ASYMPTOTIC EXPANSIONS AND DEFICIENCY

A THESIS SUBMITTED TO
THE INDIAN STATISTICAL INSTITUTE
IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DIVISION OF THEORETICAL STATISTICS
AND MATHEMATICS

BY
TAPAS KUMAR CHANDRA

CALCUTTA, MAY 1980

ACKNOWLEDGMENTS

I thank the Indian Statistical Institute for providing me with all facilities to carry out my research work.

It is a pleasure to record my indebtedness to Professor J. K. Ghosh who guided me throughout this work and helped me along at every stage of my research. I also thank him for permitting me to include our joint work in this dissertation.

I thank every member of the Division of Theoretical Statistics and Mathematics for providing the excellent academic atmosphere.

Finally, I thank Sri Arun Das for typing and Sri Dilip Chatterjee for cyclostyling of this thesis.

TAPAS K. CHANDRA

CONTENTS

LIST OF NOTATIONS

R^k	the k-dimensional Euclidean space with the usual metric topology
R^1_+	the set of all positive real numbers
$\|x\|$	the absolute value of the complex number x
x^+	the maximum of zero and the real number x
$[x]$	the greatest integer less than or equal to the real number x
$\|\cdot\|$	Euclidean Norm
$\langle , \rangle$	Euclidean Inner Product
$z^{(i)}$	the i-th component of the vector z
z^α	the product $(z^{(1)})^{\alpha(1)} \dots (z^{(k)})^{\alpha(k)}$ if z is an element of R^k and $\alpha = (\alpha(1),\dots,\alpha(k))$ is a multi-index of nonnegative integers
$s!$	the product of the integers from 1 to s
Γx	the gamma function (at x)
$\alpha!$	the product $\alpha(1)!\dots\alpha(k)!$ if $\alpha = (\alpha(1),\dots,\alpha(k))$ is a multi-index of nonnegative integers
$\binom{n}{r}$	$n!/(r!\,(n-r)!)$

$\lvert\alpha\rvert$	the sum $\alpha(1) + \dots + \alpha(k)$ if $\alpha = (\alpha(1), \dots \alpha(k))$ is a multi-index of nonnegative integers
z^1	the vector $(z^{(1)}, \dots, z^{(p)})$ if p is a <u>fixed</u> integer, $1 \leq p \leq k$, and z is an element of R^k
z^2	the vector $(z^{(p+1)}, \dots, z^{(k)})$ if p is a <u>fixed</u> integer, $1 \leq p \leq k$, and z is an element of R^k
D^i	partial differentiation with respect to the i-th coordinate variable
D^α	the operator $(D^{(1)})^{\alpha(1)} \dots (D^{(k)})^{\alpha(k)}$ if $\alpha = (\alpha(1), \dots, \alpha(k))$ is a multi-index of nonnegative integers
gradient H	$(D^1H, \dots, D^kH)$ if H is a real valued function on R^k
H'	gradient H
Hessian H	the $k \times k$ matrix $((D^iD^jH))$ if H is a real valued function on R^k
H''	Hessian H
$\# A$	the number of elements of A if A is a <u>finite</u> set

I_A	the indicator function of the set A
A^c	the complement of A
$A - B$	the set of all elements of A which <u>do not</u> belong to B
$A + B$	the symmetric difference of A and B
$A \cup B$	the union of the sets A and B
$A \cap B$	the intersection of the sets A and B
$f(A)$	the image of A under the map f if A is a subset of the domain of f
$f^{-1}(A)$	the inverse image of A under the map f if A is a subset of the range of f
I	the identity matrix
A^T	the transpose of the matrix A
A^{-1}	the inverse of the matrix A
$\det(A)$	the determinant of the square matrix A
$bd(A)$	the topological boundary of the set A
$Int(A)$	the topological interior of the set A
$sp(x;\varepsilon)$	the open sphere around x with radius $\varepsilon > 0$
$sp(A;\varepsilon)$	the union of the spheres $sp(x;\varepsilon)$, $x \in A$ if $\varepsilon > 0$

$sp(A;-\varepsilon)$	$\{x \mid sp(x;\varepsilon) \subset A\}$ if $\varepsilon > 0$
$\omega_f(x;\varepsilon)$	the oscillation of f over $sp(x;\varepsilon)$
$\omega_f(R^k)$	the oscillation of f over R^k if f is a function on R^k
E	the expection operator
$N(o;1)$	a random variable on R^1 following the normal distribution with mean zero and variance one
z_α	the upper 100α percent point of $N(o;1)$
ϕ_V	the density of the multivariate normal with mean zero and dispersion matrix V
ϕ	ϕ_I
$\Phi_V(A)$	the integral of ϕ_V over A
$\Phi(A)$	$\Phi_I(A)$
$\chi^2(\cdot;p)$	the density of the central chi-square distribution with p degrees of freedom
$\chi^2(\cdot;p,\delta)$	the density of the non-central chi-square distribution with p degrees of freedom and the noncentrality parameter $\delta(\geq 0)$
$\hat{\chi}^2(\cdot;p,\delta)$	the characteristic function of $\chi^2(\cdot;p,\delta)$
$\hat{f}$	the Fourier-Stieltjes Transform of the function f

$\hat{G}$	the Fourier-Stieltjes Transform of the distribution G
$\{\bar{P}_r\}$	the multivariate Cramér-Edgeworth polynomials
$X_n \xrightarrow{P_\theta} X$	the sequence of random vectors $\{X_n\}$ converges in distribution to the random vector X
$X_n = O_p(a_n)$	for every $\varepsilon > 0$, there exists a constant K such that $\text{Prob}(\lvert X_n\rvert > K\, a_n) < \varepsilon$ for all sufficiently large n
$X_n = o_p(a_n)$	for every $\varepsilon > 0$, $\text{Prob}(\lvert X_n\rvert > \varepsilon\, a_n) \to 0$ as $n \to \infty$
$\tilde{\theta}$	maximum likelihood estimator(s)
$\hat{\theta}$	restricted maximum likelihood estimator(s)

LIST OF ABBREVIATIONS

iff	if and only if
IID	independent and identically distributed
a. s.	almost sure
ML estimators	maximum likelihood estimators
Restricted ML estimators	maximum likelihood estimators under the null hypothesis
LR statistic	likelihood ratio statistic

INTRODUCTORY CHAPTER

1. The efficiencies introduced by E.J.G. Pitman and R.R.Bahadur are both meant to compare the asymptotic performance of statistical procedures. However there are many interesting situations where these criteria prove inadequate and further discrimination is necessary. One such attempt of further refinement is the criterion of deficiency.

This investigation was undertaken with the object of developing tools for studying deficiency of test procedures with

(i) same Pitman efficiency,

or,

(ii) same Bahadur efficiency.

Deficiency in the first case has been defined by Hodges and Lehmann (1970). Deficiency in the second case was defined by Chandra and Ghosh (1978). The latter paper, together with some applications to multivariate testing problems discussed in Chandra and Ghosh (1980b), forms Part II of this disseration. Part I is a study of valid asymptotic expansions for test statistics which include the likelihood ratio criterion, Wald's and Rao's statistics (see Rao(1965),pages 347-352). These statistics have the same limiting distribution under the null hypothesis as well as under contiguous alternatives; hence they have the same Pitman efficiency. The above-mentioned asymptotic expansions are needed to study the Pitman-deficiency of these procedures relative to each other.

Since a substantial part of this thesis is concerned with Bahadur efficiency, it is impossible to ignore the criticism of Bahadur's asymptotics by LeCam (1974) (see pages 232 and 233) as a study of "ghosts of departed quantities"; this remark has been quoted with approval by Pfanzagl (1980) (see page 27, Section 3.3c). No doubt there is some truth in this criticism but it applies equally well to all other asymptotic theories. Only the Ghosts appear at different places in different theories. In the asymptotics of Pitman (and LeCam, Pfanzagl and others), the "distance" between the null hypothesis θ_o and alternative θ_n goes to zero and so after a finitely many steps the difference between θ_o and θ_n becomes "practically" irrelevant; why should one then be interested in distinguishing such (close) hypotheses ? All asymptotics are no more than an approximation to the "finite" problem of real life. One can only hope that the ghosts provide a clue (?) to the soul if not the body of one's actual problem.

2. Recently two works on deficiency have come to our attention, namely, Albers (1974) and Kallenberg (1978). The former discusses Pitman deficiency of (mainly) <u>nonparametric</u> tests and so is not related to this thesis. It is the second

work that deserves some comments. It is about the asymptotic optimality of the likelihood ratio test with respect to two criteria ; one of them is an elaboration of the results of Bahadur (1965, 1967). Kallenberg works with the limit (as $\alpha \to 0$) of the smallest sample size $N(\alpha)$ needed to get power at least β $(0 < \beta < 1)$, the level of the test being α $(0 < \alpha < 1)$. In their paper (1978), Chandra and Ghosh also work with apparently different but equivalent quantities (see the last part of Section 3 ; see also the review of Kallenberg's monograph by Ghosh (1980)). It is therefore worth-pointing out that there is hardly any overlap ; Kallenberg is interested in cases where deficiency (in our sense) is infinity while we are interested in computing it when it is finite.

3. As noted earlier our results on asymptotic expansions are useful for studying Pitman-deficiency. To illustrate this, consider a simple hypothesis $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ where θ takes values in R^1 and three tests based on the likelihood ratio statistic, Wald's and Rao's statistics. These statistics are indistinguishable from the view-point of asymptotic efficiency. In the last paragraph of Section 6e.2, Rao (1965) raised the question of higher order discrimination between these statistics and conjectured that the statistic proposed by him is likely to be "locally" more powerful than the others (in the

second edition, this conjecture has been omitted). Peers (1971) attempted to settle the above question and concluded that the above conjecture has no ground ; in particular, his method will yield the mutual Pitman-deficiencies of these statistics to be infinity (of order $O(n^{1/2})$). The method used by Peers is, however, not quite reasonable. The defect of his method is that it compares the "cut-off points" rather than the statistics. For example, it can be shown that for a one parameter exponential family, all these tests are two-sided tests based on the <u>same</u> statistic but their relative deficiencies in the sense of Peers are $\pm\infty$. We have shown that these deficiencies are, in fact, finite; however we have not yet completed the necessary computations to settle the conjecture of Rao.

4. We now present a brief chapterwise summary of this work.

In Chapter One we consider, following Bhattacharya and Ghosh (1978), a real valued statistic $H(\bar{Z}_n)$ which is a "smooth" function of the sample mean $\bar{Z}_n$. In their paper, Bhattacharya and Ghosh restricted their attention to the case where the asymptotic distribution of

$$n^{1/2}\{H(\bar{Z}_n) - H(\mu)\}, \qquad \mu = E(\bar{Z}_n),$$

is normal. We consider here the case where

$$W_n = 2n\{H(\bar{Z}_n) - H(\mu)\}$$

is asymptotically central chi-square with degrees of freedom $= p$, say. It is shown that under certain conditions, the distribution function of W_n possesses an expansion in powers of n^{-1}, the coefficient of n^{-j} $(j \geq 1)$ being a finite linear combination of the distribution functions of central chi-squares with degrees of freedom p, $p+2$, $p+4$, etc. Under stronger conditions, it is further shown that this expansion holds uniformly in all Borel subsets of R^1. Our examples (see Section 7) and the applications to asymptotic theory discussed in the next chapter indicate that our assumptions are reasonable. At present, we do not have any example of a useful statistic, with limiting central chi-square distribution, to which our theorems do not apply.

In Chapter Two, we apply the above general theorems to get expansions (up to any degree of accuracy) for certain statistics commonly used for testing multivariate multiparameter (composite) hypotheses; for this we assume conditions similar to (but much stronger than) Cramér-Rao regularity conditions (see for example Assumptions (A_1) to (A_6) of Bhattacharya and Ghosh). In typical cases, these expansions hold uniformly over all intervals and over compact subsets of the parameter space (under H_o); they hold uniformly over all Borel subsets of R^1 in case the samples are drawn from an absolutely continuous exponential family indexed by its natural parameter. The most important example of such statistics is the likelihood ratio criterion. Our results unify those of a number of similar expansions (usually up to $o(n^{-2})$) for the likelihood ratio criterion used for testing specialised hypothesis (under normality assumptions); for references, one may consult the survey article by P.R. Krishnaiah (1978). Recently Hayakawa (1977) showed the possibility of such a general expansion (for the likelihood ratio statistic) up to $o(n^{-1})$; his method is, however, purely formal and are not necessarily valid in the sense of Bickel (1974). Under the assumptions of our theorem, we have justified the formal method used by Hayakawa. A different method, again purely formal, was used by Box in his well-known paper (1949). We have established the validity of Box's method as well.

Chapter Three extends the results of the previous chapters when the limiting distribution of W_n is non-central chi-square; this is typically the case, when the limiting 'null" distribution of W_n is central chi-suare <u>and</u> when one considers the limiting distribution of W_n under contiguous alternatives. Here too our results unify similar results available in the literature. The main theorem runs as follows : under certain conditions, the distribution function of W_n under contiguous alternatives can be expanded in powers of $n^{-1/2}$, the coefficient of $n^{-j/2}$ $(j \geq 1)$ being a finite linear combination of the distribution functions of noncentral chi-squares with same noncentrality parameter and with degrees of freedom p, p+2, p+4 etc. Our assumptions are satisfied by the (absolutely continuous) exponential family of distributions indexed by natural parameter (together with some more mild conditions).

Chapter Four developes methods to compare test procedures which are equally efficient from Bahadur's view-point. In literature such measures of discrimination are called deficiency; see, for example, Hodges and Lehmann (1970) who introduced this concept. Since two test procedures which are equally efficient by Bahadur's criterion are usually equally efficient by Cochran's criterion also, the problem of measuring deficiency has also been approached from Cochran's point of view. It is shown that approaches based on Bahadur's and Cochran's ideas lead to the same measure of deficiency if one uses limits in probability in the definition of Bahadur slopes ; the equivalence breaks down if one uses almost sure limits instead. One of the main results shows that the expansion of the significance level $\alpha_n(\beta, \theta)$, attained by a test procedure when the power at an alternative θ is held fixed, is usually of the following form:

$$\log \alpha_n(\beta;\theta) = -na(\mu(\theta)) + n^{1/2} z_\beta \sigma^*(\theta) \cdot \frac{da(\mu(\theta))}{d\mu(\theta)}$$
$$-c(k) \log n + d(\beta,\theta) + o(1).$$

Here β stands for the power of the test at θ, $c(k)$ is determined essentially by the dimension of the observations, $n^{1/2}\mu(\theta)$ and $\sigma^*(\theta)$ $(\sigma^*(\theta) > 0)$ denote respectively the asymptotic mean and standard deviation of the test statistic under θ, $a(\mu(\theta))$ is the Bahadur-slope under θ of the test statistic. It follows, in particular, that if two test procedures have the same Bahadur efficiency at θ and the associated asymptotic means and variances are also same for both of them, then the expansions of the logarithms of their significance levels will agree up to $o(1)$ terms and hence the deficiency will be finite; otherwise the deficiency will be infinity.

In Chapter Five, we compute the deficiency in some common multiparameter multivariate problems, our main interest being to compare the likelihood ratio and Bayes tests. We have developed here methods of finding expansions of the logarithms of the significance levels of Bayes tests (the corresponding expansions for the likelihood ratio tests can be obtained from Theorem 1 of Woodroofe (1978)). The main source of difficulty here is to obtain expansions (up to $o(1)$) of the logarithms of certain multidimensional large deviation probabilities. To this end, we have followed Borovkov and Rogozin (1965). The results of the last section extend those of Woodroofe (1978) and Schwarz (1978).

It may be noted that in Bahadur's as well as Cochran's approaches, one compares in effect <u>two statistics</u> rather than <u>two tests</u>. For when one specifies a test both the test statistic and the cut-off point(s) are regarded as fixed, whereas Bahadur and Cochran choose the latter so that the power $\beta(\theta)$ at the alternative θ (where a comparison is sought) is held fixed. If a weight function π is given over the alternative hypothesis, it may be more appropriate to fix the cut-off point by requiring that

$$\int \beta(\theta)\, \pi(d\theta) = \beta \qquad 0 < \beta < 1$$

However one then faces certain technical difficulties and so we have omitted it from the present discussion.

PART I

CHAPTER : ONE

VALID ASYMPTOTIC EXPANSIONS OF PERTURBED CHI-SQUARE VARIABLES

SECTION 1. INTRODUCTION

Let $\{Z_n\}_{n\geq 1}$ be a sequence of independent and identically distributed (IID) random vectors on R^k, with finite second moments. Let μ and V be respectively the mean vector and the dispersion matrix of Z_1 :

(1.1) $$\mu = E\, Z_1 \qquad V = E(Z_1 - \mu)\,(Z_1 - \mu)^T$$

where T denotes transpose. We shall assume that V is non-singular. Let H be a real valued Borel measurable function on R^k. Consider the statistic

(1.2) $$W_n^1 = n^{1/2}(H(\bar{Z}_n) - H(\mu)), \quad \bar{Z}_n = n^{-1} \sum_{j=1}^{n} Z_j \;, \quad n \geq 1.$$

It is well-known that (see Cramér (1946), page 366) if all the first-order partial derivatives of H are continuous and non-null at μ , then W_n^1 is asymptotically normal with mean zero and variance $\lambda^T V \lambda$; here $\lambda = (\lambda_1,\ldots,\lambda_k)$ with

(1.3) $$\lambda_i = \left.\frac{\partial H(z)}{\partial z^{(i)}}\right|_{z=\mu} \;, \quad 1 \leq i \leq k \,;\; z = (z^{(1)},\ldots,z^{(k)})$$

Bhattacharya and Ghosh (1978) improved this result considerably by obtaining, under further regularity conditions, Berry-Esséen bound and an asymptotic expansion in powers of $n^{-1/2}$ for the distribution function of W_n^1; they showed, among other things, that the coefficients of $n^{-j/2}$, $j \geq 1$, are polynomials (times the density of the limiting normal distribution) which depend only on the cumulants of $(Z_1 - \mu)$ of order $(j + 2)$ and less and the derivatives of H at μ of order $(j+1)$ and less. See in this connection their Theorems 1, 2 and Remark 1.1. Here we shall partially supplement their results by considering the case where λ is the null vector and

$$W_n = 2n(H(\bar{Z}_n) - H(\mu)) \tag{1.4}$$

is asymptotically distributed as a central chi-square. A statistic of this sort will be called a perturbed chi-square.

Let $L = ((\lambda_{i,j}))$ be the $k \times k$ matrix of the second-order partial derivatives of H at μ:

$$\lambda_{i,j} = \left. \frac{\partial^2 H(z)}{\partial z^{(i)} \partial z^{(j)}} \right|_{z=\mu}, \quad 1 \leq i,j \leq k \tag{1.5}$$

Assume that these partial derivatives are continuous in a neighbourhood of μ. Recall now that if the distribution of X is the k-variate normal with mean vector zero and dispersion matrix Σ, then a necessary and sufficient condition that $X^T A X$ (A is symmetric) follows a (central) χ^2 distribution is

$$\Sigma \Lambda \Sigma \Lambda \Sigma = \Sigma \Lambda \Sigma \tag{1.6}$$

and in this case, the degrees of freedom of the χ^2 is the rank of $A \Sigma$ (see Rao (1965), page 152). Clearly then a necessary and sufficient condition for W_n to be asymptotically χ^2 is that λ is the null vector, V is a non-null matrix and the equation $L^T V L = L$ holds (note that this implies L is positive semidefinite); in this case the degrees of freedom of the limiting χ^2 is the rank of $L = p$, say. Assume then that H is sufficiently smooth (i.e., that H has enough continuous derivatives in a neighbourhood of μ), that the distribution of Z_1 is smooth (i.e., Z_1 has enough number of finite moments and satisfies Cramér's condition (2.7)) and finally that the above necessary and sufficient condition holds. Under an additional technical condition on W_n (or equivalently on H), it is shown here (see Theorem 2.1, page 17) that the distribution function of W_n can be asymptotically expanded in powers of n^{-1}, the coefficient of n^{-j} $(j \geq 1)$ being a finite linear combination of distribution functions of (central) χ^2's with degrees of freedom p, $p+2$, $p+4$ etc. In case L is nonsingular, this technical condition holds but Example (7.1) shows that in general this condition cannot be relaxed without loosing the (desirable) property that in the asymptotic expansion, the coefficients of n^{-j} are linear combinations of central χ^2's. However, even when this condition does not hold the distribution function $F_n(x) = \text{Prob}(W_n \leq x)$ of W_n can possess a valid expansion (with x restricted to proper regions) in powers of n^{-1} — see, e.g., Example (7.2); at present it is not known whether such expansions have any application. It is shown in the next chapter that the likelihood ratio statistic and some of its competitors do satisfy the above technical condition under fairly general regularity assumptions so that this condition can be regarded as a natural one for expansions of perturbed χ^2's.

It is to be noted that the said expansion holds uniformly over all intervals of R^1_+. If the distribution of Z_1 is such that for some integer $m \geq 1$, $Z_1 + \cdots + Z_m$ has a nonzero absolutely continuous component with respect to Lebesgue measure on R^k (which implies that Z_1 satisfies Cramér's condition) and, moreover, W_n satisfies a stronger form of the above technical condition, then the said expansion for W_n holds uniformly over all Borel subsets of R^1_+. It will be shown in the next chapter that the likelihood ratio and other statistics do satisfy this stronger condition when the sample is drawn from an exponential family with natural parameter space (and with some more mild restrictions). Thus it appears that our main theorem, Theorem 2.1, and its variants are quite general to cover almost all statistics which are perturbed chi-squares.

Section 2 includes the statement of the main theorem of this chapter, together with some comments on its assumptions. Included also are some simple and verifiable sufficient conditions for these assumptions. The theorem is proved in Section 4, the related results being explained in Section 3. In Section 5 we state some useful variants of the main theorem in a form which is needed for the next chapter. Since the proof of the main theorem is given in meticulous details, those of the variants are omitted. Some common and frequently used methods for obtaining expansions of statistics are justified in Section 6 under the assumptions of our main theorem. The last section, Section 7, includes a few counter-examples related to the various results of this chapter. All basic notations are stated in Section 2 and part (A) of Section 4.

SECTION 2. MAIN THEOREM

(A) Statement of the main theorem:

The basic set up is described in the Introduction. Denote, in analogy with (1.3) and (1.5), the partial derivatives of H at μ by

$$(2.1) \qquad \ell_{i_1, i_2, \ldots, i_j} = (D^{i_1} D^{i_2} \ldots D^{i_j} H)(\mu)$$

$$j \geq 1, \; 1 \leq i_1, \ldots, i_j \leq k$$

where D^i denotes differentiation with respect to the i-th coordinate variable. Let s be an integer ≥ 4 and denote by $H_{s-1}(z)$ the Taylor expansion around $\mu = E\, Z_1$ of $H(z)$ up to and including terms involving the (s-1)th order derivatives of H :

$$(2.2) \qquad H_{s-1}(z) = \Sigma\, \ell_{i,j}\, z^{(i)} z^{(j)} + \ldots + \frac{2}{(s-1)!} \Sigma\, \ell_{i_1,\ldots,i_{s-1}}\, z^{(i_1)} \ldots z^{(i_{s-1})} .$$

Here and in the following z, x etc. will stand for an element of R^k and $z^{(i)}$ will denote the i-th component of z, $1 \leq i \leq k$. The symbols $\|\ \|$ and $<,>$ will denote Euclidean norm and inner product respectively, K_1, K_2 etc. will denote constants (i.e., nonrandom quantities free from n but possibly depending [redacted] on s). Let p be the rank of L. Assume without loss of generality that μ is the null vector.

With these notations, we now state our basic set of assumptions.

ASSUMPTION A_s :

(i) all derivatives of H of order s and less are continuous in a neighbourhood of μ.

(ii) the vector $\lambda = (D^1 H, \ldots, D^k H)(\mu)$ is null ;

(iii) the $k \times k$ matrix $L = ((D^i D^j H(\mu)))$ is nonnull and satisfies the equation

(2.3) $$L V L = L ;$$

(iv) if under some nonsingular linear transformation $x = Az$, $z^T L z$ becomes a positive-definite quadratic form in

$$x^1 = (x^{(1)}, \ldots, x^{(p)}), \quad p = \text{rank of } L \quad 1 \leq p \leq k,$$

then

(2.4) $$H_{s-1}(A^{-1}x) = \sum_{i,j=1}^{p} x^{(i)} x^{(j)} P_{i,j}(x) ,$$

for some polynomials $\{P_{i,j}\}$;

(v) if under some nonsingular linear transformation $x = Az$, $z^T L z$ becomes a positive-definite quadratic form in x^1, then in a bounded neighbourhood of $A\mu = 0$,

(2.5) (a) $$|H(A^{-1}x) - H_{s-1}(A^{-1}x)| \leq K_1 \|x^1\|^2 \|x\|^{s-2}$$

(2.6) (b) $$|D^i H(A^{-1}x) - D^i H_{s-1}(A^{-1}x)| \leq K_2 \|x^1\| \|x\|^{s-2}$$

$$1 \leq i \leq p .$$

Z_1 is said to satisfy Cramér's condition if

(2.7) $$\limsup_{\|t\|\to\infty} |E(\exp(i < t, Z_1 >))| < 1 .$$

Note that (2.7) implies that the dispersion matrix of Z_1 is nonsingular.

Z_1 is said to satisfy CONDITION D if

(2.8) there exist an m-dimensional vector Y and real-valued Borel measurable functions $f_1, \ldots, f_k$ on R^m such that

D(i): $Z_1^{(i)} = f_i(Y) \quad 1 \leq i \leq k$

and

D(ii): the distribution of Y has a nonzero absolutely continuous component (with respect to Lebesgue measure on R^m) whose density is positive on some open set U, the functions $f_1, \ldots, f_k$ are continuously differentiable in U and $1, f_1, \ldots, f_k$ are linearly independent as elements of the vector space of continuous functions on U.

The main Theorem can now be stated.

THEOREM 2.1. Let $\{Z_n\}_{n\geq 1}$ and H be as in the Introduction and define W_n by (1.4). Assume that, for some integer $s \geq 4$, H and Z_1 satisfy Assumptions A_s(i) - (iv) and that $E\|Z_1\|^{s-1}$ is finite.

(a) If Z_1 satisfies Cramér's condition (2.7), then there exist polynomials $\{q_r\}$ (in one variable), $1 \leq r \leq m$, whose coefficients do not depend on n such that the following expansion is valid for all $u \in R_+^1$ and is uniform in $u \in [u_o, \infty)$, $u_o > 0$:

$$P(W_n \leq u)$$

(2.9)
$$= \int_0^u \chi^2(v\,;\,p) \sum_{r=0}^{m} n^{-r} q_r(v)\,dv + \varepsilon_n$$

($q_o \equiv 1$). Here $\chi^2(v\,;\,p)$ is the density at v of a central χ^2 with p degrees of freedom, m is the greatest integer $\leq (s-3)/2$, p is the rank of L ,

(2.10)
$$\varepsilon_n = o(n^{-m}) \quad \text{if } s \text{ is odd ;}$$
$$= o(n^{-m-1/2}) \quad \text{if } s \text{ is even .}$$

Finally, u_o can be taken to be zero if $p > 1$.

(b) If, moreover, Assumption $A_s(v)$ holds and Z_1 satisfies CONDITION D (2.8), then expansion (2.9) holds uniformly over all Borel subsets B of $[o, \infty)$:

$$P(W_n \in B)$$

(2.11)
$$= \int_B \chi^2(v\,;\,p) \sum_{r=0}^{m} n^{-r} q_r(v)\,dv + \varepsilon_n \,.$$

The Theorem will be proved in Section 4.

REMARK 2.1 It will follow from the proof of Theorem 2.1 that for each $r \geq 1$ the polynomial q_r depends only on the cumulants of $(Z_1 - \mu)$ of order $(2r + 2)$ and less, and on the derivatives of H at μ of order $(2r + 2)$ and less; also the degree of the polynomial will be at most $6r$.

REMARK 2.2 Certain limited expansions for W_n can be obtained under relaxed assumptions. For convenience we shall here consider the case $p > 1$. Suppose assumptions of part (a) of the above theorem hold with $s = 4$ except that we do not assume that A_s(iv) and Cramér's condition hold. Then it can be shown that the following analogue of Berry-Esseén Theorem holds:

$$(*) \qquad \sup_{0<u<\infty} \left| \text{Prob}(W_n \leq u) - \int_0^u \chi^2(v;p)\, dv \right| \leq C\, n^{-1/2}(\log n)^{1/2}$$

where C is a suitable constant. (A proof of (*) is sketched in Remark 4.3. Whether the upper bound in (*) can be replaced by $O(n^{-1/2})$ is still under investigation.) If, moreover, A_5(i) holds and the distribution of Z_1 is <u>strongly nonlattice</u> in the sense that

$$E\{\exp(i < t_1, Z_1 >)\} < 1 \qquad \text{for all } t \neq 0,$$

then the left side of (*) is $o(n^{-1/2+\varepsilon})$ for any $\varepsilon > 0$. Finally part (a) of the theorem remains essentially true (see Example 7.5) without A_s(iv) <u>provided</u> $p > p_o(s)$, $p_o(s)$ being a suitable constant; for example one can take $p_o(5) = 5$.

(B) REMARKS ON ASSUMPTION A_s :

Assumptions A_s(ii) and (iii) ensure that the limiting distribution of W_n is a central χ^2(see the last paragraph, page 11). Assumption A_s(iv) is a technical one and ensures that the Taylor expansion (keeping only the relevant terms) of W_n, when expanded in terms of $A(\bar{Z}_n - \mu)$, is at least of degree two in the first p components of $A(\bar{Z}_n - \mu)$. Plainly, Assumption A_s(iv) and (v) hold if A_s(i) and (ii) hold <u>and</u> L is positive-definite. Assumption A_s(iv) is a natural sufficient condition on W_n which guarantees that the coefficient of n^{-r} in the expansion of $\mathrm{Prob}(W_n \leq u)$ will be a finite linear combination of chi-squares. Assumption A_s(v) is also of the same kind; it imposes restrictions on the function H itself (instead of on its Taylor expansion H_{s-1}) : roughly speaking it (combined with A_s(iv)) says that $H(z)$ contains neither any term free from the first p components of $A(z - \mu)$ nor any term linear in these p components. This statement can be made precise when H is real analytic. This and some other facts follow from Lemma 2.1 below. We need the following notations. If α is a vector of nonnegative integers, say $\alpha = (\alpha(1),\ldots,\alpha(k))$ and $z = (z^{(1)},\ldots,$ $z^{(k)})$, put

$$z^1 = (z^{(1)},\dots,z^{(p)}) \qquad z^2 = (z^{(p+1)},\dots,z^{(k)})$$

$$|\alpha| = |\alpha(1)| + \dots + |\alpha(k)| \qquad \alpha! = (\alpha(1))! \dots (\alpha(k))!$$

(2.12)

$$z^\alpha = (z^{(1)})^{\alpha(1)} \dots (z^{(k)})^{\alpha(k)}$$

$$D^\alpha = (D^1)^{\alpha(1)} \dots (D^k)^{\alpha(k)} \qquad D = (D^1,\dots,D^k)$$

REMARK 2.3 To check Assumption A_s(iv) (Assumption A_s(v)), it is enough to check (2.4) ((2.5) and (2.6) respectively) for <u>one</u> nonsingular matrix A which has the property that $z^T L z$ is positive-definite in x^1 where $x = Az$. (Proof: Let $x = A_1 z$ and $y = A_2 z$, A_1 and A_2 being nonsingular matrices, be such that $z^T L z$ is positive-definite in x^1 as well as in y^1. To show that x^1 is a linear combination of y^1. Partition $A_1 A_2^{-1}$ as follows

$$A_1 A_2^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{matrix} p \\ k-p \end{matrix}$$

$$\qquad\qquad\qquad p \qquad k-p$$

Then $x^1 = B_{11} y^1 + B_{12} y^2$. But $x^1 = 0$ if and only if $y^1 = 0$ implying that B_{12} must be null.)

Recall that a function $H: R^k \to R^k$ is said to be <u>real analytic</u> on an open set $U \subset R^k$ if H is infinitely differentiable on U and if for any x in U, there exists an open sphere $B \subset U$ with centre x such that for any y in B

$$\sum_{|\alpha|=0}^{\infty} \frac{(y-x)^{\alpha}}{\alpha!} D^{\alpha} H(x)$$

converges absolutely to $H(y)$.

LEMMA 2.1 Let U be a bounded open convex subset of R^k containing the origin and $H : U \to R^k$ be s-times continuously differentiable. Let H_{s-1} be the Taylor expansion of H around origin including the (s-1)th order derivatives of H. Then one has,

(a)

$$|H(x) - H_{s-1}(x)| \leq K_3 \|x\|^s$$

$$|D^i H(x) - D^i H_{s-1}(x)| \leq K_4 \|x\|^{s-1} \qquad 1 \leq i \leq k ;$$

(b) Let $1 \leq p \leq k$ and suppose that

(2.13) $$H(x) = 0, \quad D^i H(x) = 0 \qquad 1 \leq i \leq p$$

for all x with $x^1 = 0$; then

$$H_{s-1}(x) = \sum_{i,j=1}^{p} x^{(i)} x^{(j)} P_{i,j}(x)$$

where $\{P_{i,j}\}$ are suitable polynomials in x ;

(c) Let H be real analytic satisfying (2.13). Then

$$|H(x) - H_{s-1}(x)| \le K_5 \|x^1\|^2 \|x\|^{s-2}$$

and

$$|D^i H(x) - D^i H_{s-1}(x)| \le K_6 \|x^1\| \|x\|^{s-2} \quad 1 \le i \le p ;$$

(d) Suppose H is such that whenever $|\alpha| = s$ and $x \ne ($

(2.14) $$D^\alpha H(x) \begin{cases} \le K_7 \|x^1\|^2 \|x\|^{-2} & \text{if } |\alpha^1| = 0 , \\ \le K_8 \|x^1\| \|x\|^{-1} & \text{if } |\alpha^1| = 1 ; \end{cases}$$

then the conclusions of (c) hold.

PROOF: Recall that (see Corollary 8.3, page 58, Bhattacharya and Ranga Rao (1976))

(2.15) $$H(x) = H_{s-1}(x) + s \sum_{|\alpha|=s}^{\infty} \frac{x^\alpha}{\alpha!} \int_0^1 (1-u)^{s-1} D^\alpha H(ux)dx$$

(a) follows immediately from (2.15).

(b) Write $H_{s-1}(x)$ as a polynomial in x^1 and break up it into three parts

(2.16) $$H_{s-1}(x) = H^1_{s-1}(x) + H^2_{s-1}(x) + H^3_{s-1}(x)$$

where $H^1_{s-1}(x)$ does not involve x^1, $H^2_{s-1}(x) = \sum_{i=1}^{p} x^{(i)} Q_i(x)$

with $Q_1(x)$ free from x^1 and finally $H^3_{s-1}(x) = \sum_{i,j=1}^{p} x^{(i)} x^{(j)} \times$ $Q_{i,j}(x)$ with $Q_{i,j}(x)$ depending on both x^1 and x^2. By the first inequality of (a) and the first equality of (2.13), $|H^1_{s-1}(x)| \leq K_3 \|x\|^s$ for all x with $x^1 = 0$. As $H^1_{s-1}(x)$ is a polynomial of degree at most (s-1) it must vanish identically. Next fix an i such that $1 \leq i \leq p$. Then

$$D^i H_{s-1}(x) = Q_i(x) + \sum_{j=1}^{p} x^{(j)} Q^*_{i,j}(x)$$

where $\{Q^*_{i,j}(x)\}$ are suitable polynomials. By the second inequality of (a) and the second equality of (2.13) we then get $|Q_i(x)| \leq K_4 \|x\|^{s-1}$ for all x with $x^1 = 0$. As $Q_i(x)$ is a polynomial of degree at most (s-2), $Q_i(x) \equiv 0$. This establishes (b).

(c) Since H is analytic, so is $H-H_{s-1}$. Arguing as in (b) (with H_{s-1} replaced by $H-H_{s-1}$ and polynomials replaced by power series), the result follows.

(d) follows from (2.15) applied to H and $D^i H$ $(1 \leq i \leq p)$.

(C) REMARKS ON CONDITION D:

It is well-known that Cramér's condition is needed for any general Edgeworth type expansions for distribution functions (see, for example, Cramér (1937) and Bhattacharya and Range Rao (1976)). Condition D is stronger than Cramér's condition, and in fact it serves as a very convenient way of verifying the latter. To this end, we state the following useful lemma of Bhattacharya and Ghosh (1978) (see Lemma 2.2, page 446).

LEMMA 2.2 (Bhattacharya and Ghosh)

If Z_1 satisfies CONDITION D, then the k-fold convolution of Z_1 has a nonzero absolutely continuous component with respect to Lebesgue measure on R^k.

It is well-known that Z_1 does satisfy Cramér's condition if

(2.17) for some integer $j \geq 1$, the j-fold convolution of Z_1 has a nonzero absolutely continuous component with respect to Lebesgue measure on R^k.

In almost all applications of Lemma 2.2, the following observation (see Remark 1.2, page 437, Bhattacharya and Ghosh) proves to be useful.

REMARK 2.4 Let G be the distribution of Y where Y is as in definition of CONDITION D. If the density g, say, of the absolutely continuous part of G is such that $U \equiv \{y \mid g(y) > 0\}$ is open and $G(U) = 1$, then one may replace D(ii) in definition of CONDITION D by

(2.18) D(ii)' : $f_1, \ldots, f_k$ are continuously differentiable on U.

We shall need CONDITION D only to use the following result of Bikjalis (see Corollary 29.6, page 206 and the remark on page 207 of Bhattacharya and Ranga Rao (1976)).

THEOREM 2.2 (Bikjalis)

Let $\{Z_n\}_{n\geq 1}$ be a sequence of IID random vectors on R^k with mean zero and nonsingular dispersion matrix V. Assume that $E\|Z_1\|^{s-1}$ is finite and that (2.17) holds. Then one has

$$\text{(2.19)} \qquad \text{Prob}(n^{1/2}\bar{Z}_n \in B_k) - \int_{B_k} \xi_{s-1,n}(z)dz = o(n^{-(s-3)/2})$$

uniformly over all Borel subsets B_k of R^k. Here $\xi_{s-1,n}(z)$ is the multivariate Edgeworth expansion for $n^{1/2}\bar{Z}_n$ up to terms of order $o(n^{-(s-3)/2})$ (see (4.1) and (4.2)).

SECTION 3. AUXILIARY RESULTS

REMARK 3.1 Suppose that Assumptions A_s(iii) and (iv) hold. We may then assume that

$$\text{(3.1)} \qquad V = I, \quad z^T L z = \|z^1\|^2$$

and that (2.4) holds with $A = I$. Here I is the $k \times k$ identity matrix. (A similar remark is true for Assumptions A_s(iii) - (v).)

For a proof, note that there exists a nonsingular matrix R such that

$$R^T V^{-1} R = I , \quad R^T L R = S$$

where S is a diagonal matrix (see Rao (1965), Section 1c. 3(iii) page 37). In view of A_s(iii), S is also idempotent and hence we may assume without loss of generality that the first p diagonal elements of S are one and that the rest are zero. Then $z^T L z$ under the transformation $x = R^{-1} z$ becomes $\|x^1\|^2$ and so (2.4) holds (by A_s(iv)) with $A = R^{-1}$. Remark (2.3) completes the proof provided Z and $H(z)$ are replaced by $R^{-1}Z$ and $H(R^{-1}z)$ respectively.

The next remark will be used, essentially, at the final stage of the proof of Theorem 2.1 to show that the asymptotic expansion for W_n will be in powers of n^{-1} (instead of in powers of $n^{-1/2}$) and that the coefficient of n^{-r} $(r \geq 1)$ will be a finite linear combination of chi-squares with degrees of freedom p, $p+2$, $p+4$ etc. (instead of with degrees of freedom p, $p+1$, $p+2$ etc.). Hence it can be omitted except for the notations introduced in its first paragraph.

REMARK 3.2 Assume first that the rank p of L is > 1 and consider the transformation T_1 on R^k which sends z^1 to $(r, \theta^{(1)}, \ldots, \theta^{(p-1)})$ by means of a polar transformation and keep z^2 unaltered:

$$
\begin{aligned}
z^{(1)} &= r\cos\theta^{(1)} \cdots \cos\theta^{(p-2)}\cos\theta^{(p-1)}, \\
z^{(2)} &= r\cos\theta^{(1)} \cdots \cos\theta^{(p-2)}\sin\theta^{(p-1)}, \\
&\;\vdots \\
z^{(p-1)} &= r\cos\theta^{(1)}\sin\theta^{(2)}, \\
z^{(p)} &= r\sin\theta^{(1)}
\end{aligned} \tag{3.2}
$$

with $0 < r < \infty$ and (θ, z^2) belonging to the set

$$
\begin{aligned}
A = \{(\theta, z^2) \mid -\tfrac{\pi}{2} < \theta^{(i)} < \tfrac{\pi}{2},\ 1 \le i \le p-2, \\
0 < \theta^{(p-1)} < 2\pi,\ z^2 \in R^k\}
\end{aligned} \tag{3.3}
$$

$$\theta = (\theta^{(1)}, \ldots, \theta^{(p-1)}).$$

The Jacobian of T_1 is $r^{p-1} J(\theta)$ where

$$J(\theta) = (\cos\theta^{(1)})^{p-2}(\cos\theta^{(2)})^{p-3} \cdots \cos\theta^{(p-2)}. \tag{3.4}$$

Using (3.2), an expression of the form

$$(z^{(1)})^{\alpha(1)} \cdots (z^{(k)})^{\alpha(k)} \tag{3.5}$$

($\alpha(i)$ are nonnegative integers) can be written as

$$
\begin{aligned}
& r^{\alpha(0)} \Big(\prod_{i=1}^{p} (z^{(i)}/r)^{\alpha(i)} \Big) \Big(\prod_{i=p+1}^{k} (z^{(i)})^{\alpha(i)} \Big) \\
&= R(\alpha;\ r, \theta, z^2), \quad \text{say}
\end{aligned} \tag{3.6}
$$

where $\alpha(0) = \alpha(1) + \dots + \alpha(p)$ and $\alpha = (\alpha(0), \alpha(1), \dots, \alpha(k))$. The notation $R(\alpha) \equiv R(\alpha; r, \theta, z^2)$ will be used even if the power $\alpha(0)$ of r in (3.6) differs from $\alpha(1) + \dots + \alpha(p)$. (We shall be concerned with those $R(\alpha)$'s for which $\alpha(0)$ will differ from $\alpha(1) + \dots + \alpha(p)$ by an even integer and $\alpha(0)$ is a nonnegative integer). A finite sum of constant multiples of terms of the form $R(\alpha)$ will be denoted by $R_{ij}(r, \theta, z^2)$, $i \geq 1$, $j \geq 1$.

Recall that expression (3.5) is odd if and only if at least one of $\alpha(1), \dots, \alpha(k)$ is odd. In the same vein, we shall say that an expression of the form $R(\alpha)$ is odd if at least one of $\alpha(1), \dots, \alpha(k)$ is odd. Then, for each $r > 0$,

$$\int_A \exp(-\tfrac{1}{2} \|z^2\|^2)\, R(\alpha)\, d\theta\, dz^2 = 0 \quad \text{if} \quad R(\alpha) \quad \text{is odd} .$$

(A is defined in (3.3)). The above equation is a consequence of a symmetry argument. Specifically, let $\alpha(j)$ be odd. If $p + 1 \leq j \leq k$, replace $z^{(j)}$ only by $-z^{(j)}$ (keeping others fixed) ; if $3 \leq j \leq p$, replace $\theta^{(p-j+1)}$ only by $\phi^{(p-j+1)}$ where

$$\phi^{(p-j+1)} = \begin{cases} \theta^{(p-j+1)} & \text{if } 0 < \theta^{(p-j+1)} < \pi/2 , \\ \theta^{(p-j+1)} & \text{if } -\pi/2 < \theta^{(p-j+1)} < 0 ; \end{cases}$$

if $j = 2$, replace $\theta^{(p-1)}$ only by $\phi_1^{(p-1)}$ where

$$\phi_1^{(p-1)} = \begin{cases} -\theta^{(p-1)} + 2\pi & \text{if } 0 < \theta^{(p-1)} < \pi , \\ -\theta^{(p-1)} + 2\pi & \text{if } \pi < \theta^{(p-1)} < 2\pi , \end{cases}$$

so that $\cos \phi_1^{(p-1)} = \cos \theta^{(p-1)}$ and $\sin \phi_1^{(p-1)} = -\sin \theta^{(p-1)}$;

if $j=1$, replace $\theta^{(p-1)}$ only by $\phi_2^{(p-1)}$ where

$$\phi_2^{(p-1)} = \begin{cases} \pi - \theta^{(p-1)} & \text{if } 0 < \theta^{(p-1)} < \pi , \\ 3\pi - \theta^{(p-1)} & \text{if } \pi < \theta^{(p-1)} < 2\pi , \end{cases}$$

so that $\cos \phi_2^{(p-1)} = -\cos \theta^{(p-1)}$ and $\sin \phi_2^{(p-1)} = \sin \theta^{(p-1)}$.

More generally, we shall say that <u>the expression</u>

$$r^{\alpha(0)} \left(\prod_{i=1}^{p-1} (\cos \theta^{(i)})^{a_i} (\sin \theta^{(i)})^{b_i} \right) \left(\prod_{i=p+1}^{k} (z^{(i)})^{c_i} \right)$$

<u>is odd if at least one of the nonnegative integers</u> $b_1, \dots, b_{p-1}$, a_{p-1}, c_{p+1}, ..., c_k <u>is odd</u>. Finally say that $R_{ij}(r, \theta, z^2)$ <u>is odd if every</u> $R(\alpha)$ <u>in</u> $R_{ij}(r, \theta, z^2)$ <u>is odd</u>. Obviously, <u>if</u> $R(\alpha)$ <u>is odd, so is</u> $R(\alpha) J(\theta)$. <u>One verifies that the various</u> $R_{ij}(r, \theta, z^2)$ <u>occurring in the proof of Theorem 2.1 satisfy</u> (3.7)

below (A is defined in (3.3)) :

(3.7) $\int_A R_{ij}(r,\theta,z^2)\, J(\theta) \exp(-\frac{1}{2}\|z^2\|^2)\, d\theta\, dz^2$ is zero if j is odd and is a polynomial in r^2 (instead of r) for any $j \geq 1$.

The following lemma will be useful to verify that $R_{ij}(r,\theta,z^2)$ is odd if j is so. If $\{P_j(z)\}_{j\geq 0}$ are polynomials in k variables, say that $\{P_j(z)\}_{j\geq 0}$ enjoy the odd-even property if

(3.8) the degree of each monomial of $P_j(z)$ is odd or even according as j is odd or even .

LEMMA 3.1 Let

$$P^i(z) = \sum_{j=1}^{s-3} n^{-j/2}\, Q^i_j(z) \qquad 1 \leq i \leq k_1 \;;\; Q^i_0 \equiv 1$$

and f be a real analytic function on an open subset of R^{k_1} containing the origin. Write

$$f(P^1(z),\ldots,P^{k_1}(z)) = \sum_{j=0}^{s-3} n^{-j/2}\, Q^*_j(z) + o(n^{-(s-3)/2}) .$$

If $\{Q^i_j(z)\}_{0\leq j\leq s-3}$ enjoy the odd-even property (3.8) for each $i = 1,\ldots,k_1$, then so do $\{Q^*_j(z)\}_{0\leq j\leq s-3}$.

PROOF: Since f can be expanded in a Taylor's series and each term considered separately, it suffices to prove the lemma for the special case where $f(u_1,\ldots,u_{k_1}) = u_1 \cdots u_{k_1}$. To this end, fix a j such that $1 \leq j \leq s-3$ and note that any monomial of $Q_j^*(z)$ is obtained by multiplying some monomials (of degree $k(j_i)$, say) of $Q_{j_i}^i(z)$, $0 \leq i \leq s-3$, where

$$j_o + j_1 + \cdots + j_{s-3} = j; \tag{3.9}$$

also the degree, r say, of this monomial of $Q_j^*(z)$ will be given by

$$r = k(j_o) + k(j_1) + \cdots + k(j_{s-3}). \tag{3.10}$$

The proof will be complete once we have established that whenever $j_o, j_1, \ldots, j_{s-3}$ are nonnegative integers satisfying (3.9) and whenever $k(j_o), k(j_1),\ldots,k(j_{s-3})$ are nonnegative integers satisfying (3.11) below, r given in (3.10) will be odd or even according as j is odd or even.

(3.11) For each $i = 0,1,\ldots,s-3$, $k(j_i)$ is odd or even according as j_i is odd or even.

Put $A = \{j_i \mid j_i$ is odd, $0 \leq i \leq s-3\}$ and $B = \{j_i \mid j_i$ is even, $0 \leq i \leq s-3\}$. Then since $\sum_{i \in B} j_i$ is always even, one gets in view of (3.9)

(3.12) $\#$ A is odd or even according as j is odd or even.

Since for each $i \in B$ $k(j_i)$ is even (by (3.11)), $\sum_{i \in B} k(j_i)$ is always even. On the other hand if $i \in A$ then $k(j_i)$ is odd (by (3.11)) and so it follows from (3.12) that $\sum_{i \in A} k(j_i)$ (and hence r) is odd or even according as j is odd or even.

We next justify briefly the second half of (3.7). The special form of the density function of the multivariate normal distribution with mean zero and dispersion matrix identity and Assumption A_s(iv) about the function $h_{s-1,n}$ (see (4.4)) show respectively that the different $R(\alpha)$'s arising from the expansions (see (4.16) and (4.19)) of these two functions do satisfy the condition that $\alpha(0)$ differs from $\alpha(1) + \dots + \alpha(p)$ by an even number ; on the other hand, for the $R(\alpha)$'s arising from the Edgeworth expansion $\xi_{s-1,n}$ (see (4.2)), we in fact have $\alpha(0) = \alpha(1) + \dots + \alpha(p)$. One now need only to observe that if some $R(\alpha)$ fails to be odd, then $\alpha(1) + \dots + \alpha(p)$ will be even.

If $p = 1$, in the definition of T_1 the variable θ is not needed and one takes $r = |z^{(1)}|$; the different facts mentioned above remain true with necessary modifications. This completes Remark 3.2.

SECTION 4. PROOF OF THEOREM 2.1

(A) NOTATIONS :

Here we list the basic notations of this section; see also Remark 3.2. Let $\chi_j(t)$ be the jth cumulant of $< t, Z_1 - \mu >$ and define the Cramér-Edgeworth polynomials $\{\overline{P}_r(it)\}_{r\geq 1}$:

(4.1)
$$\overline{P}_r(it) = \sum_{\alpha=1}^{r} \{ \Sigma^* \frac{\chi_{j_1+2}(it)}{(j_1+2)!} \cdots \frac{\chi_{j_\alpha+2}(it)}{(j_\alpha+2)!} \}$$
$$\chi_j(it) = i^j \chi_j(t) \qquad t \in R^k, \; r \geq 1,$$

where the sum Σ^* is over all α-tuples of positive integers $(j_1,\ldots,j_\alpha)$ satisfying $j_1 + \cdots + j_\alpha = r$. Let ϕ_V be the density of the k-dimensional normal distribution with mean zero and dispersion matrix V. Let $\xi_{s-1,n}$ be the multivariate Edgeworth expansion of $n^{1/2}(\overline{Z}_n - \mu)$ up to $o(n^{-(s-3)/2})$:

(4.2)
$$\xi_{s-1,n} = (1 + \sum_{j=1}^{s-3} n^{-j/2}\, \overline{P}_j(-D))\, \phi_V$$

where $\overline{P}_j(-D)$ is the operator obtained by formally substituting $-D = (-D^1,\ldots, -D^k)$ for it in $\overline{P}_j(it)$, $1 \leq j \leq s-3$ (for details, see Section 2.7, pages 51-57, Bhattacharya and Ranga Rao (1976)). We note that for each $j \geq 1$, the coefficient of $n^{-j/2}$ in $\xi_{s-1,n}(z)$ is $\phi_V(z)\, P_j(z)$ where $P_j(z)$ is a polynomial in z

with the property that the degree of each monomial of $P_j(z)$ is even or odd according as j is even or odd (observe that this is stronger than the statement that if j is odd, $P_j(z)$ is odd according to the definition given in Remark 3.2). Let

$$g_n(z) = 2n(H(\mu + n^{-1/2} z) - H(\mu)) \tag{4.3}$$

and $h_{s-1,n}(z)$ be the following Taylor expansion of $g_n(z)$.

$$h_{s-1,n}(z) = 2 \sum_{j=2}^{s-1} \frac{n^{-(j-2)/2}}{j!} \sum \ell_{i_1,\ldots,i_j} z^{(i_1)} \cdots z^{(i_j)} \tag{4.4}$$

Then $W_n = g_n(n^{1/2}(\bar{Z}_n - \mu))$ and note that the coefficient of $n^{-j/2}$ in $h_{s-1,n}(z)$ is a homogeneous polynomial (in z) of degree (j+2) and hence the degree of each of its monomial is odd or even according as j is odd or even $(j \geq 1)$. Write

$$\begin{aligned} M_n &= \{z \mid \|z\|^2 \leq (s-2)\ \lambda \log n\}, \\ \lambda &= \text{largest eigen value of } V, \\ B_n &= g_n^{-1}(B) \cap M_n \subset R^k \qquad B \subset R^1 \end{aligned} \tag{4.5}$$

The following estimate of the tail probability of the multivariate normal distribution will be frequently used.

$$\int_{\|z\|^2 > s \log n} \|z\|^q \exp(-\tfrac{1}{2}\|z\|^2)dz$$

(4.6)

$$= o(n^{-s/2+\varepsilon}) \qquad \varepsilon > 0,\ q \geq 0,\ s > 0 .$$

This follows from the elementary fact that

(4.7) $$\int_{(t \log n)^{1/2}}^{\infty} r^j \exp(-\tfrac{1}{2} r^2)dr = O((\log n)^{j-1} n^{-t/2})$$

$$t > 0,\ j \geq 0$$

(B) PROOF OF PART (b) OF THEOREM 2.1 :

In view of Remark (3.1) we shall assume that (3.1) holds true and that Assumptions A_s(iv) and (v) hold with $A = I$. In the following, B will denote a Borel subset of R^1_+. One can easily verify that equations (4.8), (4.9), (4.20) and (4.22) hold uniformly over all Borel subsets of R^1_+. Now

$$\text{Prob}(W_n \varepsilon B) = \text{Prob}(n^{1/2}(\bar{Z}_n - \mu) \varepsilon g_n^{-1}(B))$$

(4.8)

$$= \int_{g_n^{-1}(B)} \xi_{s-1,n}(z)dz + o(n^{-(s-3)/2})$$

(see (4.2) and (4.3)). In the last equality we have used Bikjalis' result (see Theorem 2.2, page 26).

In the sequel we shall use the notations introduced in the first paragraph of Remark 3.2. It should be noted that the following arguments are valid provided (3.1) holds and Assumptions A_s(i) - (v) hold with $A = I$; in particular no assumption on Z_1 is needed. One can also easily verify that equations (4.12) and (4.17) hold uniformly in $z \varepsilon M_n$, and that equations (4.14) through (4.16), equations (4.18), (4.19) hold uniformly in $(r, \theta, z^2) \varepsilon T_1(M_n)$.

The transformation T_1 of Remark 3.2 together with (4.6) yields

$$\int_{g_n^{-1}(B)} \xi_{s-1,n}(z)\,dz$$

(4.9)
$$= \int_{T_1(B_n)} (2\pi)^{-k/2} r^{p-1} J(\theta) \exp(-\tfrac{1}{2}(r^2 + \|z^2\|^2)) \times \sum_{j=0}^{s-3} n^{-j/2} R_{1j}(r,\theta,z^2) \cdot dr\, d\theta\, dz^2 + o(n^{-(s-3)/2}) \qquad R_{10} \equiv 1$$

(see (4.5) and (2.12)). For convenience in notations, we shall write

(4.10)
$$R_o(r,\theta,z^2) = (2\pi)^{-k/2} r^{p-1} J(\theta) \exp(-\tfrac{1}{2}(r^2+\|z^2\|^2))$$
$$R_{ij}(r,\theta,z^2) \equiv 1 \qquad \text{if } j = 0, \; i \geq 1,$$

and further abbreviate $R_o(r,\theta,z^2)$ and $R_{ij}(r,\theta,z^2)$ as R_o and R_{ij} respectively. We apply next the transformation

$T_2(r,\theta,z^2) = (r',\theta,z^2)$ with

(4.11) $$r' = (g_n(T_1^{-1}(r,\theta,z^2)))^{1/2} \qquad (r,\theta,z^2) \in T_1(R^k)$$

Note that there exists an integer $n_o \geq 2$ such that T_2 is a C^s diffeomorphism on $T_1(M_n)$ if $n \geq n_o$. In the following $z \in M_n$ and $(r,\theta,z^2) \in T_1(M_n)$. Since by part (a) of $A_s(v)$ one has (see (4.4))

(4.12) $$g_n(z) = h_{s-1,n}(z) + \|z^1\|^2 . O((\log n/n)^{(s-2)/2})$$

and since under T_1

(4.13) $$h_{s-1,n}(z) = r^2 \sum_{j=0}^{s-3} n^{-j/2} R_{2j}$$

(use A_s (iv) and (3.1)), it follows that

(4.14) $$r' = r\left(\sum_{j=0}^{s-3} n^{-j/2} R_{3j} + O((\log n/n)^{(s-2)/2})\right).$$

(Here and in the following we have used the fact that any real analytic function, defined in an open neighbourhood of the origin, of $\sum_{j=0}^{s-3} n^{-j/2} R_{ij}(r,\theta,z^2)$ can again be expressed in the form $\sum_{j=0}^{s-3} n^{-j/2} R^*_{ij}(r,\theta,z^2) + o(n^{-(s-2)/2+\varepsilon})$ uniformly in $(r,\theta,z^2) \in T_1(M_n)$ for any $\varepsilon > 0$.)

It can be shown that the above equality can be inverted as follows

$$(4.15) \qquad r = r'\left(\sum_{j=0}^{s-3} n^{-j/2} R_{4j} + o(n^{-(s-2)/2+\varepsilon})\right) \qquad \varepsilon > 0$$

(To verify (4.15), let $r_0 = r'$ and define inductively $r_{i,n}$ as follows

$$r_{i+1,n} = r' - r_{i,n} \sum_{j=1}^{s-3} n^{-j/2} R_{3j}(r_{i,n}\,;\,\theta,\ z^2)$$

$$0 \leq i \leq s-4$$

One then verifies inductively that

$$r' - r_{i,n} \sum_{j=1}^{s-3} n^{-j/2} R_{3j}(r_{i,n}\,;\,\theta,\ z^2)$$

$$= \ r + o(n^{-(i+2)/2+\varepsilon}) \qquad 0 \leq i \leq s-4,\ \varepsilon > 0$$

uniformly in $(r, \theta, z^2) \in T_1(M_n)$ and that $r_{s-3,n}$ can be expressed, after neglecting terms of order $o(n^{-(s-3)/2})$, in the form

$$r' + r' \sum_{j=1}^{s-3} n^{-j/2} R_{4j}\ (r'\,;\,\theta,\ z^2)\ .$$

Plainly equation (4.15) holds.)

We now want to expand the integrand on the left side of (4.9) and the Jacobian, $\partial r/\partial r'$, of T_2. Clearly in view of (4.15) we can write

$$R_o(r,\theta,z^2)\sum_{j=0}^{s-3} n^{-j/2} R_{1j}(r,\theta,z^2)$$

(4.16)

$$= R_n(r',\theta,z^2)\sum_{j=0}^{s-3} n^{-j/2} R_{5j}(r',\theta,z^2) + o(n^{-(s-3)/2})$$

Now from part (b) of $A_s(v)$, one has

(4.17) $D^i g_n(z) = D^i h_{s-1,n}(z) + \|z^1\| \cdot O((\log n/n)^{(s-2)/2})$

$1 \leq i \leq p$.

Also from (3.2)

$$\left|\frac{\partial}{\partial r} g_n(T_1^{-1}(r,\theta,z^2)) - \frac{\partial}{\partial r} h_{s-1,n}(T_1^{-1}(r,\theta,z^2))\right|$$

$$\leq \sum_{i=1}^{p} |(D^i g_n)(T_1^{-1}(r,\theta,z^2)) - (D^i h_{s-1,n})(T_1^{-1}(r,\theta,z^2))|$$

$(r,\theta,z^2) \in T_1(R^k)$

Using (4.13) we therefore get

$$\frac{\partial}{\partial r} g_n(T_1^{-1}(r,\ \theta,\ z^2))$$

(4.18)

$$=2r(\sum_{j=0}^{s-3} n^{-j/2} R_{6j} + O((\log n/n)^{(s-2)/2})$$

In view of (4.11), (4.18) and (4.15),

(4.19) $\partial r/\partial r' = \sum_{j=0}^{s-3} n^{-j/2} R_{7j}(r',\theta,z^2) + o(n^{-(s-3)/2})$

Thus (4.9) can be written as

(4.20) $$\int_{g_n^{-1}(B)} \xi_{s-1,n}(z)dz = \int_{T_2T_1(B_n)} R_o \sum_{j=0}^{s-3} n^{-j/2} R_{8j} + o(n^{-(s-3)/2})$$

(The above derivation shows that the $o(n^{-(s-3)/2})$ term of (4.20) can be replaced by $o(n^{-(s-2)/2+\varepsilon})$ for any $\varepsilon > 0$). Now note that (see (4.5)) $T_2T_1(B_n) = T_2T_1(B_o) \cap T_2T_1(M_n)$ where

(4.21) $$B_o = \{(r',\theta,z^2) \in T_2T_1(R^k) \mid (r')^2 \in B\}$$

Using the continuity of the transformations T_2 and T_1 (see also (4.14)), one can choose n_o so large that if n exceeds n_o, $T_2T_1(M_n)$ contains the set $\{(r',\theta,z^2) : (r')^2 \leq (s-2) \log n - \frac{1}{2} \log 2 , \|z^2\|^2 \leq (s-2) \log n\}$ (see also Lemma 3.2, page 183 of Edwards (1973)); consequently (4.20) can be replaced by(using once again (4.6))

(4.22) $$\int_{g_n^{-1}(B)} \xi_{s-1,n}(z)\, dz$$
$$= \int_{B_o} R_o(r',\theta,z^2) \sum_{j=0}^{s-3} n^{-j/2} R_{8j}(r',\theta,z^2)dr'd\theta dz^2$$
$$+ o(n^{-(s-3)/2}).$$

Integrating now with respect to θ and z^2 and using the second half of (3.7), one can see that the integrand (except for the factor $(r')^{p-1} \exp(-\frac{1}{2}(r')^2)$) on the right side of (4.22) is a polynomial in $(r')^2$. More precisely from (3.7) one can conclude that if j is odd $\int_{B_o} R_o R_{8j} = 0$, and if $j \geq 2$ is even

$$\int_{B_o} R_o R_{8j} = \int_{(r')^2 \varepsilon B} (r')^{p-1} \exp(-\tfrac{1}{2}(r')^2)\ q_j^*((r')^2)\,dr'$$

where $q_j^*(v)$ is a suitable polynomial in $v \varepsilon R_+^1$. It is well-known that

$$\int_{B_o} R_o = \int_{(r')^2 \varepsilon B} \chi^2((r')^2 ; p)\ 2r'dr' .$$

With the following choice of

(4.23) $$q_j(v) = 2^{p/2-1}\,\Gamma(p/2)\ q^*_{j/2}(v) , \qquad j = 2,4,\ldots,$$

equation (2.11) therefore follows from (4.8) and (4.22).

(C) PROOF OF PART (a) OF THEOREM 2.1 :

As before assume that (3.1) holds and that Assumption A_s(iv) holds with $A = I$. For some technical reasons, we need the following preparatory lemmas. The first two give estimates of certain multivariate integrals. To state these lemmas, we need the following notations.

For $\varepsilon > 0$, $A \subset R^k$ and f a real valued function on R^k, let

$$sp(x\,;\,\varepsilon) = \{ z \mid \| z-x\| < \varepsilon \}$$

$$sp(A\,;\,\varepsilon) = \cup \{ sp(x\,;\,\varepsilon) \mid x \in A \}$$

(4.24) $$bd(A) = \text{the boundary of } A$$

$$\omega_f(R^k) = \sup \{ |f(y)-f(z)| : y, z \in R^k \}$$

$$\omega_f(x;\varepsilon) = \sup \{ |f(y)-f(z)| : y, z \in sp(x, \varepsilon) \}$$

LEMMA 4.1 Let s be an integer ≥ 4 and $\{ Q_j \}_{0 \leq j \leq s-3}$ polynomials in k variables ($Q_0 \equiv 1$). Let $1 \leq p \leq k$ and B denote a Borel subset of R^1_+. Then one has

(a)

(4.25) $$\int_{E_n(B)} \phi(z)dz \leq 2 \int_B \chi^2(v\,;\,p)dv + o(n^{-(s-3)/2})$$

where

$$E_n(B) = \{ z \mid \|z^1\|^2 \sum_{j=0}^{s-3} n^{-j/2} Q_j(z) \in B \}$$

(4.26) $$z^1 = (z^{(1)},\ldots,z^{(p)}) \qquad z = (z^{(1)},\ldots,z^{(k)})$$

and finally the $o(n^{-(s-3)/2})$ term does not depend on B.

(b)

(4.27) $$\int_{sp(bd(E_n(B));\varepsilon)} \phi(z)dz \leq \int_{sp(bd(B);\varepsilon((2s-3)\log n)^{1/2})} 2\chi^2(v;p)dv + o(n^{-(s-3)/2})$$

where the $o(n^{-(s-3)/2})$ term does not depend on B or ε $(0 < \varepsilon \leq 1)$.

PROOF : (a) The first part of the proof is based on the arguments analogous to those given in (4.9) through (4.22) of the proof of part (b) of Theorem 2.1 except that r' is to replaced by

$$r' = r\left(\sum_{j=0}^{s-3} n^{-j/2} Q_j(T_1^{-1}(r, \theta, z^2))\right)^{1/2}.$$

One then integrates with respect to θ and z^2, and repeatedly uses (4.7). (It is readily seen that (4.25) remains true even if the $o(n^{-(s-3)/2})$ term there is replaced by $o(n^{-(s-2)/2+\varepsilon})$ for any $\varepsilon > 0$.)

(b) For notational convenience, put

$$B(\varepsilon) = sp(bd(B) ; \varepsilon), \quad A_n = bd(E_n(B)) .$$

Note that for any $n \geq 1$, $A_n \subset E_n(bd(B))$. Get an $n_o \geq 1$ such that the Euclidean norm of the gradient of

$\|z^1\|^2 \sum_{j=0}^{s-3} n_o^{-j/2} Q_j(z)$ restricted to the set $sp(M_{n_o} ; 1)$ (see (4.5)) is less than $((2s-3) \log n)^{1/2} = \varepsilon'_n$ say. Then if $n \geq n_o$, $0 < \varepsilon \leq 1$

$$(4.28) \qquad sp(A_n ; \varepsilon) \subset (E_n(B(\varepsilon'_n)) \cap M_n) \cup M_n^c$$

where M_n^c is the complement of M_n. An appeal to part (a) (with B replaced by $B(\varepsilon_n')$) and (4.6) now establishes part (b).

This completes the proof of Lemma 4.1. (One may note that Lemma 4.1 remains true if the integrand on the left side of (4.25) or (4.27) is replaced by an expression of the form $\xi_{s-1,n}$ (with $V = I$); see (4.2).)

LEMMA 4.2 Let s be an integer ≥ 4 and let

(4.29)
$$A_n(B) = \{ z \mid h_{s-1,n}(z) \in B \}$$

(see (4.4)). Then one has uniformly over all Borel subsets B of R_+^1,

(4.30)
$$\int_{A_n(B)} \xi_{s-1,n}(z)dz = \sum_{j=0}^{m} n^{-j} \int_B \chi^2(v ; p)\, q_j(v)dv + o(n^{-(s-3)/2})$$

where $\xi_{s-1,n}$ is defined in (4.2) and the polynomials $\{ q_j(v) \}_{0 \leq j \leq m}$ and the integer m are as in Theorem 2.1.

PROOF: The arguments given in equations (4.9) through (4.23) of the proof of part (b) of Theorem 2.1 establish the lemma provided one now defines

$$r' = (h_{s-1,n}(T_1^{-1}(r, \theta, z^2)))^{1/2} .$$

(Equation (4.30) remains true even if the $o(n^{-(s-3)/2})$ term there is replaced by $o(n^{-(s-2)/2+\varepsilon})$ for any $\varepsilon > 0$.)

The next result is due to Bhattacharya (see Theorem 1.5, page 10, Bhattacharya (1977)). We use the notations of (4.24) and (4.2).

THEOREM 4.1 (Bhattacharya) Let $\{Z_n\}_{n\geq 1}$ be a sequence of IID k-dimensional random vectors with mean zero and dispersion matrix V. Assume that $E\|Z_1\|^{s-1}$ is finite for some integer $s \geq 4$ and that Z_1 satisfies Cramér's condition (2.7). Then for every real valued bounded Borel measurable function f on R^k, one has

$$|E(f(n^{1/2}\bar{Z}_n)) - \int_{R^k} f(z)\,\xi_{s-1,n}(z)\,dz|$$

(4.31)
$$\leq \delta_n\, n^{-(s-3)/2}\,\omega_f(R^k) + \int_{R^k} \omega_f(x\,;\,\exp(-dn))\,\phi_V(x)\,dx$$

where δ_n tends to zero as $n \to \infty$, d is <u>any positive constant such that</u> $d < -k^{-1}\log\theta$ where

$$\theta = \sup\{|E(\exp(i < t, Z_1 >))| : \|t\| > (16\,E\|Z_1\|^3)^{-1}\}$$

<u>and</u> moreover δ_n, d do not depend on f.

<u>We now start proving part (a) of Theorem 2.1</u> by obtaining first the asymptotic expansion of $W_n' = h_{s-1,n}(n^{1/2}(\bar{Z}_n - \mu))$ up to $o(n^{-(s-3)/2})$. Let $B_n(u) = \{z \mid h_{s-1,n}(z) \leq u\}$ ...

By part (b) of Lemma 4.1, Assumption A_s(iv) and (4.7)

$$(4.32) \qquad \int_{sp(bd(B_n(u)); \exp(-dn))} \phi(z)dz = o(n^{-(s-3)/2})$$

uniformly in $u \in R^1_+$ for any $d>0$; for this, consider separately the two cases $p > 1$ and $p = 1$. Theorem 4.1 and Lemma 4.2 together show that W'_n possesses the expansion (2.9).

Let $\delta_n(h) = h\, n^{-(s-3)/2}$ with $h > 0$. Then there exists an integer $n_o \geq 1$ such that if $n \geq n_o$, $0 < u < \infty$

$$(4.33) \qquad \{W'_n \leq u - \delta_n(h)\} \cup M_n^c \subset \{W_n \leq u\} \subset \{W'_n \leq u + \delta_n(h)\} \cup M_n^c.$$

Also since (2.9) holds for W'_n,

$$\text{Prob}(|W'_n - u| \leq \delta_n(h)) = \sum_{j=0}^{m} n^{-j} \int_{(u-\delta_n(h))^+}^{u+\delta_n(h)} \chi^2(v;p)\, q_j(v)dv + o(n^{-(s-3)/2})$$

where the $o(n^{-(s-3)/2})$ term does not depend on u or h and $(u - \delta_n(h))^+$ stands for the maximum of zero and $(u - \delta_n(h))$. The [redacted] estimate

$$\int_{(u-\delta_n(h))^+}^{u+\delta_n(h)} v^{\alpha} \chi^2(v\,;\,p)\,dv = O(\delta_n(h))$$

[redacted] is uniform in $u \in R_+^1$ if $p > 1$ and uniform in $u > u_0 > 0$ if $p = 1$. Clearly then (4.33) and the fact that (2.9) holds for W_n' complete the proof of part (a) of Theorem 2.1.

REMARK 4.1 In case $p = 1$, the expansion (2.9) holds uniformly in $u \in R_+^1$ as well provided Assumption $A_s(v)$ holds. Under this stronger condition one proceeds directly with W_n as in the proof of Theorem 2.1(b).

REMARK 4.2 Let $p > 1$. From the above proof, it is seen that in part (a) of Theorem 2.1, the expansion (2.9) holds uniformly over every family $\mathcal{A}$ of Borel subsets of R_+^1 satisfying

(4.34) $$\sup_{B \in \mathcal{A}} \int_{sp(bd\,(B);\,\varepsilon)} \chi^2(v\,;\,p)\,dv = O(\varepsilon) \quad \text{as } \varepsilon \to 0\,;$$

equation (4.33) is to be replaced by

$$\{W_n' \in sp(B\,;\,-\delta_n(h))\} \cup M_n^c$$
$$\subset \{W_n \in B\}$$
$$\subset \{W_n' \in sp(B\,;\,\delta_n(h))\} \cup M_n^c$$

where $sp(B ; -\delta_n(h)) = U\{x \mid sp(x; \delta_n(h)) \subseteq B\}$. Note that the relation $sp(B ; \delta_n(h)) - sp(B ; -\delta_n(h)) = sp(bd(B); \delta_n(h))$ is true on R^k (in fact this relation is true in any metric space every open sphere of which is connected). It also follows that the Φ-probability of the set $\{W_n \in sp(bd(B) ; \exp(-dn))\}$ is $o(n^{-(s-3)/2})$ uniformly over all Borel subsets in $\mathcal{A}$ satisfying (4.34) (Here d is any positive constant).

REMARK 4.3 Here we briefly describe one possible way of proving the inequality (*) in Remark 2.2 (see page 19). We start with Corollary 3 of Sweeting (1977) (see also inequality (2.15), page 444, Bhattacharya and Ghosh (1978)) and estimate the Φ_V-probability of ε_n-sphere ($\varepsilon_n = O(n^{-1/2})$) around the set $\{W_n \leq u\}$ repeating the arguments up to (4.28) given in Lemma 4.1(b) and then using Theorem 3.1 of Bhattacharya and Ranga Rao (1976) for each fixed z^2; to apply Theorem 3.1 one notes that for sufficiently large n, W_n is convex in z^1 on M_n.

SECTION 5. SOME VARIANTS OF MAIN THEOREM

In Theorem 2.1 we assumed that $\{Z_n\}_{n\geq 1}$ is a sequence of independent and identically distributed random variables. In most applications (e.g., to the asymptotic theory of statistics) this assumption appears to be restrictive. One however does not need the full force of the IID structure of $\{Z_n\}_{n\geq 1}$;

for example, in definition of $W_n = 2n(H(\bar{Z}_n) - H(\mu))$ the normalised deviation $n^{1/2}(\bar{Z}_n - \mu)$ based on the IID sequence $\{Z_n\}_{n\geq 1}$ can be replaced by an arbitrary sequence $\{U_n\}_{n\geq 1}$ possessing a similar Edgeworth expansion. Theorems 5.1 and 5.2 below give the precise statements. Since the proofs of the theorems run parallel to that of Theorem 2.1, they are omitted.

In almost all applications of these two theorems, we shall be supplied with the statistic W_n rather than the function H and the sequence $\{U_n\}_{n\geq 1}$. One has to choose judiciously the sequence $\{U_n\}_{n\geq 1}$ such that Assumption A_s(iv) or (v) is satisfied. This problem of choice arises typically in most statistical applications; see, for example, [redacted] Chapter Two and Example 7.2.

THEOREM 5.1 Let $\{U_n\}_{n\geq 1}$ be a sequence of k-dimensional random vectors admitting the following Edgeworth expansion uniformly over all Borel subsets B_k of R^k.

$$\text{Prob}\,(n^{1/2}(U_n - \mu) \in B_k) = \int_{B_k} \xi_{s-1,n}(z)dz + o(n^{-(s-3)/2}) \tag{5.1}$$

where $\mu \in R^k$ and

$$\xi_{s-1,n}(z) = \left(\sum_{j=0}^{s-3} n^{-j/2} P_j(z)\right)\phi_V(z) \qquad (P_0 \equiv 1), \tag{5.2}$$

$\{P_j(z)\}_{o\leq s\leq s-3}$ being suitable polynomials in z and ϕ_V being the density of the k-variate normal distribution with mean zero and dispersion matrix V. Let

(5.3) $$W_n = 2n(H(n^{1/2}(U_n - \mu)) - H(\mu)) \qquad n \geq 1$$

where H is a real valued Borel measurable function on R^k satisfying Assumptions A_s(i) - (v) for some integer $s \geq 4$ (see pages 15-16).

(i) Then there exist nonnegative integers $k_1,\ldots,k_{s-3}$ and constants $\{a_{ij}\}$ not depending on n $(0 \leq i \leq k_j,\ 0 \leq j \leq s-3)$ such that the following holds uniformly over all Borel subsets B of R^1_+:

(5.4) $$\text{Prob}(W_n \in B) = \sum_{j=0}^{s-3} n^{-j/2} \sum_{i=0}^{k_j} a_{ij} \int_B \chi^2(v\,;\,p + 2i)\,dv + o(n^{-(s-3)/2}) \qquad k_o = 1$$

(ii) If moreover $\{P_j\}_{o\leq j\leq s-3}$ enjoy the odd-even property (3.8), then

(5.5) $$a_{ij} = 0 \quad \text{for all } i,\ 0 \leq i \leq k_j,\ j \text{ odd}.$$

THEOREM 5.2 Let $\{U_n\}_{n\geq 1}$ be a sequence of k-dimensional random vectors defined on $(\Omega, \mathcal{F}, P)$ such that for some integer $s \geq 4$, Assumption (I) below holds.

Assumption (I) : There exist $\mu \in R^k$, V a nonsingular matrix and polynomials $\{P_j\}_{0\leq j\leq s-3}$ $(P_0 \equiv 1)$ in k variables such that

(5.6)
$$\sup_{B_k \in \mathcal{A}} |P(n^{1/2}(U_n - \mu) \in B_k) - \int_{B_k} \xi_{s-1,n}(z)dz| = o(n^{-(s-3)/2}) ,$$

for every family $\mathcal{A}$ of Borel subsets B_k of R^k satisfying

(5.7)
$$\sup_{B_k \in \mathcal{A}} \int_{sp(bd(B_k);\varepsilon)} \phi_V(z)dz = O(\varepsilon) \qquad \varepsilon \to 0 .$$

Here

(5.8)
$$\xi_{s-1,n}(z) = (\sum_{j=0}^{s-3} n^{-j/2} P_j(z)) \phi_V(z) .$$

Let $1 \leq p \leq k$ and define W_n' by

(5.9)
$$W_n' = n(U_n^1 - \mu^1)^T L(U_n^1 - \mu^1) + \sum_{j=1}^{s-3} n^{-j/2} Q_j(n^{1/2}(U_n - \mu))$$

where L is a positive semi-definite matrix of rank p and

$\{Q_j(z)\}_{1\le j\le s-3}$ are polynomials in z, each of which is at least of degree two in $z^1 = (z^{(1)},\dots,z^{(p)})$.

Assume that

(5.10) $$L^T V L = L .$$

Then (i), (ii) and (iii) below hold.

(i) There exist nonnegative integers $k_1,\dots,k_{s-3}$ and constants $\{a_{ij}\}$ not depending on n $(0 \le i \le k_j,\ 1 \le j \le s-3)$ such that

(5.11) $$\sup_{B \in \mathcal{B}} \Big| P(W_n' \in B) - \sum_{j=0}^{s-3} n^{-j/2} \sum_{i=0}^{k_j} a_{ij} \int_B \chi^2(v\,;\,p+2i)dv \Big| = o(n^{-(s-3)/2}) ,$$

for every family $\mathcal{B}$ of Borel subsets of R_+^1 satisfying

(5.12) $$\sup_{B\in\mathcal{B}} \int_{sp(bd(B);\varepsilon)} \chi^2(v\,;\,p)dv = O(\varepsilon) \qquad \varepsilon \to 0 .$$

(ii) If W_n is <u>any other statistic</u> defined on $(\Omega, \mathcal{A}, P)$ such that

(5.13) $$\sup_\omega \{ |W_n(\omega) - W_n'(\omega)| : n\|U_n(\omega) - \mu\|^2 > d \log n\} = o(n^{-(s-3)/2}) ;$$

then the conclusion of (i) holds with W_n' replaced by W_n (with the same choice of $\{k_j\}$ and $\{a_{ij}\}$).

(iii) If $\{P_j\}_{0\leq j\leq s-3}$ *and* $\{Q_j\}_{1\leq j\leq s-3}$ both enjoy the odd-even property (3.8), then (5.5) holds.

REMARK 5.1 Assume the set up of Theorem 5.2 except that we now have a family $\{P_\theta\}$ of probability measures and a family $\{W_n(\theta)\}$ of statistics. Suppose that Assumption (I) holds in the following uniform sense :

$$\sup_{\theta} \sup_{B_k \in \mathcal{A}} \Big| P_\theta(n^{1/2}(U_n - \mu(\theta)) \in B_k) - \int_{B_k} \xi_{s-1,n}(z\,;\,\theta)dz \Big| = o(n^{-(s-3)/2}), \tag{5.14}$$

for every family $\mathcal{A}$ of Borel subsets of R^k satisfying

$$\sup_{\theta} \sup_{B_k \in \mathcal{A}} \int_{sp(bd(B_k);\,\varepsilon)} \phi_{V(\theta)}(z)dz = O(\varepsilon) \quad \text{as } \varepsilon \to 0\,. \tag{5.15}$$

Also assume that there exists a positive constant d (free from θ) such that

$$\sup_{\theta} \sup_{\omega} \{|W_n(\omega;\theta) - W_n'(\omega;\theta)| : n\|U_n - \mu(\theta)\|^2 > d \log n\} = o(n^{-(s-3)/2}), \tag{5.16}$$

and that

(5.17) for each θ, the rank of the positive semidefinite matrix $L(\theta)$ is p, free from θ, $1 \leq p \leq k$.

Then one has

$$\sup_{\theta} \sup_{B \in \mathcal{B}} |P_\theta(W_n(\theta) \in B) - \tag{5.18}$$
$$\sum_{j=0}^{s-3} n^{-j/2} \sum_{i=0}^{k_j} a_{ij} \int_B \chi^2(v ; p + 2i)dv|$$
$$= o(n^{-(s-3)/2}),$$

for every family $\mathcal{B}$ of Borel subsets of R_+^1 satisfying (5.12).

SECTION 6. VALIDITY OF SOME COMMON FORMAL EXPANSIONS

Suppose that we are given a sequence of statistics $\{T_n\}_{n \geq 1}$. We shall say that the distribution function $F_n(x) = \text{Prob}(T_n \leq x)$ of T_n possesses an asymptotic expansion valid up to $o(n^{-r/2})$ if there exist functions $A_0, \ldots, A_r$ not depending on n such that

$$F_n(x) - \sum_{j=0}^{r} n^{-j/2} A_j(x) = o(n^{-r/2}) . \tag{6.1}$$

If moreover (6.1) holds uniformly in $x \in R^1$, we shall say that the expansion is uniformly valid up to $o(n^{-r/2})$. In this work, all the expansions obtained are uniformly valid and hence will be simply referred to as valid.

Since the first term $A_o(x)$ of the expansion (6.1) is often insufficient to get accurate approximation of $F_n(x)$, the problem of getting the higher order terms is considered by various people. The methods used are however not necessarily valid in the sense of (6.1). We shall mainly concentrate on two such formal methods, namely, the method using the approximate moments of T_n and the method using the approximate characteristic functions of T_n. Each of these methods yield explicit (and convenient) determination of the higher order terms of (6.1).

In the rest of this section, we shall assume the set up of Section 2.

(A) METHOD OF APPROXIMATE MOMENTS :

Let r be the highest degree of the polynomial

$$\alpha_{m,n}(v) = \sum_{j=0}^{m} n^{-j} q_j(v) \tag{6.2}$$

where $\{q_j\}$ and m are as in Theorem 2.1. Assume that the moments of Z_1 of order $(s-1)r$ are finite. A Taylor expansion of W_n yields the statistic

$$W_n' = h_{s-1,n}(n^{1/2}(Z_n - \mu))$$

(see (4.4)). Then the first r moments of W_n' are also finite. Expand these moments up to $o(n^{-(s-3)/2})$ and get the "approximat moments" $\delta_{i,n}$ of W_n' (or W_n) :

$$E((W_n')^i) = \delta_{i,n} + o(n^{-(s-3)/2}) \qquad 0 \leq i \leq r \tag{6.3}$$

Using Laguerre polynomials (or any other convenient method) find a polynomial

$$\hat{\alpha}_{m,n}(v) = \sum_{j=0}^{m} n^{-j} \hat{q}_j(v)$$

of degree r ($\hat{q}_j$'s do not depend on n) such that

$$\int_{R^1} v^i \chi^2(v ; p) \hat{\alpha}_{m,n}(v)dv = \delta_{i,n} \qquad 0 \leq i \leq r \tag{6.4}$$

Suppose now that the assumptions of Theorem 2.1(a) hold. Then

$$\hat{\alpha}_{m,n} = \alpha_{m,n} \tag{6.5}$$

so that the formal expansion

$$\chi^2(v ; p) \sum_{j=0}^{m} n^{-j} \hat{q}_j(v) \tag{6.6}$$

of the distribution function of W_n is valid.

PROOF OF (6.5) :

$$E((W_n')^i)$$

$$= \int_{R^k} (h_{s-1,n}(z))^i \, \xi_{s-1,n}(z)dz + o(n^{-(s-3)/2})$$

(6.7)

$$= \int_{R^1} v^i \alpha_{m,n}(v) \, \chi^2(v \, ; \, p)dv + o(n^{-(s-3)/2})$$

The first equality follows from Bhattacharya's result (Theorem 4.1 page 46); while the second can be proved by a slight modification of the proof of Theorem 2.1(a). In view of (6.3) and (6.4) and in view of the fact that both of $\hat{\alpha}_{m,n}$ and $\alpha_{m,n}$ involve terms of order $o(n^{-(s-3)/2})$, one can then conclude that

$$\int v^i \alpha_{m,n}(v) \, \chi^2(v \, ; \, p)dv = \int v^i \hat{\alpha}_{m,n}(v) \, \chi^2(v \, ; \, p)dv$$

$$0 \leq i \leq r$$

Since both of $\hat{\alpha}_{m,n}$ and $\alpha_{m,n}$ are polynomials of degree r, they must therefore be identical.

(B) INVERSION OF APPROXIMATE CHARACTERISTIC FUNCTION

To derive the asymptotic expansion for W_n it is often convenient to evaluate one of the three quantities : the exact characteristic function of W_n, one of the two possible approximate characteristic functions

$E(\exp(it\ W_n'))$ or $\int \exp(it\ h_{s-1,n}(z))\ \xi_{s-1,n}(z)dz \qquad (t \in R^1)$.

A formal inversion of the expansion up to $o(n^{-(s-3)/2})$ of any one of them will lead to a formal expansion for W_n. In general such formal inversions are not valid. Under certain conditions we justify below this formal inversion.

Suppose that the assumptions of Theorem 2.1(a) hold. Then

$$
\begin{aligned}
&E(\exp(it\ W_n)) \\
&= E(\exp(it\ W_n')) + o(n^{-(s-3)/2}) \\
&= \int_{R^k} \exp(it\ h_{s-1,n}(z))\ \xi_{s-1,n}(z)\ dz + o(n^{-(s-3)/2}) \qquad (6.8)\\
&= \int_{R^1} \exp(itv)\ \psi_{m,n}(v)\ dv + o(n^{-(s-3)/2}) \qquad t \in R^1
\end{aligned}
$$

where $\psi_{m,n}(v) = \alpha_{m,n}(v)\ \chi^2(v\ ;\ p)$ (see (6.2)). The first equality follows from the estimate of the tail probability of normalised deviation $n^{1/2}(\bar{Z}_n - \mu)$ due to von Bahr (see (2.31), page 447, Bhattacharya and Ghosh (1978)); the second follows from Bhattacharya's result (Theorem 4.1, page 46); and finally the third can be obtained by following the proof of Theorem 2.1(a).

Let

$$\hat{\psi}_{m,n}(v) = \sum_{j=0}^{m} n^{-j}\ \hat{q}_j(v)\ \chi^2(v\ ;\ p) \qquad v \in R^1$$

($\hat{q}_j$ are polynomials with coefficients free from n) be such that

$$\int_{R^1} \exp(itv)\ \hat{\psi}_{m,n}(v)\ dv$$

$$= E(\exp(it\ W_n)) + o(n^{-(s-3)/2}).$$

Then from (6.8) it follows that the Fourier-Stieltjes Transforms of $\hat{\psi}_{m,n}$ and $\psi_{m,n}$ agree up to $o(n^{-(s-3)/2})$. As each of these involves only terms of order $o(n^{-(s-3)/2})$, they must be identical. In other words, the formal expansion $\hat{\psi}_{m,n}$ for W_n is indeed valid.

SECTION 7. COUNTEREXAMPLES

This section is devoted to five examples related to the assumptions of Theorem 2.1. The first example shows that in Theorem 2.1(a), Assumption A_s(iv) cannot in general be dropped. On the other hand the second example shows that this assumption is only a sufficient condition for Theorem 2.1(a). The third example illustrates the important fact that even when Theorem 2.1 as such may not be applicable, the method employed in its proof can be applied. The fourth one points out the need to verify the odd-even property (3.8) (see Theorems 5.1 and 5.2).

EXAMPLE 7.1 Suppose that $\{Z_n\}_{n\geq 1}$ are IID two-dimensional vectors and that $Z_1^{(1)}$, $Z_1^{(2)}$ are independent $N(0;1)$. Let $H(z) = \frac{1}{2}(z^{(1)})^2 + \frac{1}{2}(z^{(2)})^3$. Then, with $s = 5$, all the assumptions of Theorem 2.1 hold except A_s(iv). We shall show that the conclusion of Theorem 2.1(a) (and hence that of Theorem 2.1(b)) does not hold.

Clearly W_n has the same distribution as $X^2 + n^{-1/2}Y^3$ where X, Y are IID $N(0,1)$. Fix a, b such that $0 < a < b < \infty$ and let $a \leq x \leq b$. Put

$$A_n = \{(w,y) \mid a \leq w \leq x, \quad y^2 < 3 \log n\}$$

$$A = \{(w,y) \mid a \leq w \leq x, \quad -\infty < y < \infty\}.$$

Then for all sufficiently large n, A_n is a subset of

$$\{(w,y) : w - n^{-1/2} y^3 > 0\}.$$

One has, uniformly in x,

$$\begin{aligned} & \text{Prob } (a \leq W_n \leq x) \\ = \; & \text{Prob } (a \leq W_n \leq x, \; Y^2 < 3 \log n) + o(n^{-1}) \\ = \; & \int_{A_n} (2\pi)^{-1} (w - n^{-1/2}y^3)^{-1/2} \exp(-\tfrac{1}{2}(w-n^{-1/2}y^3) - \tfrac{1}{2}y^2) \, dwdy \\ & + o(n^{-1}) \end{aligned}$$

$$= \int_{A_n} (2\pi)^{-1} w^{-1/2} \exp\{-\tfrac{1}{2}(w+y^2)\} f_n(w,y)\,dwdy + o(n^{-1})$$

$$= \int_{A} (2\pi)^{-1} w^{-1/2} \exp(-\tfrac{1}{2}(w+y^2)) f_n(w,y)\,dwdy + o(n^{-1})$$

$$(7.1) = \int_a^x (2\pi)^{-1/2} w^{-1/2} \exp(-\tfrac{1}{2}w)(1+n^{-1}a_1+n^{-1}\frac{a_2}{w} + n^{-1}\frac{a_3}{w^2})dw + o(n^{-1})$$

where a_2, a_3 are <u>nonzero</u> constants and

$$f_n(w,y) = (1 + \frac{n^{-1/2}y^3}{2} + \frac{n^{-1}y^6}{8})(1+\frac{n^{-1/2}y^3}{2w} + \frac{3n^{-1}y^6}{8w^2}).$$

Clearly (2.9) is incompatible with (7.1).

(If $0 < x < a$, then there exists a valid expansion of $\text{Prob}(W_n \leq x)$ but it is not quite identical with the above expansion and moreover it does not hold uniformly in x.)

REMARK 7.1 Example 7.1 suggests that even when Assumption A_s(iv) does not hold, alternative expansions in powers of n^{-1} are available. It is not however clear whether expansions of this sort may be easily obtained by some formal technique. We shall not explore this question any further.

EXAMPLE 7.2 Take Z_i's as in Example 7.1 and let $H(z) = \frac{1}{2}(z^{(1)} + (z^{(2)})^2)^2$ so that $W_n = (X + n^{-1/2} Y^2)^2$ where X, Y are as in Example 7.1. Then considering the Edgeworth expansion of $U_n = X + n^{-1/2} Y^2$ (see Theorem 5.2) it can be shown that W_n has an expansion of the form (2.9). But evidently H does not satisfy A_s(iv) (for $s = 4$).

EXAMPLE 7.3 Let Z_i's be as in Example 7.1 and take $H(z) = \frac{1}{2}\|z\|^2 + \frac{1}{2}(z^{(1)})^5 \|z\|^{-1}$ so that W_n has the same distribution as that of

$$X^2 + Y^2 + n^{-1} X^5(X^2 + Y^2)^{-1/2}$$

where X, Y are as in Example 7.1. Applying the polar transformation $X = r\cos\theta$, $Y = r\sin\theta$, W_n can be written as $r^2 + n^{-1}r^4 \times \cos^5\theta$ so that W_n possesses an expansion of the form (2.9).

EXAMPLE 7.4 Let $\{Z_i\}_{i\geq 1}$ be IID $N(0;1)$ on R^1 and put $W_n = (\sqrt{n}\,\bar{Z}_n)^2 \cdot (1 + n^{-1/2})$. Then

$$\text{Prob}(W_n \leq u) = \int_0^u \chi^2(v\,;\,1)\left(1 + \frac{v^2-1}{2\sqrt{n}}\right)dv + o(n^{-1/2}) \qquad u \in R_+^1$$

so that the coefficient of $n^{-1/2}$ does not vanish.

EXAMPLE 7.5 Let $X_1, \ldots, X_6$ be IID $N(0;1)$ and put $W_n = X_1^2 + \ldots + X_5^2 + n^{-1/2} X_6$. Then one can show using arguments of Example 7.1 that

$$P(W_n \leq u)$$
$$= \int_0^u \chi^2(v;5)dv + \frac{1}{8n}\int_0^u(\chi^2(v;5) - 2\chi^2(v;3) + \chi^2(v;1))dv$$
$$+ o(n^{-1}), \text{ uniformly in } u \in R_+^1 .$$

Thus even when $A_s(iv)$ does not hold (here $s = 4$) but $p \geq 5$, W_n can possess an asymptotic expansion (up to $o(n^{-1})$) such that the coefficient of n^{-1} is a finite linear combination of chi-squres with degrees of freedom $p-4$, $p-2$, p etc.

PART II

CHAPTER TWO

VALID ASYMPTOTIC EXPANSIONS FOR THE LIKELIHOOD RATIO AND OTHER STATISTICS

SECTION 1. INTRODUCTION AND NOTATIONS

The general theorems obtained in the previous chapter can be applied to get valid asymptotic expansions of the distribution functions of many statistics whose limiting distributions are central chi-squares. The most important example of such a statistic is the (transformed) likelihood ratio criterion proposed by Neyman and Pearson (1928); others are the statistics due to A. Wald (1945) and C.R. Rao (1948) (see Rao (1965), pages 347-352). We show under very general conditions that these three statistics indeed satisfy the assumptions of our Theorem 5.2 of Chapter One and hence that they possess an asymptotic expansion up to any degree of accuracy. We consider separately the case when the samples come from an (absolutely continuous) exponential family with natural parameter space, since the proof for the general case is quite cumbersome, is based on some tedious approximations and uses a rather deep result of Bhattacharya and Ghosh (1978); for the exponential family it is moreover shown that in typical cases the expansions are valid uniformly over all Borel subsets of R^1_+. These examples show that our assumptions are not too restrictive; in fact we do not at present know [redacted] any useful statistic whose limiting distribution is central chi-square and to which our theorems are not applicable. To apply Theorem 5.2 one has to make a natural choice for U_n; for the above three statistics it is shown here that the natural choice is the unrestricted maximum likelihood estimators together with a few of partial derivatives at θ_0 of the loglikelihood function.

It is worthwhile to mention that our theorems always yield valid expansions (see Section 6, page 55, Chapter One). All the previous expansions (for these statistics) obtained by various people are only formal. One of earliest reference is Box (1949) and a recent one is Hayakawa (1977) where further references can be found (see also Korin (1968)). Most of these expansions are obtained formally by inverting an approximate characteristic function or by equating the first few moments of the exact and approximating distributions. In general, the above procedures cannot be justified. Consequently the validity of the formal expansions in the literature has remained an interesting open problem. That under the assumptions of our main Theorem 2.1 such formal expansions are valid was proved in Section 6, Chapter One. However the expansions obtained by Box (1949) and Hayakawa (1977) do not exactly fit in the set up of Theorem 2.1 and are justified in Sections 3 and 4(C). It is to be noted that although methods of Box and Hayakawa are formal, whenever their methods are applicable they are more suitable and less tedious (for determining the expansion explicitly) than the method employed in the proof of our main theorem of Chapter One.

Below we list some of the notations to be followed in the rest of this chapter. Let $\{Y_n\}_{n\geq 1}$ be IID m-dimensional random vectors with common density $f(y;\theta)$ with respect to some sigma finite measure μ_o, where $\theta = (\theta^{(1)},\ldots,\theta^{(k)})$ takes values in some subset Θ of R^k. Let $1 \leq p \leq k$ and put

(1.1) $$\theta^1 = (\theta^{(1)},\ldots,\theta^{(p)}), \qquad \theta^2 = (\theta^{(p+1)},\ldots,\theta^{(k)}) .$$

Consider the problem of testing

(1.2) $$H_o : \theta^1 = \theta^1_o \quad \text{vs.} \quad H_1 : \theta^1 \neq \theta^1_o$$

where θ^1_o is a specified element of R^p. We write

$$L_n(\theta) = n^{-1} \sum_{j=1}^{n} \log f(y_j;\theta) \qquad L_1(\theta) = \log f(y;\theta)$$

(1.3) $$\lambda_n = 2n[\sup_{\theta\,\varepsilon\,\Theta} L_n(\theta) - \sup_{\theta\,\varepsilon\,\Theta_o} L_n(\theta)]$$

$$\Theta_o = \{\theta \,\varepsilon\, \Theta \mid \theta^1 = \theta^1_o\}$$

Thus $L_n(\theta)$ is the average loglikelihood function when the sample is $(y_1,\ldots,y_n)$ and λ_n is the transformed likelihood ratio statistic

($\exp(-\lambda_n/2)$) is the conventional likelihood ratio statistic). The unrestricted (restricted) maximum likelihood estimators will be denoted by $\tilde{\theta}_n$ ($\hat{\theta}_n$, respectively).

$$L_n(\tilde{\theta}_n) = \sup_{\theta \in \bar{\Theta}} L_n(\theta), \qquad L_n(\hat{\theta}_n) = \sup_{\theta \in \bar{\Theta}_o} L_n(\theta)$$

(1.4)
$$\hat{\theta}_n^1 = \theta_o^1 \quad , \qquad \hat{\theta}_n = (\hat{\theta}_n^1 \mid \hat{\theta}_n^2)$$

To define the criteria proposed by Wald and Rao we need some further notations :

$$L_n^{ij}(\theta) = D^i D^j L_n(\theta), \qquad I_{ij}(\theta) = E_\theta(L_n^{ij}(\theta)) \quad 1 \leq i,j \leq k$$

$$I(\theta) = ((I_{ij}(\theta)))_{1 \leq i,j \leq k}, \quad I^*_{11}(\theta) = ((I_{ij}(\theta)))_{1 \leq i,j \leq p}$$

(1.5)
$$I(\theta) = \begin{pmatrix} I^*_{11}(\theta) & I^*_{12}(\theta) \\ I^*_{21}(\theta) & I^*_{22}(\theta) \end{pmatrix}$$

$$I_{11.2}(\theta) = I^*_{11}(\theta) - I^*_{12}(\theta)\,(I^*_{22}(\theta))^{-1}\, I^*_{21}(\theta)$$

$$\phi_i(\theta) = n^{1/2} D^i L_n(\theta) \quad 1 \leq i \leq k, \qquad \phi(\theta) = (\phi_1(\theta), \ldots, \phi_k(\theta))$$

Thus $-I(\theta)$ is the conventional Fisher Information matrix when θ obtains and $\phi_i(\theta)$ is the i-th efficient score of θ. Then Wald's statistic is

(1.6)
$$W_n = -n(\tilde{\theta}^1 - \theta_o^1)^T I_{11.2}(\tilde{\theta})(\tilde{\theta}^1 - \theta_o^1)$$

and Rao's statistic is

(1.7) $$S_n = -\phi(\hat{\theta}_n)^T I^{-1}(\hat{\theta}_n)\,\phi(\hat{\theta}_n).$$

SECTION 2. EXPANSIONS FOR EXPONENTIAL DENSITIES

Let $\{Y_n\}_{n\geq 1}$ be m-dimensional random vectors with density

(2.1) $$f(y\,;\,\theta) = \exp\left(\sum_{j=1}^{k} \theta^{(i)} f_i(y) - c(\theta)\right)$$

with respect to some sigma finite measure μ_o where $f_1,\ldots,f_k$ are continuously differentiable real valued functions on R^m and θ takes values in $\bar{H}$, the natural parameter space. We assume that μ_o has a nonzero absolutely continuous component (with respect to Lebesgue measure on R^m) whose density is positive on an open set $U \subset R^m$ and $1, f_1,\ldots,f_k$ are linearly independent as elements of the vector space of continuous functions on U. Then

(2.2) $$Z_1 = (f_1(Y_1),\ldots,f_k(Y_1))$$

satisfies CONDITION D (see (2.8), Chapter One) under each $\theta \in \bar{H}$. We assume that $\bar{H}$ has a nonempty interior. Consider the testing problem (1.2) where θ_o is an interior point of $\bar{H}_o$. We want to establish the existence of a valid asymptotic expansion for the distribution of the transformed likelihood ratio statistic λ_r (defined in (1.3)) under θ_o.

THEOREM 2.1. For all integers $s \geq 4$, there exist polynomials $\{q_r\}_{0 \leq r \leq m}$ (in one variable) with coefficients free from n such that the following expansion holds uniformly over all Borel subsets B of R^1_+ :

$$(2.3) \qquad P(\lambda_n \in B ; \theta_o) = \sum_{i=0}^{m} n^{-i} \int_B \chi^2(v;p)\, q_i(v)dv + \varepsilon_n$$

where m, ε_n and $\chi^2(v;p)$ are as in Theorem 2.1 of Chapter One (see page 18).

PROOF : Without loss of generality we may take θ_o to be the origin. The likelihood equations for the unrestricted ML estimators $\tilde{\theta}$ $(\equiv \tilde{\theta}_n)$ are

$$(2.4) \qquad D^j c(\theta) = \bar{Z}_n^{(j)} \qquad 1 \leq j \leq k$$

and the corresponding equations for the restricted ML estimators $\hat{\theta}$ $(\equiv \hat{\theta}_n)$ are

$$(2.5) \qquad D^j c(\theta) = \bar{Z}_n^{(j)}\ , \quad \theta^1 = \theta_o^1\ , \quad p+1 \leq j \leq k\ .$$

Let $\mu = E(Z_1 ; \theta_o)$. If in (2.4) and (2.5) we replace $\bar{Z}_n^{(j)}$ by $\mu^{(j)}$, then they have a solution $\tilde{\theta} = \theta_o$, $\hat{\theta} = \theta_o$ respectively. Since the $k \times k$ matrix whose (i,j)th element is $D^i D^j c(\theta)$ is positive-definite, it follows by the Implicit Function Theorem (see page 272, Dieudonné (1969)) that there is a bounded convex neighbourhood N of μ such that if $\bar{Z} \in N$, then both of (2.4) and (2.5) have unique solutions $\tilde{\theta}$ and $\hat{\theta}$ respectively and tha

(2.6) $$\sup_{\theta \in \bar{\Theta}} L_n(\theta) = L_n(\tilde{\theta}), \quad \sup_{\theta \in \bar{\Theta}_o} L_n(\theta) = L_n(\hat{\theta}).$$

Since θ_o is an interior point of $(\bar{\Theta})_o$, by Chernoff's Theorem (see Theorem 3.1, page 7 of Bahadur (1971) and Bártfai (1977)) we may assume that

(2.7) $$P(\bar{Z}_n \notin N ; \theta_o) = o(n^{-s}) \quad \text{for all} \quad s > 0 .$$

Thus λ_n can be regarded as well-defined. Now in a suitable neighbourhood $N_1 (\subset N)$ of μ, $\bar{Z}_n$ and $\hat{\theta}_n$ can be written as functions of $\tilde{\theta}_n$, the functions themselves being free from n; consequently we may write using equation (2.4)

(2.8) $$\begin{aligned}\lambda_n &= 2n[\sum_{i=1}^{k} \tilde{\theta}^{(i)} \bar{Z}_n^{(i)} - c(\tilde{\theta}) - \sum_{j=p+1}^{k} \hat{\theta}^{(j)} \bar{Z}_n^{(j)} + c(\hat{\theta})] \\ &= 2n\, H(\tilde{\theta}), \quad \text{say,}\end{aligned}$$

(H is defined only on N_1).

We now apply Theorem 5.1 of Chapter One (with U_n replaced by $\tilde{\theta}_n$). Differentiating the likelihood equation (2.4) for $\tilde{\theta}$ with respect to $\bar{Z}_n^{(j)} (1 \leq j \leq k)$ one gets,

(2.9) $$I_{k\times k} = ((D^i D^j\, c(\tilde{\theta}))) \, ((\partial \tilde{\theta}^{(i)} / \partial \bar{Z}_n^{(j)}))$$

and consequently the second matrix on the right side of (2.9) (evaluated at $\mu = E(Z_1 ; \theta_o)$) is positive-definite. Since by a version of the Implicit Function Theorem, θ can be chosen to be an analytic function of $\bar{Z}_n$ and since Z_1 satisfies CONDITION D, it therefore follows from Theorem 2(a) and Remark 1.1 of Bhattacharya and Ghosh (1978) that $n^{1/2}(\tilde{\theta} - \theta_o)$ has a multivariat Edgeworth expansion which holds uniformly over all Borel subsets of R^k; in fact, as $\tilde{\theta}$ and $\bar{Z}_n$ both have the same dimension k, the expansion for $\tilde{\theta}$ is almost immediate from that for $\bar{Z}_n$. It therefore remains to verify Assumptions A_s(i) - (v) for the function H defined in (2.8). Since c is analytic(implying as before that $\tilde{\theta}$ and $\hat{\theta}$ can be taken to be analytic) H is also analytic. It is well-known that the limiting distribution of λ_n is a central chi-square with p degrees of freedom. Thus A_s(i), (ii) and (iii) hold. (That A_s(ii) and (iii) hold can as well be verified otherwise; one convenient way is to expand all functions of $\hat{\theta}$ around $\tilde{\theta}$ and use the likelihood equations; see, e.g., the second paragraph, page 85). For A_s(iv) and (v), it is sufficient to check that (see Lemma 2.1(c), Chapter One)

$$\partial H(\tilde{\theta})/\partial\tilde{\theta}^{(i)} = 0 , \quad 1 \leq i \leq p ,$$

for all $\tilde{\theta}$ such that $\tilde{\theta}^1 = \theta_o^1$. To this end, note in view of (2.4) that

$$H(\tilde{\theta}) = \sum_{j=1}^{k} \tilde{\theta}^{(j)} D^j c(\tilde{\theta}) - c(\tilde{\theta}) + \sum_{j=p+1}^{k} \hat{\theta}^{(j)} D^j c(\hat{\theta}) + c(\hat{\theta})$$

and so if $1 \leq i \leq p$

$$\partial H(\breve{\theta})/\partial\breve{\theta}^{(i)}$$

$$= \sum_{j=1}^{p} \breve{\theta}^{(j)} D^1 D^j c(\breve{\theta}) + \sum_{j=p+1}^{k} (D^j c(\hat{\theta}) - D^j c(\breve{\theta})) \frac{\partial \hat{\theta}^{(j)}}{\partial \breve{\theta}^{(i)}}$$

$$= 0 \qquad \text{if } \breve{\theta}^1 = \underset{p\times 1}{0} ;$$

in the last equality we have used the fact that when $\breve{\theta}^1 = 0$, $\hat{\theta} = \breve{\theta}$.

By Theorem 5.1 of Chapter One and (2.7), the proof of the theorem is complete.

REMARK 2.1 We have verified Assumptions A_s(iv) and (v) by writing λ_n as a function of the unrestricted ML estimators $\breve{\theta}$. Clearly λ_n can as well be regarded as a function of $\bar{Z}_n$. However considering the problem of testing the hypothesis that the population variance is one against the alternative that it is not one, the population mean being unknown and the observations coming from a normal population, one can easily verify that Assumption A_s(iv) need not hold if λ_n is regarded as a function of $\bar{Z}_n$. In case of a simple null hypothesis, these two assumptions will always hold and consequently the above theorem can be proved without using the Edgeworth expansion for $\breve{\theta}$ (There may, however, be cases where even if the null hypothesis is composite and λ_n is regarded as a function of $\bar{Z}_n$, A_s(iv) and (v) hold. This is the case when, e.g., one is testing for the mean of a normal population with unknown variance).

REMARK 2.2 The above theorem holds good (with the same proof) if λ_n is replaced by Wald's or Rao's statistic (see (1.6) and (1.7)). In fact, the case of Wald's statistic can be settled rather easily.

REMARK 2.3 It can be shown that the expansions of the distribution functions of λ_n, W_n and S_n hold uniformly over all θ_o such that θ_o^2 lies in some compact subset of Θ_o.

SECTION 3. VALIDITY OF THE METHOD OF BOX (1949)

In Section 2.1 of his (1949) paper, Box considered the problem of testing constancy of variances or covariances of k sets of p-variate normal populations and derived an 'asymptotic chi-square series solution' of the null distribution of the test statistic M (see equations (4), (5), page 320) which is a generalised form of Bartlett's statistic. We use our results of Chapter One to show that Box's asymptotic series is in fact a valid one. In the rest of this section we shall follow the notations of Box (unless otherwise stated).

We describe briefly the approach of Box. He first derives an asymptotic expansion of the logarithm of the exact characteristic function $\Phi(t)$ of ρM (ρ is a constant which may depend on μ) and uses it to deduce that

$$\Phi(t) = \Phi_n(t) + o(\mu^{-n}) \qquad t \in R^1$$

where

$$\Phi_n(t) = K(1-2it)^{-f/2} \sum_{v=0}^{n} \mu^{-v} a_v (1-2it)^{-v} ,$$

K being a constant depending on μ and $f = (k-1)p(p+1)/2$, the degrees of freedom of the limiting (chi-square) distribution of ρM. This part of Box's argument is rigorous (at least can be made rigorous without any difficulty). A formal inversion now gives an asymptotic expansion of the density $p(x)$ of ρM:

$$p(x) = p_n(x) + o(\mu^{-n})$$

$$p_n(x) = K \sum_{v=0}^{n} \mu^{-v} a_v \chi^2(x ; f + 2v)$$

where $\chi^2(. ; f+2v)$ is the density of a chi-square variable with $(f + 2v)$ degrees of freedom. This step is in general unjustifiable. Box gets the final form of his series solution by writing K asymptotically in a series of μ and rearranging the product of the two resulting series (see equation (30), page 323); obviously the last part of Box's argument can be justified easily.

To establish the validity of the above formal inversion of characteristic function, take (in Theorem 5.1) ρM to be our W_n and let $\hat{\theta}_n$ be the vector of $s_{ij\lambda}$'s (the usual unbiased estimators of the covariances between the i-th and j-th variable in the λ-th sample). Define the vector Z_1 in an obvious manner (see our relation (2.2), page 68). Finally let $\xi_{s-1,n}$ and $\xi'_{s-1,n}$ be respectively the Edgeworth expansions of $n^{1/2}(\hat{\theta}_n - \theta_o)$ and $n^{1/2}(\bar{Z}_n - E(Z_1 ; \theta_o))$ up to terms of order $o(n^{-(s-3)/2})$. In view of Section 6(B) of Chapter One, it is enough to show that

(3.1)
$$E(\exp(it\, W_n); \theta_o) = \int_{R^1} \exp(itv) \sum_{j=0}^{m} n^{-j} q_j(v) \chi^2(v ; f+2j)dv + o(n^{-(s-3)/2}).$$

It should be noted that the argument given in Section 6(B) of Chapter One does not apply. Define the functions g_n and g'_n on R^k such that (see, e.g., equation (4.3), Chapter One)

$$W_n = g_n(n^{1/2}(\hat{\theta} - \theta_o)), \quad W_n = g'_n(n^{1/2}(\bar{Z}_n - E(Z_1 ; \theta_o))).$$

Then the left side of (3.1) is

$$\int_{R^k} \exp(itg'_n(z))\, \xi'_{s-1,n}(z)dz + o(n^{-(s-3)/2})$$

$$= \int_{R^k} \exp(itg_n(\theta))\, \xi_{s-1,n}(\theta)d\theta + o(n^{-(s-3)/2})$$

The first step follows from Bhattacharya's result Theorem 4.1 (page 46 , Chapter One) while the second follows from a multivariate analogue of the first relation in equation (2.10) of

Lemma 2.1 of Bhattacharya and Ghosh (1978) (in fact, since $\hat{\theta}$ and $\bar{Z}_n$ both have the same dimension k, the proof is quite simple). Relation (3.1) now follows by arguments similar to those used in the proof of Theorem 5.2, Chapter One.

It is interesting to note that Box in the last paragraph remarked : "we see in effect we are finding a χ^2-series to the statistic M by arranging that to the order of accuracy chosen in the asymptotic series, the series will have all its cumulants identical with those of M". He, however, did not supply a proof of his remark ; for this, one would need to show that the formal differentiation of identity (18), page 322 of Box's paper is permissible. This seems difficult but an alternative proof is given below. Since all the moments of $M(= W_n)$ are finite and since the rth cumulant is a polynomial function of the first r moments $(r \geq 1)$, to establish Box's remark it is enough to show for each $r \geq 1$ that

$$(3.2) \qquad E(W_n^r) = \int_{R^1} v^r \sum_{j=0}^{m} n^{-j} q_j(v) \chi^2(v;p)dv + o(n^{-(s-3)/2}).$$

Since W_n can be bounded in absolute value by a polynomial in $s_{ij\ell}$'s, Theorem 20.1 of Bhattacharya and Ranga Rao (1976) implies that

$$(3.3) \qquad E(W_n^r I_{M_n^c}) = o(n^{-(s-3)/2})$$

where $I_{M_n^c}$ is the indicator function of the complement of $M_n = \{\|n^{1/2}(\bar{Z}_n - E(Z_1;\theta_0))\|^2 < (s-2)\lambda \log n\}$ (see (4.5), Chapter One). Let k_{s-1} be the Taylor expansion of K up to terms involving (s-1)th order derivatives where $\hat{\theta} = K(\bar{Z}_n)$. Let h_{s-1} be the similar expansion of H where $W_n = 2n(H(\hat{\theta}_n) - H(\theta_o))$.

Then

$$E(W_n^r I_{M_n}) = \int_{M_n} (h_{s-1}(k_{s-1}(z)))^r \xi'_{s-1,n}(z)dz + o(n^{-(s-3)/2})$$

$$= \int_{N_n} (h_{s-1}(\theta))^r \xi_{s-1,n}(\theta)d\theta + o(n^{-(s-3)/2})$$

(3.4) $$= \int_{R^1} v^r \sum_{j=0}^{m} n^{-j} q_j(v) \chi^2(v;p)dv + o(n^{-(s-3)/2})$$

Here $N_n = \{\|n^{1/2}(\hat{\theta}_n - \theta_o)\|^2 < (s-2)\lambda \log n\}$. Together (3.3) and (3.4) imply (3.2).

SECTION 4. EXPANSIONS IN THE GENERAL CASE

Valid asymptotic expansions for the distribution functions of the likelihood ratio statistic as well as of Wald's and Rao's statistics in the general case can be obtained, using Theorem 5.2 of Chapter One, up to any degree of accuracy <u>provided</u> suitable regularity conditions hold. Although the essential idea of the proof is quite simple, some tedious approximations are necessary to verify the technical conditions imposed in the above mentioned theorem.

Consider the testing problem described in the last part of Section 1. Below we shall consider only the expansion for the LR statistic since that for Wald's statistic or Rao's statistic is quite similar. In fact for the case of Wald's statistic the expansion is relatively simpler to arrive at (using Theorem 3 of Bhattacharya and Ghosh (1978)); the same is true for the LR and Rao's statistics when H_0 is simple. It is instructive to work out these special cases first. (See Section 5, Chandra and Ghosh(1979)).

(A) Basic Assumptions :

Let $G(\cdot\,;\theta)$ be the distribution of Y_1 under θ. Assume that Θ is open in R^k. We shall write $P(\cdot\,;\theta)$ to denote the product probability measure on the space Ω of all sequences in R^m and regard Y_n's as coordinate maps on this space. The following assumptions (see Bhattacharya and Ghosh (1978)) will be made.

ASSUMPTION B_s :

(i) There is an open subset U of R^m such that (a) for each $\theta \in \Theta$ one has $G(U;\theta) = 1$ and (b) for each $\alpha = (\alpha^{(1)},\ldots,\alpha^{(k)})$, $1 \leq |\alpha| \leq s+1$, $L_1(y;\theta)$ has a αth derivative $D^\alpha L_1(y;\theta)$ with respect to θ on $U \times \Theta$;

(ii) For each α, $1 \leq |\alpha| \leq s$, $E(|D^{\alpha} L_1(Y_1;\theta)|^s ; \theta)$ is finite and there exists an $\varepsilon > 0$ such that

$$(4.1) \qquad E\{\max_{|\theta-\theta_0|\leq\varepsilon} |D^{\alpha} L_1(Y_1;\theta)|^s ; \theta_0\} < \infty \quad \text{if} \quad |\alpha| = s+1 ;$$

(iii) For each $\theta \in \textcircled{H}$, $E(D^r L_1(Y_1;\theta);\theta) = 0$ for $1 \leq r \leq k$ and the $k \times k$ matrices

$$(4.2) \qquad \begin{array}{l} ((E(D^i D^j L_1(Y_1;\theta) ; \theta))) \\ ((E(D^i L_1(Y_1;\theta) D^j L_1(Y_1;\theta); \theta))) \end{array}$$

are nonsingular ;

(iv) The functions

$$(4.3) \qquad \begin{array}{ll} E(D^i D^j L_1(Y_1;\theta); \theta) & 1 \leq i, j \leq k \\ E(D^{\alpha} L_1(Y_1;\theta) D^{\alpha'} L_1(Y_1;\theta);\theta) & 1 \leq |\alpha|, |\alpha'| \leq s \end{array}$$

are continuous on $\textcircled{H}$;

(v) For each $\theta \in \textcircled{H}$, $f(y ; \theta)$ is strictly positive on U. Also for each $\theta \in \textcircled{H}$ and each α, $1 \leq |\alpha| \leq s$, $D^{\alpha} f(y ; \theta)$ is continuously differentiable in y on U.

(vi) The map $\theta \to G(\cdot;\theta)$ on $\textcircled{H}$ into the space of all probability measures on R^m is continuous when the latter space is given the variation norm topology.

(Remark 2.4 of Chapter One and Remark 1.6 of Bhattacharya and Ghosh will be helpful.)

Under the above assumptions, Bhattacharya and Ghosh (1978) proved that

(a) there exists a sequence of statistics $\{\hat{\theta}_n\}_{n\geq 1}$ such that the probability under θ of the event

(4.4)
$$\|\hat{\theta}_n - \theta\| < d_o\, n^{-1/2}(\log n)^{1/2}, \quad \hat{\theta}_n \text{ satisfies}$$
$$\text{the equation } D^i L_n(\theta) = 0, \quad 1 \leq i \leq k$$

is $1 + o(n^{-(s-2)/2})$.

(b) there exist polynomials $q_{r,\theta}$ (in k variables) not depending on n, such that for every sequence $\{\hat{\theta}_n\}$ satisfying the property stated in (a), one has the following asymptotic expansion :

(4.5)
$$P(n^{1/2}(\hat{\theta}_n - \theta) \in B_k ; \theta)$$
$$= \int_{B_k} (1 + \sum_{r=1}^{s-2} n^{-r/2} q_{r,\theta}(x))\, \phi_M(x)dx + o(n^{-(s-2)/2})$$

uniformly over every class $\mathcal{A}$ of Borel sets of R^k satisfying

(4.6) $$\sup_{B_k \in \mathcal{A}} \int_{sp(bd(B_k);\varepsilon)} \phi_M(x)dx = O(\varepsilon) \quad \text{as} \quad \varepsilon \to 0 .$$

Here M is the dispersion matrix of the limiting distribution of $n^{1/2}(\hat{\theta} - \theta)$. For a more precise statement, see their Theorem 3. Of course, a similar result is true for the 'restricted ML estimator' $\hat{\theta}_n$. Henceforth we shall assume familiarity with Theorem 3 (and its proof) of Bhattacharya and Ghosh (1978).

It may be noted that the existence of a measurable choice of ML estimator can be established using the selection theorems of Kuratowski and Ryll-Nardzewski and Kunugui and Novikov (see Theorem 1.4, page 462 and Theorem 3.4, page 471 of Kuratowski and Mostowski (1976); note that Theorem 3.4 remains true even if Y there is assumed to be Polish).

(B) EXPANSION FOR THE LIKELIHOOD RATIO STATISTIC :

We shall discuss first the case $p = 1$, $k = 2$ and obtain an expansion up to $o(n^{-1})$. Choose and fix a $\theta_o \in (\underline{H})_o$; assume that θ_o is the origin. The following notations will be used.

$$L_n^{i_1 \ldots i_j}(\theta) = D^{i_1} \ldots D^{i_j} L_n(\theta),$$

$$I_{i_1 \ldots i_j}(\theta) = E(L^{i_1 \ldots i_j}(\theta); \theta)$$

$$L_n^{i_1 \ldots i_j} = L_n^{i_1 \ldots i_j}(\theta_o), \quad I_{i_1 \ldots i_j} = I_{i_1 \ldots i_j}(\theta_o)$$

$$j \geq 1, \quad i_1, \ldots, i_j = 1 \text{ or } 2 .$$

An expansion of the likelihood equation for $\hat{\theta}^{(2)}$ around θ shows that

$$n^{1/2}(\hat{\theta}^{(2)} - \tilde{\theta}^{(2)})$$

$$= -n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})\, I_{21}^{(\tilde{\theta})}/I_{22}(\tilde{\theta})$$

$$+ n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})\,(n^{-1/2} R_{11} + n^{-1} R_{12})$$

$$+ n^{1/2}(\hat{\theta}^{(2)} - \tilde{\theta}^{(2)})\,(n^{-1/2} R_{21} + n^{-1} R_{22}) + o(n^{-1}),$$

on a set A_n. Here R's are polynomials in $n^{1/2}(\hat{\theta} - \tilde{\theta})$ and U_n whose coefficients do not depend on n, U_n being the vector whose components are

$$n^{1/2}(L_n^{i_1 \ldots i_j}(\tilde{\theta}) - I_{i_1 \ldots i_j}(\tilde{\theta})) \quad , \quad j = 2,3,4, \quad i_1, \ldots i_j = 1 \text{ or } 2;$$

also A_n is the set where $\hat{\theta}$ and $\tilde{\theta}$ satisfy their respective likelihood equations <u>and</u> moreover the following inequalities are true :

$$\|n^{1/2}(U_n^* - E(U_n^* ; \theta_o))\|^2, \quad \|n^{1/2}(\hat{\theta} - \theta_o)\|^2,$$

$$\|n^{1/2}(V_n - E(V_n ; \theta_o))\|^2 \text{ are each } < 3 \log n,$$

U_n^* being the vector whose components are

$$\tilde{\theta},\ L_n^{i_1 i_2},\ L_n^{i_1 i_2 i_3},\ L_n^{i_1 i_2 i_3 i_4}, \qquad i_1, i_2, i_3, i_4 = 1 \text{ or } 2$$

and V_n the vector whose components are

$$\sum_{i'=1}^{n} [\sup_{|\theta - \theta_o| \leq \varepsilon} L_n^{i_1 \ldots i_5}(\theta ; Y_{i'})] \quad i_1, \ldots, i_5 = 1 \text{ or } 2 .$$

Under the regularity assumptions stated below, it can be shown that $P(A_n ; \theta_o) = 1 + o(n^{-1})$.

One can verify that on the set A_n the following is true: for any sequence $\{t_n\}$ such that $o \leq t_n \leq 1$ and for any $i_j = 1$ or 2, $1 \leq j \leq 5$,

$$L_n^{i_1 i_2 i_3 i_4 i_5}(\theta_o + t_n(\hat{\theta} - \theta_o)) \text{ is bounded.}$$

This fact will be used repeatedly without any explicit mention.

(One can verify that the polynomials R_{ij} are given by

$$-I_{22}(\tilde{\theta})\, R_{i1} = n^{1/2}(L_n^{2i} - I_{2i}) + \frac{1}{2} \sum_{j=1}^{2} n^{1/2}(\hat{\theta}^{(j)} - \tilde{\theta}^{(j)})\, I_{2ij}$$

$$-I_{22}(\tilde{\theta})\, R_{i2} = \frac{1}{2} \sum_{j=1}^{2} n^{1/2}(\hat{\theta}^{(j)} - \tilde{\theta}^{(j)})\, n^{1/2}(L_n^{2ij} - I_{2ij})$$

$$+ \frac{1}{6} \sum_{j,k=1}^{2} n^{1/2}(\hat{\theta}^{(j)} - \tilde{\theta}^{(j)})\, n^{1/2}(\hat{\theta}^{(k)} - \tilde{\theta}^{(k)})\, I_{2ijk}$$

$$i = 1,2$$

where all functions are evaluated at $\tilde{\theta}$. One should note that although in the expressions for R's, derivatives of L_n up to <u>third</u> order are involved, to get a correct estimate of the error consideration of derivatives of L_n up to <u>fifth</u> order is necessary.)

Below all subsequent expansions are performed on the set A_n. We also need Assumption B_s(iii) (among others).

A usual iterative approximation of $n^{1/2}(\hat{\theta}^{(2)} - \tilde{\theta}^{(2)})$, starting with $-n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})\, I_{21}(\tilde{\theta})/I_{22}(\tilde{\theta})$ as the initial approximation, gives

$$n^{1/2}(\hat{\theta}^{(2)} - \tilde{\theta}^{(2)}) = n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})\Big(-\frac{I_{21}(\tilde{\theta})}{I_{22}(\tilde{\theta})} + \frac{P_1}{n^{1/2}} + \frac{P_2}{n}\Big) + o(n^{-1})$$

where P_1 and P_2 are polynomials in $n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})$ and U_n whose coefficients do not depend on n (a similar computation was done in the proof of Theorem 2.1 of Chapter One; see the proof of (4.15), page 39). We only need the fact that each successive approximation of $n^{1/2}(\hat{\theta}^{(2)} - \tilde{\theta}^{(2)})$ has a factor $n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})$ which is quite obvious). Here we have used the fact that on the set A_n, $\|U_n\|^2 = o(\log n)$. Expanding now all the partial derivatives at $\tilde{\theta}$ of L_n appearing in P_1 and P_2 around θ_o, it can be verified that

$$n^{1/2}(\hat{\theta}^{(2)} - \tilde{\theta}^{(2)}) = n^{1/2}(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})\, R_n^*(U_n^*) + o(n^{-1}) \tag{4.7}$$

where $R_n^*(U_n^*)$ is a polynomial in $n^{1/2}(U_n^* - E(U_n^* \mid \theta_o))$ whose coefficients depend on n.

Expanding λ_n around $\tilde{\theta}$ and then expanding the partial derivatives at $\tilde{\theta}$ of L_n around θ_o and finally using equation (4.7), one gets

(4.8) $$\lambda_n = n(\hat{\theta}^{(1)} - \tilde{\theta}^{(1)})^2 \, P_n^*(U_n^*) + o(n^{-1})$$

where $P_n^*(U_n^*)$ is a polynomial in $n^{1/2}(U_n^* - E(U_n^* | \theta_o))$ whose coefficients depend on n.

It may be noted that in the case of an exponential family of distributions, all the derivatives of L_n of order two or more are constants (i.e., nonrandom) and hence (4.8) can be used as an alternative way of checking Assumptions A_s(ii), (iii) and (iv).

Coming to the general (i.e., non-exponential) case, suppose that Assumptions B_s(i) - (v) with $s = 4$ hold and that the random variables

$$\{L_n^{i_1 \ldots i_j} - I_{i_1 \ldots i_j} : i_1, \ldots, i_j = 1 \text{ or } 2, \; 1 \leq j \leq 4\}$$

are linearly independent (i.e., have nonsingular dispersion matrix). Then the fact that $P(A_n ; \theta_o) = 1 + o(n^{-1})$ follows from relations (1.28), (1.29) and (2.32) of Bhattacharya and Ghosh(1978) and the analogous relations for the restricted ML estimators. Also following the proof of Theorem 3 of Bhattacharya and Ghosh we can show that the vector U_n^* have a multivariate Edgeworth expansion, i.e., Theorem 3 of Bhattacharya and Ghosh holds if the vector

$n^{1/2}(\hat{\theta} - \theta_o)$ of this theorem is replaced by the vector U_n^*. By Theorem 5.2, Chapter One, it is now immediate that the distribution function of λ_n admits of an expansion of the type stated in Theorem 5.2.

Suppose now that Assumptions B_s(i) - (v) hold (with s = 4) but that the random variables

$$T_1' = L_1^1 - I_1 , \quad T_2' = L_1^2 - I_2$$

and

$$T = \{L_1^{i_1 \cdots i_j} - I_{i_1 \cdots i_j} : i_1, \ldots, i_j = 1 \text{ or } 2, \quad 2 \leq j \leq 4\}$$

are not linearly independent. Let $y_i = n^{1/2}(L_n^i - I_i)$, i = 1,2 and let $x_1, \ldots, x_m$ stand for the set

$$\{n^{1/2}(L_n^{i_1 \cdots i_j} - I_{i_1 \cdots i_j}) : i_1, \ldots, i_j = 1 \text{ or } 2, \quad 2 \leq j \leq 4\}.$$

(We shall use these notations for this and the next paragraphs only). Then we claim

(a) that it is possible to choose $T_1, \ldots, T_r$, a subset of T, such that T_1', T_2', $T_1, \ldots, T_r$ are linearly independent ; and

(b) that on the set A_n each of $x_{r+1}, \ldots, x_m$ can be expressed up to $o(n^{-1/2})$ as a polynomial (in fact a linear combination) involving $x_1, \ldots, x_r$ (and the constant function 1) with coefficients polynomials in $n^{1/2}(\hat{\theta} - \theta_o)$. Consequently in equation (4.8) the polynomial P_n^* can be replaced by another polynomial in $n^{1/2}(\hat{\theta} - \theta_o)$ and $x_1,\ldots,x_r$; and the argument of the previous paragraph goes through.

To justify the above claims, we begin by writing <u>all</u> the linear restrictions among T_1', T_2' and $T_1,\ldots,T_m$ in the following form:

(4.9) $$c_{i1} T_1 + \cdots + c_{im} T_m + d_{i1} T_1' + d_{i2} T_2' = 0$$

for $i = 1,\ldots,r$ where r is the rank of $(C|D)$, $\underset{r\times m}{C} = ((c_{ij}))$ and $\underset{r\times 2}{D} = ((d_{ij}))$. Observe that r is also the rank of C since T_1', T_2' are linearly independent (i.e., T_1', T_2' have positive-definite dispersion matrix). Without loss of generality let the first r columns of C be linearly independent. <u>Then clearly</u> T_1', T_2' <u>and</u> $T_{r+1},\ldots,T_m$ <u>are linearly independent</u>. Also from (4.9) we get

(4.10) $$\begin{aligned} &c_{i1} x_1 + \cdots + c_{ir} x_r \\ &= -c_{i(r+1)} x_{(r+1)} - \cdots - c_{im} x_m - d_{i1} y_1 - d_{i2} y_2 . \end{aligned}$$

On the set A_n we now expand (up to $o(n^{-1/2})$) y_i around $\hat{\theta}$ and then the partial derivatives at $\hat{\theta}$ of L_n around θ_o, $i = 1,2$; equation (4.10) then implies that

$$(c_{i1} + P_{i1}(\widehat{\theta}))x_1 + \cdots + (c_{ir} + P_{ir}(\widehat{\theta}))\, x_r$$
$$= -(c_{i(r+1)} + P_{i(r+1)}(\widehat{\theta}))x_{r+1} + \cdots + (c_{im} + P_{im}(\widehat{\theta}))x_m$$
$$- d_{i1}\, P'_{i1}(\widehat{\theta}) - d_{i2}\, P'_{i2}(\widehat{\theta}) + o(n^{-1/2})$$

where for each i,j, $P_{ij}(\widehat{\theta})$ is either $(\theta_o^{(1)} - \widehat{\theta}^{(1)})$ or $(\theta_o^{(2)} - \widehat{\theta}^{(2)})$ and

$$P'_{ii'}(\widehat{\theta}) = \sum_{j=1}^{2} n^{1/2}(\theta_o^{(j)} - \widehat{\theta}^{(j)})\, I_{i'j}$$
$$+ n^{-1/2} \sum_{j,j'=1}^{2} n^{1/2}(\theta_o^{(j)} - \widehat{\theta}^{(j)})\, n^{1/2}(\theta_o^{(j')} - \widehat{\theta}^{(j')})\, I_{i'jj'}$$

$i, i' = 1,2$. Hence for all sufficiently large n, the $r \times r$ matrix $((c_{ij} + P_{ij}(\widehat{\theta})))$ is nonsingular and so we can write (up to $o(n^{-1/2})$) $x_1, \ldots, x_r$ as linear combinations of $x_{r+1}, \ldots, x_m$ and the constant function 1 with coefficients polynomials in $n^{1/2}(\widehat{\theta} - \theta_o)$.

This completes the discussion of the special case $p = 1$ and $k = 2$.

Now we shall consider briefly the general case. As before assume that $\theta_o = 0$ and that Assumptions B_s(i) - (v) hold (with $s = 2m + 2$). Apply first a nonsingular linear transformation on the parameter space which leaves the first p components of θ unchanged and which reduces $((I_{ij}))_{p+1 \le i,j \le k}$ to the identity matrix of order $(k-p)$.

Expanding $D^i L_n(\hat{\theta})$ around $\tilde{\theta}$ for $i = p+1,\ldots,k$ and replacing all the partial derivatives at $\tilde{\theta}$ of L_n by the deviations from their respective asymptotic means, one can express the likelihood equations for $\hat{\theta}^2$ as follows :

$$n^{1/2}(\hat{\theta}^{(i_1)} - \tilde{\theta}^{(i_1)}) = \sum_{i_2=1}^{k} n^{1/2}(\hat{\theta}^{(i_2)} - \tilde{\theta}^{(i_2)}) \left\{ \sum_{i_3=1}^{2m} n^{-i_3/2} R_{i_1 i_2 i_3} \right\}$$

$$+ o(n^{-m}), \quad p + 1 \leq i_1 \leq k, \; m \geq 1$$

where $\{R_{i_1 i_2 i_3}\}$ are polynomials in $n^{1/2}(\hat{\theta} - \tilde{\theta})$ <u>and</u> the normalised partial derivatives at $\tilde{\theta}$ of L_n (of order (2m+2) or less) whose coefficients do not depend on n. The above and all subsequent approximations are performed on a set of probability (under θ_o) $1 + o(n^{-m})$.

An iterative approximation of

$$n^{1/2}(\hat{\theta}^{(i_1)} - \tilde{\theta}^{(i_1)}), \quad p+1 \leq i_1 \leq k$$

where at each stage we use the approximation of $n^{1/2}(\hat{\theta}^2 - \tilde{\theta}^2)$ obtained at the previous stage and keep terms of appropriate orders of approximations, gives

$$n^{1/2}(\hat{\theta}^{(i_1)} - \tilde{\theta}^{(i_1)}) = \sum_{i_2=1}^{p} n^{1/2}(\hat{\theta}^{(i_2)} - \tilde{\theta}^{(i_2)}) \left\{ \sum_{i_3=1}^{2m} n^{-i_3/2} P_{i_1 i_2 i_3} \right\}$$

$$+ o(n^{-m})$$

where $\{P_{i_1 i_2 i_3}\}$ have the same properties as those of $\{R_{i_1 i_2 i_3}\}$ except that $\{P_{i_1 i_2 i_3}\}$ do not depend on $n^{1/2}(\hat{\theta}^2 - \tilde{\theta}^2)$. The rest of the computations is similar to that of the case $p = 1$, $k = 2$ and $m = 1$.

REMARK 4.1 Suppose that Assumptions B_s(i) - (vi) with $s = 2m+2$ hold and that the following stronger version of B_s(ii) holds:

$B_s(ii)'$: for each compact $K \subset \text{Ⓗ}_o$ and each α, $1 \leq |\alpha| \leq s$,

$$\sup_{\theta \varepsilon K} E(|D^{\alpha} L_1(Y_1; \theta)|^{s+1}; \theta) < \infty;$$

and for each compact K there exists an $\varepsilon > 0$ such that

$$\sup_{\theta' \varepsilon K} E(\max_{|\theta-\theta'| \leq \varepsilon} |D^{\alpha} L_1(Y_1;\theta)|^{s}; \theta) < \infty \text{ if } |\alpha| = s+1.$$

Then it can be shown that the above asymptotic expansion for λ_n holds uniformly in $\theta_o \varepsilon K$ for any compact $K \subset \text{Ⓗ}_o$.

(C) VALIDITY OF HAYAKAWA's EXPANSION:

Here we consider the related work of Hayakawa (1977). He has obtained an asymptotic expansion up to $o(n^{-1})$ of the distribution function of λ_n by first approximating λ_n by W_n' up to $o_p(n^{-1})$ and then inverting the resulting approximate characteristic function of W_n'. Note that

$$\int \exp(ith_{s-1}(k_{s-1}(z)))\, \xi'_{s-1,n}(z)dz$$

$$= \int_{M_n} \exp(ith_{s-1}(k_{s-1}(z)))\, \xi'_{s-1,n}(z)dz + o(n^{-(s-3)/2})$$

$$= \int_{N_n} \exp(ith_{s-1}(\theta))\, \xi_{s-1,n}(\theta)d\theta + o(n^{-(s-3)/2})$$

$$= \int_{R^1} \exp(itv) \sum_{j=0}^{m} n^{-j}\, q_j(v)\, \chi^2(v;p)dv + o(n^{-(s-3)/2})$$

(we are using the notations of the last part of Section 3, except that here we take W_n to be λ_n and $Z_1 = (L_1^{i_1 \cdots i_j} : i_1, \ldots, i_j \geq 1,\ 1 \leq j \leq 4))$. It follows arguing as in Section 6(B) of Chapter One that the asymptotic expansion given in Theorem 1 of Hayakawa is valid.

We finally remark that Hayakawa gets his expansion by expressing λ_n as a function of $\{L_n^{i_1}, L_n^{i_1 i_2}, L_n^{i_1 i_2 i_3}, L_n^{i_1 i_2 i_3 i_4}\}$ (instead of $\{\hat{\theta}, L_n^{i_1 i_2}, L_n^{i_1 i_2 i_3}, L_n^{i_1 i_2 i_3 i_4}\}$). But as observed earlier Assumption A_s(iv) then need not hold in general even on a set of probability $1 + o(n^{-1})$.

CHAPTER THREE

EXPANSIONS FOR THE LIKELIHOOD RATIO AND OTHER STATISTICS UNDER CONTIGUOUS ALTERNATIVES

SECTION 1. INTRODUCTION AND MAIN RESULT

Asymptotic expansions were obtained in the previous chapters for statistics with a limiting (central) chi-square distribution. In this chapter we shall study the corresponding case of non-central distribution under contiguous alternatives.

Let θ_o be a <u>fixed</u> element of $R^{k'}$ and $\{\theta_n\}_{n\geq 1}$ a <u>fixed</u> sequence in $R^{k'}$. Let $\{Z_n\}_{n\geq 1}$ be a sequence of k-dimensional IID random vectors. Write

$$(1.1) \quad \mu(\theta_n) = E(Z_1;\theta_n), \quad V_n = E((Z_1 - \mu(\theta_n))(Z_1 - \mu(\theta_n))^T;\theta_n)$$

$$n = 0,1,\ldots$$

where T denotes transpose.

REMARK 1.1 We shall assume later that $\{\theta_n\}_{n\geq 1}$ is "contiguous" to θ_o in the sense that the first (s-1) moments of Z_1 under θ_n have expansions (in powers of $n^{-1/2}$) up to $o(n^{-(s-3)/2})$ whose leading terms are the corresponding moments under θ_o. Suppose that $S(\theta_o)$ is a sphere around θ_o such that the distribution of Z_1 under θ <u>is defined for each</u> θ in $S(\theta_o)$. Then under enough regularity conditions the jth moment under θ will be (s-3) times continuously differentiable. In this case if θ_n is contiguous in the commonly used sense, i.e., is of the form $\theta_o + n^{-1/2} t^*$ for each $n \geq 1$ and for some $t^* \varepsilon R^{k'}$, then the Taylor expansion for the jth moment under θ_n (around θ_o) will lead to the kind of expansion required by condition (c) of Theorem. We shall also need a uniform Cramér's condition (vide (1.3)).

Let H be a real valued Borel measurable function on R^k and define W_n by equation (1.4) of Chapter One with μ there replaced by $\mu(\theta_o)$, i.e., let

$$W_n = 2n(H(\bar{Z}_n) - H(\mu(\theta_o))) \; . \tag{1.2}$$

Recall the notations and Assumptions A_s(i) - (iv) introduced in Section 2 of Chapter One. Assume without loss of generality that $\mu(\theta_o)$ is the null vector.

THEOREM 1.1 Let $\{Z_n\}_{n\geq 1}$, $\{\theta_n\}_{n\geq 0}$ and H be as above and define W_n by (1.2) above. Suppose that for some integer $s \geq 4$, the following assumptions hold :

(a) H and Z_1 satisfy Assumptions A_s(i) - (iv) with μ and V there replaced by $\mu(\theta_o)$ and V_o respectively (see pages 15-16) ;

(b) Z_1 satisfies the uniform Cramér's condition

(1.5)
$$\sup_{n \geq 0} \sup_{\|t\|>b} |E(\exp(i < t, Z_1 >) ; \theta_n)| < 1 \qquad \text{for each } b > 0 ;$$

(c) $\sup_{n \geq 0} E(\|Z_1\|^s ; \theta_n)$ is finite and for each α, $1 \leq |\alpha| \leq s-1$, $E(Z_1^{\alpha} ; \theta_n)$ admits of an expansion (in powers of $n^{-1/2}$) up to $o(n^{-(s-3)/2})$.

Then there exist nonnegative integers $k_0, \ldots, k_{s-3}$ and constants $\{P_{i,j}\}$ not depending on n $(0 \leq i \leq k_j, 0 \leq j \leq s-3)$ such that the following expansion is valid, and holds uniformly in $u \in [u_0, \infty)$, $u_0 > 0$:

(1.4)
$$P(W_n \leq u ; \theta_n) = \sum_{j=0}^{s-3} n^{-j/2} \int_0^u \Psi_j (v) dv + o(n^{-(s-3)/2})$$

One can replace u_0 by zero provided $p > 1$. Here

(1.5)
$$\Psi_j (v) = \sum_{i=0}^{k_j} P_{i,j} \chi^2(v ; p+2i, \delta)$$
$$0 \leq j \leq s-3, \quad k_0 = 0, \quad P_{0,0} \equiv 1 ;$$

(1.6)
$$p = \text{rank of Hessian at } \hat{\mu}(\theta_0) \text{ of } H ,$$
$$\delta = < t^1, V_0^{-11} t^1 > ,$$

$\chi^2(v\,;\,p+2i,\,\delta)$ being the density (at v) of a noncentral chi-square with degrees of freedom $p+2i$ and noncentrality parameter $\delta \geq 0$, V_o^{-11} the submatrix consisting of the first p rows and p columns of the inverse of V_o and finally

$$(1.7) \qquad t = \lim_n n^{1/2}(E(Z_1\,;\,\theta_n) - E(Z_1\,;\,\theta_o))$$

(see (2.12), page 21 , Chapter One).

REMARK 1.2 The Theorem in its present form is often unsuitable for statistical applications because of the assumption that W_n is a function of the mean vector $\bar{Z}_n$ based on some sequence of IID random vectors. One can however easily verify that the Theorem remains valid if the normalised deviation $n^{1/2}(\bar{Z}_n - \mu(\theta_o))$ there is replaced by $n^{1/2}(U_n - E(U_n\,;\,\theta_n))$, where $\{U_n\}_{n\geq 1}$ is an arbitrary sequence of random vectors possessing an Edgeworth expansion which is uniform with respect to θ_n .

REMARK 1.3 Suppose that assumptions (b) and (c) (with s in place of (s-1)) of the Theorem hold. Assume $A_s(i)$ and instead of the rest of A_s , assume as in Bhattacharya and Ghosh that the vector

$$\ell = (D^1H, \ldots, D^k H)\,(\mu(\theta_o))$$

is non-null. Let $W_n^1 = n^{1/2}\,(H(\bar{Z}_n) - H(\mu(\theta_o)))$. Then the distribution function of W_n^1 under θ_n has an expansion (in powers of $n^{-1/2}$) valid up to $o(n^{-(s-2)/2})$ with the leading term a normal distribution with (nonzero) mean $<\lambda, t>$ where t is as in (1.7). One may prove this by applying Theorem 2(b) of Bhattacharya and Ghosh with $P(\theta_n)$ in place of P and then expanding the coefficients in the expansion. Also it is easy to check that this expansion agrees with the formal Edgeworth expansion obtained by evaluating the first s moments of W_n^1 formally up to $o(n^{-(s-3)/2})$ by the delta-method.

REMARK 1.4 If Z_1 satisfies Cramér's condition under θ_n i.e. if

$$\sup_{\|t\| > b} |E(\exp(i < t, Z_1 >) ; \theta_n)| < 1 \qquad b \in R_+^1$$

(see Bhattacharya and Ranga Rao (1976), page 207) and if the distribution of Z_1 under θ_n converges in variation norm to that under θ_o, then the uniform Cramér's condition (1.3) holds. Under the set-up described in Remark 1.1, suppose that for each θ in $S(\theta_o)$, the distribution of Z_1 under θ_n admits of a density f_θ such that the map $\theta \to f_\theta$ is continuous in $S(\theta_o)$; then using Scheffé's theorem (see Lemma 2.1, Bhattacharya and Ranga Rao(1976)), it is easy to see that the above sufficient condition for (1.3) holds.

The proof of the Theorem is given in Section 2. In Section 3, we consider applications to the likelihood ratio and other related statistics. In Section 4, expansions under a fixed alternative for these statistics have been obtained (a) when the null hypothesis is simple; and (b) when the null hypothesis is composite and the observations are coming from an exponential family of distributions.

SECTION 2. PROOF OF THEOREM (1.1)

We assume throughout this section that the assumptions of the Theorem hold. We may then assume without loss of generality that

(2.1) $$V_o = I\ ,\quad z^T L z = \|z^1\|^2$$

and that A_s(iv) hold with $A = I$ (see Remark 3.1, Chapter One). Define g_n, $h_{s-1,n}$ and M_n by equations (4.3), (4.4) and (4.5) respectively of Chapter One (see page 35). It is well-known that for any $x \in R^k$ and any $q \geq 0$

(2.2) $$\int_{M_n^c} \|x\|^q \exp(-\tfrac{1}{2}\|z - x\|^2)dz = o(n^{-(s-3)/2})$$

We need the following three auxiliary lemmas.

LEMMA 2.1 Let assumptions (a), (b) and (c) of the Theorem hold. Then there exist $\{\phi_{i,j}\}$, not depending on n ($n \geq 0$, $j = 1,2,\ldots,s-3$) such that the following expansion is valid for $u \in R_+^1$ and is uniform in $u \in [u_o, \infty)$, $u_o>0$ (u_o can be taken to be zero if $p>1$) :

(2.3) $$P(W_n \leq u\ ;\ \theta_n) = \sum_{j=0}^{s-3} n^{-j/2} \int_0^u \phi_j(v)dv + o(n^{-(s-3)/2})$$

where

$$(2.4)\qquad \begin{aligned} \psi_j(v) &= \sum_{i=0}^{\infty} \psi_{i,j}\, \chi^2(v\,;\,p+i), \\ \psi_o(v) &= \chi^2(v\,;\,p) \qquad j \geq 1,\ v \in R_+^1 . \end{aligned}$$

Thus once this lemma is established, the Theorem would follow if one could show that ψ_j given by (2.4) can as well be expressed in the form (1.5). Lemma 2.1 will be used here to show, essentially, the existence of a valid expansion for W_n which is needed for the proof of Lemma 2.3 below.

PROOF OF LEMMA 2.1 : Let t_n be defined by

$$\mu(\theta_n) = \mu(\theta_o) + n^{-1/2}(t + t_n)$$

(see (1.7)). The Edgeworth expansion up to $o(n^{-(s-3)/2})$ of $n^{1/2}(\bar{Z}_n - \mu(\theta_n))$ under θ_m can be written as

$$\xi_{1,s,n,m}(z) = \phi_{V_m}(z)\,(1 + \sum_{j=0}^{s-3} n^{-j/2} \sum_i R_{ij}(z))$$

where $\{R_{ij}\}$ are polynomials in z with coefficients rational functions of $\{E(Z_1^{\alpha}; \theta_m) : 1 \leq |\alpha| \leq s-1\}$ with nonvanishing denominators (at least for all sufficiently large n). Because of the uniform Cramér's condition and the uniform boundedness of $E(Z_1^s\,;\,\theta_m)$, we get setting $\xi_{1,s,n}(z) = \xi_{1,s,n,n}(z)$

$$P(W_n \leq u\,;\,\theta_n) = \int_A \xi_{1,s,n}(z-t-t_n)dz + o(n^{-(s-3)/2})$$

uniformly in $u \varepsilon R^1_+$. Here A_n is the set $\{z : g_n(z) \leq u\}$, and we have used Theorem 1.5 and the first observation following its proof on page 11 of Bhattacharya (1977). For this one needs the following estimate

$$\sup_{0<u<\infty} \int_{A_n} \phi_{V_n}(z)\, dz = o(n^{-(s-3)/2})$$

(see the last remark following the proof of Theorem 2.1(b) of Chapter One).

We now make use of assumption (b) and expand R_{ij}'s and $\phi_{V_n}(z)$, getting uniformly in $u \varepsilon R^1_+$

(2.5) $$P(W_n \leq u ; \theta_n) = \int_{A_n} \xi_{2,s,n}(z) dz + o(n^{-(s-3)/2})$$

where

(2.6) $$\xi_{2,s,n}(z) = \phi(z - t) \sum_{j=0}^{s-3} n^{-j/2} P'_{i,j}(z) ,$$

$P'_{i,j}(z)$ being suitable polynomials (free from n) in z, $P'_{1,0}(z) \equiv$ We now proceed as in the proof of part (a) of Theorem 2.1 of Chapter One. Since

$$\sup_{z \varepsilon M_n} |g_n(z) - h_{s-1,n}(z)| = O(\varepsilon_n), \quad \varepsilon_n = n^{-(s-2)/2}(\log n)^{s/2}$$

and since

$$\sup_{0 < u < \infty} \left| \int_{B_n} z^{\alpha} \phi(z-t) dz \right| = O(\varepsilon_n), \quad B_n = \{z \varepsilon M_n : |h_{s-1,n}(z) - u| \leq \varepsilon_n\}$$

one gets from (2.5) and (2.2) that

$$P(W_n \leq u ; \theta_n) = \int_{A_n^*} \xi_{2,s,n}(z)dz + o(n^{-(s-3)/2})$$

uniformly in $u \in R_+^1$ where $A_n^* = \{z \in M_n : h_{s-1,n}(z) \leq u\}$. One now applies first an orthogonal transformation on $z^1 = (z^{(1)}, \ldots, z^{(p)})$ with the first transformed variable as $< t^1, z^1 > / \|t^1\|^2$, keeping the remaining z's unchanged (provided t^1 is non-null). The rest of the proof is quite similar to that of part (a) of Theorem 2.1 of Chapter One and hence is omitted.

To state the next lemma, let $j_1(\alpha)$ denote the number of odd components of the vector α of nonnegative integers.

LEMMA 2.2(a)

$$\int_{\|z^1\|^2 \leq u} z^{\alpha} \exp(-\tfrac{1}{2} \|z - t\|^2)\, dz$$

(2.7)

$$= \sum_{j=m_1}^{m_2} \alpha_j^1(t) \int_0^u \chi^2 (v ; p+2j, \delta)\, dv$$

where $\{\alpha_j^1(t)\}$ are suitable polynomials in t and

(2.8) $$m_1 = (|\alpha^1| + j_1(\alpha^1))/2, \quad m_2 = |\alpha^1|$$

(b) Let r^* be a nonnegative integer. Then

$$(2.9)\quad \begin{aligned} &(iv)^{r^*} \int_{R^1} \exp(iv\|z^1\|^2) z^{\alpha} \exp(-\tfrac{1}{2}\|z-t\|^2)\,dz \\ &= \sum_{j=m_1-r^*}^{m_2} \alpha_j^2(t)\, \hat{\chi}^2(v\,;\, p+2j,\, \delta) \end{aligned}$$

provided that $|\alpha^1| \geq 2r^*$, where $\{\alpha_j^2(t)\}$ are suitable polynomials in t and m_1, m_2 are as in (a) and $\wedge$ denotes the Fourier-Stieltjes transform.

It follows from (2.5) that Lemma 2.2 establishes the special case of the Theorem when

$$W_n = 2n(\bar{Z}_n - \mu(\theta_o))^T L(\bar{Z}_n - \mu(\theta_o))$$

PROOF OF LEMMA 2.2 : Without loss of generality we may assume that $p = k$. We need the following fact :

Fact (A) : Let, for some nonnegative integer r,

$$(2.10)\qquad g(x\,;\, b, r) = x^r \exp(-\tfrac{1}{2}(x-b)^2) \qquad x \in R^1,\ b \in R^1 .$$

Then there exist (numerical) constants $a_o, a_1, \ldots, a_m$ such that if $x \geq 0$,

$$(2.11)\quad \begin{aligned} &(g(x^{1/2}\,; b, r) + g(-x^{1/2}\,;\, b, r)) \,/\, 2x^{1/2} \\ &= \sum_{j=0}^{m} a_j\, b^{2j+j_1}\, \chi^2(x\,;\, q+2j,\, b^2) \end{aligned}$$

where $j_1 = 0$ or 1 according as r is even or odd and

$$(2.12)\qquad q = r+1+j_1 \,, \quad m = (r-j_1)/2 .$$

Fact (A) can be established as follows : the left side of (2.11) is

$$\exp\{-\tfrac{1}{2}(x+b^2)\}\sum_{i=0}^{\infty}\frac{b^{2i+j_1}}{(2i+j_1)!}\cdot x^{i+\frac{q}{2}-1},$$

while the coefficient of $x^{i+\frac{q}{2}-1}$ in the right side of (2.11) is

$$\frac{a_o+a_1(2i)+a_2(2i)(2(i-1))+\ldots.}{2^{2i+q/2}\,\Gamma(i+1)\,\Gamma(i+q/2)}\cdot b^{2i+j_1}.$$

Here $\Gamma(\cdot)$ stands for the gamma function. One therefore verifies, using the duplication formula for the gamma function, that

$$\frac{2^{2i+q/2}\,\Gamma(i+1)\,\Gamma(i+q/2)}{(2i+j_1)!}$$

is a polynomial (in i) of degree $(r-j_1)/2$. This completes the proof of Fact (A).

Observe that the integral on the left side of (2.7) can be written as

$$\int_A\prod_{i=1}^{k}\{(g_i((z^{(i)})^{1/2})+g_i(-(z^{(i)})^{1/2}))/2(z^{(i)})^{1/2}\}\,dz$$

where $A=\{z: z^{(1)}+\ldots+z^{(k)}\le u,\ z^{(1)}>0,\ldots,z^{(k)}>0\}$

and $g_i(z^{(i)})=g(z^{(i)};t^{(i)},\alpha^{(i)})$, $i\ge 1$. Fact (A), the fact that the family of noncentral chi-squares is closed under convolution and relation (2.12) complete the proof of Lemma 2.2(a).

Part (b) with $r^* = 0$ is equivalent to Part (a). The case of the general r^* follows from Part (a) and the following elementary fact :

Fact (B) : If $q \geq 2r^* + 1$ and $v \in R^1$,

$$(iv)^{r^*} \hat{\chi}^2 (v ; q, \delta) = 2^{-r^*} \sum_{j=0}^{r^*} \binom{r^*}{j} (-1)^j \hat{\chi}^2 (v; q-2j, \delta)$$

LEMMA 2.3 Suppose that the assumptions of the Theorem hold and that for each real v

$$E(\exp(iv\, W_n) ; \theta_n) = C_n(v) + o(n^{-(s-3)/2})$$

where $C_n(v)$ is the Fourier-Stieltjes transform of

$$(2.13) \qquad \sum_{j=0}^{s-3} n^{-j/2} g_j(v) \qquad v \in R^1$$

(with g_j free from n for each $j \geq 1$). Then (2.13) is the valid expansion for W_n under θ_n up to $o(n^{-(s-3)/2})$.

The proof of this lemma is omitted since it is based on arguments similar to those used in Section 6(B) of Chapter One (see also the first part of the proof of Lemma 2.1).

PROOF OF THEOREM 1.1 : We make here the convention that $P(z)$ (with or without suffixes) will stand for a polynomial in z with coefficients free from n. Proceeding as in the first part of Lemma 2.1, and using (2.1), definition of $h_{s-1,n}(z)$

and estimate (2.2), one can show that

$$E(\exp(iv\, W_n)\,;\, \Theta_n)$$

(2.14)
$$= \int_{M_n} \exp(iv\, h_{s-1,n}(z))\, \xi_{2,s,n}(z)\, dz + o(n^{-(s-3)/2})$$
$$= \int_{R^k} \exp(iv\,\|z^1\|^2)\, \psi_n^1(z)\,dz + o(n^{-(s-3)/2})$$

where

$$\psi_n^1(z) = \Big(1 + \sum_{j_1=1}^{s-3} n^{-j_1/2} \sum_{j_2=1}^{j_1} (iv)^{j_2}\, P^1_{j_1,j_2}(z)\Big)\, \xi_{2,s,n}(z)$$

($\xi_{2,s,n}$ is defined in (2.6)). We can rewrite $\psi_n^1(z)$ as follows:

(2.15)
$$\psi_n^1(z) = \sum_{j_1=0}^{s-3} n^{-j_1/2} \sum_{j_2=0}^{j_1} (iv)^{j_2}\, P^2_{j_1,j_2}(z)\, \phi(z-t)$$

($P^2_{0,0}(z) \equiv 1$). In view of Assumption A_s(iv), note that if z^α is any term of some $P^1_{j_1,j_2}(z)$, then $|\alpha^1| \geq 2j_2$. Clearly this property is inherited by [redacted] the polynomials $\{P^2_{j_1,j_2}(z)\}$. Thus <u>if we let</u> $C_n(v)$ <u>denote the integral on the right side of</u> (2.14), then Lemma 2.2(b) and (2.15) imply that

$$C_n(v) = \sum_{j_1=0}^{s-3} n^{-j_1/2} \sum_{j_2=0}^{k_{j_1}} \beta_{j_1,j_2}\, \hat{\chi}^2(v\,;\, p+2j_2,\, \delta)$$

for suitable nonnegative integers $\{k_{j_1}\}$ and suitable constants

$\{\beta_{j_1, j_2}\}$; $k_o = 0$, $\beta_{o,o} = 1$. In other words,

$$(2.16) \qquad C_n(v) = \int_{R^1} \exp(ivx) \sum_{j=0}^{s-3} n^{-j/2} \psi_j(x)dx$$

where $\{\psi_j\}$ are of the form (1.5). The definition of $C_n(v)$, relations (2.14) and (2.16) and Lemma 2.3 together complete the proof of the Theorem.

SECTION 3. APPLICATIONS

THEOREM 1.1 (suitably modified as indicated in Remark 1.2) can be used to obtain asymptotic expansions of the distribution functions of the likelihood ratio statistic, Wald's and Rao's statistics (see equations (1.3), (1.6) and (1.7) of Chapter Two) under contiguous alternatives (see Remark 1.1), provided that Assumptions B_s(i) - (vi) of Section 4(A) of Chapter Two hold and

$$\sup_{n \geq 0} E(\|Z_1\|^s ; \theta_n) < \infty$$

and $\{E(Z_1^\alpha ; \theta_n)\}$, $1 \leq |\alpha| \leq s-1$, admit [redacted] asymptotic expansions in powers of $n^{-1/2}$. Here Z_1 is the vector whose components are indexed by α , $1 \leq |\alpha| \leq s$, and the αth component is $D^\alpha L_1(Y_1 ; \theta_o)$; in particular, the dimension of Z_1 is

$$\sum_{r=1}^{s} \binom{k+r-1}{r} .$$

It should be noted that the above assumptions are satisfied by the family of exponential distributions with θ as the natural parameter, provided the assumptions made in Section 3 of Chapter Two hold.

To establish the desired expansions, one constructs a set A_n such that

(i) $P(A_n^c \,;\, \theta_n) = o(n^{-(s-3)/2})$;

and

(ii) on A_n, the statistic under consideration can be sufficiently well approximated by a W_n which is of the form (1.2) and which satisfies condition (a) of Theorem 1.1. The set A_n used in Section 4(B) of Chapter Two does this job ; in fact the only new thing to be proved is $P(A_n^c \,;\, \theta_n) = o(n^{-(s-3)/2})$ which follows easily (under θ_n, the normalised $\bar{Z}_n$ has an Edgeworth expansion). Note that the possibility of the uniform Edgeworth expansion for maximum likelihood estimators (suitably normalised) is guaranteed by Theorem 3(a) of Bhattacharya and Ghosh.

Hayakawa (1977) obtained an expansion, up to $o(n^{-1/2})$, for the likelihood ratio statistic under contiguous alternatives by a formal inversion of characteristic function. His formal expansion can be justified by suitably modifying Lemma 2.3; for details, see Section 4(C) of Chapter Two.

SECTION 4. EXPANSIONS UNDER A FIXED ALTERNATIVE

Suppose that $\{Z_n\}_{n\geq 1}$ is a sequence of IID random vectors with common distribution either P_{θ_o} or P_{θ_1}. Let $E(Z_1 ; \theta_i) = \mu(\theta_i)$ i = 0,1. Define W_n by (1.2). We want to find expansions for W_n under θ_1. To this end, assume that Assumptions A_s(i) and (ii) of Chapter One hold with

$$\lambda = (D^1 H,\ldots,D^k H)\,(\mu(\theta_o))$$

and that

$$\lambda_1 = (D^1 H,\ldots,D^k H)\,(\mu(\theta_1)) \neq 0 .$$

Then the distribution function under θ_1 of

$$n^{-1/2}\{W_n - 2n(H(\mu(\theta_1)) - H(\mu(\theta_o)))\}$$

possesses an asymptotic expansion (in powers of $n^{-1/2}$) with the leading term a normal distribution with zero mean and variance $\lambda_1^T V \lambda_1$ where

$$V = E((Z_1 - \mu(\theta_1))\,(Z_1 - \mu(\theta_1))^T ; \theta_1)$$

(V is assumed to be nonsingular). The result follows from Theorem 2(b) of Bhattacharya and Ghosh (1978). It is evident that a similar result holds under the assumptions of the last paper(i.e., under the assumption that λ and λ_1 are both nonzero).

Consider now the problem of testing a simple null hypothesis. One can apply the above result to get asymptotic expansions for the likelihood ratio and other statistics under a fixed alternative. In the last section of his paper (1977), Hayakawa has obtained formally such a result for the case of the likelihood ratio statistic. It can be shown that this formal expansion is in fact a valid one.

We assumed above that the null hypothesis is simple. Similar expansions are possible for the case of a composite null hypothesis provided the observations are coming from an exponential family of distributions with natural parameter space and provided that the ML estimators under H_o exist ; one has to express the maximum likelihood estimators under the null hypothesis in terms of the sample mean. Since Wald's statistic depends only on the unrestricted ML estimators, the asymptotic expansion for the statistic can be obtained even if the null hypothesis is composite and the parent population is not exponential.

SECTION 5. COMPARISON OF THE LR, WALD'S AND RAO'S STATISTICS

We briefly indicate how our results on asymptotic expansions can be used to study deficiency of tests which have equal Pitman efficiency. Consider the problem of testing a simple hypothesis

$H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$ where θ takes values in R^p and thre testes based on the statistics proposed by Neyman and Pearson, Wald and Rao (see the last part of Section 1). For the sake of convenience, we denote them by $\lambda_{n,1}$, $\lambda_{n,2}$, $\lambda_{n,3}$ respectively. These tests have the same Pitman efficiency and so Rao (1965) raised the question (see the last paragraph of Section 6e.2) of higher order discrimination between these statistics and conjectured that the statistic proposed by him is likely to be locally more powerful than the others (in the second edition, this conjecture has been omitted).

Suppose now that the assumptions of Section 3 hold. Then our results in Section 4 of Chapter Two show that if $\chi^2_{p,\alpha}$ is the 100α per cent point of the χ^2-distribution with p degrees of freedom, then

$$(5.1) \qquad P_{H_0}\{\lambda_{n,i} > \chi^2_{p,\alpha}\} = \alpha + o(n^{-1/2}), \quad i = 1,2,3 .$$

For contiguous alternatives of the form $\delta n^{-1/2}$, Peers (1971) has expanded formally $P_\theta(\lambda_{n,i} > \chi^2_{p,\alpha})$ up to $o(n^{-1/2})$ and noted that these expansions do <u>not</u> support Rao's conjecture. In Section 3 we have established the validity of these expansions. This fact and (5.1) imply (see Section 5 of Hodges and Lehmann (1970)) that the deficiency (as computed by Peers) of one of these tests with respect to the others is $\pm\infty$; both values being one gets from (2.5) and (2.2) that

$$P(W_n \leq u \; ; \; \theta_n) = \int_{A_n^*} \xi_{2,s,n}(z)dz + o(n^{-(s-3)/2})$$

uniformly in $u \in R_+^1$ where $A_n^* = \{z \in M_n : h_{s-1,n}(z) \leq u\}$. One now applies first an orthogonal transformation on $z^1 = (z^{(1)}, \ldots, z^{(p)})$ with the first transformed variable as $< t^1, z^1 > / \|t^1\|^2$, keeping the remaining z's unchanged (provided t^1 is non-null). The rest of the proof is quite similar to that of part (a) of Theorem 2.1 of Chapter One and hence is omitted.

To state the next lemma, let $j_1(\alpha)$ denote the number of odd components of the vector α of nonnegative integers.

LEMMA 2.2(a)

$$\int_{\|z^1\|^2 \leq u} z^{\alpha} \exp(-\tfrac{1}{2}\|z - t\|^2)\, dz$$

(2.7)

$$= \sum_{j=m_1}^{m_2} \alpha_j^1(t) \int_0^u \chi^2(v \; ; \; p+2j, \delta)\, dv$$

where $\{\alpha_j^1(t)\}_j$ are suitable polynomials in t and

(2.8) $$m_1 = (|\alpha^1| + j_1(\alpha^1))/2, \quad m_2 = |\alpha^1|$$

(b) Let r^* be a nonnegative integer. Then

Now observe that the latter region can be written in the form

(5.4) $$\{n^{1/2}\,\hat{\theta} > t_{n,1,1} \text{ or } < t_{n,1,2}\}$$

where $t_{n,1,1}$ and $t_{n,1,2}$ are determined (up to $o(n^{-1/2})$) such that the region (5.4) has the same power as that of the region (5.3). A similar remark holds for Rao's test. Note that Wald's test is equivalent to taking $\{n^{1/2}\hat{\theta} > \chi_\alpha$ or $< -\chi_\alpha\}$. Thus each of these statistics have the same power function under contiguous alternatives as a two-sided test based on $\hat{\theta}$, at least up to $o(n^{-1/2})$. In the special case of the exponential family, this equivalence holds up to $o(n^{-j/2})$ for any $j \geq 1$ and hence the relative deficiencies of these tests are zero.

To really compare the statistics, say $\lambda_{n,1}$ and $\lambda_{n,2}$, the following procedure seems more reasonable. First choose $\lambda_{n,1,0}$ such that

$$P_{H_0}(\lambda_{n,1} > \lambda_{n,1,0}) = \alpha + o(n^{-1}) .$$

One can find (vide Section 4 of Chapter Two) statistics T_1 and T_2 such that

$$\lambda_{n,1} = \{n^{1/2}\,\hat{\theta}\,(1 + n^{-1/2}\,T_1 + n^{-1}T_2)\}^2 + o_p(n^{-1}) .$$

Let

$$\alpha_1 = P_{H_o} \{ n^{1/2} \hat{\theta} (1 + n^{-1/2} T_1 + n^{-1} T_2) > (\lambda_{n,1,0})^{1/2} \}$$

$$\alpha_2 = P_{H_o} \{ n^{1/2} \hat{\theta} (1 + n^{-1/2} T_1 + n^{-1} T_2) < - (\lambda_{n,1,0})^{1/2} \}$$

Now choose $t'_{n,1}$ and $t'_{n,2}$ such that

$$P_{H_o} \{ n^{1/2} \hat{\theta} > t'_{n,1} \} = \alpha_1 + o(n^{-1}) ;$$

and

$$P_{H_o} \{ n^{1/2} \hat{\theta} < t'_{n,2} \} = \alpha_2 + o(n^{-1}) .$$

It can be shown as in the previous paragraph that

$$P_{\delta n^{-1/2}} \{ n^{1/2} \hat{\theta} > t'_{n,1} \text{ or } < t'_{n,2} \}$$

$$= P_{\delta n^{-1/2}} \{ \lambda_{n,1} > \lambda_{n,1,0} \} + o(n^{-1}) ,$$

which leads to a <u>finite</u> deficiency.

A similar development is possible for the multiparameter case but will not be presented here.

CHAPTER FOUR

COMPARISON OF TESTS WITH SAME BAHADUR-EFFICIENCY

SECTION 1. INTRODUCTION

In their paper Hodges and Lehmann (1970) studied the problem of discrimination between two statistical procedures which are, according to some criterion, equally efficient; deficiency is essentially a quantitative measure of this discrimination. In the same spirit, we have discussed here the problem of discrimination between two test procedures which have equal Bahadur-efficiency.

It is suggested by Bahadur (1967 and 1971) that in many cases alternative test procedures might be compared on the basis of the associated limiting "attained levels".

On the other hand, Cochran (1952) measured the efficiency of a test procedure by the rate of convergence to zero of its size, when the power is held fixed against a specified alternative. It is well-known that Cochran's approach to efficiency usually leads precisely to the same conclusions as Bahadur's approach does. Motivated by this fact we have introduced in Section 2 the notion of Cochran-deficiency (to be referred to as BCD for reasons explained in the next paragraph) and have shown by means of an example that at the level of deficiency, the above equivalence between Cochran's and Bahadur's viewpoints is no longer true. A necessary and sufficient condition for the existence of Cochran-deficiency is proved. In most cases this condition does not hold and so Cochran-deficiency will rarely exist. When appropriate asymptotic expansions of the significance levels are available, an "approximate" Cochran-deficiency is calculated as a compensation. Conditions under which the said expansions are valid are also investigated.

If one defines Bahadur-deficiency by comparing attained levels, then this quantity will in general be random (see Example 4.1) and very difficult to compute. In view of this one may like to go to the considerations of taking some sort of averages of Bahadur-deficiency. But since this did not turn out to be feasible, we proceed along a somewhat different route in Section 4. Using limits in probability in the definition of Bahadur slopes (rather than almost sure limits) we reinterpret Cochran-deficiency more in line with Bahadur's approach. In view of this interpretation we shall refer to Cochran-deficiency as Bahadur-Cochran-deficiency (BCD).

The following remark on the computation of (approximate) BCD (or equivalently the computation of the expansion of the size of a test procedure when its power is held fixed at β, $0 < \beta < 1$) is worth-mentioning. It consists of two distinct steps. The first step is to determine (with appropriate accuracy) "the cut-off point" of the test statistic so that the power of the test is β. This is easily done once the asymptotic distribution (up to $o(n^{-1/2})$) at a fixed alternative of the test statistic is known; in case the limiting distribution is normal, the problem of establishing the existence of such an expansion can be settled under general conditions using one of the results of Bhattacharya and Ghosh (1978); see in this connection, Section 4 of Chapter Three. The second step is typically to solve a large deviation problem; to be more specific, it is to find, essentially, a uniform asymptotic expansion (up to $o(1)$) of the logarithm of the large deviation probability; the expansion has to be uniform (in a suitable sense) since "the cut-off point" will depend on the sample size. It is this second step which is interesting and quite challenging when the parameter space is multi-dimensional. In this chapter, we shall concentrate on the "one-parameter" testing problems; here the above-mentioned problem is easy to settle using the techniques of Bahadur and Ranga Rao (1960).

SECTION 2. NOTATIONS AND PRELIMINARIES

Let $\{P(\cdot;\theta) : \theta \in \bar{H}\}$ be a family of probability distributions on some space X. Let $s = (x_1, x_2, \ldots)$ be an infinite sequence of independent observations on x. Let $T(s) \equiv \{T_n(s)\}_{n\geq 1}$ be a real valued statistic. In the next paragraph a brief synopsis of Cochran's efficiency is given; for details, consult Cochran (1952) and Bahadur (1967 and 1971).

Let $\bar{H}_o$ be a proper subset of $\bar{H}$. We are interested in testing $H_o : \theta \in \bar{H}_o$ against $H_1 : \theta \in \bar{H} - \bar{H}_o$. For this purpose, we consider a test procedure which is based on a test statistic T and which regards large values of $T_n(s)$ to be significant; i.e., the critical region W_n of the test procedure is of the form

(2.1) $$W_n = \{s : T_n(s) \geq k(n)\}.$$

Fix θ in $\bar{H} - \bar{H}_o$ and β such that $0 < \beta < 1$. Choose $\{k(n)\}_{n\geq 1}$ such that

(2.2) $$P(W_n ; \theta) = \beta \qquad \text{as} \quad n \to \infty.$$

Note that $k(n)$ will depend on β as well as on θ. Let $\alpha(n;\beta) \equiv \alpha(n;\beta,\theta)$ be the resulting size of the test :

$$\alpha(n;\beta,\theta) = \sup \{P(W_n \ni \theta_o) : \theta_o \in \Theta_o\} . \tag{2.3}$$

In typical cases the size tends to zero as n tends to infinity. Cochran argued that the rate at which $\alpha(n;\beta,\theta)$ converges to zero is an indication of asymptotic efficiency of T against θ. Equivalently, one may proceed in the following way which is more suitable for our purpose: for each δ, $0 < \delta < 1$, let $M(\delta) = M(\delta;\beta,\theta)$ be the least integer $m \geq 1$ such that $\alpha(n;\beta,\theta) < \delta$ for all $n \geq m$; and let $M(\delta) = \infty$ if no such m exists. Henceforth we shall assume that $\alpha(n;\beta) \to 0$ as $n \to \infty$, which ensures that $M(\delta)$ is finite for all δ and that $\alpha(n;\beta)$ is strictly positive for all sufficiently large n, which ensures that $M(\delta) \to \infty$ as $\delta \to 0$. The Cochran-efficiency of the test procedure, when it exists, is equal to the limit of $(M(\delta))^{-1} \log(1/\delta)$ as $\delta \to 0$.

Consider now two testing procedures based on the statistics

$$T_1(s) = \{T_{1,n}(s)\}_{n\geq 1} \text{ and } T_2(s) = \{T_{2,n}(s)\}_{n\geq 1} \text{ for}$$

the above testing problem. We want to discriminate between these two procedures when θ obtains. Let $M_i(\delta)$ be defined as above with T replaced by T_i, $i = 1,2$. Clearly the limit of $M_2(\delta)/M_1(\delta)$ as $\delta \to 0$ gives the Cochran-efficiency of T_1 relative to T_2 when θ obtains. When this efficiency is 1,

$$(M_1(\delta) - M_2(\delta))/M_1(\delta) \to 0 \quad \text{as} \quad \delta \to 0 .$$

In typical cases, however, $(M_1(\delta) - M_2(\delta))$ remains bounded as $\delta \to 0$, and so for the purpose of a more subtle distinction, one may use the limit of $(M_1(\delta) - M_2(\delta))$ as $\delta \to 0$ whenever this limit exists.

DEFINITION 2.1 The lower (upper) Bahadur-Cochran-deficiency (BCD) at θ of the first testing procedure with respect to the second is

$$\underline{D}_C(\beta, \theta) = \liminf_{\delta \to 0} (M_1(\delta) - M_2(\delta))$$

$$(\overline{D}_C(\beta, \theta) = \limsup_{\delta \to 0} (M_1(\delta) - M_2(\delta))) .$$

In case these two deficiencies are equal, we say that the BCD at θ exists and is equal to the common value.

Of course, $\underline{D}_C = \overline{D}_C = \infty$ or $-\infty$ if the limit of $M_2(\delta)/M_1(\delta)$ exists and is different from 1. The main use of deficiency is to discriminate tests for which this limit is 1. Note that although the relative Cochran-efficiency of two test procedures is usually free from β (see Proposition 11, Bahadur (1967)), their relative BCD need not be so.

SECTION 3. COCHRAN'S APPROACH

In this section, we shall discuss our problem from the standpoint of Cochran's theory of efficiency. It is proved that for the existence of a finite BCD, the size functions of the test-procedures must be related in a very special way.

(A) EXISTENCE OF A FINITE BCD.

THEOREM 3.1 Suppose that for each $i = 1,2$, $\alpha_i(n)$ is a decreasing function of n for all sufficiently large n. Then the following two conditions are equivalent :

(a) $\lim_{\delta \to 0} (M_1(\delta) - M_2(\delta))$ exists and is equal to an integer

$d \equiv d(\beta, \theta)$

(b) there exists an integer $d \equiv d(\beta, \theta)$ such that

$$\alpha_2(n) = \alpha_1(n+d) \quad \text{for all sufficiently large } n .$$

PROOF : Let $\{\alpha_i(n)\}_{n \geq 1}$ be decreasing function of n if $n \geq m$ and let (a) hold. Then there exists a $\delta_1 > 0$ such that $M_1(\delta) = d + M_2(\delta)$ if $0 < \delta < \delta_1$. Now assume that (b) does not hold ; i.e., that $\alpha_2(n_i) \neq \alpha_1(n_i + d)$ where $n_1 < n_2 < n_3 < \dots$. Choose and fix n_i such that

$$n_i > m, \quad \alpha_1(n_i + d) < \delta_1, \quad \alpha_2(n_i + d) < \delta_1 .$$

We may then assume that $\alpha_2(n_i) < \alpha_1(n_i + d)$. Then if $\delta = \alpha_1(n_i + d)$, $M_1(\delta) > n_i + d$ while $M_2(\delta) \leq n_i$; plainly the last fact contradicts (a).

REMARK 3.1 It should be noted that the main reason why the existence of a finite BCD imposes a strong condition like (b) on the sizes is the discrete nature of the quantities $M_1(\delta)$ and $M_2(\delta)$. Unfortunately, any attempt to make sizes continuous by taking resort to mixtures, as done by Hodges and Lehmann (1970), does not work here.

The above theorem is a special case of the result below. Let $\{\alpha(n)\}_{n\geq 1}$ be a decreasing sequence of real numbers in $[0,1]$. For each δ, $0 < \delta \leq 1$, let $M(\delta)$ be the smallest integer $m \geq 1$ for which $\alpha(m) < \delta$; $M(\delta) = \infty$ if there is no such m.

THEOREM 3.2 Let $\alpha(n) \to 0$ as $n \to \infty$. Then

(a) the function $\delta \to M(\delta)$ from $(0,1]$ to I_+ is a left continuous, decreasing step function (I_+ is the set of natural numbers);

(b) δ is a point of discontinuity of $M(\delta)$ if and only if

$$\alpha(M(\delta) - 1) > \delta \quad \text{or} \quad M(\delta) = 1;$$

(c) the function $\delta \to M(\delta)$ determines the sequence $\{\alpha(n)\}_{n\geq 1}$ uniquely. More precisely, let $S = \{M(\delta): M(\delta)>1, 0<\delta\leq 1\}$.

and let the elements of S, arranged in ascending order be $1 < m_1 < m_2 < \cdots$; let δ_i be such that $M(\delta_i) = m_i$ and δ_i is a point of discontinuity of $M(\delta)$. Then

$$\alpha(n) = \begin{cases} \delta_1 & \text{if } n < m_1 ; \\ \delta_i & \text{if } m_{i-1} \leq n < m_i ; \\ 0 & \text{if } S \text{ has a maximum element} \\ & m_k \text{ and } n \geq m_k . \end{cases}$$

The proof is simple and omitted.

(B) BOUNDS FOR THE UPPER AND LOWER BCD IN TERMS OF APPROXIMATE BCD :

As we know from part (A) that BCD will exist rarely, we now turn to the problem of finding bounds for the upper and lower BCD's. <u>We assume that</u>

Assumption I. Each of $\{\alpha_1(n)\}$ and $\{\alpha_2(n)\}$ is a decreasing function of n for all sufficiently large n.

Assumption II. For each $i = 1,2$, there exists a function $\{\alpha_i(x)\}$ defined for all $x \geq 1$ such that

(a) $\alpha_i(x) = \alpha_i(n)$ if x is the integer n ;

(b) $\alpha_i(x)$ is a decreasing and continuous function of x for all sufficiently large x ;

(c) there exists $d \equiv d(\beta, \theta)$ $(-\infty \leq d \leq +\infty)$ with the property that whenever the sequence $\{m(n)\}_{n\geq 1}$ satisfies the equation $\alpha_2(n) = \alpha_1(m(n))$, the limit of $(m(n) - n)$ (as $n \to \infty$) exists and equals d.

DEFINITION 3.1 The <u>approximate</u> BCD of the first testing procedure with respect to second is $d(\beta, \theta)$.

Note that the approximate BCD need not be an integer and that it depends on the particular extensions $\{\alpha_i(x)\}$ $i = 1,2$ we are using.

Define for each δ, $0 < \delta < 1$, two real numbers $M_{11}(\delta)$ and $M_{12}(\delta)$ by

$$\alpha_1(M_{11}(\delta)) = \alpha_2(M_2(\delta)-1), \quad \alpha_1(M_{12}(\delta)) = \alpha_2(M_2(\delta)).$$

Then $M_{11}(\delta) \leq M_1(\delta) \leq M_{12}(\delta) + 1$. In view of part (c) of Assumption II, one gets

$$(3.1) \qquad -[-d(\beta,\theta)] - 1 \leq \underline{D}_C(\beta,\theta) \leq \overline{D}_C(\beta,\theta) \leq [d(\beta,\theta)] + 1 ,$$

where $[t]$ stands for the greatest integer less than or equal to t.

REMARK 3.2 It follows that the BCD is $+\infty$ or $-\infty$ according as the approximate BCD is $+\infty$ or $-\infty$.

REMARK 3.3 In view of inequality (3.1), the approximate BCD can be taken as a "reasonably good" measure of deficiency; for even when BCD does not exist, the upper and lower BCD can differ by at most 2 and in typical cases (see Remark 3.4 and (E))this difference will be exactly 1. For this reason, we shall henceforth concentrate on the approximate BCD.

REMARK 3.4 If BCD exists and is finite, then it must be equal to d and so d will be an integer; d may however be an integer even if BCD does not exist (see Examples 3.1 and 3.2). If d is non-integral and finite, say $d = m+t$ ($0 < t < 1$, m is an integer), then (3.1) implies that the upper and lower BCD are respectively $m+1$ and m. On the other hand, if d is an integer, then the upper and lower BCD are respectively, d and $d-1$, or $d+1$ and d, or $d+1$ and $d-1$, according as $m(n) - n < d$ for all sufficiently large n or $m(n)-n > d$ for all sufficiently large n, or $m(n)-n > d$ and $< d$ for infinitely many values of n (for details, see Chandra and Ghosh (1978)).

(C) DETERMINATION OF APPROXIMATE BCD :

We assume throughout this section that the significance levels $\{\alpha_i(n)\}$, $i = 1,2$ of the two test procedures admit of the following asymptotic expansion :

$$\alpha_i(n;\beta,\theta) = \exp\{-n\, a_i(\beta,\theta) + n^{1/2} b_i(\beta,\theta) \tag{3.2}$$
$$+ c_i(\beta,\theta) \log n + d_i(\beta,\theta) + o_i(1)\}$$
$$a_i(\beta,\theta) > 0, \quad i = 1,2 .$$

In typical cases, $a_i(\beta,\theta)$ will be free from β ; this will be the case if Bahadur-slopes of T_1 and T_2 exist ; for a precise result, see Theorem 2 of Raghavachari (1970). Note that $M_1(\delta)/M_2(\delta) \to 1$ if and only if $a_1(\beta,\theta) = a_2(\beta,\theta)$. Henceforth we assume that the last equality holds and denote the common value of $a_1(\beta,\theta)$ and $a_2(\beta,\theta)$ by $a(\beta,\theta)$. Also in typical cases $b_i(\beta,\theta)$ will be $-z_\beta\, \sigma^*(\theta)$ times the derivative of $a(\beta,\theta)$ with respect to θ where z_β is the upper 100β per cent point of the limiting distribution of (suitably normalised) T_i and $\sigma^*(\theta)$ is the (positive) norming constant for T_i ; see equation (3.4) of Condition A in part (D). Finally, in typical cases, $c_i(\beta,\theta)$ will be free from β and θ and will depend only on the dimension of the basic model. Thus if the alternative θ is such that the two test procedures have the same Cochran efficiency, then the expansions of their sizes will usually agree up to terms of order $o(1)$.

For convenience, we shall suppress the dependence on β and θ of the quantities a, b_i, etc. The following lemma connects the two sets of assumptions made in the present and previous sections.

LEMMA 3.1 Assume that the size functions $\{\alpha_1(n)\}$ and $\{\alpha_2(n)\}$ satisfy (3.2). Then Assumptions I and II of part (B) are valid. In fact, part (a) of Assumption II holds in the following strong sense; there exist extensions $\{\alpha_i(x)\}$ $i = 1,2$ which satisfy (3.2) for non-integral values of $x(\geq 1)$ as well.

PROOF : Assume (3.2) and define $\alpha_i(x)$ by

$$(3.3) \qquad \alpha_i(x) = \exp\{\lambda \log \alpha_i(n) + \mu \log \alpha_i(n+1)\} \qquad i = 1,2$$

where $n \leq x < n+1$, $x = n\lambda + (n+1)\mu$ for some λ and μ such that $0 < \lambda \leq 1$ and $\lambda + \mu = 1$. It is plain that the lemma holds (for part (c) of Assumption II, see Theorem 3.3 below).

Henceforth we shall work with those extensions $\{\alpha_i(n)\}$ which satisfy (3.2) for all real $x(\geq 1)$. The next theorem gives the possible values of approximate BCD.

THEOREM 3.3 Let the size functions $\{\alpha_i(n)\}$ $i = 1,2$ satisfy (3.2). Then one has

(a) if $b_1 = b_2$ and $c_1 = c_2$, then $d = (d_1 - d_2)/a$;

(b) if $b_1 \neq b_2$, then d is $+\infty$ or $-\infty$ according as $b_1 > b_2$ or $b_1 < b_2$;

(c) if $b_1 = b_2$ and $c_1 \neq c_2$, then d is $+\infty$ or $-\infty$ according as $c_1 > c_2$ or $c_1 < c_2$.

The proof is quite easy and so is omitted.

REMARK 3.5 Although the approximate BCD will in general depend on the particular extensions of the sizes, it is clear from the above theorem that this dependence is slight and the value of d will not depend on the extensions so long as they satisfy (3.2) for all x (≥ 1).

REMARK 3.6. It is evident that a version of Theorem 3.1 can be proved when the asymptotic expansions of the sizes are not necessarily of the special form (3.2). We shall not discuss this point any more because we do not have, at present, any important testing procedure which does not satisfy (3.2).

(D) ON THE VALIDITY OF (3.2)

Here we shall find conditions under which the asymptotic expansion of the form (3.2) of the significance level of a test procedure is valid. We motivate ourselves by considering a test procedure in which the critical region consists of large values of the sum of some sequence of IID random variables on R^1. We have the following general (one-dimensional large deviation) result in this direction. We shall closely follow the techniques of Bahadur and Ranga Rao (1960).

THEOREM 3.4 Let $\{Y_n\}_{n\geq 1}$ be a sequence of IID random variables on R^1 with the moment generating function

$M(t) = E(\exp(t\, Y_1))$. Let

$$p_n = \text{Prob}(n^{-1/2} \sum_{i=1}^{n} Y_i > n^{1/2} \mu + q_n) \qquad n \geq 1$$

where μ is a nonzero constant and $\{q_n\}_{n \geq 1}$ is a bounded sequence of reals. Assume that

(a) the distribution of Y_1 is nonlattice;

(b) if T stands for the set $\{t : M(t) \text{ is finite}\}$, then T is a nondegenerate interval; and

(c) there exists a <u>positive</u> $t_o(\mu)$ in the interior of T such that

$$\exp(-\mu t_o(\mu))M(t_o(\mu)) = \inf \{\exp(-\mu t)M(t) : t \in T\}$$
$$= \rho(\mu) \quad \text{say} \qquad 0 < \rho(\mu) < 1.$$

Then one has,

$$\log p_n = n \log \rho(\mu) - n^{1/2} q_n \cdot \frac{d \log \rho(\mu)}{d\mu} - \frac{1}{2} \log n$$
$$- \frac{1}{2} \{\log(2\pi\, \alpha^2(\mu)) + q_n^2 \frac{d^2 \log \rho(\mu)}{d\mu^2}\} + o(1),$$

where $\alpha(\mu)$ is a positive constant; more specifically,

$$\alpha(\mu) = t_o(\mu)\, \sigma(\mu), \quad \sigma(\mu) = \{\frac{M''(t_o(\mu))}{M(t_o(\mu))} - \mu^2\}^{1/2},$$

$M''(t)$ being the second derivative (with respect to t) of $M(t)$. One may verify that

(i) $$t_o(\mu) = -\frac{d}{d\mu}(\log \rho(\mu))\,;$$

(ii) $$\sigma^2(\mu) = \left(\frac{d^2}{d\mu^2}(\log \rho(\mu))\right)^{-1}.$$

PROOF : We shall follow the ideas of Bahadur and Ranga Rao (1960). Let H_n be the distribution function of the standardized n-fold convolution of the conjugate distribution of $Y_1 - \mu$. Then $\sigma(\mu)$ is the standard deviation of this conjugate distribution. Proceeding exactly in the same way as in [redacted] Lemma 2 of the above mentioned paper, we have $p_n = \rho^n \cdot I_n$ where

$$I_n = \int_{q_n'}^{\infty} \exp(-n^{1/2}\,\alpha(\mu)\,x)\,dH_n(x),\quad q_n' = q_n/\sigma(\mu)\,.$$

One now uses Theorem 1, Chapter XV.4 of Feller (1968) and evaluates the integral I_n by direct computations; for details see Chandra and Ghosh (1978). The proof of (i) and (ii) above is easy.

REMARK 3.7 If we assume that the distribution of Y_1 satisfies Cramér's condition, we can get an expansion (in powers of $n^{-1/2}$) of $\log p_n$ similar to the one given in Theorem 2 of Bahadur and Ranga Rao (1960). More precisely, one can then show that

$$\log p_n = n \log \rho(\mu) - n^{1/2} t_o(\mu) q_n - \frac{1}{2} \log n$$
$$- \frac{1}{2}(\log(2\pi \alpha^2(\mu)) + q_n^2/\sigma^2(\mu)) + \sum_{j=0}^{2k-1} d_{n,j} n^{-j/2}$$
$$+ 0(n^{-k})$$

where $\{d_{n,1}\}$, $\{d_{n,2}\}$ etc. are suitable bounded constants.

Consider now the set-up of Section 2. Our main interest is to find an asymptotic expansion of $P(T_n > k_n ; \theta)$ where k_n is to be determined from condition (2.2). We assume that the distributions of $\{T_n\}$ under $P(\cdot ; \theta)$ and $\{P(\cdot;\theta_o);\theta_o \varepsilon \text{Ⓗ}_o\}$ satisfy the following conditions :

CONDITION A. There exist constants (free from n) $\mu(\theta)$ and $\sigma^*(\theta) > 0$ and a polynomial $q(\cdot; \theta)$ such that

$$P(T_n - n^{1/2} \mu(\theta) \leq \sigma^*(\theta) x ; \theta)$$
(3.4)
$$= \int_{-\infty}^{x} \phi(t)dt + n^{-1/2} q(x;\theta)\phi(x) + o(n^{-1/2}), \text{ uniformly in } x.$$

Here $\phi(t)$ is the density of the standard normal distribution.

CONDITION B. Whenever $\{q_n\}_{n\geq 1}$ is a bounded sequence of reals, p_n defined by

(3.5) $$p_n = \sup\{P(T_n > n^{1/2} \mu(\theta) + q_n ; \theta_o) : \theta_o \varepsilon \text{Ⓗ}_o\}$$

satisfies

(3.6) $\log p_n = -n\,a + n^{1/2}\alpha' q_n + c \log n + (\gamma_1 + q_n^2 \gamma_2) + o(1)$

for suitable constants $a, \alpha', c, \gamma_1, \gamma_2$ $(a > 0)$.

LEMMA 3.2 Assume that CONDITION A holds. Then

(3.7) $k_n = n^{1/2}\mu(\theta) + z_\beta \sigma^*(\theta) - n^{-1/2} q(z_\beta;\theta)\sigma^*(\theta) + o(n^{-1/2})$.

Here z_β is defined by

$$\int_{z_\beta}^{\infty} \phi(t)dt = \beta .$$

PROOF: Let us write

$$F_n(x) = P(T_n - n^{1/2}\mu(\theta) > \sigma^*(\theta)x \,;\, \theta),$$

$$G_n(x) = \int_{-\infty}^{x} \phi(t)dt + n^{-1/2} q(x;\theta)\,\phi(x).$$

Then $F_n(z_\beta - q(z_\beta;\theta)n^{-1/2}) = (1-\beta) + o(n^{-1/2})$. Put $d = -q(z_\beta;\theta)$, $\varepsilon_1(n) = \sup_x |F_n(x) - G_n(x)|$, $\varepsilon_2(n) = F_n(z_\beta + dn^{-1/2}) - (1-\beta)$. Choose $\varepsilon(n) \to 0$ such that $\varepsilon(n) > 0$ and $n^{1/2}\varepsilon_i(n) = o(\varepsilon(n))$ for each $i = 1,2$. Then

$$\frac{n^{1/2}}{\varepsilon(n)} \{F_n(z_\beta + (d + \varepsilon(n))\, n^{-1/2}) - (1-\beta)\}$$

$$= \frac{n^{1/2}}{\varepsilon(n)} \{G_n(z_\beta + (d+\varepsilon(n))n^{-1/2}) - G_n(z_\beta + dn^{-1/2})\} + o(1)$$

$$= G_n'(\xi_n) + o(1)$$

for some ξ_n lying between $z_\beta + d\, n^{-1/2}$ and $z_\beta + (d+\varepsilon(n))n^{-1/2}$ and so $\xi_n \to z_\beta$. As $\{G'_n(u)\}$ is bounded away from zero in a neighbourhood of z_β, one gets

$$\frac{n^{1/2}}{\varepsilon(n)} \{F_n(z_\beta + (d + \varepsilon(n))n^{-1/2}) - F_n((k_n - n^{1/2}\mu(\theta))/\sigma^*(\theta))\} > 0$$

for all sufficiently large n. This implies that

$$(k_n - n^{1/2}\, \mu(\theta))/\sigma^*(\theta) < z_\beta + (d + \varepsilon(n))\, n^{-1/2}$$

for all sufficiently large n. Similarly,

$$(k_n - n^{1/2}\, \mu(\theta))/\sigma^*(\theta) > z_\beta + (d - \varepsilon(n))n^{-1/2}$$

for all sufficiently large n. As $\varepsilon(n) \to 0$, the proof is complete.

THEOREM 3.5 Assume that CONDITIONS A and B hold. Then one has

(3.8) $$\log \alpha(n) = -na + n^{1/2}\, b + c \log n + d + o(1)$$

where

(3.9) $$b = \alpha' z_\beta\, \sigma^*(\theta), \quad d = \gamma_1 - \alpha' q(z_\beta;\theta)\sigma^*(\theta) + \gamma_2\, z_\beta^2(\sigma^*(\theta))^2.$$

The theorem is easily proved using the above lemma.

REMARK 3.8. It is well-known that $\{n^{-1/2} \sum_{i=1}^{n} Y_i\}_{n\geq 1}$ satisfies CONDITION A where $\{Y_n\}_{n\geq 1}$ is a sequence of IID non-lattice random variables with finite third moment. One of the results of Bhattacharya and Ghosh (1978) indicates that this condition is satisfied for a large collection of statistics (see Section 4 of Chapter Three). The results of the next chapter indicate that CONDITION B is also typically true.

REMARK 3.9 The proof of Lemma 3.2 shows that the above theorem remains essentially true even if in CONDITION A, (3.4) is replaced by

$$P(T_n - n^{1/2}\mu(\theta) \leq \sigma^*(\theta)\, x \,;\, \theta)$$

(3.10)

$$= \int_{-\infty}^{x} f(t)\,dt + F_1(x)\, n^{-1/2} + o(n^{-1/2}) \text{ uniformly in } x,$$

provided $f(x)$ is a strictly positive continuous density on R^1 and $F_1(x)$ is some continuous function on R^1.

(E) EXAMPLES :

Here we shall discuss two examples. In these (and many other) examples, the size functions $\{\alpha_i(n)\}_{n\geq 1}$ $i = 1,2$ admit extensions $\{\alpha_i(x) : 1 \leq x < \infty\}$ $i = 1,2$ which are continuous decreasing functions of x and moreover the following asymptotic expansions are valid :

(3.11)

$$\log \alpha_i(x;\beta,\theta) = -x\,a(\theta) + x^{1/2}\,b(\beta,\theta) - c\log x + d_i(\beta,\theta) + e(\beta,\theta)x^{-1/2} + o_i(x^{-1/2}) \quad i=$$

where $a(\theta) > 0$, $c > 0$, $b(\beta,\theta) = 0$ if and only if $\beta = 1/2$ and finally $d_1(\beta,\theta) - d_2(\beta,\theta)$ is always nonzero and does not depend on β. Also if $\beta = 1/2$,

$$
(3.12)\qquad \begin{aligned} \log \alpha_i(x;\beta,\theta) &= -x\, a(\theta) - c \log x + d_i(\tfrac{1}{2},\theta) + e(\tfrac{1}{2},\theta)x^{-1/2} \\ &\quad + f(\theta)x^{-1} + o_i(x^{-1}), \quad i = 1,2 . \end{aligned}
$$

From Theorem 3.3, the approximate BCD d can be calculated and will be free from β. To compute the upper and lower deficiencies, recall that if d is non-integral, then they are respectively $[d] + 1$ and $[d]$ and BCD cannot exist. Below we consider the case when d is integer. Then one can show that

$$
(3.13)\qquad n^{1/2}(m(n) - n - d) \to d\, b(\beta;\theta)\, (2a(\theta))^{-1} ;
$$

$$
(3.14)\qquad n(m(n)-n-d) \to -dc(a(\theta))^{-1} \quad \text{if } \beta = \tfrac{1}{2} ;
$$

Clearly then BCD cannot exist and from Remark 3.4 one can evaluate the upper and lower deficiencies. Thus the upper and lower deficiencies do depend on β but in a very weak sense. Our problem therefore reduces to show that (3.11) and (3.12) indeed hold in these examples.

The deficiency of one test procedure relative to another was defined in Section 2 for the <u>same</u> testing problem. In the following two examples, the two test procedures under comparison are for two <u>different</u> testing situations. More specifically, we want to compare one-sided test and two-sided test in what is apparently a one-sided testing problem. There are two reasons for doing this. Let Θ_1 and Θ_2 correspond to the two-sided and one-sided alternatives respectively, $\Theta_2 \subset \Theta_1$. Suppose now in a given problem the natural alternative is Θ_1 but there is some information (not entirely reliable) that the real alternative is Θ_2. In this case under the usual assumptions the likelihood ratio using Θ_2 is as Bahadur-efficient as the one using Θ_1 for all θ in Θ_2. So if Bahadur-efficiency were the only criterion, one should certainly ignore the information that Θ_2 is the real alternative. Our examples show that the choice is not so clear if one also considers the deficiency. The second reason for considering these examples is a mathematical one ; they are nontrivial and illustrate the various technical aspects of computing deficiency.

EXAMPLE 3.1 (The Normal Distribution).

Let Θ be the real line $(-\infty, +\infty)$, $\Theta_0 = \{0\}$. For $\theta \in \Theta$, let P_θ denote the normal distribution with mean θ and variance 1. Fix a positive θ.

For the testing problem H_0 that the population mean is zero against the alternative that it is non-zero, the critical region of the "best" test is given by

$$|n^{1/2}\bar{X}_n| > k_1(n) \qquad \bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i ,$$

where $k_1(n)$ is such that

(3.15) $$\beta = 1 - \Phi(k_1(n) - n^{1/2}\theta) + \Phi(-k_1(n) - n^{1/2}\theta).$$

Then its power at θ is β and its size is

(3.16) $$\alpha_1(n) = 2(1 - \Phi(k_1(n))).$$

Bahadur's (as well as Cochran's) slope of $|n^{1/2}\bar{X}_n|$ at θ is $\frac{1}{2}\theta^2$.

For the testing problem H_o that the population mean is zero against the alternative that it is positive, the critical region of the most powerful test is $\{n^{1/2}\bar{X}_n > k_2(n)\}$, where $k_2(n) = n^{1/2}\theta + z_\beta$. Its power at θ is β and its size is

(3.17) $$\alpha_2(n) = 1 - \Phi(n^{1/2}\theta + z_\beta)$$

Bahadur's slope of $n^{1/2}\bar{X}_n$ at θ is $\frac{1}{2}\theta^2$.

Thus the two test procedures are equally efficient when θ obtains. At the level of deficiency however their performances are different (as indicated at the beginning of (E)).

We want now to show that (3.11) and (3.12) hold with $a(\theta) = \theta^2/2$, $b(\beta, \theta) = -z_\beta\theta$, $c = -1/2$, $d_2(\beta,\theta) = -\frac{1}{2}(z_\beta^2 + \log(2\pi\theta^2))$, $d_1(\beta,\theta) = d_2(\beta,\theta) + \log 2$, $e(\beta,\theta) = -z_\beta\theta^{-1}$,

$f(\theta) = -\theta^{-2}$. In fact, one can assume that (3.15), (3.16) and (3.17) hold even if the integer n is replaced by the real $x(>0)$. That $\{\alpha_2(x)\}$ satisfies (3.11) and (3.12) is easy. To show the same for $\{\alpha_1(x)\}$, it suffices to show that $\exp(x\theta^2)\,(k_1(x) - x^{1/2}\theta - z_\beta)$ and $\exp(x\theta^2)\,(\log \alpha_1(x) - \log \alpha_2(x) - \log 2)$ tend to zero as $x \to \infty$. Since these are easy to establish, the details are omitted (see Chandra and Ghosh (1978)).

REMARK 3.10 One can easily see that although the above test procedures are equivalent from Bahadur's point of view, the Pitman efficiency of the first one with respect to the second is < 1; in fact, the Pitman efficiency at the alternatives $n^{-1/2}\theta(\theta > 0)$ is $(z_\alpha - z_{\beta(\theta)})^2\,\theta^{-2}$ where

$$\beta(\theta) = \int_{z_{\alpha/2}-\theta}^{\infty} \phi(t)\,dt + \int_{z_{\alpha/2}+\theta}^{\infty} \phi(t)\,dt \qquad (\alpha < \beta(\theta) < 1) \tag{3.18}$$

REMARK 3.11 Consider the above testing problem except that the observations are now taken from a one-parameter exponential family with density (with respect to some absolutely continuous sigma-finite measure)

$$f_\theta(x) = \exp(\theta x - c(\theta))$$

where θ takes values in the natural parameter space. Assume that the natural parameter space is an open set containing the origin and that $c(0) = c'(0) = 0$. Then the deficiency of the first testing procedure with respect to the second (i.e., the likelihood ratio test when the alternative is $\theta > 0$) is

$$\frac{\log(1 + \sqrt{c''(\hat{\theta}(x_o))/c''(\hat{\theta}(x_1))}\ )}{\theta c'(\theta) - c(\theta)} ;$$

here c' and c'' denote respectively the first and second derivative of c, $x_o > 0$ and $x_1 < 0$ are defined (uniquely) by

$$\phi(x_o) = \phi(x_1) = \theta\, c'(\theta) - c(\theta)$$

where

$$\phi(x) = x\,\hat{\theta}(x) - c(\hat{\theta}(x))$$

and $\hat{\theta}(x)$ is the unique solution of the equation $x = c'(\theta)$. The proof depends on the results of the next chapter and hence is omitted.

EXAMPLE 3.2 (The Student's t distribution)

Here $\textcircled{H} = \{(\mu,\sigma) : -\infty < \mu < \infty,\ 0 < \sigma < \infty\}$, $\textcircled{H}_o = \{0\} \times (0,\infty)$. For $\theta = (\mu,\sigma)$ in $\textcircled{H}$, let P_θ stand for the normal distribution with mean μ and variance σ^2. Fix a $\theta_1 = (\mu_1, \sigma_1)$ such that $\mu_1 > 0$, $\sigma_1 > 0$; put $\mu = \mu_1 \sigma_1^{-1}$, $\theta = (\mu,1)$ and $\theta_o = (o,1)$.

For the testing problem H_o that the population mean is zero against the alternative that it is non-zero (the population variance being unknown), the "best" test is based on the critical region $\{T_1(n) > k_1(n)\}$ where $T_1(n) = |n^{1/2}\,\bar{X}_n\, s_n^{-1}|$, $ns_n^2 = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ and

$$\beta = P_{\theta_1}(T_1(n) > k_1(n)).$$

Its power at θ_1 is β and size is

$$\alpha_1(n) = 2P_{\theta_o}(n^{1/2}\,\bar{X}_n\, s_n^{-1} > k_1(n))$$

Bahadur's slope at θ_1 of T_1 is $\frac{1}{2}\log\,(1 + \mu^2)$.

If now the alternative hypothesis is taken as "the mean is positive and the variance is unknown", then the "best" test is based on the critical region $\{T_2(n) > k_2(n)\}$ where $T_2(n) = n^{1/2}\,\bar{X}_n\, s_n^{-1}$ and $k_2(n)$ is such that $\beta = P_{\theta_1}(T_2(n) > k_2(n))$. Its power at θ_1 is β and size is

$$\alpha_2(n) = P_{\theta_o}(n^{1/2}\,\bar{X}_n\, s_n^{-1} > k_2(n))$$

Bahadur's slope at θ_1 of T_2 is $\frac{1}{2}\log\,(1 + \mu^2)$.

Here too the two test procedures are equally efficient, though their deficiency is not zero. We shall show that $\{\alpha_1(n)\}$ and $\{\alpha_2(n)\}$ satisfy (3.11) and (3.12) with $a(\theta_1) = \frac{1}{2}\log(1+\mu^2)$, $b(\beta,\theta_1) = -z_\beta(1 + \frac{1}{2}\mu^2)^{1/2}\,\mu(1+\mu^2)^{-1}$ and

$d_1(\beta, \theta_1) - d_2(\beta ; \theta_1) = \log 2$. The extensions $\{\alpha_i(x)\}$ can be defined as in the proof of Lemma 3.1. To show that $\{\alpha_2(n)\}$ satisfies (3.11), one proceeds as in the proof of Lemma 3.2 and gets constants d_1 and d_2 (we do not need the exact values of d_1 and d_2) such that

(3.19) $$k_2(n) = n^{1/2}\mu + (1 + \tfrac{1}{2}\mu^2)^{1/2}(z_\beta + d_1 n^{-1/2} + d_2 n^{-1}) + o(n^{-1});$$

here one uses the fact that there exist two polynomials P_1 and P_2 whose coefficients are free from n such that

$$P_{\theta_1}(T_2(n) - n^{1/2}\mu > u(1+\tfrac{1}{2}\mu^2)^{1/2})$$

$$= \int_u^\infty \phi(t)dt + (P_1(u)n^{-1/2} + P_2(u)\, n^{-1})\, \phi(u) + o(n^{-1})$$

uniformly in $u \in R^1$ (see Bhattacharya and Ghosh (1978)). Recall that under θ_o, $(n-1)^{1/2}\, n^{-1/2}\, T_2(n)$ follows Student's t-distribution with (n-1) degrees of freedom. A repeated application of integration by parts (vide Chandra and Ghosh (1978)) then yields the desired expansion for $\log \alpha_2(n)$. Consider now the case of $\{\alpha_1(n)\}$. Since under θ_1, $(T_2(n) - n^{1/2}\mu)(1+\tfrac{1}{2}\mu^2)^{-1/2}$ converges weakly to the standard normal distribution, $P_{\theta_1}(T_2(n) < -k_1(n))$ is $o(n^{-1})$; this implies that $k_1(n)$ also satisfies (3.19) and so

$$\log \alpha_1(n) = \log 2 + \log \alpha_2(n) + o(n^{-1/2}).$$

This completes the proof of the fact that $\{\alpha_1(n)\}$ also satisfies (3.11) and that $d_1(\beta ; \theta_1) - d_2(\beta ; \theta_1) = \log 2$. The proof of the fact that the sizes satisfy (3.12) should now be clear.

SECTION 4. BAHADUR'S APPROACH

In this section we shall consider two possible ways of measuring deficiency from the standpoint of Bahadur's theory of efficiency (see Bahadur (1960) and (1971)). It is shown by means of an example that Bahadur-deficiency in the strong sense need not exist even when BCD exists. A new interpretation of the latter is suggested in part (B).

(A) BAHADUR-DEFICIENCY IN THE STRONG SENSE.

Assume the set-up of Section 2. For each real t, let

$$F_{in}(t) = \sup\{P(T_i(n) > t ; \theta_0) : \theta_0 \in (\bar{H})_0\}$$

(4.1) $$L_{in}(s) = F_{in}(T_i(s_1, \ldots, s_n)).$$

For each δ, $0 < \delta < 1$, and for each s, let $N_i(\delta ; s)$ be the least integer $m \geq 1$ such that $L_{in}(s) < \delta$ for all $n \geq m$; and let $N_i(\delta ; s) = \infty$ if no such m exists $(i = 1,2)$. For some interesting properties of $N_i(\delta, s)$, see Section 7 of Bahadur (1971).

DEFINITION 4.1 The random lower (upper) Bahadur-deficiency at θ of the first testing procedure with respect to the second is

$$\underline{D}_B(\beta ; \theta) = (\text{a.s. } P_\theta) \liminf_{\delta \to 0} (N_1(\delta,s) - N_2(\delta,s))$$

$$(\overline{D}_B(\beta ; \theta) = (\text{a.s. } P_\theta) \limsup_{\delta \to 0} (N_1(\delta,s) - N_2(\delta,s))).$$

In case the above two deficiencies are equal, we say that the Bahadur-deficiency exists and is equal to the common value. As in the case of BCD, the main use of studying these random deficiencies is to discriminate tests with the same Bahadur-efficiency i.e., tests for which the (a.s. P_θ) limit of $N_2(\delta,s)/N_1(\delta,s)$ is 1.

In this approach, the main source of difficulty is that the quantities

$$\sup\{L_{in}(s) : n \geq m\} \qquad m \geq 1$$

are difficult to expand; any possible expansion would seem to depend on the particular sample sequence considered.

EXAMPLE 4.1 (*The Uniform Distribution*). Let θ be such that $0 < \theta < 1$; let $f_1(x)$ and $f_2(x)$ be respectively the densities of the uniform distributions over $[0, \theta]$ and $[0,1]$. Consider the problem of testing

$$H_0 : f = f_2 \qquad \text{vs.} \qquad H_1 : f = f_1$$

on the basis of the following two statistics :

$$T_1(n) = \max(x_4, x_6, \ldots, x_{2n})$$

$$T_2(n) = \max(x_1, x_3, \ldots, x_{2n-1}) \quad n \geq 2 .$$

Then clearly $\alpha_1(n) = \beta\, \theta^{n-1}$ and $\alpha_2(n) = \beta\, \theta^n$ so that the BCD exists and equals 1 for all β.

We are going to show that the (a.s. or stochastic) limit of $(N_1(\delta,s) - N_2(\delta,s))$, if it exists, cannot be degenerate. Clearly

$$P_\theta(N_1(\delta,s) = m) = P_\theta(N_2(\delta,s) = m+1) \quad m \geq 2,$$

and the distribution function of $N_2(\delta,s)$ under θ is given by

$$P_\theta(N_2(\delta,s)=m) = \begin{cases} \delta\, \theta^{1-p} \exp\left(\left(\sum_{j=m+1}^{p-1} j^{-1}\right) \log \delta\right) & \text{if } m \leq p-2 \\ \delta\, \theta^{1-p} & \text{if } m = p-1 \\ 1 & \text{if } m = p \end{cases}$$

where $p \equiv p(\theta;\delta)$ is the <u>integer</u> satisfying

$$\frac{\log \delta}{\log \theta} \leq p(\theta ; \delta) < \frac{\log \delta}{\log \theta} + 1 .$$

The next lemma studies the weak convergence under P_θ of $p(\theta ; \delta) - N_2(\delta, s)$ as $\delta \to 0$.

LEMMA 4.1 For each c, $0 \le c \le 1$, let X_c be a random variable such that $P_\theta(X_c = 0) = 1 - \theta^c$ and $P_\theta(X_c = i) = (1 - \theta)\,\theta^{c+i-1}$, $i \ge 1$. Let $e(\delta)$ be the excess over $(p(\theta\,;\,\delta)-1)$ of $\log \delta/\log \theta$, $0 < e(\delta) \le 1$. Then

(a) if $e(\delta_n) \to 0$ and $\delta_n \to 0$, $p(\theta\,;\,\delta_n) - N_2(\delta_n\,;\,s)$ converges weakly to X_c under P_θ;

(b) if $p(\theta\,;\,\delta_n) - N_2(\delta_n\,;\,s)$ converges weakly to X under P_θ and $\delta_n \to 0$, $\{e(\delta_n)\}_{n \ge 1}$ is a convergent sequence; moreover X can be taken to be X_c where c is the limit of $e(\delta_n)$.

PROOF: (a) By definition of $e(\delta_n)$,

$$\delta_n\, \theta^{1-p(\theta;\delta_n)} = \theta^{e(\delta_n)} \to \theta^c \quad \text{as} \quad n \to \infty .$$

Also the definition of $p(\theta\,;\,\delta)$ implies that for each $k \ge 1$,

$$\theta^{-1}\, \delta_n^{1/(p(\theta;\delta_n)-k)} \ge \theta^{k/(p(\theta;\delta_n)-k)}$$

and

$$\theta^{-1}\, \delta_n^{1/(p(\theta;\delta_n)-k)} < \theta^{(k-1)/(p(\theta;\delta_n)-k)} .$$

Thus for each $k \ge 1$, $\delta_n^{1/(p(\theta;\delta_n)-k)} \longrightarrow \theta$ as $n \to \infty$.

Consequently one has for each $m \ge 0$

$$P_\theta(p(\theta\,;\,\delta_n) - N_2(\delta_n\,;\,s) = m) \to P_\theta(X_c = m);$$

which completes the proof of (a).

(b) As $\{e(\delta_n)\}_{n\geq 1}$ is a bounded sequence, it has a convergent subsequence. By part (a), every convergent subsequence of it converges to the same number. Part (b) is therefore immediate.

It follows from the above lemma that the (a.s. P_θ or stochastic) limit of $N_1(\delta;s) - N_2(\delta;s)$ cannot be degenerate; to see this, one needs only to note that $N_1(\delta;s)$ and $N_2(\delta;s)$ are independent and then to use Theorem 2, Chapter VIII.3 of Feller (1968). Thus in this example one cannot hope to get a single numerical value of deficiency from Bahadur's point of view.

(B) ANOTHER INTERPRETATION OF BCD :

So far we have considered asymptotic efficiencies of test procedures in terms of almost sure convergence only. However it is possible to discuss the same with almost sure convergence replaced by convergence-in-probability; indeed this was the approach of Bahadur (1960). Although Bahadur's (exact) slopes are easier to interpret in the former case, the (approximate) slopes based on convergence-in-probability are not only easier to use but perhaps more basic and stable. For these reasons, we now consider the following alternative measure of deficiency from Bahadur's point of view.

The following result is due to Raghavachari (1970); see his Theorem 2. The set-up is same as that given in Section 2.

LEMMA 4.2 (Raghavachari)

For all β, $0 < \beta < 1$, one has

$$\lim_{n\to\infty} n^{-1} \log \alpha(n;\beta,\theta) = -c(\theta),\ 0 < c(\theta) < \infty$$

if and only if

$$n^{-1} \log L_n(s) \xrightarrow{P_\theta} -c(\theta) .$$

This fact leads to the following definition.

DEFINITION 4.2 Fix a $\theta \notin (\mathrm{H})_0$, an ε with $0 < \varepsilon < 1$ and a δ with $0 < \delta < 1$. Then $V(\delta;\varepsilon,\theta)$ is the smallest integer $m \geq 1$ such that whenever $n \geq m$

$$P_\theta(L_n(s) < \delta) > 1 - \varepsilon ;$$

and let $V(\delta;\varepsilon,\theta) = +\infty$ if no such m exists.

The next lemma gives the asymptotic behaviour of $V(\delta;\varepsilon,\theta)$ as $\delta \to 0$.

LEMMA 4.3 Assume that

$$(4.2) \qquad n^{-1} \log L_n(s) \xrightarrow{P_\theta} -c(\theta) \qquad 0 < c(\theta) < \infty .$$

Then for each ε, $0 < \varepsilon < 1$ one has

(4.3) $$\frac{\log \delta}{V(\delta;\ \varepsilon,\ \theta)} \longrightarrow -\, c(\theta) \qquad \text{as} \quad \delta \rightarrow 0 .$$

PROOF: Fix a θ and an ε. We abbreviate $V(\delta;\varepsilon,\theta)$ and $c(\theta)$ as $V(\delta)$ and c respectively. Also for notational convenience we shall write $L_n(s)$ as $L(n)$.

Since c is finite, $V(\delta)$ tends to ∞ as δ tends to zero. Clearly then it suffices to prove that

(4.4) $$\liminf_{\delta \rightarrow 0} [-\log \delta/(V(\delta)-1)] \geq c ;$$

and

(4.5) $$\limsup_{\delta \rightarrow 0} [-\log \delta/V(\delta)] \leq c .$$

To prove (4.4) we assume, by way of contradiction, that (4.4) is false. Then there exist an η and a sequence $\{\delta_n\}$ such that $0 < \eta < c$, $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ and

$$-\frac{\log \delta_n}{V(\delta_n)-1} < c - \eta .$$

From the definition of $V(\delta)$, we have

$$P_\theta\left(-\frac{\log \delta_n}{V(\delta_n)-1} \geq -\frac{\log L(V(\delta_n)-1)}{V(\delta_n)-1}\right) \geq \varepsilon .$$

So certainly

$$P_\theta\left(c - \eta \geq \frac{\log L(V(\delta_n)-1)}{V(\delta_n)-1}\right) \geq \varepsilon$$

which contradicts Assumption (4.2).

The proof of (4.5) is similar.

Lemma 4.3 suggests the following measure of deficiency; consider the set-up of Section 2 and define $V_1(\delta;\varepsilon,\theta)$ and $V_2(\delta;\varepsilon,\theta)$ similarly using $L_{1n}(s)$ and $L_{2n}(s)$ for $L_n(s)$ in Definition 4.2.

DEFINITION 4.3 Fix a $\theta \not\in (\bar{H})_o$ and an ε, $0 < \varepsilon < 1$. Then the lower (upper) deficiency at θ of the first testing procedure with respect to the second is

$$\liminf_{\delta \to 0} (V_1(\delta;\varepsilon,\theta) - V_2(\delta;\varepsilon,\theta))$$

$$(\limsup_{\delta \to 0} (V_1(\delta;\varepsilon,\theta) - V_2(\delta;\varepsilon,\theta))).$$

Let $F_{in}(t)$ be a strictly decreasing continuous function of t, $i = 1,2$. For each $\theta \not\in (\bar{H})_o$, we make the same assumption about $P_\theta(T_{in} > t)$. For each δ, $0 < \delta < 1$, let $t_{in}(\delta) = F_{in}^{-1}(\delta)$. Consider the sequence of tests $\phi_{in}(\delta)$:

Reject H_o if and only if $T_{in} > t_{in}(\delta)$.

Then the error of first kind for this test is δ. We denote its power by β^*_{in}.

Fix $\theta \not\in (\bar{H})_o$ and define, for each $\beta, 0 < \beta < 1$, the test $\psi_{in}(\beta)$:

Reject H_o if and only if $T_{in} > c_{in}(\beta)$

where $c_{in}(\beta)$ is such that $P_\theta(T_{in} > c_{in}(\beta)) = \beta$. Let its error of first kind be denoted by α_{in}.

Using the tests $\phi_{in}(\beta)$ define $M_i(\delta; \beta, \theta)$ as in Section 2. Then,

$$M_i(\delta\,;\,\beta,\,\theta) = V_i(\delta\,;\,1-\beta,\,\theta). \tag{4.6}$$

To see this, note that if $n \geq V_i(\delta\,;\,1-\beta,\,\theta)$, then by definition of $V_i(\delta; 1-\beta, \theta)$, the tests ϕ_{in} have error of first kind $= \delta$ and power at $\theta > \beta$. Hence for $n \geq V_i(\delta; 1-\beta, \theta)$ the tests ϕ_{in} which have power $= \beta$, must have error of first kind $\alpha_{in} < \delta$. By definition of $M_i(\delta; \beta, \theta)$, this means that $M_i(\delta; \beta, \theta) \leq V_i(\delta; 1-\beta, \theta)$. Similarly the reverse inequality can be proved.

Thus BCD, upper, lower or approximate, agrees with the corresponding notion as may be defined using $V_i(\delta; \varepsilon, \theta)$.

CHAPTER FIVE

DEFICIENCY FOR MULTIVARIATE TESTING PROBLEMS

SECTION 1. INTRODUCTION AND NOTATIONS

In the previous chapter we considered the problem of discriminating tests with same (Bahadur-) efficiency; the examples considered there concern with essentially the "one parameter" case. Here we shall compare the likelihood ratio test with other common (and equivalent) test procedures for the multiparameter exponential family. For this the main problem is to find expansions up to $o(1)$ of the logarithms of multi-dimensional large deviation probabilities (see the last paragraph of the introduction of the previous chapter); to this end we have described a method which is fairly general and also may work even for some non-exponential families. Some closely related works are Borovkov and Rogozin (1965) and Woodroofe (1978); some of the results of Bahadur and Zabell also have peripheral connections (but we do not explicitely need them).

Consider a simple null hypothesis $H_0 : \theta = \theta_0$, a composite alternative $H_1 : \theta \in \Theta_1$, two test statistics $T_{1,n}$ and $T_{2,n}$ and critical regions of the form $\{T_{i,n} > t_{i,n}\}$. Let $\alpha_{i,n}$ and $\beta_{i,n}(\theta)$, $i = 1,2$, denote the error of first kind and the power function respectively. Throughout this chapter, θ will denote a fixed element of $(\overline{H})_1$ and we shall choose $t_{i,n}$ such that $\beta_{1,n}(\theta) = \beta_{2,n}(\theta) = \beta$, β being fixed and strictly between zero and one. As shown in the previous chapter, the following expansions for $\alpha_{i,n}(\beta;\theta)$, $i = 1,2$, will be valid under quite general conditions:

$$\log \alpha_{i,n}(\beta,\theta) = -na(\beta,\theta) + n^{1/2} b_i(\beta,\theta) + (\log n)c_i(\beta,\theta) + d_i(\beta,\theta) + o(1), \quad 0 < a(\beta,\theta) < \infty . \tag{1.1}$$

Then for sufficiently small δ, there exist $n_1(\delta)$ and $n_2(\delta)$ such that $\alpha_{i,x} = \delta$ if and only if $x = n_i(\delta)$; moreover, the limit of $(n_1(\delta) - n_2(\delta))$ as $\delta \to 0$ exists and equals $\lambda = (d_1(\beta,\theta) - d_2(\beta,\theta))/a(\beta,\theta)$ if $b_1(\beta,\theta) = b_2(\beta,\theta)$ and $c_1(\beta,\theta) = c_2(\beta,\theta)$; vide Lemma 3.1 and Theorem 3.3 of Chapter Four. This limit is the approximate Bahadur-Cochran deficiency (of $T_{1,n}$ with respect to $T_{2,n}$) to be abbreviated henceforth as deficiency. Another simple interpretation of this deficiency is provided by the relation

$$\alpha_{2,n} = (\exp(-a\lambda) + o(1))\, \alpha_{1,n} , \tag{1.2}$$

showing how large $\alpha_{1,n}$ is compared with $\alpha_{2,n}$. In the subsequent pages we calculate this quantity (or equivalently show that the sizes indeed satisfy (1.1)) in some common multiparameter multivariate problems.

In Section 2 we have a sample of size n from a k-variate normal population with the mean vector $\theta = (\theta^{(1)},\ldots,\theta^{(k)})$ and identity as the dispersion matrix; the null hypothesis $H_o : \theta = o$ is tested against $H_1 : \theta^{(i)} > 0$, $i = 1,2,\ldots,k$. Here $T_{2,n}$ is the likelihood ratio statistic and $T_{1,n}$ is the likelihood ratio statistic against the unrestricted alternative $H_1' : \theta \neq 0$. The deficiency at θ is

(1.3) $$\lambda = k(2 \log 2) ||\theta||^{-2}$$

and what is more illuminating in this case, $\alpha_{2,n} = (2^{-k}+o(1))\alpha_{1,n}$. A similar result holds when $k = 2$ and the alternative is restricted to a non-convex set

(1.4) $$\{\theta \mid \theta^{(1)} > 0 \text{ and } \theta^{(2)} > 0 \text{, or, } \theta^{(1)} < 0 \text{ and } \theta^{(2)} < 0\}$$

In Section 3 we consider a sample of size n from a bivariate normal population with the mean vector θ and dispersion matrix identity; the null hypothesis is $H_o : \theta = 0$ to be tested against $H_1 : \theta \neq 0$. Here $T_{1,n}$ is the likelihood ratio statistic and

(1.5) $$T_{2,n} = \int f_{\theta,n} \pi(\theta) \, d\theta / f_o$$

where $f_{\theta,n}$ is the joint density under θ and π is a prior density; the second procedure will be referred to as the Bayes test. Assuming that $\pi(\theta)$ is continuous and positive everywhere we first approximate $T_{2,n}$ and then approximate $t_{2,n}$ (here Lemma 3.2 of Chapter Four is needed). Under additional conditions, $\alpha_{2,n}$ is evaluated; the value of $\alpha_{1,n}$ remains the same as in Section 1 and so the deficiency can be calculated. It turns out that the deficiency of the likelihood ratio test is less than or equal to

$$(1.6) \qquad \|\theta\|^{-2} \log\{\pi D_{\phi}^{2} g(\|\theta\|^{2}, \phi_{o})\}$$

where $-\frac{1}{2}g$ is the logarithm of the prior density written in terms of the polar coordinates, (i.e.,

$$(1.7) \qquad g(r,\phi) = -2 \log \pi(r^{1/2} \cos \phi, r^{1/2} \sin \phi))$$

$D_{\phi}^{2} g$ denotes the second order partial derivative of g with respect to ϕ and ϕ_o is the value of ϕ at which, for fixed $r = \|\theta\|^2$, g attains its minimum. Thus the deficiency at θ depends on the curvature of g on the circle that passes through θ and has origin at the centre. If the prior density is bivariate normal with zero mean vector, the variances equal to σ^2 and correlation coefficient $\rho \neq 0$, then the deficiency is

$$(1.8) \qquad \|\theta\|^{-2} [- b(\theta^{(1)} \mp \theta^{(2)})^2 + \log \{\pi b\|\theta\|^2\}]$$

according as ρ is positive or negative; here $b = |\rho|/\sigma^2(1-\rho^2)$. (Here g attains its minimum at two values of ϕ and hence a slight modification of (1.6) is needed.)

In Section 4, we extend the results of Section 3 to linear exponential densities on R^k

(1.9) $$f_\theta(x) = \exp\{\sum_{i=1}^{k} \theta^{(i)} x^{(i)} - c(\theta)\}$$

with respect to an absolutely continuous measure with θ varying over the natural parameter space which is assumed to be an open set. We test $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ and define $T_{1,n}$ and $T_{2,n}$ as before. Under the assumptions that π is positive everywhere and twice continuously differentiable we show that an approximate Bayes test is given by the critical region $\{\bar{X}_n \in S_n\}$ where

(1.10) $$S_n : \langle \hat{\theta}, \bar{X}_n \rangle - c(\hat{\theta}) - n^{-1}(\sigma(\hat{\theta}) - \log \pi(\hat{\theta})) > k'_{2,n}$$

Here $\hat{\theta} \equiv \hat{\theta}(\bar{X}_n)$ is the maximum likelihood estimator and

(1.11) $$\sigma(\theta) = \frac{1}{2} \log(\det(c''(\theta))),$$

$c''(\theta)$ being the $k \times k$ matrix of the second order partial derivatives of c. This follows from the following identity

(1.12) $$\begin{aligned} T_{2,n} &= \int \exp\{n \langle \theta, \bar{X}_n \rangle - nc(\theta) + \log \pi(\theta)\} d\theta \\ &= (2\pi)^{k/2} n^{-k/2} \exp\{n \langle \hat{\theta}, \bar{X}_n \rangle - c(\hat{\theta}) - \sigma(\hat{\theta}) + \\ &\qquad \log \pi(\hat{\theta})\} (1 + o(1)) \end{aligned}$$

where the $o(1)$ term goes to zero as $n \to \infty$ uniformly over compact sets of $\bar{X}_n$; incidentally equation (1.12) may be regarded as a refinement of the main result of Schwarz (1978), under conditions stronger than his. Under some technical conditions, $\alpha_{2,n}$ has been found. Using Theorem 1 of Woodroofe (1978), deficiency can then be readily computed and bounded by a certain integral involving the curvature of $\sigma(\hat{\theta}(x)) - \log \pi(\hat{\theta}(x))$ for fixed $\phi(x) = \langle \hat{\theta}(x), x \rangle - c(\hat{\theta}(x))$. Section 2 contains the special case where the set M_n defined below is zero dimensional.

(1.13) $$M_n = \{x \in S_n \mid \phi(x) = \phi_{min}\}$$

(1.14) $$\phi_{min} = \inf\{\phi(x) \mid x \in S_n\}$$

In Section 4, only the main steps of the computations will be given since the arguments are quite similar (at least conceptually) to those given in Section 3. One may note here that the function $\phi(x)$ is the negative of the point entropy (and hence is the Fenchel Transform $c^*(x)$) introduced by Bahadur and Zabell (1979).

The remarks of Section 4 are intended to clarify the technical assumptions. In particular the final remark indicates how our results are related to Theorem 1 of Woodroofe (1978) and can be used to get expansions for the large deviation probabilities for a class of statistics which includes the likelihood ratio criterion.

Results of the same type for composite hypotheses have been obtained, but they require even more technical conditions and so are not mentioned here.

IN THE SUBSEQUENT SECTIONS WE SHALL ADHERE TO THE NOTATIONS INTRODUCED ABOVE UNLESS OTHERWISE STATED.

SECTION 2. NORMAL WITH RESTRICTED MEAN VECTOR

We shall first find the size of the unrestricted likelihood ratio test. The critical region of this test is $\{\|\bar{X}\|^2 > k_{1,n}\}$. Since $\|\bar{X}\|^2$, when suitably normalised, has an Edgeworth expansion in powers of $n^{-1/2}$, there exists a constant $d(k)$ (free from n) such that

$$(2.1) \qquad k_{1,n} = \|\theta\|^2 + 2\|\theta\| z_\beta n^{-1/2} + 2\|\theta\| d(k) n^{-1} + o(n^{-1})$$

$(d(k) = (z_\beta^2 + k - 1)/2\|\theta\|)$. Now using polar transformation one can verify that

$$\alpha_{1,n} = \frac{(nk_{1,n})^{\frac{k}{2}-1}\exp(-\frac{n}{2}k_{1,n})}{2^{k/2}\Gamma(k/2)} \int_0^\infty \frac{2}{n}\cdot e^{-u/2}\left(1+\frac{u}{k_{1,n}}\right)^{\frac{k}{2}-1} du$$

(2.2)

$$= \frac{(nk_{1,n}/2)^{k/2-1}\exp(-nk_{1,n}/2)}{\Gamma(k/2)}(1+o(1))$$

We now come to the case of the restricted likelihood ratio test. Its critical region is of the form

$$\{\|\bar{X}\| > k_{2,n},\ \ \bar{X}^{(i)} > 0,\ \ i = 1,2,\ldots,k\}$$

$$\cup\ \{\sum_{i\notin J}(\bar{X}^{(i)})^2 > k_{2,n},\ \ \bar{X}^{(i)} < 0\ \text{ iff }\ i \in J\}$$

where the union is taken over all nonempty proper subsets J of $\{1,\ldots,k\}$. One sees immediately from the well-known estimate of the tail of the standard normal distribution that

$$P_\theta(\|\bar{X}\|^2 > k_{2,n}) - \beta \to 0 \qquad \text{as} \quad n \to \infty$$

at an exponential rate. Consequently one can verify that

$$k_{2,n} - k_{1,n} = o(n^{-i}) \qquad \text{for all} \quad i \geq 1.$$

It is then evident that

$$P_{\theta=0}(\|\bar{X}\|^2 > k_{2,n},\ \ \bar{X}^{(i)} > 0\ \ i = 1,\ldots,k) = (2^{-k}+o(1))\alpha_{1,n}$$

and that for any J

$$P_{\theta=0}\left(\sum_{i\notin J} (\overline{X}^{(i)})^2 > k_{2,n}\ ,\quad \overline{X}^{(i)} < 0 \quad \text{iff} \quad i \in J\right)$$

$$= 2^{-j} \text{ (the right side of (2.2) with } k \text{ replaced by } (k-j))$$

$$= o(\alpha_{1,n})\ ;$$

here j is the number of elements of J. Thus $\alpha_{2,n} = (2^{-k}+o(1))\alpha_{1,n}$ implying that the deficiency of the unrestricted likelihood ratio test with respect to the restricted likelihood ratio test is as given in (1.3).

We now consider the above problem except that the new alternative hypothesis specifies a non-convex set of θ. Specifically let $k = 2$ and H_1 be as in (1.4). Then it is easy to verify that the critical region is the <u>complement</u> of the <u>bounded convex</u> set :

$$\{\|\overline{X}\|^2 \leq k_n\ ,\quad \text{if } \overline{X}^{(1)}\,\overline{X}^{(2)} > 0\}$$

$$\cup \ \{|\overline{X}^{(1)}| \leq k_n^{1/2} \text{ and } |\overline{X}^{(2)}| \leq k_n^{1/2} \text{ if } \overline{X}^{(1)}\,\overline{X}^{(2)} < 0\}$$

(k_n being a suitable positive constant). Proceeding as above, one can verify that $\alpha_{2,n} = (2^{-1} + o(1))\ \alpha_{1,n}$ and so the deficiency is $\|\theta\|^{-2}\log 2$.

SECTION 3. COMPARISON OF BAYES AND LIKELIHOOD RATIO TESTS FOR THE MEAN VECTOR OF A BIVARIATE NORMAL POPULATION

To get an asymptotic expansion of $\alpha_{2,n}$, it is convenient to write the Bayes critical region (see 1.5), in the form

$$(3.1) \qquad 2n^{-1}\log((2\pi)^{-1} n T_{2,n}) > k_{2,n} .$$

The reason for writing the critical region in the form will become clear from (3.2) below. Now

$$T_{2,n} = \exp(\tfrac{n}{2}\|\overline{X}\|^2) \int \exp(-\tfrac{n}{2}\|\theta\|^2 + \log \pi(\theta + \overline{X}))\,d\theta$$

$$= \exp(\tfrac{n}{2}\|\overline{X}\|^2) \Big\{ \int_0^{\delta}\int_0^{2\pi} 2^{-1} \exp(-\tfrac{nr}{2} + \log \pi(\overline{X}^{(1)}$$

$$+ r^{1/2}\cos\phi,\ \overline{X}^{(2)} + r^{1/2}\sin\phi)) + O(\exp(-n\delta/2))\Big\}$$

$$(3.2) \qquad = 2\pi n^{-1}\exp(\tfrac{n}{2}\|\overline{X}\|^2 + \log \pi(\overline{X}))\,(1 + o(1)),$$

where the $o(1)$ term is uniformly small on compact sets of $\overline{X}$. (Here we assume that $\pi(\theta)$ is positive and continuous everywhere.) We therefore consider an approximate Bayes test whose critical region is $\{ T'_{2,n} > k'_{2,n} \}$ where

$$(3.3) \qquad T'_{2,n} = \|\overline{X}\|^2 + 2n^{-1}\log \pi(\overline{X})$$

and $k'_{2,n}$ is determined so that its power at θ is $\beta + o(n^{-1/2})$. To get an expansion for $k'_{2,n}$, observe that

$$T'_{2,n} = \|\theta\|^2 + 2n^{-1}\log \pi(\theta) + n^{-1/2}\{2\|\theta\|\, U + n^{-1/2}(U^2+V^2)\} + o_p(n^{-1}).$$

where U, V are (appropriate linear) functions of $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ and are IID $N(0;1)$. Clearly then the statistic $T'_{2,n}$ has an Edgeworth expansion in powers of $n^{-1/2}$. There exists therefore a constant d such that

(3.4) $$n^{1/2}(k'_{2,n} - \|\theta\|^2 - \frac{2}{n}\log \pi(\theta))/2\|\theta\| = z_\beta + n^{-1/2} d + o(n^{-1/2})$$

(vide Lemma 3.2 of Chapter Four). It is to be noted that d does not depend on the prior π; in fact d can alternatively be determined from the condition that

$$\text{Prob}(U + (U^2 + V^2)/(2\|\theta\|\, n^{1/2}) > z_\beta + n^{-1/2} d) = \beta + o(n^{-1/2})$$

where U, V are IID $N(0;1)$ ($d = (z_\beta^2 + 1)/2\|\theta\|$).

To evaluate the size $\alpha'_{2,n}$ of the approximate test, we apply the standard polar transformation and get

$$\alpha'_{2,n} = n(4\pi)^{-1} \int\int_{r - n^{-1}g(r,\phi) \geq k'_{2,n}} \exp(-nr/2)\, dr\, d\phi$$

where $g(r, \phi)$ is as in (1.7). Let

$$r_{o,n} = \inf\{r : 0 < r < \infty,\ 0 < \phi < 2\pi,\ r - n^{-1} g(r,\phi) \geq k'_{2,n}\}$$

and fix $\delta > 0$. Then

$$\alpha'_{2,n} = n(4\pi)^{-1} \exp(-n\, r_{o,n}/2)\{\int_0^{\delta} \int_{A_n} \exp(-nr/2)\, d\phi\, dr \tag{3.5}$$
$$+ O(\exp(-n\delta/2))\}$$

where

$$A_n = \{\phi : 0 < \phi < 2\pi,\ r_{o,n} + r - n^{-1} g\,(r_{o,n} + r, \phi) \geq k'_{2,n}\}.$$

<u>We now want to replace</u> A_n <u>by a suitable (approximating) interval of</u> ϕ. Note that $\{r_{o,n}\}$ is bounded and the infimum is attained on the boundary and hence

(i) $r_{o,n} > 0$ and $r_{o,n} - k'_{2,n} \to 0$ as $n \to \infty$.

We now assume that

(ii) there exists a unique ϕ, say $\phi_{o,n}$, such that $r_{o,n} - n^{-1}\, g(r_{o,n}, \phi) = k'_{2,n}$ and that $\phi_{o,n} \to \phi_o$ (say), $0 < \phi_o < 2\pi$;

(iii) g is twice continuously differentiable with respect to r and ϕ; and finally

(iv) $D^2_{\phi}\, g\,(\|\theta\|^2, \phi_o) > 0$.

Then by (i), (ii) and (iii) and equation (3.4), one can conclude that

$$r_{o,n} = \|\theta\|^2 + 2\|\theta\| \{z_\beta\, n^{-1/2} + d\, n^{-1}\} + 2n^{-1}\{\log \pi(\theta) - \tag{3.6}$$
$$\log \pi(\|\theta\| \cos \phi_o,\ \|\theta\| \sin \phi_o)\} + o(n^{-1})$$

and that a Taylor expansion of g around $(r_{o,n}, \phi_{o,n})$ yields

$$r_{o,n} + r - n^{-1}\, g(r_{o,n} + r, \phi)$$
$$= k'_{2,n} + r\{1 - n^{-1}\, D_r\, g \cdot (1 + o_r(1))\}$$
$$- (2n)^{-1}\, (\phi - \phi_{o,n})^2\, D_\phi^2\, g \cdot (1 + o_r(1)) ,$$

where $D_r\, g$ and $D_\phi^2\, g$ denote respectively the first order and second order partial derivative of g with respect to r and ϕ and $o_r(1)$ denotes a term which goes to zero uniformly in n as $r \to 0$; here we have used the fact that the partial derivative $D_\phi\, g$ of g with respect to ϕ vanishes at $(r_{o,n}, \phi_{o,n})$. Above and henceforth all partial derivatives of g are evaluated at $(r_{o,n}, \phi_{o,n})$. In view of the last equality, we can approximate A_n by the interval of ϕ

$$|\phi - \phi_{o,n}| \leq t_n\, (nr)^{1/2}$$

where

$$t_n = 2^{1/2}\, (1 - n^{-1} D_\phi\, g)^{1/2}\, (D_\phi^2\, g)^{-1/2} \tag{3.7}$$

In fact, A_n contains and is contained in intervals of the type

$\phi_{o,n} \pm t_n(nr)^{1/2}\,(1 + o_r(1))$. Thus from (3.5) one gets

$$\alpha'_{2,n} = (2\pi)^{-1/2}\, t_n \exp(-nr_{o,n}/2)(1 + o(1))$$

(3.8) $$= \exp\{-\tfrac{n}{2}\|\theta\|^2 - \|\theta\|\, z_\beta\, n^{-1/2} - \|\theta\|\, d - \tfrac{1}{2}\log t - \log\pi(\theta) + \log\pi(\|\theta\|\cos\phi_o, \|\theta\|\sin\phi_o)\}.(1+ o(1))$$

where

(3.9) $$t = \pi\, D_\phi^2\, g(\|\theta\|^2, \phi_o)$$

(g is defined in (1.7)).

We now return to the (exact) Bayes test with power β. Fix $\delta > 0$ and choose a compact set C such that

(3.10)
(a) $P_\theta(\bar{X} \notin C) = o(n^{-1/2})$;
(b) $P_{\theta=0}(\bar{X} \notin C) = O(\exp(-\tfrac{n}{2}(\|\theta\|^2 + \delta_1)))$ $\qquad \delta_1 > 0$.

Then for any $\varepsilon > 0$ the set

$$\{2n^{-1}\log((2\pi)^{-1}\, n\, T_{2,n}) > k_{2,n}\} \cap \{\bar{X} \in C\}$$

lies between the sets

$$\{T'_{2,n} > k_{2,n} \pm 2\|\theta\|\, \varepsilon\, n^{-1}\} \cap \{\bar{X} \in C\}.$$

Also (3.2) and (3.10a) imply that $T_{2,n}$ and T'_{2n} (in appropriate normalised form) have same Edgeworth expansion up to $o(n^{-1/2})$. From this one derives as before that

(3.11) $$k_{2,n} = k'_{2,n} + o(n^{-1}).$$

It follows from the above facts and the expansion for the size of the approximate Bayes test that

$$\begin{aligned}&(\text{the right side of (3.8) with } d \text{ replaced by } (d+\varepsilon)) \\ &+ O(\exp(-\tfrac{n}{2}(\|\theta\|^2 + \delta_1))) \\ &\leq \alpha_{2,n} \\ &\leq (\text{the right side of (3.8) with } d \text{ replaced by } (d-\varepsilon)) \\ &\quad + O(\exp(-\tfrac{n}{2}(\|\theta\|^2 + \delta_1))),\end{aligned}$$

which shows that (3.8) can be taken as the expansion for $\alpha_{2,n}$ as well.

The last fact and equations (2.2) and (2.1) together imply that the deficiency of the lielihood ratio test with respect to the Bayes test is

$$(3.12) \qquad \|\theta\|^{-2}[-2 \log \{\pi(\|\theta\| \cos \phi_o, \|\theta\| \sin \phi_o)/ \pi(\theta)\} + \log t]$$

(t is defined in (3.9)). From our assumptions, it follows that for fixed $\|\theta\|$, the function $\pi(\|\theta\| \cos \phi_o, \|\theta\| \sin \phi_o)$ is maximized at $\phi = \phi_o$. Hence the deficiency (3.12) is less than or equal to the expression (1.6).

If in Assumption (ii) we have $\phi_o = 0$ or 2π, then the above analysis goes through except for a minor change ; in particular (3.12) gives the deficiency provided one redefines t as follows :

$$t = 2\pi \, D_\phi^2 \, g(\|\theta\|^2 , \phi_o) .$$

If in Assumption (ii) we have $D_\phi^{2i} \, g(\|\theta\|^2 , \phi_o) = 0$ for $i = 1,\ldots,p-1$, but $D_\phi^{2p} \, g(\|\theta\|^2, \phi_o) > 0$ <u>and</u>

$$D_\phi^{2i} \, g(r_{o,n}, \phi_{o,n}) = 0 \quad \text{for} \quad i = 1,\ldots,p-1, \quad p \geq 1$$

(and Assumption (iii) is changed accordingly), then the expansion of the size of the Bayes test will still be given by (3.8) <u>provided</u> one redefines t as follows :

$$t = \frac{\pi^2 \, (D_\phi^{2p} \, g)^{1/p}}{((2p)!)^{1/p} \, (\Gamma(1/2p+1))^2 \, 2^{1/p}} \qquad p \geq 1$$

If Assumption (ii) holds with $\phi^1_{o,n},\ldots,\phi^J_{o,n}$ (instead of a unique $\phi_{o,n}$) satisfying similar assumptions and converging to $\phi^1_o,\ldots,\phi^J_o$ respectively, then the deficiency will be a sum of J terms, obtained by replacing ϕ_o by $\phi^1_o,\ldots,\phi^J_o$ in (3.12).

In case the prior is a bivariate normal with the mean vector 0, variances equal to σ^2 and correlation coefficient $\rho \neq 0$, one can directly work with the Bayes test (instead of the approximate Bayes test) and easily verify that

(a) the critical region of the Bayes test is

$$\|\bar{X}\|^2 + 2\rho \, (n\sigma^2(1-\rho^2) +1)^{-1} \, \bar{X}^{(1)} \, \bar{X}^{(2)} > k_{2,n} \, ;$$

(b) $J = 2$, for all $n \geq 1$, $\phi^1_{o,n} = \pi/4$ or $3\pi/4$ and $\phi^2_{o,n} = 5\pi/4$ or $7\pi/4$ according as ρ is positive or negative and

$$r_{o,n} = \frac{k_{2,n}}{1 + |\rho|\ (n\sigma^2(1-\rho^2)+1)^{-1}} ;$$

(c) the size of the Bayes test is

$$\frac{\exp(-n\ r_{o,n}/2)\ \sigma(1-\rho^2)^{1/2}}{||\theta||\ (\pi|\rho|)^{1/2}} ;$$

(d) the deficiency (3.12) is as given in (1.8).

A numerical investigation, where we tabulated the deficiency (1.8) for $\sigma^2 = 1$, $\rho = \pm .8, \pm .6, \pm .4, \pm .2$ and $\theta^{(1)}, \theta^{(2)}$ $\pm 3.0, \pm 2.0, \pm 1.5, \pm 1.0, \pm 0.5$, shows that the deficiency lies between (2.5, -4.2), (0.8, -1.7), (0.6, -1.6) and (0.3, -2.7) respectively for the above values of ρ. Thus there is no significant difference between the Bayes test and the likelihood ratio test ; when $\rho = 0$, these two tests are identical. This may be interpreted as a sort of closeness of the likelihood ratio test statistic to the Bayes test statistic. Our computations show that the deficiency is negative for most pairs of $\theta^{(1)}$ and $\theta^{(2)}$; in this connection one may note that the deficiency is indeed negative for all θ such that $||\theta||^2 < 1/\pi b$ where

$b = \sigma^{-2}|\rho|\ (1-\rho^2)^{-1}$ and so $b \to \infty$ as $|\rho| \to 1$ and $b \to 0$ as $\rho \to 0$.

So far we have assumed that the dispersion matrix of the population of the observations is identity. If we now let

$$\Sigma = \begin{pmatrix} \sigma^{*2} & \sigma^{*2}\rho^* \\ \sigma^{*2}\rho^* & \sigma^{*2} \end{pmatrix} \qquad 0 < \sigma^* < \infty$$

be the dispersion matrix (σ^* and ρ^* are known), then applying first an orthogonal transformation $Y = AX$ on the samples so that the dispersion matrix of Y becomes identity, one can easily verify that the deficiency is still given by (3.12) except that the prior π is replaced by $\pi(\theta^*)/\sigma^{*2}(1-\rho^{*2})^{1/2}$ where

$$\theta^{*(1)} = 2^{-1/2}\sigma^*\{(1+\rho^*)^{1/2}\theta^{(1)} + (1-\rho^*)^{1/2}\theta^{(2)}\}$$

$$\theta^{*(2)} = 2^{-1/2}\sigma^*\{(1+\rho^*)^{1/2}\theta^{(1)} - (1-\rho^*)^{1/2}\theta^{(2)}\}.$$

(Think of θ as $E(Y)$.)

For instance in the above special case, one can verify the following statements :

(a) the critical region of the exact Bayes test is

$$\|\bar{Y}\|^2 - \frac{(\bar{Y}^{(1)})^2}{n\sigma_1^2+1} - \frac{(\bar{Y}^{(2)})^2}{n\sigma_2^2+1} \geq k_{2,n}$$

where

$$\sigma_1^2 = \sigma^2(1+\rho)/\sigma^{*2}(1+\rho^*)$$

$$\sigma_2^2 = \sigma^2(1-\rho)/\sigma^{*2}(1-\rho^*) \; ; \qquad \rho \neq \rho^*$$

(b) $r_{o,n} = k_{2,n}/(1 - \frac{1}{n\,\sigma_1^2})$ or $k_{2,n}/(1 - \frac{1}{n\,\sigma_2^2})$ according as $\sigma_1^2 > \sigma_2^2$ or $\sigma_1^2 < \sigma_2^2$;

(c) the size of the Bayes test is

$$(2/\pi)^{1/2} \frac{\exp(-n\, r_{o,n}/2)}{\|\theta\| \, (|\sigma_2^{-2} - \sigma_1^{-2}|)^{1/2}} \; ;$$

(d) the deficiency is

$$\frac{1}{\|\theta\|^2} \{ -\alpha \, |\sigma_2^{-2} - \sigma_1^{-2}| + \log(\tfrac{\pi}{2} \|\theta\|^2 \, |\sigma_2^{-2} - \sigma_1^{-2}|) \} , \tag{3.13}$$

where $\alpha = (\theta^{*(2)})^2$ or $(\theta^{*(1)})^2$ according as $\sigma_1^2 > \sigma_2^2$ or $\sigma_1^2 < \sigma_2^2$.

Note that if $\sigma^* = 1$ and $\rho^* = 0$, then the expression (3.13) agrees with (1.8) as it should be (observe that in this case, $|\sigma_2^{-2} - \sigma_1^{-2}| = 2b$).

For priors whose densities are not positive everywhere, the deficiency may be finite or infinity and in case it is finite, the value may considerably differ from (3.12). For example, if the prior is Lebesgue measure on the positive quadrant and is zero elsewhere, then it can be shown that the deficiency at θ (lying in the positive quadrant) is $-4 \log 2/\|\theta\|^2$, which is same as the negative of the deficiency of Section 2 (with $k = 2$). On the other hand if the prior is degenerate at θ, then

$$\alpha_{2,n} = \frac{\exp\{-\frac{1}{2}(n^{1/2}\|\theta\| + z_\beta)^2\}}{(2\pi n)^{1/2}\|\theta\|}(1 + o(1))$$

and so the deficiency at θ is ∞.

SECTION 4. BAYES TEST FOR THE EXPONENTIAL FAMILY

For the exponential family (1.9), assume that the natural parameter space $\bar{H}$ is open, $c(\theta)$ is strictly convex and that for some n_o, $\bar{X}_n$ lies in the set of possible expectations of the family almost surely (under θ_o) for all $n \geq n_o$. The last condition ensures that the maximum likelihood estimator $\hat{\theta} = \hat{\theta}(\bar{X}_n)$ given by the (unique) solution of the equation

$$c'(\hat{\theta}) = \bar{X}_n \tag{4.1}$$

is well-defined; here $c'(\theta)$ is the $k \times 1$ vector of the first order partial derivatives of $c(\theta)$. Assume without loss of generality that

(4.2) $$\theta_o = 0, \quad c(\theta_o) = 0, \quad c'(\theta_o) = 0.$$

Let $I(\theta\,;\,\theta')$ be the Kullback-Leibler information number of θ with respect to θ':

(4.3) $$I(\theta\,;\,\theta') = \langle \theta - \theta', c'(\theta) \rangle - c(\theta) + c(\theta').$$

Recall that $E_\theta(X) = c'(\theta)$ and the dispersion matrix under θ of X is $c''(\theta)$.

Consider now the Bayes test procedure as introduced in the introduction and observe that

$$T_{2,n} = \int \exp(n \langle \theta, \bar{X} \rangle - nc(\theta) + \log \pi(\theta))d\theta$$

$$= \exp(n \langle \hat{\theta}, \bar{X} \rangle - nc(\hat{\theta}))$$

$$\times \int \exp(n \langle \theta, \bar{X} \rangle - nc(\theta + \hat{\theta}) + nc(\hat{\theta})$$

$$+ \log\pi(\theta + \hat{\theta}))\, d\theta$$

$$= \exp(n\,\phi(\bar{X}) - \sigma(\hat{\theta}))\left\{\int_{0<t(\theta)\leq\delta} \exp(nt(\theta)\right.$$

$$\left. + \log \pi(\hat{\theta} + B^{-1}\theta))\, d\theta + O(\exp(-n\delta))\right\}.$$

Here B is a matrix such that $B^TB = c''(\hat{\theta})$, $\sigma(\theta)$ is as in (1.11) and

$$\phi(x) = \sup_\theta \{\langle \theta, x \rangle - c(\theta)\}$$

$$= \langle \hat{\theta}(x), x \rangle - c(\hat{\theta}(x))$$

(4.4) $$t(\theta) = \langle B^{-1}\theta, \bar{X} \rangle - c(\hat{\theta} + B^{-1}\theta) + c(\hat{\theta}).$$

One now applies the polar transformation and observes that the set $\{0 < t(\theta) < \delta\}$ lies between two sets of the form $\{\|\theta\|^2 \leq \delta'\}$, $\delta' = O(\delta)$; for this one uses the convexity of $t(\theta)$. Thus the identity (1.12) is established.

We therefore consider the approximate Bayes test which rejects H_o if and only if $\bar{X} \in S_n$:

$$S_n = \{x : \phi(x) - n^{-1} G(\hat{\theta}(x)) \geq k'_{2,n}\}$$

(4.5)

$$G(\theta) = \sigma(\theta) - \log \pi(\theta) .$$

The constant $k'_{2,n}$ is determined so that the power at θ is $\beta + o(n^{-1/2})$. It follows from Theorem 2 of Bhattacharya and Ghosh (see also Theorem 2.1 of Chapter Two) that under θ, $\phi(\bar{X})$ has an Edgeworth expansion in powers of $n^{-1/2}$. Expanding ϕ, one gets, as in Sections 2 and 3, a constant d (free from n) such that

$$P_\theta \{ n^{1/2}(< \bar{X} - E_\theta(\bar{X}), \theta > + < \hat{\theta} - \theta, \bar{X} - E_\theta(\bar{X}) >$$

$$- \frac{1}{2} < \hat{\theta} - \theta, c''(\theta) (\hat{\theta} - \theta) >) > z_\beta + n^{-1/2} d\}$$

$$= \beta + o(n^{-1/2}).$$

It also follows that

$$k'_{2,n} = I(\theta;\theta_o) + (< \theta, c''(\theta)\theta >)^{1/2} (z_\beta n^{-1/2} + d n^{-1})$$

(4.6)

$$-(\log \pi(\theta) - \sigma(\theta))n^{-1} + o(n^{-1}) ,$$

To evaluate the size $\alpha'_{2,n}$ of the approximate test, we assume as in Woodroofe (1978) that there exists an integer n_1 such that the vector of sample totals has a bounded continuous density for all $n \geq n_1$; this condition is needed so that the local limit theorem of Borovkov and Rogozin (1965) is valid. Let

$$(4.7) \qquad \phi_{min} = \inf\{\phi(x) : \phi(x) - n^{-1}G_1(x) \geq k'_{2,n}\},$$

where the subscript 1 in G_1 refers to the composition of G with θ as a function of $\bar{X}$ (i.e., $G_1(x) = G(\theta(x))$); similar conventions will be used below for other functions also. Now using the techniques of Borovkov and Rogozin (1965) (see also Proposition 1, Section 2 of Woodroofe (1978)), we get

$$(4.8) \qquad \alpha'_{2,n} = (n/2\pi)^{k/2}\left\{\int_{A_{n,\delta}} \exp(-n\,\phi(x) - \sigma_1(x))dx + O(\exp(-n(\phi_{min} + \delta)))\right\}(1 + o(1)),$$

where

$$(4.9) \qquad A_{n,\delta} = \{x \mid \phi(x) - n^{-1}G_1(x) \geq k'_{2,n},\ 0 \leq \phi(x) - \phi_{min} \leq \delta\}.$$

Note that $A_{n,\delta}$ is compact and so on it $\sigma_1(x)$ is bounded above.

To evaluate the integral on the right side of (4.8), we need to make a few assumptions. The following remark explains why they are plausible. Let

(4.10) $$M_n = \{x \in S_n \mid \phi(x) = \phi_{min}\}.$$

REMARK 4.1 Since ϕ is strictly convex and its global minimum is attained outside S_n, M_n is a subset of the boundary of S_n. Using Lagrangian multipliers, one expects that M_n may alternatively be obtained from the dual problem of minimising G_1 subject to the condition that $\phi = \phi_{min}$. Since ϕ_{min} converges to $I(\theta ; \theta_o)$ as $n \to \infty$, one can consider also the problem of minimising G_1 subject to the condition that $\phi = I(\theta ; \theta_o)$. Let M^* be the set of points where this restricted minimum of G_1 is attained. One would then expect that M_n will "converge" to a (unique nonempty) subset M of M^*; since ϕ is strictly convex, M_n and M^* are (nonempty and) compact. We assume below that M is also compact. Let M be a $(k-r-1)$-dimensional submanifold of the $(k-1)$-dimensional manifold $N = \{x \mid \phi(x) = I(\theta ; \theta_o)\}$ (in Section 2, M_n and M are zero-dimensional). Get a finite open cover $U_1,\ldots,U_m$ of M satisfying the following conditions: for a fixed U_i, there exists a coordinate system $\eta_i(x) = (\eta_i^{(1)}(x),\ldots,\eta_i^{(k)}(x))$ such that

(a) $x \to \eta_i(x)$ is a diffeomorphism on U_i;

(b) $\eta_i^{(k)} = \phi$; and

(c) $\eta_i^{(1)},\ldots,\eta_i^{(k-1)}$ are local coordinates for the manifold N and $M \cap U_i = \{x \in N \cap U_i \mid \eta^{(1)}(x)=\ldots=\eta^{(k-1)}(x)=0\}$

In view of Remark 4.3 below, we may assume without loss of generality that there is only one such coordinate system η which is a diffeomorphism on a neighbourhood V of M. Put

(4.11) $$\eta^1 = (\eta^{(1)},\ldots,\eta^{(k-r-1)}) \text{ and } \eta^2 = (\eta^{(k-r)},\ldots,\eta^{(k-1)}).$$

Since M_n converges to M, it seems reasonable to assume that for all sufficiently large n, the dimension of M_n is also $(k-r-1)$ and that it is in fact of the form

(4.12) $$M_n = \{x \in V \mid \eta^{(k)}(x) = \phi_{min},\ \eta^2(x) = C_n(\eta^1(x),\ \eta^{(k)}(x))\}$$

where

(4.13) $C_n(\eta^1(x),\eta^{(k)}(x)) \in R^r$ tends to zero uniformly in V.

Then on M, the partial derivatives of G_1 with respect to $\eta^{(1)},\ldots,\eta^{(k-1)}$ are zero and the corresponding $(k-1) \times (k-1)$ matrix of the second-order partial derivatives of G_1 is positive semidefinite. Let G_1^* denote the Hessian of G_1 with respect to η^2 i.e.,

(4.14) $$G_1^* = ((D^iD^jG_1))_{(k-r) \leq i,j \leq (k-1)}$$

We assume below that G_1^* is positive definite on V, and hence on M_n for all sufficiently large n.

We now state these assumptions more formally. Assume that on $A_{n,\delta}$ (see (4.9)), there exists a one-to-one thrice continuously differentiable transformation $x \rightarrow \eta(x)$ such that

(a) $\eta^{(k)} = \phi$;

(b) equation (4.12) and (4.13) hold for all sufficiently large n ;

(c) the Hessian $G_1^*(\eta^1, G_n(\eta^1, \eta^{(k)}), \phi_{min})$ of G_1 with respect to η^2 is positive definite for all η ;

(d) the elements of the Jacobian matrix are bounded.

Here and in the following we use the same notations σ_1, G_1 etc., even after a change of variables. (The assumption that there exists a set M to which M_n converges will be made later (vide (4.20)).

If I_n denotes the integral on the right side of (4.8), then I_n can be written as

$$\int_{\phi_{min}}^{\phi_{min}+\delta} \exp(-n\,\eta^{(k)}) \int_{D_{n,\delta}(\eta^{(k)})} \left(\int_{B_{n,\delta}(\eta^1,\eta^{(k)})} \exp(-\sigma_1(\eta))\, J(\eta)\, d\eta^2 \right) d\eta^1 d\eta^{(k)} ; \tag{4.15}$$

here $J(\eta)$ is the Jacobian of the transformation and

$$D_{n,\delta}(\eta^{(k)}) = \text{the projection of } A_{n,\delta}(\eta^{(k)}) \text{ to } \eta^1 \text{ space,}$$

(4.16) $$B_{n,\delta}(\eta^1,\eta^{(k)}) = \text{the section at } \eta^1 \text{ of } A_{n,\delta}(\eta^{(k)}),$$

$$A_{n,\delta}(\eta^{(k)}) = \text{the section at } \eta^{(k)} \text{ of the image } \eta(A_{n,\delta}) \text{ of } A_{n,\delta} \text{ under } \eta .$$

For any $\eta \varepsilon \eta(A_{n,\delta})$ one gets, by expanding G_1 around $(\eta^1, C_n(\eta^1, \eta^{(k)}), \phi_{min}) \varepsilon \eta(M_n)$,

(4.17) $$\phi(x) - n^{-1} G_1(x)$$
$$= k'_{2,n} + (\eta^{(k)} - \phi_{min})(1 - n^{-1} D^k G_1 \cdot (1 + o(1)))$$
$$-(2n)^{-1} < \eta^2 - C_n(\eta^1,\eta^{(k)}), (G_1^* + o(1))(\eta^2 - C_n(\eta^1, \eta^{(k)}) > ,$$

where the $o(1)$ terms go to zero uniformly in n <u>and</u> η^1 as $\delta \to 0$, all the derivatives of G_1 are evaluated at $(\eta^1, C_n(\eta^1, \eta^{(k)}), \phi_{min})$ and G_1^* is defined in (4.14). (Here we have used the fact that $D^i G_1 = 0$ for $i = (k-r), \dots, (k-1)$, that $\phi(x)$ attains its infimum on the boundary of S_n and that the following inclusion holds because of Assumption (d) :

(4.18) $$\eta(A_{n,\delta}) \subset sp(\eta(M_n); \delta'), \quad \delta' = O(\delta), \ \delta' \text{ free from } n)$$

By Assumption (c) and (4.17) one therefore gets

$$\int_{B_{n,\delta}(\eta^1,\eta^{(k)})} \exp(-\sigma_1(\eta))J(\eta)d\eta^2$$

$$= (1+o(1)) \times \int_{<\eta^2-c_n(\eta^1,\eta^{(k)}),G_1^*\cdot(\eta^2-c_n(\eta^1,\eta^{(k)}))> \leq a_n} \exp(-\sigma_1)J$$

(4.19)
$$= \frac{\exp(-\sigma_1)J}{(\det(G_1^*))^{1/2}} \cdot \frac{(\pi a_n)^{r/2}}{\Gamma(r/2+1)} (1 + o(1)),$$

where

$$a_n = 2n(\eta^{(k)} - \phi_{min})(1 - n^{-1} D^k G_1)$$

and all the functions are evaluated at $(\eta^1, c_n(\eta^1, \eta^{(k)}), \phi_{min})$.

Now we note that

$$D_{n,\delta}(\phi_{min}) = \{\eta^1 \mid (\eta^1, c_n(\eta^1,\eta^{(k)}), \phi_{min}) \in \eta(M_n)\}.$$

We shall denote the set on the right side by $\eta^1(M_n)$. Let M be a nonempty compact subset of R^k such that as $n \to \infty$

(4.20) the Hausdorff distance of M_n and $M \to 0$

and

(4.21) $\lambda(\eta^1(M_n) + \eta^1(M)) \to 0$,

where λ is $(k-r-1)$-dimensional Lebesgue measure, $+$ denotes symmetric difference and $\eta^1(M)$ is the projection of M to η^1 space (for the definition of the Hausdorff distance, see page 19, Bhattacharya and Ranga Rao (1976)). We note that the strict convexity of ϕ, the fact that the origin lies outside S_n and Assumption (d) imply that

(4.22) the Hausdorff distance of $D_{n,\delta}(\eta^{(k)})$ and $D_{n,\delta}(\phi_{min})$ $\to 0$ as $\delta \to 0$ uniformly in n and $\eta^{(k)}$.

We also assume that as $\delta \to 0$

(4.23) $\lambda(D_{n,\delta}(\eta^{(k)}) + D_{n,\delta}(\phi_{min})) \to 0$ uniformly in n and $\eta^{(k)}$.

Note that under (4.20) and (4.22), Assumption (c) can be deduced (using compactness of $\eta(M)$) from Assumption (c') below:

(c') the Hessian $G_1^*(\eta^1, 0, I(\theta ; \theta_0))$ is positive definite for all $\eta^1 \varepsilon \eta^1(M_n)$.

We now replace Assumption (c) by (stronger) Assumption (c'). We finally assume that

(e) the Jacobian determinant of the transformation $\eta(x)$ is positive and bounded away from zero on $\eta(M)$.

Using (4.22) and (4.23) and equation (4.19), the integral I_n can be reduced to

$$n^{-1}(2\pi)^{r/2} \exp(-n\,\phi_{min})\,(1+o(1)). \int_{\eta^1(M_n)} \frac{\exp(-\sigma_1)J}{(\det G_1^*)^{1/2}}\, d\eta^1 .$$

Finally using (4.20) and (4.21) we get

$$\alpha'_{2,n} = n^{k/2-1} (2\pi)^{-(k-r)/2} \exp(-n\, \phi_{min}) \; (1+o(1))$$

$$\times \int_{\eta^1(M)} \frac{\exp(-\sigma_1) J}{(\det(G_1^*))^{1/2}} \, d\eta^1 , \tag{4.24}$$

where in the integrand all functions are evaluated at $(\eta^1, 0, I(\theta;\theta_0))$. It can now be seen as in Section 3 that the size of the Bayes test has the same asymptotic expansion as that of $\alpha'_{2,n}$.

REMARK 4.2 Let $C_{n,\eta}(k)$ and C be nonempty compact subsets of R^{k-r-1} such that as $n \to \infty$,

(i) the Hausdorff distance of $C_{n,\eta}(k)$ and C tends to zero uniformly in $\eta^{(k)}$;

(ii) the Hausdorff distance of the boundaries of $C_{n,\eta}(k)$ and C tends to zero uniformly in $\eta^{(k)}$;

and

(iii) $\lambda(bd(C)) = 0$,

λ being (k-r-1)-dimensional Lebesgue measure. Then one has

$$\lambda(C_{n,\eta}(k) + C) \to 0 \qquad n \to \infty ,$$

uniformly in $\eta^{(k)}$. This result may be used to check conditions like (4.21) and (4.23).

For a proof, fix $a > 0$ and get an $\varepsilon > 0$ such that

$$\lambda(sp(C ; \varepsilon) - C) < a/2 , \quad \lambda(sp(bd(C); 2\varepsilon)) < a/2 .$$

Choose $N \geq 1$ such that $n \geq N$ implies

$$d(bd(C_{n,\eta}(k)), \ bd(C)) < \varepsilon/2 , \quad d(C_{n,\eta}(k) , C) < \varepsilon/2 .$$

Here d stands for the Hausdorff distance. Now if

$d(C_{n,\eta}(k), C) < \varepsilon$,

$$\begin{aligned} \lambda(C_{n,\eta}(k) + C) &\leq \lambda(sp(C_{n,\eta}(k); \varepsilon) - C_{n,\eta}(k)) \\ &\quad + \lambda(sp(C ; \varepsilon) - C) \qquad (4.25) \\ &\leq \lambda(sp(C_{n,\eta}(k); \varepsilon) - C_{n,\eta}(k)) + a/2 . \end{aligned}$$

We next observe that for any $n \geq 1$

$$sp(C_{n,\eta}(k) ; \varepsilon) - C_{n,\eta}(k) \subset sp(bd(C_{n,\eta}(k)); \varepsilon) ; \qquad (4.26)$$

(to see this, let x be an element of the set on the left side; as x lies outside $C_{n,\eta}(k)$ <u>and</u> $C_{n,\eta}(k)$ is closed, there exists $y \in bd(C_{n,\eta}(k))$ such that $d(x,y) = d(x, C_{n,\eta}(k))$; as $x \in sp(C_{n,\eta}(k) ; \varepsilon)$, one can then conclude that $d(x,y) < \varepsilon$ and hence that (4.26) is true). Thus if $n \geq N$

$$\lambda(sp(C_{n,\eta}(k); \varepsilon) - C_{n,\eta}(k)) \leq \lambda(sp(bd(C); 2\varepsilon)) < a/2$$

The last fact and (4.25) together complete the proof.

REMARK 4.3 Suppose our assumptions do not hold for S_n but there exists a finite open cover $U_1, \ldots, U_m$ of $\{\phi(x) \leq I(\theta ; \theta_o) + \delta\}$ and the assumptions are true for $S_{n,i} = S_n \cap U_i$. Then we can write S_n as a finite disjoint union of sets $S'_{n,j}$ for each of which our assumptions hold. In this case, the final result (4.24) remains true.

REMARK 4.4 If instead of Assumption (e), we assume that as $\eta^2 \to 0$,

$$J(\eta^1, \eta^2, \eta^{(k)}) / J_1(\eta^1, \eta^{(k)}) \; \psi(\eta^2) \to 1$$

where $\psi(c\,\eta^2) = c^{\gamma}\,\psi(\eta^2)$ for some γ, $\gamma > 0$, then (4.24) holds with a new integrand.

EXAMPLE : Consider a tri-variate normal population with mean $\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)})$ and dispersion matrix identity. Let the prior density be $\pi(\theta) = \exp\{-a(\theta^{(3)})^2/2\}$, $a > 0$. The critical region of the Bayes test is $\{\|\bar{X}\|^2 - b_n(\bar{X}^{(3)})^2 \geq k_{2,n}\}$ where

$$b_n = (n+a-1)a/2(n(n+a)),$$

$$k_{2,n} = \|\theta\|^2 + 2\|\theta\| \, (z_\beta \, n^{-1/2} + dn^{-1}) + o(n^{-1}),$$

d being a suitable constant. The size of the test is

$$\alpha_{2,n} = (n/2\pi)^{3/2} \int \exp(-nr/2) \frac{r^{1/2}}{2} \sin\phi_1 \, d\phi_1 \, d\phi_2 \, dr,$$

the integral being taken over the region

$$r(1-b_n \sin^2 \phi_1) \geq k_{2,n}, \; 0 < r < \infty, \; 0 < \phi_1 < \pi, \; 0 < \phi_2 < 2\pi .$$

Here $M_n = \{(\phi_1, \phi_2) : \phi_1 = 0 \text{ or } \phi_2 = \pi\}$ and is free from n; the dimension of $M_n = 1$; $r_{o,n}$, the smallest r in the above region, is $k_{2,n}$ and $\gamma = 1$ (see Remark 4.4). One can verify that

$$\alpha_{2,n} = \frac{2 \exp(-nk_{2,n}/2) \, n^{1/2}}{(2\pi \, k_{2,n})^{1/2}} \quad (1+o(1)).$$

REMARK 4.5 To see how our results are related to Theorem 1 of Woodroofe (1978), consider instead of S_n a set

$$S = \{x \mid \phi(x) - G(x) \geq k\}, \quad k < G(0)$$

where G is a thrice continuously differentiable real valued function. Define, in analogy with (4.7) and (4.10),

$$\phi_{min} = \inf \{\phi(x) \mid x \in S\}, \quad M = \{x \in S \mid \phi(x) = \phi_{min}\}$$

and suppose that there exists a coordinate system $\eta(x)$ with properties analogous to (a), (b), (c), (d) and (e); in particular $\eta^2 = 0$ on M. Assume furthermore that

$$1 - D^k G > 0 \quad \text{on} \quad M . \tag{4.27}$$

It can then be shown that

$$P_{\theta_o}(S) = n^{(k-r)/2-1} (2\pi)^{(k-r)/2} \exp(-n\,\phi_{min})(1+o(1))$$

$$\times \int_{\eta(M)} \frac{J \exp(-\sigma)\, (1 - D^k G)^{r/2}}{(\det G^*)^{1/2}}\, d\eta^1 , \tag{4.28}$$

where (k-r-1) is the dimension of M and in the integrand all functions are evaluated at $(\eta^1, 0, \phi_{min})$; the definitions of σ, J etc. should be obvious.

In Theorem 1 of Woodroofe, G is his ϕ_o and one can take η^2 as his ω_o (here $D^k G = 0$ on M).

The assumptions made here can be relaxed as in Remark 4.3. In fact both of (4.24) and (4.28) can be suitably modified to cover cases where M is a finite disjoint union of manifolds of different dimensions.

REFERENCES

Albers, W. (1974) : Asymptotic Expansions and the Deficiency Concept of Statistics. Mathematical Centre Tracts, 58 Amsterdam.

Bahadur, R.R. (1960) : Stochastic comparison of tests, Ann. Math. Stat., 31, 276-295.

________ (1965) : An optimal property of the likelihood ratio statistic, Proc. Fifth Berkeley Symp., 1, Univ. of California Press.

________ (1967) : Rates of convergence of estimates and test statistics, Ann. Math. Stat., 38, 303-324.

________ (1971) : Some Limit Theorems in Statistics. SIAM, Philadelphia.

Bahadur, R.R. and Ranga Rao, R. (1960) : On deviations of the sample mean, Ann. Math. Stat., 31, 1015-1027.

Bahadur, R.R. and Zabell, S.L. (1979) : Large deviations of the sample mean in general vector spaces, Ann. Probability, 7, 587-621.

Bártfai, P. (1977) : On the multivariate Chernoff theorem, Preprint, Budapest.

Bhattacharya, R.N. (1977) : Refinements of the multivariate central limit theorem and applications, Ann. Probability, 5, 1-27.

Bhattacharya, R.N. and Ghosh, J.K. (1978) : On the validity of the formal Edgeworth expansions, Ann. Stat., 6, 434-451.

Bhattacharya, R.N. and Ranga Rao, R. (1976) : Normal Approximation and Asymptotic Expansions. Wiley, New York.

Bickel, P.J. (1974) : Edgeworth expansions in nonparametric statistics, Ann. Stat., 2, 1-20.

Borovkov, A.A. and Rogozin, B.A. (1965) : On the multidimensional central limit theorem, Theor. Probability Appl., 10, 55-62.

Box, G.E.P. (1949) : A general distribution theory for a class of likelihood criteria, Biometrika, 36, 317-346.

Chandra, T.K. and Ghosh, J.K. (1978) : Comparison of tests with same Bahadur-efficiency, Sankhyā, Ser.A, 40, 253-277.

__________ (1979) : Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables, Sankhyā, Ser. A, 41, 22-47.

__________ (1980a): Valid asymptotic expansions for the likelihood ratio and other statistics under contiguous alternatives, Sankhyā, Ser. A, 42, to appear.

__________ (1980b): Deficiency for multiparameter testing problems, Sixtieth Birthday Volume of C.R.Rao, to appear.

Cochran, W.G. (1952) : The χ^2-goodness of fit. Ann. Math. Stat., 23, 493-507.

Cramér, H. (1937) : Random Variables and Probability Distributions. Cambridge Tracts in Math., Cambridge.

Cramér, H. (1946) : Mathematical Methods of Statistics. Princeton University Press, Princeton.

Dieudonné, J. (1960) : Foundations of Modern Analysis. Academic Press, New York.

Edwards, C.H. Jr. (1973) : Advanced Calculus of Several Variables. Academic Press, New York.

Feller, W. (1960) : An Introduction to Probability Theory and Its Applications. Vol.1, Wiley, New York.

________ (1968) : An Introduction to Probability Theory and Its Applications. Vol.2, Wiley, New York.

Ghosh, J.K. (1980) : Review of the book "Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families", by W.C.M. Kallenberg, Sankhyā, to appear.

Ghosh, J.K., Sinha, B.K. and Subramanyam, K. (1979): Edgeworth expansions for Fisher-consistent estimators and second order efficiency, Tech. Report No. 7/79, Indian Statistical Institute.

Hayakawa, T. (1977) : The likelihood ratio criterion and the asymptotic expansion of its distribution, Ann. Inst. Statist. Math., 29, Part A, 359-378.

________ (1975) : The likelihood ratio criterion for a composite hypothesis under a local alternative, Biometrika, 62, 451-460.

Hodges, J.L. (Jr.) and Lehmann, E.L. (1970) : Deficiency, Ann. Math. Stat., 41, 783-801.

Kallenberg, W.C.M. (1977) : Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families. Mathematical Centre Tracts, 78, Amsterdam.

Korin, B.P. (1968) : On the distribution of a statistic used for testing a covariance matrix, Biometrika, 55, 171-178.

Krishnaiah, P.R. (1978) : Some recent developments on real multivariate distributions, Developments in Statistics, Vol. 1, 135-169.

Kuratowski, K. and Mostowski, A. (1976) : Set Theory. North-Holland Publishing Company, Amsterdam.

LeCam, L. (1974) : Notes on Asymptotic Methods in Statistical Decision Theory. Centre de Recherches Mathématiques, Université de Montréal.

Pfanzagl, J. (1980) : Asymptotic expansions in parametric decision theory, Developments in Statistics, 1-138, to appear.

Peers, H.W. (1971) : Likelihood ratio and associated test criteria, Biometrika, 58, 577-587.

Raghavachari, M. (1970) : On a theorem of Bahadur on the rate of convergence of test statistics, Ann. Math. Stat., 41, 1695-1699.

Rao, C.R. (1965) : Linear Statistical Inference and Its Applications. Wiley, New York.

Schwarz, G. (1978) : Estimating the dimension of a model, Ann. Stat., 6, 461-464.

Sweeting, T.J. (1977) : Speeds of convergence for the multi-dimensional central limit theorem, Ann. Probability, 5, 28-41.

Woodroofe, M. (1979) : Large deviations of likelihood ratio statistics with applications to sequential analysis, Ann. Stat., 6, 72-84.