Ib M. Skovgaard

TRANSFORMATION OF AN EDGEWORTH EXPANSION BY A SEQUENCE OF SMOOTH FUNCTIONS.

Preprint No. 2

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

February 1980

ABSTRACT. Based on the assumption that a sequence of distributions can be approximated to a certain order of accuracy by an Edgeworth series, it is proved that such an expansion may be transformed by a sequence of smooth functions of the corresponding random vectors to yield another Edgeworth expansion of the resulting sequence of distributions. This expansion may be calculated from the moments obtained by the delta method, i.e. the moments formally calculated from a Taylor series expansion omitting terms of low order. It is briefly sketched how this method may be used to obtain Edgeworth expansions of maximum likelihood estimators based on independent but not identically distributed random vectors. A simple example of this kind is given.

Key words: delta method, Edgeworth expansion, smooth transformation.

1. Introduction.

The aim of the present paper is to provide the tools for Edgeworth expansions of distributions of statistics in non-standard cases, in particular connected with independent, but not identically distributed observations. The main result, presented in Section 3, is a generalisation of a resent result by Bhattacharya & Ghosh (1978) concerning transformation of an Edgeworth expansion by a smooth function. The line of proof used to establish Theorem 3.2 is the one used in the above mentioned paper.

The usefulness of the theorem emerges from the following example. Let P_θ be a distribution with density $\exp\{<\theta,x> - \psi(\theta)\}$ with respect to some measure μ on $\mathbb{R}^k$, $x,\theta \in \mathbb{R}^k$, and let $X_1, X_2, \ldots$ be independent, X_i distributed as P_{θ_i}, $\theta_i = a_i(\beta)$, $\beta \in \mathbb{R}^m$ and $a_i : \mathbb{R}^m \to \mathbb{R}^k$ linear. The maximum likelihood estimate $\hat{\beta}_n$ of β based on $X_1,\ldots,X_n$ is given by the equation

$$\tau_n(\hat{\beta}_n) = \sum_{i=1}^{n} a_i^*(x_i)$$

where a_i^* is the adjoint of a_i, and $\tau_n(\beta) = \sum_{i=1}^{n} a_i^* E_\beta\{x_i\}$. As a sum of independent random variables the distribution of $\sum_{i=1}^{n} a_i^*(X_i)$ can often be expanded in an Edgeworth series. Thus, if we are able to transform such an expansion (by τ_n^{-1}), we can obtain an expansion of the distribution of $\hat{\beta}_n$. An example of this kind is given in Section 6. In Skovgaard (1980) the results are applied in greater generality to obtain an expansion of the distribution of the maximum likelihood estimator.

Section 2 provides the notation used in the paper. In Section 3 we present the main theorem, and a lemma, which, besides being used in the proof, may be used to verify the assumptions of the theorem. The content of the theorem, which generalizes Theorem 2 in Bhattacharya & Ghosh (1978), is that a valid Edgeworth expansion may be transformed by a sequence of sufficiently smooth functions to yield another valid Edgeworth expansion. This may be done by the so-called delta method, in which moments of a function of a random vector are formally calculated from a Taylor series expansion of the function. The conditions needed are essentially conditions on the derivatives of the functions. The proof and the calculation of the final expansion is much relieved by a theorem of Leonov & Shiryaev (1959) on the calculation of cumulants of polynomials of random vectors. This result is reviewed in the appendix.

Section 4 contains proofs of the results in Section 3 and some lemmas, which may be of independent interest. In Section 5 we reformulate the classical results on Edgeworth expansions of densities in terms of cumulants and characteristic functions of the sequence of distributions to be expanded. This is done merely by going through the classical proofs and picking out the necessary assumptions. These assumptions are at an "earlier stage" than usual, in the sense that the classical results may be proved by veryfying the assumptions in the cases usually considered. However, the derived theorem may be used in other cases as well, e.g. dependent observations, or independent observations not satisfying the usual moment conditions.

2. Notation.

To a large extent we shall follow the notation used in Bhattacharya & Rao (1976). The major exception is that we replace partial derivatives by differential forms. When $x = (x_1, \dots, x_k)$ and $y \in (y_1, \dots, y_k)$, $x, y \in \mathbb{R}^k$, we define

$$< x,y > = \sum_{i=1}^{k} x_i y_i \quad , \quad \| x \| = < x,x >^{\frac{1}{2}} \tag{2.1}$$

$$x'^m = (x, \dots, x) \in (\mathbb{R}^k)^m \quad , \quad m \in \mathbb{N} \tag{2.2}$$

$\mathcal{B}_k$ denotes the Borel system on $\mathbb{R}^k$.

Let f be a function of E into F, where E and F are Euclidean vector spaces. The class $C^s(E,F)$ is the class of s times continuously differentiable functions. For $f \in C^s(E,F)$ we define $D^p f(x) : E^p \to F$, the p'th differential of f taken at x, as the symmetric p-linear function satisfying

$$D^p f(x)(t'^p) = \frac{d^p}{dh^p} f(x + ht)\Big|_{h=0} \quad , \quad t \in E, \ h \in \mathbb{R} \tag{2.3}$$

$B_p(E,F)$ is the class of symmetric p-linear functions of E^p into F, and if $A \in B_p(E,F)$ we define

$$\| A \| = \sup\{A(x'^p) \mid \| x \| \leq 1\} \tag{2.4}$$

If $A \in B_2(E, \mathbb{R})$ we shall not distinguish between this and the function (also denoted by A) $A : E \to E$ defined by $< A(x), y > = A(x,y)$, $x,y \in E$; thus $A \in \text{Hom}(E,E)$, the class of linear functions of E into E. If $A \in \text{Hom}(E,F)$, A^* denotes its adjoint, i.e. $< A(x),y> = < x, A^*(y) >$. 1_E denotes the identity map on E.

Cumulants and moments are defined as multi-linear symmetric forms, e.g. μ_p, the p'th moment of the random vector $X \in \mathbb{R}^k$, is given by

$$\mu_p(t'^p) = E\{ \langle t,x \rangle^p\}, \quad t \in \mathbb{R}^k \tag{2.5}$$

where $E\{\ldots\}$ denotes expectation. By linearity we extend the definition of multivariate forms to allow for multiplication by complex numbers, e.g. if $A \in B_p(E,\mathbb{R})$ and $x \in E$, then $A((ix)'^p) = i^p A(x'^p)$.

Polynomials of E into F are mappings of the form

$$x \to A_0 + A_1(x) + \ldots + A_n(x'^n) \quad , \quad x \in E \tag{2.6}$$

where $A_0 \in F$, $A_i \in B_i(E,F)$, $i = 1,\ldots,n$. n is the degree of the polynomial and the A's are the coefficients.

The standard normal density on $\mathbb{R}^k$ is denoted ϕ and $\phi_{\mu,\Sigma}$ is the normal density with mean $\mu \in \mathbb{R}^k$, and variance $\Sigma \in B_2(\mathbb{R}^k, \mathbb{R})$.

The Cramér-Edgeworth polynomials $\tilde{P}_r$ are as usual defined by the formal identity

$$\sum_{r=0}^{\infty} u^r \tilde{P}_r(z : \{\chi_j\}) = \exp\{ \sum_{r=1}^{\infty} u^r \chi_{r+2}(z'^{r+2})/(r+2)!\} \tag{2.7}$$

where (χ_j), $j \in \mathbb{N}$ are the cumulants of a distribution. Also, if $\Sigma \in B_2(\mathbb{R}^k, \mathbb{R})$ is regular, $P_r(-\phi_{0,\Sigma} : \{\chi_j\})$ is the density of the finite signed measure with characteristic function $\tilde{P}_r(it : \{\chi_j\}) \cdot \exp\{-\frac{1}{2}\Sigma(t,t)\}$, obtained by formally substituting the differential operator for $(-it)$ in $\tilde{P}_r(it : \{\chi_j\})$, and using this on $\phi_{0,\Sigma}$. In particular $P_r(-\phi_{0,\Sigma} : \{\chi_j\})(x)$ is a polynomial in $x \in \mathbb{R}^k$ multiplied by $\phi_{0,\Sigma}(x)$.

The order symbols o and O are unless otherwise stated used as $n \to \infty$, i.e. if f and g are functions of $\mathbb{N}$ into normed spaces, then $f = O(g)$ if there exists a constant $C > 0$ such that $\| f(n) \| \leq C \| g(n) \|$ for all $n \in \mathbb{N}$, and $f = o(g)$ if to each $\varepsilon > 0$ an $n_0 \in \mathbb{N}$ exists, such that $\| f(n) \| \leq \varepsilon \| g(n) \|$ when $n \geq n_0$.

3. Main Results.

Suppose we have established an asymptotic expansion of the distributions of (U_n), $n \in \mathbb{N}$ of the form

$$P\{U_n \in B\} = \int_B \xi_n(u)\,du + o(\beta_{s,n}) \quad \text{as } n \to \infty$$

$$\text{uniformly in } B \in \mathcal{B}_k \tag{3.1}$$

where $s \geq 3$, $E\{\|U_n\|^s\} < +\infty$,

$$\xi_n(u) = \sum_{r=0}^{s-2} P_r(-\phi : \{\chi_{\nu,n}\})(u) \tag{3.2}$$

$$\beta_{s,n} = (\sup\ \{\|\chi_{\nu,n}\|^{1/(\nu-2)} \mid 3 \leq \nu \leq s\})^{s-2} = o(1) \tag{3.3}$$

and $(\chi_{\nu,n})$, $1 \leq \nu \leq s$ are the cumulants of U_n, satisfying $\chi_{1,n} = 0$, $\chi_{2,n} = 1_{\mathbb{R}^k}$. The reader may think of U_n as a normalized sum of independent identically distributed random vectors, in which case $\beta_{s,n}$ is of order $n^{-(s-2)/2}$, since $\chi_{\nu,n}$ is $O(n^{-(\nu-2)/2})$.

Consider a sequence (g_n), $n \in \mathbb{N}$ of functions of $\mathbb{R}^k$ into $\mathbb{R}^m$, $m \leq k$, p times differentiable at zero, $(p \geq 2)$, and satisfying

$$g_n(0) = 0 \quad \text{and} \quad Dg_n(0) \quad \text{is of rank } m \tag{3.4}$$

and suppose, that we want to approximate the distribution of $g_n(U_n)$ by some expansion as $n \to \infty$. Obviously, in this expansion we can only hope for an error term, which is at most as accurate as the error term in (3.1).

Define

$$f_n = B_n^{-1} g_n\ , \quad B_n^2 = (Dg_n(0))(Dg_n(0))^* \tag{3.5}$$

such that $f_n(U_n)$ has asymptotic variance $1_{\mathbb{R}^m}$. We shall prove, that under certain conditions on the derivatives of g_n, a valid expansion of the distribution of $f_n(U_n)$ may be constructed from the moments of $f_n(U_n)$ obtained formally by the delta method. Consider the Taylor series expansion of $f_n(U_n)$

$$Y_n = \sum_{j=1}^{p-1} \frac{1}{j!} (D^j f_n(0))(U_n'^j) \tag{3.6}$$

Since Y_n is a polynomial in U_n, it is in principle easy to compute its moments. An Edgeworth expansion constructed from these moments may in turn be used as an approximation to the distribution of $f_n(U_n)$. There are, however, some problems. First of all the computation may involve moments of U_n of higher order than s. This problem is resolved by defining the formal cumulants of U_n, such that the j'th formal cumulant of U_n is identical to the cumulant of U_n, when $j \leq s$, otherwise it is zero. Also we define the formal moments of U_n (and thereby Y_n) in terms of the formal cumulants by the well known formula connecting moments and cumulants.

Let $(\kappa_{\nu,n})$, $1 \leq \nu \leq q$, $n \in \mathbb{N}$ be the cumulants of Y_n computed from the first q formal moments of Y_n. Also let

$$\tilde{\kappa}_{\nu,n} = \kappa_{\nu,n} + o(\beta_{s,n}) \quad , \quad 1 \leq \nu \leq q \tag{3.7}$$

be some approximations to these cumulants. Thus if the cumulants are calculated as a sum of terms, e.g. by the method of Leonov & Shiryaev (1959) described in the appendix, terms of order $o(\beta_{s,n})$ may be omitted. In particular the cumulants $\tilde{\kappa}_{\nu,n}$ satisfying (1.10) and (1.11) in Bhattacharya & Ghosh (1978) satisfy (3.7).

Let η_n be the density of the finite signed measure with characteristic function

$$\hat{\eta}_n(t) = \exp\{i < t, \tilde{\kappa}_{1,n} > - \tfrac{1}{2}\tilde{\kappa}_{2,n}(t,t)\} \sum_{r=0}^{q-2} \tilde{P}_r(it : \{\tilde{\kappa}_{\nu,n}\}) \quad (3.8)$$

when $\tilde{\kappa}_{2,n}$ is regular; otherwise we may define η_n as identically zero.

The validity of the delta method and the problem, how to choose q is solved by the Theorem 3.2 below. Also, notice that (3.7) determines the accuracy to be used, when calculating moments of Y_n.

Define

$$\rho_n(\alpha) = ((2+\alpha) \log \beta_{s,n}^{-1})^{\frac{1}{2}} \quad , \quad \alpha > 0 \quad (3.9)$$

$$H_n(\alpha) = \{t \in \mathbb{R}^k \mid \|t\| \leq \rho_n(\alpha)\} \quad , \quad \alpha > 0 \quad (3.10)$$

such that, according to Lemma 4.1, integrals of polynomials multiplied by ξ_n outside $H_n(\alpha)$ are $o(\beta_{s,n})$.

<u>Assumptions 3.1</u>. <u>There exists an</u> $\alpha > 0$, <u>such that for all sufficiently large</u> n, <u>we have</u>

I. f_n <u>is</u> p <u>times continuously differentiable on</u> $H_n(\alpha)$, <u>and</u>

$$\sup\{\|D^p f_n(t)\| \mid t \in H_n(\alpha)\} = o(\beta_{s,n}) \quad (3.11)$$

II. <u>Define</u> $\lambda_n = \sup\{(\|D^j f_n(0)\|/j!)^{1/(j-1)} \mid 2 \leq j \leq p-1\}$. <u>Then</u>

$$\lambda_n^{p-1} = o(\beta_{s,n}) \quad (3.12)$$

Theorem 3.2. Suppose, that Assumptions 3.1 are fulfilled. Then, if $\lambda_n^{q-1} = o(\beta_{s,n})$ and $q \geq s$

$$P\{f_n(U_n) \in B\} = \int_B \eta_n(u)\,du + o(\beta_{s,n})$$

$$\text{uniformly in } B \in \mathcal{B}_m \qquad (3.13)$$

In particular, (3.13) holds if $q = \max\{p,s\}$.

Remark 3.3. In Bhattacharya & Ghosh (1978), U_n equals $\sqrt{n} \cdot \overline{X}_n$, where $\overline{X}_n$ is an average of n i.i.d. random vectors, and $f_n(U_n) = \sqrt{n}\, f(\overline{X}_n)$, such that $\|D^j f_n(t)\|$ is of order $\sqrt{n}^{-(j-1)}$, where defined. If the distribution of $\overline{X}_n$ can be expanded in an Edgeworth series of the form (3.1) with $\beta_{s,n}$ of order $n^{-(s-2)/2}$, then (3.13) holds with $q = p = s$. Also, in general, when p is chosen as the smallest integer, not less than s, satisfying (3.12), then we obtain the remarkably simple result, that one more cumulant is to be calculated, than the number of terms in the Taylor series expansion of $f_n(U_n)$.

Remark 3.4. From the proof of the theorem it is easily seen, that if (3.1) only holds uniformly over any class $\mathcal{C}$ of Borel sets satisfying

$$\sup_{B \in \mathcal{C}} \int_{(\partial B)^\varepsilon} \phi(u)\,du = O(\varepsilon) \quad \text{as } \varepsilon \to 0$$

where $(\partial B)^\varepsilon = \{t \in \mathbb{R}^k \mid \exists\, u \in B : \|t-u\| \leq \varepsilon\}$, then (3.13) also holds uniformly over any such class.

Lemma 3.5. Let Assumptions 3.1 be fulfilled and $k = m$, then for n large enough, f_n restricted to $H_n(\alpha)$ is one-to-one on its image, which contains $H_n(\alpha_1)$ for some $\alpha_1 > 0$. The inverse function also satisfies Assumptions 3.1. Conversely, if (h_n), $n \in \mathbb{N}$ is a sequence

<u>of functions, locally (around zero) inverse to (f_n), and satisfying the Assumptions 3.1, then also (f_n) satisfies the assumptions.</u>

<u>Remark 3.6</u>. The lemma may be helpful in proving (3.11) and (3.12) in applications, where the transformations are given in terms of (f_n^{-1}), e.g. the example in the introduction. If the functions (f_n) are analytic on $(H_n(\alpha))$, and if $\tilde{\lambda}_n = \sup\{(\| D^j f_n(0) \| / j!)^{1/(j-1)} \mid 2 \leq j\}$ satisfies $\tilde{\lambda}_n^{p-1} = o(\beta_{s,n})$, then (3.11) and (3.12) are fulfilled. (3.11) is verified by noting that $\rho_n(\alpha)$ is less than the radius of convergence of the Taylor series expansion of f_n, when n is sufficiently large; thus $D^p f_n(t)$ can be written as a convergent series. Evaluating this, the result follows. The condition $\tilde{\lambda}_n^{p-1} = o(\beta_{s,n})$ may be verified by verifying the same condition for the functions (f_n^{-1}). This is seen from Lemma 4.3.

<u>Remark 3.7</u>. It is seen from the proof, that the normalization (3.5) of $g_n(U_n)$ may be replaced by any other normalization giving asymptotic variance one. Also the index $n \in \mathbb{N}$ appearing throughout may be replaced by any set directed to the right, e.g. a projective system.

<u>Remark 3.8</u>. In applications the functions (f_n) may be stochastic. In this case the conditions (3.11) and (3.12) need only be fulfilled with probability $1 - o(\beta_{s,n})$.

4. Lemmas and proofs.

First, we prove some auxiliary lemmas, and then we return to the proofs of Lemma 3.5 and Theorem 3.2.

Lemma 4.1. *Let* $A \in B_p(\mathbb{R}^k, \mathbb{R})$, $p \in \mathbb{N}$. *Then*

$$\int_{\|u\| \geq c} |A(u'^p)|\, \phi(u)\, du$$

$$\leq \|A\| \exp\{-\tfrac{1}{2}c^2\}(2^{p/2} c^{k+p-2} + 2^{(k+2p-2)/2}\, \Gamma(\tfrac{k+p}{2}))/\Gamma(\tfrac{k}{2}) \tag{4.1}$$

where Γ *is the gamma function.*

Proof. Use the inequality $|A(u'^p)| \leq \|A\| \, \|u\|^p$ and transform to a one-dimensional integral in $\|u\|$. Using the inequality $(c + \|u\| - c)^\alpha \leq (2c)^\alpha + (2(\|u\| - c))^\alpha$, $\alpha > 0$, the result follows by integration. □

Lemma 4.2. *The finite signed measures with densities* (ξ_n) *given by* (3.2) *have moments* $(\alpha_{j,n})$, $j,n \in \mathbb{N}$ *of all orders, satisfying*

$$\|\alpha_{j,n} - \mu_{j,n}\| = \begin{cases} 0 & \text{if } 1 \leq j \leq s \\ O(\beta_{s,n}^{(s-1)/(s-2)}) & \text{otherwise} \end{cases} \tag{4.2}$$

where $(\mu_{j,n})$, $j \in \mathbb{N}$ *are the formal moments of* U_n.

Proof. By differentiating $\exp\{-\tfrac{1}{2}\|z\|^2\} \sum_{r=0}^{s-2} \tilde{P}_r(z:\{\chi_{\nu,n}\})$, j times, it is seen, that the value at zero coincides with the general formula for moments in terms of cumulants, except for terms of order $O(\beta_{s,n}^{(s-1)/(s-2)})$, when the cumulants are the formal cumulants of U_n.

Lemma 4.3. *Let* $h_n : \mathbb{R}^k \to \mathbb{R}^k$ *be* s *times differentiable at* $t \in \mathbb{R}^k$, $s \geq 2$, *and suppose that* $Dh_n(t)$ *is regular. Then* h_n *is one-to-one in a* neighbourhood of t, and if h_n^{-1} is an inverse of h_n restricted to this set, then h_n^{-1} is s times differentiable at $u = h_n(t)$, and for $2 \leq j \leq s$

$$D^j h_n^{-1}(u) =$$

$$\sum_{\nu \in T(j-1)} [(j-1+\Sigma\nu_i)!/\prod_{i=1}^{j-1} \nu_i!(i+1)!^{\nu_i}] Dh_n(t)^{-1} \circ S[D_2^{,\nu_1},\ldots,D_j^{,\nu_{j-1}}] \tag{4.3}$$

where $T(j-1) = \{(\nu_1,\ldots,\nu_{j-1}) \mid \sum_{i=1}^{j-1} i\nu_i = j-1\}$,
$D_i = -D^i h_n(t) \circ (Dh_n(t)^{-1})^{,i} \in B_i(\mathbb{R}^k, \mathbb{R}^k)$ and $S[D_2^{,\nu_1},\ldots,D_j^{,\nu_{j-1}}]$ is some convex combination of compositions of the $\Sigma\nu_i$ arguments If $k = 1$ any such composition is just the product, such that S itself reduces to the product.

Comment. An example is $S[D_2^{,3}](x^{,4})$
$= \frac{1}{5} D_2(D_2(x,x)^{,2}) + \frac{4}{5} D_2(x,D_2(x,D_2(x,x)))$. A general formula is not known, but for our application the important thing is that
$\| S[D_2^{,\nu_1},\ldots,D_j^{,\nu_{j-1}}] \| \leq \| D_2 \|^{\nu_1} \ldots \| D_j \|^{\nu_{j-1}}$.

Proof. The theorem is well known except for formula (4.3). We shall only give the proof in the case $k = 1$, since the general case is proved in the same way. The only difference arises because of the non-commutativity of the D_j's, but since we have not stated the form of S, this causes no trouble. The theorem is proved by induction. The formula is easily verified for $j = 1,2$. Let $j > 2$ and suppose the formula holds for $D^{j-1}h_n^{-1}(u)$. Replacing t by $h_n^{-1}(u)$ we may differentiate both sides with respect to u to obtain $D^j h_n^{-1}(u)$. Obviously, each term in the formula for $D^j h_n^{-1}(u)$ will be of the form

$$c(\nu)(Dh_n(t)^{-1})^{j+\Sigma\nu_i}(-D^2h_n(t))^{\nu_1}\dots(-D^jh_n(t))^{\nu_{j-1}}$$

with $\nu \in T(j-1)$. Such a term occurs only as derivative of a corresponding term with ν replaced by $\mu \in T(j-2)$ satisfying either $\mu_2 = \nu_2 - 1$, $\mu_i = \nu_i$ for $i \geq 3$ or $\mu_i = \nu_{i+1}$, $\mu_{i+1} = \nu_{i+1} - 1$ for some $i \geq 2$ and $\mu_k = \nu_k$ for $k \neq i, i+1$. Using the formula (4.3) for $c(\mu)$ and adding the contributions from different μ's proves the formula. □

Lemma 4.4. Suppose Assumptions 3.1 are fulfilled, and that $Dg_n(0) = 1_{\mathbb{R}^k}$ $(k = m)$. Then with α as in these

(a) $$f_n(t) = t + \sum_{j=2}^{p-1} \frac{1}{j!}(D^j f_n(0))(t'^j) + \|t\|^p o(\beta_{s,n}) = t + o(1)$$

uniformly in $t \in H_n(\alpha)$ (4.4)

(b) $$Df_n(t) = 1_{\mathbb{R}^k} + \sum_{j=2}^{p-1} \frac{1}{(j-1)!}(D^j f_n(0))(t'^{j-1}) + \|t\|^{p-1}\, o(\beta_{s,n})$$

$$= 1_{\mathbb{R}^k} + o(1)$$ uniformly in $t \in H_n(\alpha)$ (4.5)

Proof. The first equation in (4.4) and (4.5) is merely a Taylor-series expansion, estimating the error term by (3.11). The second equation follows by noting that $\rho_n(\alpha) \to \infty$ slower than any power of $\beta_{s,n}$, and that $\lambda_n = o(\beta_{s,n}^{1/(p-1)})$.

□

Proof of Lemma 3.5. Suppose (f_n) satisfies the Assumptions 3.1. By (4.5) an $n_0 \in \mathbb{N}$ exists, such that for $n \geq n_0$ $\sup\{\|Df_n(t) - 1_{\mathbb{R}^k}\| \mid t \in H_n(\alpha)\} \leq \frac{1}{2}$. Hence, if $t_1, t_2 \in H_n(\alpha)$

$$\|f_n(t_1) - f_n(t_2)\| = \|t_1 - t_2\| + \tfrac{1}{2}\theta\|t_1 - t_2\| \geq \tfrac{1}{2}\|t_1 - t_2\|,$$

where $\theta \in [-1,1]$, proving that f_n is one-to-one. Next, if $\alpha_1 < \alpha$ and H^0 denotes the interior of H, (4.4) shows that for n large $H_n(\alpha_1) \cap f_n(H_n^0(\alpha)) = H_n(\alpha_1) \cap f_n(H_n(\alpha))$, which is closed relative to $H_n(\alpha_1)$. Also, this set is open relative to $H_n(\alpha_1)$ when $n \geq n_0$, because $Df_n(t)$ is regular on $H_n^0(\alpha)$. Since $H_n(\alpha_1)$ is connected, we conclude that $H_n(\alpha_1) \subseteq f_n(H_n(\alpha))$. Thus an inverse f_n^{-1} of f_n's restriction to $H_n(\alpha)$ exists, and is defined on $H_n(\alpha_1)$. Also, by (4.3) and (3.12) $\| D^j f_n^{-1}(0) \| = O(\lambda_n^{j-1})$, if $2 \leq j \leq p-1$. It is seen, e.g. from (4.4), that $\| D^{i+1} f_n(t) \| = O(\lambda_n^{i-1})$ uniformly in $t \in H_n(\alpha)$, $1 \leq i \leq p-2$, such that Lemma 4.3 together with (3.11) and (3.12) implies

$$\| D^p f_n^{-1}(u) \| = O(\lambda_n^{p-1}) + o(\beta_{s,n}) = o(\beta_{s,n}) \quad , \quad u \in f_n(H_n(\alpha)) \ .$$

The second half of the lemma follows from the first. □

Lemma 4.5. *Suppose $P_{1,n}$ and $P_{2,n}$ are sequences of signed measures with densities on $\mathbb{R}^k$*

$$\left(\sum_{j=0}^{M} A_{j,n}^{(i)}(x'^j) \right) \phi(x) \quad , \quad i = 1,2 \ , \ n \in \mathbb{N} \ , \ A_{j,n}^{(i)} \in B_j(\mathbb{R}^k, \mathbb{R}) \tag{4.6}$$

with respect to the Lebesque measure. Assume, that all integrals of the form

$$\int_{\mathbb{R}^k} A(x'^p) \ d(P_{1,n} - P_{2,n})(x) \quad , \ p \in \mathbb{N} \ , \ A \in B_p(\mathbb{R}^k, \mathbb{R})$$

are $O(\varepsilon_n)$ *as* $n \to \infty$, *where* (ε_n) *is a sequence of non-negative numbers. Then*

$$|P_{1,n}(B) - P_{2,n}(B)| = O(\varepsilon_n) \quad \text{uniformly in } B \in \mathcal{B}_k \tag{4.7}$$

Proof. Writing $\sum_{j=0}^{M} (A_{j,n}^{(1)} - A_{j,n}^{(2)})(x'^j)$ in terms of multivariate Hermite-polynomials, the orthogonal property of these, combined with the assumption, ensures that $\|A_{j,n}^{(1)} - A_{j,n}^{(2)}\| = O(\varepsilon_n)$ as $n \to \infty$. Thus

$$|P_{1,n}(B) - P_{2,n}(B)| = |\int_B \sum_{j=0}^{M} (A_{j,n}^{(1)} - A_{j,n}^{(2)})(x'^j)\phi(x)dx|$$

$$\leq \sum_{j=0}^{M} \|A_{j,n}^{(1)} - A_{j,n}^{(2)}\| \int_{\mathbb{R}^k} \|x\|^j \phi(x)dx = O(\varepsilon_n)$$

uniformly in $B \in \mathcal{B}_k$. □

Sketch of the proof of Theorem 3.2. First assume $Dg_n(0) = 1_{\mathbb{R}^k}$, $(k = m)$. (a) By transforming the density ξ_n it is proved that an expansion, $(\tilde{\eta}_n)$ say, of the distribution of $f_n(U_n)$ exists, $(\tilde{\eta}_n)$ of the form (4.6) with an error term of order $o(\beta_{s,n})$. (b) Moments computed in $(\tilde{\eta}_n)$ are equivalent (i.e. identical except for an error term of order $o(\beta_{s,n})$) to the (ξ_n) moments of the Taylor series approximation Y_n of $f_n(U_n)$. (c) By Lemma 4.2 the (ξ_n) moments of Y_n are equivalent to the formal moments of Y_n. (d) The formal moments are equivalent to the (η_n) moments, where (η_n) is given by (3.8). This result is obtained by use of a result by Leonov & Shiryaev (1959) described in the appendix, on cumulants of polynomials of random vectors in terms of the original cumulants. (e) By Lemma 4.5 the theorem follows in the case $Dg_n(0) = 1_{\mathbb{R}^k}$ and the general case is proved by supplementing the map f_n by a linear mapping p_n, such that (f_n, p_n) is locally one-to-one. Using this result the general theorem follows by a linear transformation of the expansion.

Lemma 4.6. Let the Assumptions 3.1 be fulfilled and suppose $Dg_n(0) = 1_{\mathbb{R}^k}$. Then

$$P\{f_n(U_n) \in B\} = \int_B (1 + Q_{1,n}(z))\phi(z)dz + o(\beta_{s,n})$$

$$\text{uniformly in } B \in \mathcal{B}_k \tag{4.8}$$

where $(Q_{1,n})$, $n \in \mathbb{N}$ are polynomials of bounded degree, no constant term and coefficients, which are $o(1)$ as $n \to \infty$. Also, if $A \in B_q(\mathbb{R}^k, \mathbb{R})$, $q \in \mathbb{N}$ we have

$$\int_{\mathbb{R}^k} A(u'^q)\xi_n(u)du = \int_{\mathbb{R}^k} A(\psi_n(z)'^q)(1 + Q_{1,n}(z))\phi(z)dz + o(\beta_{s,n}) \tag{4.9}$$

where (ψ_n) is given by

$$\psi_n(z) = z + \sum_{j=2}^{p-1} \frac{1}{j!}(D^j f_n^{-1}(0))(z'^j) \tag{4.10}$$

and (f_n^{-1}) are inverse functions of (f_n) restricted to $H_n(\alpha)$.

Proof. First notice, that by Lemma 3.5, f_n's restriction to $H_n(\alpha)$, $\tilde{f}_n$ say, has an inverse f_n^{-1} satisfying Assumptions 3.1 for some $\alpha_1 > 0$ (for n large enough). Thus, using (3.1) and Lemma 4.1

$$P\{f_n(U_n) \in B\} = \int_{f_n^{-1}(B)} \xi_n(u)du + o(\beta_{s,n})$$

$$= \int_{B_n} \xi_n(f_n^{-1}(z))|\det Df_n^{-1}(z)|dz + o(\beta_{s,n}) \tag{4.11}$$

where $B_n = B \cap H_n(\alpha_1)$, and det A means the determinant of A.

In the sequel let $(Q_{i,n})$ denote polynomials of bounded degree, without constant terms, and with bounded coefficients as $n \to \infty$, i fixed. By expanding $f_n^{-1}(z) = \psi_n(z) + o(\beta_{s,n})$, and expanding $Df_n^{-1}(z)$ as in (4.5) the last integral in (4.11) may be written

$$\int_{B_n} \xi_n(z+Q_{2,n}(z)+o(\beta_{s,n})\|z\|^p)\,|\det(1_{\mathbb{R}^k}+Q_{3,n}(z)+o(\beta_{s,n})\|z\|^{p-1})|\,dz$$

$$= \int_{B_n} \xi_n(z+Q_{2,n}(z)+o(\beta_{s,n})\|z\|^p)(1+Q_{4,n}(z)+o(\beta_{s,n})Q_{5,n}(z))\,dz \tag{4.12}$$

where the coefficients of $(Q_{2,n})$, $(Q_{3,n})$ are $O(\lambda_n)$ and those of $(Q_{4,n})$ are $o(1)$, $(Q_{3,n})$ taking values in $\mathrm{Hom}(\mathbb{R}^k, \mathbb{R}^k)$. Notice, that when n is large (4.5) assures that the determinant is positive. By expanding the exponent in $\xi_n(f_n^{-1}(z))$ we obtain

$$\xi_n(z+Q_{2,n}(z)+o(\beta_{s,n})\|z\|^p)$$
$$= (2\pi)^{-k/2}\exp\{-\tfrac{1}{2}\|z\|^2\}(1+Q_{6,n}(z)+o(\beta_{s,n})Q_{7,n}(z)) \tag{4.13}$$

where the coefficients of $(Q_{6,n})$ are $o(1)$, and $(Q_{6,n})$ is of bounded degree, because $\lambda_n^{p-1} = o(\beta_{s,n})$, such that only finitely many of the terms of the expansion of the exponent enters $(Q_{6,n})$. Combining (4.12) and (4.13) with (4.11) and noting that

$$\int_{\mathbb{R}^k \setminus H_n(\alpha_1)} (1+Q_{1,n}(z))\phi(z)\,dz = o(\beta_{s,n})$$

we obtain (4.8).

(4.9) follows in the same way by expanding

$$A(u'^q) = A(\psi_n(z)'^q) + Q_{8,n}(z)o(\beta_{s,n}) \,. \qquad \square$$

<u>Lemma 4.7</u>. <u>With notation and assumptions as in Lemma 4.6, we have</u>

$$\int_{\mathbb{R}^k} A(\tau_n(u)'^q)\xi_n(u)du = \int_{\mathbb{R}^k} A(z'^q)(1+Q_{1,n}(z))\phi(z)dz + o(\beta_{s,n})$$

(4.14)

<u>where</u> $A \in B_q(\mathbb{R}^k, \mathbb{R})$ <u>and</u>

$$\tau_n(u) = u + \sum_{j=2}^{p-1} \frac{1}{j!}(D^j f_n(0))(u'^j) \tag{4.15}$$

<u>Proof</u>. Since $A(\tau_n(u)'^q)$ is a polynomial in u, (4.9) shows that

$$\int_{\mathbb{R}^k} A(\tau_n(u)'^q)\xi_n(u)$$

$$= \int_{\mathbb{R}^k} A((\tau_n o\psi_n)(z)'^q)(1+Q_{1,n}(z))\phi(z)dz + o(\beta_{s,n})$$

Now, $D(\tau_n o\psi_n)(0) = 1_{\mathbb{R}^k}$ and $D^j(\tau_n o\psi_n)(0) = 0$ if $2 \le j \le p-1$. Also, by the assumptions, Lemma 3.5 and the formula for differentials for composite functions (see Federer (1969), 3.1.11)

$$\| D^\alpha(\tau_n 0\psi_n)(0)\| \le \alpha! \sum_{\nu\in T(\alpha)} \|D^{\Sigma\nu_i}\tau_n(0)\| \prod_{i=1}^{\alpha} \frac{1}{\nu_i!}(\| D^i\psi_n(0)\|/i!)^{\nu_i}$$

$$\le \alpha! \sum_{\nu\in T(\alpha)} (\Sigma\nu_i)!\lambda_n^{\Sigma\nu_i-1} \prod_{i=1}^{\alpha}\frac{1}{\nu_i!}(c\lambda_n^{i-1})^{\nu_i} \le c(\alpha)\lambda_n^{\alpha-1} ,$$

where $\alpha \in \mathbb{N}$, c and $c(\alpha)$ are constants, and $T(\alpha)$ is given in Lemma 4.3.

Hence, all derivatives except the first are $o(\beta_{s,n})$, and since $\tau_n \circ \psi_n$ is a polynomial, we have

$$A((\tau_n \circ \psi_n)(y)'^q) = A(y'^q) + Q_{9,n}(y)o(\beta_{s,n})$$

from which the result follows. □

Proof of Theorem 3.2 in the case $Dg_n(0) = 1_{\mathbb{R}^k}$. As in Lemma 4.6 the exponent of (η_n) may be expanded, showing the existence of densities of the form (4.6) having all moments equivalent to those of (η_n). By Lemma 4.5, it remains only to prove, that the moments of (η_n) are equivalent to the formal moments of $(Y_n) = (\tau_n(U_n))$, and hence, by Lemma 4.7, to those of the expansion obtained in Lemma 4.6. In Leonov & Shiryaev (1959) it is described how to calculate cumulants of polynomials in terms of the original cumulants (see the appendix). By use of this method, it is not hard to prove, that the formal cumulants $(\kappa_{\nu,n})$ of (Y_n) satisfy

$$\kappa_{\nu,n} = O(\max\{\beta_{s,n}^{(\nu-2)/(s-2)} \; , \; \lambda_n^{\nu-2}\}) \; , \quad \nu \geq 2 \qquad (4.16)$$

Recall, that q is the number of cumulants entering the expansion (η_n), and note that if $3 \leq \nu \leq q$, $\kappa_{\nu,n}^{(q-1)/(\nu-2)} = o(\beta_{s,n})$ if $q \geq s$ and $\lambda_n^{q-1} = o(\beta_{s,n})$. Since (η_n) is constructed from the first q formal cumulants of Y_n, except for terms of order $o(\beta_{s,n})$ (see (3.7)), Lemma 4.2 shows, that the (η_n) moments deviate at most $O(\sup\{\kappa_{\nu,n}^{(q-1)/(\nu-2)} \mid 3 \leq \nu < q\}) = o(\beta_{s,n})$ from the moments of Y_n obtained from the formal cumulants $(\kappa_{\nu,n})$ changing $\kappa_{\nu,n}$ to zero if $\nu > q$. Since $\kappa_{\nu,n} = o(\beta_{s,n})$ if $\nu > q$, this proves the theorem in the case $Dg_n(0) = 1_{\mathbb{R}^k}$. □

Proof of Theorem 3.2 in the general case. If $m < k$, define for each $n \in \mathbb{N}$ a linear mapping $p_n : \mathbb{R}^k \to \mathbb{R}^{k-m}$ as an orthogonal projection on a subspace of dimension $k - m$ composed with a linear isometry of the subspace into $\mathbb{R}^{k-m}$, such that

$$p_n^{-1}\{0\} \cap (Dg_n(0))^{-1}\{0\} = \{0\} \qquad (4.17)$$

Notice, that $(Dg_n, p_n)(0) = (Dg_n(0), p_n)$ is regular, and define

$$h_n = \begin{cases} (Dg_n(0), p_n)^{-1} (g_n, p_n) : \mathbb{R}^k \to \mathbb{R}^k & \text{if } m < k \\ Dg_n(0)^{-1} g_n : \mathbb{R}^k \to \mathbb{R}^k & \text{if } m = k \end{cases} \qquad (4.18)$$

such that $Z_n = h_n(U_n)$ has asymptotic variance $1_{\mathbb{R}^k}$. It is easily seen, that the Assumptions 3.1 are fulfilled for the sequence (h_n) if they are for (f_n). Thus the expansion corresponding to (3.13) is valid for the sequence (Z_n). $f_n(U_n)$ is a linear function, T_n say, of Z_n. If $(\sigma_{j,n})$ are the formal cumulants of the Taylor series approximation of Z_n, and $(\tilde{\sigma}_{j,n})$ the approximations corresponding to (3.7), we may transform the expansion (3.13) of (Z_n) by T_n, to yield a valid expansion of the distribution of $f_n(U_n)$ with characteristic function

$$\exp\{i < T_n^*(t), \tilde{\sigma}_{1,n} > -\tfrac{1}{2}\tilde{\sigma}_{2,n}(T_n^*(t)'^2)\} \sum_{r=0}^{q-2} \tilde{P}_r(iT_n^*(t):\{\tilde{\sigma}_{\nu,n}\})$$

$$= \exp\{i < t, \tilde{\kappa}_{1,n} > -\tfrac{1}{2}\tilde{\kappa}_{2,n}(t,t)\} \sum_{r=0}^{q-2} \tilde{P}_r(it:\{\tilde{\kappa}_{\nu,n}\})$$

where $\tilde{\kappa}_{\nu,n}(t'^{\nu}) = \tilde{\sigma}_{\nu,n}(T_n^*(t)'^{\nu})$. Since this coincides with (3.8), and $(\tilde{\kappa}_{\nu,n})$ is easily seen to satisfy (3.7), the theorem is proved. □

5. Edgeworth expansions.

In this section we modify some results in Bhattacharya & Rao (1976) concerning asymptotic expansions of characteristic functions and densities. The purpose is to obtain conditions for a sequence of distributions to possess an Edgeworth expansion in terms of cumulants and characteristic functions of the sequence, rather than assuming that it is a distribution of normalized sums of independent random variables satisfying certain moment conditions.

Consider a sequence (U_n), $n \in \mathbb{N}$, of statistics on $\mathbb{R}^k$ having finite s'th moments for some fixed $s \geq 3$, and cumulants $(\chi_{\nu,n})$, $1 \leq \nu \leq s$, $n \in \mathbb{N}$ satisfying

$$\chi_{1,n} = 0 \quad , \quad \chi_{2,n} = 1_{\mathbb{R}^k} \quad , \quad n \in \mathbb{N} \tag{5.1}$$

$$\beta_{s,n} = (\max\{\| \chi_{\nu,n}\|^{1/(\nu-2)} \quad , \quad 3 \leq \nu \leq s\})^{s-2} = o(1) \tag{5.2}$$

Let $\hat{G}_n : \mathbb{R}^k \to \mathbb{C}$ denote the characteristic function of U_n.

Theorem 5.1. Let (a_n), (ε_n), $n \in \mathbb{N}$ be sequences of positive numbers satisfying $\beta_{s,n} \leq \varepsilon_n \leq a_n^{-(s-2)} = o(1)$, and assume that for some $\delta_1 > 0$, $\log \hat{G}_n(t)$ is defined on the set $\{\|t\| < \delta_1 a_n\}$ for all $n \in \mathbb{N}$, and

$$\sup\{\| D^s \log \hat{G}_n(t) \| \,\big|\, \|t\| < \delta_1 a_n\} = O(\varepsilon_n) \tag{5.3}$$

Then, if $0 \leq \alpha \leq s$, we have for some $\delta > 0$

$$\| D^{\alpha}[\hat{G}_n(t) - \exp\{-\tfrac{1}{2}\|t\|^2\} \sum_{r=0}^{s-3} \tilde{P}_r(it:\{\chi_{\nu,n}\})]\|$$

$$\leq \varepsilon_n(\|t\|^{s-\alpha}+\|t\|^{3(s-2)+\alpha}) \exp\{-\tfrac{1}{4}\|t\|^2\} O(1) \qquad (5.4)$$

uniformly on the set $\|t\| < \delta\, a_n$.

Proof. Very much like Bhattacharya & Rao (1976), Theorem 9.9. The main differences are, that we use Euclidean norms of differentials and cumulants, and that we assume (5.3) instead of some other assumptions implying it. □

Remark 5.2. In the ordinary case $U_n = B_n^{-\frac{1}{2}}(X_1 + \ldots + X_n)$, where $X_1, X_2, \ldots$ is a sequence of independent random vectors, and $B_n = \sum_{i=1}^{n} V\{X_i\}$, then (5.3) can be shown to hold with (ε_n) of order $(n^{-(s-2)/2})$ and (a_n) of order $(n^{(s-2)/2s})$ if $\|B_n^{-1}\| = O(n^{-1})$ $\sum_{j=1}^{n} E\{\|X_j\|^s\} = O(n)$.

Theorem 5.3. *Let the assumptions of Theorem 5.1 be fulfilled, and suppose that*

I. $$a_n/((2+\eta)\log \varepsilon_n^{-1})^{\frac{1}{2}} \to \infty \quad \textit{as } n \to \infty \qquad (5.5)$$

for some $\eta > 0$.

II. *For any* $\delta > 0$

$$\int_{\|t\| \geq \delta a_n} |\hat{G}_n(t)|\, dt = O(\varepsilon_n) \qquad (5.6)$$

and

$$\int_{\|t\| \geq \delta a_n} \|D^s \hat{G}_n(t)\|\, dt = O(\varepsilon_n) \qquad (5.7)$$

Then for n *sufficiently large, the density* $f_n(x)$ *of the distribution of* U_n *exists, and*

$$(1+\|x\|^s)\,|f_n(x)-\sum_{r=0}^{s-3} P_r(-\phi:\{\chi_{\nu,n}\})(x)| = O(\varepsilon_n) \quad (5.8)$$

uniformly in $x \in \mathbb{R}^k$.

Proof. (5.6) implies that the integral of $|\hat{G}_n(t)|$ is finite for n large, and hence that the density f_n exists. Let $\|v\|=1$, $v \in \mathbb{R}^k$ and define for $\alpha = 0,s$

$$h_n^{(\alpha)}(x,v) = \langle x,v\rangle^{\alpha}\,(f_n(x) - \sum_{r=0}^{s-3} P_r(-\phi:\{\chi_{\nu,n}\})(x))$$

$$\hat{h}_n^{(\alpha)}(t) = (-i)^{\alpha} D^{\alpha}[\hat{G}_n(t) - \sum_{r=0}^{s-3} \tilde{P}_r(it:\{\chi_{\nu,n}\})\exp\{-\tfrac{1}{2}\|t\|^2\}]$$

such that

$$h_n^{(\alpha)}(x,v) = (2\pi)^{-k} \int_{\mathbb{R}^k} \exp\{-i\langle x,t\rangle\}\hat{h}_n^{(\alpha)}(t)(v'^{\alpha})\,dt$$

Thus (5.8) is bounded by

$$(2\pi)^{-k} \int (|\hat{h}_n^{(0)}(t)| + \|\hat{h}_n^{(s)}(t)\|)\,dt .$$

Since the integral of (5.4) is $O(\varepsilon_n)$ it remains only to estimate the integral outside the set $\{\|t\| < \delta a_n\}$. But, since (5.5) and Lemma 4.1 assures that

$$\int_{\|t\| \geq \delta a_n} \|D^{\alpha} \sum_{r=0}^{s-3} \tilde{P}_r(it:\{\chi_{\nu,n}\})\exp\{-\tfrac{1}{2}\|t\|^2\}\|\,dt = O(\varepsilon_n) ,$$

the theorem follows from (5.6) and (5.7) . □

Remark 5.4. If the assumptions of Theorem 5.3 hold for some $s \geq k+1$, then a uniform expansion of the distribution of U_n over all Borel sets follows, because the integral of $(1+\|x\|^{k+1})^{-1}$ is finite. In the case $s \leq k$ truncation techniques must be used.

The conditions (5.6) and (5.7) are usually the critical ones. To facilitate applications we shall give a lemma which provides sufficient conditions for (5.6) and (5.7) to hold in the case where U_n is a normalized sum of independent variables.

Let $X_1, X_2, \ldots$ be mutually independent random vectors in $\mathbb{R}^k$ with mean zero, characteristic functions $g_j : \mathbb{R}^k \to \mathbb{C}$, $j \in \mathbb{N}$ and cumulants $(\kappa_{\nu,j})$, $\nu = 1,\ldots,s$; $j \in \mathbb{N}$. Define

$$U_n = \Sigma_n^{-\frac{1}{2}} (X_1 + \ldots + X_n) \tag{5.9}$$

where $\Sigma_n = \sum_{j=1}^{n} \kappa_{2,j}$. The characteristic function of U_n is given by

$$\hat{G}_n(t) = \prod_{j=1}^{n} g_j(\Sigma_n^{-\frac{1}{2}}(t)) \tag{5.10}$$

and its cumulants are

$$\chi_{\nu,n}(t'^{\nu}) = \sum_{j=1}^{n} \kappa_{\nu,j}(\Sigma_n^{-\frac{1}{2}}(t)'^{\nu}) \quad , \quad t \in \mathbb{R}^k \; , \; \nu = 1,\ldots,s$$

<u>Lemma 5.5</u>. <u>Let</u> (U_n) <u>be given by</u> (5.9), <u>and assume that</u>

I. <u>An integer</u> p <u>and a finite set</u> $M \subset \mathbb{N}$ <u>with at least</u> $p+s$ <u>elements exists</u>, <u>such that for each set</u> $M_1 \subseteq M$ <u>with</u> p <u>elements</u>

$$\int \prod_{j \in M_1} |g_j(t)| \, dt < \infty \tag{5.11}$$

II. <u>A constant</u> $K > 0$ <u>exists</u>, <u>such that</u>

$$\gamma_n = \inf\{ \sum_{j=1}^{n} (1 - |g_j(t)|^2) \mid \|t\| \geq K\}$$

<u>satisfies</u>

$$\gamma_n / (1 + \|\Sigma_n^{-1}\|^{-1} a_n^{-2}) \log(\sqrt{\det\Sigma_n}\, n^s/\varepsilon_n) \to \infty \tag{5.12}$$

III. $\max\{\|\kappa_{\nu,j}\| \; \|\Sigma_n^{-1}\|^{\nu/2} \mid 1 \le j \le n,\ 1 \le \nu \le s_0\} = O(1)$ (5.13)

<u>where</u> $s_0 \ge s$, s_0 <u>even</u>.

<u>Then (5.6) and (5.7) hold</u>.

<u>Proof</u>. Let $u = \Sigma_n^{-\frac{1}{2}}(t)$, $d_n = \|\Sigma_n^{-\frac{1}{2}}\| a_n \delta$ and note, that $\|u\| \ge d_n$ implies $\|t\| \ge a_n \delta$. We then have

$$\int_{\|t\| \ge a_n \delta} |\hat{G}_n(t)| dt \le \int_{\|u\| \ge d_n} (\det \Sigma_n)^{\frac{1}{2}} \prod_{j=1}^{n} |g_j(u)| \, du$$

$$\le (\det \Sigma_n)^{\frac{1}{2}} (\sup\{\prod_{j \notin M} |g_j(u)| \mid \|u\| \ge d_n\}) \int_{\|u\| \ge d_n} \prod_{j \in M} |g_j(u)| \, du \tag{5.14}$$

where M is the set in the Assumption I of the lemma. The last integral is bounded by assumption, and

$$\prod_{j \notin M} |g_j(u)| \le \exp\{-\tfrac{1}{2} \sum_{j \notin M} (1 - |g_j(u)|^2)\} .$$

Using the inequality $1 - |g_j(2t)|^2 \le 4(1 - |g_j(t)|^2)$, which holds in general for characteristic functions, we conclude, that if $0 < d_n \le K$, $\|u\| \ge d_n$ and m is suitably chosen, then

$$\sum_{j=1}^{n} (1 - |g_j(u)|^2) \ge 4^{-m} \sum_{j=1}^{n} (1 - |g_j(2^m u)|^2)$$

$$\ge (d_n^2/4K^2) \inf\{\sum_{j=1}^{n} (1 - |g_j(t)|^2) \mid \|t\| \ge K\}$$

$$= \gamma_n d_n^2/4K^2 . \tag{5.15}$$

If $d_n > K$, the left hand side of (5.15) is bounded below by γ_n, hence

$$\prod_{j\notin M} |g_j(u)| \le \exp\{-\tfrac{1}{2}\sum_{j=1}^{n}(1-|g_j(u)|^2)\}\ \exp\{\tfrac{1}{2}|M|\}$$

$$\le \exp\{-\tfrac{1}{2}\gamma_n \min(1, d_n^2/4K^2)\}\ \exp\{\tfrac{1}{2}|M|\}$$

where $|M|$ is the cardinality of M. (5.6) now follows from (5.14) using (5.12).

Now, $(D^s\hat{G}_n(t))(v'^s)$ is a sum of n^s terms of the form

$$(\prod_{j\notin D} g_j(\Sigma_n^{-\frac{1}{2}}(t))) \prod_{j\in D} (D^{k_j} g_j(\Sigma_n^{-\frac{1}{2}}(t)))(\Sigma_n^{-\frac{1}{2}}(v)'^{k_j}) \qquad (5.16)$$

where $D \subseteq \mathbb{N}$ contains at most s elements and $\sum_D k_j = s$. By Assumption I

$$n^s \prod_{j\notin D} |g_j(\Sigma_n^{-\frac{1}{2}}(t))| = n^s (\prod_{j\in M\setminus D} |g_j(\Sigma_n^{-\frac{1}{2}}(t))|) \prod_{j\notin M\cup D} |g_j(\Sigma_n^{-\frac{1}{2}}(t))|$$

$$(5.17)$$

where the first product is integrable uniformly in D. By arguments similar to those above, we see that the second product is $O(\varepsilon_n)n^{-s}(\det \Sigma_n)^{-\frac{1}{2}}$. Since the integral of the first product is $O(\sqrt{\det \Sigma_n})$, (5.17) is $O(\varepsilon_n)$. Consider

$$\prod_{j\in D} \| D^{k_j} g_j(\Sigma_n^{-\frac{1}{2}}(t))\| \le \prod_{j\in D} \rho_{k_j,j} \le \max_{1\le j\le n} (\rho_{s_0,j})^{s/s_0}$$

$$(5.18)$$

where $\rho_{\nu,j} = \sup\{E(|\langle X_j, t\rangle|^{\nu}) \mid \|t\| \le 1\}$, and $s_0 \ge s$. When s_0 is even

$$E(|\langle X_j,t\rangle|^{s_0}) = E(\langle X_j,t\rangle^{s_0})$$

is a linear combination of terms of the form

$$(\kappa_{1,j}(t))^{p_1} \cdots (\kappa_{s_0,j}(t'^{s_0}))^{p_{s_0}}, \quad \sum_{i=1}^{s_0} i p_i = s_0$$

where the product is some (symmetric) product of the multilinear forms. By Assumption III it now follows, that (5.16) is $O(\varepsilon_n)$, when integrated over the set $\|t\| \geq \delta a_n$, $\delta > 0$. The lemma is proved. □

Remark 5.6. In the usual situation, where all the eigenvalues of Σ_n are of the same order, and $\|\Sigma_n\|$ is of order n, ε_n of order $n^{-(s-2)/2}$ and (det Σ_n) of order n^k, the supremum in Assumption II should increase faster than log n, so that roughly the n'th term should be $o(n^{-1})$. E.g., the assumptions are easily verified if $X_1, X_2, \ldots$ are identically distributed with a distribution satisfying Cramer's condition.

6. An example. Let $X_1, X_2, \ldots$ be mutually independent real random variables, $(\alpha + \beta t_i)X_i$ being distributed as gamma with shape parameter $\omega_i > 0$, where $\alpha > 0$ and $\beta > 0$ are the unknown parameters, and $t_1, t_2, \ldots$ and $\omega_1, \omega_2, \ldots$ are sequences of known positive real numbers. Our purpose is under simple conditions to derive an expansion of the distribution of the maximum likelihood estimator (MLE) $(\hat{\alpha}, \hat{\beta})$, based on $X_1, \ldots, X_n$. Most of the quantities defined in the sequel will depend on n, but we shall usually omit n as index or argument.

$(T_1, T_2) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} t_i X_i)$ is a sufficient statistic with cumulants given by

$$K_\nu((u,v)'^\nu) = (\nu - 1)! \sum_{i=1}^{n} \omega_i \left(\frac{u}{\alpha}\rho_i + \frac{v}{\beta}(1 - \rho_i)\right)^\nu ,$$

$$(u,v) \in \mathbb{R}^2, \quad 1 \leq \nu < \infty \qquad (6.1)$$

where $\rho_i = \alpha/(\alpha + \beta t_i) \in]0,1[$. Notice, that the cumulants depend on α and β, in particular we shall consider the function

$$\tau_n : \mathbb{R}_+^2 \to \mathbb{R}, \quad \tau_n(\alpha,\beta) = K_1 .$$

Also, notice that K_2 is the Fisher-information of (α,β).

Let m_n and M_n be the smallest and largest eigenvalue of K_2, respectively. We shall assume, that the sequence $\omega_1, \omega_2, \ldots$ is bounded, and that $(\log n)/m_n \to 0$ as $n \to \infty$. The first assumption causes no loss of generality, since an X_i corresponding to a large ω_i might be considered as a sum of sufficiently many independent X_{ij}'s with the same scale parameter. The second assumption avoids a detailed investigation of the behaviour of the t_i's, but it may be proved, that the condition is satisfied either for all pairs (α,β) or for none.

Our first goal is to derive an Edgeworth expansion of the distribution of (T_1,T_2) and next to transform it to an expansion of $(\hat{\alpha},\hat{\beta})$ on noting, that

$$(\hat{\alpha},\hat{\beta}) = \tau_n^{-1}(T_1,T_2) \tag{6.2}$$

with probability one, if K_2 is regular, i.e. $m_n > 0$. In the sequel let n be such, that this is so. Define $U_n = K_2^{-\frac{1}{2}}((T_1,T_2) - K_1)$ as some normalization of (T_1,T_2). Its cumulants (χ_ν) then satisfy

$$\begin{aligned} |\chi_\nu((u,v)'^\nu)| &= |K_\nu(K_2^{-\frac{1}{2}}(u,v))'^\nu| \\ &\leq (\nu-1)! \; \|(u,v)\|^2 \left[\left(\frac{1}{\alpha}+\frac{1}{\beta}\right) \|K_2^{-\frac{1}{2}}((u,v))\|\right]^{\nu-2} \\ &\leq (\nu-1)! \; \|(u,v)\|^\nu \left[\left(\frac{1}{\alpha}+\frac{1}{\beta}\right) m_n^{-\frac{1}{2}}\right]^{\nu-2} \end{aligned} \tag{6.3}$$

showing, that the cumulant generating function, being analytic in a neighbourhood of the origin, has radius of convergence greater than $\delta m_n^{\frac{1}{2}}$ for some $\delta > 0$ independent of n. An immediate consequence is, that if $a_n = m_n^{\frac{1}{2}}$, $\varepsilon_n = a_n^{-(s-2)}$ then the assumptions of Theorem 5.1 and (5.5) are fulfilled for all $s \geq 3$. We shall use Lemma 5.5 to prove (5.6) and (5.7). Since (ω_i) is bounded the cumulants $(\kappa_{\nu,i})$ of (X_i,t_iX_i) are also bounded and consequently satisfy (5.13). Also (5.11) is easy; just take M as the first $p+s$ integers, where p is chosen such that the sum of the ω's over any subset with p elements is at least one. Such a p exists, otherwise M_n would be bounded. To prove (5.12), notice that

$$\sum_{i=1}^{n} (1 - |g_j(u,v)|^2) = \sum_{i=1}^{n} \left(1 - \left(1 + \left(\frac{u}{\alpha}\rho_i + \frac{v}{\beta}(1-\rho_i)\right)^2\right)^{-\omega_i/2}\right)$$

is decreasing in any direction of (u,v) away from the origin, so that we may bound it on any set $\|(u,v)\| \geq C$ by its values on the set $\|(u,v)\| = C$. For a suitably small $\varepsilon > 0$ take C_ε such that $(\frac{u}{\alpha}\rho_i + \frac{v}{\beta}(1-\rho_i))^2 \leq \varepsilon$ if $\|(u,v)\| = C_\varepsilon$. Then with $\|(u,v)\| = C_\varepsilon$ we have

$$\sum_{i=1}^{n} (1 - |g_j(u,v)|^2) \geq \sum_{i=1}^{n} \tfrac{1}{2}\omega_i (\frac{u}{\alpha}\rho_i + \frac{v}{\beta}(1-\rho_i))^2 (1 - c\varepsilon)$$

$$= \tfrac{1}{2} K_2((u,v)'^2)(1-c\varepsilon) \quad \text{for some } c > o,\ c\,\varepsilon < 1 .$$

But since $K_2((u,v)'^2)$ tends to infinity at least at the rate of n, and $\log n/m_n \to 0$ as $n \to \infty$, it is now easy to see, that (5.12) is satisfied, because $M_n = O(n)$. Thus (U_n) satisfies (3.1) - (3.3) for any $s \geq 3$ with $\beta_{s,n} = O(\lambda_n^{-(s-2)/2})$.

It now turns out to be a trivial matter to verify Assumptions 3.1 with $p = s$ and λ_n proportional to m_n. The reason is, that the j'th derivative of τ_n is exactly the $(j+1)$'th cumulant K_{j+1} of (T_1, T_2). This is a general phenomena in exponential families. The normalized function f_n given in (3.5) may be defined by

$$f_n(U_n) = K_2^{-\frac{1}{2}}((\hat{\alpha},\hat{\beta}) - (\alpha,\beta)) = K_2^{\frac{1}{2}}(\tau_n^{-1}(K_1 + K_2^{\frac{1}{2}}(U_n)) - (\alpha,\beta))$$

such that if $x \in \mathbb{R}^2$

$$f_n^{-1}(x) = K_2^{-\frac{1}{2}}(\tau_n((\alpha,\beta) + K_2^{-\frac{1}{2}}(x)) - K_1)$$

$$D^j f_n^{-1}(o)((u,v)'^j) = K_2^{-\frac{1}{2}} K_{j+1}((K_2^{-\frac{1}{2}}(u,v))'^j)$$

implying that $\|D^j f_n^{-1}(o)\| = \|\chi_{j+1}\| = O(\lambda_n^{-(j-1)/2})$, $j \geq 1$. Since f_n^{-1} equals τ_n except for a linear transformation (3.11) follows from (5.3) proving that Assumptions 3.1 hold (f_n^{-1}). Thus, by Lemma 3.5, Assumptions 3.1 hold for (f_n) too, and Theorem 3.2 may be applied to provide an Edgeworth expansion of $f_n(U_n)$, or equivalently of $(\hat{\alpha},\hat{\beta})$. We shall not go through the trivial calculation of the expansion, but refer the reader to Skovgaard (1980), where this expansion is obtained for a much larger class of cases.

Remark 6.1. It may be proved, that $m_n \to \infty$ if and only if $\Sigma\ \omega_i(t_i - \bar{t})^2 \to \infty$ and $\Sigma\ \omega_i/t_i^2 \to \infty$, where $\bar{t} = \Sigma\ \omega_i t_i/\Sigma\ \omega_i$. Along the lines of this paper it follows, that this condition implies the consistency and asymptotic normality of $K_2^{\frac{1}{2}}(\hat{\alpha},\hat{\beta})$. Also, the condition is obviously necessary, since otherwise the information is bounded.

7. Appendix.

In Leonov & Shiryaev (1959) a method of calculating cumulants of polynomials in terms of the original cumulants was derived. We shall shortly review the method. Consider the table

$$\begin{matrix} (1,1),\dots,(1,n_1) \\ \\ (k,1),\dots,(k,n_k) \end{matrix} \tag{7.1}$$

and let $P_1 \cup \dots \cup P_m$ be a partition of its entries. Two sets, P_1 and P_2 say, of the partition is said to hook, if there exists an $i \in \{1,\dots,k\}$ and a pair $j_1, j_2 \in \{1,\dots,n_i\}$, such that $(i,j_1) \in P_1$ and $(i,j_2) \in P_2$. Two sets (P_1 and P_2) are said to communicate, if there exists a sequence $P_{m_1},\dots,P_{m_s}$; $P_1 = P_{m_1}$, $P_2 = P_{m_s}$, such that P_{m_i} and $P_{m_{i+1}}$ hook for $i = 1,\dots,s-1$. The partition is said to be indecomposable if any pair of the partition communicates.

Consider a polynomial

$$Y = A_1(X) + A_2(X'^2) + \dots + A_n(X'^n) \tag{7.2}$$

where $X \in \mathbb{R}^k$ is a random vector having finite $(n\alpha)$'th absolute moment, and $A_j \in B_j(\mathbb{R}^k, \mathbb{R}^m)$, $j = 1,\dots,n$.

We shall be concerned with the calculation of κ_α, the α'th cumulant of Y. Since κ_α is determined by the values $\kappa_\alpha(t'^\alpha)$, $t \in \mathbb{R}^m$, which is the α'th cumulant of $< t,Y >$, we may assume that Y is one-dimensional. Raise Y to the power α and consider one of its terms, say

$$Z = A_{n_1}(X'^{n_1}) \dots A_{n_\alpha}(X'^{n_\alpha}) \tag{7.3}$$

Arrange the X's in a table of the form

$$\begin{array}{c}(1,1),\ldots,(1,n_1)\\ \\ (\alpha,1),\ldots,(\alpha,n_\alpha)\end{array} \qquad (7.4)$$

in which each entry corresponds to one X. Let $\pi = P_1 \cup \ldots \cup P_m$ be a partition of the table (7.4), and define for each P_j, $\chi(P_j)$ as the $|P_j|$'th cumulant of X, where $|P_j|$ is the number of elements of P_j. Also let $\chi(\pi)$ be the multilinear form of $(\mathbb{R}^k)^{n_1 + \ldots + n_\alpha}$ into $\mathbb{R}$, given by

$$(\chi(\pi))(t_{11},\ldots,t_{1n_1},\ldots,t_{\alpha 1},\ldots,t_{\alpha n_\alpha}$$

$$= \prod_{j=1}^{m} (\chi(P_j))((t_{lm}),(lm) \in P_j)$$

where $(t_{11},\ldots,t_{1n},\ldots,t_{\alpha 1},\ldots,t_{\alpha n_\alpha}) \in (\mathbb{R}^k)^{n_1 + \ldots + n_\alpha}$.

The result of Leonov & Shiryaev (1959) is stated in the following lemma.

Lemma 7.1. *With notation as above*

$$\kappa_\alpha = \Sigma^* \sum_{\pi}^{**} < \chi(\pi), \ (A_{n_1},\ldots,A_{n_\alpha}) > \qquad (7.5)$$

where Σ^* *is the summation over all sequences* $(n_1,\ldots,n_\alpha) \in \{1,\ldots,n\}^\alpha$, Σ^{**} *is the summation over all indecomposable partitions of* (7.4), *and when* A,B *are* N-*linear mappings of* $(\mathbb{R}^k)^N \to \mathbb{R}$, *we define*

$$< A,B > = \Sigma \ A(e_1,\ldots,e_N) \ B(e_1,\ldots,e_N)$$

where the summation is over all ordered sequences of the canoni-

<u>cal base-vectors in</u> $\mathbb{R}^k$.

In Leonov & Shiryaev (1959) the result is formulated in terms of coordinates, which is convenient for computations, but the result above is convenient for theoretical considerations. In particular it is seen, that the terms of the sum in (7.5) are bounded by $\prod_{j=1}^{m} \|\chi(P_j)\| \prod_{i=1}^{\alpha} \|A_{n_i}\|$, a result which is used in the proof of Theorem 3.2, combined with the fact that only indecomposable partitions contribute to κ_α. For computations usually the first six cumulants will do. These may be found in James & Mayne (1962).

Acknowledgements.

This work was supported by the Danish Natural Science Research Council. Also, I would like to thank Steffen L. Lauritzen for fruitful discussions on the subject and for comments on an earlier version of the manuscript and a referee for pointing out several errors and ambiguities.

References.

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

Bhattacharaya, R.N. & Rao, R.R. (1976). Normal Approximations and Asymptotic Expansions. Wiley, New York.

Federer, H. (1969). Geometric Measure Theory. Springer, New York.

James, G.S. & Maine, A.J. (1962). Cumulants of functions of random variables. Sankhya Ser. A, 24, 47-54.

Leonov, V.P. & Shiryaev, A.N. (1959). On a method of calculation of semi-invariants. Theor. Probability Appl. 4, 319-329.

Skovgaard, I.M. (1980). Edgeworth expansions of the distributions of maximum likelihood estimators. Preprint, Inst. of Math. Stat., Univ. of Copenhagen.

PREPRINTS 1979

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Edwards, David: Large Sample Tests for Stationarity and Reversibility in Finite Markov Chains.

No. 2 Andersson, Steen A. : Distribution of Maximal Invariants Using Proper Action and Quotient Measures.

No. 3 Johansen, Søren: A Note on the Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression.

No. 4 Björnsson, Ottó: Four Simple Charcterizations of Standard Borel Spaces.

No. 5 Johansen, Søren: Some Comments on Robustness

No. 6 Hald, Anders: T.N. Thiele's Contributions to Statistics.

No. 7 Jacobsen, Martin: Markov Chains: Birth and Death Times with Conditional Independence.

PREPRINTS 1980

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Olkin, Ingram and Væth, Michael: Maximum Likelihood Estimation in a Two-way Analysis of Variance with Correlated Errors in One Classification.

No. 2 Skovgaard, Ib M.: Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions.