Sankhyā : The Indian Journal of Statistics
1979, Volume 41, Series A, Pts. 1 and 2, pp. 22–47.

VALID ASYMPTOTIC EXPANSIONS FOR THE LIKELIHOOD RATIO STATISTIC AND OTHER PERTURBED CHI-SQUARE VARIABLES

By TAPAS K. CHANDRA and J. K. GHOSH
Indian Statistical Institute

SUMMARY. Let $\{Z_n\}_{n\geqslant 1}$ be a sequence of random vectors. Under certain conditions, distributions of statistics which are smooth functions of the mean vector $\bar{Z}_n$ and whose asymptotic distributions are central Chi-square are shown to possess asymptotic expansions in powers of n^{-1}. As applications, asymptotic expansions of the null distributions of the likelihood ratio statistic, Wald's and Rao's statistics are obtained. The results proved here supplement the recent work of Bhattacharya and Ghosh (1978) and also justify the validity of the formal expansions obtained by Box (1949) and Hayakawa (1977).

1. Introduction

Following Bhattacharya and Ghosh (1978)—hereafter abbreviated as BG—we consider a random variable (r.v.) $H(\bar{Z}_n)$ where $\bar{Z}_n = n^{-1}\sum_1^n \bar{Z}_i$, $\{\bar{Z}_n\}_{n\geqslant 1}$ is a sequence of independent and identically distributed (i.i.d.) k-dimensional random vectors with mean vector $\mu = E(Z_1)$ and *nonsingular dispersion matrix* $V = E(Z_1-\mu)(Z_1-\mu)^T$ and H is a real-valued measurable function on R^k. If H is continuously differentiable at μ and the vector l of the first-order partial derivatives of H at μ is non-null, then it is well-known that $\sqrt{n}(H(\bar{Z}_n)-H(\mu))$ is asymptotically normal with mean zero and variance $l^T V l$. Under much stronger conditions on H and Z_1, BG improved this result considerably by obtaining an asymptotic expansion for the distribution function of $\sqrt{n}(H(\bar{Z}_n)-H(\mu))$; see in this connection their Theorem 2 and Remark 1.1. We shall partially supplement this result by considering the case where l is the null vector and

$$W_n = 2n(H(\bar{Z}_n)-H(\mu)) \qquad \dots \quad (1.1)$$

is asymptotically distributed as a (central) χ^2. *A statistic of this sort will be called a perturbed* χ^2.

Let L be the matrix of second-order partial derivatives of H at μ. If the second-order partial derivatives of H are continuous in a neighbourhood of μ, then a necessary and sufficient condition for W_n to be asymptotically χ^2 is that l is the null vector and $L^T V L = L$ (see Rao, 1965, page 152). *Assume* then that H is sufficiently smooth (i.e. that H has enough continuous derivatives in a neighbourhood of μ) and that the above necessary and sufficient condition holds. Under an additional technical condition we show the distribution function of W_n can be expanded asymptotically in a series of χ^2-integrals. For a precise statement see Theorem 1. Note that if L is non-singular the technical condition holds but Example 2.1 shows in general this condition cannot be relaxed. In this connection, see also Remark 2.8.

As an application of our main result, we consider the likelihood ratio (LR) statistic Λ_n and the *transformed* LR *statistic* $-2 \log \Lambda_n = \lambda_n$, say. It is well-known that under certain conditions the statistic λ_n is asymptotically distributed as a χ^2 variable. Various people have sought to improve this approximation by *formally* expanding the distribution function of λ_n. One of the earliest references is Box (1949) and a recent one is Hayakawa (1977) where further references can be found. Most of these expansions are obtained formally either by inverting an approximate characteristic function (c.f.) or by equating the first few moments of the exact and approximating distribution functions. In general, formal expansions obtained this way are not valid as asymptotic expansions in the sense of Bickel (1974). Consequently the validity of the formal expansions in the literature has remained an interesting open problem. Using our Theorems the formal expansions are justified in Sections 3 and 4. In Section 3 we prove that if we have an absolutely continuous (with respect to Lebesgue measure on R^{p+q}) exponential density with natural parameters $\theta_1, \ldots, \theta_{p+q}$ and we are testing $H_0(\theta_1 = \ldots = \theta_p = 0)$, then the formal expansion of $P(\lambda_n \epsilon B)$ (B is a Borel set) is valid uniformly over all Borel sets. It is noted that this justifies the expansion of Box (1949). In Section 4 we establish the validity of the expansion for $P(\lambda_n < u)$ under very general conditions; the details of the proof are omitted since they are tiresome and routine. Under the conditions of this section Hayakawa's expansion can be justified. Some other applications are considered in Section 5.

Results concerning the expansions of perturbed χ^2 under contiguous alternatives have been obtained and will be published elsewhere.

2. MAIN THEOREM AND RELATED RESULTS

We continue with the notations of Section 1. Denote the partial derivatives of H at μ by

$$l_{i_1 i_2 \ldots i_p} = (D^{i_1} D^{i_2} \ldots D^{i_p} H)(\mu),\ 1 \leqslant i_1, \ldots, i_p \leqslant k,$$

where D^i stands for differentiation with respect to the i-th coordinate. Thus l is the vector $(l_1, \ldots, l_k)$ and L is the $k \times k$ matrix $((l_{ij}))$. *In this section, s always stands for some integer $\geqslant 4$; $\mathcal{B}$ will stand for the set of all Borel subsets of R^1; the symbols* $\|.\|$ and $<\,,\,>$ will denote *Euclidean norm* and *inner product* respectively. Let p be the rank of L $(0 \leqslant p \leqslant k)$. Without loss of generality assume that $\mu = 0$.

We now state an assumption needed for our main theorem.

Assumption A_s: (i) all the derivatives of H of order s and less are continuous in a neighbourhood of μ;

(ii) the vector l is null;

(iii) the matrix L is non-null and satisfies the equation $LVL = L$;

(iv) if under some nonsingular linear transformation $x = Az$, $x^T = (x^{(1)}, \ldots, x^{(k)})$, $z^T L z$ becomes a positive-definite quadratic form in $(x^{(1)}, \ldots, x^{(p)})$, then,

$$H_{s-1}(A^{-1}x) = \sum_{i,j=1}^{p} x^{(i)} x^{(j)} P_{ij}(x) \qquad \ldots \quad (2.1)$$

for some polynomials P_{ij} where $H_{s-1}(z)$ is the Taylor expansion around $\mu = 0$ of $H(z)$ up to and including the $(s-1)$-th order derivatives of H;

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

$$\text{(a)} \quad |H(A^{-1}x) - H_{s-1}(A^{-1}x)| \leqslant K_1 \|x^1\|^2 \|x\|^{s-2}$$

$$\text{(b)} \quad |D^i H(A^{-1}x) - D^i H_{s-1}(A^{-1}x)| \leqslant K_2 \|x^1\| \|x\|^{s-2} \quad 1 \leq i \leq p \qquad \ldots \quad (2.2)$$

where K_1 and K_2 are constants.

The condition (iv) is a technical one and ensures that the Taylor expansion of W_n, when expressed in terms of $(x - A\mu)$, is at least of degree two in the first p components. Similar interpretation can be given to the condition A_s(v).

Note that A_s(iv) and (v) are true if A_s(i) and (ii) hold and L is positive-definite.

Remark 2.1 : To check A_s(iv)-(A_s(v)), it is enough to check (2.1) for *one* non-singular matrix A which has the property that $z^T L z$ is a positive definite quadratic form in $(x^{(1)}, \ldots, x^{(p)})$.

The following lemma provides a *sufficient condition* for A_s(iv) and (v) for all s (with $A = I$ for notational convenience).

Lemma 2.1 : *If* A_s(i) *holds and if*

$$H(x) = 0, \quad D^i H(x) = 0, \quad 1 \leqslant i \leqslant p \qquad \ldots \quad (2.3)$$

for all $x = (0, x^2)^T$ *the first* p *components being zero and the remaining ones lying in a* $(k-p)$*-dimensional neighbourhood of* 0, *then*

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

for suitable polynomials $\{p_{ij}\}$;

If, moreover, H *is (real) analytic in a neighbourhood of* 0, *then there exists a neighbourhood of* 0, *in which the following two inequalities hold :*

$$|H(x) - H_{s-1}(x)| \leqslant k_1 \|x^1\|^2 \|x\|^{s-2}$$

and

$$|D^i H(x) - D^i H_{s-1}(x)| \leqslant k_2 \|x^1\| \, \|x\|^{s-2} \qquad 1 \leqslant i \leqslant p$$

where k_1 and k_2 constants.

The proof of the lemma is omitted.

A *sufficient condition* for A_s(v) which does not involve analyticity of H can be obtained from the following result : if A_s(i) holds and if there exists a neighbourhood of 0 in which

$$|D^\nu H(x)| \leqslant k_3 \|x^1\|^2 / \|x\|^2$$

whenever $|\nu| = s$ and $|\nu^1| = 0$ while

$$|D^\nu H(x)| \leqslant k_4 \|x^1\| / \|x\|$$

whenever $|\nu| = s$ and $|\nu^1| = 1$, then the conclusion of part (b) of the above lemma holds. Here K_3 and K_4 are suitable constants and if $\nu = (\nu^{(1)}, \ldots, \nu^{(k)})$ is a multi-index of nonnegative integers, then

$$D^\nu = (D^1)^{\nu(1)} \ldots (D^k)^{\nu(k)}$$

$$|\nu| = \nu^{(1)} + \ldots + \nu^{(k)}$$

$$|\nu^1| = \nu^{(1)} + \ldots + \nu^{(p)}$$

To see this, one expresses the remainder for $H-H_{s-1}$ and $D_i(H-H_{s-1})$ in the form given in Corollary 8.3, page 85 of Bhattacharya and Rao (1976).

We shall say that Z_1 satisfies *condition D* if there exist an m dimensional vector Y_1 and real-valued Borel measurable functions $f_1, \ldots, f_k$ such that

(i) $Z_1^{(i)} = f_i(Y_1)$, $1 \leqslant i \leqslant k$; and

(ii) the distribution of Y_1 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. ... (2.4)

The significance of condition D is clear from Lemma 2.2 of BG.

Theorem 1 : *Suppose that for some integer $s \geqslant 4$, H and Z_1 satisfy the assumptions A_s (i)—(iv) and that $E\|Z_1\|^{s-1}$ is finite.*

(a) *If in addition the assumption A_s (v) holds and Z_1 satisfies condition D, then there exist polynomials ψ_r (in one variable) whose coefficients do not depend on n, such that*

$$sup \left\{P|(W_n \epsilon B) - \int_B \psi_{m,p}| : B \epsilon \mathcal{B}\right\} = \epsilon_n \qquad \ldots (2.5)$$

where

$$\epsilon_n = o(n^{-m}) \quad \text{if } s = 2m+3;$$

$$= o(n^{-m-1}) \quad \text{if } s = 2m+4,$$

$$\psi_{m,n}(v) = f_{\chi^2,p}(v) \sum_{r=0}^{m} n^{-r}\psi_r(v) \quad (\psi_0 = 1) \qquad \ldots (2.6)$$

m is the greatest integer less than or equal to $(s-3)/2$. and $f_{\chi^2,p}$ is the density of a (central) χ^2-variable with p degrees of freedom (p = rank of L).

(b) *If Z_1 satisfies Cramer's condition*

$$\lim_{\|t\| \to \infty} sup \; |E(exp\{i < t, Z_1 > \})| < 1, \qquad \ldots (2.7)$$

then the conclusion of (a) holds if the left hand side of (2.5) is replaced by

$$\sup_{u \epsilon R^1} \{|P(W_n \leqslant u) - \int_{-\infty}^{u} \psi_{m,n}|\}$$

Remark 2.2 : It will follow from the proof of Theorem 1 that one does not need the full force of the i.i.d. structure of $\{Z_n\}_{n \geq 1}$. Thus in the definition of $W_n = 2n[H(\overline{Z}_n) - H(\mu)]$, the normalised deviation $n^{\frac{1}{2}}(\overline{Z}_n - \mu)$ based on some sequence of i.i.d. random vectors can be replaced by an arbitrary sequence $\{\overline{Z}_n\}_{n \geq 1}$ which has a similar multivariate Edgeworth expansion (see the paragraph preceding the proof of Theorem 1(a)) and which satisfies the equation (2.30) (or the equation (2.20)) of BG.

Remark 2.3 : In most applications, we shall use the extended version of Theorem 1 as indicated in the previous remark. We shall usually be supplied with the statistic W_n rather than the function H and the sequence of r.v.s. $\{\overline{Z}_n\}_{n \geq 1}$. One has to choose judiciously H and $\{\overline{Z}_n\}_{n \geq 1}$ such that the condition A_s(iv) (or A_s(v)) is satisfied. This problem of choice arises typically in most statistical applications including those considered in Sections 3, 4 and 5. See in this connection Example 2.2 and Remark 3.1.

Proof of Theorem 1 : Before proving Theorem 1 we make two remarks and fix some notations.

Remark 2.4 : Suppose that the assumption A_s holds. *We may (and do) assume that* $V = I$ (the identity matrix), L is a diagonal matrix whose first p diagonal elements are one and the rest are zero and relations (2.1) and (2.2) hold with $A = I$. For a proof, note that we can get hold of a nonsingular matrix R such that $R^T V^{-1} R = I$ and $R^T L R = S$ where S is a diagonal matrix (see Rao, 1965, page 37). In view of A_s(iii), S is also idempotent. Consequently we may assume without loss of generality that the first p diagonal elements of S are one and the rest are zero. Then $z^T L z$ under the transformation $x = R^{-1} z$ becomes $\sum_{i=1}^{p} (x^{(i)})^2$ and so (2.1) and (2.2) hold with $A = R^{-1}$ (see Remark 2.1). Now instead of the vector Z consider the vector $R^{-1}X$ and redefine H accordingly.

Remark 2.5 : Consider the transformation T_1 which sends $z^1 = (z^{(1)}, \ldots, z^{(p)})$ to $(r, \theta^{(1)}, \ldots, \theta^{(p-1)})$ by means of a polar transformation and keep $z^2 = (z^{(p+1)}, \ldots, z^{(k)})$ unaltered

$$z^{(1)} = r \prod_{i=1}^{p-1} \cos \theta^{(i)}$$

$$z^{(j)} = r \sin \theta^{(p-j+1)} \prod_{i=1}^{p-j} \cos \theta^{(i)} \qquad 2 \leqslant j \leqslant p$$

where

$$0 < r < \infty, \quad -\pi/2 < \theta^{(i)} < \pi/2 \qquad i = 1, 2, \ldots, p-2$$

and

$$0 \leqslant \theta^{(p-1)} \leqslant 2\pi.$$

An expression of the form $\prod_{i=1}^{k} (z^{(i)})^{\alpha_i}$ where α_i are nonnegative integers can be written as

$$r^{\alpha_0} \left(\prod_{i=1}^{p} (z^{(i)}/r)^{\alpha_i} \right) \left(\prod_{i=p+1}^{k} (z^{(i)})^{\alpha_i} \right)$$

$$= R(\alpha, r, \theta, z^2) = R(\alpha).$$

where $\alpha = (\alpha_0, \ldots, \alpha_k)$, $\alpha_0 = \sum_{i=1}^{p} \alpha_i$. The notation $R(\alpha)$ will be used even if $\alpha_0 \neq \sum_{1}^{p} \alpha_j$; we shall be concerned with those $R(\alpha)$'s for which α_0 will differ from $\sum_{i=1}^{p} \alpha_i$ by an even number.

Say that $R(\alpha)$ is *odd* if at least one of $\alpha_1, \ldots, \alpha_k$ is odd. Then the integral of $\exp \{-\frac{1}{2}\|Z^2\|^2\}$ $R(\alpha)$ with respect to (θ, z^2) is zero if $R(\alpha)$ is odd. More generally we say that the expression

$$r^{a_0} \left[\prod_{i=1}^{p-1} (\cos \theta^{(i)})^{a_i} (\sin \theta^{(i)})^{b_i} \right] \left[\prod_{i=p+1}^{k} (z^{(i)})^{c_i} \right]$$

is *odd* if at least one of $b_0, \ldots, b_{p-1}, a_{p-1}, c_{p+1}, \ldots, c_k$ is odd. Note that the Jacobian of the transformation T_1 is

$$r^{p-1} \prod_{i=1}^{p-2} (\cos \theta^{(i)})^{p-i-1} = r^{p-1} J(\theta) \quad \text{(say)}$$

and that if $R(\alpha)$ is odd, then $R(\alpha)J(\theta)$, is also odd.

Finally by $R_{ij}(r, \theta, z^2)$ we shall denote a finite sum of constant multiples of terms of the form $R(\alpha)$ and say that $R_{ij}(r, \theta\ z^2)$ is *odd* if every such $R(\alpha)$ is odd. One verifies that the various $R_{ij}(r, \theta, z^2)$ occuring in the proof of Theorem 1 have the property that $R_{ij}(r, \theta, z^2)$ is odd if j is odd and that if j is even and $R_{ij}(r, \theta, z^2)$ includes some $R(\alpha)$ (or constant multiple of $R(\alpha)$) which fails to be odd, then the power α_0 of r in that $R(\alpha)$ will be even.

A few notations will now be introduced. In view of Remark 2.4 we shall assume $v = I$, $z^T L z = \sum_{i=1}^{p} (z^{(i)})^2$ and (2.1) and (2.2) holds with $A = I$. Let

$$\xi_{s-1,n}(z) = [1+\sum_{r=1}^{s-3} n^{-r/2}\widetilde{P}_0(-D)]\phi(z)$$

be the *multivariate Edgeworth expansion* of $n^{\frac{1}{2}}(\overline{Z}_n-\mu)$ where ϕ is the normal density on R^k with mean zero and dispersion matrix I, $\{\widetilde{P}_r(i\,t)\}$ are the Cramer-Edgeworth polynomials (for a definition, see (2.1) of BG) and finally, $\widetilde{P}_r(-D)$ is the operator obtained by formally substituting $-D = (-D^1, \ldots, -D^k)$ for it in $\widetilde{P}_r(i\,t)$. We note that *the coefficient of $n^{-j/2}$ in $\xi_{s-1,n}(z)$ is $\phi(z)P_j(z)$ where $P_j(z)$ is a polynomial in z with the property that the degree of each term of $P_j(z)$ is even or odd according as j is even or odd.*

Let $g_n(z) = 2n\{H(\mu+n^{-\frac{1}{2}}z)-H(\mu)\}$ and $h_{s-1}(z)$ be a Taylor expansion of $g_n(z)$ i.e.,

$$h_{s-1}(z) = 2\sum_{j=2}^{s-1} \frac{n^{(j-2)/2}}{j\,!} \sum l_{i_1,\ldots,i_j} z^{(i_1)} z^{(i_2)} \ldots z^{(i_j)}$$

with $z = (z^{(1)}, \ldots, z^{k)})$.

Put

$$M_n = \{z : \|z\|^2 < (s-1) \log n\}.$$

For $B \subset R^1$, put $B_n = g_n^{-1}(B) \subset R^k$.

We shall first prove Theorem 1(a). By Lemma 2.2 of BG and in view of the fact that for all integers $q \geqslant 0$, the integral of $\|z\|^q \exp\{-\|z\|^2/2\}$ over the complement of M_n is $o(n^{-(s-2)/2})$, it is enough to exhibit $\psi_{m,n}$ of the form (2.6) such that

$$\sup\left\{\left|\int_{B_n \cap M_n} \xi_{s-1,n} - \int_B \psi_{m,n}\right| : B\epsilon\mathcal{B}\right\} = \epsilon_n. \qquad \ldots \quad (2.8)$$

Below we assume $p > 1$; the special case $p = 1$ needs only a slight modification of the following argument and is omitted.

Using the transformation T_1 of Remark 2.5 we get

$$\int_{B_n \cap M_n} \xi_{s-1,n} = \int_{T_1(B_n \cap M_n)} r^{p-1} J(\theta) \left[1+\sum_{j=1}^{s-3} n^{-j/2} R_{1j}(r, \theta, z^2)\right]$$
$$\times \exp\{-(r^2+\|z^2\|^2)/2\}. \qquad \ldots \quad (2.9)$$

Apply next the transformation

$$T_2(r, \theta, z^2) = (r', \theta, z^2) \text{ with } r' = (g_n(T_1^{-1}(r, \theta, z^2)))^{\frac{1}{2}}.$$

Note that T_2 is a diffeo-morphism on $T_1(M_n)$. Since by part (a) of A_s(v)

$$g_n(z) = h_{s-1}(z)+\|z^1\|^2 \cdot o(n^{-s-3)/2},$$

uniformly on M_n and under T_1,

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

(use A_s(iv)), it can be shown that

$$r = r'\left(1+\sum_{j=1}^{s-3} n^{-j/2} R_{3j}(r', \theta, z^2)+o(n^{-(s-3)/2}\right) \qquad \ldots \quad (2.10)$$

uniformly on $T_2\, T_1\, (M_n)$.

By part (b) of A_s (v) $D^l g_n(z) = D^l h_{s-1}(z)+\|z^1\|o(n^{-(s-3)/2})$ *uniformly* on M_n and so

$$\frac{\partial}{\partial r} g_n(T_1^{-1}(r, \theta, z^2)) = \frac{\partial}{\partial r} h_{s-1}(T_1^{-1}(r, \theta, z^2))+o(n^{-(s-3)/2})$$

$$= 2r\left(1+\sum_{j=1}^{s-3} n^{-s/2} R_{4j}(r, \theta, z^2))+o(n^{-(s-3)/2}\right) \quad \ldots \quad (2.11)$$

uniformly on $T_1(M_n)$. Finally the Jacobian of T_2 is

$$\frac{\partial r}{\partial r'} = 2r'(\partial g_n(T_1^{-1}(r, \theta, z^2))/\partial r)^{-1}. \qquad \ldots \quad (2.12)$$

From (2.9), (2.10), (2.11) and (2.12), one gets

$$\int_{B_n \cap M_n} \xi_{s-1,n} = \int (r')^{p-1} J(\theta)\left[1+\sum_{1}^{s-3} n^{-j/2} R_{5j}(r', \theta, z^2)\right]$$
$$\{(r, \theta, z^2) : (r')^2 \epsilon B\} \cap T_2 T_1(M_n)$$
$$\exp\{-[(r')^2+\|z^2\|^2]/2\}+o(n^{-(s-3)/2}), \qquad \ldots \quad (2.13)$$

uniformly over all Borel subsets B of R^1. Since

$$T_2T_1(M_n) \supset \{(r', 0, z^2) : (r')^2+\|z^2\|^2 < (s-3/2)\log n\}$$

for all sufficiently large n, one can replace (2.13) by

$$\int_{B_n \cap M_n} \xi_{s-1,\,n} = \int_{\{(r',0,z^2):(r')^2 \epsilon B\}} (r')^{p-1}J(\theta)\left[1+\sum_{j=1}^{s-3} n^{-j/2}R_{bj}(r',\theta,z^2)\right]$$
$$\exp\{-[(r')^2+\|z^2\|^2]/2\}+o(n^{-(s-3)/2})$$

uniformly over all Borel subsets B of R^1. From this and using Remark 2.6, it is easy to get (2.8). This completes the proof of Theorem 1(a).

We now come to the proof of Theorem 1(b). First note, because of the estimate

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

uniformly on M_n, that

$$\sup_{u \epsilon R^1}\left|\int_{\{z:g_n(z)<u\}\cap M_n} \xi_{s-1,n} - \int_{\{z:h_{s-1}(z)<u\}\cap M_n} \xi_{s-1,n}\right| = O(n^{-(s-2)/2}(\log n)^{s/2}.$$

Thus it is enough to exihibit $\psi_{m,n}$ (using the equation 2.20, of BG) of the form (2.6) such that

$$\sup_{u \epsilon R^1}\left|\int_{\{z:h_{s-1}(z)<u\}\cap M_n} \xi_{s-1,\,n} - \int_{-\infty}^{u} \psi_{m,\,n}\right| = \epsilon_n.$$

One now applies the transformation T_2T_1 except that here one defines

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

and arrives at the equation (2.10). The rest of the proof is similar to that of part (a).

This completes the proof of Theorem 1.

Remark 2.6 : An explicit method of determining the polynomials ψ_r is described below. Suppose that the assumptions of Theorem 1(b) hold. Let r be the highest degree of the polynomial $\alpha_{m,n}(v) = \psi_{m,n}(v)/f_{\chi^2,p}(v)$ (see (2.6)). Assume that the moments of Z_1 of order $(s-1)r$ are finite. Compute the first r moments of $W_n' = h_{s-1}(n^{\frac{1}{2}}(\overline{Z}_n-\mu))$ up to $o(n^{-(s-3)/2})$ and use Laguerre polynomials to find a polynomial (of degree r) $\hat{\alpha}_{m,n}$ such that

$$\int v^i f_{\chi^2,p}(v)\, \hat{\alpha}_{m,n}(v)\ dv = E(W_n'^i)+o(n^{-(s-3)/2})$$

for $i = 0, 1, \ldots, r$. Then $\hat{\alpha}_{m,n} = \alpha_{m,n}$. This is so because

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

$$= \int_{R^1} v^i \psi_{m,n}(v)dv + o(n^{-(s-3)/2}). \quad \ldots \quad (2.14)$$

(The first equality follows from Theorem 20.1 of Bhattacharya and Rao, 1976, while the second can be proved by a slight refinement of Theorem 1(b)). Consequently,

$$\int v^i f_{\chi^2,p}(v)\hat{\alpha}_{m,n}(v) = \int v^i f_{\chi^2,p}(x)\, \alpha_{m,n}(v)dv, \ 0 \leqslant i \leqslant r.$$

Both of $\hat{\alpha}_{m,n}$ and $\alpha_{m,n}$ being polynomials of degree r, must be therefore indentical.

Remark 2.7 : Under certain conditions, the inversion of a formal expansion for the characteristic function can be justified as follows : Suppose that the assumptions of Theorem 1(b) hold. Then

$$E(\exp(itW_n)) = E(\exp(itW_n')) + o(n^{-(s-3)/2})$$

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

$$= \int \exp(itv)\psi_{m,n}(v)dv + o(n^{-(s-3)/2}). \quad \ldots \quad (2.15)$$

(The first equality follows from von Bahr's result (vide 2.31, page 447 of BG); the second one follows from Theorem 20.1 of Bhattacharya and Rao, 1976, and finally the third equality can be obtained by following the proof of Theorem 1(b)).

Let

$$\hat{\psi}_{m,n}(v) = f_{\chi^2,p}(v) \sum_{j=0}^{m} n^{-j/2}\, \hat{\psi}_j(v)$$

($\hat{\psi}_j$'s are polynomials in v) be an 'asymptotic expansion' for the d.f. of W_n such that

$$\int \exp(itv)\hat{\psi}_{m,n}(v)dv = E(\exp(itW_n)) + o(n^{-(s-3)/2}).$$

Then the definition of $\hat{\psi}_{m,n}$ and (2.15) imply that the Fourier-Stieltjes Transforms of $\hat{\psi}_{m,n}$ and $\psi_{m,n}$ are identical and so $\hat{\psi}_{m,n} = \psi_{m,n}$.

Thus the expansions obtained formally by inverting asymptotic expansions of the exact characteristic functions $E(e^{itW_n})$ or of the approximate characteristic functions namely $E(e^{itW'_n})$ and $\int \exp(ith_{s-1})\, \xi_{s,n-1}$ of W_n are valid.

We conclude this section with the following counter examples.

Example 2.1 : Suppose that $\{Z_n\}_{n \geqslant 1}$ are i.i.d. 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 all the assumptions of Theorem 1(a) hold except A_s(iv). We shall show that the conclusion of Theorem 1(b) does not hold (and hence that of Theorem 1(a) also does not hold).

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

$$A_n = \{(w, y) : a \leqslant w \leqslant x, |y| < (3 \log n)^{1/2}\}$$

and

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

Then one has, *uniformly in* x,

$$P(a \leqslant W_n \leqslant x)$$

$$= P(a \leqslant W_n \leqslant x, |Y| < (3 \log n)^{1/2}) + o(n^{-1})$$

$$= \text{const.} \int_{B_n} (w - n^{-1/2} y^3)^{-1/2} \exp\{-\tfrac{1}{2}(w - n^{-1/2} y^3) - \tfrac{1}{2} y^2\} dw dy + o(n^{-1})$$

$$= \text{const} \int_{A_n} w^{-1/2} \exp\{-(w + y^2)/2\} \left(1 + \frac{n^{-1/2} y^3}{2} + \frac{n^{-1} y^6}{8}\right)$$

$$\cdot \left(1 + \frac{n^{-1/2} y^3}{2w} + \frac{3n^{-1} y^6}{8w^2}\right) dw dy + o(n^{-1})$$

$$= \text{const.} \int_{\substack{a \leqslant w \leqslant x \\ -\infty < y < \infty}} w^{-1/2} \exp\{-(w + y^2)/2\} \left(1 + \frac{n^{-1/2} y^3}{2} + \frac{n^{-1} y^6}{8}\right)$$

$$\cdot \left(1 + \frac{n^{-1/2} y^3}{2w} + \frac{3n^{-1} y^6}{8w^2}\right) dw dy + o(n^{-1})$$

$$= \text{const.} \int_{a \leqslant w \leqslant x} w^{-1/2} \exp(-w/2) \left(1 + n^{-1} a_1 + \frac{n^{-1} a_2}{w} + \frac{n^{-1} a_3}{w^2}\right) dw dy + o(n^{-1})$$

where a_1, a_2, a_3 are non-zero constants. Clearly the conclusion of Theorem 1(b) is incompatible with the last equality.

Example 2.2 : Take Z_i's as in Example 2.1 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 2.1. Then, it can be shown using remark 2.3 that W_n has an expansion of the type asserted in Theorem 1. But evidently H does not satisfy A_1(iv).

Remark 2.8 : Example 2.1 suggests that even when the condition A_1(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 here.

3. Likelihood ratio statistic for the case of exponential densities

Let $\{Y_n\}_{n \geq 1}$ be i.i.d. m-dimensional random vectors with common exponential density, $f_\theta(y) = \exp\left\{c(\theta)+\sum_{i=1}^{k} \theta^{(i)} f_i(y)\right\}$ with respect to some σ-finite measure μ where $f_1, \ldots, f_k$ are continuously differentiable real-valued functions. Let Θ be the natural parameter space (vide Lehmann, 1959, page 51). *Assume* that μ has an absolutely continouous component in 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 $Z_1 = (f_1(Y_1), f_2(Y_1), \ldots, f_k(Y_1))$ satisfies condition D (see (2.4)) under each $\theta \epsilon \Theta$.

We assume that Θ has a nonempty interior and without loss of generality that the origin is an interior point of Θ. Consider the testing problem

$$H_0(\theta^{(1)} = \ldots = \theta^{(p)} = 0) \text{ vs. } H_1(\theta^{(1)} \neq 0, \text{ or } \theta^{(2)} \neq 0, \ldots \text{ or } \theta^{(p)} \neq 0)$$

$(1 \leqslant p \leqslant k)$. The transformed LR statistic is $\lambda_n = 2[\sup_{\theta \epsilon \Theta} L_n(\theta) - \sup_{\theta \epsilon \Theta_0} L_n(\theta)]$ where $L_n(\theta) = \sum_{i=1}^{n} \log f_\theta(Y_i)$ is the *log-likelihood* function and $\Theta_0 = \{\theta : \theta^{(1)} = \ldots = \theta^{(p)} = 0\}$. Let θ_0 belong to the interior of Θ_0. We want to establish the existence of an asymptotic expansion for the distribution of λ_n under θ_0.

Theorem 2 : *For all integers $s \geqslant 4$, there exists $\psi_{s,n}$ of the form* (2.6) *such that*

$$\sup_B \{|P_{\theta_0}(\lambda_n \epsilon B) - \int_B \psi_{s,n}|\} = o(n^{-(s-3)/2})$$

where the supremum is over all Borel subsets B of R^1.

Proof : Without loss of generality we may take θ_0 to be the origin. The likelihood equations for the unrestricted m.l.e. $\tilde{\theta}$ are

$$\frac{\partial}{\partial \theta^{(j)}} c(\theta)+\overline{Z^{(j)}} = 0, \quad 1 \leqslant j \leqslant k, \qquad \ldots \quad (3.1)$$

and the corresponding equations for the m.l.e. $\hat{\theta}^2 (\equiv \hat{\theta}^2_n)$ under Θ_0 are

$$\frac{\partial}{\partial \theta^{(j)}} c(0, \theta^2)+\bar{Z}^{(j)} = 0, \quad p+1 \leqslant j \leqslant k, \qquad \ldots \quad (3.2)$$

where for $\theta = (\theta^{(1)}, \ldots, \theta^{(k)})$, we let $\theta^2 = (\theta^{(p+1)}, \ldots, \theta^{(k)})$. Let $\mu^{(j)} = E_{\theta_0}(Z_1^{(j)})$ and $\mu = (\mu^{(1)}, \ldots, \mu^{(k)})$. If in (3.1) and (3.2) we replace $\overline{Z^{(j)}}$ by $\mu^{(j)}$, then they have a solution $\tilde{\theta} = 0$, $\hat{\theta}^2 = 0$ respectively. Since the $k \times k$ matrix whose (i, j)-th element is $\partial^2 c(\theta)/\partial\theta^{(i)}\partial\theta^{(j)} = E_\theta(\partial^2 L_1(\theta)/\partial\theta^{(i)}\partial\theta^{(j)})$ is negative definite, it follows by the Implicit Function Theorem (see page 272 of Dieudonne, 1969) that there is a bounded neighbourhood N of μ where both of (3.1) and (3.2) have unique solutions $\tilde{\theta}$ and $\hat{\theta}^2$ respectively and

$$\sup_{\theta \in \Theta_0} L_n(\theta) = L_n(0, \hat{\theta}^2), \quad \sup_{\theta \in \Theta} L_n(\theta) = L_n(\tilde{\theta}).$$

Since θ_0 is an interior point of Θ_0, by Chernoff's theorem we may assume

$$P_{\theta_0}(\bar{Z}_n \notin N) = o(n^{-(s-3)/2}), \ \forall s \geqslant 4.$$

In a suitable neighbourhood $(\subset N)$ of μ, $\bar{Z}_n$ and $\hat{\theta}^2$ can be written as functions of $\tilde{\theta}$, the functions themselves being free from n and hence we may write, using the equation (3.1),

$$\begin{aligned}\lambda_n &= 2n \left\{ c(\tilde{\theta})+ \sum_{i=1}^{k} \tilde{\theta}^{(i)}\bar{Z}^{(i)} - c(0, \hat{\theta}^2) - \sum_{j=p+1}^{k} \hat{\theta}^{(j)}\bar{Z}^{(j)} \right\} \\ &= 2nH(\hat{\theta}), \quad \text{say.}\end{aligned}$$

Differentiating the likelihood equations (3.1) for $\tilde{\theta}$ with respect to $\bar{Z}_n^{(j)}$ $(1 \leqslant j \leqslant k)$, one gets

$$I_{k\times k}+((\partial^2 c(\tilde{\theta})/\partial\tilde{\theta}^{(i)}\partial\tilde{\theta}^{(j)}))_{k\times k}((\partial\tilde{\theta}^{(j)}/\partial\bar{Z}_n^{(i)}))_{k\times k} = 0,$$

and consequently, the matrix $((\partial\tilde{\theta}^{(j)}/\partial\bar{Z}_n^{(i)}))$ when evaluated at $\mu = E_{\theta_0}(Z_1)$ (which implies that $\tilde{\theta} = 0$) is nonsingular. Since $\tilde{\theta}$ is an analytic function of

$\bar{Z}_n$ and Z_1 satisfies condition D, it therefore follows from Theorem 2(a) and Remark 1.1 of BG that $\sqrt{n}(\bar{\theta}-\theta_0)$ has a multivariate Edgeworth expansion which holds uniformly over all Borel subsets of R^k; in particular the analogue of equation (2.30) of BG holds. Thus in view of Theorem 1(a) and Remark 2.2, Theorem 2 will follow if we can check the assumptions A_s((i)—(v)) for the function H defined above. Since c is analytic, by a version of the Implicit Function Theorem, H is also analytic. It is well-known and in fact not difficult to check that

$$\frac{\partial H}{\partial \theta^{(i)}}\bigg|_{\theta_0} = 0, \quad i = 1, \ldots, k;$$

$$\frac{\partial^2 H}{\partial \theta^{(i)}\partial \theta^{(j)}}\bigg|_{\theta_0} = 0, \quad \text{if } i \text{ or } j > p$$

and

$$\left(\left(\frac{\partial^2 H}{\partial \theta^{(i)}\partial \theta^{(j)}}\bigg|_{\theta_0}\right)\right)_{i,j=1,\ldots,p}$$

$$= \left(\left(\frac{\partial^2 c(\theta)}{\partial \theta^{(i)}\partial \theta^{(j)}}\bigg|_{\theta_0}\right)\right)^{-1}_{i,j=1,\ldots,p}$$

Thus A_s(i)-(ii) hold. For A_s(iv) and (v), it is sufficient to check that (see Lemma 2.1)

$$\frac{\partial H}{\partial \theta^{(i)}} = 0, \qquad \text{for} \quad i = 1, 2, \ldots, p$$

for all $\theta = (0, \theta^2)$ where θ^2 lies in some neihbourhood of θ_0^2. For this one writes, using the likelihood equations for $\tilde{\theta}$,

$$H(\tilde{\theta}) = c(\tilde{\theta}) - \sum_{i=1}^{k} \tilde{\theta}^{(i)} \frac{\partial c(\tilde{\theta})}{\partial \tilde{\theta}^{(i)}} - c(0, \theta^2) - \sum_{j=p+1}^{k} \theta^{(j)} \frac{\partial c(\tilde{\theta})}{\partial \tilde{\theta}^{(j)}}$$

and verifies that $\partial H(\tilde{\theta})/\partial \tilde{\theta}^{(i)} \equiv 0$ for all $\tilde{\theta} = (0, \tilde{\theta}^2)$. Note that when $\tilde{\theta}_1 = 0$, $\tilde{\theta}^2 = \bar{\theta}^2$.

Remark 3.1 : We have verified the assumptions A_4(iv) and (v) by writing λ_n as a function of the unrestricted m.l.e. $\tilde{\theta}$. Clearly λ_n can as well be expressed as a function of $\bar{Z}_n$. However, considering the problem of testing the hypothesis that the population variance is 1 against that it is not 1, the population mean being unknown and the observations coming from a normal population, one can easily verify that the assumption A_4(iv) need not hold if λ_n is regarded as a function of $\bar{Z}_n$. In case of a simple null hypothesis, those two assumptions will always hold and consequently Theorem 2 can be proved without using the multivariate Edgeworth expansion of $\tilde{\theta}$.

Box (1949) considered the test of constancy of variances and covariances of k-sets of p-variate normal populations and derived an 'asymptotic χ^2-series solution' of the null distribution of the test statistic M which is a generalised form of Bartlett's statistic. We use our results of Section 2 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's paper.* We describe briefly the approach of Box; for details, see Section 2.1, pages 320-323 of his paper. Box first derives an asymptotic expansion of the logarithm of the exact c.f. $\phi(t)$ of M and uses it to deduce that $\phi(t) = \phi_n(t)+o(\mu^{-n})$ 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 χ^2-distribution of M. (This part of Box's argument can be made rigorous, without any difficulty). A *formal* inversion now gives an asymptotic expansion of the density $p(x)$ of M :

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

where $p_n(x) = K \sum_{v=0}^{n} \mu^{-v}a_v$ (density of a χ^2-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 series (see equation (30), page 323 of Box's paper); obviously this last part of Box's argument can be justified easily.

For justification, take M to be W_n. Put $\bar{\theta}_n = K(\bar{Z}_n)$ and let $\xi_{s-1,n}$ and $\xi'_{s-1,n}$ be respectively the Edgeworth expansions of $n^{\frac{1}{2}}(\bar{\theta}_n - \theta_0)$ and $n^{\frac{1}{2}}(\bar{Z}_n - \nu)$ where $\nu = E_{\theta_0}(Z_1)$. In view of Remark 2.7, it is enough to show that

$$E(\exp(itW_n)) = \int_{R^1} \exp(itv)\psi_{m,n}(v)dv + o(n^{-(s-3)/2}).$$

$$\text{L.H.S.} = \int_{R^k} \exp\{itg'_n(n^{\frac{1}{2}}(\bar{Z}_n - \nu))\}\xi'_{s-1,n} + o(n^{-(s-3)/2})$$

$$= \int_{R^k} \exp\{itg_n(n^{\frac{1}{2}}(\bar{\theta}_n - \theta_0))\}\xi_{s-1,n} + o(n^{-(s-3)/2})$$

$$= \int_{R^1} \exp(itv)\psi_{m,n}(v) + o(n^{-(s-3)/2}).$$

Here g'_n is defined by the equation $W_n = g'_n(n^{\frac{1}{2}}(\bar{Z}_n - \nu))$. The first equality follows from von Bahr's result and Theorem 20.1 of (Bhattacharya and Ranga Rao, 1976); the proof of the second one is a multivariate extension of that of Lemma 2.1 of BG, while that of the third one is similar to the proof of Theorem 1(b).

It is interesting to note that Box in the last paragraph of Section 2.1 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, didnot supply a proof of his remark, for this one would need to show that the formal differentiation of the identity (18), page 322 of Box's paper is permissible. Since all the moments of $M(= W_n)$ are finite and since the r-th cumulant is a polynomial function of the first r moments $(r \geqslant 1)$, to establish Box's remark it is enough to show that

$$E(W_n^r) = \int v^r \psi_{m,n}(v)dv + o(n^{-(s-3)/2}), \quad r \geqslant 1. \qquad \ldots \quad (3.3)$$

Since W_n can be bounded in absolute value by a polynomial in S_{ijl}'s. Theorem 20.1 of (Bhattacharya and Ranga Rao, 1976) implies that

$$E(W_n^r I_{M_n^c}) = o(n^{-(s-3)/2}). \qquad \ldots \quad (3.4)$$

Put $W_n^* = h_{s-1}(k_{s-1}(n^{\frac{1}{2}}(Z_n - \nu)))$, k_{s-1} being the Taylor expansion of g'_n (up to $(s-1)$-th order derivatives).

Then

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

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

$$= \int v^r \psi_{m,n}(v) dv + o(n^{-(s-3)/2}) \quad \ldots \quad (3.5)$$

here $N_n = \{\|n^{\frac{1}{2}}(\bar{\theta}_n - \theta_0)\| < ((k-1)\log n)^{\frac{1}{2}}\}$. Together (3.4) and (3.5) imply (3.3).

4. Likelihood ratio statistic in the general case

A valid asymptotic expansion for the distribution function of the likelihood ratio statistic in the general case can be obtained up to any order of accuracy provided suitable regularity conditions hold. We shall however outline only a sketch of the argument.

Consider the same testing problem as in Section 3 (but we do not assume that the observations $\{Y_i\}_{i \geq 1}$ are coming from an exponential family of distributions). Let $\hat{\theta}^2 (\equiv \hat{\theta}_n^2)$ denote as before the m.l.e. when H_0 is assumed and $\bar{\theta}(\equiv \theta_n)$ the unrestricted m.l.e. Put $\underset{p\times 1}{\theta^1} = \theta_0'$; for the explanation of other notations, Section 3 should be consulted.

For convenience in notations, we shall discuss first the case $p = 1$, $k = 2$ and obtain our expansion up to $o(n^{-1})$. Choose and fix θ_0 whose first component is zero. All probability statements made in this section refer to P_{θ_0}, the product probability under θ_0 on the set of all infinite sequence of observations. We shall assume without loss of generality that $\theta_0 = 0$.

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

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

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

$$+ n^{1/2}(\hat{\theta}^{(1)} - \bar{\theta}^{(1)}) \left\{ \frac{R_{11}}{\sqrt{n}} + \frac{R_{12}}{n} \right\} \quad \ldots \quad (4.1)$$

$$+ n^{1/2}(\hat{\theta}^{(2)} - \bar{\theta}^{(2)}) \left\{ \frac{R_{21}}{\sqrt{n}} + \frac{R_{22}}{n} \right\}$$

$$+ o(n^{-1}).$$

on a set A_n. Here R's are polynomials in $n^{1/2}(\hat{\theta}-\tilde{\theta})$ and

$$U_n \equiv \{n^{1/2}(L_n^{i_1\ldots i_j}(\tilde{\theta})-I_{i_1\ldots i_j}(\theta_0)) : j=2,\ 3,\ 4,\ i_1, \ldots, i_j = 1 \text{ or } 2\}$$

whose coefficients do not depend on n; for any $j \geqslant 1$ and $i_1, \ldots, i_j = 1$ or 2,

$$L_n^{i_1\ldots i_j}(\theta) = n^{-1}(D^{i_1} \ldots D^{i_j} L_n)(\theta),$$

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

$$I_{i_1\ldots i_j} = I_{i_1\ldots i_j}(\theta_0),$$

$$L_n^{i_1\ldots i_j} = L_n^{i_1\ldots i_j}(\theta_0),$$

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

$$\|n^{1/2}(U_n^*-E_{\theta_0}(U_n^*))\|^2 \leqslant 3 \log n,$$

$$\|n^{1/2}(\hat{\theta}-\theta_0)\|^2 \leqslant 3 \log n,$$

$$\|n^{1/2}(V_n-E_{\theta_0}(V_n))\|^2 \leqslant 3 \log n,$$

U_n^* being the vector whose components are

$$\tilde{\theta},\ L_n^{i_1 i_2},\ L_n^{i_1 i_2 i_3},\ L_n^{i_1 i_2 i_3 i_4},$$

$(i_1, i_2, i_3, i_4 = 1$ or $2)$ and V_n the vector whose components are

$$\sum_{i'=1}^{n} \left\{ \sup_{|\theta-\theta_0| < \epsilon} L_1^{i_1\ldots i_5}(\theta\,;\, Y_{i'}) \right\};$$

$(i_1, \ldots, i_5 = 1$ or $2)$. Under the regularity assumptions stated below, it can be shown that

$$P_{\theta_0}(A_n) = 1-o(n^{-1}).$$

One can verify that on the set A_n the following is true: for any sequence $\{t_n\}$ such that $0 \leqslant t_n \leqslant 1$ and for any $i_j = 1$ or 2, $1 \leqslant j \leqslant 5$,

$$L_n^{i_1 i_2 i_3 i_4 i_5}(\theta_0+t_n(\tilde{\theta}-\theta_0)) \text{ is bounded.}$$

This fact will be repeatedly used without any explicit mention.

Below all subsequent expansions are performed on the set A_n. Also we shall assume that the matrix $((I_{ij}))$ is negative-definite and that the vector (I_j) is null. We shall say that $I_{i_1 \ldots i_j}$ is the asymptotic mean of $L_n^{i_1 \ldots i_j}(\bar{\theta})$.

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

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

$$= n^{1/2}(\hat{\theta}^{(1)}-\bar{\theta}^{(1)})\left\{-\frac{I_{21}}{I_{22}}+\frac{P_1}{\sqrt{n}}+\frac{P_2}{n}\right\}+o(n^{-1})$$

where P_1 and P_2 are polynomials in $n^{1/2}(\hat{\theta}^{(1)}-\bar{\theta}^{(1)})$ and U_n whose coefficients do not depend on n. 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 $\bar{\theta}$ of L_n appearing in P_1 and P_2 around θ_0, it can be verified that

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

$$= n^{1/2}(\hat{\theta}^{(1)}-\bar{\theta}^{(1)})R_n^*(U_n^*)+o(n^{-1}) \qquad \ldots \quad (4.2)$$

where R_n^* is a polynomial in $n^{1/2}(U_n^*-E_{\theta_0}(U_n^*))$ whose coefficients depend on n.

Expanding λ_n around $\bar{\theta}$ and then expanding the partial derivatives at $\bar{\theta}$ of L_n around θ_0 and finally using (4.2), one gets

$$\lambda_n = 2\{L_n(\bar{\theta})-L_n(\hat{\theta})\}$$

$$= n(\hat{\theta}^{(1)}-\bar{\theta}^{(1)})^2 P_n^*(U_n^*+o(n^{-1}) \qquad \ldots \quad (4.3)$$

where P_n^* is a polynomial in $n^{1/2}$ $(U_n^*-E_{\theta_0}(U_n^*))$ whose coefficients depend on n. Let

$$\lambda_n' = n(\hat{\theta}^{(1)}-\bar{\theta}^{(1)})^2 P_n^*(U_n^*).$$

Then on the set A_n the condition A_s (iv) holds for λ_n' and λ_n can be approximated by λ_n' up to terms of order $o(n^{-1})$.

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., non-random) and hence (4.3) can be used as an alternative way of checking the condition A_3(iv).

Coming to the general (i.e., non-exponential) case, *suppose that the assumptions* (A_1) *to* (A_4) *and* (A_6) *of Theorem 3 of BG with* $s = 4$ *hold where in* (A_2) *we do not impose the conditions ensuring uniformity in* $\theta_0 \in K$ *and that the random variables*

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

are linearly independent (*i.e., have nonsingular dispersion matrix*). Then the fact that $P_{\theta_0}(A_n) = 1 - o(n^{-1})$ follows from relations (1.28), (1.29) and (2.32) of BG and the analoguus relations for the restricted m.l.e. $\hat{\theta}$. Also following the proof of Theorem 3 of BG we can show that the vector U_n^* have a multivariate Edgeworth expansion *i.e.*, Theorem 3 of BG holds if the vector $n^{1/2}(\hat{\theta} - \theta_0)$ of this theorem is replaced by the vector U_n^*. By our Theorem 1(b) (applied to λ_n'), Remark 2.2 and the equation (4.3), the expansion up to $o(n^{-1})$ of the type stated in Theorem 1(b) is valid for the transformed likelihood ratio statistic.

Suppose now that the assumptions (A_1) to (A_4) and (A_6) of BG hold with $s = 4$ *but that the random vectors*

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

and

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

are not linearly independent. Let

$$y_i = n^{1/2}(L_n^i - I_i), \quad i = 1, 2$$

and let $x_1, \ldots, x_m$ stand for the set

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

(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 coefficient polynomials $n^{1/2}(\tilde{\theta}-\theta_0)$. Consequently in the equation (4.3) the polynomial P_n^* can be replaced by another polynomial in $\{n^{1/2}(\tilde{\theta}-\theta_0), x_1, \ldots, x_r\}$ and the argument of the previous paragraph goes through.

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

$$c_{i1}T_1+\ldots+c_{im}T_m+d_{i1}T_1'+d_{i2}T_2' = 0 \qquad \ldots \quad (4.4)$$

for $i = 1, \ldots, r$ where r is the rank of $(C|D)$, $\underset{r\times m}{C} = ((c_{ij}))$, $\underset{r\times 2}{D} = ((d_{ij}))$. Observe that r is also the rank of C since T_1', T_2' are linearly independent (i.e., T_1' and T_2' have positive-definite dispersion matrix). Without loss of generality, let the first r columns of C be linearly independent. *Then clearly T_1', T_2' and $T_{r+1}, \ldots, T_m$ are linearly independent.* Also from (4.4) we get

$$\begin{aligned} & c_{i1}x_1+\ldots+c_{ir}x_r \\ &= -c_{i(r+1)}x_{r+1}-\ldots-c_{im}x_m-d_{i1}y_1-d_{i2}\,y_2. \end{aligned} \qquad \ldots \quad (4.5)$$

On the set A_n we now expand (up to $o(n^{-1/2})$) y_i around $\tilde{\theta}$ and then the partial derivatives at $\tilde{\theta}$ of L_n around θ_0, $i = 1, 2$; (4.5) then implies that

$$\begin{aligned} & (c_{i1}+P_{i1}(\tilde{\theta}))x_1+\ldots+(c_{ir}+P_{ir}(\tilde{\theta}))x_r \\ &= -(c_{i(r+1)}+P_{i(r+1)}(\tilde{\theta}))\,x_{r+1}+\ldots+(c_{im}+P_{im}(\tilde{\theta}))x_m \\ &\quad -d_{i1}P_{i1}'(\tilde{\theta})-d_{i2}P_{i2}'(\tilde{\theta})+o(n^{-1/2}), \end{aligned}$$

where for each i, j, $P_{ij}(\tilde{\theta})$ is either $(\theta_0^{(1)}-\tilde{\theta}^{(1)})$ or $(\theta_0^{(2)}-\tilde{\theta}^{(2)})$ and

$$\begin{aligned} P_{ii'}'(\theta) &= \sum_{j=1}^{2} n^{1/2}(\theta_0^{(j)}-\tilde{\theta}^{(j)})I_{i'j} \\ &\quad + \frac{1}{\sqrt{n}}\cdot \sum_{j,j'=1}^{2} n^{1/2}(\theta_0^{(j)}-\tilde{\theta}^{(j)})n^{1/2}(\theta_0^{(j')}-\tilde{\theta}^{(j')})I_{i'jj'}, \end{aligned}$$

$i' = 1$ or 2. Hence for all sufficiently large n, the $r\times r$ matrix $((c_{ij}+P_{ij}(\tilde{\theta})))$ 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 coefficient polynomials in* $n^{1/2}(\hat{\theta}-\theta_0)$.

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 $\theta_0 = 0$ and that the $k\times k$ matrix $((I_{ij}))$ is negative-definite. Apply first a nonsingular linear transformation on the parameter space which leaves the first p components of θ unchanged and which reduces

$$((I_{ij}))_{i,j} = {}_{p+1,\ldots,k}$$

to the identity matrix of order $(k-p)$.

Expanding (D^iL_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 functions for $\hat{\theta}^2$ as follows :

$$n^{1/2}(\theta^{(i_1)}-\tilde{\theta}^{(i_1)})$$

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

$(p+1 \leqslant i_1 \leqslant k,\ m \geqslant 1)$ where $\{R_{i_1 i_2 i_3}\}$ are polynomials in $n^{1/2}(\hat{\theta}-\tilde{\theta})$ and 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 with P_{θ_0}-probability $1-o(n^{-m})$.

An iterative approximation of

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

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

$$n^{1/2}(\theta^{(i_1)}-\tilde{\theta}^{(i_1)})$$

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

$(p+1 \leqslant i_1 \leqslant k)$ 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 $n^{1/2}(\theta^2 - \bar{\theta}^2)$. The rest of the computations is similar to that of the case $p = 1$, $k = 2$ and $m = 1$ and requires similar assumptions.

Remark 4.1 : Suppose that the assumptions (A_1) to (A_6) of BG with $s = 2m+2$ hold. Then it can be shown that the above asymptotic expansion of the distribution function of the transformed likelihood ratio statistic up to $o(n^{-m})$ holds uniformly in $\theta_0 \epsilon K$ for any compact K.

In the rest of this section 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' upto $o_p(n^{-1})$, and finally inverting formally the resulting approximate characteristic function or W_n'. Since

$$\int_{R^k} \exp\{ith_{s-1}(k_{s-1}(z))\}\, \xi'_{s-1,n}(z)dz$$

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

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

$$= \int_{R^1} \exp\{itv\}\, \psi_{m,n}(v)dv + o(n^{-(s-3)/2})$$

(we are using the notations of the last part of Section 3 except that here we take W_n to be λ_n and $Z_1 = (L_1^{i_1 \ldots i_j} : i_1, \ldots, i_j \geqslant 1,\ 1 \leqslant j \leqslant 4))$, it follows arguing as in Remark 2.7 that the asymptotic expansion given in Theorem 1 of Hayakawa is a valid one.

We finally remark that Hayakawa gets his expansion by expressing λ_n as functions 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 $\{\bar{\theta}, L_n^{i_1 i_2}, L_n^{i_1 i_2 i_3}, L^{i_1 i_2 i_3 i_4}\}$). But as observed earlier (see Remark 3.1) the assumption A_s(iv) then need not hold in general even on a set of P_{θ_0}-probability $1 - o(n^{-1})$.

5. Other applications

Consider the same testing problem as in Section 4. Here we discuss briefly how the asymptotic expansions for the statistics proposed by Wald (1945) and Rao (1948) (vide Rao 1965, pages 347-352) as alternatives to the likelihood ratio statistic can be obtained up to any order of approximation under the regularity conditions stated in Section 4. Below we use the notations of Section 4.

Wald's statistic is

$$W_n = -n(\tilde{\theta}^1 - \theta_0^1)^T I_{11\cdot 2}(\tilde{\theta})(\tilde{\theta}^1 - \theta_0^1)$$

where

$$I_{11\cdot 2}(\theta) = I_{11}^*(\theta) - I_{12}^*(\theta)(I_{22}^*(\theta))^{-1} I_{21}^*(\theta)$$

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

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

It evident from the expansions of $I_{ij}^*(\theta)$ around θ_0 that the condition A_2(iv) is satisfied if we regard W_n as a function of $\tilde{\theta}$ and the appropriate partial derivatives of L_n at θ_0. So by Theorem 1(b) and Remark 2.2 one gets an asymptotic expansion of $P_{\theta_0}(W_n \leqslant u)$.

Rao's statistic is

$$S_n = -\sum_{i,j=1}^{k} \phi_i(\hat{\theta})\phi_j(\hat{\theta}) I^{ij}(\hat{\theta})$$

where $\phi_i(\theta)$ is *the i-th efficient score of* θ, *i.e.*,

$$\phi_i(\theta) = n^{-1/2} \partial L_n(\theta)/\partial\theta^{(i)}$$

and

$$((I^{ij}(\theta)))_{k\times k} = ((I_{ij}(\theta)))^{-1}_{k\times k};$$

recall that $\underset{p\times 1}{\hat{\theta}^1} = \theta_0^1$. For convenience, assume that $p = 1$ and $k = 2$. Expanding $\phi_i(\hat{\theta})$ as well as $I^{ij}(\hat{\theta})$ around $\bar{\theta}$ and then expanding the partial derivatives at $\tilde{\theta}$ of L_n around θ_0 and finally using the equation (4.2), one gets an asymptotic expansion of $P_{\theta_0}(S_n \leqslant u)$ up to $o(n^{-1})$; of course one can get an asymptotic expansion for S_n which is valid up to any degree of accuracy.

The case of testing the more general composite hypothesis considered in Section 6e.3 of Rao (1965) can be reduced, under appropriate assumptions which guarantee reparametrization in a suitable neighbourhood of θ_0, to the case discussed here.

References

Bhattacharya, R. N. and Ghosh, J. K. (1978): On the validity of the formal Edgeworth expansion. *Ann. Statist.*, **6**, 434-451.

Bhattacharya, R. N. and Ranga Rao, R. (1976): *Normal Approximations and Asymptotic Expansions*, Wiley, New York.

Bickel, P. J. (1974): Edgeworth expansions in nonparametric statistics. *Ann. Statist.*, **2**, 1-20.

Box, G. E. P. (1949): A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317-346.

Dieudonne, J. (1969): *Foundations of Modern Analysis*, Academic Press, New York.

Hayakawa, Takesi (1977): The likelihood ratio criterion and the asymptotic expansion of its distribution. *Ann. Inst. Statist. Math.*, **29**, Part A, 359-378.

Lehmann, E. L. (1959): *Testing Statistical Hypotheses*, Wiley, New York.

Rao, C. R. (1965): *Linear Statistical Inference And Its Applications*, Wiley, New York.

Paper received : October, 1978.
Revised : February, 1979.