The Annals of Statistics
1978, Vol. 6, No. 2, 434–451

ON THE VALIDITY OF THE FORMAL EDGEWORTH EXPANSION

By R. N. Bhattacharya[1] and J. K. Ghosh

The University of Arizona, the University of Pittsburgh and the Indian Statistical Institute

Let $\{Y_n\}_{n\geq 1}$ be a sequence of i.i.d. m-dimensional random vectors, and let $f_1, \cdots, f_k$ be real-valued Borel measurable functions on R^m. Assume that $Z_n = (f_1(Y_n), \cdots, f_k(Y_n))$ has finite moments of order $s \geq 3$. Rates of convergence to normality and asymptotic expansions of distributions of statistics of the form $W_n = n^{\frac{1}{2}}[H(\bar{Z}) - H(\mu)]$ are obtained for functions H on R^k having continuous derivatives of order s in a neighborhood of $\mu = EZ_1$. This asymptotic expansion is shown to be identical with a formal Edgeworth expansion of the distribution function of W_n. This settles a conjecture of Wallace (1958). The class of statistics considered includes all appropriately smooth functions of sample moments. An application yields asymptotic expansions of distributions of maximum likelihood estimators and, more generally, minimum contrast estimators of vector parameters under readily verifiable distributional assumptions.

1. Introduction. Consider a sequence of independent and identically distributed m-dimensional random vectors $\{Y_n\}_{n\geq 1}$. Let $f_1, \cdots, f_k$ be real-valued Borel measurable functions on R^m. Consider the statistic

$$W_n = n^{\frac{1}{2}}(H(\bar{Z}) - H(\mu)) \tag{1.1}$$

where H is a real-valued Borel measurable function on R^k, and

$$Z_n = (f_1(Y_n), \cdots, f_k(Y_n)), \qquad \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i, \qquad \mu = EZ_n. \tag{1.2}$$

Note that all functions of sample moments are of the form $H(\bar{Z})$. For example, $H(\bar{Z})$ becomes the bivariate sample correlation coefficient if one takes $m = 2$, $k = 5$, $f_1(y) = y^{(1)}$, $f_2(y) = y^{(2)}$, $f_3(y) = (y^{(1)})^2$, $f_4(y) = (y^{(2)})^2$, $f_5(y) = y^{(1)}y^{(2)}$ (for $y = (y^{(1)}, y^{(2)})$), $H(z) = (z^{(5)} - z^{(1)}z^{(2)})(z^{(3)} - (z^{(1)})^2)^{-\frac{1}{2}}(z^{(4)} - (z^{(2)})^2)^{-\frac{1}{2}}$ for $z = (z^{(1)}, \cdots, z^{(5)})$ belonging to a neighborhood N of $\mu = (EY_1^{(1)}, EY_1^{(2)}, E(Y_1^{(1)})^2, E(Y_1^{(2)})^2, (EY_1^{(1)}Y_1^{(2)}))$ contained in the set $\{z \in R^5 : z^{(3)} > (z^{(1)})^2, z^{(4)} > (z^{(2)})^2, -1 < H(z) < 1\}$; H may be defined arbitrarily outside N.

It is well known (see Cramér (1946), page 366, and Wilks (1962), page 260) that if Z_1 has finite second moments and H is continuously differentiable in a neighborhood of μ, then W_n has a limiting normal distribution with mean zero

Received November 1976; revised July 1977.

[1] Research supported by NSF Grant MCS76-06118.

AMS 1970 *subject classifications.* Primary 62E20; Secondary 62G05, 62G10, 62G20.

Key words and phrases. Asymptotic expansion, delta method, Cramér's condition, minimum contrast estimators.

and variance

$$\sigma^2 = \sum_{i,j=1}^{k} v_{ij} l_i l_j \tag{1.3}$$

where $V = ((v_{ij}))$ is the *dispersion matrix* of Z_1 and

$$l_i = (D_i H)(\mu) = \frac{\partial H(z)}{\partial z^{(i)}}\bigg|_{z=\mu} \qquad 1 \leq i \leq k;\ z = (z^{(1)}, \cdots, z^{(k)}) . \tag{1.4}$$

Throughout this article *it is assumed that σ^2 is positive*. As a first refinement of asymptotic normality one has

THEOREM 1. *If Z_1 has finite third moments and if all third order derivatives of H are continuous in a neighborhood of $\mu = EZ_1$, then*

$$\sup_{B \in \mathscr{B}} |\mathrm{Prob}\,(W_n \in B) - \int_B \phi_{0,\sigma^2}(v)\,dv| = O(n^{-1/2}) \tag{1.5}$$

for every class $\mathscr{B}$ of Borel sets satisfying

$$\sup_{B \in \mathscr{B}} \int_{(\partial B)^\varepsilon} \phi_{0,\sigma^2}(v)\,dv = O(\varepsilon) \quad (\varepsilon \downarrow 0) . \tag{1.6}$$

Here ∂B is the boundary of B, $(\partial B)^\varepsilon$ is the ε-neighborhood of B, and

$$\phi_{0,\sigma^2}(v) = (2\pi\sigma^2)^{-1/2} \exp\{-v^2/(2\sigma^2)\} \qquad -\infty < v < \infty . \tag{1.7}$$

It is important to note that the mean $H(\mu)$ and the variance σ^2/n of the asymptotic distribution of $H(\bar{Z})$ are *not* the mean and variance of $H(\bar{Z})$. Indeed, in many common examples (e.g., the t-statistic, the sample correlation) the mean and higher moments of $H(\bar{Z})$ may not even be finite. This feature of the problem shows up in a more serious manner when one attempts an asymptotic expansion going beyond (1.5). It is common practice among applied statisticians to calculate "approximate moments" of W_n by expanding $H(\bar{Z})$ around μ, keeping a certain number of terms, raising to an appropriate power and taking expectations term by term. This is the so-called *delta method*. These "approximate moments" are sometimes used to obtain a formal Edgeworth expansion of the distribution function of W_n. It was conjectured by Wallace (1958) (also see Bickel (1974)) that such a formal expansion would be valid if suitable assumptions were made. One of the principal aims in this article is to prove that a more precisely formulated version of this conjecture, as described in the following paragraphs, is valid. As pointed out by Wallace, such a formal expansion is easier to compute compared to the alternative procedure of reducing a multivariate Edgeworth expansion to a univariate one.

Denote the derivatives of H at μ by

$$l_{i_1,\ldots,i_p} = (D_{i_1} D_{i_2} \cdots D_{i_p} H)(\mu) \qquad 1 \leq i_1, \cdots, i_p \leq k , \tag{1.8}$$

where D_i denotes *differentiation with respect to the* ith *coordinate*. A Taylor expansion of W_n yields the statistic

$$W_n' = n^{1/2}\{\sum_{i=1}^{k} l_i(\bar{Z}^{(i)} - \mu^{(i)}) + \tfrac{1}{2} \sum_{i,j} l_{i,j}(\bar{Z}^{(i)} - \mu^{(i)})(\bar{Z}^{(j)} - \mu^{(j)}) + \cdots + \frac{1}{(s-1)!} \sum l_{i_1,\ldots,i_{s-1}}(\bar{Z}^{(i_1)} - \mu^{(i_1)}) \cdots (\bar{Z}^{(i_{s-1})} - \mu^{(i_{s-1})})\} . \tag{1.9}$$

Since $W_n - W_n' = o_p(n^{-(s-2)/2})$, one may expect that an asymptotic expansion of the distribution function of W_n' may coincide with that of W_n. Also, it is easy to check that (if Z_1 has sufficiently many finite moments) the jth *cumulant* $\kappa_{j,n}$ of W_n' is given by

$$\kappa_{j,n} = \bar{\kappa}_{j,n} + o(n^{-(s-2)/2}) \qquad j \geq 1\,, \tag{1.10}$$

where

$$\begin{aligned} \bar{\kappa}_{j,n} &= \textstyle\sum_{i=1}^{s-2} n^{-i/2} b_{j,i} && \text{if } \ j \neq 2\,, \\ &= \sigma^2 + \textstyle\sum_{i=1}^{s-2} n^{-i/2} b_{2,i} && \text{if } \ j = 2\,, \end{aligned} \tag{1.11}$$

and $b_{j,i}$'s depend only on appropriate moments of Z_1 and on derivatives of H at μ of orders $s-1$ and less. We refer to $\bar{\kappa}_{j,n}$ as "*approximate cumulants*" of W_n' (or W_n). The expression

$$\exp\left\{it\bar{\kappa}_{1,n} + \frac{(it)^2}{2}(\bar{\kappa}_{2,n} - \sigma^2) + \textstyle\sum_{j=3}^{s} \frac{(it)^j}{j!}\bar{\kappa}_{j,n}\right\}\exp\{-\sigma^2 t^2/2\} \tag{1.12}$$

is an approximation of the characteristic function of W_n' (or W_n). Expanding the first exponential factor one may reduce (1.12) to

$$\begin{aligned} &\exp\{-\sigma^2 t^2/2\}[1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2}\pi_r(it)] + o(n^{-(s-2)/2}) \\ &\qquad = \hat{\psi}_{s,n}(t) + o(n^{-(s-2)/2})\,, \end{aligned} \tag{1.13}$$

say, where π_r's are polynomials whose coefficients do not depend on n. The *formal Edgeworth expansion* $\Psi_{s,n}$ of the distribution function of W_n is defined by

$$\begin{aligned} \psi_{s,n}(v) &= \left[1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2}\pi_r\left(-\frac{d}{dv}\right)\right]\phi_{\sigma^2}(v)\,, \\ \Psi_{s,n}(u) &= \textstyle\int_{-\infty}^{u} \psi_{s,n}(v)\,dv\,. \end{aligned} \tag{1.14}$$

Note that the Fourier–Stieltjes transform of $\Psi_{s,n}$ is $\hat{\psi}_{s,n}$.

To state the next result let $|\cdot|$, $\langle\ ,\ \rangle$ denote *Euclidean norm* and *inner product*, respectively.

THEOREM 2. *Assume that, for some integer $s \geq 3$, all the derivatives of H of orders s and less are continuous in a neighborhood of $\mu = EZ_1$ and that $E|Z_1|^s$ is finite.*

(a) *If, in addition,* (i) *the distribution of Y_1 has a nonzero absolutely continuous component (with respect to Lebesgue measure on R^m) and* (ii) *the density of this component is strictly positive on some nonempty open set U on which $f_1, \cdots, f_k$ are continuously differentiable and $1, f_1, \cdots, f_k$ are linearly independent (as elements of the vector space of continuous functions on U), then*

$$\sup_B |\mathrm{Prob}\,(W_n \in B) - \textstyle\int_B \psi_{s,n}(v)\,dv| = o(n^{-(s-2)/2})\,, \tag{1.15}$$

where the supremum is over all Borel sets B.

(b) *If, instead of* (a), *it is merely assumed that*

$$\limsup_{|t|\to\infty} |E(\exp\{i\langle t, Z_1\rangle\})| < 1\,, \tag{1.16}$$

then the relation

$$\text{(1.17)} \qquad \sup_{B \in \mathscr{B}} |\text{Prob}\,(W_n \in B) - \int_B \psi_{s,n}(v)\,dv| = o(n^{-(s-2)/2})$$

holds uniformly over every class $\mathscr{B}$ *of Borel sets satisfying* (1.6).

REMARK 1.1. Theorems 1 and 2 extend in a straightforward manner to vector-valued $H(z) = (H_1(z), \cdots, H_p(z))$ provided that the dispersion matrix $M = \Sigma V \Sigma'$ of $(\langle Z_1, \text{grad } H_1(\mu)\rangle, \cdots, \langle Z_1, \text{grad } H_p(\mu)\rangle)$ is nonsingular. Here Σ is the $p \times k$ matrix whose rth row is grad $H_r(\mu) = (D_1 H_r(\mu), \cdots, D_k H_r(\mu))$. In this case one must replace $\psi_{s,n}$ by

$$\text{(1.18)} \qquad [1 + \sum_{r=1}^{s-2} n^{-r/2} \tilde{\pi}_r(-D)]\phi_M(x) \qquad x \in R^p,$$

where ϕ_M is the normal density on R^p with mean zero and dispersion M, $\tilde{\pi}_r$ is a polynomial in p variables (whose coefficients do not depend on n), and $-D = (-D_1, \cdots, -D_p)$. There is virtually no difference in the proofs for vector-valued H, apart from an additional complexity in notation.

REMARK 1.2. Let G denote the distribution of Y_1. If the density g, say, of the absolutely continuous part of G is such that $U_1 \equiv \{y : g(y) > 0\}$ is open and $G(U_1) = 1$, then one may replace (ii) in the statement of Theorem 2(a) by (ii)': $f_1, \cdots, f_k$ *are continuously differentiable* on U_1. For, in this case, the functions $1, f_1, \cdots, f_k$ are linearly dependent as continuous functions on U_1 if and only if $1, f_1(Y_1), \cdots, f_k(Y_1)$ are linearly dependent as elements of the L^2 space of random variables, and, as explained in the first paragraph of Section 2, one may always replace $\{1, f_1, \cdots, f_k\}$ by a maximal linearly independent set $\{1, f_{i_1}, \cdots, f_{i_{k'}}\}$ $(1 \leqq k' \leqq k)$.

REMARK 1.3. Assuming, in addition to the hypothesis of Theorem 2(a), that f_i's are analytic, Chibishov (1972) proved that an asymptotic expansion

$$\text{Prob}\,(W_n \in C) - \int_C [1 + \sum_{r=1}^{s-2} n^{-r/2} q_r(x)]\phi_M(x)\,dx = o(n^{-(s-2)/2})$$

holds uniformly over all measurable convex sets C (intervals, in case H is real). For the special case of polynomial H he was able to prove that this expansion was uniform over all Borel sets. For many applications (see, e.g., Theorem 3) analyticity of f_i's is a severe restriction. Also, he was not concerned with the problem of identifying this expansion with the formal Edgeworth expansion.

REMARK 1.4. Note that in Theorem 2 we only require $E|Z_1|^s < \infty$, whereas an algebraic computation of the moments of W_n' yields expressions for $\kappa_{j,n}$ $(1 \leqq j \leqq s)$ as polynomials in $n^{-\frac{1}{2}}$ whose coefficients are (polynomial) functions of moments of Z_1 of orders up to $s(s-1)$. This apparent anomaly is resolved by the fact that the "approximate cumulants" $\tilde{\kappa}_{j,n}$, $1 \leqq j \leqq s$, only involve moments (of Z_1) of orders s and less so that (1.14) is well defined. In the course of proving Theorem 2 it is first shown that under the hypothesis of Theorem 2(b) there exists an asymptotic expansion of the distribution function of W_n in the form

(1.19)
$$F_n(u) + o(n^{-(s-2)/2}),$$
$$F_n(u) = \int_{-\infty}^{u} [1 + \sum_{r=1}^{s-2} n^{-r/2} q_r(v)]\phi_{\sigma^2}(v)\, dv,$$

where q_r's are polynomials. The coefficients of q_r $(1 \leq r \leq s-2)$ are polynomials in the moments of Z_1 of orders s and less, and the coefficients of these last polynomials are constants which do not depend on the distribution of Z_1. It is next shown that, in case Z_1 has finite moments of all orders,

(1.20)
$$q_r(v)\phi_{\sigma^2}(v) = \pi_r\left(-\frac{d}{dv}\right)\phi_{\sigma^2}(v) \qquad 1 \leq r \leq s-2.$$

It follows that π_r's $(1 \leq r \leq s-2)$ depend only on those moments of Z_1 which are of orders s and less, and the same is, therefore, true of $\bar{k}_{j,n}$ $(1 \leq j \leq s)$. In view of (1.11)—(1.13), and (1.21) below, the jth moment of $\Psi_{s,n}$ $(j \geq 0)$ differs from that computed from $\bar{k}_{j,n}$ (using the familiar relations between moments and cumulants) by $o(n^{-(s-2)/2})$. In other words, under the hypothesis of Theorem 2(b) it is a valid procedure to compute moments of the asymptotic expansion by the so-called *delta method* in which W_n' is raised to a power, expectations taken term by term (formally) and terms of order $o(n^{-(s-2)/2})$ neglected. Expansions of moments as well as expectations of other smooth functions of W_n' (and of W_n, if it has enough moments) are valid *solely* under moment conditions on Z_1 (see Götze and Hipp (1977)), and these expansions may be obtained by integrating the smooth function with respect to the formal Edgeworth expansion $\Psi_{s,n}$, *even when the distribution function of W_n does not admit an expansion*. Finally, the proof of the identification (1.20) depends crucially on the following important combinatorial result of James (1955), (1958), and James and Mayne (1962):

(1.21)
$$\kappa_{j,n} = O(n^{-(j-2)/2}) \qquad j \geq 3,$$

which holds if $E|Z_1|^{j(s-1)} < \infty$. There may, however, be statistics whose cumulants satisfy (1.10), (1.11), but not (1.21). Consider such a statistic T_n, assume (for simplicity) that it has finite moments of all orders, and define, for each $r \geq 3$, the polynomials $\pi_{j,r}$ by

(1.22)
$$\exp\left\{it\bar{k}_{1,n} + \frac{(it)^2}{2}(\bar{k}_{2,n} - \sigma^2) + \sum_{j=3}^{r}\frac{(it)^j}{j!}\bar{k}_{j,n}\right\}\exp\{-\sigma^2t^2/2\}$$
$$= \exp\{-\sigma^2t^2/2\}[1 + \sum_{j=1}^{s-2} n^{-j/2}\pi_{j,r}(it)] + o(n^{-(s-2)/2})$$
$$= \hat{\psi}_{s,r,n}(t) + o(n^{-(s-2)/2}),$$

say. Define the *formal Edgeworth expansion of type* (r, s) by

(1.23)
$$\Psi_{s,r,n}(u) = \int_{-\infty}^{u}\left[1 + \sum_{j=1}^{s-2} n^{-j/2}\pi_{j,r}\left(-\frac{d}{dv}\right)\right]\phi_{\sigma^2}(v)\, dv.$$

It is easy to see from (1.22) that the polynomials $\pi_{j,r}$ have no constant terms, and

$\hat{\psi}_{s,r,n}(0) = 1$. It follows that there exists a *smallest integer* r_0 such that

$$\frac{d^j}{dt^j} \log \hat{\psi}_{s,r,n}(t)\Big|_{t=0} = o(n^{-(s-2)/2}) \qquad \text{if } j > r_0 . \tag{1.24}$$

If now the distribution function of the statistic T_n has a valid asymptotic expansion given by (1.19), then the same procedure as used in verifying (1.20) leads to the conclusion: $F_n = \Psi_{s,r,n}$ *if and only if* $r \geq r_0$.

REMARK 1.5. Theorem 2, incidentally, justifies the remark made in Ghosh and Subramanyam (1974), page 356, that their $E^{\prime}(T_n - \theta_0)^2$ is the second moment of an Edgeworth expansion.

AN APPLICATION. We now apply Theorem 2(a) for vector-valued H (see Remark 1.1) to obtain asymptotic expansions of distributions of a class of statistics including *maximum likelihood estimators* and the so-called *minimum contrast estimators* for vector parameters.

Let $\{Y_n\}_{n\geq 1}$ be a sequence of i.i.d. m-dimensional random vectors whose common distribution G_θ is parametrized by $\theta = (\theta^{(1)}, \cdots, \theta^{(p)})$ belonging to an open subset Θ of R^p. For each θ let $f(y; \theta)$ be an extended real-valued Borel measurable function on R^m. For nonnegative integral vectors $\nu = (\nu^{(1)}, \cdots, \nu^{(p)})$ write $|\nu| = \nu^{(1)} + \cdots + \nu^{(p)}$, $\nu! = \nu^{(1)}! \cdots \nu^{(p)}!$, and let $D^\nu = (D_1)^{\nu^{(1)}} \cdots (D_p)^{\nu^{(p)}}$ denote the νth *derivative with respect to* θ. We shall write P_θ to denote the product probability measure on the space of all sequences in R^m and regard Y_n's as coordinate maps on this space. Expectation with respect to P_θ will be denoted by E_θ. The following assumptions will be made:

(A_1) There is an open subset U of R^m such that (i) for each $\theta \in \Theta$ one has $G_\theta(U) = 1$, and (ii) for each ν, $1 \leq |\nu| \leq s + 1$, $f(y; \theta)$ has a νth derivative $D^\nu f(y; \theta)$ with respect to θ on $U \times \Theta$.

(A_2) For each compact $K \subset \Theta$ and each ν, $1 \leq |\nu| \leq s$, $\sup_{\theta_0 \in K} E_{\theta_0}|D^\nu f(Y_1; \theta_0)|^{s+1} < \infty$; and for each compact K there exists $\varepsilon > 0$ such that $\sup_{\theta_0 \in K} E_{\theta_0}(\max_{|\theta-\theta_0|\leq\varepsilon} |D^\nu f(Y_1; \theta)|)^s < \infty$ if $|\nu| = s + 1$.

(A_3) For each $\theta_0 \in \Theta$, $E_{\theta_0} D_r f(Y_1; \theta_0) = 0$ for $1 \leq r \leq p$, and the matrices

$$\begin{aligned} I(\theta_0) &= ((-E_{\theta_0} D_i D_r f(Y_1; \theta_0))) , \\ D(\theta_0) &= ((E_{\theta_0}(D_i f(Y_1; \theta_0) \cdot D_r f(Y_1; \theta_0)))) \end{aligned} \tag{1.25}$$

are nonsingular.

(A_4) The functions $I(\theta)$, $E_\theta(D^\nu f(Y_1; \theta) \cdot D^{\nu'} f(Y_1; \theta))$, $1 \leq |\nu|, |\nu'| \leq s$, are continuous on Θ.

(A_5) The map $\theta \to G_\theta$ on Θ into the space of all probability measures on (the Borel sigma field of) R^m is continuous when the latter space is given the (variation) norm topology.

(A_6) For each $\theta \in \Theta$, G_θ has a nonzero absolutely continuous component (with respect to Lebesgue measure) whose density has a version $g(y; \theta)$ which is strictly positive on U. Also, for each θ and each ν, $1 \leq |\nu| \leq s$, $D^\nu f(y; \theta)$ is continuously differentiable in y on U.

Now write

$$(1.26)\qquad L_n(\theta) = \sum_{j=1}^{n} f(Y_j;\theta)\,, \qquad L_1(\theta) = f(Y_1;\theta)\,,$$

and consider the p equations

$$(1.27)\qquad \begin{aligned} 0 &= \frac{1}{n} D_r L_n(\theta_0) + \frac{1}{n} \sum_{i=1}^{p} (\theta^{(i)} - \theta_0^{(i)}) D_i D_r L_n(\theta_0) + \cdots \\ &\quad + \frac{1}{n} \sum_{|\nu|=s-1} \frac{(\theta-\theta_0)^\nu}{\nu!} D^\nu D_r L_n(\theta_0) + R_{n,r}(\theta) \\ &= \frac{1}{n} D_r L_n(\theta)\,, \qquad\qquad 1 \leq r \leq p\,, \end{aligned}$$

where $x^\nu = (x^{(1)})^{\nu^{(1)}} \cdots (x^{(p)})^{\nu^{(p)}}$ for $x = (x^{(1)}, \cdots, x^{(p)}) \in R^p$, and $R_{n,r}(\theta)$ is the usual remainder in the Taylor expansion, so that

$$(1.28)\qquad |R_{n,r}(\theta)| \leq \frac{c(s,p)}{n} |\theta - \theta_0|^s \max_{|\nu|=s+1} \sup_{|\theta'-\theta_0| \leq |\theta-\theta_0|} |D^\nu L_n(\theta')|\,.$$

The statistics $\hat\theta_n$ considered below are measurable maps on the probability space into some compactification of Θ.

THEOREM 3.

(a) *Assume* (A_1)—(A_4) *hold for some* $s \geq 3$. *There exists a sequence of statistics* $\{\hat\theta_n\}_{n\geq 1}$ *such that for every compact* $K \subset \Theta$

$$(1.29)\qquad \begin{aligned} \inf_{\theta_0 \in K} P_{\theta_0}(|\hat\theta_n - \theta_0| < d_0 n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}},\ \hat\theta_n \text{ solves } (1.27)) \\ = 1 - o(n^{-(s-2)/2})\,, \end{aligned}$$

where d_0 *is a constant which may depend on* K.

(b) *If* (A_1)—(A_4) *hold, then there exist polynomials* q_{r,θ_0} *(in* p *variables), not depending on* n, *such that for every sequence* $\{\hat\theta_n\}_{n\geq 1}$ *satisfying* (1.29) *and every compact* $K \subset \Theta$ *one has the asymptotic expansion*

$$(1.30)\qquad \begin{aligned} \sup_{\theta_0 \in K} |P_{\theta_0}(n^{\frac{1}{2}}(\hat\theta_n - \theta_0) \in B) - \textstyle\int_B [1 + \sum_{r=1}^{s-2} n^{-r/2} q_{r,\theta_0}(x)] \phi_M(x)\, dx| \\ = o(n^{-(s-2)/2}) \end{aligned}$$

uniformly over every class $\mathscr{B}$ *of Borel sets of* R^p *satisfying*

$$(1.31)\qquad \sup_{\theta_0 \in K} \sup_{B \in \mathscr{B}} \int_{(\partial B)^\epsilon} \phi_M(x)\, dx = O(\epsilon) \qquad \text{as} \quad \epsilon \downarrow 0\,.$$

Here $M = I^{-1}(\theta_0) D(\theta_0) I^{-1}(\theta_0)$, *where* $I(\theta_0)$, $D(\theta_0)$ *are defined by* (1.25). *Also, the coefficients of the polynomials* q_{r,θ_0} *are themselves polynomials in the moments of* $D^\nu L_1(\theta_0)$, $1 \leq |\nu| \leq s$, *under* P_{θ_0}, *and are consequently bounded on compacts.*

REMARK 1.6. Theorem 3 is actually proved under the weaker hypothesis (A_1)—(A_3) and (in place of (A_4)) $(A_4)'$: *the distribution of* Z_1 *under* P_θ *satisfies Cramér's condition* (1.16), *for each* θ. Under this latter condition, and for one-dimensional parameters, relations similar to (1.30) were established (with analogous regularity assumptions) for the class of intervals, in place of general $\mathscr{B}$

satisfying (1.31), by Pfanzagl (1973b, Theorem 1) and Chibishov (1973b). Pfanzagl also provided a verifiable condition (see [21], page 1012) under which his distributional assumption may be checked. The situation is more complex in the multiparameter case. For this case Chibishov (1972, 1973a) was able to prove a result analogous to (1.30) for the special class of all measurable convex sets (which, of course, satisfies (1.31); see [4], page 24) under the additional assumption that $D^\nu f(y;\theta)$, $1 \leqq |\nu| \leqq s$, be analytic in y. In the present context this assumption is severely restrictive. Note that assumption (A_4) provides a simple verifiable sufficient condition for the validity of $(A_4)'$ (see Lemma 2.2 and Remark 1.2). Finally, it is also possible (see the proof in Section 2) to replace the continuity conditions in (A_4) by 'boundedness' conditions (as, e.g., in Pfanzagl (1973a)).

REMARK 1.7. Under assumptions (A_1)—(A_4) with $s = 3$ one may easily prove (using Theorem 1 for vector H, instead of Theorem 2) that the error of normal approximation is $O(n^{-\frac{1}{2}})$ uniformly over every compact $K \subset \Theta$ and every class $\mathscr{B}$ satisfying (1.31). However, for the special class of all Borel measurable convex sets such a result has been proved by Pfanzagl (1973b).

REMARK 1.8. Assume that for some $s \geqq 2$ one has (A_1), (A_2), $E_{\theta_0}|D^\nu f(Y_1;\theta_0)|^s < \infty$ for $1 \leqq |\nu| \leqq s$, and $E_{\theta_0}(\max_{|\theta-\theta_0|\leq\epsilon} |D^\nu f(Y_1;\theta)|)^s < \infty$ for some $\epsilon > 0$ and all ν with $|\nu| = s + 1$. Then one may prove using (1.27), (1.28) and the law of the iterated logarithm that there exists an a.s. (P_{θ_0}) finite integer-valued random variable $N(\cdot)$ such that with P_{θ_0}- probability one for $n > N(\cdot)$ one has

$$(1.32)\qquad \left|\frac{1}{n} D_r L_n(\theta_0)\right| \leqq d_1 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\,,$$

$$\left|\frac{1}{n} D^\nu L_n(\theta_0) - E_{\theta_0} D^\nu f(Y_1;\theta_0)\right| \leqq d_1 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} \qquad 2 \leqq |\nu| \leqq s\,,$$

$$|R_{n,r}(\theta)| \leqq |\theta - \theta_0|^s(d_2 + d_1 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}})$$
$$\text{for all}\quad \theta \quad \text{satisfying} \quad |\theta - \theta_0| \leqq \epsilon \qquad 1 \leqq r \leqq p\,,$$

for any positive constant d_1 and a suitable constant d_2. Using the Brouwer fixed point theorem, as in the proof of Theorem 3(a), one can then show that there exists a sequence of statistics $\{\hat\theta_n\}_{n\geq1}$ such that for every $d > 0$ with P_{θ_0}- probability one

$$(1.33)\qquad |\hat\theta_n - \theta_0| < dn^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} \qquad \text{and} \quad \hat\theta_n \quad \text{solves (1.27) if} \quad n > N(\cdot)\,.$$

If, due to some additional structure (e.g., convexity or concavity of $L_n(\theta)$ as a function θ for every n, a.s. (P_{θ_0})), the equations (1.27) have at most one solution for each n (a.s. (P_{θ_0})), then of course one may define $\hat\theta_n$ to be this solution when it exists and arbitrarily (measurably) if it does not, and such a $\hat\theta_n$ will satisfy (1.33) with P_{θ_0}- probability one (*strong consistency*) and, under the hypothesis (A_1)—(A_4) will also admit the asymptotic expansion (1.30). Finally, we consider the so-called minimum contrast estimators (see Pfanzagl (1973b)). It is known (see [21], Lemma 3, which admits extension to $p > 1$) that for such estimators $\hat{\theta}_n$, say, one has, under certain regularity conditions,

$$\sup_{\theta_0 \in K} P_{\theta_0}(|\hat{\theta}_n - \theta_0| > d'(\theta_0) n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) = o(n^{-(s-2)/2}) \tag{1.34}$$

for every compact $K \subset \Theta$. Here d' is bounded on compacts. Since $\hat{\theta}_n$ minimizes (or maximizes) $L_n(\theta)$ it follows that (1.29) holds. Augmenting these regularity conditions, if necessary, so that (A_1)—(A_4) hold one has (1.30). Conditions not significantly different from (A_1)—(A_4) are generally included among these regularity conditions. Finally, the reason for not restricting the context of Theorem 3 to minimum contrast estimators is that in its present from this theorem also applies to problems, e.g., in mathematical economics (see Bhattacharya and Majumdar (1973)), in which $\hat{\theta}_n$ is not a statistical estimator.

Among the earliest results on asymptotic expansion of some special functions of sample moments we refer to Hsu (1945) who obtained an asymptotic expansion for the sample variance.

For relations with questions concerning asymptotic efficiencies of statistical estimators we refer to Pfanzagl (1973a), Ghosh and Subramanyam (1974), and Ghosh, Sinha and Wieand (1977).

Some of the results of this article in weaker form were announced earlier in Bhattacharya (1977). It may be noted that the entire Section 4 of that article ([2]) was based on joint work by the authors.

2. Proofs. For proving Theorems 1 and 2 it will be assumed, without any essential loss of generality, that the *dispersion matrix* V of Z_n *is nonsingular*. For, if V is singular, then $1, f_1(Y_n), \cdots, f_k(Y_n)$ are linearly dependent when considered as elements of the L^2 space of random variables. Then there exist a maximal integer k' and distinct indices $i_1, \cdots, i_{k'}$ among $1, 2, \cdots, k$ such that $1, f_{i_1}, \cdots, f_{i_{k'}}$ are linearly independent. Defining $\tilde{Z}_n = (f_{i_1}(Y_n), \cdots, f_{i_{k'}}(Y_n))$ one can define a function H' defined on $R^{k'}$ and as smooth as H such that $H'(\bar{\tilde{Z}}) = H(\bar{Z})$ where $\bar{\tilde{Z}} = (1/n) \sum_{i=1}^n \tilde{Z}_i$. In view of the positivity of σ^2, $k' \geqq 1$.

Throughout the letters c, d will denote constants (i.e., nonrandom numbers not depending on n, x, z, u, or v).

Let $\chi_j(t)$ denote the jth *cumulant of* $\langle t, Z_1 - \mu \rangle = \sum_{r=1}^k t^{(r)}(Z_1 - \mu)^{(r)}$, and introduce the *Cramér–Edgeworth polynomials*

$$\tilde{P}_r(it) = \sum_{\nu=1}^r \left\{ \sum{}^* \frac{\chi_{j_1+2}(it)}{(j_1 + 2)!} \cdots \frac{\chi_{j_\nu+2}(it)}{(j_\nu + 2)!} \right\} \tag{2.1}$$

$$\chi_j(it) = i^j \chi_j(t) \qquad t \in R^k;\ r = 1, 2, \cdots,$$

where the sum $\sum^*$ is over all p-tuples of positive integers $(j_1, \cdots, j_p)$ satisfying $\sum_{i=1}^p j_i = r$. Letting D_i denote *differentiation with respect to the* ith *coordinate*, write

$$D = (D_1, \cdots, D_k). \tag{2.2}$$

Then $\tilde{P}_r(-D)$ is a differential operator. Write

(2.3) $$\phi_V(z) = (2\pi)^{-k/2}(\det V)^{-1/2}\exp\{-\tfrac{1}{2}\langle z, V^{-1}z\rangle\}\,,$$
$$\xi_{s,n}(z) = [1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2}\tilde{P}_r(-D)]\phi_V(z) \qquad z \in R^k\,.$$

Define the functions

$$g_n(z) = n^{1/2}[H(\mu + n^{-1/2}z) - H(\mu)]\,, \qquad h_1(z) = \textstyle\sum_{i=1}^{k} l_i z^{(i)}\,,$$

(2.4) $$h_{s-1}(z) = \textstyle\sum l_i z^{(i)} + \frac{1}{2}n^{-1/2}\sum l_{i,j} z^{(i)}z^{(j)} + \cdots + \frac{1}{(s-1)!}\, n^{-(s-2)/2}\sum l_{i_1,\dots,i_{s-1}} z^{(i_1)}\cdots z^{(i_{s-1})}$$
$$z = (z^{(1)}, \cdots, z^{(k)}) \in R^k\,.$$

Note that h_{s-1} is a Taylor expansion of g_n and write

(2.5) $$W_n = g_n(n^{-1/2}(Z-\mu))\,, \qquad W_n' = h_{s-1}(n^{-1/2}(Z-\mu))\,.$$

Define the maps

(2.6) $$T(z) = (z^{(1)}, \cdots, z^{(k-1)}, g_n(z))\,, \qquad T_p(z) = (z^{(1)}, \cdots, z^{(k-1)}, h_p(z))$$

where $p = 1$ or $s-1$. *Assume* without loss of generality that $l_k > 0$. For the following discussion n_0 is an integer such that for $n > n_0$ the map $T_{s-1}(T)$ is a $C^\infty(C^s)$ diffeomorphism on the set

(2.7) $$M_n = \{|z| < ((s-1)\Lambda \log n)^{1/2}\}$$

onto its image. Here Λ is the *largest eigenvalue of* V.

LEMMA 2.1. *Assume* $\rho_s = E|Z_1|^s < \infty$ *and that all derivatives of* H *of orders* s *and less are continuous in a neighborhood of* $\mu = EZ_1$, *for some* $s \geq 3$. *Then there exist polynomials* q_r *(in* one variable), *whose coefficients do not depend on* n, *such that uniformly over all Borel subsets* B *of* R^1 *one has*

(2.8) $$\int_{\{g_n(z)\in B\}} \xi_{s,n}(z)\,dz = \int_B dF_n(u) + o(n^{-(s-2)/2})\,,$$

where

(2.9) $$F_n(u) = \int_{-\infty}^{u} [1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2}q_r(v)]\phi_{\sigma^2}(v)\,dv \qquad u \in R^1\,.$$

Also, for all nonnegative integers j

(2.10) $$\int_{M_n} g_n^j(z)\xi_{s,n}(z)\,dz = \int_{-\infty}^{\infty} u^j\,dF_n(u) + o(n^{-(s-2)/2})\,,$$
$$\int_{R^k} h_{s-1}^j(z)\xi_{s,n}(z)\,dz = \int_{-\infty}^{\infty} u^j\,dF_n(u) + o(n^{-(s-2)/2})\,.$$

PROOF. By the change of variables $x = T_1^{-1}T(z)$, the first integral in (2.8), when restricted to the set M_n, becomes

(2.11) $$\int_{\{h_1(x)\in B\}\cap T_1^{-1}T(M_n)} \xi_{s,n}(T^{-1}T_1(x))[l_k/D_k g_n(T^{-1}T_1(x))]\,dx\,.$$

Now the elements of the Jacobian matrix of $T(z)$ and those of the inverse of this matrix, as well as their derivatives of orders $s-1$ and less, are bounded on M_n by constants independent of n. Hence a Taylor expansion yields

(2.12) $$(T^{-1}T_1(x))^{(k)} - x^{(k)} = (T^{-1}T_1(x))^{(k)} - (T^{-1}T(x))^{(k)}$$
$$= \textstyle\sum_{r=1}^{s-2} n^{-r/2}p_r(x) + R(|x|)\cdot o(n^{-(s-2)/2})\,,$$

where p_r's are polynomials in k variables and R is a polynomial in one variable whose coefficients do not depend on n; and the factor $o(n^{-(s-2)/2})$ does not involve x. Using (2.12) and the fact that $(T^{-1}T_1(x))^{(i)} = x^{(i)}$ for $1 \leq i \leq k-1$, one reduces (2.11) to

$$\int_{\{h_1(x) \in B\} \cap T_1^{-1}T(M_n)} [1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2} p_r'(x)] \phi_V(x)\, dx + o(n^{-(s-2)/2}), \tag{2.13}$$

where p_r''s are polynomials (in k variables) whose coefficients do not depend on n. Since $T_1^{-1}T(M_n) \supset \{|x| < ((s-\frac{3}{2})\Lambda \log n)^{1/2}\}$ if $n > n_0$, (2.13) reduces to

$$\int_{\{h_1(x) \in B\}} [1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2} p_r'(x)] \phi_V(x)\, dx + o(n^{-(s-2)/2}).$$

Recall that $h_1(x) = \sum l_j x^{(j)} = \langle l, x\rangle$ and write

$$G_n(u) = \int_{\{\langle l, x\rangle \leq u\}} [1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2} p_r'(x)] \phi_V(x)\, dx \qquad u \in R^1.$$

The Fourier–Stieltjes transform of G_n is

$$[1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2} p_r'(-iD)] \phi_V(tl) = [1 + \textstyle\sum_{r=1}^{s-2} n^{-r/2} q_r'(it)] \exp\left\{-\frac{\sigma^2 t^2}{2}\right\}$$

where q_r''s are polynomials (in one variable) whose coefficients do not depend on n. Define

$$q_r(v) = \left[q_r'\left(-\frac{d}{dv}\right) \phi_{\sigma^2}(v)\right] \Big/ \phi_{\sigma^2}(v)$$

to complete the proof of (2.8). The first relation in (2.10) is proved in the same manner, while the second follows from the first and the inequalities

$$\begin{aligned} &\sup_{z \in M_n} |g_n^j(z) - h_{s-1}^j(z)| \leq d_1 n^{-(s-1)/2} (\log n)^{s/2}, \\ &\int_{\{z \notin M_n\}} h_{s-1}^j(z) \xi_{s,n}(z)\, dz = o(n^{-(s-2)/2}) \qquad j \geq 0. \end{aligned} \tag{2.14}$$

$\square$

Proof of Theorem 1. Let Q_n denote the *distribution of* $n^{1/2}(Z - \mu)$ and let Φ_V be the *k-variate normal distribution with mean zero and dispersion matrix* V. It follows from a recent result of Sweeting (1977), Corollary 3 (also see [4], pages 160–162) that

$$\begin{aligned} &|Q_n(A) - \Phi_V(A)| \leq c_1 \lambda^{-3/2} \rho_3 n^{-1/2} + c_2 \Phi_V((\partial A)^{\epsilon_n}), \\ &\epsilon_n = c_3 \Lambda^{1/2} \lambda^{-3/2} \rho_3 n^{-1/2} \qquad \rho_3 = E|Z_1|^3. \end{aligned} \tag{2.15}$$

Here λ is the smallest (and Λ the largest) eigenvalue of V. Fix $B \in \mathscr{B}'$, where $\mathscr{B}'$ satisfies (1.6), and in (2.15) take

$$A = \{z \in R^k : g_n(z) \in B\}. \tag{2.16}$$

Since g_n is continuous,

$$\partial A \subset \{z \in R^k : g_n(z) \in \partial B\}. \tag{2.17}$$

Now if $z \in (\partial A)^\epsilon$, then there exists z' such that $g_n(z') \in \partial B$ and $|z - z'| < \epsilon$. If, in addition, $z \in M_n$ (see (2.7)), then $|g_n(z) - g_n(z')| \leq d'\epsilon$, where d' is an upper bound of $|\text{grad } g_n|$ on M_n^ϵ (the ϵ-neighborhood of M_n). Since the Φ_V-probability of the complement of M_n is $o(n^{-(s-2)/2})$, it follows that

$$\Phi_V((\partial A)^\epsilon) \leq \Phi_V(\{g_n(z) \in (\partial B)^{d'\epsilon}\}) + o(n^{-(s-2)/2}) \qquad 0 < \epsilon \leq 1. \tag{2.18}$$

But by Lemma 2.1 (relation (2.8)) one has

$$\begin{aligned}\Phi_V(\{g_n(z)\in(\partial B)^{d'\epsilon}\}) &= \int_{\{g_n\in(\partial B)^{d'\epsilon}\}} \xi_{s,n}(z)\,dz + o(n^{-(s-2)/2}) \\ &= \int_{(\partial B)^{d'\epsilon}} \phi_{0,V}(v)\,dv + o(n^{-(s-2)/2}) \qquad (2.19)\\ &= O(\epsilon) + o(n^{-(s-2)/2})\end{aligned}$$

if $\rho_s = E|Z_1|^s$ is finite. Taking $s = 3$ and using (1.6), (2.18) and (2.19) the right side of (2.15) is estimated as $O(n^{-\frac{1}{2}})$ uniformly over $\mathscr{B}$. Again use Lemma 2.1, this time for B itself, to complete the proof of Theorem 1.

Proof of Theorem 2. We first prove part (b) of Theorem 2. From a general result on asymptotic expansion under Cramer's condition (1.16) (see [4], Corollary 20.2, page 214) and the estimates (2.18), (2.19) it follows that

$$\sup_{B\in\mathscr{B}} |Q_n(A) - \int_A \xi_{s,n}(z)\,dz| = o(n^{-(s-2)/2}) \qquad (2.20)$$

where $\mathscr{B}$ satisfies (1.6) and A is defined by (2.16). Now use Lemma 2.1 to estimate the integral. It remains to identify F_n and $\Psi_{s,n}$ (see (1.14)). First assume that Z_1 is *bounded*. Since $W_n' = h_{s-1}(n^{\frac{1}{2}}(\bar{Z} - \mu))$ is a polynomial in $n^{\frac{1}{2}}(\bar{Z} - \mu)$ it follows from the asymptotic expansions of moments of Q_n, i.e., of the derivatives of its characteristic function at zero (see [4], Theorem 9.9, page 77), that

$$EW_n'^j = \int_{R^k} h_{s-1}^j(z)\xi_{s,n}(z)\,dz + o(n^{-(s-2)/2}) \qquad j \geqq 0\,. \qquad (2.21)$$

By Lemma 2.1 (second relation in (2.10)) one then has

$$EW_n'^j = \int_{-\infty}^{\infty} u^j\,dF_n(u) + o(n^{-(s-2)/2}) \qquad j \geqq 0\,. \qquad (2.22)$$

On the other hand, the expression (1.12) differs from $\tilde{\psi}_{s,n}$ by $o(n^{-(s-2)/2})$ uniformly on a compact neighborhood of zero, say $\{|t| \leqq 1\}$. Also, according to a result due to James (1955), (1958), and James and Mayne (1962), the cumulants of W_n' satisfy

$$\kappa_{j,n} = O(n^{-(j-2)/2}) \qquad j \geqq 3\,, \qquad (2.23)$$

so that, the "approximate cumulants" $\bar{\kappa}_{j,n}$ (see (1.11)) satisfy

$$\bar{\kappa}_{j,n} = \kappa_{j,n} + o(n^{-(s-2)/2}) \qquad j \geqq 1\,, \qquad (2.24)$$

taking $\bar{\kappa}_{j,n} = 0$ for $j > s$. Hence (1.12) differs from the characteristic function of W_n' by $o(n^{-(s-2)/2})$ uniformly on $\{|t| \leqq 1\}$. Therefore,

$$\sup_{|t|\leqq 1} |\tilde{\psi}_{s,n}(t) - E(\exp\{itW_n'\})| = o(n^{-(s-2)/2})\,. \qquad (2.25)$$

By the familiar inequality of Cauchy for derivatives of analytic functions, derivatives of $\tilde{\psi}_{s,n}$ at zero differ from those of $E(\exp\{itW_n'\})$ by $o(n^{-(s-2)/2})$, proving

$$EW_n'^j = \int_{-\infty}^{\infty} u^j\,d\Psi_{s,n}(u) + o(n^{-(s-2)/2}) \qquad j \geqq 0\,. \qquad (2.26)$$

Together (2.22) and (2.26) imply

$$\int_{-\infty}^{\infty} u^j\,dF_n(u) - \int_{-\infty}^{\infty} u^j\,d\Psi_{s,n}(u) = o(n^{-(s-2)/2}) \qquad j \geqq 0\,. \qquad (2.27)$$

Since neither F_n nor $\Psi_{s,n}$ involve terms of order $o(n^{-(s-2)/2})$,

$$\int_{-\infty}^{\infty} u^j\,dF_n(u) = \int_{-\infty}^{\infty} u^j\,d\Psi_{s,n}(u) \qquad j \geqq 0\,. \qquad (2.28)$$

Now the Fourier–Stieltjes transforms of F_n and $\Psi_{s,n}$ are (extendable to) entire functions on the complex plane whose values and derivatives of all orders coincide at the origin. Hence $F_n = \Psi_{s,n}$, completing the proof of Theorem 2(b) in case Z_1 is bounded. We now proceed with the general case. Recall the polynomials κ_r defined by (1.13) and write

$$\text{(2.29)} \qquad \bar{q}_r(v) = \left[\kappa_r\left(-\frac{d}{dv}\right)\phi_{0,V}(v)\right]\Big/\phi_{0,V}(v)$$
$$= \text{coeff. of } \quad n^{-r/2} \quad \text{in} \quad \psi_{s,n}\,.$$

Both q_r and $\bar{q}_r$ are polynomials in the *cumulants of* Z_1 *of orders* s *and less.* Denoting the vector of all these cumulants by γ_s write $q_r(\gamma_s)$, $\bar{q}_r(\gamma_s)$ to denote this functional dependence. For $c > 0$ define the truncated random vector Z_{1c} to be equal to Z_1 if $|Z_1| \leq c$ and zero if $|Z_1| > c$. We can choose c so large that the characteristic function of Z_{1c} satisfies Cramér's condition (1.16). Let $\gamma_{s,c}$ denote the vector of all cumulants of Z_{1c} of orders s and less. Since Z_{1c} is a bounded random vector, $q_r(\gamma_{s,c}) = \bar{q}_r(\gamma_{s,c})$. Since $\gamma_{s,c} \to \gamma_s$ as $c \to \infty$ (and q_r, $\bar{q}_r$ are continuous in γ_s), one gets $q_r(\gamma_s) = \bar{q}_r(\gamma_s)$. Proof of Theorem 2(b) is complete.

In order to prove Theorem 2(a) it is now enough to show that, under the given hypothesis,

$$\text{(2.30)} \qquad \text{Prob}\,(n^{\frac{1}{2}}(\bar{Z} - \mu) \in A) = \textstyle\int_A \xi_{s,n}(z)\,dz + o(n^{-(s-2)/2})$$

uniformly over all Borel subsets A of R^k. By a result of Bikjalis (1968) this will follow if we can show that there exists an integer p such that $Z_1 + \cdots + Z_p$ has a nonzero absolutely continuous component with respect to Lebesgue measure on R^k. The following result shows that this is true with $p = k$.

LEMMA 2.2. *Assume that* G *has a nonzero absolutely continuous component (with respect to Lebesgue measure on* R^m*) whose density is positive on some open ball* B *in which the functions* f_i $(1 \leq i \leq k)$ *are continuously differentiable and in which* 1, $f_1, \cdots, f_k$ *are linearly independent as elements of the vector space of continuous functions on* B. *Then* Q_1^{*k} *has a nonzero absolutely continuous component.*

PROOF. To show that the distribution of $Z_1 + \cdots + Z_k = (\sum_1^k f_1(Y_j), \cdots, \sum_1^k f_k(Y_j))$ has a nonzero absolutely continuous component under the given hypothesis define the map (on R^{mk} into R^k)

$$F(y_1, \cdots, y_k) = (\textstyle\sum_1^k f_1(y_j), \cdots, \sum_1^k f_k(y_j))$$
$$y_j = (y_j^{(1)}, \cdots, y_j^{(m)}) \in R^m,\ 1 \leq j \leq k\,.$$

The Jacobian matrix of this map will be denoted by $J_{k,m}$. This matrix may be displayed as $J_{k,m} = [A_1 A_2 \cdots A_k]$, where A_j is a $k \times m$ matrix whose ith row is $(\text{grad} f_i)(y_j)$. Clearly, it is enough to show that $J_{k,m}$ has rank k at some $(y_1, \cdots, y_k)$ with y_j in the open ball B for all j. We shall prove this by induction on k (keeping m fixed). Suppose then, as induction hypothesis, that $J_{k_0-1,m}(a_1, \cdots, a_{k_0-1})$ has rank $k_0 - 1$ for some $k_0 - 1 \geq 1$ and for some $(a_1, \cdots, a_{k_0-1})$ with a_j in B for all j. Note that the submatrix formed by the first $(k_0 - 1)$ rows and

$(k_0 - 1)m$ columns of $J_{k_0,m}(a_1, \cdots, a_{k_0-1}, y)$ is $J_{k_0-1,m}(a_1, \cdots, a_{k_0-1})$, while its last m columns are given by $A_{k_0}(y)$, and the first $(k_0 - 1)m$ elements of its last row are formed by grad $f_{k_0}(a_1), \cdots,$ grad $f_{k_0}(a_{k_0-1})$.

Let $E_1, \cdots, E_{k_0-1}$ be $(k_0 - 1)$ linearly independent columns among the first $(k_0 - 1)m$ columns of $J_{k_0,m}$ (which exist by the induction hypothesis). Let $C_1, C_2, \cdots, C_m$ be the $(k_0 \times k_0)$ submatrices of $J_{k_0,m}$ formed by augmenting $E_1, E_2, \cdots, E_{k_0-1}$ by the first, second, $\cdots$, mth columns of $A_{k_0}(y)$, respectively. If rank of $J_{k_0,m}(a_1, \cdots, a_{k_0-1}, y)$ is less than k_0 for all y in B, then the determinants of $C_1, \cdots, C_m$ must vanish for all y in B, i.e.,

$$d_1 \frac{\partial f_1(y)}{\partial y^{(i)}} + \cdots + d_{k_0} \frac{\partial f_{k_0}(y)}{\partial y^{(i)}} = 0 \qquad \text{for} \quad i = 1, \cdots, m\,, \quad \text{and} \quad y \in B\,.$$

Here d_j is $(-1)^j$ times the determinant of the submatrix of $J_{k_0,m}$ comprising the columns $E_1, \cdots, E_{k_0-1}$ minus the jth row. Since $d_{k_0} \neq 0$, by induction hypothesis, the above relations are equivalent to saying that the gradient of (the nonzero linear combination) $\sum_1^{k_0} d_j f_j(y)$ vanishes identically in B. This means that $\sum d_j f_j$ is constant on every line segment contained in B; since B is connected, this means that there exists a number d_0 such that $\sum_1^{k_0} d_j f_j(y) = d_0$ for all y in B contradicting the hypothesis of linear independence of $1, f_1, \cdots, f_{k_0}$ in B. Hence there must exist a_{k_0} in B such that $J_{k_0,m}(a_1, \cdots, a_{k_0-1}, a_{k_0})$ has rank k_0. The proof is now completed by noting that the hypothesis of linear independence of $1, f_1$ in B implies that grad f_1 does not vanish identically in B, so that the induction hypothesis is true for $k_0 - 1 = 1$. □

The above lemma improves Lemma 1.4 in [2]. The main idea behind the proof is contained in Dynkin (1951), Theorem 2.

Proof of Theorem 3. We shall need an estimate of tail probabilities due to von Bahr (1967). Let $\{Z_n\}_{n \geq 1}$ be a sequence of i.i.d. random vectors each with mean μ and dispersion matrix V. Let Λ denote the largest eigenvalue of V. Then, if $E|Z_1|^s < \infty$ for some integer $s \geq 3$,

$$\text{(2.31)} \qquad \text{Prob}\,(|n^{\frac{1}{2}}(\bar{Z} - \mu)| > ((s - 1)\Lambda \log n)^{\frac{1}{2}}) \leq dn^{-(s-2)/2}(\log n)^{-s/2}$$

where $\bar{Z} = n^{-1}(Z_1 + \cdots + Z_n)$, and d is bounded on any bounded set of values of Λ.

Fix $\theta_0 \in \Theta$. In view of (2.31), the assumptions (A_1)—(A_4) and inequality (1.28) imply that there are constants d_1, d_2, d_3 such that

$$P_{\theta_0}\left(\left|\frac{1}{n} D_r L_n(\theta_0)\right| > d_1 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\right) \leq d_2(\log n)^{-s/2} n^{-(s-2)/2} \qquad 1 \leq r \leq p\,,$$

$$\text{(2.32)} \qquad P_{\theta_0}\left(\left|\frac{1}{n} D^\nu D_r L_n(\theta_0) - E_{\theta_0} D^\nu D_r L_1(\theta_0)\right| > d_1 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\right) \leq d_2(\log n)^{-s/2} n^{-(s-2)/2} \qquad 1 \leq |\nu| \leq s - 1\,,$$

$$P_{\theta_0}(|R_{n,r}(\theta)| > |\theta - \theta_0|^s (d_3 + d_1 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}})) \leq d_2(\log n)^{-s/2} n^{-(s-2)/2}\,.$$

Therefore, on a set having P_{θ_0}- probability at least $1 - d_4(\log n)^{-s/2}n^{-(s-2)/2}$ one may rewrite (1.27) as

$$(2.33) \qquad (\theta - \theta_0) = (I(\theta_0) + \eta_n)^{-1}\Big[\delta_n + \textstyle\sum_{2\leq|\nu|\leq s-1} \frac{1}{\nu!}(\theta - \theta_0)^\nu E_{\theta_0} D^\nu D_r L_1(\theta_0)$$
$$+ d_3|\theta - \theta_0|^s \varepsilon_n\Big],$$

where η_n is a random matrix and δ_n is a random vector each having norm less than $d_5 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$ and ε_n is a random vector of norm less than one. Note that there exists a sufficiently large positive constant d_6 and a (nonrandom) integer n_0 such that if $n > n_0$ and $|\theta - \theta_0| \leq d_6 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$, the right side of (2.33) is less than $d_6 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$. It then follows from the Brouwer fixed point theorem (see Milnor (1965), page 14) applied to the expression on the right side of (2.33) (regarded as a function of $\theta - \theta_0$) that there exists a statistic $\hat{\theta}_n$ such that

$$(2.34) \qquad P_{\theta_0}(|\hat{\theta}_n - \theta_0| < d_6 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}},\ \hat{\theta}_n \text{ solves } (1.27))$$
$$\geq 1 - d_4(\log n)^{-s/2}n^{-(s-2)/2}$$

To obtain an asymptotic expansion of the distribution of $\hat{\theta}_n$, first define

$$(2.35) \qquad f_\nu(y) = D^\nu \log f(y; \theta_0), \qquad Z_n^{(\nu)} = f_\nu(Y_n) \qquad 1 \leq |\nu| \leq s$$

Consider the random vectors $Z_n = (Z_n^{(\nu)})_{1\leq|\nu|\leq s}$ whose coordinates are indexed by ν's. The dimension of Z_n is $k = \sum_{j=1}^{s} \binom{p+j-1}{j}$. From the definition of $\hat{\theta}_n$ one has, outside a set of probability at most $o(n^{-(s-2)/2})$,

$$(2.36) \qquad 0 = \frac{1}{n} D_r L_n(\hat{\theta}_n) = Z^{(e_r)} + \textstyle\sum_{|\nu|=1}^{s-1} \frac{1}{\nu!} Z^{(e_r+\nu)}(\hat{\theta}_n - \theta_0)^\nu + R_{n,r}(\hat{\theta}_n)$$
$$1 \leq r \leq p,$$

where the rth coordinate of e_r is one and other coordinates zero. Now consider the p equations

$$(2.37) \qquad 0 = z^{(e_r)} + \textstyle\sum_{|\nu|=1}^{s-1} \frac{1}{\nu!} z^{(e_r+\nu)}(\theta - \theta_0)^\nu \equiv P(\theta, z; r) \qquad 1 \leq r \leq p,$$

in the $p + k$ variables θ, z. These equations have a solution at $\theta = \theta_0$, $z = \mu$, where $\mu = EZ_1$, i.e.,

$$(2.38) \qquad \mu^{(e_r)} = 0 \qquad 1 \leq r \leq p,$$
$$\mu^{(\nu)} = E_{\theta_0} D^\nu \log f(Y_1; \theta_0) \qquad 2 \leq |\nu| \leq s$$

Also, since $I(\theta_0)$ is nonsingular, the p vectors $(D_1 P(\theta_0, \mu; r)), \cdots, (D_p P(\theta_0, \mu; r))$, $1 \leq r \leq p$, are linearly independent. Therefore, by the implicit function theorem, there is a neighborhood N of μ and p uniquely defined real-valued infinitely differentiable functions H_i $(1 \leq i \leq p)$ on N such that $\theta = H(z) = (H_1(z), \cdots, H_p(z))$ satisfies (2.37) for $z \in N$, and $\theta_0 = H(\mu)$. By (2.32), $|Z^{(e_r)} + R_{n,r}(\hat{\theta}_n)| < d_7 n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$ with P_{θ_0}- probability $1 - o(n^{-(s-2)/2})$. Therefore, by (2.36) and the uniqueness part of the implicit function theorem, with P_{θ_0}- probability $1 - o(n^{-(s-2)/2})$ one has

$$\hat{\theta}_n = H(\mathbf{Z}') \quad \text{with} \quad \mathbf{Z}'^{(\nu)} = \mathbf{Z}^{(\nu)} \quad \text{for} \quad 2 \leqq |\nu| \leqq s\,,$$

$$(2.39) \qquad\qquad = \mathbf{Z}^{(e_r)} + R_{n,r}(\hat{\theta}_n) \quad \text{for} \quad \nu = e_r\,, \quad 1 \leqq r \leqq p\,.$$

Therefore, by (2.32) and (2.34), there are constants d_8, d_9 such that

$$P_{\theta_0}(|n^{\frac{1}{2}}[H(\mathbf{Z}) - H(\mu)] - n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)| \leqq d_8(\log n)^{s/2} n^{-(s-1)/2})$$

$$(2.40) \qquad = P_{\theta_0}(|H(\mathbf{Z}') - H(\mathbf{Z})| = |R_{n,r}(\hat{\theta}_n)| \leqq d_8(\log n)^{s/2} n^{-s/2})$$

$$\geqq 1 - d_9(\log n)^{-s/2} n^{-(s-2)/2}\,.$$

In view of (A_6) (and Remark 1.2) Lemma 2.2 applies, so that Theorem 2 yields, for vector H (see Remark 1.1),

$$(2.41) \qquad P_{\theta_0}(n^{\frac{1}{2}}[H(\mathbf{Z}) - H(\mu)] \in B) = \int_B \psi_{s,n}(x)\,dx + o(n^{-(s-2)/2})$$

uniformly over all Borel sets B. Here $\psi_{s,n}$ is given by (1.18) with $M = I^{-1}(\theta_0)D(\theta_0)I^{-1}(\theta_0)$, where $I(\theta_0)$ and $D(\theta_0)$ are defined by (1.25). This evaluation of M follows from (2.33), (2.36), or, alternatively, from a computation of grad $H_r(\mu)$, $1 \leqq r \leqq p$, obtained from inverting the Jacobian matrix (at (θ_0, μ)) of the transformation whose first p coordinate functions are given by the right side of (2.37) and the remaining coordinate functions by $z^{(\nu)}$, $1 \leqq |\nu| \leqq s$. Finally, if $\mathscr{B}$ satisfies (1.31), then it is simple to check that

$$(2.42) \qquad \sup_{B \in \mathscr{B}} \int_{(\partial B)^\epsilon} |\psi_{s,n}(x)|\,dx \leqq d_{10}\epsilon + o(n^{-(s-2)/2}) \qquad 0 \leqq \epsilon \leqq 1\,.$$

Relations (2.40)—(2.42), with $\epsilon = d_8(\log n)^{s/2} n^{-(s-1)/2}$, now complete the proof excepting for the uniformity over compacts. By assumptions (A_1)—(A_6), the constants d_8, d_9, d_{10} are bounded on compact K (since so are d_1—d_7). The term $o(n^{-(s-2)/2})$ in (2.41) is uniform on compact K for $B \in \mathscr{B}$ due to the uniformity of the error of approximation of the distribution Q_n of $n^{\frac{1}{2}}(\mathbf{Z} - \mu)$ by its Edgeworth expansion, assuming, without loss of generality (see Remark 1.2), that the dispersion matrix of Z_1 is nonsingular. Note that we have only made use of (2.41) uniformly over $\mathscr{B}$. For this it is sufficient (see Theorem 2(b)) that Z_1 satisfies Cramér's condition (1.16). Assumptions (A_3) and (A_4) now imply that this condition holds uniformly on compacts K in an appropriate sense (see the first observation in [2] following (1.50), page 11). □

There appears to have grown in recent times a considerable amount of applied work, especially in econometrics, on the formal Edgeworth expansion. See, for example, Chambers (1967), Phillips (1977), Sargan (1976), and references contained in these articles. It may be noted that the conditions imposed by Chambers (1967) (Section 2.2) on the characteristic function of the statistic are not sufficient to insure the existence of a valid asymptotic expansion. Besides, such conditions imposed directly on the statistic are extremely hard to verify, at least in the context of the present article.

Acknowledgment. The authors are indebted to Professor I. R. Savage and the referee for several helpful suggestions.

REFERENCES

[1] BAHR, B. VON (1967). On the central limit theorem in R_k. *Ark. Mat.* 7 61–69.

[2] BHATTACHARYA, R. N. (1977). Refinements of the multidimensional central limit theorem and applications. *Ann. Probability* 5 1-27.

[3] BHATTACHARYA, R. N. and MAJUMDAR, M. (1973). Random exchange economies. *J. Econ. Theory* 6 37–67.

[4] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

[5] BICKEL, P. J. (1974). Edgeworth expansions in nonparametric statistics. *Ann. Statist.* 2 1-20.

[6] BIKJALIS, A. (1968). Asymptotic expansions of distribution functions and density functions of sums of independent and identically distributed random vectors. *Litovsk. Mat. Sb.* 6 405–422 (in Russian).

[7] CHAMBERS, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* 54 367-383.

[8] CHIBISHOV, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theor. Probability Appl.* 17 620–630.

[9] CHIBISHOV, D. M. (1973a). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theor. Probability Appl.* 18 295–303.

[10] CHIBISHOV, D. M. (1973b). An asymptotic expansion for the distribution of sums of a special form with an application to minimum contrast estimates. *Theor. Probability Appl.* 18 649–661.

[11] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[12] DYNKIN, E. B. (1951). Necessary and sufficient statistics for a family of probability distributions. *Uspehi Mat. Nauk.* 6 No. 1 (41) 68–90.

[13] GHOSH, J. K., SINHA, B. K. and WIEAND, H. S. (1977). Second order efficiency of the mle with respect to any bounded bowl-shaped loss function (to appear).

[14] GHOSH, J. K. and SUBRAMANYAM, K. (1974). Second order efficiency of maximum likelihood estimators. *Sankhyā Ser. A* 36 325–358.

[15] GÖTZE, F. and HIPP, C. (1977). Asymptotic expansions in the central limit theorem under moment conditions (to appear).

[16] HSU, P. L. (1945). The approximate distributions of the mean and variance of a sample of independent variables. *Ann. Math. Statist.* 16 1–29.

[17] JAMES, G. S. (1955). Cumulants of a transformed variate. *Biometrika* 42 529–531.

[18] JAMES, G. S. (1958). On moments and cumulants of systems of statistics. *Sankhyā* 20 1-30.

[19] JAMES, G. S. and MAYNE, ALAN J. (1962). Cumulants of functions of random variables. *Sankhyā Ser. A* 24 47-54.

[20] MILNOR, J. W. (1965). *Topology from the Differentiable View Point.* Univ. Press of Virginia, Charlottesville.

[21] PFANZAGL, J. (1973a). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* 1 993–1026.

[22] PFANZAGL, J. (1973b). The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 25 171-198.

[23] PHILLIPS, P. C. B. (1977). A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. *Econometrica* 45 1517-1534.

[24] SARGAN, J. D. (1976). Econometric estimators and the Edgeworth expansion. *Econometrica* 44 421–448.

[25] SWEETING, T. J. (1977). Speeds of convergence for the multidimensional central limit theorem. *Ann. Probability* 5 28–41.

[26] WALLACE, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 635–654.

[27] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
BUILDING NO. 89
THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721

INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
CALCUTTA 7000 35 INDIA