for negative ω and some positive ω if $\alpha \neq \beta$, and highlighted some interesting points about bivariate distributions which have non-negative marginal variables.

CHAPTER 4

THE CIRCULAR STRUCTURAL MODEL

1. Introduction

Suppose we have a finite set of pairs of observations $(X_1, Y_1), \ldots, (X_N, Y_N)$, which are independent. Further suppose there is a set of N underlying unobservable variables (U_i, V_i) such that

$$\begin{aligned} X_i &= U_i + \varepsilon_i , \\ Y_i &= V_i + \nu_i , \end{aligned} \qquad i = 1, 2, \ldots, N , \tag{4.1}$$

where $\{(\varepsilon_1, \nu_1), \ldots\}$ is a sequence of independent bivariate random variables with zero means and finite variances. U_i and V_i are assumed connected by a mathematical functional relationship involving a vector of parameters $\boldsymbol{\zeta}$; let us say

$$k(U_i, V_i, \zeta) = 0 . \tag{4.2}$$

The problem is to estimate consistently $\boldsymbol{\theta}$, the vector of structural parameters, which includes $\boldsymbol{\zeta}$ and the parameters of the bivariate distribution of (ε_i, ν_i) . For example the relationship could be of a linear nature of the form, $V_i = \alpha + \beta U_i$. If the U_i's are random variables then (4.1) is a structural relationship; if the U_i's form a sequence of fixed numbers then (4.1) is a functional relationship. Structural and functional relationships are particularly interesting because they introduce the possibility of non-identifiability of $\boldsymbol{\theta}$, essentially meaning that there are more parameters to estimate than statistics available to estimate them. There has been much literature about this problem, especially for the linear relationship. For a review of the linear relationship case see Moran (1971) and also Dolby (1976).

Here we will consider the problem of estimating consistently the parameters of a circular structural relationship, but will touch briefly on the allied functional relationship. Estimation in the functional relationship has been previously considered by Robinson (1961) using least squares, and by Chan (1965) using the method of minimum squared distance, which is essentially the same as least squares. Chan also considered consistency of these estimators. The circular model is defined as follows:

$$\begin{aligned} X_i &= \xi + \rho \cos \tau_i + \varepsilon_i \\ & \qquad\qquad\qquad\qquad i = 1, 2, \ldots, N , \\ Y_i &= \eta + \rho \sin \tau_i + \nu_i , \end{aligned} \tag{4.3}$$

where ε_i and ν_i are independent and identically distributed normal variables with variance σ^2. $\boldsymbol{\theta} = (\xi, \eta, \rho, \sigma^2)$ is the vector of structural parameters. The structural relationship in which the τ_i's are independent and identically distributed is the model discussed here. Before estimation can be considered it is necessary to establish the identifiability of the structure.

2. Identifiability of the Structural Model

It is assumed that the τ_i's are independent and identically distributed random variables with the same unknown cumulative distribution function $G(\tau)$. There is no necessity for $G(\tau)$ to be parametric with a finite parameter set, as an assumption of this kind would reduce the problem of consistency to the classical case considered by Wald (1949).

A parameter is said to be identifiable if it can be determined uniquely from the knowledge of the distribution of the observed random variables, here X_i and Y_i, or equivalently from their joint characteristic function. Note that the concept of identifiability applies to the distribution itself rather than to a finite set of observations from the distribution. The joint characteristic function of X_i and Y_i is

$$\chi_{X,Y}(s, t) = \exp\left(-\tfrac{1}{2}s^2\sigma^2-\tfrac{1}{2}t^2\sigma^2+i\xi s+i\eta t\right)\psi(s, t) ,$$

where

$$\psi(s, t) = \int_{[0,2\pi)} \exp(is\rho \cos \tau + it\rho \sin \tau)dG(\tau) ,$$

using the known characteristic function of the normal distribution.

Consider the one dimensional characteristic function found by putting $t = 0$,

$$\chi_X(s) = \exp\left(-\tfrac{1}{2}s^2\sigma^2 + i\xi s\right)\psi(s) ,$$

where,

$$\psi(s) = \int_{[0,2\pi)} \exp(is\rho \cos \tau)dG(\tau) .$$

$\exp(i\xi s)\psi(s)$ is the characteristic function of a distribution contained in the bounded interval $[\xi-\rho, \xi+\rho]$ and is therefore analytic and has no normal factors. Let $\exp\left(-\tfrac{1}{2}s^2\sigma_1^2\right)$ be the characteristic function of the largest normal factor in the distribution defined by $\chi_X(s)$. From the above $\sigma_1^2 = \sigma^2$, and $\exp(i\xi s)\psi(s)$ is uniquely determined by $\chi_X(s)$. Similarly the distribution of $\eta + \rho \sin \tau$ is uniquely determined by $\chi_Y(t)$, the characteristic function found by putting $s = 0$ in $\chi_{X,Y}(s, t)$. Now consider the variable given by

$$v = (1-\lambda)(\xi+\rho \cos \tau) + \lambda(\eta+\rho \sin \tau) .$$

This has characteristic function

$$\exp\left[-\tfrac{1}{2}\{(1-\lambda)^2+\lambda^2\}\sigma^2 t^2 + i\{\xi(1-\lambda)+\lambda\eta\}t\right]\psi_1(t) ,$$

where

$$\psi_1(t) = \int_{[0,2\pi)} \exp[it\{(1-\lambda)\rho \cos \tau + \lambda\rho \sin \tau\}]dG(\tau) .$$

As above it is clear that the distribution of v can be found uniquely from $\chi\left((1-\lambda)t, \lambda t\right)$ as long as λ is finite. By Cramér and Wold (1936) the probability distribution on the plane is then uniquely determined from $\chi(s, t)$. For identifiability the circle must also be uniquely determined, and this will only be the case if $G(\tau)$ is a distribution function whose support consists of at least three points.

Although identification is necessary for estimation more conditions may be required for any particular estimation procedure, for example estimation by maximum likelihood.

There are actually four different estimation problems possible, and only for the third is the above analysis needed. $G(\tau)$ could be known exactly, for example $G(\tau)$ could be the uniform distribution. Here all the assumptions of Wald (1949) are satisfied and the problem reduces to the classical maximum likelihood estimation of θ . This will be discussed in section 4. $G(\tau)$ could be known except for a fixed number of parameters denoted by υ . Again the assumptions of Wald are satisfied and the problem reduces to the maximum likelihood estimation of θ and υ , a slightly larger but still finite parameter set. Thirdly suppose $G(\tau)$ is not known. Estimation of G then requires a parameter space which is a function space. This is discussed in the next section. The final case is the functional relationship where nothing is known about the τ_i's , and does not involve a $G(\tau)$. Here the τ_i's , called incidental or nuisance parameters, must be eliminated from the likelihood using for example the approach of Andersen (1973). However it can be shown that no such method is applicable to the circular functional relationship and a different approach must be employed there. This will be briefly mentioned in section 12.

3. Maximum Likelihood Estimation for the Structural Model

Kiefer and Wolfowitz (1956) showed that under certain assumptions the maximum likelihood estimators of the structural parameters are strongly consistent, and the maximum likelihood estimator of G , the distribution function of the τ_i's , converges to G at every point of continuity of G with probability one. The essential idea of their proof is based on Wald's 1949 proof of the consistency of the maximum likelihood estimator of a parameter, and the assumptions here bear a marked similarity to those of Wald. It is however a deeper result, since one of the parameters here is actually a function, so the theory involves estimation of a point in a function space. As before write $\theta = (\xi, \eta, \rho, \sigma^2)$, and let $f(X, Y|\theta, \tau)$ denote the probability density of (X, Y) given parameters θ and a fixed value of the random variable τ . Then

$$f(X_i, Y_i|\theta, \tau_i)$$

$$= (2\pi\sigma^2)^{-1}\exp\left[-\tfrac{1}{2}\sigma^{-2}\left[(X_i-\xi-\rho\cos\tau_i)^2+(Y_i-\eta-\rho\sin\tau_i)^2\right]\right] . \quad (4.4)$$

The five assumptions of Kiefer and Wolfowitz must be verified.

ASSUMPTION 1. $f(X, Y|\theta, \tau)$ is a density with respect to a σ-finite measure on a Euclidean space on which (X, Y) is the generic point. In fact here f is Riemann integrable with respect to X and Y , so this assumption holds.

Let Ω be the space of possible values of θ , and let $T = [0, 2\pi)$ be the space of values of τ . f is jointly measurable in (X, Y) and τ for each θ . Let $\theta_t^{(s)}$ $(1 \le s \le k)$ be the sth component of the point θ_t in Ω . Here $k = 4$, since there are four structural parameters. Also note that $G(\tau)$ is of bounded variation on T , hence is Riemann integrable on T . Let $\Gamma = \{G\}$ be a given space of cumulative distribution functions of τ . We could start by assuming Γ to be the set of all distribution functions on T , reducing this set as necessary. Let θ_0 and G_0 be respectively, the true value of the parameter θ , and the true distribution of τ . It is assumed that $\theta_0 \in \Omega$ and $G_0 \in \Gamma$. Let $\kappa = (\theta, G)$ be the generic point in $\Omega \times \Gamma$, and $\kappa_0 = (\theta_0, G_0)$. Define

$$f(X, Y|\kappa) = \int_T f(X, Y|\theta, \tau)dG(\tau) .$$

In the space $\Omega \times \Gamma$ define the metric

$$\delta(\kappa_1, \kappa_2) = \delta([\theta_1, G_1], [\theta_2, G_2])$$

$$= \sum_{s=1}^{4} \left|\operatorname{artan} \theta_1^{(s)} - \operatorname{artan} \theta_2^{(s)}\right| + \int_T |G_1(\tau)-G_2(\tau)|d\tau . \qquad (4.5)$$

Except for the deletion of the term $\exp - |\tau|$ in the integral, this definition appears in Kiefer and Wolfowitz (1956) and also in Wolfowitz (1957) as a convenient definition of distance rather than the only definition that can be employed. Let $\overline{\Omega} \times \overline{\Gamma}$ be the completed space of $\Omega \times \Gamma$, the space together with the limits of its Cauchy sequences in the sense of the metric (4.5). We also require $\overline{\Omega}$ and $\overline{\Gamma}$ to be compact, so that $\overline{\Omega} \times \overline{\Gamma}$ is a compact metric space by Tychonoff's theorem (see Kingman and Taylor [(1966), p. 39]).

Now look at Ω and $\overline{\Omega}$. The set of real numbers R is not compact. By Kingman and Taylor [(1966), p. 34], R can be compactified by adjoining the two points $+\infty$ and $-\infty$. Also all closed and bounded intervals of R are compact, which will prove useful later. Let

$$\Omega = \{\theta : \xi, \eta \in R, \rho \in (R \backslash R^-), \text{ and } \sigma^2 \ge c > 0\} ;$$

then Ω is compact by Tychonoff's theorem. Notice that we have assumed that $\sigma^2 \ge c$ where c is some strictly positive number. If we had merely assumed $\sigma^2 > 0$, then to compactify Ω we must also include the point $\sigma^2 = 0$, which creates great difficulties in the other assumptions. We start by assuming Γ is the set of all distributions on the interval $[0, 2\pi)$ and note that $\overline{\Gamma}$ must therefore include at least some distribution functions on $\overline{T} = [0, 2\pi]$. To consider $\overline{\Gamma}$ further we will look at the assumptions of Kiefer and Wolfowitz which potentially restrict Γ .

ASSUMPTION 2 requires extending the definition of $f(X, Y|\kappa)$ so that the range of κ will be $\overline{\Omega} \times \overline{\Gamma}$ and so that for any $\{\kappa_i\}$ and κ^* in $\overline{\Omega} \times \overline{\Gamma}$, $\kappa_i \to \kappa^*$ defined by $\delta(\kappa_i, \kappa^*) \to 0$ implies $f(X_i, Y_i|\kappa_i) \to f(X, Y|\kappa^*)$ except perhaps on a set of (X, Y) whose probability is zero according to the probability density $f(X, Y|\kappa_0)$. The exceptional set of (X, Y) may depend on κ^* , and $f(X, Y|\kappa^*)$ need not be a probability density function. Put $f(X, Y|\theta, \tau) = 0$, whenever $\xi = \pm\infty$, $\eta = \pm\infty$ and $\sigma^2 = +\infty$.

Now consider

$$G_i(\tau) = \begin{cases} \frac{1}{2} , & 0 \le \tau < 2\pi(1 - 1/i) , \quad i \ge 1 , \\ 1 , & 2\pi(1 - 1/i) \le \tau < 2\pi ; \end{cases}$$

then $\delta(G_i, G^*) \to 0$ where $G^*(\tau)$ has a jump of $\frac{1}{2}$ at $\tau = 0$ and a jump of $\frac{1}{2}$ at $\tau = 2\pi$ and is not therefore in Γ . Since $f(X, Y|\kappa)$ is the integral of $f(X, Y|\theta, \tau)$ with respect to $G(\tau)$ over T it is easy to show that

$$f(X, Y|\kappa_i) \not\to f(X, Y|\kappa^*) .$$

We can show that

$$f(X, Y|\kappa_i) \to \left(2\pi\sigma^2\right)^{-1}\exp\left\{-\tfrac{1}{2}\sigma^{-2}\left[\left(X_i-\xi-\rho\right)^2+\left(Y_i-\eta\right)^2\right]\right\}$$

whereas

$$f(X, Y|\kappa^*) = \left(4\pi\sigma^2\right)^{-1}\exp\left\{-\tfrac{1}{2}\sigma^{-2}\left[\left(X_i-\xi-\rho\right)^2+\left(Y_i-\eta\right)^2\right]\right\} ,$$

and thus it does not converge for any (X, Y) . Certainly it does not converge for a set (X, Y) whose probability is positive according to the probability density $f(X, Y|\kappa_0)$. Therefore we should extend the definition of $f(X, Y|\kappa)$ to be the integral of $f(X, Y|\theta, \tau)$ with respect to $G(\tau)$ over the set $\overline{T}$, instead of T.

We want

$$\int_T f(X, Y|\theta_i, \tau)dG_i(\tau) \to \int_{\overline{T}} f(X, Y|\theta^*, \tau)dG^*(\tau) \tag{4.6}$$

if $\delta([\theta_i, G_i], [\theta^*, G^*]) \to 0$.

$\overline{\Omega} \times \overline{\Gamma}$ is a compact metric space and is therefore separable. From Graves (1956), (4.6) is satisfied since

(a) $G_i(\tau)$ and $G^*(\tau)$ are of uniformly bounded variation on $\overline{T}$,

(b) $f(X, Y|\theta_i, \tau)$ and $f(X, Y|\theta^*, \tau)$ are continuous functions of τ on $\overline{T}$,

(c) since $\overline{\Omega} \times \overline{\Gamma}$ is separable, if $\delta([\theta_i, G_i], [\theta^*, G^*]) \to 0$, then $\delta(\theta_i, \theta^*) \to 0$. $f(X, Y|\theta, \tau)$ is a continuous function of $\boldsymbol{\theta}$ for all τ on $\overline{T}$, so that it can be shown that

$$\lim_{i\to\infty} f(X, Y|\theta_i, \tau) = f(X, Y|\theta^*, \tau)$$

uniformly on $\overline{T}$.

(d)
$$\begin{aligned}\delta(G_i, G^*) &= \int_T |G_i(\tau)-G^*(\tau)|d\tau \\ &\geq \int_0^a |G_i(\tau)-G^*(\tau)|d\tau \\ &\geq \left|\int_0^a \{G_i(\tau)-G^*(\tau)\}d\tau\right| \\ &= \left|\int_0^a G_i(\tau)d\tau - \int_0^a G^*(\tau)d\tau\right| \geq 0 .\end{aligned}$$

As $i \to \infty$, $\delta(G_i, G^*) \to 0$, so for any a in $[0, 2\pi)$,

$$\lim_{i\to\infty} \int_0^a G_i(\tau)d\tau = \int_0^a G^*(\tau)d\tau .$$

By the continuity of $\int_0^t G_i(\tau)d\tau$ regarded as a function of $G(\tau)$,

$\lim_{i\to\infty} G_i(2\pi) = G^*(2\pi)$ and $\lim_{i\to\infty} G_i(0) = G^*(0)$.

ASSUMPTION 4 is the identifiability assumption. If κ_1 in $\bar{\Omega} \times \bar{\Gamma}$ is different from κ_0 , the true value of (θ, G) then for at least one $z = (z_1, z_2)$,

$$\int_{-\infty}^{z_1} dx \int_{-\infty}^{z_2} f(x, y|\kappa_1)dy \neq \int_{-\infty}^{z_1} dx \int_{-\infty}^{z_2} f(x, y|\kappa_0)dy .$$

This is not quite the same thing as requiring that the structure be identifiable for all κ in $\bar{\Omega} \times \bar{\Gamma}$, but requires more than identifiability for κ in $\Omega \times \Gamma$. From section 2 one of the crucial steps of the identifiability of the normal factor relies on $[\xi-\rho, \xi+\rho]$ being a bounded interval. This will only be the case if ρ is finite. Therefore Ω must be restricted so that $\rho \leq \rho^*$, where ρ^* is large but finite. Further Γ must be restricted to the set of distribution functions whose support on T is not less than three points. In section two we showed that for identifiability of θ difficulty only occurs if we do not have three distinct points on the circle. However $f(x, y|\kappa_1)$ and $f(x, y|\kappa_2)$ can only be indistinguishable for all x and y if both $G_1(\tau)$ and $G_2(\tau)$ have less than three points of support on T , or $\theta_1 = \theta_2$ and G_1 has a non-zero jump at $\tau = 0$. The latter causes problems if $dG_1(\tau)$ and $dG_2(\tau)$ are identical on $T\backslash\{0\}$ but $G_1(\tau)$ has a non-zero jump at $\tau = 0$ and $G_2(\tau)$ has a non-zero jump of the same size at $\tau = 2\pi$. One solution is to restrict Γ to distribution functions which do not have a non-zero jump at $\tau = 0$. Another alternative is to recognize that $\tau = 0$ and $\tau = 2\pi$ is the same point and not try to distinguish between them. We will assume that the first holds. Also if κ_0 is the correct "parameter" value G_0 has more than two points of support, and we can distinguish between $f(x, y|\kappa_0)$ and $f(x, y|\kappa_1)$ where G_1 has less than three points of support. Thus this assumption is valid here.

ASSUMPTION 3. For any κ in $\bar{\Omega} \times \bar{\Gamma}$ and any $\zeta > 0$, $\nu(X, Y|\kappa, \zeta)$ is a measurable function of (X, Y) where

$$\nu(X, Y|\kappa, \zeta) = \sup f(X, Y|\kappa')$$

where the supremum is taken over all κ' in $\overline{\Omega} \times \overline{\Gamma}$ for which $\delta(\kappa, \kappa') < \zeta$. From Graves (1956),

$$\left|\int_{\overline{T}} f(X, Y|\theta, \tau)dG(\tau)\right| \leq LV_{\overline{T}}(G)$$

where $|f| \leq L$ and $V_{\overline{T}}(G)$ is the total variation of G on $\overline{T}$. If $G(\tau) = G_1(\tau) - G_2(\tau)$, where $|G_1(\tau)-G_2(\tau)| < \varepsilon$, then

$$V_{\overline{T}}(G) \leq 2\pi\varepsilon .$$

Thus

$$\left|\int_{\overline{T}} f(X, Y|\theta, \tau)dG_1(\tau) - \int_{\overline{T}} f(X, Y|\theta, \tau)dG_2(\tau)\right| \leq 2L\pi\varepsilon ,$$

whenever

$$|G_1(\tau)-G_2(\tau)| < \varepsilon .$$

So $f(X, Y|\theta, G)$ is continuous in G . Since $\sigma^2 \geq c$, then $f(X, Y|\theta, G)$ is also continuous in (θ, G) and assumption 3 holds, since $f(X, Y|\theta, G)$ is also bounded.

ASSUMPTION 5. For any κ in $\overline{\Omega} \times \overline{\Gamma}$ we have as $\zeta \downarrow 0$,

$$\lim \iint \left\{\log\left[\frac{\nu(X,Y|\kappa,\zeta)}{f(X,Y|\kappa_0)}\right]\right\}^+ f(X, Y|\kappa_0)dXdY < \infty .$$

With the assumption that $\sigma^2 \geq c$, this clearly holds.

A maximum likelihood estimate is defined as a pair $(\hat{\theta}_N, \hat{G}_N) \in \overline{\Omega} \times \overline{\Gamma}$ such that

$$\prod_{i=1}^{N} f(X_i, Y_i|\hat{\theta}_N, \hat{G}_N) \geq \prod_{i=1}^{N} f(X_i, Y_i|\theta, G) \tag{4.7}$$

for all $(\theta, G) \in \overline{\Omega} \times \overline{\Gamma}$, if such a $(\hat{\theta}_N, \hat{G}_N)$ exists. Kiefer and Wolfowitz (1956) also say that if any maximum likelihood estimator defined by (4.7) exists, then it is consistent. They do not give conditions which ensure the existence of maximum likelihood estimators, but do say that reasonable conditions are not difficult to formulate. They also do not consider uniqueness of the maximum likelihood estimators.

Here, for fixed (X_i, Y_i) , $f(X_i, Y_i|\theta, G)$ is a real valued continuous function of (θ, G) , where (θ, G) is an element of a compact space $\bar{\Omega} \times \bar{\Gamma}$. By Graves (1956) a real valued continuous function defined on a compact metric space has attainable maximum and minimum values. This also applies to the likelihood function $\prod_{i=1}^{N} f(X_i, Y_i | \theta, G)$, so $(\hat{\theta}_N, \hat{G}_N)$ does exist in this case.

The major difficulty with this approach in practice is that $\hat{G}_N$ must still be found, and there is no general method for doing this. The difficulty in estimating $\hat{G}_N$ is akin to the estimation of a mixing distribution. For example Choi and Bulgren (1968) gave a procedure for estimating a mixing distribution where the observed random variable is univariate. Their procedure still requires finding an estimator of the mixing distribution from a function space so that the resultant distribution of the random variable is in some sense close to the empiric distribution of the observed random variable. To define what they mean by close, they need a metric and since their argument is based on the general concept of Wolfowitz (1957), the difficulties are very similar to those described above.

4. Specific Distribution Assumed for τ_i

The conditional probability density of (X_i, Y_i) given τ_i is given by

$$f\{(X_i, Y_i) | \tau_i\}$$
$$= (2\pi\sigma^2)^{-1} \exp\left[-\tfrac{1}{2}\sigma^{-2}\left\{(X_i-\xi-\rho \cos \tau_i)^2+(Y_i-\eta-\rho \sin \tau_i)^2\right\}\right] . \quad (4.8)$$

It is interesting to examine the case where τ has a uniform distribution on the interval $[0, 2\pi)$. Then by integration of (4.8) we get

$$f(X_i, Y_i) = (2\pi\sigma^2)^{-1} \exp\left[-\tfrac{1}{2}\sigma^{-2}\left\{(X_i-\xi)^2+(Y_i-\eta)^2+\rho^2\right\}\right] I_0(r_i\rho/\sigma^2) ,$$

where $r_i = \left\{(X_i-\xi)^2+(Y_i-\eta)^2\right\}^{\frac{1}{2}}$ and $I_\nu(z)$ is the modified Bessel function of the first kind and of order ν ,

$$I_\nu(z) = \sum_{m=0}^{\infty} (\tfrac{1}{2}z)^{2m+\nu}/\{m!\Gamma(m+\nu+1)\} .$$

If we have a sample of size N, then the log-likelihood is given by

$$l = -N \log 2\pi - N \log \sigma^2 - \tfrac{1}{2}\sigma^{-2} \sum_{i=1}^{N} \left\{(X_i-\xi)^2+(Y_i-\eta)^2+\rho^2\right\}$$
$$+ \sum_{i=1}^{N} \log I_0\left(r_i\rho/\sigma^2\right) . \quad (4.9)$$

Then the maximum likelihood estimates $\hat{\theta} = (\hat{\xi}, \hat{\eta}, \hat{\rho}, \hat{\sigma}^2)$ are given by the solutions of the equations:

$$\sum_{i=1}^{N} (X_i-\xi) - \rho \sum_{i=1}^{N} \{(X_i-\xi)/r_i\} I_{1,0}\left(r_i\rho/\sigma^2\right) = 0 ; \quad (4.10)$$

$$\sum_{i=1}^{N} (Y_i-\eta) - \rho \sum_{i=1}^{N} \{(Y_i-\eta)/r_i\} I_{1,0}\left(r_i\rho/\sigma^2\right) = 0 ; \quad (4.11)$$

$$\rho - N^{-1} \sum_{i=1}^{N} r_i I_{1,0}\left(r_i\rho/\sigma^2\right) = 0 ; \quad (4.12)$$

and

$$\sigma^2 - (2N)^{-1} \sum_{i=1}^{N} \left(r_i^2-\rho^2\right) = 0 , \quad (4.13)$$

where we write

$$I_{a,b}(z) = I_a(z)/I_b(z) ,$$

and since

$$I_0'(z) = I_1(z) .$$

These equations are not immediately soluble and would have to be solved iteratively. Notice that $\bar{X} = N^{-1} \sum X_i$ and $\bar{Y} = N^{-1} \sum Y_i$ are also consistent estimators of ξ and η, and thus $(\bar{X}, \bar{Y})$ could provide a convenient initial value for $(\hat{\xi}, \hat{\eta})$ in any iterative procedure.

A further aspect that must be considered is whether or not there exists a unique solution to the set of likelihood equations. One obvious solution of (4.10) to (4.13) is $(\hat{\xi}, \hat{\eta}, \hat{\rho}, \hat{\sigma}^2) = (\bar{X}, \bar{Y}, 0, \sigma^{*2})$ where

$$\sigma^{*2} = (2N)^{-1} \sum_{i=1}^{N} \left\{(X_i-\bar{X})^2+(Y_i-\bar{Y})^2\right\} .$$

Using the property obvious from the series,

$$\lim_{z \to 0} z^{-1} I_{1,0}(z) = \tfrac{1}{2} ,$$

it can be shown that the second and third derivatives of l with respect to ρ vanish at $(\overline{X}, \overline{Y}, 0, \sigma^{*2})$ unless σ^{*2} is zero. Further the fourth derivative of l with respect to ρ is negative at this point, indicating that the point $(\overline{X}, \overline{Y}, 0, \sigma^{*2})$ is a local maximum in the ρ direction. By analysis of the ρ derivatives of l we can not eliminate the possibility that $(\overline{X}, \overline{Y}, 0, \sigma^{*2})$ is a local maximum of the likelihood surface. Next we assume ξ and η known, and look at the (ρ, σ^2) plane. At $(\rho, \sigma^2) = (0, \sigma^{*2})$ both $\partial^2 l/\partial\rho^2$ and $\partial^2 l/\partial\rho\partial(\sigma^2)$ vanish. Therefore

$$D = \left(\partial^2 l/\partial\rho^2\right)\left(\partial^2 l/\partial(\sigma^2)^2\right) - \left(\partial^2 l/\partial\rho\partial\sigma^2\right)^2 = 0 .$$

No conclusion regarding maximum or minimum of the function can be drawn from the vanishing of this quadratic form in the second derivatives, at the critical point. It is simpler to analyse the likelihood directly rather than consider higher order derivatives. If this is done we can see that $(0, \sigma^{*2})$ is not a local maximum of the likelihood surface in the (ρ, σ^2) plane, and is not therefore a local maximum in the $(\xi, \eta, \rho, \sigma^2)$ space.

The problem of uniqueness still remains unclear. Further examination of the likelihood equations is difficult because of the complicated nature of the equations and the problem will be difficult to resolve.

5. The Behaviour of $I_{1,0}(z)$

Before continuing with the examination of the estimators we need to know more about the behaviour of $I_{1,0}(z)$. From Hartman and Watson (1974) it is known that $I_{1,0}(z)$ is a continuous distribution function on $z \geq 0$ and a monotonically increasing function of z . Both $I_0(z)$ and $I_1(z)$ have been tabulated for z up to $z = 20$ in the British Association Mathematical Tables, volume VI (1950) and volume X (1952). For z greater than 20 , the asymptotic series given in Abramowitz and Stegun [(1972), eq. 9.7.1] can be used. Actually this series is also sufficiently accurate for a range of z values less than 20 . Also computer algorithms for calculating $I_\nu(z)$ for $z \geq 0$ do exist (see for example Amos *et al* (1977)). For computation of the ratio $I_{1,0}(z)$, Gautschi and Slavik (1978) remark that the algorithm used by Amos *et al* is equivalent to Gauss' continued fraction, whereas it may be advantageous to base the algorithm on the lesser known continued fraction of Perron especially if $z \gg 0$ (see Gautschi and Slavik for details).

Amos (1974) also examined bounds for $I_{\nu+1}(z)/I_\nu(z)$. If $\nu = 0$ these yield

$$z/\left\{1+\left(z^2+1\right)^{\frac{1}{2}}\right\} \leq I_{1,0}(z) \leq z/\left\{z^2+4\right\}^{\frac{1}{2}}$$

and

$$z/\left\{\tfrac{1}{2}+\left(z^2+\tfrac{9}{4}\right)^{\frac{1}{2}}\right\} \leq I_{1,0}(z) \leq z/\left\{\tfrac{1}{2}+\left(z^2+\tfrac{1}{4}\right)^{\frac{1}{2}}\right\} .$$

These bounds are quite sharp and have the correct asymptotic behaviour as $z \to 0$ and $z \to \infty$. Using the asymptotic series for $I_\nu(z)$ for large z from Abramowitz and Stegun (1972),

$$I_{1,0}(z) \doteqdot 1 - (2z)^{-1} - \left(8z^2\right)^{-1} - \left(8z^3\right)^{-1} \ldots .$$

For small z, using the definition of $I_\nu(z)$, the following convergent series can be established:

$$I_{1,0}(z) = \tfrac{1}{2}z\left\{1-\tfrac{1}{8}z^2+\tfrac{1}{48}z^4+O\left(z^6\right)\right\} .$$

These approximations will help in the iterative solution of the likelihood equations. The approximation actually used would depend on the most likely range of z.

6. The Asymptotic Dispersion Matrix in the Uniform Case

It is useful to write

$$X_i - \xi = r_i \cos \alpha_i \quad \text{and} \quad Y_i - \eta = r_i \sin \alpha_i .$$

r_i and α_i are unobservable, being functions of the observations and the unknown parameters. r_i/σ has a non-central $\chi_{(2)}$ distribution with non-centrality parameter $\lambda^2 = \rho^2/\sigma^2$, and α_i has a uniform distribution on $[0, 2\pi)$.

The asymptotic dispersion matrix of the maximum likelihood estimators can be derived in the usual way from the second derivatives of the log-likelihood l . To aid this differentiation a preliminary result, obtained using Watson [(1944), p. 79] is given:

$$dI_{a,a-1}(z)/dz = 1 - (2a-1)I_{a,a-1}(z) - I^2_{a,a-1}(z) .$$

Also let $z_i = r_i\rho/\sigma^2$ and $M_i = \left\{1-z_i^{-1}I_{1,0}\left(z_i\right)-I^2_{1,0}\left(z_i\right)\right\}$, then the second derivatives of l are

$$\partial^2 l/\partial\xi^2 = -N\sigma^{-2} + \rho^2 \sum_{i=1}^{N} \{1-2 \cos \alpha_i\}I_{1,0}\left(z_i\right)/z_i + \rho^2\sigma^{-4} \sum_{i=1}^{N} \cos^2\alpha_i\left\{1-I^2_{1,0}\left(z_i\right)\right\} , \quad (4.14)$$

$$\partial^2 l/\partial\xi\partial\eta = \rho^2\sigma^{-4} \sum_{i=1}^{N} \sin \alpha_i \cos \alpha_i\{M_i-I_{1,0}\left(z_i\right)/z_i\} , \quad (4.15)$$

$$\partial^2 l/\partial\xi\partial\rho = -\sigma^{-4} \sum_{i=1}^{N} r_i \cos \alpha_i\{1-I^2_{1,0}\left(z_i\right)\} , \quad (4.16)$$

$$\partial^2 l/\partial\xi\partial\sigma^2 = -\sigma^{-2}(\partial l/\partial\xi) + \rho\sigma^{-4} \sum_{i=1}^{N} z_iM_i \cos \alpha_i , \quad (4.17)$$

$$\partial^2 l/\partial\rho^2 = -N\sigma^{-2} + \rho^{-2} \sum_{i=1}^{N} M_iz_i^2 , \quad (4.18)$$

$$\partial^2 l/\partial\rho\partial\sigma^2 = -\sigma^{-2}(\partial l/\partial\rho) - \rho^{-1}\sigma^{-2} \sum_{i=1}^{N} M_iz_i^2 , \quad (4.19)$$

and

$$\partial^2 l/\partial\left(\sigma^2\right)^2 = N\sigma^{-4} - \sigma^{-6} \sum_{i=1}^{N} \left(r_i^2+\rho^2\right) + 2\sigma^{-4} \sum_{i=1}^{N} z_iI_{1,0}\left(z_i\right) + \sigma^{-4} \sum_{i=1}^{N} z_i^2M_i . \quad (4.20)$$

Similarly $\partial^2 l/\partial\eta^2$, $\partial l/\partial\eta\partial\rho$ and $\partial^2 l/\partial\eta\partial\sigma^2$ can be obtained.

Now

$$f\left(X_i, Y_i\right)dX_idY_i = r_i\sigma^{-2}\exp\left\{-\tfrac{1}{2}\sigma^{-2}\left(r_i^2+\rho^2\right)\right\}I_0\left(r_i\rho/\sigma^2\right)dr_i(2\pi)^{-1}d\alpha_i ,$$

and if we have continuous functions h_1, h_2 and h_3 such that

$$h_1(X_i,\ Y_i) = h_2(r_i)h_3(\alpha_i)\ ,$$

where all these functions have finite expectations, then since r_i and α_i are independent

$$E\{h_1(X_i,\ Y_i)\} = E_r\{h_2(r_i)\}E_\alpha\{h_3(\alpha_i)\}\ .$$

At the true parameter value, θ_0 , $E(\cos\alpha_i) = 0$, so the expectations of (4.15), (4.16) and (4.17) are all zero. So $\hat{\xi}$ and $\hat{\eta}$ are asymptotically uncorrelated with each other and with $\hat{\rho}$ and $\hat{\sigma}^2$. Furthermore

$$E\left(\cos^2\alpha_i\right) = \tfrac{1}{2}\ ,$$

so the expectation of (4.14) can be further reduced giving,

$$E\left(-\partial^2 l/\partial\xi^2\right) = N\sigma^{-2} - \tfrac{1}{2}N\rho^2\sigma^{-4}(1-\delta)\ ,$$

and

$$\mathrm{Var}(\hat{\xi}) \simeq \sigma^2N^{-1}\{1-\tfrac{1}{2}\rho^2\sigma^{-2}(1-\delta)\}^{-1}\ ,$$

where

$$\delta = E\left\{I^2_{1,0}\left(r_i\rho/\sigma^2\right)\right\}\ .$$

δ is given exactly by the integral

$$\delta = \sigma^{-2}e^{-\rho^2/2\sigma^2}\int_0^\infty re^{-r^2/2\sigma^2}I_{1,0}\left(r\rho/\sigma^2\right)I_1\left(r\rho/\sigma^2\right)dr\ .$$

Since $\hat{\xi}$ and $\hat{\eta}$ are uncorrelated with $\hat{\rho}$ and $\hat{\sigma}^2$ the asymptotic dispersion matrix of the latter two estimators is the same as if ξ and η were known. If ξ and η were known then we would be discussing the asymptotic dispersion matrix of the maximum likelihood estimates of the noncentrality and scale parameters of a non-central chi distribution with known degrees of freedom m , where here $m = 2$. Thus some of the results obtained here are applicable to this noncentral chi distribution.

To evaluate (4.18), (4.19) and (4.20) we need

$$E\left(r_i^2\right) = \rho^2 + 2\sigma^2\ ,$$

and

$$E\left\{r_i I_{1,0}\left(r_i \rho/\sigma^2\right)\right\} = \sigma^{-2} e^{-\rho^2/2\sigma^2} \int_0^\infty r^2 e^{-r^2/2\sigma^2} I_1\left(r\rho/\sigma^2\right) dr$$

$$= \rho ,$$

from (4.12) since at the true parameter value θ_0 the expected values of the first derivatives of the log-likelihood are zero. Also we require

$$\beta = E\left\{r_i^2 I_{1,0}^2\left(r_i \rho/\sigma^2\right)\right\}$$

$$= \sigma^{-2} e^{-\rho^2/2\sigma^2} \int_0^\infty r^3 e^{-r^2/2\sigma^2} I_{1,0}\left(r\rho/\sigma^2\right) I_1\left(r\rho/\sigma^2\right) dr .$$

Integrals of this type do not seem to have appeared in the literature. However since the integral is essentially a function of a single parameter it could be tabulated. Let $z = r\rho/\sigma^2$ and $A = \sigma^2/\rho^2 = \lambda^{-2}$; then

$$\beta = \rho^2 A^3 \exp\left(-\tfrac{1}{2}A^{-1}\right) \int_0^\infty z^3 \exp\left(-\tfrac{1}{2}Az^2\right) I_{1,0}(z) I_1(z) dz . \tag{4.21}$$

Similarly δ can be written as

$$\delta = A \exp\left(-\tfrac{1}{2}A^{-1}\right) \int_0^\infty z \exp\left(-\tfrac{1}{2}Az^2\right) I_{1,0}(z) I_1(z) dz . \tag{4.22}$$

The variances and covariances of $\hat{\rho}$ and $\hat{\sigma}^2$ can thus be written

$$\mathrm{Var}(\hat{\rho}) = \sigma^2 N^{-1} \Delta^{-1}\{A-1+A^{-1}(\beta/\rho^2 - 1)\} ,$$

$$\mathrm{Cov}\left(\hat{\rho}, \hat{\sigma}^2\right) = \rho\sigma^2 N^{-1} \Delta^{-1}\{(\beta/\rho^2 - 1)-A\} ,$$

and

$$\mathrm{Var}\left(\hat{\sigma}^2\right) = \sigma^4 N^{-1} \Delta^{-1}\{(\beta/\rho^2 - 1)\} ,$$

where

$$\Delta = \left(\beta/\rho^2 - 1\right)\left(1+A^{-1}\right) - 1 .$$

Both δ and β have to be obtained by numerical integration and this will be discussed next.

7. Numerical Evaluation of δ and β

Before the numerical integration of expressions (4.21) and (4.22) we obtain some rough preliminary bounds for δ and β. Actually these bounds are quite sharp. They are surprisingly accurate for large ρ^2/σ^2 and will be used in the discussion of efficiency of other estimators of ρ and ξ.

Since $I_{1,0}(r\rho/\sigma^2)$ is monotonically increasing between zero and unity,

$$0 \le I_{1,0}^2(r\rho/\sigma^2) \le I_{1,0}(r\rho/\sigma^2) \le 1 .$$

Using this and the inequality $E(z^2) > (E(z))^2$, we obtain,

$$\left[E\left\{I_{1,0}(r\rho/\sigma^2)\right\}\right]^2 \le \delta \le E\left\{I_{1,0}(r\rho/\sigma^2)\right\} ,$$

where

$$E\left\{I_{1,0}(r\rho/\sigma^2)\right\} = \sigma^{-2}e^{-\rho^2/2\sigma^2} \int_0^\infty re^{-r^2/2\sigma^2} I_1(r\rho/\sigma^2)dr . \qquad (4.23)$$

Expressions like (4.23) can be evaluated by expanding the modified Bessel function as a power series and integrating term-by-term. Then using the definition of the confluent hypergeometric function,

$${}_1F_1[a;\ b;\ z] = \{\Gamma(b)/\Gamma(a)\} \sum_{n=0}^{\infty} \{\Gamma(a+n)/\Gamma(b+n)\}z^n/n! ,$$

from Slater [(1960), eq. 1.1.8] and Kummer's first theorem (Slater [(1960), eq. 1.4.1]),

$${}_1F_1[a;\ b;\ z] = e^{-z}\,{}_1F_1[b-a;\ b;\ -z] ,$$

we obtain

$$E\left\{I_{1,0}(r\rho/\sigma^2)\right\} = (\rho^2/2\sigma^2)^{\frac{1}{2}}\Gamma(3/2)\,{}_1F_1[\tfrac{1}{2};\ 2;\ -\tfrac{1}{2}\rho^2/\sigma^2] .$$

Tables of the confluent hypergeometric function are available (see for example Slater (1960)). For $\gamma = \frac{1}{2}\rho^2/\sigma^2$ large, using Slater [(1960), eq. 4.1.2], we derive an asymptotic series for ${}_1F_1[\frac{1}{2};\ 2;\ -\gamma]$ to give

$$1 - \tfrac{1}{2}\gamma^{-1} - \gamma^{-2}/8 - 3\gamma^{-3}/16 + \ldots \le \delta$$
$$\le 1 - \gamma^{-1}/4 - 3\gamma^{-2}/32 - 15\gamma^{-3}/128 . \qquad (4.24a)$$

Similarly the following bounds for β/ρ^2 can be obtained:

$$1 \le \beta/\rho^2 \le 1 + 3\gamma^{-1}/4 - 3\gamma^{-2}/32 - \ldots \quad (4.25a)$$

However neither of these bounds is sufficiently accurate to be used as an approximation to β .

These inequalities can be improved upon by a more complicated analysis of the integrals (see Appendix A), but even then the improvement is slight. We obtain

$$1 - \tfrac{1}{2}\gamma^{-1} - \gamma^{-2}/8 - 3\gamma^{-3}/16 \ldots \le \delta \le 1 - 3\gamma^{-1}/8 - 11\gamma^{-2}/128 + \ldots \quad (4.24b)$$

and

$$1 + \gamma^{-1}/8 + 93\gamma^{-2}/128 + \ldots \le \beta/\rho^2 \le 1 + 5\gamma^{-1}/8 - 27\gamma^{-2}/128 + \ldots \quad (4.25b)$$

Actually a more accurate lower bound for β can be obtained by noting that Δ must be positive and so

$$\beta/\rho^2 > 1 + [1 + \rho^2/\sigma^2]^{-1} ,$$

and for γ large this is asymptotically

$$\beta/\rho^2 > 1 + \tfrac{1}{2}\gamma^{-1} - \gamma^{-2}/4 + \gamma^{-3}/8 \ldots \quad (4.26)$$

To evaluate (4.21) and (4.22) the Gregory formula for numerical integration was used. This replaces the integral by the sum

$$S = h\left\{\tfrac{1}{2}f(0) + \sum_{k=1}^{\infty} f(kh)\right\} + \text{correction term} ,$$

where $f(z)$ is the integrand, h is the interval width and the correction term involves differences between ordinates near the origin. This form of Gregory's formula was used because it does not require ordinates beyond the limits of integration, in particular it does not involve values of the modified Bessel function for negative arguments, and also only requires differences between values of the integrand at neighbouring points, not derivatives as in the Euler-Maclaurin formula for example.

The values of A and h considered were $A = 0.01(0.01)0.1(0.1)1.0$ and $h = 0.05, 0.1, 0.2, 0.5$ and 1.0 . The infinite upper bound was replaced by Kh where $f(z)$ is sufficiently small for $z > Kh$. The value of Kh varied considerably with A . For example Kh varied from less than 8 for $A = 1.0$ to around 180 for $A = 0.01$. The actual numerical integration was performed by using the values of $I_0(z)$ and $I_1(z)$ from the British Association Mathematical Tables, volume X (1952) for $z \leq 20$ and for $z > 20$ by using the first six terms in the asymptotic expansion for $I_\nu(z)$ given by Abramowitz and Stegun [(1972), eq. 9.7.1]. Additional terms in this expansion only produced a small order difference in the calculations. This accuracy was not needed for δ but for A small, β/ρ^2 was required to a fairly high degree of accuracy since it did come close to the lower bound given by (4.26).

Six significant figures were decided upon for β/ρ^2 since it was impractical to get a higher accuracy, even though this probably left minor errors in the third significant figures of the entries of the asymptotic dispersion matrix. The value of h used was 0.1 . δ was calculated simultaneously with β/ρ^2 and the same values of h were used even though δ was not needed to the same degree of accuracy, and for small A larger values of h sufficed. For $A = 1.0$, 0.9 and 0.8 , the calculations for δ and β/ρ^2 were checked using $h = 0.05$ and the values of $I_0(z)$ and $I_1(z)$ from the British Association Mathematical Tables, volume VI (1950). These values of A had the smallest effective range of integration and required smaller interval widths to obtain the same accuracy as the results for small A . The values with $h = 0.05$ agreed with those for $h = 0.1$ to the required accuracy. Also for h small, the correction term was of a lower order of magnitude than the accuracy required, and was ignored. For values of δ and corresponding values of $\mathrm{Var}(\hat{\xi})$ see Table V and for values of β/ρ^2 refer to Table VI.

$A = \lambda^{-2}$	δ	$N\ \mathrm{Var}(\hat{\xi})/\sigma^2$	A	δ	$N\ \mathrm{Var}(\hat{\xi})/\sigma^2$
1.0	0.34515	1.487	0.09	0.90447	2.131
0.9	0.37130	1.537	0.08	0.91577	2.112
0.8	0.40178	1.597	0.07	0.92689	2.093
0.7	0.43772	1.671	0.06	0.93781	2.076
0.6	0.48070	1.763	0.05	0.94854	2.060
0.5	0.53290	1.877	0.04	0.95910	2.046
0.4	0.59731	2.014	0.03	0.96951	2.033
0.3	0.67780	2.160	0.02	0.97979	2.021
0.2	0.77788	2.249	0.01	0.98995^-	2.010
0.1	0.89302	2.150			

TABLE V

A	β/ρ^2	$N\ \mathrm{Var}(\hat{\rho})/\sigma^2$	$N\ \mathrm{Cov}(\hat{\rho},\ \hat{\sigma}^2)/\rho\sigma^2$	$N\ \mathrm{Var}(\hat{\sigma}^2)/\sigma^4$
1.0	1.52145	12.16	-11.16	12.16
0.9	1.49569	9.70	-8.70	10.67
0.8	1.46683	7.61	-6.61	9.27
0.7	1.43424	5.87	-4.87	7.96
0.6	1.39712	4.44	-3.44	6.73
0.5	1.35442	3.30	-2.30	5.60
0.4	1.30476	2.43	-1.43	4.57
0.3	1.24634	1.80	-0.795	3.65
0.2	1.17695	1.37	-0.374	2.87
0.1	1.09467	1.13	-0.129	2.29
0.09	1.08572	1.11	-0.112	2.25
0.08	1.07664	1.10	-9.70×10^{-2}	2.21
0.07	1.06745	1.08	-8.22×10^{-2}	2.17
0.06	1.05814	1.07	-6.85×10^{-2}	2.14
0.05	1.04871	1.06	-5.63×10^{-2}	2.13
0.04	1.03918	1.04	-4.36×10^{-2}	2.10
0.03	1.02953	1.03	-3.23×10^{-2}	2.08
0.02	1.01980	1.02	-2.04×10^{-2}	2.02
0.01	1.00995	1.01	-1.01×10^{-2}	2.01

TABLE VI

Figure 4.1 gives β/ρ^2 and also $1 + [1 + \rho^2/\sigma^2]^{-1}$ and $1 + \frac{5}{8}\gamma^{-1} - \frac{27}{128}\gamma^{-2}$, as functions of A.

It is the variance of $\hat{\xi}$ for constant σ^2 and increasing ρ that is tabulated. Since σ^2 is constant and ρ increasing, one might expect $N\ \mathrm{Var}(\hat{\xi})/\sigma^2$ to increase also. In Table V, $N\ \mathrm{Var}(\hat{\xi})/\sigma^2$ increases until A is around 0.2, then decreases. This apparent anomaly can be explained as follows. For A sufficiently small it is easily shown that δ is close to the lower bound given by (4.24b), and this implies that for A small

$$N\ \mathrm{Var}(\hat{\xi})/\sigma^2 \simeq 2 + A + \tfrac{1}{2}A^2 + \dots .$$

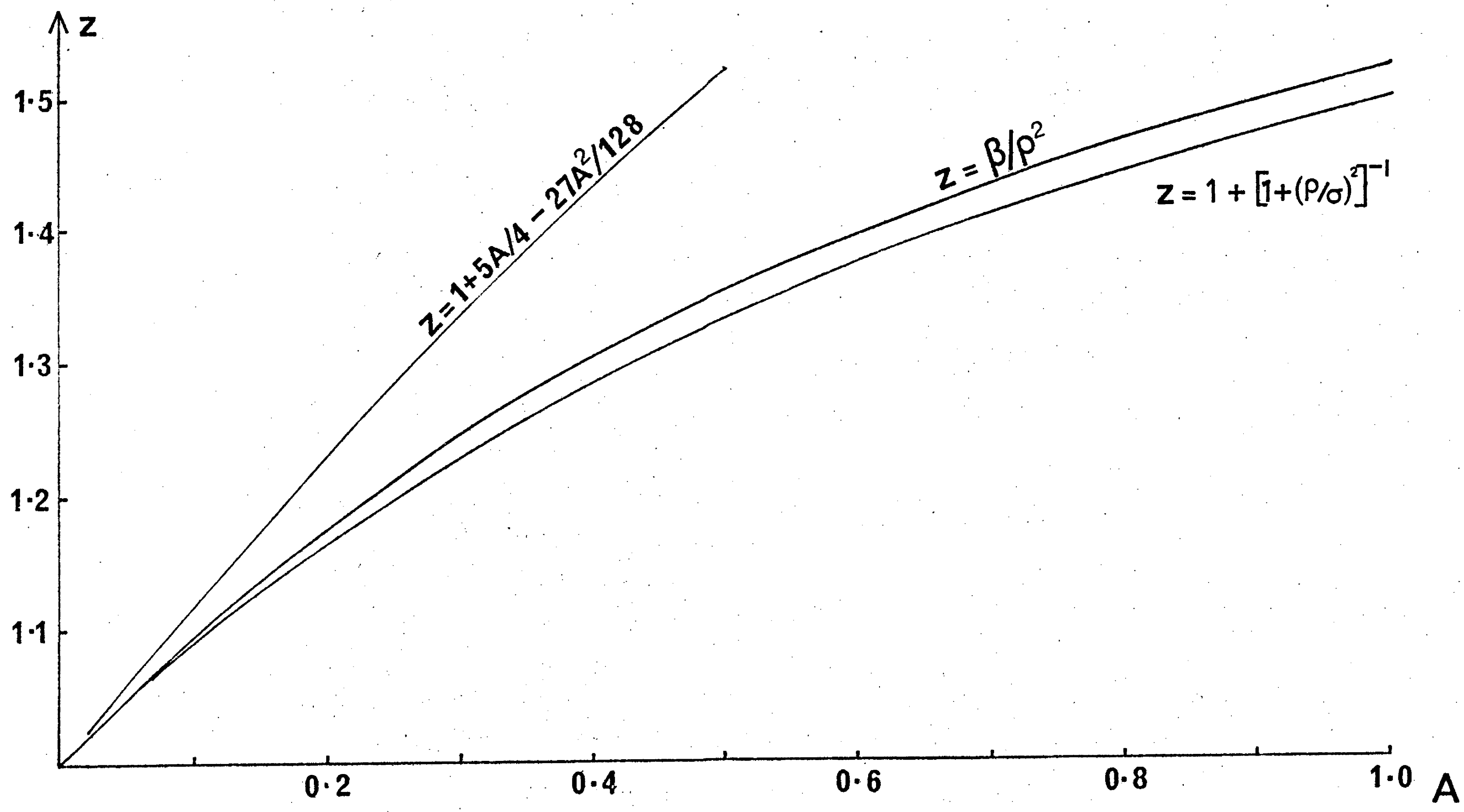

z
1·5
1·4
1·3
1·2
1·1
0·2
0·4
0·6
0·8
1·0
A
$z=1+5A/4-27A^2/128$
$z=\beta/\rho^2$
$z=1+[1+(\rho/\sigma)^2]^{-1}$

Fig. 4.1

For small A , $N\,\mathrm{Var}(\hat{\xi})/\sigma^2$ approaches 2 from above as A decreases. However for larger A the appropriate bound is

$$\pi\left(\rho^2/8\sigma^2\right)\left\{{}_1F_1[\tfrac{1}{2};\ 2;\ -\gamma]\right\}^2 \leq \delta\ .$$

If we use the recurrence relationships and tables of the confluent hypergeometric function from Slater (1960) to evaluate the above lower bound for δ we can show that since

$$N\,\mathrm{Var}(\hat{\xi})/\sigma^2 = \left[1 - \tfrac{1}{2}\rho^2/\sigma^2(1-\delta)\right]^{-1} ,$$

an upper bound for $N\,\mathrm{Var}(\hat{\xi})/\sigma^2$ when $A = 1.0$, is 1.5 . This confirms that the trend apparent in Table V does occur.

Notice that β/ρ^2 is very close to the lower bound given by (4.26) for γ large. The actual determinant of the dispersion matrix of $\hat{\rho}$ and $\hat{\sigma}^2$ is proportional to $\rho^2\Delta/\sigma^2$; hence for γ large the determinant is not extremely small. Unfortunately the lower bound given by (4.26) is strictly a lower bound and can not be used to examine the asymptotic behaviour of $\mathrm{Var}(\hat{\rho})$ since it makes Δ equal to zero.

Also for A near 1 the standard deviation of $\hat{\rho}$ is reasonably large for moderate N , and we would require a larger value of N than for A small before we could consider that the asymptotic normal distribution for $\hat{\rho}$ is valid.

8. Efficiency of $\overline{X}$

$(\overline{X}, \overline{Y})$ is a consistent and unbiased estimator of (ξ, η) . It is interesting to compare the efficiency of $\overline{X}$ as an estimator of ξ with that of $\hat{\xi}$. To examine this we need

$$E(\overline{X}) = \xi\ ,$$

and

$$\mathrm{Var}(X) = \sigma^2 + \rho^2\mathrm{Var}(\cos\tau) = \sigma^2 + \tfrac{1}{2}\rho^2\ ,$$

and therefore

$$\mathrm{Var}(\overline{X}) = \sigma^2 N^{-1}\left(1 + \tfrac{1}{2}\rho^2/\sigma^2\right)\ .$$

The distribution of $\hat{\xi}$ must be symmetric around ξ due to the symmetric estimation procedure, hence $\hat{\xi}$ must be unbiased. To compare $\overline{X}$ and $\hat{\xi}$ we use the notion of asymptotic relative efficiency in the Pitman sense as in Chapter 2. Using this we find that

$$\mathrm{ARE}(\overline{X}, \hat{\xi}) = [\{1 - \tfrac{1}{2}\rho^2/\sigma^2(1-\delta)\}\{1 + \tfrac{1}{2}\rho^2/\sigma^2\}]^{-1} ,$$

and this is easily evaluated using the calculated values of δ. The $\mathrm{ARE}(\overline{X}, \hat{\xi})$ decreases slowly from 0.991 at $A = 1.0$ to 0.810 at $A = 0.3$, then more rapidly to 0.358 at $A = 0.1$ and down to 0.039 at $A = 0.01$.

For ρ^2/σ^2 reasonably large we have already found an asymptotic expression for $\mathrm{Var}(\hat{\xi})$, and using this we obtain

$$\begin{aligned}\mathrm{ARE}(\overline{X}, \hat{\xi}) &\simeq 2\gamma^{-1} - \tfrac{3}{2}\gamma^{-2} \ldots \\ &\simeq 4A - 6A^2 \ldots \quad \text{for small } A .\end{aligned}$$

As A decreases so does $\mathrm{ARE}(\overline{X}, \hat{\xi})$. This is easily explained. In estimating $\hat{\xi}$ allowance is made for (X, Y) being distributed around the circumference of a circle, whereas no such allowance is made in determining ξ using $\overline{X}$. For large ρ^2/σ^2 the moment estimators are inefficient although consistent and unbiased, and

$$\lim_{A\to 0} \mathrm{ARE}(\overline{X}, \hat{\xi}) = 0 .$$

On the other hand, the circle is poorly defined for large A and the asymptotic efficiency of $\overline{X}$ relative to $\hat{\xi}$ is high.

9. Asymptotic Behaviour of $\hat{\rho}$ and $\hat{\sigma}^2$

It is interesting to examine the asymptotic behaviour of $\hat{\rho}$ and $\hat{\sigma}^2$ as γ tends to infinity, but to do this an accurate approximation for β/ρ^2 is needed. This is possible because β/ρ^2 is in the Cramér-Rao lower bound for the variance of $\hat{\rho}$ where ξ, η and σ^2 are known, that is

$$\mathrm{Var}(\hat{\rho}) \simeq \sigma^2 N^{-1}[\beta/\rho^2 - 1]^{-1}\sigma^2/\rho^2 ,$$

and a very efficient estimator of ρ exists in this case, namely $\overline{r}$. Assuming ξ, η and σ^2 to be known, then

$$E(r_i) = \sigma^{-2}\exp(-\tfrac{1}{2}\rho^2/\sigma^2) \int_0^\infty r^2\exp(-\tfrac{1}{2}r^2/\sigma^2) I_0(r\rho/\sigma^2)\,dr ,$$

and this can be evaluated in a similar manner to expression (4.23) to give

$$E(r_i) = (2\sigma^2)^{\frac{1}{2}}\Gamma(3/2)\,{}_1F_1[-\tfrac{1}{2};\ 1;\ -\gamma]\ .$$

Let $\bar{r} = N^{-1}\sum_{i=1}^{N} r_i$; then $\bar{r}$ is a biased estimator of ρ . In fact since $I_0(r\rho/\sigma^2) > I_1(r\rho/\sigma^2)$ for all r , it is easily seen that

$$E(\bar{r}) = E(r_i) > \rho$$

independently of N . However we can use $\bar{r}$ to find an accurate lower bound for β/ρ^2 , since

$$\mathrm{ARE}(\bar{r},\ \hat{\rho}) \le 1\ .$$

$E(\bar{r})$ and $\mathrm{Var}(\bar{r})$ can be evaluated for certain γ using the recurrence relations for ${}_1F_1[a;\ b;\ z]$ and the tables from Slater (1960). Also $\partial E(\bar{r})/\partial\rho$ can be calculated in a similar manner after using Slater [(1960), eq. 2.1.1],

$$d\left({}_1F_1[a;\ b;\ z]\right)/dz = a\ {}_1F_1[a+1;\ b+1;\ z]/b\ ,$$

to differentiate $E(\bar{r})$ with respect to ρ . Also $\hat{\rho}$ is an unbiased estimator of ρ . Even though $\bar{r}$ is wildly biased for large A , for example $E(\bar{r}) = 1.549\rho$ for $A = 1.0$, the $\mathrm{ARE}(\bar{r},\ \hat{\rho})$ is above 0.986 for all A less than 1.0 . Thus we can evaluate a lower bound for β/ρ^2 which is accurate for all A .

For γ large, using Slater [(1960), eq. 4.1.2], asymptotic expressions for $E(\bar{r})$ and $\partial E(\bar{r})/\partial\rho$ can be found and thus an asymptotic expression for β/ρ^2 established,

$$\beta/\rho^2 \ge 1 + \tfrac{1}{2}\gamma^{-1} - \gamma^{-2}/8 - \gamma^{-3}/32\ \ldots\ . \tag{4.27}$$

This bound for β/ρ^2 is only about

$$\gamma^{-2}/8 - 5\gamma^{-3}/32$$

above that given by (4.26) and is actually quite accurate for small A .

Returning to the case under consideration where ξ, η and σ^2 are not known, using (4.27) for large γ ,

$$\Delta \simeq \gamma^{-1}/4 - 3\gamma^{-2}/16 ,$$

$$N \operatorname{Var}(\hat{\rho})/\sigma^2 \simeq 1 + \tfrac{1}{2}\gamma^{-1} ,$$

$$N \operatorname{Var}\left(\hat{\sigma}^2\right)/\sigma^4 \simeq 2 + \gamma^{-1} ,$$

and

$$N \operatorname{Corr}\left(\hat{\rho}, \hat{\sigma}^2\right) \simeq -\tfrac{1}{2}\gamma^{-1} .$$

Thus as ρ^2/σ^2 increases, $\operatorname{Var}(\hat{\rho})$ tends to $\sigma^2 N^{-1}$, $\operatorname{Var}\left(\hat{\sigma}^2\right)$ tends to $2\sigma^4 N^{-1}$ and $\operatorname{Corr}\left(\hat{\rho}, \hat{\sigma}^2\right)$ tends to zero, for fixed N. In fact as ρ/σ tends to infinity, $I_{1,0}\left(r\rho/\sigma^2\right)$ tends to unity for all non-zero r_i so that

$$\hat{\rho}/\bar{r} \to 1$$

as expected.

For large ρ^2/σ^2 when r is non-zero

$$I_{1,0}\left(r\rho/\sigma^2\right) \simeq \left(1 - \sigma^2/2\rho r - O\left(\rho^{-2}/\sigma\right)\right) ,$$

so

$$\hat{\sigma}^2 \to \tfrac{1}{2}N^{-1}\left\{\sum_{i=1}^{N} \left[r_i^2 - \bar{r}^2 + \sigma^2\bar{r}/\rho \ \ldots\right]\right\} ;$$

thus

$$\hat{\sigma}^2 \to N^{-1} \sum_{i=1}^{N} \left(r_i - \bar{r}\right)^2 . \tag{4.28}$$

Notice the similarity between these estimators and their asymptotic variances, and the estimators of the mean and variance in samples from a normal distribution with mean ρ and variance σ^2.

If $\mu = E(r)$ and $\nu = \operatorname{Var}(r)$, then from Jensen (1969), $(r-\mu)/\nu$ does tend to a standard normal distribution as γ tends to infinity.

10. Other Estimators of ρ

If (ξ, η) is known then r_i/σ has a noncentral chi distribution with two degrees of freedom and non-centrality parameter ρ^2/σ^2. If in addition σ^2 is known, ρ can be estimated using the second sample moment of r, and when σ^2 is unknown using the second and fourth sample moment of r. These estimators are

$$\rho_1 = \left\{N^{-1} \sum_{i=1}^{N} r_i^2 - 2\sigma^2\right\}^{\frac{1}{2}} = v^{\frac{1}{2}} ,$$

and

$$\rho_2 = \left\{2\left(N^{-1} \sum_{i=1}^{N} r_i^2\right)^2 - N^{-1} \sum_{i=1}^{N} r_i^4\right\}^{\frac{1}{4}} = u^{\frac{1}{4}} .$$

To examine their asymptotic relative efficiencies with respect to $\hat{\rho}$ we need the moments of ρ_1 and ρ_2. The easiest way to find these moments is via the moments of u and v.

To find the "mean" and "variance" of a function g of a random variable z, we assume that $g(z)$ can be expanded in a Taylor series about $E(z) = \mu$,

$$g(z) = g(\mu) + (z-\mu)g'(\mu) + \tfrac{1}{2}(z-\mu)^2 g''(\mu) + \ldots ,$$

then taking expectations of both sides

$$E\{g(z)\} \simeq g(\mu) + \tfrac{1}{2}g''(\mu)\mathrm{Var}(z) + \ldots$$

and

$$\mathrm{Var}\{g(z)\} \simeq [g'(\mu)]^2\mathrm{Var}(z) + \ldots .$$

In general these approximations are only rough because they ignore terms involving higher order central moments of z. However here the higher order central moments are of small order, and so the approximations are fairly accurate.

First consider ρ_1 to order N^{-1},

$$E(v) = \rho^2 ,$$

and

$$\mathrm{Var}(v) = 4\sigma^2 N^{-1}\left(\rho^2+\sigma^2\right) .$$

Thus

$$E(\rho_1) \simeq \rho + O(N^{-1}) \; ,$$

and

$$\mathrm{Var}(\rho_1) \simeq \sigma^2 N^{-1}(1 + \sigma^2/\rho^2) \; .$$

Since σ^2 is assumed known, $\hat{\rho}$ is unbiased and

$$\mathrm{Var}(\hat{\rho}) = \sigma^2 N^{-1}(\sigma^2/\rho^2)(\beta/\rho^2 - 1)^{-1} \; ,$$

then

$$\mathrm{ARE}(\rho_1, \hat{\rho}) \simeq (\sigma^2/\rho^2)(\beta/\rho^2 - 1)^{-1}(1 + \sigma^2/\rho^2)^{-1} \; .$$

Additionally for large γ using (4.27),

$$\mathrm{ARE}(\rho_1, \hat{\rho}) \simeq 1 - \gamma^{-1}/4 \; .$$

To examine ρ_2 we need

$$E(u) = \rho^4 + 8N^{-1}\sigma^2\rho^2 + 8N^{-1}\sigma^4 \; ,$$

and we find $\{E(u)\}^{\frac{1}{4}}/\rho$ by the binomial series expansion, if N is large, and thus show

$$E(\rho_2) \simeq \rho + O(N^{-1})$$

and

$$\mathrm{Var}(\rho_2) = \sigma^2 N^{-1}(1 + \sigma^2/\rho^2 + 8\sigma^4/\rho^4 + 4\sigma^6/\rho^6) + O(N^{-2}) \; .$$

Here of course σ^2 is unknown and for large γ,

$$\mathrm{Var}(\hat{\rho}) \simeq \sigma^2 N^{-1}(1+\tfrac{1}{2}\gamma^{-1}) \; ;$$

so

$$\mathrm{ARE}(\rho_2, \hat{\rho}) \simeq 1 - 3\gamma^{-2}/2 \; .$$

If (ξ, η) is not known then the estimation is more complicated. If we have estimates (ξ^*, η^*) of (ξ, η) then in ρ_1 and ρ_2, r_i can be replaced by r_i^* where

$$r_i^* = \left\{\left(X_i-\xi^*\right)^2+\left(Y_i-\eta^*\right)^2\right\}^{\frac{1}{2}} .$$

Since ξ^* and η^* are functions of all the X_i and Y_i, the r_i^* will be identically, but not independently distributed, and the distribution of r_i^* is unlikely to have a nice form. If we do use estimators of the same form as ρ_1 and ρ_2 then the easiest estimators of ξ^* and η^* to use are $\overline{X}$ and $\overline{Y}$. Since these are inefficient we would not necessarily expect the resultant estimators of ρ to be efficient. In fact

$$E\left(\rho_1^*\right) \simeq \rho + O\left(N^{-1}\right)$$

and

$$\text{Var}\left(\rho_1^*\right) \simeq \sigma^2 N^{-1}\left(1 + \sigma^2/\rho^2\right)$$

so that the asymptotic relative efficiency of ρ_1^* with respect to $\hat{\rho}$ is the same as that of ρ_1. However if we carry out the same procedure for ρ_2^* we find that

$$E\left(\rho_2^*\right) \simeq \rho + O\left(N^{-1}\right) ,$$

so that ρ_2^* is asymptotically unbiased but

$$\text{Var}\left(\rho_2^*\right) \simeq \tfrac{1}{2}N^{-1}\{\rho^2 + \sigma^2\left(6 + 6\sigma^2/\rho^2 + 16\sigma^4/\rho^4 + 8\sigma^6/\rho^6\right)\} ;$$

thus for γ large

$$\text{ARE}\left(\rho_2^*, \hat{\rho}\right) \simeq 2\sigma^2/\rho^2 .$$

If we are only interested in estimation of ρ and σ^2 then perhaps it is better to consider statistics which do not involve estimation of ξ and η. For example consider d_{ij}^2, the square of the distance between two points $\left(x_i, x_i\right)$ and $\left(x_j, x_j\right)$, where

$$d_{ij}^2 = \left\{\left(x_i-x_j\right)^2+\left(y_i-y_j\right)^2\right\} = \{\left(x_i-\xi\right)-\left(x_j-\xi\right)\}^2 + \{\left(y_i-\eta\right)-\left(y_j-\eta\right)\}^2 .$$

Let

$$D = N^{-1}(N-1)^{-1} \sum_{i<j}\sum d_{ij}^2 ,$$

and

$$S = N^{-1}(N-1)^{-1} \sum_{i<j}\sum d_{ij}^4 ,$$

then

$$E(D) = \rho^2 + 2\sigma^2$$

and

$$E(S) = 3\rho^4 + 16\rho^2\sigma^2 + 16\sigma^4 .$$

Therefore we can find consistent estimators of ρ and σ^2 using D and S namely,

$$\rho_4 = \left(4D^2-S\right)^{\frac{1}{4}} = W^{\frac{1}{4}}$$

and

$$\sigma_4^2 = \tfrac{1}{2}\left(D-\rho_4^2\right) .$$

Note that these estimators are not unbiased. Also let

$$\rho_3 = \left(D-2\sigma^2\right)^{\frac{1}{2}} ,$$

be the estimator if σ^2 is known. Then

$$E\left(\rho_3\right) \simeq \rho + O\left(N^{-1}\right) ,$$

and

$$\mathrm{Var}\left(\rho_3\right) \simeq \sigma^2 N^{-1}\left(1 + \sigma^2/\rho^2\right) .$$

and thus this estimator is just as efficient as ρ_1 and ρ_1^*. Also

$$E\left(\rho_4\right) = \rho + O\left(N^{-1}\right) ,$$

and

$$\mathrm{Var}\left(\rho_4\right) \simeq \sigma^2 N^{-1}\{1 + \sigma^2/\rho^2 + 8\sigma^4/\rho^4 + 4\sigma^6/\rho^6\} ,$$

and therefore is just as efficient as ρ_2, and much more efficient than ρ_2^*. Since ρ_3 and ρ_4 are also easier to evaluate when (ξ, η) is unknown, they should be used in that case. One disadvantage with the form of the estimator given for σ_4^2 is that it could possibly be negative and the modified estimator $\max\left\{\sigma_4^2,\ 0\right\}$ should be used.

11. Use of an Approximation to $I_{1,0}\left(r\rho/\sigma^2\right)$

In the numerical solution of the maximum likelihood equations, $I_{1,0}\left(r\rho/\sigma^2\right)$ may be replaced by an approximation. In many of the cases considered $I_{1,0}\left(r\rho/\sigma^2\right)$ will be close to unity for most r, and it is interesting to examine the effect of substituting unity for $I_{1,0}\left(r\rho/\sigma^2\right)$ in $\partial l/\partial\xi$, $\partial l/\partial\eta$, $\partial l/\partial\rho$ and $\partial l/\partial\sigma^2$. These modified derivative equations yield

$$\sum_{i=1}^{N}\left(X_i-\xi\right) - \rho\sum_{i=1}^{N}\left(X_i-\xi\right)/r_i = 0 ,$$

$$\sum_{i=1}^{N}\left(Y_i-\eta\right) - \rho\sum_{i=1}^{N}\left(Y_i-\eta\right)/r_i = 0 ,$$

$$\rho = \bar{r} ,$$

and

$$\sigma^2 = (2N)^{-1}\sum_{i=1}^{N}\left(r_i-\bar{r}\right)^2 .$$

Let $\theta^* = \left(\xi^*,\ \eta^*,\ \rho^*,\ \sigma^{*2}\right)$ be the solutions of the above equations. Notice that if σ^{*2} is replaced by $2\sigma^{*2}$, which by (4.28) is a far better estimator, the above is akin to assuming that r_i given ξ and η has a normal distribution with mean ρ and variance σ^2. As has been already shown $E(\bar{r}) > \rho$, where the bias is independent of N. Also as N goes to infinity, $\mathrm{Var}(\bar{r})$ goes to zero, so ρ^* can not be a consistent estimator of ρ. However if γ is large enough the bias in $E(\bar{r})$ is small.

Difficulties occur in finding the variances of these starred estimators. If we have an estimating equation of the form

$$\sum_{i=1}^{N} g\left(X_i,\ Y_i,\ \mu^*\right) = 0 ,$$

where

$$E[g(X_i, Y_i, \mu)] = 0$$

and μ is the unknown parameter, one method of finding the variance of μ^* is to use a Taylor series expansion of $g(X_i, Y_i, \mu^*)$ around the point $\mu^* = \mu$ and proceed in a similar manner to finding the variance of the maximum likelihood estimator. This is essentially what was done by Neyman and Scott (1948) in a more general context. For this type of procedure we need the expected value of $|\partial^2 g/\partial\mu^2|$ to be finite. For our estimating equations we have for example

$$\partial^2 g/\partial\xi^2 = 3\rho\sigma^{-2} \sum_{i=1}^{N} r_i^{-2}\left[(X_i-\xi)/r_i - (X_i-\xi)^3/r_i^3\right] .$$

$(X_i-\xi)/r_i$ is $\cos \alpha_i$, so this can be written

$$\partial^2 g/\partial\xi^2 = 3\rho\sigma^{-2} \sum_{i=1}^{N} r_i^{-2}\left(\cos \alpha_i - \cos^3\alpha_i\right) .$$

From Feller [(1971), lemma 3, p. 151], if z_i , $i = 1, \ldots, N$, are independently distributed then $E\left|\sum_{i=1}^{N} z_i\right|$ exists if and only if $E|z_i|$ exists for all $i = 1, 2, \ldots, N$. Consequently

$$E|\partial^2 g/\partial\xi^2|$$

exists, if and only if,

$$E\left|r_i^{-2}\left\{\cos \alpha_i - \cos^3\alpha_i\right\}\right|$$

exists. Since α_i and r_i are independently distributed here, this is just

$$E\left(r_i^{-2}\right)E\left|\cos \alpha_i - \cos^3\alpha_i\right| .$$

The second term $E\left|\cos \alpha_i - \cos^3\alpha_i\right|$ is positive. The first term is given by

$$\lim_{c\to 0^+} \left\{\sigma^{-2}e^{-\rho^2/2\sigma^2} \int_c^\infty r^{-1}e^{-r^2/2\sigma^2} I_0\left(r\rho/\sigma^2\right)dr\right\} ,$$

and this is not finite. This is quite easy to see. $I_0(r\rho/\sigma^2)$ is always greater than one so

$$\int_c^\infty r^{-1}e^{-r^2/2\sigma^2} I_0(r\rho/\sigma^2)\,dr \geq \int_c^\infty r^{-1}e^{-r^2/2\sigma^2}\,dr \,.$$

Without loss of generality assume $2\sigma^2$ equals one in the right hand integral and make a change of variable to $z = r^2$. Then

$$\int_c^\infty r^{-1}e^{-r^2}\,dr = \tfrac{1}{2}\int_{c^{\frac{1}{2}}}^\infty z^{-1}e^{-z}\,dz$$

$$= \tfrac{1}{2}E_1\left(c^{\frac{1}{2}}\right) ,$$

by the definition of the exponential integral. The limit as c goes to zero of $E_1\left(c^{\frac{1}{2}}\right)$ is not bounded. Thus $E|\partial^2 g/\partial\xi^2|$ is not bounded.

The same thing occurs in the equation estimating ξ in the allied functional relationship considered by Chan (1965), namely,

$$\sum_{i=1}^{N} \left(X_i - \xi\right)\left(1 - \bar{r}/r_i\right) = 0 \,.$$

If we call the left hand side of this equation $h(\xi)$ it can be shown that $\partial^2 h/\partial\xi^2$ includes $\partial^2 g/\partial\xi^2$ as a factor. Chan based his analysis on Neyman and Scott's paper, and we have just shown that one of the conditions of Neyman and Scott do not hold for this equation.

Although the conditions of Neyman and Scott are only sufficient and not necessary for consistency, we do not have any systematic method of finding the variance of ξ^*. In the case of the true likelihood equations (4.10) to (4.13) an important property of $I_{1,0}(z)$ namely

$$\lim_{z\to 0} z^{-1} I_{1,0}(z) = \tfrac{1}{2} ,$$

ensures that the higher derivatives of the estimating equations are suitably bounded near the origin. A more appropriate approximation to $I_{1,0}(z)$ would be

$$I_{1,0}(z) = \begin{cases} 1 , & z \geq c , \\ 0 , & z < c , \end{cases}$$

where c is some suitably chosen positive number. Then we require the values of the integrals from c to ∞, rather than from 0 to ∞. If we assume c is a small positive number independent of N, and if ρ and σ^2 are known, then approximately

$$\mathrm{Var}(\xi^*) \simeq \sigma^2 N^{-1} W^{-1}$$

where

$$W = 1 - \tfrac{1}{2}\Gamma(\tfrac{1}{2})\gamma^{\frac{1}{2}}\,{}_1F_1[\tfrac{1}{2};\ 1;\ -\gamma]\ .$$

For γ large this is not much different from $\mathrm{Var}(\hat{\xi})$ as expected. For $A = 0.1$ the variances of ξ^* and $\hat{\xi}$ differ very slightly. However for $A = 1$, $\mathrm{Var}(\xi^*) \simeq 1.98\sigma^2 N^{-1}$ whereas $\mathrm{Var}(\hat{\xi}) \simeq 1.49\sigma^2 N^{-1}$ and for $A = 0.5$, $\mathrm{Var}(\xi^*) \simeq 2.33\sigma^2 N^{-1}$ whereas $\mathrm{Var}(\hat{\xi}) \simeq 1.88\sigma^2 N^{-1}$, indicating that the approximation ξ^* is less accurate for large A, as expected. Therefore the approximation would really only be useful if A were small.

We could also consider the approximation

$$I_{1,0}(z) = \begin{cases} 1 - 1/2z\ , & z \geq c\ , \\ 0\ , & z < c\ . \end{cases}$$

One disadvantage of this approximation is that now all four likelihood equations have to be iterated, and this increases the amount of work that has to be done.

12. The Circular Functional Relationship

Returning to equation (4.3) we now consider the allied functional relationship where the τ_i's are unknown for all $i = 1, 2, \ldots, N$. Here we consider the sequence $\{\tau_i\}_{i=1}^{N}$ to consist of fixed quantities, and the problem is to estimate θ consistently.

One method of eliminating τ_i is to obtain the maximum relative likelihood function of θ. This involves replacing τ_i by its maximum likelihood estimate $\hat{\tau}_i = \hat{\tau}_i(\theta)$. The likelihood is

$$l = -N \log 2\pi - N \log \sigma^2 - \tfrac{1}{2}\sigma^{-2} \sum_{i=1}^{N} \left\{ \left(X_i - \xi - \rho \cos \tau_i\right)^2 + \left(Y_i - \eta - \rho \sin \tau_i\right)^2 \right\}\ ,$$

and taking the derivative with respect to τ_i we obtain

$$\partial l/\partial\tau_i = -\sigma^{-2}\{(X_i-\xi-\rho\cos\tau_i)\rho\sin\tau_i - (Y_i-\eta-\rho\sin\tau_i)\rho\cos\tau_i\} .$$

Provided ρ is greater than zero, and using the equality $\cos^2\tau_i + \sin^2\tau_i = 1$, the above equation implies that

$$\hat{\cos}\,\tau_i = (X_i-\xi)/r_i \quad\text{and}\quad \hat{\sin}\,\tau_i = (Y_i-\eta)/r_i .$$

The maximum likelihood estimates are given by the solutions of

$$\sum_{i=1}^{N}(X_i-\xi) - \rho\sum_{i=1}^{N}(X_i-\xi)/r_i = 0 ,$$

$$\sum_{i=1}^{N}(Y_i-\eta) - \rho\sum_{i=1}^{N}(Y_i-\eta)/r_i = 0 ,$$

$$\hat{\rho} = \bar{r} ,$$

and

$$\hat{\sigma}^2 = (2N)^{-1}\sum_{i=1}^{N}(r_i-\bar{r})^2 .$$

Since $E(\bar{r}) > \rho$ at θ_0 , where the bias is independent of N and $\mathrm{Var}(\bar{r})$ tends to zero as N tends to infinity, $\hat{\rho}$ is inconsistent. It is easy to show that $\hat{\sigma}^2$ is inconsistent also. Chan (1965) tried to overcome this difficulty by assuming that ρ and σ^2 are incidental parameters as well as τ_i , and found the maximum relative likelihood function of (ξ, η) . As indicated in section 11 his estimators of ξ and η are still inconsistent.

At θ_0 ,

$$E\left[(\hat{\cos}\,\tau_i)^2+(\hat{\sin}\,\tau_i)^2\right] = 1 ,$$

but $\mathrm{Var}(\hat{\cos}\,\tau_i)$ and $\mathrm{Var}(\hat{\sin}\,\tau_i)$ are non-zero, so that

$$[E(\hat{\cos}\,\tau_i)]^2 + [E[\hat{\sin}\,\tau_i]]^2 < 1 = \sin^2\tau_i + \cos^2\tau_i .$$

The estimators of $\cos\tau_i$ and $\sin\tau_i$ are biased. In addition

$$E[\hat{\cos}\tau_i] = E_r[E(\cos\alpha_i|r)]$$
$$= E_r\left[I_{1,0}\left(r_i\rho/\sigma^2\right)\cos\tau_i\right] ,$$

since conditional on r_i , α_i has a von Mises distribution with mean τ_i and shape parameter $\rho r_i/\sigma^2$ (for more details see the latter part of this section).

$$\left|I_{1,0}\left(r_i\rho/\sigma^2\right)\right| < 1$$

if both r_i and ρ are positive, and the bias here only depends on the ratio σ^2/ρ^2 . We can conclude that

$$|E[\hat{\cos}\tau_i]| < \cos\tau_i .$$

Other methods of dealing with this situation are described among others by Andersen (1973) and Kalbfleisch and Sprott (1970). Briefly these methods rely on finding statistics which are ancillary or sufficient for the incidental parameters τ_i . By a change of variables from (X_i, Y_i) to (r_i, α_i) using the transformations

$$X_i - \xi = r_i\cos\alpha_i \quad\text{and}\quad Y_i - \eta = r_i\sin\alpha_i ,$$

here we are only interested in finding statistics which are ancillary for τ_i . For this we require two conditions as defined by Kalbfleisch and Sprott [(1970), p. 178]. These are

M1. A non-singular transformation taking

$$X_1, Y_1, X_2, Y_2, \ldots, X_N, Y_N \text{ to } r_1, r_2, \ldots, r_N, \alpha_1, \alpha_2, \ldots, \alpha_N ,$$

such that

$$\left(\prod_i dX_i dY_i\right) f(X_1, Y_1, \ldots, X_N, Y_N; \theta, \tau) = \left\{f(r_1, r_2, \ldots, r_N; \theta)\prod_i dr_i\right\}$$
$$\times\left\{f(\alpha_1, \alpha_2, \ldots, \alpha_N; \theta, \tau \mid r_1, \ldots, r_N)\prod_i d\alpha_i\right\} ,$$

and

M2. The second factor on the right hand side contains no available information concerning θ in the absence of knowledge of τ . Here the conditions of Kalbfleisch and Sprott have been rewritten in terms of the transformation from (X_i, Y_i) to (r_i, α_i) which is the obvious transformation to use here. The difficulty is condition M2; what exactly is meant by "no available information concerning θ "? More usually this factor will contain information about θ and use of $f(r_i|\theta)$ in forming the likelihood will only give an approximate marginal likelihood, as pointed out by Kalbfleisch and Sprott (1973). Sprott (1975) and Godambe (1976) give methods of establishing this condition, but the above density is not of the simple type required for their analyses.

Now

$$f(r_i, \alpha_i)dr_i d\alpha_i = r_i\sigma^{-2}\exp\left[-\tfrac{1}{2}\sigma^{-2}\left(r_i^2+\rho^2\right)\right]I_0\left(r_i\rho/\sigma^2\right)dr_i$$
$$\times \left[2\pi I_0\left(r_i\rho/\sigma^2\right)\right]^{-1}\exp\left[\rho r\sigma^{-2}\cos\left(\alpha_i-\tau_i\right)\right]d\alpha_i \ .$$

This certainly satisfies M1. Notice that just taking the first part of the right hand side is equivalent to assuming a rectangular distribution for α_i , which is not always reasonable. To establish M2, one possibility is to show that in some sense $r_1, r_2, \ldots, r_N$ are "sufficient" for θ by finding suitable transformations transitive on $X_1, Y_1, \ldots, X_N, Y_N$ and $\tau_1, \tau_2, \ldots, \tau_N$ which leave $r_1, r_2, \ldots, r_N$ and θ fixed. A transformation transitive on the set $\{\tau_1, \tau_2, \ldots, \tau_N\}$ maps each τ_i onto τ_j an element of the same set, where each element of the original set must be present in the final set. We want transformations which will map $(X_i, Y_i) \to (X_j, Y_j)$, $i \neq j$, $i, j = 1, 2, \ldots, N$. The only transformations leaving $r_1, \ldots, r_N$ and θ fixed are rotations about the point (ξ, η) , but unless $r_1, r_2, \ldots, r_N$ are equal these transformations will not map $(X_i, Y_i) \to (X_j, Y_j)$.

The distribution of α_i given r_i is

$$f(\alpha_i|r_i, \tau_i, \theta) = \left[2\pi I_0\left(r_i\rho/\sigma^2\right)\right]^{-1}\exp\left[\rho r_i\sigma^{-2}\cos\left(\alpha_i-\tau_i\right)\right] ,$$

that is, α_i given r_i and τ_i has a von Mises distribution with mean τ_i and shape parameter $r_i\rho/\sigma^2$. The distance $\alpha_i - \tau_i$ contains some information about the ratio $\rho r_i/\sigma^2$.

The approach of Kalbfleisch and Sprott is not applicable here, and thus questions about the validity of their method do not arise. It should be noted however, that some doubts about their analysis have been raised expecially with respect to determining whether or not condition M2 holds in practice.

As shown earlier $\cos\hat{}\tau_i$ is a biased estimator of $\cos\tau_i$. It might be more profitable to seek an unbiased estimator of $\cos\tau_i$ other than $\rho^{-1}(X_i-\xi)$. In fact there is a whole family of such estimators, and they are of the form

$$\cos\tau_i^* = \left(\frac{X_i-\xi}{r_i}\right)\frac{r_i^{\nu}}{\rho^{\nu}}\frac{I_\nu\left(r_i\rho/\sigma^2\right)}{I_1\left(r_i\rho/\sigma^2\right)}(k+1)^{\nu+1}\exp\left\{-\frac{k}{2\sigma^2}\left[r_i^2-\rho^2/(1+k)\right]\right\} ,$$

where $\nu > -1$ and $k \geq 0$. We require further restrictions on ν , such as $\nu \geq +1$, if we are to apply the theory of Neyman and Scott (1948):

$$E\left(\cos\tau_i^*|r_i, \tau_i\right) = \cos\tau_i r_i^{\nu} I_{\nu,0}\left(r_i\rho/\sigma^2\right)(k+1)^{\nu+1}\rho^{-\nu}\exp\left\{-\frac{k}{2\sigma^2}\left(r_i^2+\frac{\rho^2}{1+k}\right)\right\} ,$$

and using a modified form of Watson [(1944), p. 394],

$$E\left(\cos\tau_i^*\right) = \cos\tau_i .$$

In this way estimators for ξ and η can be derived:

$$\sum_{i=1}^{N}\left(X_i-\xi\right) - \rho\sum_{i=1}^{N}\cos\tau_i^* = 0 ,$$

$$\sum_{i=1}^{N}\left(Y_i-\eta\right) - \rho\sum_{i=1}^{N}\sin\tau_i^* = 0 ,$$

$$\rho - \sum_{i=1}^{N} r_i I_{1,0}\left(r_i\rho/\sigma^2\right) = 0$$

and

$$\sigma^2 - (2N)^{-1}\sum_{i=1}^{N}\left(r_i^2-\rho^2\right) = 0 .$$

To decide upon an optimal ν and k value would be extremely messy.

Neyman and Scott require conditions on higher derivatives to be satisfied, and these should be checked. To do this it is easier to set $k = 0$ and ν integral. Later we will look at $\nu = 2$ and $k = 0$ to see how the resulting estimators perform in this case. In particular we are interested to find out whether or not any iterative procedure based on the above equations is stable given certain underlying distributions of τ_i .

13. Departure from Uniformity of τ_i

Returning to equation (4.8) it is interesting to examine the effects on the estimating equations for $\boldsymbol{\theta}$ if we assume that the distribution of τ_i departs from uniformity. Let τ_i , $i = 1, 2, \ldots, N$, have the von Mises distribution $M(\mu, \kappa)$ so that

$$g(\tau_i) = \left(2\pi I_0(\kappa)\right)^{-1}\exp\left[\kappa \cos\left(\tau_i-\mu\right)\right] \ , \quad 0 \le \tau_i < 2\pi \ .$$

As κ tends to zero this distribution tends to the uniform distribution on $[0, 2\pi)$. Consequently small κ represents a minor departure from uniformity. By integration of (4.8) using Gradshteyn and Ryzhik [(1965), p. 488] we obtain

$$f(X_i, Y_i) = \left[2\pi\sigma^2 I_0(\kappa)\right]^{-1} \exp\left[-\tfrac{1}{2}\sigma^{-2}\left[(X_i-\xi)^2+(Y_i-\eta)^2+\rho^2\right]\right] I_0\left[\left(p_i^2+q_i^2\right)^{\frac{1}{2}}\right] \ ,$$

where

$$p_i = \kappa \cos \mu + \left(X_i-\xi\right)\rho/\sigma^2 = \kappa \cos \mu + \rho r_i \cos \alpha_i/\sigma^2$$

and

$$q_i = \kappa \sin \mu + \left(Y_i-\eta\right)\rho/\sigma^2 = \kappa \sin \mu + \rho r_i \sin \alpha_i/\sigma^2 \ .$$

The log-likelihood is

$$l = -N \log_e 2\pi - N \log_e \sigma^2 - N \log_e I_0(\kappa) - \tfrac{1}{2}\sigma^{-2} \sum_{i=1}^{N} \left(r_i^2+\rho^2\right)$$
$$+ \sum_{i=1}^{N} \log_e I_0\left[\left(p_i^2+q_i^2\right)^{\frac{1}{2}}\right] \ .$$

The maximum likelihood estimates $\hat{\theta} = \left(\hat{\kappa}, \hat{\xi}, \hat{\eta}, \hat{\rho}, \hat{\sigma}^2, \hat{\mu}\right)$ are given by the solutions of the equations

$$I_{1,0}(\kappa) = N^{-1} \sum_{i=1}^{N} \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu})\left(p_i \cos \mu + q_i \sin \mu\right) ,$$

$$\sum_{i=1}^{N} \left(X_i - \xi\right) - \rho \sum_{i=1}^{N} p_i \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu}) = 0 ,$$

$$\sum_{i=1}^{N} \left(Y_i - \xi\right) - \rho \sum_{i=1}^{N} q_i \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu}) = 0 ,$$

$$\rho = N^{-1} \sum_{i=1}^{N} \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu})\left[p_i\left(X_i - \xi\right) + q_i\left(Y_i - \eta\right)\right],$$

$$\sigma^2 = (2N)^{-1} \sum_{i=1}^{N} \left(r_i^2 - \rho^2\right) ,$$

and

$$\sin \mu \sum_{i=1}^{N} p_i \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu}) - \cos \mu \sum_{i=1}^{N} q_i \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu}) = 0 .$$

The latter equation can also be written

$$\sin \mu \sum_{i=1}^{N} \left(X_i - \xi\right) \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu}) - \cos \mu \sum_{i=1}^{N} \left(Y_i - \eta\right) \tilde{\nu}^{-1} I_{1,0}(\tilde{\nu}) = 0 ,$$

where

$$\tilde{\nu} = \left[\kappa^2 + r_i^2 \rho^2 / \sigma^4 + 2\kappa r_i \rho \sigma^{-2} \cos\left(\alpha_i - \mu\right)\right]^{\frac{1}{2}} .$$

Using the equality $\sin^2\mu + \cos^2\mu = 1$ and the second and third of the above estimating equations we derive

$$\cos \mu = \sum_{i=1}^{N} \left(X_i - \xi\right) \Big/ \left\{\left[\sum_{i=1}^{N} \left(X_i - \xi\right)\right]^2 + \left[\sum_{i=1}^{N} \left(Y_i - \xi\right)\right]^2\right\}^{\frac{1}{2}} .$$

In addition from the first equation

$$I_{1,0}(\kappa) = \rho^{-1} N^{-1} \left\{\left[\sum_{i=1}^{N} \left(X_i - \xi\right)\right]^2 + \left[\sum_{i=1}^{N} \left(Y_i - \eta\right)\right]^2\right\}^{\frac{1}{2}} .$$

From Watson [(1944), p. 365],

$$\tilde{\nu}^{-1} I_1(\tilde{\nu}) = 2 \sum_{m=0}^{\infty} (-1)^m (m+1) \frac{I_{m+1}(\kappa)}{\kappa} \frac{I_{m+1}\left(r_i \rho / \sigma^2\right)}{\left(r_i \rho / \sigma^2\right)} C_m^1\left[-\cos\left(\alpha_i - \mu\right)\right] ,$$

and

$$I_0(\tilde{\nu}) = \sum_{m=0}^{\infty} \varepsilon_m I_m(\kappa) I_m\left(r_i\rho/\sigma^2\right)\cos\left[m\left(\alpha_i-\mu\right)\right] ,$$

where $\varepsilon_m = 1$ if $m = 0$ and 2 otherwise and the $C_m^1(x)$ are Gegenbauer polynomials defined as the coefficients in the expansion

$$\left(1-2tx+t^2\right)^{-1} = \sum_{m=0}^{\infty} C_m^1(x)t^m ,$$

or

$$C_m^1(x) = \sum_{k=0}^{[m/2]} (-1)^k 2^{m-2k} \frac{\Gamma(m+1-k)}{k!(m-2k)} x^{m-2k}$$

and

$$C_0^1(x) = 1 .$$

$[m/2]$ is defined as the largest integer contained in $(m/2)$. If we are only interested in the case κ very small, then $I_s(\kappa) \sim (\frac{1}{2}\kappa)^s/\Gamma(s+1)$ as κ tends to zero. Using this, $\tilde{\nu}^{-1}I_1(\tilde{\nu})$ and $I_0(\tilde{\nu})$ can be expanded out as polynomial series in κ and consequently we can find an approximation to $\tilde{\nu}^{-1}I_{1,0}(\tilde{\nu})$. To first order this is

$$\tilde{\nu}^{-1}I_{1,0}(\tilde{\nu}) = \left(r_i\rho/\sigma^2\right)^{-1}\left[I_{1,0}\left(r_i\rho/\sigma^2\right)+\left\{2I_{2,0}\left(r_i\rho/\sigma^2\right)-I_{1,0}\left(r_i\rho/\sigma^2\right)\right\}\kappa\cos\left(\alpha_i-\mu\right)\right]$$

and therefore we obtain

$$\sum_{i=1}^{N}\left(\frac{X_i-\xi}{r_i}\right)\left\{r_i-I_{1,0}\left(r_i\rho/\sigma^2\right)\right\} - \kappa\rho\sum_{i=1}^{N}\frac{X_i-\xi}{r_i}\cos\left(\alpha_i-\mu\right)$$
$$\times\left[2I_{2,0}\left(r_i\rho/\sigma^2\right)-I_{1,0}^2\left(r_i\rho/\sigma^2\right)\right] - \kappa\sigma^2\cos\mu\sum_{i=1}^{N} r_i^{-1}I_{1,0}\left(r_i\rho/\sigma^2\right) = 0 ,$$

and

$$\rho - N^{-1}\sum_{i=1}^{N} r_i I_{1,0}\left(r_i\rho/\sigma^2\right) + \kappa N^{-1}\sum_{i=1}^{N} r_i\cos\left(\alpha_i-\mu\right)$$
$$\times\left[2I_{2,0}\left(r_i\rho/\sigma^2\right)-I_{1,0}^2\left(r_i\rho/\sigma^2\right)\right] + \kappa\sigma^2\rho^{-1}N^{-1}\sum_{i=1}^{N}\cos\left(\alpha_i-\mu\right)I_{1,0}\left(r_i\rho/\sigma^2\right) = 0 .$$

Even in this simplified form, these equations would be extremely difficult to solve, and this will not be attempted here. However we can see the effect of the non-uniformity of τ_i . The estimation procedure is no longer symmetric, since now there is a preferred direction for τ_i . Also the distribution of α_i is no longer independent of r_i , and the parameters of this distribution must be estimated jointly with the circular structural parameters.

14. The Elliptic Structural Model

The elliptic structural model is defined as follows. Let

$$\begin{aligned} X_i &= \xi + p(t_i-\varphi)\cos t_i + \varepsilon_i \,, \\ Y_i &= \eta + p(t_i-\varphi)\sin t_i + \nu_i \,, \end{aligned} \qquad i = 1, 2, \ldots ,$$

where (ε_i, ν_i) is defined as before. (ξ, η) is the centre of the ellipse. Let $2a$ be the length of the major axis, $2b$ the length of the minor axis, $+\varphi$ the angle between the x-axis and the major axis, and t_i is the angle between the observation (X_i, Y_i) and the x-axis in the anticlockwise direction. $p(t_i-\varphi)$ is the radial length of the ellipse at this point. We also define the eccentricity k by the equality $b^2 = a^2(1-k^2)$, $0 \le k < 1$. e is the more usual notation, but k is used here to save confusion with the transcendental number also denoted by e . This model reduces to the circular model when $k = 0$. Using the parametric equation for the ellipse

$$p(t_i-\varphi) = b\left[1-k^2\sin^2(t_i-\varphi)\right]^{-\frac{1}{2}} ,$$

we can show that

$$[E(X_i-\xi)]^2/a^2 + [E(Y_i-\eta)]^2/b^2 = 1 .$$

The circular model has four structural parameters ξ, η, ρ and σ^2 , whereas here we have six, ξ, η, a, b, φ and σ^2 , or alternatively ξ, η, a, k, φ and σ^2 . Therefore we would expect the analysis to be more complicated. Here the ellipse has been specified by its polar coordinates since then the angle t_i has a physical meaning for the ellipse as described above. The more usual parametric specification, for $\varphi = 0$, is given by

$$\left.\begin{aligned} X_i &= \xi + a\cos\tau_i + \varepsilon_i\,, \\ Y_i &= \eta + b\sin\tau_i + \nu_i\,, \end{aligned}\right\} \quad i = 1, 2, \ldots .$$

Here τ_i is called the angle of eccentricity. If we consider two concentric circles with respective radii a and b, then τ_i is the angle that connects a point on the inner circle which has y coordinate $E(Y_i)$ with a point on the outer circle which has x coordinate $E(X_i)$.

In the circular model we assumed $t_i \overset{d}{\sim} R[0,\ 2\pi)$, and found $f(X_i,\ Y_i)$ by integrating $f(X_i,\ Y_i|t_i)$. For the ellipse we want to assume a distribution for t_i which reduces to $R[0,\ 2\pi)$ as k tends to zero. There are a large number of such distributions of which two deserve particular attention.

If t_i has a probability distribution function

$$g_1(t_i) = [4E(k)]^{-1}\left[1-k^2\sin^2(t_i-\varphi)\right]^{\frac{1}{2}},$$

where $E(k)$ is the complete elliptic integral of the second kind; this has the geometrical interpretation that the probability that t_i falls in equal arc lengths on the ellipse are equal. If t_i has the probability distribution function given by

$$g_2(t_i) = [4K(k)]^{-1}\left[1-k^2\sin^2(t_i-\varphi)\right]^{-\frac{1}{2}},$$

where $K(k)$ is the complete elliptic integral of the first kind, this has the geometrical interpretation that the probability t_i falls in a certain arc length is proportional to the area subtended at the centre of the ellipse by that arc. The appropriate distribution would depend on the physical interpretation of the angle t_i. In addition we could imagine that $t_i \overset{d}{\sim} R[0,\ 2\pi)$ giving

$$g_3(t_i) = (2\pi)^{-1}.$$

As in the circular model it is convenient to write

$$X_i - \xi = r_i(\alpha_i-\varphi)\cos\alpha_i ,$$

$$Y_i - \eta = r_i(\alpha_i-\varphi)\sin\alpha_i ,$$

where

$$r_i = \left\{(X_i-\xi)^2+(Y_i-\eta)^2\right\}^{\frac{1}{2}} .$$

Then $f(X, Y_i)$ is given by

$$f(X_i, Y_i) = \left(2\pi\sigma^2\right)^{-1}\exp\left(-\tfrac{1}{2}\sigma^{-2}r_i^2\right)$$

$$\times \int_0^{2\pi} \exp\left[-\tfrac{1}{2}\sigma^{-2}\left\{p^2(t_i-\varphi)-2r_ip(t_i-\varphi)\cos(t_i-\alpha_i)\right\}\right]g(t_i)dt_i .$$

Unfortunately no matter which $g(t)$ is used it does not appear that the right hand side can be simplified further, even with the use of Jacobian elliptic functions from say Byrd and Friedman (1971). To use any such table of integrals of Jacobian elliptic functions we would have to expand out the exponential function as a power series and integrate term-by-term. Even if such integrals were available, the answer would be far too messy to be of any practical use. It does not appear that the t_i's can be eliminated analytically from the probability distribution function and consequently the maximum likelihood estimators for the structural parameters of the ellipse can not be found. If k is close to zero we could use the binomial theorem, and the approximation to the exponential function for small exponent to find an approximation to the probability density function, using Gradshteyn and Ryzhik [(1965), eq. 3.937], where say $g(t) = g_1(t)$. This gives a very complicated expression involving $I_0(rb/\sigma^2)$, $I_1(rb/\sigma^2)$ and $I_2(rb/\sigma^2)$. Even this expression is too messy for practical use. Notice that similar problems arise in integrating $f(X_i, Y_i|t_i)$ if we consider a circular structural model where the variances of ε_i and ν_i are different.

We can, of course, calculate the moments of X_i and Y_i . Doing this we find that $\overline{X} = N^{-1}\sum X_i$ and $\overline{Y} = N^{-1}\sum Y_i$ are unbiased and consistent estimators of ξ and η respectively. Unfortunately the higher moments of both X_i and Y_i depend on the orientation of the ellipse as well as the eccentricity k. Even this simple estimator of the centre (ξ, η) has moments which are difficult to handle.

Here the distribution of r_i/σ given t_i is a non-central chi distribution with two degrees of freedom and non-centrality parameter $p^2(t_i-\varphi)/\sigma^2$. However the moments of r_i do not involve the orientation φ. We will assume that $g(t) = g_1(t)$, and then

$$E(r_i) = (\tfrac{1}{2}\pi)^{\frac{1}{2}}\sigma[E(k)]^{-1} \sum_{n=0}^{\infty} \frac{(-\frac{1}{2})_n}{(1)_n n!} \frac{(-1)^n b^{2n}}{(2\sigma^2)^n} \int_0^{\pi/2} (1-k^2\sin^2 t)^{\frac{1}{2}-n} dt \ .$$

This can be simplified using Byrd and Friedman [(1971), eq. 314 and 315],

$$E(r) = \frac{(\frac{1}{2}\pi)^{\frac{1}{2}}\sigma}{E(k)} \left[\left\{\sum_{n=0}^{\infty} \frac{(-\frac{1}{2})_n}{(1)_n n!} \frac{(-1)^n b^{2n}}{(2\sigma^2)^n} \int_0^K nd^{2n}u\,du\right\} + E(k)\right] ,$$

where $nd(u)$ is a Jacobian elliptic function. Again each integral in this sum must be integrated iteratively and the solution would be extremely messy. As in the circular model, the even moments of r_i are far easier to deal with; for example,

$$E(r^2) = 2\sigma^2 + b^2K(k)/E(k) \ .$$

Both $K(k)$ and $E(k)$ can be written in terms of hypergeometric functions (see Byrd and Friedman [(1971), p. 15]). Also

$$E(r^4) = 8\sigma^4 + 8\sigma^2b^2\ K(k)/E(k) + b(1-k^2)^{-1} \ .$$

The even moments of r involve three parameters and not two as in the circular model, so basing statistics on the moments of r^2 would be extremely difficult.

Here we have two radial parameters a and b, or b and k, but only one radial estimator r. If σ^2/b^2 is small, $\bar{r}$ is an estimator of the ratio $b\pi/2E(k)$, and for small k we can show that

$$(ab)^{\frac{1}{2}} < b\pi/2E(k) < \tfrac{1}{2}(a+b) \ .$$

$\bar{r}$ can not be used in the same manner as it was used for the circular model. This means that a different approach to the problem of estimating the parameters would have to be found, and there is no obvious extension to estimating the parameters of the circular model that is applicable here.

15. Comments and Conclusions

In this chapter we examined the circular structural model. After introducing the model and proving identifiability we went on to examine the conditions for maximum likelihood estimation. We briefly indicated that the model would be very hard to solve in practice and then examined a special case. For this special case the angular variables were assumed to have a rectangular distribution on the interval $[0, 2\pi)$. Maximum likelihood estimation of the structural parameters was then considered. A large number of other estimators were also considered, mainly estimators for ρ and σ^2. For ξ and η only the obvious alternative estimators $\overline{X}$ and $\overline{Y}$ have been considered. Assuming ξ, η and σ^2 to be known it was discovered that $\overline{r}$ is a biased but very efficient estimator of ρ. Therefore $\overline{r}$ can be used to find an accurate bound for β/ρ^2. Numerical values of β/ρ^2 were obtained for a certain range of σ^2/ρ^2 values, and it was found that at least for σ^2/ρ^2 small, the bound for β/ρ^2 found using $\overline{r}$ was an accurate approximation to β/ρ^2. Thus we can obtain approximate algebraic expressions for the variances and covariances of $\hat{\rho}$ and $\hat{\sigma}^2$, if σ^2/ρ^2 is small.

If ξ and η are known, ρ and σ^2 can be estimated using the second and fourth sample moments of r_i. It was shown that the resulting estimators are asymptotically unbiased, and quite efficient. If ξ and η are not known, then we can consider estimating them and substituting these estimates into the previous estimators of ρ and σ^2. The estimator of ρ is asymptotically unbiased but if σ^2 is unknown it is inefficient at least for σ^2/ρ^2 small. To overcome this, estimators based on the square of the distance between pairs of points (X_i, Y_i) and (X_j, Y_j) were considered. These estimators are asymptotically unbiased and efficient. If the sample size is large, one disadvantage of these estimators is that they involve a large number of calculations, of the order N^2. If σ^2/ρ^2 is small it might be quicker to use the maximum likelihood estimators where $I_{1,0}(z)$ has been replaced by an approximation.

The use of two approximations to $I_{1,0}(z)$ was discussed and comments made about finding the variances of the resulting estimators. Estimators proposed by Chan (1965) were also briefly mentioned. The circular functional relationship was discussed and a set of consistent estimating equations was found in this case. In addition two variations on the circular structural model were mentioned. In the first the effect on the maximum likelihood estimating equations of non-uniformity of the distribution of τ_i was examined. In the second the intractibality of the elliptic structural model was demonstrated. The special case of the circular structural model discussed in this chapter may have appeared at first sight to be a simple case. However as the two variations on the model indicate most generalizations of it become intractable. Even assuming different variances for ε_i and ν_i results in difficulties in integrating equation (4.8).

If ξ and η are known then estimation of ρ and σ^2 is just estimation of the scale and non-centrality parameter for a non-central chi distribution with known degrees of freedom, $m = 2$. In the next chapter we generalize this estimation to general m where m is still known.

CHAPTER 5

MAXIMUM LIKELIHOOD ESTIMATION IN THE NON-CENTRAL CHI-DISTRIBUTION

1. Introduction

In the analysis of the circular structural model it was necessary to consider the estimation of the scale and non-centrality parameters of a non-central chi-distribution with known degrees of freedom m. We considered the case $m = 2$, but there is no difficulty in extending the analysis to general m.

Johnson (1962) considered the general folded normal distribution, the case $m = 1$. Leone *et al* (1961) and Elandt (1961) considered moment estimators for both the scale and non-centrality parameters in the case $m = 1$. For more general m, the scale parameter is usually assumed to be known. Meyer (1967) obtained the maximum likelihood estimator of the non-centrality parameter for known scale parameter and $m = 2$. He concentrated on proving the existence and uniqueness of the likelihood estimator, and did not consider its variance. Dwivedi and Pandey (1975) extended Meyer's argument to general m and found an approximation to the maximum likelihood estimator. Perlman and Rasmussen (1975) and Neff and Strawderman (1976) examined the moment estimator for the non-centrality parameter and based improved estimators on this.

In this chapter we derive the variance of the maximum likelihood estimator of the non-centrality parameter where the degrees of freedom is known, both where the scale parameter is known and unknown. Also comparisons are made with the moment estimator.

2. The Model

Let

$$X_j = \rho \sin \varphi_1 \sin \varphi_2 \ldots \sin \varphi_{j-1} \cos \varphi_j + \varepsilon_j \ , \quad j = 1, 2, \ldots, m-1 \ , \tag{5.1}$$

and

$$X_m = \rho \sin \varphi_1 \sin \varphi_2 \ldots \sin \varphi_{m-1} + \varepsilon_m \ ,$$

where $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)$ is an m-dimensional multivariate normal distribution with $E\varepsilon_j = 0$, $E(\varepsilon_j \varepsilon_i) = 0$, $i \neq j$, and $E\left(\varepsilon_j^2\right) = \sigma^2$,

$i, j = 1, 2, \ldots, m$; $0 \le \varphi_j \le \pi$, $j = 1, 2, \ldots, m-2$ and $0 \le \varphi_{m-1} \le 2\pi$. Then

$$r^2/\sigma^2 = \sum_{j=1}^{m} X_j^2/\sigma^2$$

has a non-central chi-squared distribution with non-centrality parameter $\lambda^2 = \rho^2/\sigma^2$. The probability distribution function of r is known to be

$$f(r) = r^{\frac{1}{2}m}\rho^{-\frac{1}{2}m+1}\sigma^{-2}\exp\{-\tfrac{1}{2}\sigma^{-2}(r^2+\rho^2)\}I_{\frac{1}{2}m-1}(r\rho/\sigma^2) \ . \tag{5.2}$$

Here the non-central chi distribution is considered rather than the non-central chi-squared distribution to avoid the use of a square root in the argument of the modified Bessel function. There is no difference in the results. Also although λ^2 is the non-centrality parameter, ρ is an important parameter and will be considered as well. Notice that although the model has been derived for m integral, (5.2) also holds for general m greater than zero.

3. Maximum Likelihood Estimation (m known)

The log-likelihood for a sample of size N is

$$l = \tfrac{1}{2}m \sum_{i=1}^{N} \ln r_i - (\tfrac{1}{2}m+1)N \ln \rho - N \ln \sigma^2 - \tfrac{1}{2}\sigma^{-2} \sum_{i=1}^{N} \left(r_i^2+\rho^2\right) + \sum_{i=1}^{N} \ln I_{\frac{1}{2}m-1}\left(r\rho/\sigma^2\right) \ .$$

Taking the derivative with respect to ρ we obtain

$$\partial l/\partial\rho = -(\tfrac{1}{2}m+1)N\rho^{-1} - N\rho\sigma^{-2} + \sigma^{-2} \sum_{i=1}^{N} r_i I'_{\frac{1}{2}m-1}\left(r_i\rho/\sigma^2\right)/I_{\frac{1}{2}m-1}\left(r_i\rho/\sigma^2\right) ,$$

and using the equality from Watson [(1944), p. 79],

$$I'_v(z) = I_{v+1}(z) + vz^{-1}I_v(z) ,$$

the expression for $\partial l/\partial\rho$ reduces to

$$\partial l/\partial\rho = -N\rho\sigma^{-2} + \sigma^{-2} \sum_{i=1}^{N} r_i I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r_i\rho/\sigma^2\right) .$$

Thus the maximum likelihood equations are

$$\rho - N^{-1} \sum_{i=1}^{N} r_i I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r_i \rho/\sigma^2\right) = 0 , \tag{5.3}$$

and

$$\sigma^2 = (mN)^{-1} \sum_{i=1}^{N} \left(r_i^2-\rho^2\right) = 0 . \tag{5.4}$$

Alternatively $\hat{\lambda}$ is the solution of the equation

$$\lambda - N^{-1} \sum_{i=1}^{N} \left(r_i/\sigma\right) I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r_i \lambda/\sigma\right) = 0 ,$$

where

$$\sigma^2 = \left(mN+\lambda^2\right)^{-1} \sum_{i=1}^{N} r_i^2 .$$

Later the differences between these two parameterizations of the model will be discussed.

Neither set of equations is immediately soluble and would have to be solved iteratively, using perhaps

$$\bar{r} = N^{-1} \sum r_i \quad \text{and} \quad \sigma^2 = N^{-1} \sum \left(r_i-\bar{r}\right)^2$$

as initial values.

The case $m = 2$ has been given in detail in chapter 4, and since the case with general m is similar, less details will be given here.

4. Behaviour of $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$

Hartman and Watson (1974) proved that $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$ is a monotonically increasing distribution function for $z \geq 0$, when $m \geq 2$. This does not include the case $m = 1$, but from Abramowitz and Stegun [(1972), eq. 10.2.13],

$$I_{\frac{1}{2},-\frac{1}{2}}(z) = \tanh z .$$

$\tanh z$ is a monotonically increasing distribution function for $z \geq 0$, as desired. Also from the series definition of the modified Bessel function we can show that

$$\lim_{z\to 0} z^{-1} I_{\frac{1}{2}m,\frac{1}{2}m-1}(z) = m^{-1} .$$

For odd m , $I_{\frac{1}{2}m}$ is expressible in terms of hyperbolic functions (see Abramowitz and Stegun [(1972), eq. 10.2.24]), and tabulation is not needed. For example

$$I_{\frac{3}{2},\frac{1}{2}}(z) = \coth z - z^{-1} ,$$

and this is quite easy to evaluate. For m even, $I_{\frac{1}{2}m}$ has only been sketchily tabulated, except for $m = 0$ and $m = 2$. For large z and $m \geq 2$, using the asymptotic series for $I_\nu(z)$ from Abramowitz and Stegun [(1972), eq. 9.7.1] we derive

$$I_{\frac{1}{2}m,\frac{1}{2}m-1}(z) \simeq 1 - (m-1)(2z)^{-1} + (m-1)(m-3)\left(8z^2\right)^{-1} + (m-1)(m-2)\left(8z^3\right)^{-1} + \ldots$$

For m odd and z small, a convergent series for $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$ can be established using the ascending series for $I_{n+\frac{1}{2}}(z)$ given by Abramowitz and Stegun [(1972), eq. 10.2.5]. As before there is no approximation which is valid over the entire range of z .

5. The Asymptotic Dispersion Matrix

Define

$$\begin{aligned}\beta &= E\left[r^2 I^2_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)\right] \\ &= \sigma^{-2}\rho^{-\frac{1}{2}m+1}e^{-\rho^2/2\sigma^2}\int_0^\infty r^{\frac{1}{2}m+2}e^{-r^2/2\sigma^2} I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right) I_{\frac{1}{2}m}\left(r\rho/\sigma^2\right) dr .\end{aligned}$$

Let $z = r\rho/\sigma^2$ and $A = \sigma^2/\rho^2 = \lambda^{-2}$ as before, then

$$\beta/\rho^2 = A^{\frac{1}{2}m+2}\exp\left(-\tfrac{1}{2}A^{-1}\right)\int_0^\infty z^{\frac{1}{2}m+2}\exp\left(-Az^2/2\right) I_{\frac{1}{2}m,\frac{1}{2}m-1}(z) I_{\frac{1}{2}m}(z)\, dz . \quad (5.5)$$

The expectations of the second derivatives of the log-likelihood yield

$$\mathrm{Var}(\hat{\rho}) \simeq \sigma^2 N^{-1}\Delta^{-1}\left\{\tfrac{1}{2}mA-1+A^{-1}\left(\beta/\rho^2 - 1\right)\right\} ,$$

$$\mathrm{Cov}\left(\hat{\rho},\, \hat{\sigma}^2\right) \simeq \rho\sigma^2 N^{-1}\Delta^{-1}\left\{\left(\beta/\rho^2 - 1\right)-A\right\} ,$$

and

$$\mathrm{Var}\left(\hat{\sigma}^2\right) \simeq \sigma^4 N^{-1}\Delta^{-1}\left(\beta/\rho^2 - 1\right) ,$$

where

$$\Delta = \left(\beta/\rho^2 - 1\right)\left(\tfrac{1}{2}m+A^{-1}\right) - 1 \ .$$

Since $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$ is a monotonically increasing function of z between zero and unity, β/ρ^2 is bounded by

$$\left\{E\left[rI_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)\right]\right\}^2 \le \beta/\rho^2 \le E\left[r^2 I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)\right]$$

However neither of these bounds yields an accurate approximation to β/ρ^2 . As before a more appropriate lower bound can be derived by using the inequality $\Delta > 0$, and a more appropriate upper bound by a careful examination of the integral defining β/ρ^2 (see Appendix A). These bounds are

$$1 + \tfrac{1}{2}\gamma^{-1} - m\gamma^{-2}/8 + m^2\gamma^{-3}/32 \ \ldots$$
$$\le \beta/\rho^2 \le 1 + 5\gamma^{-1}/8 - (20m-13)\gamma^{-2}/128 \ . \qquad (5.6)$$

To evaluate β/ρ^2 using (5.5) the Gregory formula for integration was used, as in the case $m = 2$. The values of m considered were $m = 1, 3, 5$ and 7 . These odd values were much easier to use than even values because $I_\nu(z)$ can be expressed in terms of hyperbolic functions if 2ν is odd. The same values for A and h , the interval width, were considered as for the case $m = 2$. Much wider interval widths sufficed here to get the desired six decimal place accuracy. β/ρ^2 could probably be found to greater accuracy, but this was not seriously attempted. Again β/ρ^2 is close to the lower bound given by (5.6) at least for A small. (See Table VI and Figure 5.1 for the calculated values of β/ρ^2 .)

Now

$$E(r) = \sigma^{-2}\rho^{-\frac{1}{2}m+1}e^{-\rho^2/2\sigma^2}\int_0^\infty r^{\frac{1}{2}m+1}e^{-r^2/2\sigma^2} I_{\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)dr$$
$$= \left(2\sigma^2\right)^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}m+\frac{1}{2})}{\Gamma(\frac{1}{2}m)} \ {}_1F_1[-\tfrac{1}{2};\ \tfrac{1}{2}m;\ -\gamma] \ ,$$

using an analysis similar to the one used to examine (4.23). Since $I_{\frac{1}{2}m-1}\left(r\rho/\sigma^2\right) \gg I_{\frac{1}{2}m}\left(r\rho/\sigma^2\right)$ for all $r > 0$, it is easy to show that $E(\bar{r}) > \rho$, independently of N . However as before we can use $\bar{r}$ to find an improved lower bound for β/ρ^2 via the inequality $\mathrm{ARE}(\bar{r},\ \hat{\rho}) \le 1$, for

A/m	1	2	3	5	7
1.0	1.73391	1.52145^-	1.40914	1.28825^-	1.22321
0.9	1.68632	1.49569	1.39266	1.27973	1.21797
0.8	1.63458	1.46683	1.37384	1.26977	1.21178
0.7	1.57809	1.43424	1.35213	1.25796	1.20432
0.6	1.51616	1.39712	1.32680	1.24375^-	1.19518
0.5	1.44796	1.35442	1.29682	1.22629	1.18367
0.4	1.37262	1.30476	1.26074	1.20432	1.16876
0.3	1.28933	1.24634	1.21646	1.17577	1.14863
0.2	1.19787	1.17695	1.16094	1.13714	1.11991
0.1	1.09996	1.09467	1.09001	1.08211	1.07559
0.09	1.08998	1.08572	1.08190	1.07533	1.06981
0.08	1.07999	1.07664	1.07360	1.06827	1.06371
0.07	1.06999_8	1.06745^-	1.06510	1.06091	1.05727
0.06	1.05999_{95}	1.05814	1.05640	1.05325^-	1.05045^-
0.05	1.04999_{94}	1.04871	1.04750	1.04526	1.04323
0.04	1.03999_{9997}	1.03918	1.03840	1.03693	1.03558
0.03	1.03000	1.02953	1.02910	1.02825	1.02746
0.02	1.02000	1.01980	1.01960	1.01922	1.01885^-
0.01	1.01000	1.00995^-	1.00990	1.00980	1.00970

β/ρ^2

TABLE VI

σ^2 known. We require $\gamma >> m$ for the approximation to the confluent hypergeometric function from Slater [(1960), eq. 4.1.2] to be valid. For γ large, this analysis yields

$$\beta/\rho^2 \geq 1 + \tfrac{1}{2}\gamma^{-1} - (m-1)\gamma^{-2}/8 + (m-1)(m-2)\gamma^{-3}/32 \;\ldots\; . \tag{5.7}$$

This bound is only about

$$\gamma^{-2}/8 - (4m-3)\gamma^{-3}/32 \;\ldots\; ,$$

above the bound given by (5.6) which is the lower limit for the determinant Δ to be positive. The right hand side of (5.7) is quite accurate as an approximation if A is small.

When $m = 1$ additional sharper bounds can be obtained. For $m = 1$ the approximation to the confluent hypergeometric function used to obtain an

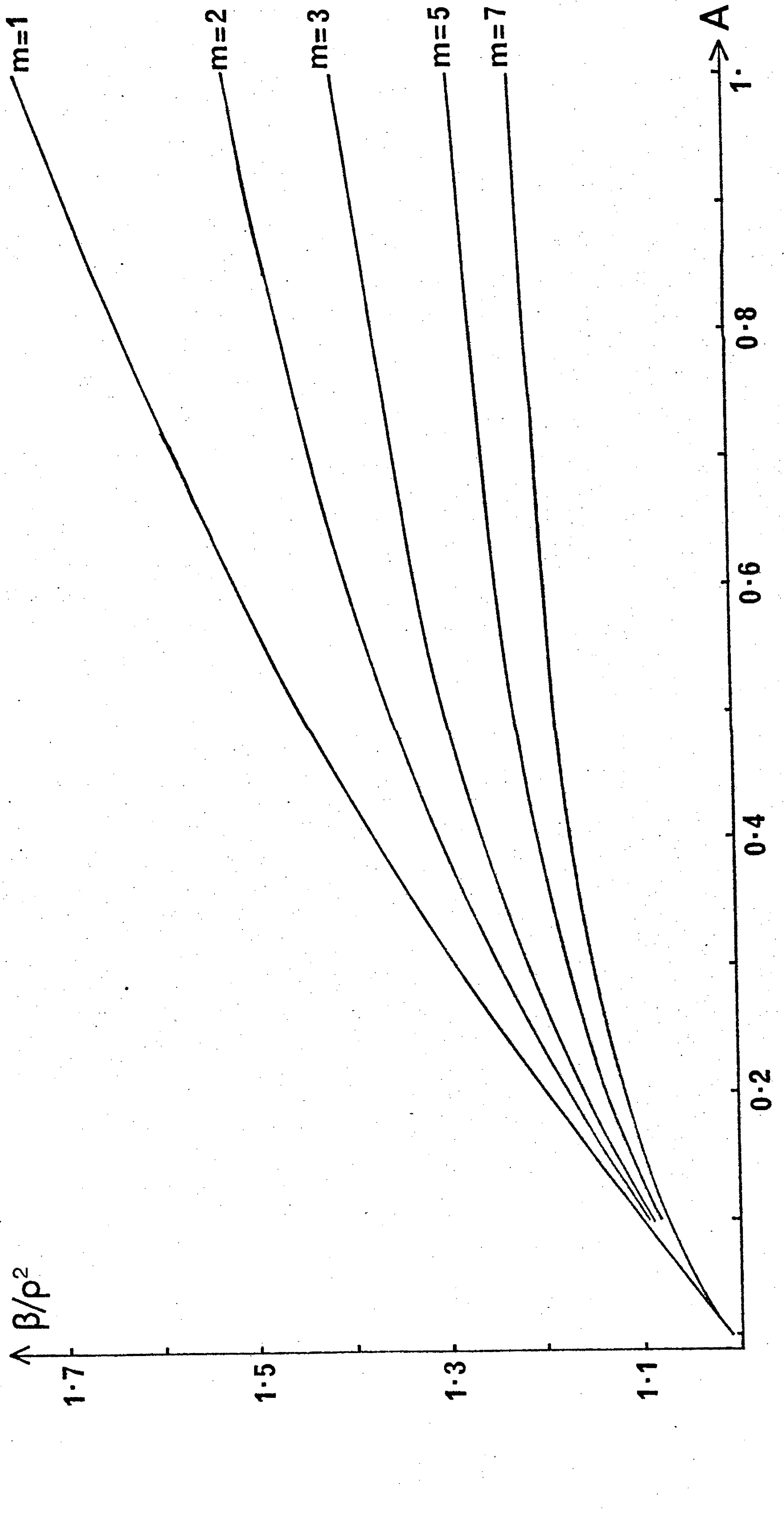


Fig. 5·1

approximate series expansion for $E(\bar{r})$ leads to the approximation,

$$E(\bar{r}) \simeq \rho \ .$$

This is valid for γ large, but we are really interested in the deviation of $E(\bar{r})$ from ρ. This deviation is of smaller order than γ^{-1} and further analysis is required. Returning to the exact expression for $E(\bar{r})$,

$$E(\bar{r}) = \left(2\sigma^2\right)^{\frac{1}{2}}\Gamma(\tfrac{1}{2})^{-1}{}_1F_1[-\tfrac{1}{2};\ \tfrac{1}{2};\ -\gamma] \ .$$

Applying the recurrence relationship for ${}_1F_1[a;\ b;\ z]$ from Abramowitz and Stegun [(1972), eq. 13.4.4], we get

$${}_1F_1[-\tfrac{1}{2};\ \tfrac{1}{2};\ -\gamma] = {}_1F_1[\tfrac{1}{2};\ \tfrac{1}{2};\ -\gamma] + \left(\rho^2/\sigma^2\right)\ {}_1F_1[\tfrac{1}{2};\ \tfrac{3}{2};\ -\gamma] \ .$$

Then using Abramowitz and Stegun [(1972, eqs. 13.6.12 and 13.6.19] we get

$${}_1F_1[-\tfrac{1}{2};\ \tfrac{1}{2};\ -\gamma] = \exp(-\gamma) + (\rho/\sigma)(\pi/2)^{\frac{1}{2}}\mathrm{erf}\left(2^{-\frac{1}{2}}\rho/\sigma\right) \ ,$$

where

$$\mathrm{erf}(x) = (2/\pi)^{\frac{1}{2}} \int_0^x \exp\left(-t^2\right)dt \ ,$$

is the error function. Substituting back into the expression for $E(\bar{r})$ we obtain

$$E(\bar{r}) = \rho\left[\mathrm{erf}\left(2^{-\frac{1}{2}}\rho/\sigma\right)+\left(\rho^2/2\sigma^2\right)^{-\frac{1}{2}}\pi^{-\frac{1}{2}}\exp\left(-\rho^2/2\sigma^2\right)\right] \ .$$

Using

$$\mathrm{erf}(x) = 2\Phi\left(2^{\frac{1}{2}}x\right) - 1 \ ,$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function,

$$E(\bar{r}) = \rho\left[2\Phi(\rho/\sigma)-1+\left(\rho^2/2\sigma^2\right)^{-\frac{1}{2}}\pi^{-\frac{1}{2}}\exp\left(-\rho^2/2\sigma^2\right)\right] \ .$$

Differentiating with respect to ρ gives

$$\partial E(\bar{r})/\partial\rho = 2\Phi(\rho/\sigma) - 1 \ ,$$

and again using the inequality $\mathrm{ARE}(\bar{r},\ \hat{\rho}) \le 1$ where σ^2 is known we derive the inequality

$$\beta/\rho^2 \ge 1 + \left(\sigma^2/\rho^2\right)[\partial E(\bar{r})/\partial\rho]^2\left[\rho^2/\sigma^2 + 1 - \sigma^{-2}[E(\bar{r})]^2\right]^{-1} \ . \qquad (5.8)$$

$A = \sigma^2/\rho^2$	1.0	0.5	0.2	0.1
β/ρ^2	1.73391	1.44796	1.19787	1.09996
R.H.S. of (5.8)	1.729	1.4470	1.19779	1.09995

The right hand side of (5.8) is an excellent approximation for β/ρ^2 indicating that $\mathrm{ARE}(\bar{r}, \hat{\rho})$ is high over the entire range of A considered here.

If we use the approximation to β/ρ^2 given by (5.7),

$$N\ \mathrm{Var}(\hat{\sigma})/\sigma^2 \simeq \tfrac{1}{2}[1 - (m-1)\gamma^{-1} + \ldots] ,$$

$$N\ \mathrm{Var}(\hat{\lambda}) \simeq \gamma + 1$$

and

$$\mathrm{Corr}(\hat{\lambda}, \hat{\sigma}) \simeq [1-\tfrac{1}{2}(m-1)\gamma^{-1}][1-(m-2)\gamma^{-1}]^{\frac{1}{2}} .$$

This agrees with the asymptotic formulae obtained by Johnson (1962) for $m = 1$, although in this case more accurate approximations can be found using (5.8). Notice the high correlation between $\hat{\lambda}$ and $\hat{\sigma}$ for large γ. In fact this high correlation exists for all values of A, at least for $m = 2$ and $m = 3$. This suggests reparametrization of the model. For example it can be shown that $\hat{\rho}$ and $\hat{\sigma}^2$ are only highly correlated for large A, and

$$\mathrm{Corr}(\hat{\rho}, \hat{\sigma}^2) \to 0 \quad \text{as} \quad \gamma \to \infty ,$$

indicating that this is perhaps a better parameterization.

6. Comparisons with Moment Estimators

If σ^2 is known, the moment estimator of λ^2 is given by

$$\lambda_1^2 = N^{-1} \sum_{i=1}^{N} r_i^2/\sigma^2 - m ,$$

and has variance,

$$4N^{-1}[\tfrac{1}{2}m + \rho^2/\sigma^2] .$$

The Cramér-Rao lower bound for the variance of $\hat{\lambda}^2$ is

$$4N^{-1}[\beta/\rho^2 - 1]^{-1} .$$

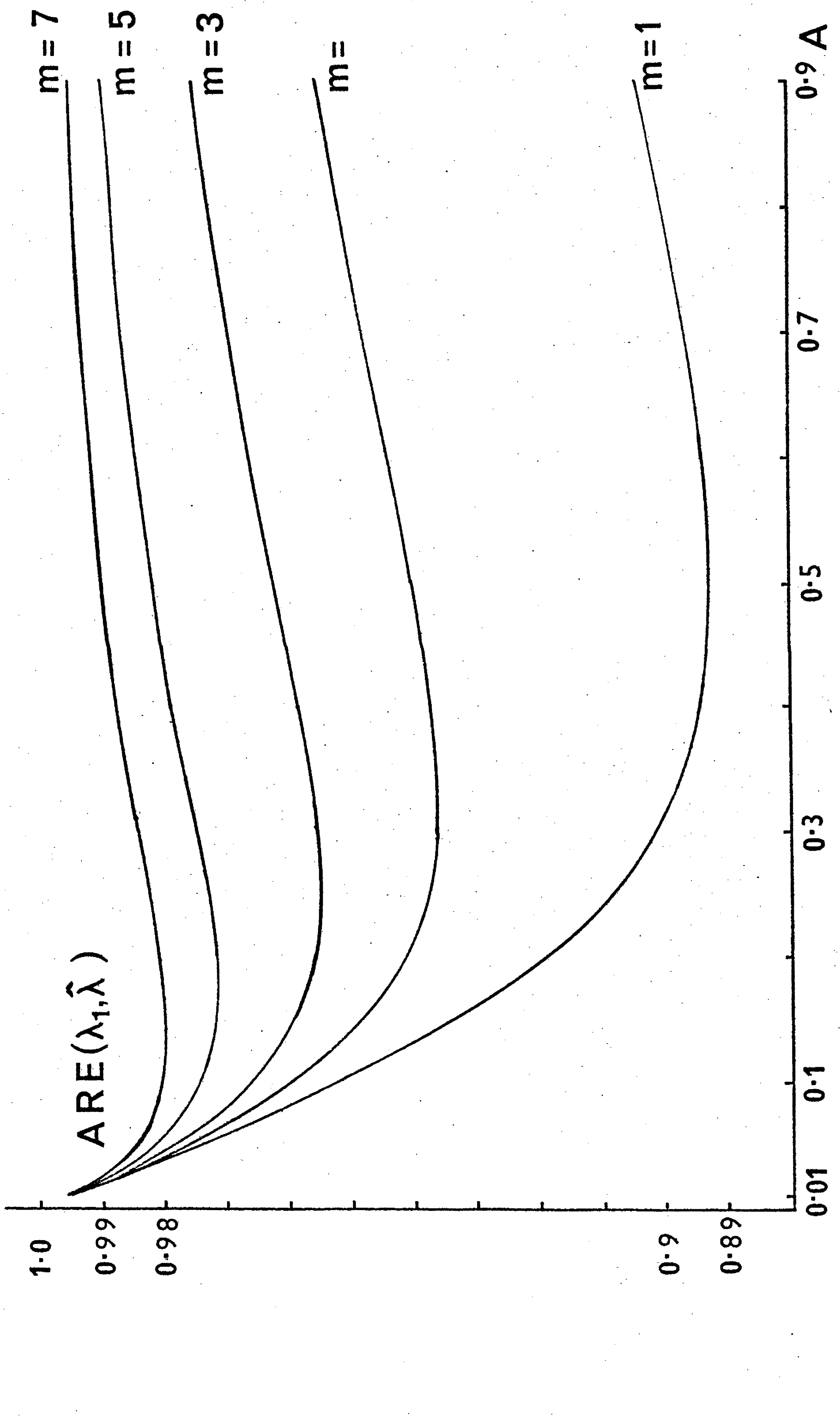

ARE(λ_1,$\hat{\lambda}$)
m = 7
m = 5
m = 3
m =
m=1
A
1·0
0·99
0·98
0·9
0·89
0·01
0·1
0·3
0·5
0·7
0·9

Fig. 5.2

For large γ and $\gamma >> m$,

$$\mathrm{ARE}\left(\lambda_1^2, \hat{\lambda}^2\right) \simeq 1 - \gamma^{-1}/4 + (2m-3)\gamma^{-3}/32 \ \dots \ .$$

In fact $\mathrm{ARE}\left(\lambda_1^2, \hat{\lambda}^2\right)$ is high for the range of A and m considered here. For each m , the minimum occurs in the middle of the A range, and for each A , $\mathrm{ARE}\left(\lambda_1^2, \hat{\lambda}^2\right)$ increases with m ; see Figure 5.2. The only disadvantage of the estimator λ_1^2 is that it is not necessarily positive. To overcome this drawback most authors suggest using the modified estimator $\max\left(\lambda_1^2, 0\right)$. Ifram (1975) discusses this modified estimator and gives a method of finding its bias.

At least for the case $\lambda^2 >> m$ and m small it is apparent that not much improvement on the moment estimator is possible. However $\gamma \approx m$ and $m >> \gamma$ must also be considered.

If σ^2 is unknown, r_i/σ has a non-central chi distribution with both scale and non-centrality parameters unknown. The moment estimator for $\rho = \sigma\left(\lambda^2\right)^{\frac{1}{2}}$ is

$$\rho_2 = \left[\left(N^{-1} \sum_{i=1}^{N} r_i^2\right)^2 (1+\tfrac{1}{2}m) - \tfrac{1}{2}mN^{-1} \sum_{i=1}^{N} r_i^4\right]^{\frac{1}{4}} ,$$

and the case when m is two has been examined in chapter 4, section 10. It tends to efficiency, at least for λ^2 large,

$$\mathrm{ARE}\left(\rho_2, \hat{\rho}\right) \simeq 1 - 3m^2\gamma^{-2}/8 + \dots \ .$$

More work is required to evaluate the moment estimator for λ^2 since it involves an estimator for σ^2 as well as one for ρ . However one would expect the results for large λ^2 to be similar to those for ρ_2 .

7. The Case $m >> \gamma$

As before we can show that

$$1 \le \beta/\rho^2 \le \left(\rho^2/2\sigma^2\right) \frac{\Gamma(\tfrac{1}{2}m+\tfrac{3}{2})}{\Gamma(\tfrac{1}{2}m+1)} \quad {}_1F_1\left[-\tfrac{1}{2};\ \tfrac{1}{2}m+1;\ -\rho^2/2\sigma^2\right] .$$

The asymptotic expression used previously for the confluent hypergeometric function is inappropriate here. Instead we use the approximation valid for $m \gg \gamma$ from Slater [(1960), eq. 4.3.1] to give

$${}_1F_1[-\tfrac{1}{2};\ \tfrac{1}{2}m+1;\ -\gamma] = 1 + \gamma/(m+2) - \gamma^2/2(m+2)(m+4) + O\left(m^{-3}\right) .$$

Luke [(1975), p. 11] gives an expansion for the ratio of two gamma functions $\Gamma(z+a)/\Gamma(z+b)$ in terms of generalized Bernoulli polynomials. Using the expressions for these generalized Bernoulli polynomials from Luke [(1975), p. 505] we derive

$$\Gamma(\tfrac{1}{2}m+\tfrac{3}{2})/\Gamma(\tfrac{1}{2}m+1) = (\tfrac{1}{2}m)^{\frac{1}{2}}\left\{1 + 3m^{-1}/4 - 7m^{-2}/16 + O\left(m^{-3}\right)\right\} .$$

Thus we obtain the bounds

$$1 \le \beta/\rho^2 \le \left(\rho^2/2\sigma^2\right)^{-\frac{1}{2}}(\tfrac{1}{2}m)^{\frac{1}{2}}\left\{1+O\left(m^{-1}\right)\right\} .$$

Neither of these bounds is of any use as an approximation to β/ρ^2 . Since we know that $\Delta > 0$, we can use this inequality to show that

$$\beta/\rho^2 > 1 + 2m^{-1} .$$

The difficulty in finding an approximation to β/ρ^2 arises because for $m \gg \gamma$, the range of z which contributes most to the integral defining β , is z small and for z small the approximation $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z) \approx 1$ is inappropriate. We require an approximation to $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$ which is accurate for z small. Dwivedi and Pandey (1975) showed that $z^{-1}I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$ is a monotonically decreasing function bounded above by m^{-1} . Consequently

$$I_{\frac{1}{2}m,\frac{1}{2}m-1}(z) < zm^{-1} .$$

Substitution of this inequality into the integral defining β yields

$$\beta/\rho^2 < 1 + 2m^{-1} + 2\gamma m^{-1} \ldots .$$

Further we can show that

$$I_{\frac{1}{2}m,\frac{1}{2}m-1}(z) \simeq zm^{-1}\left\{1 - z^2/m(m+2) + 2z^4/m^2(m+2)(m+4) \ \ldots\right\}$$

for $m >> z$, but this approximation is not accurate enough since it leads to the approximation

$$\beta/\rho^2 \simeq 1 + 2m^{-1} - 8\gamma m^{-2} - 16\gamma^3 m^{-2} ,$$

but we know that $\beta/\rho^2 > 1 + 2m^{-1}$ for Δ to be positive.

Previously we used $\overline{r}$ to obtain an approximation to β/ρ^2 , since $\overline{r}$ has expectation close to ρ when γ is large and much greater than m . However when m is much greater than γ ,

$$E(\overline{r}) \simeq \left(m\sigma^2\right)^{\frac{1}{2}}\{1 - 4m^{-1} + \gamma m^{-1} + \ldots\} ,$$

and this does not approximately equal ρ for m large. $\overline{r}$ is not a good estimator of ρ when m is large, and it is quite difficult to find an approximation for β/ρ^2 in this case.

Using the bound $\beta/\rho^2 > 1 + 2m^{-1}$ we can conclude that if σ^2 is known,

$$\text{ARE}\left(\lambda_1^2, \hat{\lambda}\right) > 1 - 4\gamma m^{-1} ,$$

where $m >> \gamma$, which indicates that the asymptotic relative efficiency of the moment estimator is high for large m .

Similar problems occur when attempting to find an approximation for β/ρ^2 in the case where $\gamma \approx m$, and this case will not be considered here. From Figure 5.2 we would expect the asymptotic relative efficiency of the moment estimator λ_1^2 with respect to $\hat{\lambda}^2$ to remain fairly high.

8. m-Dimensional Hyperspherical Structural Model

We can also consider estimating the parameters of an m-dimensional hyperspherical relationship. This may only be of practical relevance for $m \leq 3$. The model is similar to (5.1) and is

$$X_j - \xi_j = \rho \sin \varphi_1 \sin \varphi_2 \ldots \sin \varphi_{j-1} \cos \varphi_j + \varepsilon_j , \quad j = 1, 2, \ldots, m-1 ,$$

and

$$X_m - \xi_m = \rho \sin \varphi_1 \sin \varphi_2 \ldots \sin \varphi_{m-1} + \varepsilon_m ,$$

where in addition to the assumptions associated with (5.1), $\left(\xi_1, \xi_2, \ldots, \xi_m\right)$ is the centre of the hypersphere. If $\left(\varphi_1, \ldots, \varphi_{m-1}\right)$ is uniform over the whole sphere, then it has probability density function given by

$$g(\varphi_1, \ldots, \varphi_{m-1}) = \frac{\Gamma(\frac{1}{2}m)}{2\pi^{\frac{1}{2}m}} \sin^{m-2}\varphi_1 \sin^{m-3}\varphi_2 \ldots \sin \varphi_{m-2} \qquad (5.9)$$

for $m \geq 2$. To integrate $f(X_1, \ldots, X_m)$ with respect to $\varphi_1, \ldots, \varphi_{m-1}$ a change of variable is made. Let

$$X_j - \xi_j = r \sin \psi_1 \sin \psi_2 \ldots \sin \psi_{j-1} \cos \psi_j , \quad j = 1, 2, \ldots, m-1 ,$$

and

$$X_m - \xi_m = r \sin \psi_1 \sin \psi_2 \ldots \sin \psi_{m-1} .$$

Then r/σ has a non-central $\chi_{(m)}$ distribution with non-centrality parameter ρ^2/σ^2 . From Downs (1966) $(\psi_1, \psi_2, \ldots, \psi_{m-1}|r)$ has an m-dimensional von Mises distribution over the hypersphere. The actual evaluation of $f(X_1, X_2, \ldots, X_m)$ by integration is achieved by rotating the $(\psi_1, \ldots, \psi_{m-1})$ axes. Then

$$f(X_{i1}, X_{i2}, \ldots, X_{im}) = \Gamma(\tfrac{1}{2}m)2^{-1}\pi^{-\frac{1}{2}m}r_i^{-\frac{1}{2}m+1}\rho^{-\frac{1}{2}m+1}\sigma^{-2}$$
$$\exp\left[-\tfrac{1}{2}\sigma^{-2}\left\{\sum_{j=1}^{m} (X_{ij}-\xi_j)^2+\rho^2\right\}\right] I_{\frac{1}{2}m-1}\left\{r_i\rho/\sigma^2\right\} .$$

Then the maximum likelihood equations for ξ_j are the same as (4.10) with $I_{1,0}\left\{r_i\rho/\sigma^2\right\}$ replaced by $I_{\frac{1}{2}m,\frac{1}{2}m-1}(r\rho/\sigma^2)$ and $\hat{\rho}$ and $\hat{\sigma}^2$ given by the solutions of (5.3) and (5.4). Further it can be shown that

$$\mathrm{Var}(\hat{\xi}) = \sigma^2 N^{-1}\delta_1^{-1} ,$$

where

$$\delta_1 = 1 - m^{-1}(\rho^2/\sigma^2)(1-\delta) ,$$

and

$$\delta = E\left[I^2_{\frac{1}{2}m,\frac{1}{2}m-1}(r\rho/\sigma^2)\right] ,$$

that is

$$\delta = \sigma^{-2}\rho^{-\frac{1}{2}m+1} \int_0^\infty r^{\frac{1}{2}m} e^{-r^2/2\sigma^2} I_{\frac{1}{2}m,\frac{1}{2}m-1}(r\rho/\sigma^2) I_{\frac{1}{2}m}(r\rho/\sigma^2)\,dr . \qquad (5.10)$$

δ can be dealt with in a similar manner as for the case $m = 2$. The inequality equivalent to (4.24b) is

$$1 - \tfrac{1}{2}(m-1)\gamma^{-1} + (m-1)(3m-8)\gamma^{-2}/16 + \ldots \le \delta \le 1 - \tfrac{1}{2}(m-\tfrac{5}{4})\gamma^{-1}$$
$$+ \left(32m^2-124m+109\right)\gamma^{-2}/128 - \left(128m^3-1040m^2+2602m-1965\right)\gamma^{-3}/1024 \;\ldots \quad (5.11)$$

The lower bound does not hold for $m = 1$ and a more precise argument is needed. As before δ can be numerically integrated using Gregory's formula, and this was done simultaneously with the calculations for β/ρ^2 (for numerical values see Table VII and Figure 5.3).

A/m	1	2	3	5	7
1.0	0.55040	0.34515$^-$	0.25426	0.16766	0.12535$^-$
0.9	0.58398	0.37130	0.27551	0.18308	0.13743
0.8	0.62184	0.40178	0.30067	0.20163	0.15210
0.7	0.66476	0.43772	0.33091	0.22438	0.17029
0.6	0.71355	0.48070	0.36795$^-$	0.25295$^-$	0.19343
0.5	0.76898	0.53290	0.41432	0.28988	0.22388
0.4	0.83121	0.59731	0.47397	0.33948	0.26573
0.3	0.89826	0.67780	0.55317	0.40952	0.32686
0.2	0.96154	0.77788	0.66179	0.51548	0.42445$^-$
0.1	0.99759	0.89302	0.81192	0.69083	0.60317
0.09	0.99868	0.90447	0.82938	0.71445$^-$	0.62930
0.08	0.99937	0.91577	0.84721	0.73948	0.65764
0.07	0.99976	0.92689	0.86539	0.76603	0.68844
0.06	0.99993	0.93781	0.88388	0.79416	0.72199
0.05	0.99999	0.94854	0.90265	0.82395$^-$	0.75860
0.04	0.99999_9	0.95910	0.92167	0.85547	0.79859
0.03	1.00000	0.96951	0.94093	0.88878	0.84233
0.02	1.00000	0.97979	0.96041	0.92393	0.89021
0.01	1.00000	0.98995$^-$	0.98010	0.96099	0.94263

δ

TABLE VII

In this model, the case $m = 1$ corresponds to a mixture of two normal distributions with means at $\xi + \rho$ and $\xi - \rho$. The uniform distribution over the hypersphere corresponds to having a mixing probability of $\frac{1}{2}$ for

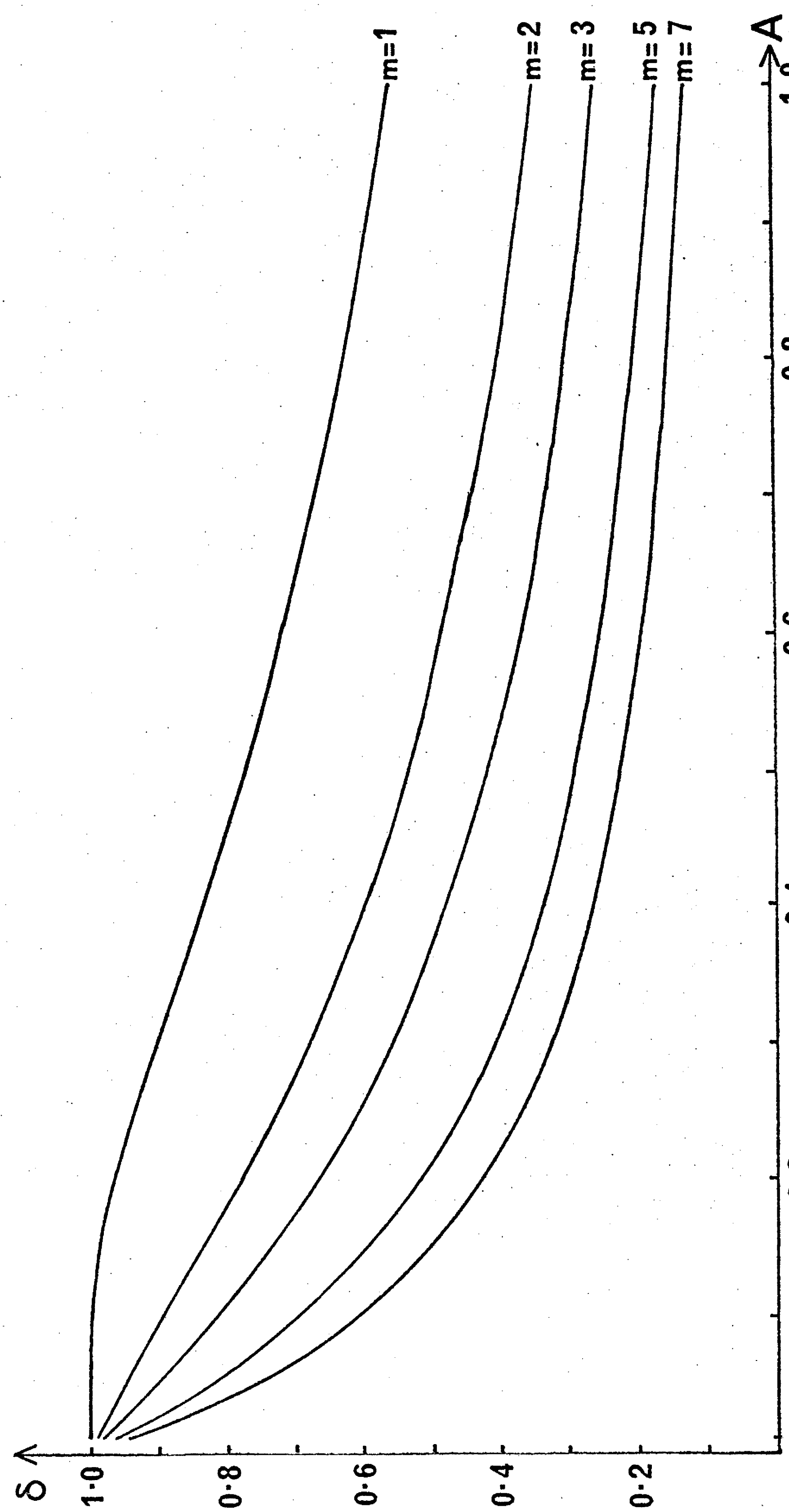


Fig. 5·3

each distribution.

The more accurate bound for δ when $m = 1$ can be derived as follows:

$$E\left[I_{\frac{1}{2},-\frac{1}{2}}\left(r\rho/\sigma^2\right)\right]^2 \le \delta \le E\left[I_{\frac{1}{2},-\frac{1}{2}}\left(r\rho/\sigma^2\right)\right] ,$$

where

$$E\left[I_{\frac{1}{2},-\frac{1}{2}}\left(r\rho/\sigma^2\right)\right] = \gamma^{\frac{1}{2}}e^{-\gamma}[\Gamma(\tfrac{3}{2})]^{-1}{}_1F_1[1;\ \tfrac{3}{2};\ \gamma] .$$

Use of the approximation to the confluent hypergeometric function from Slater [(1960), eq. 4.1.2] would give

$${}_1F_1[1;\ \tfrac{3}{2};\ \gamma] \simeq \Gamma(\tfrac{3}{2})\gamma^{-\frac{1}{2}}e^{\gamma} ,$$

which leads to the approximation $\delta \simeq 1$. But from Abramowitz and Stegun [(1972), eq. 13.6.19]

$${}_1F_1[1;\ \tfrac{3}{2};\ \gamma] = e^{-\gamma}{}_1F_1[\tfrac{1}{2};\ \tfrac{3}{2};\ -\gamma]$$

$$= e^{-\gamma}\pi^{\frac{1}{2}}2^{-1}\gamma^{-\frac{1}{2}}\mathrm{erf}\left(\gamma^{\frac{1}{2}}\right) .$$

Since $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{1}{2}\pi^{\frac{1}{2}}$,

$$E\left[I_{-\frac{1}{2},\frac{1}{2}}\left(r\rho/\sigma^2\right)\right] = \mathrm{erf}\left(2^{-\frac{1}{2}}\rho/\sigma\right) .$$

Using the equality $\mathrm{erf}(x) = 2\Phi\left(2^{\frac{1}{2}}x\right) - 1$, bounds for δ are

$$\{2\Phi(\rho/\sigma)-1\}^2 \le \delta \le \{2\Phi(\rho/\sigma)-1\} .$$

These bounds hold for all ρ/σ but are only accurate approximations to δ for γ large:

$A = \sigma^2/\rho^2$	1.0	0.5	0.2	0.1	0.05
δ	0.550	0.769	0.962	0.9976	0.99999
$2\Phi(\rho/\sigma) - 1$	0.683	0.843	0.975	0.9984	0.99999
$[2\Phi(\rho/\sigma)-1]^2$	0.466	0.710	0.950	0.9969	0.99998

We are also interested in $\overline{X}_j$ as an estimator of ξ_j . It can be shown that

$$\mathrm{Var}\left(\overline{X}_j\right) = \sigma^2N^{-1}\left\{1+m^{-1}\left(\rho^2/\sigma^2\right)\right\} ,$$

and for γ large we can show that

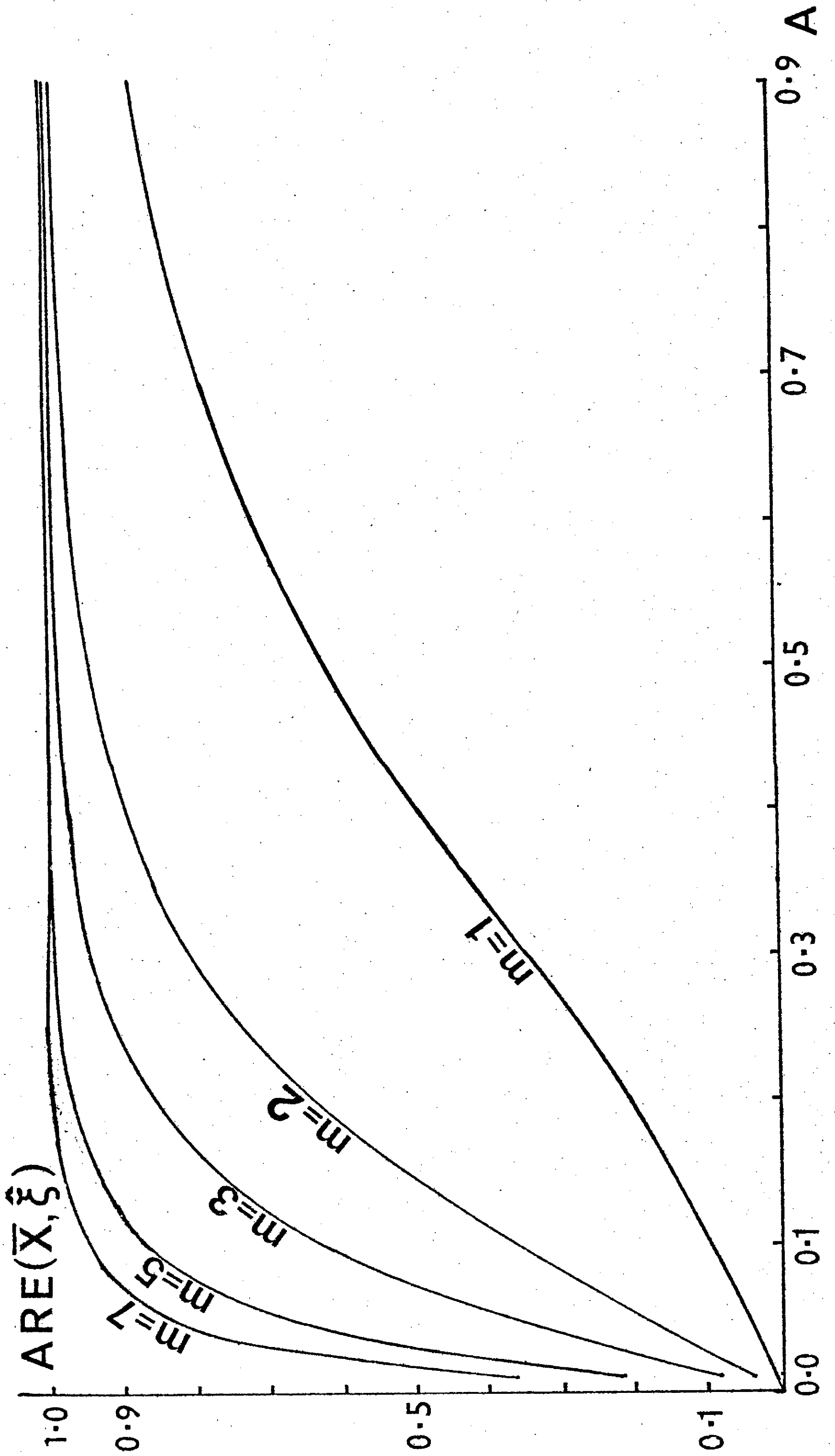


Fig. 5.4

$$\operatorname{Var}\left(\hat{\xi}_j\right) \simeq \sigma^2 N^{-1}\left\{m - m(m-1)(3m-8)\gamma^{-1}/8 \ldots\right\} .$$

Notice that $3m - 8$ is positive for $m \geq 3$, so as γ increases $N \operatorname{Var}\left(\hat{\xi}_j\right)/\sigma^2$ increases to m if $m \geq 3$, only decreasing to m in the case $m = 2$. For $\gamma >> m$, and γ large,

$$\operatorname{ARE}\left(\bar{X}_j, \hat{\xi}_j\right) \simeq \tfrac{1}{2}m^2\gamma^{-1} .$$

As γ tends to infinity, $\operatorname{ARE}\left(\bar{X}_j, \hat{\xi}_j\right)$ tends to zero. As before, the actual value of $\operatorname{ARE}\left(\bar{X}_j, \hat{\xi}_j\right)$ can be calculated for those values of $A = \lambda^{-2}$ for which δ has been evaluated numerically (see Figure 5.4).

9. Comments

We could continue further and examine the effects on the maximum likelihood estimators of replacing $I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)$ by an approximation. We could also consider extending the estimators ρ_3, ρ_4 and σ_4^2 to the case of general m. However we would expect results similar to those for $m = 2$. For example if $\mathbf{x}_i = \left(x_{i1}, x_{i2}, \ldots, x_{im}\right)$, then

$$D = N^{-1}(N-1)^{-1} \sum_{i<j}\sum d_{ij}^2$$

where

$$d_{ij}^2 = \sum_{k=1}^{m} \left\{\left(x_{ik}-\xi_k\right)-\left(x_{jk}-\xi_k\right)\right\}^2 ,$$

and

$$E(D) = \rho^2 + m\sigma^2 .$$

Then if σ^2 is known,

$$\rho_3 = \left(D-m\sigma^2\right)^{\frac{1}{2}} .$$

Also

$$S = N^{-1}(N-1)^{-1} \sum_{i<j}\sum d_{ij}^4 ,$$

then

$$E(S) = 2m(m+2)\sigma^4 + 4(m+2)\rho^2\sigma^2 + 2(m+1)m^{-1}\rho^4 .$$

Hence

$$\rho_4 = \left[(m+2)D^2 - m/2\ S\right]^{\frac{1}{4}} ,$$

and

$$\sigma_4^2 = m^{-1}\left(D-\rho_4^2\right) .$$

To calculate $E(S)$ we require $E\left[\left(X_{ik}-\xi_k\right)^2\left(X_{il}-\xi_l\right)^2\right]$, and due to the symmetry of the system this involves an integral like

$$\int \cdots \int \sin^2\varphi_1 \cos^2\varphi_1 \cos^2\varphi_2 g\left(\varphi_1, \ldots, \varphi_{m-1}\right) d\varphi_1 \ldots d\varphi_{m-1} ,$$

where $k = 1$ and $l = 2$, and $g\left(\varphi_1, \ldots, \varphi_{m-1}\right)$ is defined by (5.9). This integral is equivalent to

$$\frac{\Gamma(\frac{1}{2}m)}{2\pi^{\frac{1}{2}m}} \int_0^\pi \sin^m\varphi_1 \cos^2\varphi_1 d\varphi_1 \int_0^\pi \sin^{m-3}\varphi_2 \cos^2\varphi_2 d\varphi_2$$
$$\times \int \cdots \int g\left(\varphi_3, \varphi_4, \ldots, \varphi_{m-1}\right) d\varphi_3 \ldots d\varphi_{m-1} ,$$

for $m \geq 3$. The right hand multiple integral is just $2\pi^{\frac{1}{2}m-1}/\Gamma(\frac{1}{2}m-1)$, and the remaining double integral can be calculated using the equality $\sin^2 z = 1 - \cos^2 z$ and equation 2.511 from Gradshteyn and Ryzhik (1965). For $m = 2$, we simply needed

$$(2\pi)^{-1} \int_0^{2\pi} \sin^2\varphi_1 \cos^2\varphi_1 d\varphi_1 = \tfrac{1}{8} .$$

Using the above results we can show

$$E\left(\rho_3\right) \simeq \rho + O\left(N^{-1}\right) ,$$

and

$$\operatorname{Var}\left(\rho_3\right) \simeq \sigma^2 N^{-1}\left(1 + \frac{m}{2}\sigma^2/\rho^2\right) .$$

ρ_3 is an efficient estimator of ρ if $\gamma >> m$, as expected. The variance for ρ_4 is extremely tedious to find analytically since it requires terms like $E(S^2)$ and $E(D^2S)$. The special case $m = 2$ has already been considered in chapter 2. There the estimators were found to be unbiased asymptotically and efficient and we would not expect that the results would differ markedly in the general case.

D and S both contain $mN(N-1)/2$ terms and hence if N is large a considerable amount of work is required to evaluate the estimators. If σ^2/ρ^2 is small enough then it would probably be easier to estimate the parameters by using one of the modified maximum likelihood procedures.

CHAPTER 6

STONE CIRCLES AND SIMULATION STUDIES

1. Introduction

As yet not many practical aspects of fitting circles to data have been considered, nor illustrations of the various methods of parameter estimation using actual or simulated data made. One source of actual data is the carefully measured stone ring data of Alexander Thom and in the next few sections fitting of this data will be discussed, as well as some discussion of some of the methods used by Thom and others to estimate the circular parameters. One of the important practical aspects of data fitting that has not been discussed here is the problem of using asymptotic properties of estimators for finite sample sizes. In most of this work we have been concerned with asymptotic properties of estimators. We have assumed that the theoretical sample size N is large enough for asymptotic theory to hold, without really defining how large N should be. For example we might say that the bias term is of order N^{-1}, but this may not necessarily be insignificant if a small sample is involved. To make a thorough and proper examination of the effect of smaller order terms, these terms should be examined in more detail. However the amount of work required to do this could be prohibitive. For example, for maximum likelihood estimation we would be required to include all the third derivative terms, and for four parameters that involves a large number of terms. An alternative is to carry out some simulation studies in which we set up a set of circles with the same known centre, radius and variance each with N points, and independently estimate the circular parameters. Then we can examine the properties of these estimators in a sample of size N, by estimating the means variances, and perhaps skewness and kurtosis. The results of such a simulation study are given in section 6.

2. Stone Circles

In the British Isles there are a large number of megalithic stone structures for which the stones appear to be on, or close to, the circumference of a circle or ellipse. There is a considerable amount of literature concerned with the reasons for the erection of these structures. One question considered is the possible astronomical significance of various alignments of the stones. Another is Alexander Thom's theory that many of the dimensions of the megalithic sites can be expressed as simple multiples of a megalithic yard of 0.83 m (or 2.72 ft). For a general survey and bibliography of the literature on stone sites see for example Thom (1967), Hadingham (1975) and Burl (1976). For discussion of the megalithic yard see Thom (1955), Broadbent (1955), Kendall (1974) and Freeman (1976). Before any discussion of the theoretical significance of the dimensions of these stone circles it is necessary to estimate the circular parameters accurately and this is considered here.

From Burl (1976) we can see that 963 stone circles have so far been discovered in the British Isles, of which 660 have "known" diameters. "Known" in this context means that the structure has been at least roughly surveyed and the diameter estimated. Most of these have diameters of between 10 ft and 100 ft, although it is clear that at least initially many diameter estimates were rounded off to the nearest 5 or 10 ft (see Burl [(1976), p. 40]). Thom was one of the few who attempted accurate and systematic surveying of the stones and has the most reliable data. Here we confine ourselves to examining circular stone rings; leaving aside egg-shaped and elliptical structures. Also we will ignore such problems as fallen stones, re-erected stones and the problem as to where on the stone the (X, Y) observation should be taken. Thom took his observations from the centre of the stones. The difficulty arises here because the stones are not uniform in size or shape, so that measuring from the wrong point on the stones will deform the circle.

3. Thom's "Least Square" Method

Assuming that an initial survey has been made, the next step is to construct the circle, or other shape formed by the stones. One of the methods used seems to be a theoretical construction with central pegs and loops of rope. For example Angell (1977) gives a polygonal method for constructing the outlines of many non-elliptical rings. Thom (1955) uses what he called the "least squares" estimation procedure.

An initial circle is constructed with centre 0 and radius R. The perpendicular distance of the stone from this circle is denoted by $|\Delta_i|$, where Δ_i is regarded as positive if stone i is outside the circle, negative otherwise. Thom then assumes that the size of the outline is rescaled so that R becomes $R + \delta R$ and the position of the centre is perturbed a small amount to (ξ, η) . The modified perpendicular distance Δ_i' is given by

$$\Delta_i' = \Delta_i - \delta R - \xi \cos \theta_i - \eta \sin \theta_i ,$$

where θ_i is the angle between the x-axis and the stone i . If

$$\boldsymbol{\beta}' = (\delta R, \xi, \eta)$$

then

$$\boldsymbol{\beta}_T = (X'X)^{-1}X'Y ,$$

where

$$(X'X) = \begin{bmatrix} N & \sum \cos \theta_i & \sum \sin \theta_i \\ \sum \cos \theta_i & \sum \cos^2\theta_i & \sum \cos \theta_i \sin \theta_i \\ \sum \sin \theta_i & \sum \cos \theta_i \sin \theta_i & \sum \sin^2\theta_i \end{bmatrix} ,$$

and

$$(X'Y)' = \left\{\sum \Delta_i, \sum \Delta_i \cos \theta_i, \sum \Delta_i \sin \theta_i\right\} .$$

The underlying error structure assumed is that Δ_i is normally distributed with zero mean. Essentially Thom assumes that the errors are in the radial direction with an additional independent error around the circumference of the circle. This does not allow for any error in erecting the stones. Thom's estimation procedure is not least squares as claimed but more resembles a first iteration towards a least squares type solution. However when σ^2 is small enough Thom's procedure would be an accurate approximation to the least squares solution, and also a good approximation to the maximum likelihood solution of the model described in chapter 4, section 4.

Thom's analysis becomes more complicated for a composite ring such as Avebury. In Thom, Thom and Foord (1976) they estimate more than four circular arcs, but presuppose that the centres of three of these arcs fall on the vertices of a right-angled triangle with side ratios 3:4:5 . We will not consider this estimation further, except to remark that it has come in for some criticism, see Angell and Barber (1977) and Freeman (1977).

If there are stones around the entire circumference Thom [(1967), p. 35] gave an alternative method which can be used as an accurate approximation to his least squares method. Assuming that a preliminary circle has been fitted to the data, the circle is divided into four quadrants, labelled for convenience ne, se, sw and nw . The mean of Δ_i for each quadrant is found, δ_{ne}, δ_{se}, δ_{sw} and δ_{nw} respectively. The required radius is the initial radius plus one quarter of the sum of these means and the required centre is the initial centre moved north-east by $\frac{1}{2}(\delta_{ne}-\delta_{sw})$ and northwest by $\frac{1}{2}(\delta_{nw}-\delta_{se})$.

Thom and Thom (1973) used the above approximation in estimating the diameter of the ring of Brogar. There they also set out more clearly the hypothesis that the stones were placed at equal angles apart along the circular arc. This does seem to be borne out by the data. We might then try to model this distribution by say $\tau_i \overset{d}{\sim} iM^{-1} + \varphi + t_i$ where M is the number of stones initially laid out, φ is the angular deviation of the first stone from the x-axis, and t_i is a zero mean symmetric random variable. Then as before we could assume normal errors in both the X and Y directions. Now the distribution of α_i given r_i and τ_i where

$$X_i - \xi = r_i \cos \alpha_i \quad \text{and} \quad Y_i - \eta = r_i \sin \alpha_i$$

is von Mises with probability distribution function

$$f(\alpha_i | r_i, \tau_i) = \left[2\pi I_0\left(r_i\rho/\sigma^2\right)\right]^{-1} \exp\left\{-r_i\rho\sigma^{-2}\cos\left(\alpha_i-\tau_i\right)\right\} .$$

Unless $\tau_i \overset{d}{\sim} R[0, 2\pi)$, the distribution of α_i depends on r_i, ρ and σ^2 . If ρ/σ is large, the probability distribution function of α_i is extremely peaked around τ_i , and if ρ/σ is sufficiently large the dispersion of α_i around τ_i can probably be ignored, and we can regard the angular variate as being independent of the radial variate.

4. AML1 and Angell and Barber's Circle Fitting Algorithm

In addition to the stone circle data, there are two sets of data simulated by Chan (1965). For the first $\tau_i \overset{d}{\sim} R[0,\ 2\pi)$ as we have previously assumed, but for the second $\tau_i \overset{d}{\sim} R[0,\ \pi/2)$, for which Chan's estimators are not even claimed to be consistent. These data sets will be referred to as Chan A and Chan B, respectively. Chan's estimators will be given a subscript c . In addition we have a further simulated data set with $\tau_i \overset{d}{\sim} R[0,\ 2\pi)$ which will be referred to as set C . In the next section we will consider the data of Avebury from Thom *et al* (1976) and of Brogar from Thom and Thom (1973). All this data can be found in Appendix B.

To find the solutions for Chan's equations,

$$\sum_{i=1}^{N} \left(X_i-\xi\right)\left(1 - \bar{r}/r_i\right) = 0 \ ,$$

and

$$\sum_{i=1}^{N} \left(Y_i-\eta\right)\left(1 - \bar{r}/r_i\right) = 0 \ ,$$

we can use the Newton-Raphson iterative technique for non-linear simultaneous equations, see for example Whittaker and Robinson [(1932), p. 90]. This is quite easy to program and converges very rapidly even on a small computer. Notice that there appears to be a slight error on page 54 of Chan's paper. $(pt-gs)$ should be $(pt+gs)$ in the equations for $\xi^{(k+1)}$ and $\eta^{(k+1)}$. For ρ and σ^2 Chan uses the estimators

$$\rho_c^2 = \left[\left(N^{-1} \sum_{i=1}^{N} r_i\right)^2 - \sigma_c^2\right]^{\frac{1}{2}} ,$$

and

$$\sigma_c^2 = (N-3)^{-1} \sum_{i=1}^{N} \left(r_i-\bar{r}\right)^2 \ .$$

For data sets A and C , Chan's estimators for ξ and η are approximately maximum likelihood, since $z_i = r_i\rho/\sigma^2$ is large enough for the approximation $I_{1,0}\left(z_i\right) \simeq 1$ to be nearly true for all i . But ρ_c should be replaced by $\bar{r}$ and σ_c^2 by

$$\sigma^2 = N^{-1} \sum_{i=1}^{N} \left(r_i - \bar{r}\right)^2 .$$

These latter estimators will be referred to as approximate maximum likelihood 1 (AML1).

Set A. $N = 12$

	ξ	η	ρ	σ^2
Real Values	0	0	10	1
Chan's Est.	-0.7986	0.1045	9.6435	0.4980
AML1	-0.7986	0.1045	9.6695	0.3735
ρ_3 $\left(\sigma^2 = 1\right)$	-	-	9.8885	-
ρ_4, σ_4^2	-	-	10.0595	-0.7055*
A & B	-0.7534	0.1534	9.6790	0.3976
A & B (mod)	-0.7534	0.1534	9.6596	0.3764

σ_4^2 is negative which is the main disadvantage of such moment based estimators. A & B represent estimators based on Angell and Barber's (1977) circle-fitting algorithm. These estimators are surprisingly accurate. Angell and Barber considered the relationship

$$H(x, y) = x^2 + y^2 + fx + gy + c ,$$

and

$$M = \sum_{i=1}^{N} \left\{H^2\left(x_i, y_i\right)\right\} - N^{-1}\left\{\sum_{i=1}^{N} H\left(x_i, y_i\right)\right\}^2 ,$$

then minimized M with respect to f and g . c is obtained from

$$c_a = -N^{-1}\left[\sum x_i^2 + \sum y_i^2 + f_a \sum x_i + g_a \sum y_i\right] .$$

This is just the minimization of

$$N^{-1}\left\{\sum r_i^4 - N^{-1}\left(\sum r_i^2\right)^2\right\} ,$$

with respect to 2ξ and 2η . Therefore

$$\xi_a = -f_a/2 , \quad \eta_a = -g_a/2 ,$$

$$\rho_a = \left\{\left(f_a^2 + g_a^2\right)/4 - c_a\right\}^{\frac{1}{2}} ,$$

and we will let

$$\sigma_a^2 = \tfrac{1}{2}\left[\left\{(N-1)^{-1}M+\rho_a^4\right\}^{\frac{1}{2}}-\rho_a^2\right] ,$$

the latter since

$$E(M) = 4(N-1)\sigma^2\left(\rho^2+\sigma^2\right) .$$

Using Neyman and Scott's (1948) method of finding the variance of consistent estimators

$$\mathrm{Var}\left(\xi_a\right) = 2\sigma^2N^{-1}\left(1 + 5\sigma^2/\rho^2\right)$$

if $\tau_i \overset{d}{\sim} R[0,\ 2\pi)$. For ρ^2/σ^2 large, this compares favourably with the maximum likelihood estimator which has variance

$$\mathrm{Var}(\hat{\xi}) = 2\sigma^2N^{-1}\left(1 + \tfrac{1}{2}\sigma^2/\rho^2\right) ,$$

and so

$$\mathrm{ARE}\left(\xi_a,\ \hat{\xi}\right) = 1 - \tfrac{9}{2}\sigma^2/\rho^2 .$$

If $\sigma^2/\rho^2 = 0.01$, $\mathrm{ARE}\left(\xi_a,\ \hat{\xi}\right) = 0.95^-$. Surprisingly this circle fitting algorithm is quite an efficient estimator of the centre of the circle.

If we make the appropriate substitutions we can see that

$$\rho_a = \left\{N^{-1}\sum_{i=1}^{N} r_i^2\right\}^{\frac{1}{2}} ,$$

and this is independent of the distribution of τ_i . If ξ and η are known

$$E\left(\rho_a\right) = \rho\left(1 + 2\sigma^2/\rho^2\right)^{\frac{1}{2}} + O\left(N^{-1}\right).$$

This is inconsistent but if σ^2/ρ^2 is small $E\left(\rho_a\right)$ is approximately equal to ρ . Now

$$\mathrm{Var}\left(\rho_a\right) = \sigma^2N^{-1}\left(1 + \sigma^2/\rho^2\right)\left(1 + 2\sigma^2/\rho^2\right)^{-1} .$$

If σ^2 is known then the asymptotic efficiency of ρ_a with respect to $\hat{\rho}$ is close to one. However for this case we already have an efficient estimator of ρ_a, namely $\bar{r}$ and if ξ and η are known,

$$E(\bar{r}) \approx \rho\left(1 + \tfrac{1}{2}\sigma^2/\rho^2\right) ,$$

if σ^2/ρ^2 is small, whereas $\left(1 + 2\sigma^2/\rho^2\right)^{\frac{1}{2}} \simeq 1 + \sigma^2/\rho^2$. So the bias of $\bar{r}$ is less than that of ρ_a . We would expect that if ξ and η are not known the substitution of ξ_a and η_a into $\bar{r}$ would produce a better estimator than ρ_a . This together with

$$\sigma_a^{*2} = N^{-1} \sum_{i=1}^{N} \left(r_i^2 - \rho_a^{*2}\right)$$

constitute the modified Angell and Barber estimators. For data set A, ρ_a underestimates ρ and thus ρ_a^* underestimates ρ even further. However in general ρ_a^* ought to be less biased, and in the simulation studies in section 7 we show that ρ_a^* is less biased in general, than ρ_a .

σ_a^2 is biased but the bias is small if σ^2/ρ^2 is small. If σ^2/ρ^2 is small, then approximately

$$E\left(\sigma_a^2\right) = \sigma^2\left(1 + 2\sigma^2/\rho^2\right) .$$

However later we will see that σ_a^2 has some rather undesirable properties in small and medium sized samples, especially if σ^2/ρ^2 is small.

Note that the estimators of ξ in particular appear to be quite biased. This is not the case, in fact for $N = 12$ the small sample variance of $\hat{\xi}$ is around 0.2 (by simulation) and the results for set A are within 95% confidence limits based on this variance.

SET C $N = 20$

	ξ	η	ρ	σ^2
Real Values	0	0	10	1
Chan's Est.	-0.0703	-0.0246	10.283	0.9365
AML1	-0.0703	-0.0246	10.3281	0.7960
ρ_3 $(\sigma^2 = 1)$	-	-	10.1596	-
ρ_4, σ_4^2	-	-	10.1926	0.6642
A & B	-0.016	0.005	10.37	0.8275
A & B (mod)	-0.016	0.005	10.33	0.7976
AML2	-0.0871	0.0122	10.279	0.7967

AML2 is "approximate maximum likelihood 2 " and is obtained by letting $I_{1,0}(z) \simeq 1 - \frac{1}{2}z$ in the maximum likelihood estimating equations. Again the non-linear Newton-Raphson iterative technique is used to find the solution. The estimators are more difficult to find than AML1. For AML1 only two equations have to be solved iteratively and this can be achieved on a fairly small machine. Also not many iterations are needed to obtain accurate solutions. For AML2 we have to solve all four maximum likelihood equations iteratively, which is slightly more difficult to program and requires more iterations than AML1 for the solution to be as accurate. However this routine still converges fairly rapidly.

Even though σ^2/ρ^2 is very small we have improved the estimate of ρ by using the better approximation to $I_{1,0}(z)$. Actually with σ^2/ρ^2 so small, the improvement in ρ is fairly marginal, but we will see later that as the ratio σ^2/ρ^2 increases the improvement in estimation of ρ becomes more marked.

Again all estimating procedures give good estimates, except perhaps σ_4^2 which is on the low side. Notice that whereas for AML1,

$$\sigma^2 = N^{-1} \sum_{i=1}^{N} \left(r_i - \bar{r}\right)^2 ,$$

for AML2, the maximum likelihood estimator

$$\sigma^2 = (2N)^{-1} \sum_{i=1}^{N} \left(r_i^2 - \rho^2\right) ,$$

is used. The asymptotic variances are

$$\mathrm{Var}(\hat{\xi}) = 0.1 \quad , \quad \mathrm{sd}(\hat{\xi}) \simeq 0.32 \ ,$$

$$\mathrm{Var}(\hat{\rho}) \simeq 0.05 \ , \quad \mathrm{sd}(\hat{\rho}) \simeq 0.22 \ ,$$

and

$$\mathrm{Var}\left(\hat{\sigma}^2\right) \simeq 0.1 \ , \quad \mathrm{sd}\left(\hat{\sigma}^2\right) \simeq 0.32 \ .$$

Notice that the standard deviation of $\hat{\sigma}^2$ is

$$\mathrm{sd}\left(\hat{\sigma}^2\right) = 2^{\frac{1}{2}}N^{-\frac{1}{2}}\sigma^2 \ .$$

N	10	20	30	33	35	50	100
$2^{\frac{1}{2}}N^{-\frac{1}{2}}$	0.447	0.316	0.258	0.246	0.239	0.2	0.141

It is clear that unless N is fairly large even the maximum likelihood estimator of σ^2 could be inaccurate. This is unfortunate as it makes inference based on $\hat{\sigma}^2$ difficult.

SET B $\quad N = 12$

	ξ	η	ρ	σ^2
Real Values	0	0	20	1
Chan's Est.	0.1601	-1.3399	21.2247	0.5412
AML1	0.1601	-1.3399	21.24	0.4600
A & B	1.3094	0.4413	19.1505	0.4913
A & B (mod)	1.3094	0.4413	19.1387	0.4528
FR1	1.7782	0.4109	19.1755	0.4509
FR2	0.2244	-1.2838	21.1537	0.4061

For data set B, $\tau_i \overset{d}{\sim} R[0, \pi/2)$ and this does not satisfy the assumptions of our model. The actual maximum likelihood equations for this model are estremely messy. The integral

$$\int_0^{\pi/2} \exp\left(p \cos \tau_i + q \sin \tau_i\right) d\tau_i$$

is not available in closed form. Because of this it was decided to use the functional relationship estimating equations given in chapter 4, where $\nu = 2$ and $k = 0$. Thus we have

$$N^{-1}\sum_{i=1}^{N}\left(X_i-\xi\right) - \rho^{-1}N^{-1}\sum_{i=1}^{N}\left(X_i-\xi\right)r_iI_{2,1}\left(r_i\rho/\sigma^2\right) = 0 \ ,$$

$$N^{-1}\sum_{i=1}^{N}\left(Y_i-\eta\right) - \rho^{-1}N^{-1}\sum_{i=1}^{N}\left(Y_i-\eta\right)r_iI_{2,1}\left(r_i\rho/\sigma^2\right) = 0 \ ,$$

$$\rho - N^{-1}\sum_{i=1}^{N} r_iI_{1,0}\left(r_i\rho/\sigma^2\right) = 0 \ ,$$

and

$$\sigma^2 - (2N)^{-1}\sum_{i=1}^{N}\left(r_i^2-\rho^2\right) = 0 \ .$$

If we make the approximations $I_{2,1}(z) \simeq 1$ and $I_{1,0}(z) \simeq 1$ in these equations the resulting estimation equations will be referred to as functional relationship 1, that is FR1. If we make the approximations $I_{2,1}(z) \simeq 1 - 3/2z$ and $I_{1,0}(z) \simeq 1 = 1/2z$ in the above equations, the resulting estimation equations will be referred to as FR2. The estimates based on these estimating equations are included here for interest. More detailed comments on these estimation procedures will be made later in this chapter. Here we will merely remark that the approximations made to obtain FR1 and FR2 from the functional relationship estimating equations are similar to those made to obtain AML1 and AML2 from the maximum likelihood estimating equations.

5. Ring of Brogar and Avebury

Brogar with stones 7 & 8 N = 35

	ξ	η	ρ	σ^2
Thom	0.357	-0.548	170.01	-
A & B	0.2861	-0.4526	169.99	3.2168
AML1	0.27	-0.47	169.97	3.1851
ρ_4, σ_4^2			174.38	-450

without stones 7 & 8 N = 33

	ξ	η	ρ	σ^2
Thom			170.33	
A & B	0.8390	-0.0417	170.41	1.3964
AML1	0.8466	0.0309	170.41	1.3574
ρ_4, σ_4^2			172.40	-260

Thom's values are taken from Thom and Thom (1973). A number of comments should be made about the analysis in this paper. The analysis is not entirely consistent especially in deciding into which quadrant a stone should be placed. Also stones 18 and 40 are marked as being present and an angular measurement is given, but no radii, and could not be included in the analysis.

Thom indicates that the stones consist of flat slabs of up to 15ft (4.6 m) high with average thickness not much over 9 inches (0.23 cm). The mean in quadrant approximation is used to fit the circle, with an initial radial estimate of 170 ft (51.8 m). 0.30 ft (0.091 m) is given as the standard deviation of the radial estimate obtained by this method. A justification, based on a 19th century survey of the ring, is given to support the hypothesis that stones 7 and 8 have been markedly affected by re-erection. The circle was re-estimated omitting these stones, and the revised estimate of the standard deviation of the estimate of the radius drops to 0.22 ft (0.067 m). The circle fitting algorithm of Angell and Barber, and maximum likelihood estimation give essentially the same estimates of ξ, η and ρ. σ^2/ρ^2 is extremely small, much less than 0.01 and so the approximation $I_{1,0}(z) \simeq 1$ is valid here. The drop in the estimate of σ^2 by the omission of stones 7 and 8 tends to support the hypothesis of re-erection.

Let us consider the second analysis, $N = 33$, and take the maximum likelihood estimates. The 95% confidence interval for the diameter based on $\hat{\sigma}^2$ is 340.82 ± 0.56 ft. A more conservative confidence interval can be found by noting that $\mathrm{sd}(\hat{\sigma}) = 0.247\sigma^2 \simeq \frac{1}{4}\sigma^2$. So the 95% confidence interval for $\hat{\sigma}^2$ is plus and minus half the variance, that is $\pm\frac{1}{2}\sigma^2$. At worst therefore $\hat{\sigma}^2 \simeq \frac{1}{2}\sigma^2$, so at most $\sigma^2 \simeq 2\hat{\sigma}^2$. The 95% confidence interval for $\hat{\rho}$ using $\sigma^2 \simeq 2\hat{\sigma}^2$ is 170.41 ± 0.56 ft, or for the diameter 340.82 ± 1.12 ft. The width of this conservative interval is 2.24 ft which is more than 0.8 of a megalithic yard. Not surprisingly the confidence interval for the diameter does include the point 340 ft = 125 MY. In fact we really ought to find the joint probability ellipsoid for $\hat{\rho}$ and $\hat{\sigma}^2$, but this would be very difficult.

Another interesting point is the comparatively poor estimation by the estimators ρ_4 and σ_4^2. In both cases σ_4^2 is wildly negative and ρ_4 has a positive bias. Later we will see that $N = 35$ is a small number for the estimation of ρ_4 and more especially σ_4^2; we are not in an asymptotic situation where N is large and therefore small sample properties of ρ_4 and σ_4^2 dominate. In small samples the variance of σ_4^2 increases markedly as the radius ρ increases.

Thom *et al* (1976) assume a complicated structure for the stone ring at Avebury, which is imposed upon the data before any estimation of the parameters. They maintain that these assumptions are justified because the resulting estimate of the variance of the radius is small. Angell and Barber (1977) and Freeman (1977) comment on this estimation procedure. Angell and Barber estimate the centre and radius of the four arcs using their circle-fitting algorithm, and Freeman uses least squares assuming that for each arc r_i given ξ and η has a normal distribution with mean ρ and variance σ^2. Freeman also perturbs the last decimal place of the observations using equal probabilities of adding +1, 0 or -1 to that figure and again estimates the parameters. These slight perturbations have a perceptible effect on the estimators. One reason for this is that to each of the data sets we are only fitting an arc of about 20° instead of the whole circle. As Angell and Barber (1977) point out, the error factors are of the same order of magnitude for a large range of radii values; the likelihood function is fairly "flat".

Although the results have been given in terms of residual sums of squares to conform with the results given by Freeman, the respective estimates for σ^2 using least squares are 3.20, 3.07, 6.30 and 0.41. These indicate that the ratio σ^2/ρ^2 is very small for all four arcs and so the least square estimates given by Freeman are approximately maximum likelihood.

	ξ	η	ρ	Residual S.S.
CENTRE C	$N = 7$			
Thom	520.9	720.8	707.7	32.19
Angell & Barber	538.6	585.0	573.0	~25
Freeman L.S.	530.8	651.0	638.8	24.40
Freeman P.L.S.	531.5	647.4	635.2	22.52
F.R.2	*	*	*	
CENTRE W	$N = 16$			
Thom	1612.8	1697.1	2041.5	63.21
Angell & Barber	1375.9	1446.8	1697.5	~52
Freeman L.S.	1472.0	1553.4	1840.4	49.20
Freeman P.L.S.	1475.5	1557.6	1845.7	48.50
F.R.2	1471.9	1553.3	1840.4	
CENTRE A	$N = 10$			
Thom	723.2	538.6	707.7	72.21
Angell & Barber	667.1	561.1	648.6	~78
Freeman L.S.	795.0	516.5	782.8	63.04
Freeman P.L.S.	802.2	513.8	790.5	79.05
F.R.2	745.0	516.5	782.8	
CENTRE B	$N = 7$			
Thom	586.6	386.9	707.7	19.32
Angell & Barber	503.2	548.5	527.4	~3.4
Freeman L.S.	512.7	533.1	545.5	2.89
Freeman P.L.S.	514.3	529.4	549.3	2.78
F.R.2	*	*	*	

FR2 results from the functional relationship estimating equations. Notice that there is little or no difference between the estimates obtained by this method and those obtained by least squares in the two cases where the iteration procedure for FR2 converged. For both centres C and B the iterative solutions using initial values, the least squares estimates and initial values, Angell and Barber's estimators failed to converge. Even worse they actually diverged away from the initial values. Although one might argue that seven data points are rather few to estimate four parameters, the other estimation procedures indicate that the points do lie on an arc. The failure of FR2 to converge to any solution must throw doubt on the stability of the estimating equations for short arc lengths.

6. Simulation Studies: AML2

Assume we have an estimation procedure which estimates a population parameter θ from a sample of size N. We know the asymptotic properties of this estimator, $\tilde{\theta}$, as N goes to infinity but wish to examine the small sample properties of $\tilde{\theta}$. One method of doing this is to simulate M independent sets of samples of size N, then independently estimate $\tilde{\theta}$ from each sample. If M is large enough we can then examine the sample distribution of $\tilde{\theta}_i$, $i = 1, 2, \ldots, M$, to see how much it differs from the asymptotic distribution of $\tilde{\theta}$. The most correct method of testing the hypothesis that the sample distribution does not differ from the asymptotic distribution, except by random fluctuations, is to base the test statistic on the sample distribution function of $\tilde{\theta}$, $F_M(\tilde{\theta})$ and use one of the goodness-of-fit tests. For example if the asymptotic distribution of $\tilde{\theta}$ is $N(\theta, v^2)$ we could use the Kolmogorov-Smirnov statistic

$$D_M = \sup_{\tilde{\theta}} \left|F_M(\tilde{\theta}) - \Phi\{(\tilde{\theta}-\theta)/v\}\right| .$$

One disadvantage of using D_M for testing is that we may have to calculate $\Phi(\cdot)$ for every $\tilde{\theta}_i$, $i = 1, 2, \ldots, M$. This would be tedious especially if M is large, and although $\Phi(\cdot)$ is well tabulated, in practice we would have to use one of the computer approximations to $\Phi(\cdot)$. Also if we discover that the sample distribution of $\tilde{\theta}$ does deviate significantly from the asymptotic distribution we would still be left with the problem of examining how the sample distribution differs. In addition we have a large number of parameter estimators to examine, not all of which have an asymptotic distribution but only some asymptotic moments and we wish to examine the small sample properties of the moments of the estimators. Since we also wish to test the sample moments of the estimators for deviations from the moments of the asymptotic distribution we require M to be reasonably large. The amount of work required to calculate the D_M statistics would be very large and an alternative test must be considered.

We could consider calculating the first four k statistics of the sample distribution of $\tilde{\theta}_i$, $i = 1, 2, \ldots, M$. Then

$$E(k_j) = \kappa_j , \quad j = 1, 2, 3, 4 ,$$

where κ_j is the jth cumulant of the distribution of $\tilde{\theta}_i$ based on a sample of size N . An important point should be noted here. There is nothing in the maximum likelihood theory to suggest that a maximum likelihood estimator has any moments. If $\tilde{\theta}_i$ is a maximum likelihood estimator then the theory merely says that the limiting distribution as N goes to infinity, of $\sqrt{N}\,(\tilde{\theta}-\theta)$ is normal with zero mean and variance v^2 . $\tilde{\theta}_i$ need not have a mean or variance, for example see Cox and Hinkley [(1974), p. 295]. However we can still examine the sample distribution of $\tilde{\theta}_i$ and its k-statistics and see how much the k- statistics differ from those of the asymptotic distribution and its cumulants.

From Kendall and Stuart [(1977), p. 325], in normal samples k_2 , the sample variance is independent of $k_p/k_2^{p/2}$, $p = 3, 4, \ldots$, hence this ratio has mean zero for all p . Letting

$$\gamma_1 = k_3/k_2^{3/2}$$

and

$$\gamma_2 = k_4/k_2^2 ,$$

then γ_1 is a measure of the skewness of the sample distribution and γ_2 is a measure of its kurtosis. From above, if $\tilde{\theta}_i$ were normal with mean θ and variance v^2 then

$$E\left[\gamma_1(\tilde{\theta})\right] = 0 = E\left[\gamma_2(\tilde{\theta})\right] .$$

To first order in normal samples,

$$\text{Var}\left(\gamma_1\right) = 6M^{-1}$$

and

$$\text{Var}\left(\gamma_2\right) = 24M^{-1} ,$$

consequently unless M is moderately large it would be difficult to distinguish minor deviations in skew and kurtosis from a null hypothesis of normality. Similarly it could be difficult to distinguish a small bias term in k_1 and k_2 from statistical fluctuations if M were not sufficiently large. We must keep in mind that a small order bias, especially in the k_2

statistic, could be introduced by the convergence criteria used in our iterative procedures. More importantly we might expect a systematic bias if an approximation method is used. For example we would expect the estimator of the radius in the AML1 procedure to have a small positive bias.

We are not really interested in an exhaustive study to detect minor deviations from normality and small order bias terms; we have too many estimators to contemplate this. Rather we are interested in a smaller scale study to detect gross deviations from theoretical behaviour and gain a rough idea of how large N has to be before we can assume that the asymptotic behaviour of the estimator will hold. Also we wish to examine estimators for which no theoretical asymptotic distribution has been postulated such as ρ_3, ρ_4 and σ_4^2, and the estimators of Angell and Barber, to see how far their distributions deviate from normality and to determine their small sample behaviour to some extent.

A large number of tests for departure from normality have been postulated, see for example Pearson *et al* (1977). One simple test of this type, quoted by Pearson *et al*, which uses the sample skewness and kurtosis statistics, γ_1 and γ_2, is the R-test. It requires having good approximations to the probability integrals of γ_1 and γ_2 in normal samples. These can be obtained from D'Agostino and Pearson (1973) and Pearson and Hartley (1970).

If $\pm A(\alpha')$ are the upper and lower $100\alpha'\%$ points of γ_1 and $B(\alpha')$ and $C(\alpha')$ are the similar lower and upper points for γ_2, then the four points with co-ordinates $\{-A(\alpha'), B(\alpha')\}$, $\{+A(\alpha'), B(\alpha')\}$, $\{-A(\alpha'), C(\alpha')\}$, $\{+A(\alpha'), C(\alpha')\}$ form a rectangle. If γ_1 and γ_2 were independent in sampling from a normal population, the overall probability of a (γ_1, γ_2) point falling outside this rectangle would be

$$\alpha = 4\{\alpha'-(\alpha')^2\} ,$$

$$\alpha' = \tfrac{1}{2}\{1-(1-\alpha)^{\frac{1}{2}}\} .$$

Because γ_1 and γ_2 are not independent, a smaller percentage of points than α fall outside the rectangle, so that this test is conservative. Note that γ_1 and γ_2 are very nearly independent in samples of size greater than 100. Pearson *et al* do give corrections for $\alpha = 0.05$ and

$\alpha = 0.10$ when $M = 20$, 50 and 100 .

The computer used for the simulations was the Univac 1100/42 at the Australian National University. The circles were generated by independently generating sets of $3N$ uniform deviates $\{T_i, U_i, V_i\}_{i=1}^{N}$ using RAND. The sequence $\{\tau_i\}_{i=1}^{N}$ was generated using the equality,

$$\tau_i = 2\pi T_i \ , \quad i = 1, 2, \ldots, N \ ,$$

and the ε_i and ν_i normal deviates were generated using the polar method

$$\varepsilon_i = \left(-2 \ln U_i\right)^{\frac{1}{2}}\cos\left(2\pi V_i\right) \ ,$$

and

$$\nu_i = \left(-2 \ln U_i\right)^{\frac{1}{2}}\sin\left(2\pi V_i\right) \ .$$

Double precision was used to maintain accuracy in the iterative procedures. The pseudo-random number generator RAND was programmed by R. Brent and is not to be confused with RANDU. RANDU generates sets of pseudo-uniform variates by a congruence scheme of the form

$$X_{i+1} = \left(aX_{i+1}+b\right)\bmod 2^{35} \ ,$$

where $a = 5^{15} = 3051758125$ and $b = 7261067085$. Using RANDU does not satisfactorily generate independent and unbiased $\left(\varepsilon_i, \nu_i\right)$ pairs, even from independent sets $\{U_i\}_{i=1}^{N}$ and $\{V_i\}_{i=1}^{N}$.

First we will examine AML2 for $M = 800$ and use the uncorrected conservative R-test to test for departures from normality. For this value of M the conservative test will be very nearly an exact test. If we take $\alpha' = 0.01$, then α is just under 0.04 , and $\pm A(\alpha') = 0.202$, $B(\alpha') = -0.35$ and $C(\alpha') = 0.46$. We will call this rectangle R_1 . In fact not many other values of $A(\alpha')$, $B(\alpha')$ and $C(\alpha')$ have been tabulated. Pearson *et al* (1977) give a relationship between γ_1 and a standard unit normal variate Z ,

$$Z\left(\gamma_1\right) = \delta \sinh^{-1}\left(\gamma_1/\lambda\right) \ ,$$

and give a table of δ and λ^{-1} values for a wide range of M . Thus we can evaluate $A(\alpha')$ for any α' value.

α'	0.01	0.005	0.001	0.0001
$A(\alpha')$	0.202	0.224	0.270	0.327

No such expression has been derived for the non-symmetric variable γ_2. Table 34C from Pearson and Hartley (1970) only gives $B(\alpha')$ and $C(\alpha')$ values for $\alpha' = 0.01$ and $\alpha' = 0.05$. In addition D'Agostino and Pearson (1973) give $B(\alpha')$ and $C(\alpha')$ for a larger range of α' values, but only for $M \leq 200$. We will still set up R_2 which has $\alpha' = 0.005$ and $\alpha = 0.02$, R_3 with $\alpha' = 0.001$ and $\alpha = 0.004$ and R_4 with $\alpha' = 0.0001$ and $\alpha = 0.0004$.

Circles were simulated with $\rho = 5.0$, 10.0 and 20.0, $\sigma^2 = 1.0$, $\xi = \eta = 0$, and for $N = 12$, 20, 30, 40, 50 and 75. We wish to examine $\hat{\xi}$, $\hat{\eta}$, $\hat{\rho}$ and $\hat{\sigma}^2$ for all these circles and sample sizes. We must keep in mind the dangers of applying a large number of tests; some will be significant by chance and this is one reason for selecting an extreme α value. For each value of N the sets of circles with $\rho = 20$ and $\rho = 10$ were generated with the same (ε_i, ν_i) deviates as the sets of circles with $\rho = 5$. This was done to enable a closer examination of bias in any of the estimators as ρ increases, without requiring the bias terms to be statistically significant. If, for example, the bias in $\hat{\rho}$ decreased as ρ increased, this would indicate that the approximation method was estimating ρ more accurately as σ^2/ρ^2 decreased. As indicated previously, the study was not intended to be exhaustive. The joint normality of $\hat{\rho}$ and $\hat{\sigma}^2$ was not tested, and neither was the independence of $\hat{\xi}$ and $\hat{\eta}$. Some of the results of the simulation studies are tabulated in Appendix D.

As a result of the approximation procedure there is probably a slight positive bias in $\hat{\rho}$ which decreases with increasing ρ. As long as the distribution of $\hat{\rho}$ has finite mean and variance, then even if $\hat{\rho}$ is not normally distributed, by the central limit theorem the distribution of $k_1(\hat{\rho})$ will be close to normal. The theoretical standard standard deviation of this k_1 statistic based on the asymptotic variance of $\hat{\rho}$ is $\mathrm{sd}(\hat{\rho})M^{-\frac{1}{2}}$. The only $k_1(\hat{\rho})$ which is outside the two-sided 95% confidence interval centred on ρ is when $(N, \rho) = (30, 5)$ where $k_1(\hat{\rho}) = 5.0136$ and the theoretical standard deviation based on the asymptotic variance of $\hat{\rho}$ is 0.0066. If we had sufficient confidence that the approximation method would result in a positive bias for $\hat{\rho}$ we should really only consider a one-sided test, and in this instance the above case would still be significant.

The apparent bias in $\hat{\rho}$ does decrease as ρ increases, except for $N = 12$ when the bias is negative and increases with ρ. The apparent bias in $\hat{\rho}$ also decreases as N increases, at least for N greater than 12. There is a marked negative bias in $k_1\left(\hat{\sigma}^2\right)$ which is significant for all N and ρ. This bias does decrease with increasing ρ and N. For example for $(N, \rho) = (12, 5)$, $k_1\left(\hat{\sigma}^2\right) = 0.7382$, and if the asymptotic variance of $\hat{\sigma}^2$ held, then the standard deviation of this k_1 statistic would be 0.0148, and $k_1\left(\hat{\sigma}^2\right)$ is hence more than 17 standard deviations below the expected value of 1.0000. For $N = 75$ and $\rho = 20$, $k_1\left(\hat{\sigma}^2\right) = 0.9577$ where the predicted value of the standard deviation of k_1 is 0.0058. $k_1\left(\hat{\sigma}^2\right)$ is still 7 standard deviations below its expected value. The sample variance of $\hat{\sigma}^2$ is less than the asymptotic variance, so use of the former in calculating the standard deviation of k_1 would not alter the situation. This bias is at least partially due to the approximation $I_{1,0}(z) \approx 1 - 1/2z$ used, and the resulting positive bias in the estimation of ρ. Also

$$\hat{\sigma}^2 = (2N)^{-1} \sum_{i=1}^{N} \left(r_i^2 - \hat{\rho}^2\right)$$

and

$$E\left(\hat{\rho}^2\right) = [E(\hat{\rho})]^2 + \mathrm{Var}(\hat{\rho}) ,$$

so $\hat{\rho}^2$ is only an asymptotically unbiased estimator of ρ^2.

Even though an approximation method is used, due to considerations of symmetry, $\hat{\xi}$ and $\hat{\eta}$ using AML2 should be unbiased. For all N and ρ, both $k_1(\hat{\xi})$ and $k_1(\hat{\eta})$ are well within the required bounds.

As indicated previously, the k_2 statistics are the "sample" variances of the maximum likelihood estimators and give unbiased estimates of the variance of the maximum likelihood estimators. If the asymptotic normal distribution of the maximum likelihood estimator holds, then since M is large we could assume that the distribution of k_2 is also very close to normal. With $M = 800$, the standard deviation of the k_2 statistics would then be 1/20 th of their expected value. Before examining the k_2 statistics we should look at γ_1 and γ_2 to see if the assumptions of normality of the estimators are violated.

For all ρ and $N = 12$ and 50, $\gamma_1(\hat{\sigma}^2) > 0.202$ and $\gamma_2(\hat{\sigma}^2) > 0.46$, indicating that $\hat{\sigma}^2$ has a distribution that is positively skewed and leptokurtic. For all ρ and $N = 20, 30, 40$ and 75, $\gamma_1(\hat{\sigma}^2)$ exceeds 0.202, decreasing with increasing ρ and also decreasing with N, except that $\gamma_1(\hat{\sigma}^2)$ is lower when $N = 50$ than it is when $N = 75$. For $(N, \rho) = (75, 20)$, $\gamma_1(\hat{\sigma}^2) = 0.325$, which is still outside R_4. The distribution of $\hat{\sigma}^2$ can not be regarded as being normal.

For all N and ρ, $(\gamma_1(\hat{\rho}), \gamma_2(\hat{\rho}))$ is inside R_1. Let us also consider the rectangle R_0 where $A(\alpha') = \pm 0.142$ and $B(\alpha') = -0.26$, and $C(\alpha') = 0.29$. Here $\alpha' = 0.05$, and thus $\alpha = 0.19$. Then $(\gamma_1(\hat{\rho}), \gamma_2(\hat{\rho}))$ lies inside R_0 for all N and ρ except $N = 20$. For $N = 20$, $\gamma_1(\hat{\rho})$ is -0.158 when $\rho = 5$, -0.156 when $\rho = 10$ and -0.153 when $\rho = 20$. We would not reject the hypothesis that $\hat{\rho}$ is approximately normally distributed. For $N = 12$, $\gamma_2(\hat{\eta}) = 1.503$ when $\rho = 5$, 0.824 when $\rho = 10$ and 0.626 when $\rho = 20$, and for $N = 20$, $\gamma_2(\hat{\eta}) = 0.583$ when $\rho = 5$, and these are all outside R_1. For all other (N, ρ) we would not reject the hypothesis that $\hat{\xi}$ and $\hat{\eta}$ are normal. Notice that by symmetry $\hat{\xi}$ and $\hat{\eta}$ would have the same distribution.

For $N \geq 30$ we can regard the distributions of $\hat{\xi}$, $\hat{\eta}$ and $\hat{\rho}$ as being reasonably close to normal. Even though $\hat{\xi}$, $\hat{\eta}$ and $\hat{\rho}$ are not normally distributed for $N = 12$ and 20, we can still compare the k_2 statistics with the expected asymptotic values. Although we have shown that the distributions of some of the estimators, notably $\hat{\sigma}^2$, are non-normal we have not investigated the question of robustness; that is how non-normal does a population have to be before procedures derived theoretically using the normal assumption become invalid. Pearson and Please (1975) give some answers to this question. In drawing up confidence intervals for k_2 statistics the following observation should be made. If $\tilde{\theta}_i$ has a distribution with variance v^2 then

$$E\left(k_2(\tilde{\theta})/v^2\right) = 1$$

and

$$\mathrm{Var}\left(k_2(\tilde{\theta})/v^2\right) = \{\gamma_2 + 2 + 2/(M-1)\}M^{-1} \rightarrow \left(\gamma_2+2\right)M^{-1} .$$

The shape of the distribution of $k_2(\tilde{\theta})$ tends to normality as M increases to infinity, but the variance of this distribution is dependent on γ_2 . If the distribution of $\tilde{\theta}$ is markedly leptokurtic, then the variance of $k_2(\tilde{\theta})$ will be larger than we might expect if we had assumed a normal distribution for $\tilde{\theta}$.

For $N = 12$, $\hat{\eta}$ and therefore $\hat{\xi}$ have significant γ_2 statistics, but even taking this into account $k_2(\hat{\xi})$ and $k_2(\hat{\eta})$ are both more than two standard deviations above their expected values. This is also true for $k_2(\hat{\eta})$ when $N = 20$, and $N = 30$, and for $k_2(\hat{\xi})$ when $N = 40$, for all ρ . For $N = 12$ and all ρ , $k_2(\hat{\rho})$ is more than five standard deviations above its expected value. For example when $N = 12$ and $\rho = 5$, $k_2(\hat{\rho}) = 0.1111$, whereas the expected value is 0.0870 and has standard deviation 0.0044 .

$k_2\left(\hat{\sigma}^2\right)$ is a more difficult statistic to evaluate, since $\hat{\sigma}^2$ has marked leptokurtic properties; for $(N, \rho) = (12, 5)$, $\gamma_2\left(\hat{\sigma}^2\right) = 1.33$. However for all $N \leq 50$, and all ρ , $k_2\left(\hat{\sigma}^2\right)$ is more than two standard deviations below its expected value, whereas for $N = 75$ it is around one standard deviation below its expected value. The negative bias in $k_2\left(\hat{\sigma}^2\right)$

is more than 5 standard deviations when $N = 12$, and decreases with increasing N.

The bias in $k_2(\hat{\sigma}^2)$ decreases marginally with increasing ρ. The bias in $k_2(\hat{\sigma}^2)$ is probably due to the approximation method and can not really be regarded as a defect. The bias in $k_1(\hat{\sigma}^2)$ is far more serious.

For $N \geq 20$, $k_2(\hat{\rho})$ is well within the 95% confidence limits set around the asymptotic variance of $\hat{\rho}$ although it is probably still a little on the high side for $N = 20$. $\hat{\rho}$ seems to approach normality faster than $\hat{\xi}$ and $\hat{\eta}$ which approach normality faster than $\hat{\sigma}^2$. Even for $N = 75$, $\hat{\sigma}^2$ is biased and skew, and the bias of $\hat{\sigma}^2$ is a little worrying since an accurate estimate of $\hat{\sigma}^2$ is needed in practice to find confidence intervals for the other parameter estimates.

7. Simulation Studies: AML1 and Angell and Barber's Circle Fitting Algorithm

The same sets of circles that were generated to test AML2 were used to test AML1 and Angell and Barber's circle fitting algorithm. AML1 will be discussed first and compared with AML2.

It was decided to ignore the results for $(N, \rho) = (12, 5)$. For this set of circles all the γ_1 and γ_2 statistics are so highly significant, indicating gross departures from normality that it was decided not to test the k_1 and k_2 statistics. Otherwise the results for γ_1 and γ_2 show a similar pattern to the estimators derived using AML2. For $N \geq 20$, $\hat{\xi}$, $\hat{\eta}$ and $\hat{\rho}$ do not exhibit skewness or kurtosis. For $N = 12$ and 50, $\hat{\sigma}^2$ exhibits both positive skewness and positive kurtosis, while for $N = 20$, 30, 40 and 75, $\hat{\sigma}^2$ exhibits positive skewness.

Again from symmetry we would expect $\hat{\xi}$ and $\hat{\eta}$ using AML1 to be unbiased, and this was indeed the case. The actual k_1 statistics obtained were similar to those obtained using AML2. As expected $k_1(\hat{\rho})$ is definitely biased. The bias does decrease as ρ decreases but is statistically significant for all N and ρ. For $N = 20$ and $\rho = 5$,

$k_1(\hat{\rho}) = 5.114$, for $\rho = 10$ it is 10.056 and for $\rho = 20$ it is 20.028 , whereas the standard deviation of k_1 based on the asymptotic theoretical distribution of $\hat{\rho}$ is around 0.008 . For $N = 75$, the corresponding values of k_1 are 5.106 for $\rho = 5$, 10.053 for $\rho = 10$ and 20.027 for $\rho = 20$ whereas the standard deviation of k_1 is around 0.004 . Of course here $\hat{\rho}$ is $\bar{r}$, with ξ and η replaced by unbiased estimators. If ξ and η were known and σ^2/ρ^2 small

$$E(\bar{r}) = \rho[1 + \tfrac{1}{2}\sigma^2/\rho^2 + \tfrac{1}{8}\sigma^4/\rho^4 \ldots]$$

which is 20.025 when $\rho = 20$ and $\sigma^2 = 1$. The bias in $\hat{\rho}$ when $\rho = 20$ is small compared with ρ and is also fairly small compared with the standard deviation of $\hat{\rho}$, which is $\left(1.0125\ \sigma^2 N^{-1}\right)^{\frac{1}{2}}$, if N is reasonably small. For example if $N = 75$, $\mathrm{sd}(\hat{\rho}) = 0.116$. The standard deviation of $\hat{\rho}$ is of the order of 0.025 for N greater than 1600 . Confidence intervals based on $\hat{\rho}$ would probably include the correct value, but not the correct probability content. For larger values of σ^2/ρ^2 problems could arise in setting up confidence intervals for ρ based on $\hat{\rho}$. For example for $\sigma^2/\rho^2 = 0.01$, the bias in $\hat{\rho}$ is about double the bias in $\hat{\rho}$ for $\sigma^2/\rho^2 = 0.0025$, that is around 0.05 , whereas the standard deviation of $\hat{\rho}$ is $\left(1.05\ \sigma^2 N^{-1}\right)^{\frac{1}{2}}$.

Again $k_1\left(\hat{\sigma}^2\right)$ is significantly biased for all N and ρ . $k_1\left(\hat{\sigma}^2\right)$ tends to be smaller than the corresponding value derived using AML2 especially for smaller circles, but this effect is very small and practically non-existent for $\rho = 20$.

The k_2 statistics for $\hat{\xi}$ and $\hat{\eta}$ tend to be slightly bigger than the corresponding values under AML2, this effect tending to disappear for increasing ρ . Whereas the k_2 statistics for $\hat{\rho}$ tend to be smaller. For all N considered and $\rho = 5$ both $k_2(\hat{\xi})$ and $k_2(\hat{\eta})$ are significantly bigger than their asymptotically expected values. For $\rho = 10$ and $\rho = 20$ and $N = 30, 40$ and 75 either one or both of $k_2(\hat{\xi})$ and $k_2(\hat{\eta})$ is significantly bigger than its asymptotically expected value. For example for $N = 75$, $\rho = 5$, $k_2(\hat{\xi}) = 0.030$ and $k_2(\hat{\eta}) = 0.031$, whereas the asymptotic expected value is 0.027 with standard deviation around 0.001 . When $\rho = 20$, $k_2(\hat{\xi}) = 0.028$ and $k_2(\hat{\eta}) = 0.029^+$, whereas the asymptotic expected value is 0.027^- with standard deviation around 0.001 . We can see that for $(N, \rho) = (75, 20)$ the k_2 statistics are just significantly too large but $k_2(\hat{\xi}) + k_2(\hat{\eta})$ is not significantly too big. The k_2 statistics for $\hat{\rho}$ tend to be slightly smaller than the corresponding values under AML2. This is to be expected, since certainly if ξ and η are known $\bar{r}$ has a smaller variance than $\hat{\rho}$.

The k_2 statistics for $\hat{\sigma}^2$ are similar to those for $\hat{\sigma}^2$ using AML2, except they tend to be slightly smaller for $\rho = 5.0$. $\gamma_1(\hat{\sigma}^2)$ also has very similar behaviour to that in AML2. It could be marginally smaller for $\rho = 5$, but is certainly of the same order of magnitude.

As expected AML1 is not suitable for small circles, even for $\sigma^2/\rho^2 = 0.04$ especially if the sample size is small. As the sample size increases for σ^2/ρ^2 small, and as ρ increases, the properties of the estimators approach those of AML2. However care must be exercised in basing any confidence intervals for ρ on $\hat{\rho}$; some account must be taken of the bias in $\hat{\rho}$. AML1 does have the advantage that only two parameters, ξ and η , are iterated instead of four, and hence it converges much faster than AML2. However AML2 converges fairly quickly for large N .

No asymptotic distributions have been assumed for the estimators obtained by using Angell and Barber's circle fitting algorithm. There are actually six estimators that we will consider here, ξ_a, η_a, ρ_a, σ_a^2, ρ_a^* and σ_a^{*2} . These have all been defined in section 4. We also have some theoretical moments of these estimators.

Assuming that the distribution of the estimator does have a finite mean and variance, by the central limit theorem we can regard the k_1 statistics as being normally distributed. We will consider that an estimator is significantly biased if it is more than two standard deviations from the actual circular parameter. Then none of the $k_1(\xi_a)$ or $k_1(\eta_a)$

statistics are significantly biased.

If ξ and η are known, then the expected bias in ρ is of order $+1/\rho$. In fact $k_1(\rho_a)$ is lower than $\rho + 1/\rho$ for all N and ρ, except $(N, \rho) = (30, 20)$ and $(N, \rho) = (50, 20)$. It is clear that as N increases, $k_1(\rho_a)$ does tend to $\rho + 1/\rho$. This bias in $k_1(\rho_a)$ is significant. Since we expected ρ_a to be positively biased a one-sided test is more appropriate here.

For $N = 75$ and $\rho = 5$, $k_1(\rho_a) = 5.192$, for $\rho = 10$, $k_1(\rho_a) = 10.098$ and for $\rho = 20$, $k_1(\rho_a) = 20.050^-$. The standard deviation of $k_1(\rho_a)$ for $N = 75$ is around 0.004. The bias of ρ_a^* is about half that of ρ_a as expected, and from this point of view one might expect ρ_a^* to be a better estimator of the radius.

$k_1\left(\sigma_a^2\right)$ is more difficult to analyse since we have not calculated any statistics of the distribution of σ_a^2. Some trends in $k_1\left(\sigma_a^2\right)$ are apparent. $k_1\left(\sigma_a^2\right)$ appears to be lowest for $\rho = 5$, but does not differ greatly between $\rho = 10$ and $\rho = 20$ for the same N value. It also tends to increase with N for the same ρ: $k_1\left(\sigma_a^2\right)$ underestimates σ^2, but underestimates σ^2 less than does $\hat{\sigma}^2$ under AML2. Except for $N = 12$, $k_1\left(\sigma_a^{*2}\right) < k_1\left(\hat{\sigma}^2\right)$ (under AML2) < 1. Even for N as large as 75 there is still some difference between $k_1\left(\sigma_a^{*2}\right)$ and $k_1\left(\sigma_a^2\right)$. So both σ_a^2 and σ_a^{*2} underestimate σ^2, but as far as bias is concerned σ_a^2 performs better than $\hat{\sigma}^2$, whereas σ_a^{*2} does not perform as well as either.

Before discussing the k_2 statistics, it is necessary to look at γ_1 and γ_2. ρ_a does not exhibit skewness or kurtosis for any of the N and ρ values considered, and neither does ρ_a^*. $\gamma_2(\xi_a)$ is significant for $N = 12$ and $\rho = 5$ and 10; for $N \geq 20$ or $\rho = 20$, ξ_a and η_a do not exhibit skewness or kurtosis. $\gamma_1\left(\sigma_a^2\right)$ is significant for all N and

ρ , and $\gamma_2\left(\sigma_a^2\right)$ is significant for $N = 12$, when $\rho = 5$ and 10 and for $N = 20$ when $\rho = 5$, the latter γ_2 value being just significant. $\gamma_1\left(\sigma_a^{*2}\right)$ is significant for all N when $\rho = 5$, and additionally for all ρ when $N = 20$ and for $\rho = 10$ when $N = 75$. σ_a^{*2} shows about the same skewness as $\hat{\sigma}^2$ under AML2, and appears to be more skew and leptokurtic than σ_a^2 .

We have not derived an asymptotic variance for either σ_a^2 or σ_a^{*2} , but instead will compare the k_2 statistics with the asymptotic values predicted by maximum likelihood. $k_2\left(\sigma_a^{*2}\right)$ seems to increase markedly as the ratio σ^2/ρ^2 decreases. For example for $N = 50$ and $\rho = 5$, $k_2\left(\sigma_a^{*2}\right) = 0.1130$, for $\rho = 10$, $k_2\left(\sigma_a^{*2}\right) = 0.338$ and for $\rho = 20$, $k_2\left(\sigma_a^{*2}\right) = 1.234$. No correction is required to the variance of this statistic as $\gamma_2\left(\sigma_a^{*2}\right)$ is quite small, and thus the standard deviation of k_2 is one twentieth of its expected value. Compare these k_2 statistics with those for σ_a^2 ; for $\rho = 5$, $k_2\left(\sigma_a^2\right) = 0.039$, for $\rho = 10$, $k_2\left(\sigma_a^2\right) = 0.0390$ and for $\rho = 20$, $k_2\left(\sigma_a^2\right) = 0.038$. Obviously the variance of σ_a^{*2} is much greater than that of σ_a^2 , even when $\sigma^2/\rho^2 = 0.04$. So σ_a^2 must be the preferred estimator. $k_2\left(\sigma_a^2\right)$ is greater than $k_2\left(\hat{\sigma}^2\right)$ under AML2. In fact except when $N = 12$ and $N = 30$ the $k_2\left(\sigma_a^2\right)$ are well within the 95% confidence limits set for the k_2 statistics of the asymptotic distribution of $\hat{\sigma}^2$.

On the other hand although $k_2\left(\rho_a^*\right)$ is greater than $k_2\left(\rho_a\right)$, the difference is only about ten percent for small N , and decreases as N increases. For N as large as 75 there is very little difference between $k_2\left(\rho_a^*\right)$ and $k_2\left(\rho_a\right)$. Here the bias in ρ_a would be the deciding factor making the use of ρ_a^* more desirable. This leads to a most peculiar estimation scheme where we use σ_a^2 but not ρ_a and ρ_a^* but not σ_a^{*2} .

Probably σ_a^2 should be dropped in favour of some other variance estimator based on ρ_a^* especially for N reasonably large.

Both ρ_a and ρ_a^* have good variance properties compared with $\hat{\rho}$ under AML2. For example when $(N, \rho) = (75, 20)$, $k_2(\hat{\rho}) = 0.0126$ whereas $k_2(\rho_a) = 0.0125$ and $k_2(\rho_a^*) = 0.0130$, and when $(N, \rho) = (75, 5)$, $k_2(\hat{\rho}) = 0.0132$ whereas $k_2(\rho_a) = 0.0121$ and $k_2(\rho_a^*) = 0.0128$. The variance of ρ_a is

$$\mathrm{Var}(\rho_a) = \sigma^2 N^{-1}(1 + \sigma^2/\rho^2)(1 + 2\sigma^2/\rho^2)^{-1} ,$$

from section 4, and this is less than the variance of $\hat{\rho}$. For example for $(N, \rho) = (75, 5)$, $\mathrm{Var}(\hat{\rho}) = 0.0139$ whereas $\mathrm{Var}(\rho_a) = 0.0128$. Again because of bias considerations $\hat{\rho}$ would be preferred if we require a very accurate estimate of ρ, however if σ^2/ρ^2 is extremely small ρ_a^* might be preferred as being quicker to calculate. Except for $N = 12$ and probably $N = 20$, $k_2(\rho_a)$ is not significantly different from its expected value. As before, to be significant the observed k_2 statistics needs to differ from its expected value by ten percent.

Both $k_2(\xi_a)$ and $k_2(\eta_a)$ drop to within a respectable distance of their expected value for N around 50. As anticipated ξ_a and η_a are efficient estimators of ξ and η if σ^2/ρ^2 is small, becoming less efficient as the ratio σ^2/ρ^2 increases.

Angell and Barber's circle fitting algorithm would be a good method of estimating the parameters if σ^2/ρ^2 were extremely small. It estimates the circular parameters as well as AML1, as long as ρ_a^* and σ_a^2 are used. However for $\sigma^2/\rho^2 = 0.04$, AML1 performs better, giving estimators with smaller variances.

8. Simulation Studies: ρ_3, ρ_4 and σ_4^2

As before we use the same sets of circles that were generated for AML2, and calculate the moment type estimators ρ_3, ρ_4 and σ_4^2 . These estimators do not have a theoretical asymptotic distribution, However the theoretical asymptotic mean and variance properties of ρ_3 and ρ_4 seemed to suggest that they would be suitable estimators. ρ_3 was calculated assuming that the variance, σ^2 , was known to be 1.0 . It became apparent that ρ_3, ρ_4 and σ_4^2 have rather undesirable small sample properties. Their behaviour improved with increasing N but worsened as ρ increased.

ρ_3 is supposed to be asymptotically unbiased with variance

$$\mathrm{Var}(\rho_3) \simeq \sigma^2 N^{-1}(1 + \sigma^2/\rho^2) .$$

For $(N, \rho) = (12, 5)$, $k_1(\rho_3) = 4.976$, which is significantly lower than the expected value of 5 . For $(N, \rho) = (20, 5)$, $k_1(\rho_3) = 4.987$ which is within the 95% confidence limits of the asymptotically expected value. Also unlike the case for $N = 12$, $\gamma_1(\rho_3)$ and $\gamma_2(\rho_3)$ are well within R_1 . No correction need be made to the variance of $k_2(\rho_3)$. $k_2(\rho_3) = 0.0661$ which is significantly higher than the expected value of 0.0520 . As the ratio σ^2/ρ^2 decreases the situation worsens. For $(N, \rho) = (20, 20)$, $k_1(\rho_3) = 19.976$ which is significantly low, $\gamma_1(\rho_3) = -1.026$ and $\gamma_2(\rho_3) = 1.24$. The distribution of ρ_3 is exhibiting negative skewness and positive kurtosis and a correction to the variance of $k_2(\rho_3)$ has to be made here. The observed value is $k_2(\rho_3) = 0.2941$ whereas the expected value is 0.0501 . The standard deviation of $k_2(\rho_3)$ is thus $(\gamma_2+2)^{\frac{1}{2}} * 0.0501 * 0.05$. We can see that $k_2(\rho_3)$ is obviously much greater than the predicted value. The trend is for $|\gamma_1(\rho_3)|$ and $|\gamma_2(\rho_3)|$ to decrease for fixed ρ as N increases, but increase for fixed N as ρ increases. At $N = 75$, when $\rho = 5$, $(\gamma_1(\rho_3), \gamma_2(\rho_3)) = (-0.079, -0.097)$, when $\rho = 10$ it is $(-0.330, 0.422)$ and when $\rho = 20$ it is $(-1.07, 2.73)$. The $k_1(\rho_3)$ statistics for each of these ρ values are 4.999 , 9.999 and 19.999 respectively. All of these are close to their asymptotically expected values. The $k_2(\rho_3)$ statistics are 0.0143, 0.0178 and 0.0330 respectively where the expected values are 0.0139, 0.0135 and 0.0134 respectively. Only the first of these observed k_2 statistics is not significantly higher than expected, even after making the correction to the standard deviation for non-zero γ_2 . ρ_3 is only useful as an estimator if N is large and ρ is small.

The same trend is apparent for ρ_4 except that it has significant skew and kurtosis for $\rho = 5$ and N as high as 50 , and significant skewness for $\rho = 5$ and $N = 75$. At $(N, \rho) = (50, 5)$, $k_1(\rho_4) = 5.017$, $\gamma_1(\rho_4) = -0.276$ and $\gamma_2(\rho_4) = 0.502$. The expected value of $k_2(\rho_4)$ is 0.0211 , whereas the observed value is 0.302 , which is significantly higher than expected. At $(N, \rho) = (75, 5)$, $k_1(\rho_4) = 5.006$, $\gamma_1(\rho_4) = -0.31$ and $\gamma_2(\rho_4) = 0.24$. The expected value of $k_2(\rho_4)$ is 0.0140 whereas the observed value is 0.0192 , again significantly higher than expected.

The situation for σ_4^2 is even worse. Even for $(N, \rho) = (75, 5)$, $\left(\gamma_1\left(\sigma_4^2\right), \gamma_2\left(\sigma_4^2\right)\right) = (0.93, 1.89)$, so that σ_4^2 is still positively skewed and kurtic. $k_1\left(\sigma_4^2\right) = 0.9620$ indicating that if biasedness is considered, σ_4^2 is much better than most estimators of the variance for this N and ρ . The observed $k_2\left(\sigma_4^2\right)$ value is 0.0771 which is quite high compared with the variances of the other estimators of σ^2 . As ρ increases the situation worsens. For $(N, \rho) = (75, 20)$, $k_1\left(\sigma_4^2\right) = 0.9474$, $k_2\left(\sigma_4^2\right) = 9.592$, $\gamma_1\left(\sigma_4^2\right) = 1.77$, and $\gamma_2\left(\sigma_4^2\right) = 5.29$. Although σ_4^2 does not appear to be very biased its variance is enormous, and its distribution is obviously non-normal. As indicated previously, the situation worsens as N decreases. Indeed for $(N, \rho) = (12, 20)$, $k_1\left(\sigma_4^2\right) = 0.148$, $k_2\left(\sigma_4^2\right) = 395.9$, $\gamma_1\left(\sigma_4^2\right) = 2.21$ and $\gamma_2\left(\sigma_4^2\right) = 8.76$.

Unless N is very large, ρ_4 and σ_4^2 are unsuitable estimators of ρ

and σ^2 , especially if the ratio σ^2/ρ^2 is very small. One possible explanation for this is that the estimators are derived from equations such as

$$E(D) = \rho^2 + 2\sigma^2$$

and

$$E(S) = 3\rho^4 + 16\rho^2\sigma^2 + 16\sigma^4 .$$

That is

$$\rho_4 = \left(4D^2-S\right)^{\frac{1}{4}} ,$$

and

$$\sigma_4^2 = \tfrac{1}{2}\left(D-\rho_4^2\right) .$$

If σ^2/ρ^2 is small, then σ_4^2 is found by taking the difference of two relatively large numbers, and is thus prone to error. These estimators are probably best if ρ^2 and σ^2 are of similar orders of magnitude. Although the estimators are asymptotically unbiased with the required variances, if the ratio σ^2/ρ^2 is small asymptotically here does mean for very large N . This probably explains the negative estimates of σ_4^2 obtained for the Ring of Brogar, in what looks like an ideal estimation problem.

9. Simulation Studies: Other Estimation Procedures

One estimation procedure not yet considered is FR2, and here we will compare FR2 with AML1 and AML2. We will not consider FR1 here. FR1 is an approximation to FR2 which is a slight more crude approximation than AML1 is to AML2. AML2 makes the approximation $I_1/I_0 \simeq 1 - \tfrac{1}{2}z^{-1}$, whereas AML1 uses $I_{1,0}(z) \simeq 1$; on the other hand FR2 makes the approximation $I_2/I_1 \simeq 1 - z^{-1}$, $I_{1,0}(z) \simeq 1 - \tfrac{1}{2}z^{-1}$, whereas FR1 uses $I_{2,1} \simeq 1$ and $I_{1,0} \simeq 1$. In theory we would expect to compare FR1 with AML1 and FR2 with AML2, but this does not appear to be the case, as will become evident. Only $\rho = 20$ and $\rho = 10$ were considered, and for the moment we will concentrate on the case $\rho = 20$. One difficulty that did occur was that AML2 was very slow to converge if $\tau_i \overset{d}{\sim} R[0, \pi)$ or $\tau_i \overset{d}{\sim} R[0, \pi/2)$, and comments on this will be made as we proceed.

If the distribution of τ_i was really $R[0, 2\pi)$ then the estimators obtained using FR2, and their variances and higher sample statistics, were very similar to those obtained using AML2. The results using FR2 are thus also similar to those obtained using AML1.

If the distribution of τ_i was really $R[0, 3\pi/2)$, the statistics for FR2 were very similar to those for AML1, except that ρ under AML1 is more biased. For example for $N = 75$, $k_1(\hat{\rho}) = 20.031$ whereas $k_1(\rho_f) = 20.007$. In this respect FR2 is preferable to AML1 since it converges almost as fast as AML1. For N small the estimators had slightly higher variances under FR2, and AML1, than under AML2, the difference decreasing as N increased. Also for $N = 12$, ξ_f, η_f and ρ_f had significant skew and kurtosis statistics whereas the estimators from AML2 did not. Moreover the convergence of AML2 is considerably slower and for $N \geq 20$ the gain in accuracy would not be worth the larger amount of work required to obtain the estimators. Other convergence problems also arise for AML2. For example the matrix, inverted to find the updated estimators from the previous values, can become ill-conditioned, that is its determinant can become very small, or the procedure may fail to converge in a specified number of iterations.

If instead $\tau_i \overset{d}{\sim} R[0, \pi)$ great difficulties occurred in getting AML2 to converge to the same degree of accuracy as FR2 or AML1 in say 100 iterations. For this reason AML2 was rejected as a reasonable procedure to use if $\tau_i \overset{d}{\sim} R[0, \pi)$. The same problem arose for $\tau_i \overset{d}{\sim} R[0, \pi/2]$. For this reason FR2 was compared with AML1. In fact the results for the two estimation procedures for $\tau_i \overset{d}{\sim} R[0, \pi)$ were remarkably similar, although the k_2 statistics for FR2 may be marginally smaller, and of course the estimation of ρ by FR2 is much less biased. Of course the k_2 statistics are not the same as they would be if $\tau_i \overset{d}{\sim} R[0, 2\pi)$. For example $k_2(\xi_f) = 0.027$, $k_2(\eta_f) = 0.159$, $k_2(\rho_f) = 0.079$ and $k_2(\sigma_f^2) = 0.024$, for $N = 75$ whereas the associated values for AML2 when $\tau_i \overset{d}{\sim} R[0, 2\pi)$ are $k_2(\hat{\xi}) = 0.028$, $k_2(\hat{\eta}) = 0.029$, $k_2(\hat{\rho}) = 0.013$ and $k_2(\hat{\sigma}^2) = 0.026$. Thus we can see that the variances of the estimators of η and ρ have increased dramatically. If $\tau_i \overset{d}{\sim} R[0, \pi/2)$ then we have convergence problems for FR2 and for small N. For $N = 12$, about 40 percent of the FR2 iterations either failed to converge or became ill-conditioned whereas only 5 percent of the AML1 iterations had similar problems. For $N = 75$ all the AML1 iterations converged whereas about 11 percent of the FR2 iterations ran into difficulties. Even though ξ, η and ρ are more biased under AML1, this estimation procedure is preferred at least for N large. For N less than or equal to 20 all the statistics are so grossly non-normal that very little can be deduced. However for larger N, although the estimators are still non-normal, the AML1 estimators have much smaller variances. Also when $N = 75$, all the γ_1 statistics of the FR2 estimators have modulus greater than 4 and all the γ_2 statistics are greater than 30 indicating marked skewness and kurtosis properties, whereas for AML1, $|\gamma_1|$ is around 0.4 and γ_2 is not significant for any of the four estimators. Again the k_2 statistics are much higher than they would be if $\tau_i \overset{d}{\sim} R[0, 2\pi)$, indicating that the variances of the estimators is much larger. For example when $N = 75$, under AML1, $k_2(\hat{\xi}) = 1.051$, $k_2(\hat{\eta}) = 1.015$, $k_2(\hat{\rho}) = 1.647$ and $k_2(\hat{\sigma}^2) = 0.024$. The variances are much greater than they would be if $\tau_i \overset{d}{\sim} R[0, \pi)$.

As pointed out by Angell and Barber (1977), if we try to estimate circular parameters from a small arc segment we will run into difficulties. The variance estimate will not vary much for a range of radii values, and we would expect that any likelihood surface would be fairly flat for a range of radii values making convergence difficult. Similarly the variances of most other estimators would be fairly large, and we would not expect good estimation under any scheme. We have already indicated that if $\tau_i \overset{d}{\sim} R[0, \pi/2)$ the exact likelihood is difficult to handle and is probably of not much use. We would conclude that great care needs to be taken in the use of any estimators estimated from short circular arcs.

For $\tau_i \overset{d}{\sim} R[0, 3\pi/2)$ and $\tau_i \overset{d}{\sim} R[0, \pi)$ when $\rho = 10$, FR2 is again the preferred estimation method. If $\tau_i \overset{d}{\sim} R[0, \pi/2)$, FR2 has even greater convergence failure problems. Again for small N , the k_2 statistics of both AML1 and FR2 procedures are very large. Also ρ under AML1 is quite biased, even for $N = 40$; $k_1(\hat{\rho}) = 10.52$, where $k_2(\hat{\rho}) = 5.26$, $\gamma_1(\hat{\rho}) = 1.54$ and $\gamma_2(\hat{\rho}) = 6.62$. Great care would need to be taken in making conclusions for any of the estimation procedures.

10. Simulation Studies: Conclusion

It could not be maintained that we have carried out a definitive or exhaustive study of the possible estimators for estimating the parameters of the circular model. However we have shown that in most practical cases estimators like ρ_3, ρ_4 and σ_4^2 have very undesirable properties. FR2 is not going to be useful in estimating the parameters if only a segment of the circle is present, as was hoped, but if a half of the circle or more is present FR2 is better than AML1, having a less biased estimate of ρ , and is better than AML2 since it converges much faster. The approximations AML1 and AML2 did not throw up any nasty surprises, and Angell and Barber's circle fitting algorithm performed quite well as long as σ^2/ρ^2 was sufficiently small. Also we gained a rough idea of how big N needs to be before we can assume that the estimators are approximately normally distributed. This value of N could be smaller than that indicated, since the study was fairly conservative. One remaining problem is how to ensure that the variance σ^2 is not underestimated. The estimate of σ^2 may need to be inflated slightly to eliminate the negative bias, but this will not be examined here except to point out that if the estimate of σ^2 is inflated, suitable adjustments to the variance of the estimator need to be made.

APPENDIX A

MORE BOUNDS FOR δ AND β , $m = 2$

We now attempt to improve the bounds for δ given by equation (4.24a). The difficulty in dealing with the integral defining δ , (4.22) is the factor $I_{1,0}(z)$. Using the relationship from Watson [(1944), p. 77],

$$I_\nu(z) = \exp\left(-\tfrac{1}{2}\pi\nu i\right)J_\nu[z\ \exp(\tfrac{1}{2}i\pi)]\ , \quad -\pi < \arg z < \pi/2\ ,$$

we can rewrite the equation in Magnus, Oberhettinger and Soni [(1966), p. 147] to obtain

$$I_{1,0}(z) = 2z \sum_{n=1}^{\infty} \left(z^2+v_{0,n}^2\right)^{-1} \tag{A1}$$

where $v_{0,n}$ are the zeros of $J_0(z)$ in the half plane $z > 0$, arranged according to non-decreasing real parts. From Watson [(1944), p. 483] the $v_{0,n}$ are all real.

Substitution of (A1) into (4.22) gives

$$\delta = 2\sigma^{-2}\exp(-\gamma) \sum_{n=1}^{\infty} \int_0^{\infty} r^2\left[r^2+\alpha_n^2\right]^{-1} \exp\left(-\tfrac{1}{2}\sigma^{-2}r^2\right)I_1\left(r\rho/\sigma^2\right)dr \tag{A2}$$

where $\gamma = \rho^2/2\sigma^2$ and $\alpha_n^2 = v_{0,n}^2\sigma^4/\rho^2$ for all n . By introducing a new variable t , it can be seen that

$$\delta = 2\rho^{-1}e^{-\gamma} \sum_{n=1}^{\infty} \int_0^{\infty} r^2 e^{-\frac{1}{2}\sigma^{-2}r^2} I_1\left(r\rho/\sigma^2\right) \int_0^{\infty} \exp\left\{-\left(r^2+\alpha_n^2\right)t\right\}dt$$

$$= 2\rho^{-1}e^{-\gamma} \sum_{n=1}^{\infty} \int_0^{\infty} \exp\left(-\alpha_n^2 t\right)dt \int_0^{\infty} r^2\exp\{-r^2\left(t+\tfrac{1}{2}\sigma^{-2}\right)\}I_1\left(r\rho/\sigma^2\right)dt\ .$$

The inner integral can be evaluated in the same manner as (4.23) to give

$$\delta = \left(2\sigma^2\right)^{-1}e^{-\gamma} \sum_{n=1}^{\infty} \int_0^{\infty} \exp\left(-\alpha_n^2 t\right)\left(t+\tfrac{1}{2}\sigma^{-2}\right)^{-2} * \exp\{-\tfrac{1}{4}\rho^2\sigma^{-4}\left(t+\tfrac{1}{2}\sigma^{-2}\right)^{-1}\}dt\ .$$

If we let $s = 2\sigma^2\left(t+\tfrac{1}{2}\sigma^{-2}\right)$ and $\kappa_n = \alpha_n^2/2\sigma^2$,

$$\delta = e^{-\gamma} \sum_{n=1}^{\infty} \exp\left(\kappa_n\right) \int_1^{\infty} s^{-2} \exp\left[-\kappa_n s+\gamma s^{-1}\right] ds \ .$$

Expanding $\exp\left(\gamma s^{-1}\right)$ as a power series in s^{-1} we obtain,

$$\delta = e^{-\gamma} \sum_{n=1}^{\infty} \exp\left(\kappa_n\right) \sum_{j=0}^{\infty} \gamma^j/j! \ E_{j+2}\left(\kappa_n\right) \ , \tag{A3}$$

where by definition

$$E_n(x) = \int_1^{\infty} \exp(-xt)t^{-n} dt \ , \quad n = 1, 2, \ldots \ ,$$

is the generalized exponential integral. The double sum in (A3) does not seem to have appeared in the literature. However the following inequality is available from Abramowitz and Stegun [(1972), eq. 5.1.19],

$$e^{-x}(x+n)^{-1} \leq E_n(x) \leq e^{-x}(x+n-1) \ .$$

Thus δ can be bounded above and below by

$$e^{-\gamma} \sum_{n=1}^{\infty} \sum_{j=0}^{\infty} \left(j+2+\kappa_n\right)^{-1} \gamma^j/j! \leq \delta \leq e^{-\gamma} \sum_{n=1}^{\infty} \sum_{j=0}^{\infty} \left(j+1+\kappa_n\right)^{-1} \gamma^j/j! \ . \tag{A4}$$

Initially we study the left hand side of this inequality. If first we try summing over j , by definition of the confluent hypergeometric function we obtain,

$$e^{-\gamma} \sum_{n=1}^{\infty} \left(2+\kappa_n\right)^{-1} {}_1F_1\left[2+\kappa_n; \ 3+\kappa_n; \ \gamma\right] \ . \tag{A5}$$

Difficulties occur in finding an adequate approximation for this confluent hypergeometric function. From Watson [(1944), p. 504], $v_{0,n}$ occur approximately π units apart; $v_{0,1} \simeq \frac{3}{4}\pi$ and as $n \to \infty$, $v_{0,n} \to (n-\frac{1}{4})\pi$. Hence κ_n is much smaller than γ for small n , and much larger for large n , if γ is reasonably large. Thus no approximations of the confluent hypergeometric function are suitable for all n .

An alternative is to sum over n first instead of j . Let $\mu_j^2 = j + 2$, then the left hand side of (A4) becomes

$$e^{-\gamma} \sum_{j=0}^{\infty} \gamma^j/j! \ (4\gamma) \sum_{n=1}^{\infty} \left(4\mu_j^2\gamma+v_{0,n}^2\right)^{-1} \ ,$$

and by (A1) this is just

$$e^{-\gamma} \sum_{j=0}^{\infty} \left(\gamma^j/j!\right)\left(\gamma^{\frac{1}{2}}/\mu_j\right) I_{1,0}\left(2\mu_j\gamma^{\frac{1}{2}}\right) .$$

Using the approximation for $I_{1,0}(z)$ given in Chapter 4, Section 5, we obtain

$$e^{-\gamma} \sum_{j=0}^{\infty} \gamma^j/j! \ \{\gamma^{\frac{1}{2}}(j+2)^{-\frac{1}{2}} - \tfrac{1}{4}(j+2)^{-1} - \gamma^{-\frac{1}{2}}\tfrac{1}{32}(j+2)^{-\frac{3}{2}} \ldots\} . \tag{A6}$$

For the approximation to the upper bound, $(j+2)$ in (A6) is replaced by $(j+1)$. To use the approximation for $I_{1,0}(z)$ we require $2\mu_j\gamma^{\frac{1}{2}}$ to be moderately large, and this will not hold for some values of $A = \sigma^2/\rho^2$ previously considered.

To evaluate (A6) we require sums of the form

$$\sum_{j=0}^{\infty} \left(\gamma^j/j!\right)(j+d)^{-s} .$$

Barnes (1906) gave asymptotic expansions for large $|\gamma|$ for sums of this kind where d is not zero or a negative integer. Since we are only interested in the leading order terms, the only relevant part of Barnes solution is

$$e^{\gamma}\gamma^{-s} \sum_{n=0}^{\infty} c_n \frac{\Gamma(1-s)}{\Gamma(1-s-n)} \gamma^{-n} , \tag{A7}$$

where

$$[-y^{-1}\log(1-y)]^{s-1}(1-y)^{d-1} = \sum_{n=1}^{\infty} c_n(-y)^n , \tag{A8}$$

is used to determine c_n. From Hansen [(1975), 6.12.1 and 52.1.5], c_n can be expressed in terms of a sum of Stirling numbers if s is an integer, and this is used for $s = 1, 2, \ldots$. Otherwise from Erdélyi [(1953b), 19.8 (7)]

$$[-y^{-1}\log(1-y)]^{s-1} = 1 + (s-1)y \sum_{n=0}^{\infty} \psi_n(s-1+n)y^n \tag{A9}$$

where $\psi_n(\cdot)$ are the Stirling polynomials. Explicit expressions for the first seven Stirling polynomials can be found in Nielsen [(1965), §29], for example $\psi_0(x) = \frac{1}{2}$, $\psi_1(x) = (3x+2)/24$ and $\psi_2(x) = x(x+1)/48$. Also by the binomial theorem,

$$(1-y)^{d-1} = 1 - (d-1)y + (d-1)(d-2)y^2/2! - (d-1)(d-2)(d-3)y^3/3! \; \ldots \; . \qquad \text{(A10)}$$

By substitution of (A9) and (A10) into (A8), the c_n's can be determined iteratively for non-integral s . Then each of the successive terms in (A6) can be evaluated using (A7) where $d = 2$. We are only interested in leading terms, since (A6) is only valid for $(2j+4)^{\frac{1}{2}}\rho/\sigma$ not too small. This procedure can then be repeated for the right hand side of (A4) to give

$$1 - 7\gamma^{-1}/8 + 45\gamma^{-2}/128 + \ldots \le \delta \le 1 - 3\gamma^{-1}/8 - 11\gamma^{-2}/128 \; \ldots \; . \qquad \text{(A11)}$$

One surprising aspect is that this bound is wider than the bound given for δ by (4.24a). This indicates that tighter bounds for $E_n(x)$ are required. However (A11) does not fully contain the bounds given by (4.24a), so that the combined inequalities yield a much sharper bound,

$$1 - \tfrac{1}{2}\gamma^{-1} - \gamma^{-2}/8 - 3\gamma^{-3}/16 \; \ldots \le \delta \le 1 - 3\gamma^{-1}/8 - 11\gamma^{-2}/128 \; \ldots \; . \qquad \text{(A12)}$$

If these bounds are compared with the actual δ values then at least for $A \le 0.1$, $\left\{E\left[I_{1,0}\left(r\rho/\sigma^2\right)\right]\right\}^2$ is an excellent approximation to δ if sufficient terms in the expansion are considered, successive terms in the expansion not decreasing very rapidly. This seems to indicate that the preceding analysis was unnecessary, since our rough lower bound for δ is such a good approximation to δ , an upper bound is a bit superfluous. However since the upper and lower bounds in (A12) have a separation of less than $\frac{1}{8}\gamma^{-1}$ these bounds provide a check that no major errors were made in the numerical integration of δ .

Now the integral for β can be analysed in the same manner giving

$$\beta = 4\sigma^2 e^{-\gamma} \sum_{n=1}^{\infty} \exp\left(\kappa_n\right) \sum_{j=0}^{\infty} \left(\gamma^j/j!\right)\left\{E_{j+3}\left(\kappa_n\right) + \gamma E_{j+4}\left(\kappa_n\right)/(1+j)\right\} . \qquad \text{(A13)}$$

Again the expression for β can be bounded above and below using the bounds for $E_{j+3}(\cdot)$ and $E_{j+4}(\cdot)$. Continuing with the analysis as before where $(2j+6)^{\frac{1}{2}}\rho/\sigma$ is not too small, we obtain

$$1 + \gamma^{-1}/8 + 93\gamma^{-2}/128 + \ldots \le \beta/\rho^2 \le 1 + 5\gamma^{-1}/8 - 27\gamma^{-2}/128 \ldots .$$

These bounds are completely contained in the bounds for β/ρ^2 given by (4.25a) and hence we have improved bounds for β . The lower bound is still far too small, since we require $\Delta > 0$. We can improve the lower bound by using the inequality $\Delta > 0$ and the resultant bounds are

$$1 + \tfrac{1}{2}\gamma^{-1} - \gamma^{-2}/4 + \gamma^{-3}/8 \ldots \le \beta/\rho^2 \le 1 + 5\gamma^{-1}/8 - 27\gamma^{-2}/128 \ldots .$$

The upper bound is of little use as an approximation to β/ρ^2 , again indicating that this analysis is of little value except as a check on the numerical integration.

More Bounds for δ and β . General m

The series expansion for $I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)$ is obtained in the same manner as (A1). The equation is the same as that of (A1) except that $v_{0,n}$ are replaced by $v_{\frac{1}{2}m-1,n}$ where these are the zeros of $z^{-\frac{1}{2}m+1}J_{\frac{1}{2}m-1}(z)$ in the half-plane $z > 0$ arranged in increasing order. By Watson [(1944), p. 483] they are real and positive. Letting

$$\alpha_n^2 = v_{\frac{1}{2}m-1,n}^2 \cdot \sigma^4/\rho^4$$

we carry out the analysis as before to obtain

$$\delta = e^{-\gamma} \sum_{n=1}^{\infty} \exp\left(\alpha_n^2/2\sigma^2\right) \sum_{j=0}^{\infty} \left(\gamma^j/j!\right) E_{\frac{1}{2}m+1+j}\left(\alpha_n^2/2\sigma^2\right) ,$$

where we define

$$E_{n+\frac{1}{2}}(x) = \int_1^{\infty} t^{-(n+\frac{1}{2})}\exp(-xt)dt , \quad n = 1, 2, 3, \ldots .$$

For m even, $E_{\frac{1}{2}m+1+j}(x)$ are the familiar generalized exponential integrals for which Hopf (1934) found the bounds used in the previous section. In a similar manner, Hopf's proof can be extended to show that

$$e^{-x}(x+n+\tfrac{1}{2})^{-1} \le E_{n+\frac{1}{2}}(x) \le e^{-x}(x+n-\tfrac{1}{2})^{-1} , \quad n = 1, 2, \ldots .$$

Thus δ can be bounded above and below as before. If we examine the smaller of these bounds we obtain

$$e^{-\gamma} \sum_{j=0}^{\infty} \left(\gamma^j/j!\right) \left(\gamma^{\frac{1}{2}}/\mu_j\right) I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(2\mu_j\gamma^{\frac{1}{2}}\right) \leq \delta$$

where $\mu_j^2 = j + \frac{1}{2}m + 1$. Now using the asymptotic expression for $I_{\frac{1}{2}m,\frac{1}{2}m-1}(z)$ for z large, we can put the above expression into the same form as before. Again using the equation of Barnes (1906) and determining the constants c_n by similar methods as before, we obtain the bounds

$$1 - (4m-1)\gamma^{-1}/8 + \left(32m^2-44m+15\right)\gamma^{-2}/128 - \left(128m^3-496m^2+442m-9\right)\gamma^{-3}/1024$$
$$\leq \delta \leq 1 - (4m-5)\gamma^{-1}/8 + \left(32m^2-124m+109\right)\gamma^{-2}/128$$
$$- \left(128m^3-1040m^2+2602m-1965\right)\gamma^{-3}/1024 .$$

As in the case $m = 2$, the lower bound found by this method is less than $\left\{E\left[I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)\right]\right\}^2$ but the upper bound is less than $E\left[I_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)\right]$. Therefore combining the two bounds we obtain a narrower bound.

If $\frac{1}{2}m >> \gamma$, then after obtaining the bounds for γ , by substituting the bounds for $E_{\frac{1}{2}m+1+j}(\cdot)$, we sum over j first instead of n to obtain

$$e^{-\gamma} \sum_{n=1}^{\infty} \left(\tfrac{1}{2}m+1+\kappa_n\right)^{-1} {}_1F_1\left[\tfrac{1}{2}m+1+\kappa_n;\ \tfrac{1}{2}m+2+\kappa_n;\ \gamma\right] .$$

In this case a suitable approximation for the confluent hypergeometric function exists, but the resulting series would still have to be summed over n , and then an adequate approximation to α_n^2 would have to be found.

β can be evaluated in a similar manner to obtain

$$\beta = \left(2\sigma^2\right)\left(\tfrac{1}{2}m+1\right)e^{-\gamma} \sum_{j=0}^{\infty} \left(\gamma^j/j!\right) \sum_{n=1}^{\infty} \exp\left(\kappa_n\right)$$
$$\times \left[E_{\frac{1}{2}m+2+j}\left(\kappa_n\right) + \gamma/\left(\tfrac{1}{2}m+j\right) E_{\frac{1}{2}m+3+j}\left(\kappa_n\right)\right] .$$

Substituting the bounds for $E_{\frac{1}{2}m+2+j}(\cdot)$ and $E_{\frac{1}{2}m+3+j}(\cdot)$ and evaluating the resultant expressions as before yields

$$1 + \gamma^{-1}/8 + \ldots \leq \beta/\rho^2 \leq 1 + 5\gamma^{-1}/8 - (20m-13)\gamma^{-2}/128 .$$

As in the case $m = 2$, the lower bound for β/ρ^2 is far too small, although it is an improvement on the bound obtained from the inequality

$$\left\{E\left[rI_{\frac{1}{2}m,\frac{1}{2}m-1}\left(r\rho/\sigma^2\right)\right]\right\}^2 \le \beta/\rho^2 \ .$$

As before we can obtain an extremely good lower bound for β/ρ^2 using the inequality $\Delta > 0$, indicating that the above analysis is not good enough. What is needed are more accurate and usable bounds for $E_{n+\frac{1}{2}}(x)$ and $E_n(x)$, so that much tighter bounds for both δ and β/ρ^2 can be derived.

APPENDIX B

CIRCLE DATA

Chan A $\xi = \eta = 0$, $\rho = 10$, $\sigma^2 = 1$, $\tau_i \overset{d}{\sim} R[0, 2\pi)$, $N = 12$

x	-10.219	-8.615	-9.901	5.226	8.330	6.622
y	-2.984	-5.935	1.993	5.952	-2.968	5.237
x	9.220	0.771	-9.328	1.229	5.733	-3.016
y	3.683	-9.118	-4.632	9.949	8.513	9.286

ref: Chan (1965)

Chan B $\xi = \eta = 0$, $\rho = 20$, $\sigma^2 = 1$, $\tau_i \overset{d}{\sim} R[0, \pi/2)$, $N = 12$

x	4.287	19.889	5.372	1.611	20.744	5.703
y	18.856	7.158	18.711	20.396	-0.070	19.733
x	16.154	21.702	19.360	10.699	3.057	17.211
y	11.128	3.279	6.980	16.700	19.752	13.220

ref: Chan (1965)

Set C $\xi = \eta = 0$, $\rho = 10$, $\sigma^2 = 1$, $\tau_i \overset{d}{\sim} R[0, 2\pi)$, $N = 20$

x	-2.4935	-10.2950	9.3715	5.0234	5.3855	-4.5053	
y	8.6590	-0.7254	3.0871	9.4207	9.6710	7.8987	
x	-9.6905	-10.0125	3.9712	-9.3855	-5.0392	-7.4598	-3.0047
y	-1.6878	-4.7040	-7.8059	5.6540	8.9226	8.9872	8.9272
x	6.5473	-1.7604	7.1042	-3.4269	12.0623	3.0074	-10.5866
y	-6.8319	10.5623	6.7061	-10.6143	1.6118	-10.2697	1.1602

Brogar. ref: Thom and Thom (1973).

Point data actually given in the form β and r^* where $r^* = (x^2+y^2)^{\frac{1}{2}} - 170$ and $\theta = (6n+\beta)$, (x, y) obtained by cartesian co-ordinate conversion, and given in feet.

Stone	β	r^*	x	y
1	0.4	1.0	169.93	19.06
2	-0.5	0.5	167.08	33.99
3	-0.9	-0.5	162.01	49.84
{7	-0.2	-6.0	122.26	109.31
8	-0.8	-4.6	112.38	121.36
10	1.5	1.0	81.59	150.28
12	-0.3	1.3	53.79	162.64
14	-0.6	0.2	19.56	169.07
15	0.4	0.4	-1.19	170.40
16	0.9	0.3	-20.46	169.07
18	1.7	?	-	-
19	-1.4	0.9	-65.68	157.78
20	-4.1	0.4	-74.43	153.28
22	1.8	0.0	-117.66	122.70
24	-1.6	1.4	-136.16	104.10
28	-1.2	-2.9	-162.69	38.16
29	1.8	-0.1	-169.44	12.44
30	1.0	2.0	-171.97	-3.00
32	0.7	0.0	-165.84	-37.37
33	0.5	-0.6	-160.65	-53.75
34	0.7	-1.8	-152.81	-70.29
36	0.6	-1.7	-135.11	-100.34
37	0.5	-1.6	-124.16	-113.77
38	1.2	-1.4	-110.17	-127.63
39	1.4	-0.6	-96.19	-139.44
40	0.1	?	-	-
41	1.0	-0.4	-66.27	-156.12
44	0.7	1.3	-15.82	-170.57
45	0.2	2.1	6.01	-172.10
46	0.4	1.4	19.11	-170.33
47	0.7	1.7	37.75	-167.50
48	0.7	2.8	55.40	-163.68
49	1.2	0.6	72.64	-154.36
50	1.6	-0.4	88.87	-144.45
53	-1.1	0.9	124.78	-116.77
56	-2.0	-0.4	152.44	-74.35
58	0.1	2.6	168.89	-35.59

Avebury. Ref: Thom, Thom and Foord (1976)

Centre C

Stone	x	y
1	733.7	44.0
3	659.7	28.0
4	624.2	19.3
5	588.4	13.9
6	551.6	12.3
7	515.1	9.5
8	478.0	16.6

Centre W

Stone	x	y
9	445.3	23.4
10	413.8	46.2
11	377.9	74.1
12	357.1	94.1
13	327.7	112.4
14	300.6	136.2
15	272.0	158.8
16	243.5	183.0
17	216.3	205.0
18	188.9	229.8
19	163.5	255.5
20	140.0	285.0
21	120.6	305.7
22	103.1	323.1
23	85.9	344.0
24	61.8	371.3

Centre A

Stone	x	y
30	19.3	624.4
31	24.9	663.0
32	33.3	698.3
33	43.7	731.3
34	55.5	764.4
35	62.9	790.1
36	69.2	815.0
37	85.0	849.8
38	98.5	884.6
39	123.6	910.5

Centre B

Stone	x	y
40	146.8	936.9
41	175.2	962.4
42	206.7	984.7
43	237.6	1002.9
44	270.3	1022.5
45	292.5	1031.2
46	315.8	1042.0

3 additional stones

50	461.1	1085.4
68	1033.4	946.2
98	769.9	64.9

See also Figure 1 of Thom *et al.* From this figure it can be seen that stone number 98 is adjacent to stone number 1 , and it could be argued that stone 98 belongs to the arc with centre C . The breakup of the stones into the various groups appears to have been achieved by presupposing that the arc lengths produced were close to an integral number of rods, a megalithic rod being 2.5 megalithic yards. The (x, y) co-ordinates are given in feet, as they are in the paper of Thom *et al.* The paper also describes which stones are upright, and which were discovered in 1936 by excavation and marked by concrete plinths.

APPENDIX C

If X is infinitely divisible then it is not true in general that $Y \overset{d}{\sim} 1/X$ is also infinitely divisible. Thorin (1977a) proved that the Pareto distribution is infinitely divisible. The Pareto distribution has probability density function

$$f_X(x) = ak^a x^{-a-1} \ , \quad a > 0 \ , \quad x \geq k > 0 \ .$$

If we let $Y \overset{d}{\sim} 1/X$ then $0 \leq Y \leq k^{-1}$, and since Y has finite support it cannot be infinitely divisible. However there seem to be some positive random variables which are both infinitely divisible and closed under convolution, whose reciprocals are also infinitely divisible. For example if X has an inverse Gaussian distribution (see Johnson and Kotz [(1970), p. 139]) then X is infinitely divisible and closed under convolution by inspection of the characteristic function. If Y is the reciprocal of X, where X has an inverse Gaussian distribution then by inspection of the characteristic function (see Johnson and Kotz [(1970), p. 149]), Y is also infinitely divisible.

The gamma distribution is known to be both infinitely divisible and closed under convolution. $Y \overset{d}{\sim} \left(\chi_k^2\right)^{-1}$ was proved to be infinitely divisible as a part of the proof that the t-distribution is infinitely divisible. For references see chapter 3, section 2. It is not difficult to extend this proof to include $Y \overset{d}{\sim} 1/X$ where X is a gamma distribution.

If we were interested in proving any such conjecture for positive random variables which are both infinitely divisible and closed under convolution, we would have to restrict ourselves to considering variables for which $F(0^+) = 0$. This eliminates both the Poisson and negative binomial distribution from consideration. However we will not prove any such conjecture, but just indicate an interesting phenomenon which might be worthy of further study, and at the same time indicating how useful the Laplace transform is in establishing infinite divisibility.

Bessel Function Distribution. If X has the Bessel function distribution,

$$f_X(x) = e^{-x}(\rho/x)I_\rho(x) \ , \quad \rho > 0 \ .$$

This is infinitely divisible (see Feller (1966)). If $Y \overset{d}{\sim} 1/X$ then it is easy to show that

$$f_Y(y) = y^{-2} f_X\left(y^{-1}\right) ,$$

and the Laplace transform of y is given by

$$\begin{aligned} L_Y(t) &= E\left(\exp(-ty)\right) \\ &= \int_0^\infty \exp(-ty)\exp\left(y^{-1}\right)(\rho/y) I_\rho\left(y^{-1}\right) dy . \end{aligned}$$

Using Oberhettinger and Badii [(1973), p. 159, eq. 15.55]

$$L_Y(t) = 2\rho I_\rho\left((2t)^{\frac{1}{2}}\right) K_\rho\left((2t)^{\frac{1}{2}}\right) .$$

We will show that Y is infinitely divisible using the following theorem from Feller [(1971), p. 450] or alternatively Schoenberg (1938). The function $L(t)$ is the Laplace transform of an infinitely divisible distribution if and only if $L(t) = \exp\left(-h(t)\right)$ where h is a function such that $h(0) = 0$, and h has a completely monotone derivative, that is

$$(-1)^n \frac{d^n h'(t)}{dt^n} \geq 0 \quad \text{for} \quad n = 1, 2, \ldots .$$

Here

$$h'(t) = \frac{L_Y'(t)}{L_Y(t)} = t^{-\frac{1}{2}} \frac{I_\rho'\left((2t)^{\frac{1}{2}}\right)}{I_\rho\left((2t)^{\frac{1}{2}}\right)} + t^{-\frac{1}{2}} \frac{K_\rho'\left((2t)^{\frac{1}{2}}\right)}{K_\rho\left((2t)^{\frac{1}{2}}\right)} ,$$

and $t^{-\frac{1}{2}} K_\rho'\left[(2t)^{\frac{1}{2}}\right] / K_\rho\left[(2t)^{\frac{1}{2}}\right]$ is completely monotonic; this was proved as a part of the proof that $\left(\chi_k^2\right)^{-1}$ is infinitely divisible (see for example Ismail (1977)). Consequently it is sufficient to prove that $t^{-\frac{1}{2}} I_\rho'\left[(2t)^{\frac{1}{2}}\right] / I_\rho\left[(2t)^{\frac{1}{2}}\right]$ is completely monotonic.

$$I_\rho'\left[(2t)^{\frac{1}{2}}\right] = I_{\rho+1}\left[(2t)^{\frac{1}{2}}\right] + \rho(2t)^{-\frac{1}{2}} I_\rho\left[(2t)^{\frac{1}{2}}\right] ,$$

therefore

$$t^{-\frac{1}{2}} I_\rho'\left[(2t)^{\frac{1}{2}}\right] / I_\rho\left[(2t)^{\frac{1}{2}}\right] = t^{-\frac{1}{2}} I_{\rho+1}\left[(2t)^{\frac{1}{2}}\right] / I_\rho\left[(2t)^{\frac{1}{2}}\right] + \rho/\left(2^{\frac{1}{2}} t\right) ,$$

where $\rho/(2^{\frac{1}{2}}t)$ is completely monotonic. Hence it suffices to prove that $t^{-\frac{1}{2}}I_{\rho+1}[(2t)^{\frac{1}{2}}]/I_{\rho}[(2t)^{\frac{1}{2}}]$ is completely monotonic. From Magnus, Oberhettinger and Soni [(1966), p. 147] we can derive

$$I_{\nu+1}(z)/I_{\nu}(z) = 2z\sum_{n=1}^{\infty}\left(z^2+\alpha_{\nu,n}^2\right)^{-1} ,$$

where $\alpha_{\nu,n}$ are the zeros of $z^{-\nu}J_{\nu}(z)$ in the half plane $z > 0$, arranged according to non-decreasing real parts. From Watson [(1944), p. 483] these $\alpha_{\nu,n}$ are all real for $\nu > -1$. Thus

$$t^{-\frac{1}{2}}I_{\rho+1}\left(t^{\frac{1}{2}}\right)/I_{\rho}\left(t^{\frac{1}{2}}\right) = 2\sum_{n=1}^{\infty}\left(t+\alpha_{\rho,n}^2\right)^{-1} ,$$

and since $\alpha_{\rho,n}$ are real, $t+\alpha_{\rho,n}^2$ is positive for positive t, and obviously

$$(-1)^k\frac{d^k}{dt^k}\left[t^{-\frac{1}{2}}I_{\rho+1}\left(t^{\frac{1}{2}}\right)/I_{\rho}\left(t^{\frac{1}{2}}\right)\right] = 2\sum_{n=1}^{\infty}k!\left(t+\alpha_{\rho,n}^2\right)^{-k-1} > 0 .$$

We have proved that Y is infinitely divisible.

Non-central chi-squared distribution. If X has a non-central chi-squared distribution then X has probability density function

$$f_X(x) = \tfrac{1}{2}(x/\lambda)^{\frac{1}{4}(v-2)}I_{\frac{1}{2}(v-2)}[(\lambda x)^{\frac{1}{2}}]\exp\left(-(\lambda+x)/2\right) .$$

Let $\rho = \frac{1}{2}(v-2)$ and $\alpha = \lambda/2$, then

$$f_X(x) = \tfrac{1}{2}(x/2\alpha)^{\rho/2}I_{\rho}[(2\alpha x)^{\frac{1}{2}}]\exp[-\alpha - x/2] .$$

From Oberhettinger and Badii [(1973), p. 153, eq. 15.24] this has Laplace transform

$$L_X(t) = (t+1)^{-\rho-1}\exp\{-\alpha + \alpha/(t+1)\} .$$

Feller [(1971), p. 438] also gave this Laplace transform. It can be written as

$$L_X(t) = \exp\{-\alpha + \alpha/(t+1) - (\rho+1)\log_e(t+1)\} .$$

If $L_X(t) = \exp\{-h(t)\}$, then

$$h(t) = \alpha - \alpha/(t+1) + (\rho+1)\log_e(t+1)$$

and

$$h(0) = 0 \ .$$

The derivative of h is

$$h'(t) = \alpha(t+1)^{-2} + (\rho+1)(t+1)^{-1} \ .$$

Therefore

$$(-1)^k d^k h'(t)/dt^k = \alpha(k+1)!(t+1)^{-k-2} + (\rho+1)k!(t+1)^{-k-1} > 0 \ ,$$

and X is infinitely divisible. If Y is the reciprocal of X then

$$L_Y(t) = E[\exp(-t/X)] \ .$$

If the Bessel function is expanded as an infinite series and integration is performed term-by-term then using Oberhettinger and Badii [(1973), p. 41, eq. 5.34], we obtain

$$L_Y(t) = (2\alpha)^{-\rho}(2t)^{\frac{1}{2}}e^{-\alpha} \sum_{m=0}^{\infty} \frac{(\alpha t/2^{\frac{1}{2}})^{m+\rho}}{m!\Gamma(m+\rho+1)} K_{m+\rho+1}[(2t)^{\frac{1}{2}}] \ ,$$

where $\rho = \frac{1}{2}(v-2)$ and $\alpha = \lambda/2$. Unfortunately this Laplace transform does not seem to be available in closed form so we will not proceed any further. Note that $K_\nu(x)$ is defined in Watson [(1944), p. 78].

In the examples given above the Laplace transforms are much easier to find than the characteristic functions, and it is far easier to establish infinite divisibility using theorems on the Laplace transform rather than attempting to prove that the necessary and sufficient conditions for the characteristic function hold.

APPENDIX D

Some of the results of the simulation studies described in chapter 6 are given here. Only the k_1 and k_2 statistics for the estimators of ξ and η are given since their γ_1 and γ_2 statistics are fairly uninteresting. To compare k_1 and k_2 sample statistics with their asymptotic values we recall some comments made in chapter 6, section 6. If the distribution of an estimator $\tilde{\theta}$ has finite mean and variance then by the central limit theorem $k_1(\tilde{\theta})$ will have a distribution which is close to normal. If $\mathrm{sd}(\tilde{\theta})$ is the theoretical standard deviation of the distribution of $\tilde{\theta}$ then

$$\mathrm{sd}\ k_1(\tilde{\theta}) = \mathrm{sd}(\tilde{\theta})M^{-\frac{1}{2}} = \sqrt{2}\ \mathrm{sd}(\tilde{\theta})/40 ,$$

since $M = 800$. In addition

$$\mathrm{sd}\ k_2(\tilde{\theta}) \simeq \left(\gamma_2+2\right)^{\frac{1}{2}}\mathrm{sd}(\tilde{\theta})M^{-\frac{1}{2}} = \left(1 + \gamma_2/2\right)^{\frac{1}{2}}\mathrm{sd}(\tilde{\theta})/20 ,$$

where γ_2 is the kurtosis of the distribution of $\tilde{\theta}$. The asymptotic values for $k_2(\tilde{\theta})$ have been given previously. For example from chapter 4, $N\ \mathrm{Var}(\hat{\xi}) = 2.046$ if $\rho = 5$, 2.010 if $\rho = 10$ and 2.0025 if $\rho = 20$. Similarly $N\ \mathrm{Var}(\hat{\rho}) = 1.04$ if $\rho = 5$, 1.01 if $\rho = 10$ and 1.00 if $\rho = 20$, and $N\ \mathrm{Var}\left(\hat{\sigma}^2\right) = 2.10$ if $\rho = 5$, 2.01 if $\rho = 10$ and 2.00 if $\rho = 20$. The asymptotic values of the k_2 statistics for Angell and Barber's estimators can be found in chapter 6, section 4. The basic test for normality used is described in chapter 6, section 6, where rectangle R_1 is

$$\left\{-0.202 \le \gamma_1(\tilde{\theta}) \le 0.202,\ -0.35 \le \gamma_2(\tilde{\theta}) \le 0.46\right\} .$$

Only the results for $\tau_i \overset{d}{\sim} R[0,\ 2\pi)$ have been tabulated here. For $\tau_i \overset{d}{\sim} R[0,\ 3\pi/2)$ or $R[0,\ \pi)$ or $R[0,\ \pi/2)$ the trends have been described in chapter 6, section 9. The actual numerical results are not very interesting in themselves and it was decided to omit them.

	$\tilde{\theta}$	AML2 $k_1(\tilde{\theta})$	AML2 $k_2(\tilde{\theta})$	AML1 $k_1(\tilde{\theta})$	AML1 $k_2(\tilde{\theta})$	A & B $k_1(\tilde{\theta})$	A & B $k_2(\tilde{\theta})$
N = 12							
ρ = 5	ξ	-0.025	0.213	-0.002	0.831	-0.027	0.231
	η	-0.018	0.225	0.021	0.240	-0.017	0.243
ρ = 10	ξ	-0.020	0.208	-0.022	0.210	-0.023	0.215
	η	-0.016	0.215	-0.017	0.218	-0.016	0.222
ρ = 20	ξ	-0.019	0.208	-0.020	0.208	-0.021	0.210
	η	-0.014	0.211	-0.015	0.212	-0.015	0.215
N = 20							
ρ = 5	ξ	-0.015	0.111	-0.016	0.116	-0.014	0.123
	η	-0.012	0.130	-0.010	0.133	-0.010	0.145
ρ = 10	ξ	-0.013	0.108	-0.013	0.109	-0.013	0.112
	η	-0.012	0.125	-0.011	0.125	-0.012	0.130
ρ = 20	ξ	-0.012	0.108	-0.012	0.108	-0.012	0.109
	η	-0.012	0.123	-0.011	0.122	-0.012	0.124
N = 30							
ρ = 5	ξ	-0.009	0.073	-0.009	0.077	-0.010	0.083
	η	-0.013	0.081	-0.012	0.086	-0.010	0.091
ρ = 10	ξ	-0.005	0.073	-0.005	0.074	-0.006	0.076
	η	-0.012	0.080	-0.011	0.081	-0.010	0.083
ρ = 20	ξ	-0.003	0.073	-0.003	0.074	-0.004	0.074
	η	-0.010	0.079	-0.010	0.080	-0.010	0.080
N = 40							
ρ = 5	ξ	-0.001	0.056	0.000	0.060	0.001	0.064
	η	-0.011	0.055	-0.010	0.058	-0.009	0.064
ρ = 10	ξ	0.002	0.056	0.003	0.057	0.002	0.058
	η	-0.011	0.054	-0.010	0.054	0.010	0.057
ρ = 20	ξ	0.004	0.056	0.004	0.057	0.004	0.057
	η	-0.010	0.053	-0.010	0.053	-0.010	0.054
N = 50							
ρ = 5	ξ	0.004	0.042	0.004	0.045	0.004	0.049
	η	-0.011	0.044	-0.010	0.047	-0.007	0.050
ρ = 10	ξ	0.006	0.043	0.006	0.044	0.006	0.045
	η	-0.010	0.043	-0.010	0.044	-0.008	0.045
ρ = 20	ξ	0.007	0.044	0.007	0.044	0.007	0.044
	η	-0.010	0.043	-0.009	0.043	-0.009	0.043
N = 75							
ρ = 5	ξ	0.002	0.029	0.001	0.030	0.000	0.033
	η	-0.004	0.029	-0.003	0.031	0.001	0.032
ρ = 10	ξ	0.003	0.028	0.003	0.029	0.002	0.029
	η	-0.003	0.029	-0.002	0.030	-0.001	0.030
ρ = 20	ξ	0.004	0.028	0.004	0.028	0.003	0.028
	η	-0.003	0.029	-0.002	0.029	-0.001	0.029

TABLE D.1

$N = 12$	$\tilde{\theta}$	AML2 $\hat{\rho}$	AML1 $\hat{\rho}$	A & B ρ_a	A & B ρ_a^*
$\rho = 5$	$k_1(\tilde{\theta})$	4.999	5.134	5.146	5.071
	$k_2(\tilde{\theta})$	0.111	0.382	0.101	0.116
$\rho = 10$	$k_1(\tilde{\theta})$	9.991	10.048	10.063	10.024
	$k_2(\tilde{\theta})$	0.108	0.106	0.105	0.118
$\rho = 20$	$k_1(\tilde{\theta})$	19.988	20.016	20.024	20.003
	$k_2(\tilde{\theta})$	0.106	0.107	0.106	0.119
$N = 20$					
$\rho = 5$	$k_1(\tilde{\theta})$	5.007	5.114	5.174	5.095
	$k_2(\tilde{\theta})$	0.056	0.053	0.052	0.057
$\rho = 10$	$k_1(\tilde{\theta})$	10.003	10.056	10.086	10.047
	$k_2(\tilde{\theta})$	0.055	0.054	0.054	0.058
$\rho = 20$	$k_1(\tilde{\theta})$	20.001	20.028	20.043	20.025
	$k_2(\tilde{\theta})$	0.055	0.055	0.054	0.058
$N = 30$					
$\rho = 5$	$k_1(\tilde{\theta})$	5.014	5.117	5.190	5.104
	$k_2(\tilde{\theta})$	0.035	0.033	0.032	0.035
$\rho = 10$	$k_1(\tilde{\theta})$	10.011	10.063	10.099	10.055
	$k_2(\tilde{\theta})$	0.034	0.033	0.033	0.035
$\rho = 20$	$k_1(\tilde{\theta})$	20.010	20.036	20.054	20.032
	$k_2(\tilde{\theta})$	0.034	0.034	0.033	0.035

TABLE D.2

The sample statistics for AML2 $\hat{\rho}$ already conform fairly closely to the asymptotic expected values for $\hat{\rho}$. For larger N the variances of all four estimators are almost the same, but the bias apparent in the k_1 statistics for AML1 $\hat{\rho}$, ρ_a and ρ_a^* for $N = 30$ are maintained for all other values of N studied.

$N = 12$		$\tilde{\theta}$	$k_1(\tilde{\theta})$	$k_2(\tilde{\theta})$	$\gamma_1(\tilde{\theta})$	$\gamma_2(\tilde{\theta})$
$\rho = 5$	AML2	$\hat{\sigma}^2$	0.738	0.118	0.950	1.33
	AML1	$\hat{\sigma}^2$	0.727	0.125	1.971	12.9
	A & B	σ_a^2	0.771	0.116	0.027	0.08
		σ_a^{*2}	0.754	1.236	0.354	0.45
$\rho = 10$	AML2	$\hat{\sigma}^2$	0.744	0.120	0.918	1.12
	AML1	$\hat{\sigma}^2$	0.741	0.117	0.905	1.09
	A & B	σ_a^2	0.812	0.142	0.870	0.97
		σ_a^{*2}	0.778	4.567	0.165	0.16
$\rho = 20$	AML2	$\hat{\sigma}^2$	0.745	0.120	0.906	1.06
	AML1	$\hat{\sigma}^2$	0.744	0.119	0.903	1.05
	A & B	σ_a^2	0.817	0.106	0.276	0.07
		σ_a^{*2}	0.820	17.93	0.231	-0.19
$N = 20$						
$\rho = 5$	AML2	$\hat{\sigma}^2$	0.840	0.081	0.511	0.29
	AML1	$\hat{\sigma}^2$	0.828	0.077	0.488	0.26
	A & B	σ_a^2	0.904	0.100	0.647	0.46
		σ_a^{*2}	0.802	0.510	0.579	1.00
$\rho = 10$	AML2	$\hat{\sigma}^2$	0.847	0.082	0.512	0.23
	AML1	$\hat{\sigma}^2$	0.844	0.081	0.506	0.22
	A & B	σ_a^2	0.920	0.099	0.563	0.34
		σ_a^{*2}	0.769	1.802	0.362	0.80
$\rho = 20$	AML2	$\hat{\sigma}^2$	0.849	0.082	0.518	0.23
	AML1	$\hat{\sigma}^2$	0.848	0.082	0.517	0.23
	A & B	σ_a^2	0.912	0.096	0.534	0.28
		σ_a^{*2}	0.695	6.993	0.232	0.72

TABLE D.3a

	$\tilde{\theta}$	AML2 $\hat{\sigma}^2$	AML1 $\hat{\sigma}^2$	A & B σ_a^2	A & B σ_a^{*2}
N = 30					
ρ = 5	$k_1(\tilde{\theta})$	0.894	0.883	0.906	0.888
	$k_2(\tilde{\theta})$	0.055	0.052	0.060	0.272
	$\gamma_1(\tilde{\theta})$	0.412	0.390	0.534	0.276
ρ = 10	$k_1(\tilde{\theta})$	0.901	0.898	0.938	0.888
	$k_2(\tilde{\theta})$	0.055	0.054	0.060	0.908
	$\gamma_1(\tilde{\theta})$	0.400	0.395	0.435	0.150
ρ = 20	$k_1(\tilde{\theta})$	0.902	0.902	0.941	0.873
	$k_2(\tilde{\theta})$	0.055	0.055	0.060	3.446
	$\gamma_1(\tilde{\theta})$	0.394	0.393	0.399	0.061
N = 40					
ρ = 5	$k_1(\tilde{\theta})$	0.912	0.901	0.947	0.876
	$k_2(\tilde{\theta})$	0.044	0.042	0.051	0.158
	$\gamma_1(\tilde{\theta})$	0.417	0.405	0.522	0.387
ρ = 10	$k_1(\tilde{\theta})$	0.920	0.917	0.967	0.853
	$k_2(\tilde{\theta})$	0.044	0.044	0.050	0.502
	$\gamma_1(\tilde{\theta})$	0.397	0.394	0.441	0.216
ρ = 20	$k_1(\tilde{\theta})$	0.922	0.921	0.962	0.790
	$k_2(\tilde{\theta})$	0.044	0.044	0.048	1.880
	$\gamma_1(\tilde{\theta})$	0.391	0.391	0.405	0.088

	$\tilde{\theta}$	AML2 $\hat{\sigma}^2$	AML1 $\hat{\sigma}^2$	A & B σ_a^2	A & B σ_a^{*2}
N = 50					
ρ = 5	$k_1(\tilde{\theta})$	0.929	0.919	0.949	0.915
	$k_2(\tilde{\theta})$	0.034	0.034	0.039	0.113
	$\gamma_1(\tilde{\theta})$	0.332	0.310	0.276	0.231
	$\gamma_2(\tilde{\theta})$	0.61	0.55	0.07	-0.19
ρ = 10	$k_1(\tilde{\theta})$	0.936	0.934	0.973	0.910
	$k_2(\tilde{\theta})$	0.036	0.035	0.039	0.338
	$\gamma_1(\tilde{\theta})$	0.284	0.279	0.248	0.096
	$\gamma_2(\tilde{\theta})$	0.52	0.51	0.23	-0.24
ρ = 20	$k_1(\tilde{\theta})$	0.937	0.937	0.970	0.884
	$k_2(\tilde{\theta})$	0.036	0.036	0.038	1.234
	$\gamma_1(\tilde{\theta})$	0.268	0.267	0.246	0.005
	$\gamma_2(\tilde{\theta})$	0.48	0.47	0.33	-0.26
N = 75					
ρ = 5	$k_1(\tilde{\theta})$	0.949	0.939	0.941	0.939
	$k_2(\tilde{\theta})$	0.026	0.025	0.028	0.060
	$\gamma_1(\tilde{\theta})$	0.368	0.357	0.321	0.357
ρ = 10	$k_1(\tilde{\theta})$	0.956	0.954	0.975	0.945
	$k_2(\tilde{\theta})$	0.026	0.026	0.028	0.166
	$\gamma_1(\tilde{\theta})$	0.338	0.335	0.300	0.267
ρ = 20	$k_1(\tilde{\theta})$	0.958	0.957	0.978	0.937
	$k_2(\tilde{\theta})$	0.026	0.026	0.027	0.582
	$\gamma_1(\tilde{\theta})$	0.325	0.325	0.300	0.206

TABLE D.3b

N	ρ	$\tilde\theta$	$k_1(\tilde\theta)$	$k_2(\tilde\theta)$	$\gamma_1(\tilde\theta)$	$\gamma_2(\tilde\theta)$
$N = 12$	$\rho = 5$	ρ_3	4.976	0.144	-0.406	0.55
		ρ_4	5.010	0.425	-2.812	17.01
		σ_4^2	0.690	2.062	2.211	8.42
	$\rho = 10$	ρ_3	9.982	0.273	-0.939	1.72
		ρ_4	9.975	1.234	-3.288	20.71
		σ_4^2	0.582	25.71	2.210	8.71
	$\rho = 20$	ρ_3	19.989	0.791	-1.413	2.99
		ρ_4	19.929	4.75	-3.890	27.37
		σ_4^2	0.148	395.9	2.212	8.76
$N = 20$	$\rho = 5$	ρ_3	4.987	0.066	-0.151	-0.03
		ρ_4	5.010	0.120	-0.383	0.18
		σ_4^2	0.857	0.683	0.984	1.26
	$\rho = 10$	ρ_3	9.985	0.110	-0.516	0.51
		ρ_4	9.985	0.322	-0.929	0.93
		σ_4^2	0.892	8.05	1.057	0.95
	$\rho = 20$	ρ_3	19.976	0.294	-1.026	1.24
		ρ_4	19.953	1.149	-1.219	1.48
		σ_4^2	1.023	124.06	1.102	1.04
$N = 30$	$\rho = 5$	ρ_3	5.006	0.038	-0.156	0.06
		ρ_4	5.029	0.064	-0.506	0.66
		σ_4^2	0.875	0.347	1.168	1.84
	$\rho = 10$	ρ_3	10.011	0.055	-0.450	0.47
		ρ_4	10.025	0.148	-1.132	2.23
		σ_4^2	0.810	3.70	1.464	3.12
	$\rho = 20$	ρ_3	20.018	0.130	-1.044	1.67
		ρ_4	20.032	0.492	-1.543	3.69
		σ_4^2	0.536	54.52	1.528	3.54
$N = 40$	$\rho = 5$	ρ_3	5.001	0.029	-0.091	-0.14
		ρ_4	5.018	0.041	-0.401	1.13
		σ_4^2	0.909	0.196	1.403	5.73
	$\rho = 10$	ρ_3	10.004	0.038	-0.265	0.17
		ρ_4	10.016	0.084	-1.127	3.85
		σ_4^2	0.860	1.923	1.796	7.20
	$\rho = 20$	ρ_3	20.010	0.077	-0.950	2.00
		ρ_4	20.023	0.263	-1.690	5.92
		σ_4^2	0.645	28.29	1.720	5.87
$N = 50$	$\rho = 5$	ρ_3	5.003	0.022	-0.089	0.10
		ρ_4	5.017	0.030	-0.276	0.50
		σ_4^2	0.928	0.133	1.036	1.98
	$\rho = 10$	ρ_3	10.007	0.027	-0.309	0.46
		ρ_4	10.018	0.058	-0.821	1.35
		σ_4^2	0.876	1.270	1.470	3.13
	$\rho = 20$	ρ_3	20.013	0.054	-0.874	1.40
		ρ_4	20.028	0.172	-1.314	2.41
		σ_4^2	0.647	18.40	1.474	2.90
$N = 75$	$\rho = 5$	ρ_3	4.999	0.014	-0.079	-0.10
		ρ_4	5.006	0.019	-0.307	0.24
		σ_4^2	0.962	0.077	0.932	1.89
	$\rho = 10$	ρ_3	10.000	0.018	-0.330	0.42
		ρ_4	10.003	0.035	-0.992	2.46
		σ_4^2	0.961	0.659	1.673	4.92
	$\rho = 20$	ρ_3	19.999	0.033	-1.074	2.73
		ρ_4	20.000	0.099	-1.703	5.39
		σ_4^2	0.947	9.592	1.770	5.29

TABLE D.4

REFERENCES

Abramowitz, M., and Stegun, I.A. (Eds) (1972). *Handbook of Mathematical Functions*. National Bureau of Standards. Dover: New York. [2.4, 3.2, 3.4, 3.5, 3.6, 4.5, 4.7, 5.4, 5.5, 5.8, Appendix A]

Amos, D.E. (1974). Computation of modified Bessel functions and their ratios. *Math. Comput.* 28, 239-252. [4.5]

Amos, D.E., Daniel, S.L. and Weston, M.K. (1977). Algorithm 511 CDC 6600 Subroutines IBESS and JBESS for Bessel functions $I_\nu(x)$ and $J_\nu(x)$, $x \geq 0$, $\nu \geq 0$. *ACM Trans. Math. Software* 3, 93-95. [4.5]

Andersen, E.B. (1973). *Conditional Inference and Models for Measuring*. Copenhagen, Mentalhygiejnisk Forlag. [4.2, 4.12]

Anderson, D.A. (1977). Statistical inference for a certain class of bivariate distributions. *J. Mult. Anal.* 7, 560-572.

Angell, I.O. (1977). Are stone circles circles? *Science and Archaeology*, No. 19, 16-19. [6.3]

Angell, I.O. and Barber, J. (1977). An algorithm for fitting circles and ellipses to megalithic stone rings. *Science and Archaeology*, No. 20, 11-16. [6.3, 6.4, 6.5]

Barnes, E.W. (1906). The asymptotic expansion of integral functions defined by Taylor's series. *Phil. Trans. Roy. Soc. A* 206, 249-297. [Appendix A]

Bhuchongkul, S. (1964). Non-parametric tests for independence. *Ann. Math. Statist.* 35, 138-149. [2.3]

Bierens de Haan, D. (1939). *Nouvelles Tables D'Intégrales Définies*. Stechert: New York. [3.4]

British Association Mathematical Tables Vol. VI (1950). *Bessel Functions Part I. Functions of Orders Zero and Unity*. Cambridge University Press. [4.5, 4.7]

British Association Mathematical Tables Vol. X (1952). *Bessel Functions Part II. Functions of Positive Integer Order*. Cambridge University Press. [4.5, 4.7]

Broadbent, S.R. (1955). Quantum hypotheses. *Biometrika* 42, 45-57. [6.2]

Burl, Aubrey (1976). *The Stone Circles of the British Isles*. Yale University Press: New Haven and London. [6.2]

Byrd, P.F. and Friedman, M.D. (1971). *Handbook of Elliptic Integrals for Engineers and Scientists*. 2nd ed. Springer-Verlag: New York. [4.14]

Campbell, J.T. (1934). The Poisson correlation function. *Proc. Edin. Math. Soc.* 4, 18-26. [3.7]

Chan, Nai Ng. (1965). On circular functional relationships. *J. Roy. Statist. Soc. B* 27, 45-56. [4.1, 4.11, 4.12, 6.4]

Cheriyan, K.C. (1941). A bivariate correlated gamma type distribution function. *J. Indian Math. Soc.* 5, 133-144. [3.3]

Choi, K. and Bulgren, W.G. (1968). An estimation procedure for mixtures of distributions. *J. Roy. Statist. Soc. B* 30, 444-460. [4.3]

Cook, M.B. (1951). Two applications of bivariate k-statistics. *Biometrika* 38, 368-376. [2.3]

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall: London. [2.3, 6.6]

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *J. London Math. Soc.* 11, 209-294. [4.2]

D'Agostino, R. and Pearson, E.S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika* 60, 613-622. [6.6]

David, F.N. and Fix, E. (1961). Rank correlation and regression in non-normal surfaces. *Proceedings of the Fourth Berkeley Sympos. Math. Statist. Prob.* Volume 1, 177-197. [3.3]

Dolby, G.R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. *Biometrika* 63, 39-50. [4.1]

Downs, T.D. (1966). Some relationships among the von Mises distributions of different dimensions. *Biometrika* 53, 269-272. [5.8]

Dussauchoy, A. and Berland, R. (1972). Lois gamma à deux dimensions. *C.R. Acad. Sci, Paris* 274A, 1946-1949. [3.3]

Dwass, M. and Teicher, H. (1957). On infinitely divisible random vectors. *Ann. Math. Statist.* 28, 461-470. [3.7]

Dwivedi, T.D. and Pandey, J.N. (1975). A note on Meyer's maximum likelihood estimate of the non-centrality parameter of non-central χ^2 variate. *Sankhyā B* 37, 453-456. [5.1, 5.7]

Elandt, R.C. (1961). The folded normal distribution: Two methods of estimating parameters from moments. *Technometrics* 3, 551-562. [5.1]

Erdélyi, A. (1953). *Higher Transcendental Functions.* Volume II. McGraw-Hill: New York. [3.3, 3.4, 3.5]

Erdélyi, A. (1953b). *Higher Transcendental Functions.* Volume III. McGraw-Hill: New York. [Appendix A]

Eyraud, H. (1936). Les principes de la mesure des corrélations. *Ann. Univ. Lyon A* 1, 30-47. [3.2]

Feller, W. (1966). Infinitely divisible distributions and Bessel functions associated with random walks. *J. Soc. Indust. Appl. Math.* 14, 864-875. [3.8]

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Volume II, 2nd Ed. Wiley: New York. [4.11, Appendix C].

Fieller, E.C., Hartley, H.O. and Pearson, E.S. (1957). Tests for rank correlation coefficients I. *Biometrika* 44, 470-481. [2.3]

Fieller, E.C. and Pearson, E.S. (1961). Tests for rank correlation coefficients II. *Biometrika* 48, 29-40. [2.3]

Fisz, Marek (1962). Infinitely divisible distributions: Recent results and applications. *Ann. Math. Statist.* 33, 68-84. [3.1]

Fraser, D.A.S. (1957). *Nonparametric Methods in Statistics.* Wiley: New York. [2.3]

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. A* (3) 14, 53-77. [2.3]

Freeman, P.R. (1976). A Bayesian analysis of the megalithic yard. *J. Roy. Statist. Soc. A* 139, 20-55. [6.2]

Freeman, P.R. (1977). Note: Thom's survey of the Avebury ring (with comment by Alexander Thom). *Journal for the History of Astronomy* 8, 134-136. [6.3, 6.5]

Gautschi, W. and Slavik, J. (1978). On the computation of modified Bessel function ratios. *Math. of Comput.* 32, 865-875. [4.5]

Gayen, A.K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika* 38, 219-247. [2.3]

Ghosh, B.K. (1966). Asymptotic expansions for the moments of the distribution of correlation coefficients. *Biometrika* 53, 258-262. [2.3]

Gnedenko, B.V. and Kolmogorov, A.N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley: Cambridge, Mass. [3.1]

Godambe, V.P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* 63, 277-284. [4.12]

Goovaerts, M.J., D'Hooge, L. and de Pril, N. (1978). On the infinite divisibility of the ratio of two gamma distributed variables. *Stoch. Proc. Appl.* 7, 291-298. [3.2]

Gradshteyn, I.S. and Ryzhik, I.W. (1965). *Tables of Integrals, Series and Products.* 4th Ed. Academic Press: New York and London. [4.13, 4.14, 5.9]

Graves, L.M. (1956). *The Theory of Functions of Real Variables.* 2nd Ed. McGraw-Hill: New York. [4.3]

Griffiths, R.C. (1969a). The canonical correlation coefficient of bivariate gamma distributions. *Ann. Math. Statist.* 40, 1401-1408. [3.3, 3.5]

Griffiths, R.C. (1969b). A class of infinitely divisible bivariate gamma distributions. *Sankhyā A* 31, 473-476. [3.3]

Grosswald, E. (1976a). The Student t-distribution of an odd degree of freedom is infinitely divisible. *Ann. Prob.* 4, 680-683. [3.2]

Grosswald, E. (1976b). The Student t-distribution of any degrees of freedom is infinitely divisible. *Z. Wahrscheinlichkeitstheorie und Verb. Gebiete* 36, 103-109. [3.2]

Gumbel, E.J. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc.* 56, 335-349. [3.2]

Hadingham, E. (1975). *Circles and Standing Stones.* Heinemann: London. [6.2]

Hansen, E.R. (1975). *A Table of Series and Products.* Prentice-Hall, Englewood Cliffs, N.J. [Appendix A]

Hartman, P. and Watson, G.S. (1974). Normal distribution on spheres. *Ann. Prob.* 2, 593-607. [4.5, 5.4]

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik des Universität Berlin.* 5, 179-233. [2.3]

Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika* 51, 241-245. [3.7]

Hopf, E. (1934). *Mathematical Problems of Radiative Equilibrium.* Cambridge Tracts in Mathematics and Mathematical Physics, No. 31. Cambridge University Press. [Appendix A]

Ifram, A.F. (1975). Inadmissibility and unbiasedness of some estimates of non-centrality parameters. *Bull. internat. Statist. Inst. Proc. 40th Session*, Warsaw 46, No. 3, 409-412. [5.6]

Ismail, M.E.H. (1977). Bessel functions and the infinite divisibility of the Student t-distribution. *Ann. Prob.* 5, 582-585. [3.2]

Ismail, M.E.H. and Kelker, D.H. (1976). The Bessel polynomials and the student t-distribution. *SIAM J. Math. Anal.* 7, 82-91. [3.2]

Jensen, D.R. (1969). Limit properties of non-central multivariate Rayleigh and chi-square distributions. *SIAM J. Appl. Math.* 17, 802-814. [4.9]

Jensen, D.R. (1976). Gaussian approximation to bivariate Rayleigh distributions. *J. Statist. Comput. Simul.* 4, 259-267. [3.5]

Johnson, N.L. (1949). Bivariate distributions based on simple translational systems. *Biometrika* 36, 297-304. [2.5]

Johnson, N.L. (1962). The folded normal distribution: Accuracy of estimation by maximum likelihood. *Technometrics* 4, 249-256. [5.1, 5.5]

Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions*, Volume 1. Houghton-Mifflin: Boston. [3.2, Appendix C].

Johnson, N.L. and Kotz, S. (1972). *Continuous Multivariate Distributions*. Wiley: New York. [2.1, 3.2]

Kahaner, D.K. and Waterman, M.S. (1978). Asymptotics for gamma and beta percentiles. *J. Statist. Comput. Simul.* 6, 315-316. [3.6]

Kalbfleisch, J.D. and Sprott, D.A. (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. B* 32, 179-208. [4.12]

Kalbfleisch, J.D. and Sprott, D.A. (1973). Marginal and conditional likelihoods. *Sankhyā A* 35, 311-328. [4.12]

Kelker, D. (1971). Infinite divisibility and variance mixtures of the normal distribution. *Ann. Math. Statist.* 42, 802-808. [3.2]

Kendall, D.G. (1974). Hunting quanta. *Philos. Trans. Roy. Soc. London A* 276, 231-266. [6.2]

Kendall, M.G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Volume 1, 4th Ed. Griffin: London. [6.6]

Kibble, W.F. (1941). A two variate gamma type distribution. *Sankhyā* 5, 137-150. [3.3]

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27, 887-906. [4.3]

Kimeldorf, G. and Sampson, A. (1975a). One parameter families of bivariate distributions with fixed marginals. *Commun. in Statist.* 4, 293-301. [2.1]

Kimeldorf, G. and Sampson, A. (1975b). Uniform representations of bivariate distributions. *Commun. in Statist.* 4, 617-627. [2.1, 2.3]

Kingman, J.F.C. and Taylor, S.J. (1966). *Introduction to Measure and Probability Theory*. Cambridge University Press. [4.3]

Konijn, H.S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.* 27, 300-323. [2.3]

Kowalski, C.J. and Tarter, M.E. (1969). Coordinate transformations to normality and the power of normal tests for independence. *Biometrika* 56, 139-148. [2.3]

Lancaster, H.O. (1957). Some properties of the bivariate normal distributior considered in the form of a contingency table. *Biometrika* 44, 289-292. [2.3, 3.4, 3.5]

Leone, F.C., Nelson, L.S. and Nottingham, R.B. (1961). The folded normal distribution. *Technometrics* 3, 543-550. [5.1]

Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatoires*, 2nd Ed. Gauthier-Villars: Paris. [3.1]

Lévy, P. (1948). The arithmetical character of the Wishart distribution. *Proc. Camb. Phil. Soc.* 44, 295-297. [3.8]

Luke, Y.L. (1975). *Mathematical Functions and Their Approximations.* Academic Press: New York. [5.7]

McKay, A.T. (1934). Sampling from batches. *J. Roy. Statist. Soc. B* 1, 207-216. [3.3]

Magnus, W., Oberhettinger, F. and Soni, R.P. (1966). *Formulas and Theorems for the Special Functions of Physics*, 3rd Ed. Springer-Verlag: Berlin. [Appendix A and C]

Mardia, K.V. (1970a). *Families of Bivariate Distributions.* Griffin: London. [2.1]

Mardia, K.V. (1970b). A translation family of bivariate distributions and Fréchet's bounds. *Sankhyā A* 32, 119-122. [2.3]

Maung, K. (1941). Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children. *Ann. Eugen. Lond.* 11, 189-223. [2.3, 3.4, 3.5]

Meyer, P.L. (1967). The maximum likelihood estimate of the non-centrality parameter of a non-central χ^2 variate. *J. Amer. Statist. Assoc.* 62, 1258-1264. [5.1]

Moran, P.A.P. (1958). Approximate relations between series and integrals. *Maths. Tables and Other Aids to Computation* 12, 34-37. [2.4]

Moran, P.A.P. (1969). Statistical inference with bivariate gamma distributions. *Biometrika* 56, 627-634. [2.1, 3.1]

Moran, P.A.P. (1971). Estimating structural and functional relationships. *J. Mult. Anal.* 1, 232-255. [4.1]

Morgenstern, D. (1956). Einfache Beispiel zweidimensionaler Verteilungen. *Mitt. Math. Statist.* 8, 234-235. [3.2]

Nataf, A. (1962). Détermination des distributions de probabilities dont les marges sont données. *C.R. Acad. Sci. Paris* 255, 42-43. [2.1]

Neff, N. and Strawderman, W.E. (1976). Further remarks on estimating the parameters of a non-central chi-square distribution. *Commun. Statist. - Theory and Method A* 5, 65-76. [5.1]

Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32. [4.11, 4.12, 6.4]

Nielsen, Niels. (1965). *Die Gammafunction.* Chelsea Publishing Company: Bronx, New York. [Appendix A]

Oberhettinger, F. and Badii, L. (1973). *Tables of Laplace Transforms.* Springer-Verlag: Berlin and New York. [Appendix C]

Pearson, E.S., D'Agostino, R.B. and Bowman, K.O. (1977). Tests for departure from normality: Comparison of power. *Biometrika* 64, 231-246. [6.6]

Pearson, E.S. and Hartley, H.O. (Eds) (1970). *Biometrika Tables for Statisticians*, Volume 1, 3rd Ed. Cambridge University Press. [6.6]

Pearson, E.S. and Please, N.W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika* 62, 223-241. [6.6]

Perlman, M.D. and Rasmussen, U. (1975). Some remarks on estimating a non-centrality parameter. *Commun. Statist.* 4, 455-468. [5.1]

Pitman, E.J.G. (1937). Significance tests which may be applied to samples from any population II. The correlation coefficient. *J. Roy. Statist. Soc. Suppl.* 4, 225-232. [2.3]

Robinson, S.M. (1961). Fitting spheres by the method of least squares. *Comm. Ass. Computing Machinery* 4, 491. [4.1]

Sarmanov, I.O. (1970). Gamma correlation process and its properties. *Soviet Math. Dokl.* 11 (2), 320-322. [3.3]

Schoenberg, I.J. (1938). Metric spaces and completely monotone functions. *Ann. Math.* 39, 811-841. [Appendix C]

Slater, L.J. (1960). *Confluent Hypergeometric Functions.* Cambridge University Press. [4.7, 4.9, 5.5, 5.7, 5.8]

Slater, L.J. (1966). *Generalized Hypergeometric Functions.* Cambridge University Press. [2.3]

Sprott, D.A. (1975). Marginal and conditional sufficiency. *Biometrika* 62, 599-605. [4.12]

Steutel, F.W. (1973). Some recent results in infinite divisibility. *Stoch. Proc. Appl.* 1, 125-143. [3.1]

Stuart, A. (1954). The asymptotic relative efficiencies of tests and the derivatives of their power function. *Skand. Aktuarietids.* 37, 163-169. [2.3]

Thom, A. (1955). A statistical examination of the megalithic sites in Britain. *J. Roy. Statist. Soc. A* 118, 275-295. [6.2, 6.3]

Thom, A. (1967). *Megalithic Sites in Britain.* Clarendon Press, Oxford. [6.2, 6.3]

Thom, A. and Thom, A.S. (1973). A megalithic lunar observatory in Orkney: The ring of Brogar and its cairns. *J. Hist. Astronomy* 4, 11-23. [6.3, 6.4, 6.5]

Thom, A., Thom, A.S. and Foord, T.R. (1976). Avebury (1): A new assessment of the geometry and metrology of the ring. *J. Hist. Astronomy* 7, 183-192. [6.3, 6.4, 6.5]

Thorin, O. (1977a). On the infinite divisibility of the Pareto distribution. *Scand. Actuarial J.*, 31-40. [3.2, Appendix C]

Thorin, O. (1977b). On the infinite divisibility of the lognormal distribution. *Scand. Actuarial J.*, 121-148. [3.2]

U.S. National Bureau of Standards (1948). *Tables of Probability Functions*, Volume II, 2nd Ed. U.S. Printing Office: Washington. [2.4]

Vere-Jones, D. (1967). The infinite divisibility of a bivariate gamma distribution. *Sankhyā A* 29, 421-422. [3.3]

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Inst. Math. Statist.* **20**, 595-601. [2.1, 4.2, 4.3]

Watson, G.N. (1944). *A Treatise on the Theory of Bessel Functions*, 2nd Ed. Cambridge University Press. [4.6, 4.12, 4.13, 5.3, Appendix A, Appendix C]

Whitt, Ward (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4**, 1280-1289. [2.3, 3.2]

Whittaker, E.T. and Robinson, G. (1932). *The Calculus of Observation. A Treatise on Numerical Mathematics*, 2nd Ed. Blackie: Glasgow. [6.4]

Wilson, E.B. and Hilferty, M.M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci. Washington* **17**, 684-688. [3.5]

Wolfowitz, J. (1957). The minimum distance method. *Ann. Math. Statist.* **28**, 78-88. [4.3]