Ib M. Skovgaard

Edgeworth Expansions in Statistics

Preprint
October
1981
12

Institute of Mathematical Statistics
University of Copenhagen

Ib M. Skovgaard*

EDGEWORTH EXPANSIONS IN STATISTICS

Preprint No. 12

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

October 1981

CONTENTS

Section 1. Introduction.

This work is a thesis for the licentiat degree at the Institute of Mathematical Statistics, University of Copenhagen, and it consists of Section 1 - 3, Appendices A and B and three separately written papers. This unusual form of the thesis requires some explanation. The most important parts of this work are the three papers, which, although theoretically based on the general theory of Edgeworth expansions (Section 2 - 3), were written before these chapters. They are strongly connected in subject but written independently, all three depending heavily on the classical theory of Edgeworth expansions, of which an excellent and comprehensive account is given in Battacharya & Rao (1976). The reason, that I have chosen to include an account of this theory as Section 2 - 3 here is, of course, that some changes were desirable for our purpose. This slight change in scope will be explained further below; the most important things are the introduction of a multivariate notation, which makes calculations easier and makes a "directional" approach possible, and the separation of the theorems from their applications to sums of independent variables. Also distinction is being made between the construction of the Edgeworth approximations to a given measure as described in Section 2, and in Section 3 their properties as asymptotic expansions for certain sequences of measures.

In the three papers the notation has been chosen as a compromise between mathematical convenience and traditions of statistical literature, e.g. by avoiding the use of the tensor product. Here, I have made it an object to use the notation, that I find most convenient for the development and presentation of the theory, even though it may take a while for the reader to get used to it. The Appendices A and B should be sufficient to explain the notation, but further knowledge of multivariate algebra is helpful when doing the calculations.

I have not rewritten the three papers in this notation, because that would still not have finished the work. There are many subjects left out, e.g. expansions of the likelihood ratio test statistics and confidence regions, and at least the last two papers may be seen merely as examples of applications of the Edgeworth expansions. Also, based on the version of the classical expansions given in section 2 and 3, there should be possibilities of improving the theory outlined in the papers. Thus, an extension of the work would be more fruitful than a revision of the papers.

Section 2 contains the basic definitions of the multivariate Hermite- and Cramér-Edgeworth polynomials together with the construction of the Edgeworth approximations. The definitions agree with Chambers (1967) except for notational differences and their consequences on the various concepts. A theorem is given on the moments of the Edgeworth measures, but otherwise no attention is given to their behaviour.

In Section 3 the appearance of the Edgeworth approximations as expansions is investigated and conditions are given for the error of the density and of the distribution to tend to zero at a certain rate. Applications to sums of independent variables are considered in the form of theorems applying to these cases. Compared to the classical theory as outlined in Bhattacharya & Rao (1976) the technique of proof has been changed to allow for different rates of increase in the eigenvalues of the variance-covariance matrix, at the cost of the introduction of a logarithmic factor in the error bound of a power series expansion. This change does not alter the validity of the expansion, and if the expansion of any order is valid, it makes no difference at all, because the next term of the expansion then determines the limiting behaviour of the error.

Some notations, concepts and theorems used in Sections 2 and 3 are explained in Appendices A and B. Appendix A explains the use of the tensor product, which is used only to make the notation simpler; no deep algebraic results are needed for our purpose. Appendix B on differentiability contains a brief intoduction to differentiability and differentials and gives some theorems on how to calculate higher differentials of functions derived from other functions by composition, inversion, etc.

In the paper "Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions" it is shown, how an Edgeworth expansion of a sequence of distributions, obtained e.g. by the classical theory, may be transformed to an Edgeworth expansion (or rather a sequence of Edgeworth approximations with an error, which tends to zero at a certain rate) of the sequence of distributions obtained by non-linear transformations of the original sequence. The technique used is based on that in Bhattacharya & Ghosh (1978), where the result corresponding to transformation of an average of non-i.i.d. random variables is proved. Section 5 of the paper may be disregarded in this context, since an improved version is contained in Section 2 and 3 here. A natural extension would be also to consider asymptotically quadratic transformations, which would be relevant for expanding distributions of test statistic. In Chandra & Ghosh (1979) this is done for the cases corresponding to Bhattacharya & Ghosh (1978) and the extension to other cases is fairly obvious.

The paper "Edgeworth Expansions of the Distributions of Maximum Likelihood Estimators in the General (Non I.I.D.) Case" uses the results of the first paper to derive Edgeworth expansions for maximum likelihood estimators under conditions, that are not restricted to particular situations, such as replications of an experiment, and it is shown how the results may be applied to provide expansions, e.g. for the cases of normal regression models. Some explicit formulae are given to facilitate the computations. Similar results for the i.i.d. case have been proved in the one-dimensional case by Pfanzagl (1973), and in the multivariate case by Bhattacharya & Ghosh (1978). Ivanov (1976) has proved the validity of the expansion for the normal regression models, but not computed the expansion. Other papers on this subject are Chambers (1967), Chibisov (1972, 1973a, 1973b).

As a more special application the paper "A Second-order Investigation of Asymptotic Ancillarity" uses the Edgeworth expansions to investigate the conditional approach suggested in Efron & Hinkley (1978) in terms of asymptotic properties. The conditional distribution of the maximum likelihood estimator is approximated to second order by an Edgeworth approximation, which is explicitely calculated and some comparisons between observed and expected Fisher informations are made. The possibility of expanding conditional distributions in Edgeworth series has been pointed out by Michel (1979).

It is a pleasure to thank my supervisor Steffen L. Lauritzen for many discussions and helpful suggestions during the course of my work, and the Institute of Mathematical Statistics, University of Copenhagen, for their support. Also, I wish to thank the Department of Statistics at University of California, Berkeley, for their hospitality during the academic year 1980/81 and Ina Buhl for typing the manuscript.

Section 2. Edgeworth approximations.

In this section we shall construct the (multivariate) Edgeworth measures, used to approximate probability measures, in particular for sums of independent random variables. The Edgeworth measures are finite signed measures, absolutely continuous with respect to the Lebesgue measure, and their densities take the form of polynomials multiplied by a normal density.

Let E be a finite dimensional real vector space, $p = \dim E$, and Δ an inner product on E, i.e. a bilinear, positively definite, symmetric mapping of $E \times E$ into E. The canonical Lebesgue measure λ_Δ on the Euclidan space (E,Δ) assigns mass one to a unit cube.

The normal distribution on E with center $x_0 \in E$ and variance $\Delta^{-1} \in E \otimes E$ has the density

$$\varphi_{x_0,\,\Delta}(x) = (2\pi)^{-p/2} \exp\left\{-\tfrac{1}{2}\Delta(x-x_0,\ x-x_0)\right\} \tag{2.1}$$

with respect to λ_Δ. If $x_0 = 0$, we shall often write φ_Δ or just φ, if it is obvious, what Δ is.

Next, we define the Hermite polynomials. Recall, that in the one-dimensional case $(E=\mathbb{R})$, the k'th Hermite polynomial H_k is defined by

$$H_k(x) = (-1)^k \left[\frac{d^k}{dx^k}\varphi(x)\right] / \varphi(x) \tag{2.2}$$

In the general case we define the k'th Hermite polynomial $H_{k,\Delta}$ on (E,Δ) by

$$H_{k,\Delta} : E^{\otimes k} \to \mathrm{Pol}\,(E, \mathbb{R})$$

$$H_{k,\Delta}(A,x) = (-1)^k < D^k\varphi_\Delta(x)\ ,\ A > / \varphi_\Delta(x),\ x \in E\ ,\ A \in E^{\otimes k} \tag{2.3}$$

For fixed A this is a polynomial in x. If $E=\mathbb{R}$ the usual Hermite polynomials (2.2) are obtained by letting $A=1$. $H_{k,\Delta}$ is linear and symmetric in A, i.e. if $y_1,\dots, y_k \in E$ and σ is a permutation on $\{1,\dots,k\}$ then $H_{k,\Delta}(y_1\otimes\cdots\otimes y_k\ , x) = H_{k,\Delta}(y_{\sigma(1)}\otimes\cdots\otimes y_{\sigma(k)},x)$.

Thus, as a function of A , $H_{k,\Delta}$ is determined by its values on tensors of the form $A = y^{\otimes k}$, $y\in E$. These may be written

$$H_{k,\Delta}(y^{\otimes k}\ , x) = (-1)^k \left[\frac{d}{dh^k}\, \varphi_\Delta(x+hy)\Big|_{h=0} \right] / \varphi_\Delta(x) \tag{2.4}$$

which is determined in a simple way from the one-dimensional Hermite-polynomial.

The use of the Hermite polynomials is based on the following simple form of corresponding characteristic functions.

Lemma 2.1. The characteristic function of the measure with density $H_{k,\Delta}(A,x)\ \varphi_\Delta(x)$ with respect to λ_Δ is the function

$$t \to i^k < A\ , t^{\otimes k} > \exp\{-\tfrac{1}{2}\Delta^{-1}(t,t)\}\ , \ t\in E^* \tag{2.5}$$

Proof. It is sufficient to show the result with $A = y^{\otimes k}$. In that case, the result is an easy consequence of the well-known one-dimensional version, see e.g. Petrov (1975) Ch VI.1.

□

Another important property of the Hermite polynomials is their orthogonality property.

Theorem 2.2. Let A and B be symmetric tensors in E^k (i.e. of the form $\Sigma\, y_j^{\otimes k}$, and let $\Delta^{\otimes k} : E^{\otimes k} \to (E^*)^{\otimes k} \simeq (E^{\otimes k})^*$ be the k'th tensor product of $\Delta: E \to E^*$ with itself. Then

$$\int H_{k,\Delta}(A,x)\, H_{k,\Delta}(B,x)\, \varphi_\Delta(x)\, d\lambda_\Delta(x)$$

$$= \begin{cases} k!\ \langle \Delta^{\otimes k}(A)\ ,\ B \rangle & \text{if } k=m \\ 0 & \text{if } k\neq m \end{cases} \qquad (2.6)$$

Remark 2.3. Since $H_{k,\Delta}(A,x)$ is symmetric in $A\in E^{\otimes k}$, (2.6) also determines the values of integrals of the same form, but with arbitrary A and $B\in E^{\otimes k}$. One just has to replace A and B by their symmetric equivalents modulus the space generated by the permutations.

To give a coordinate analogue of (2.6), assume that Δ is the usual inner product on $E=\mathbb{R}^p$ (equivalently choose an orthonormal basis with respect to Δ), and consider

$$H_{k_1\ldots k_p}(x) = H_{k,\Delta}(e_1^{\otimes k_1}\otimes\ldots\otimes e_p^{\otimes k_p}\ ,\ x)$$

$$= (-1)^k \left[\left(\frac{d}{dx_1}\right)^{k_1} \ldots \left(\frac{d}{dx_p}\right)^{k_p} \varphi(x)\right] / \varphi(x)\ ,\ k=k_1+\ldots+k_p \qquad (2.7)$$

where $(e_1,\ldots,e_p)$ is the orthonormal basis (Δ). Then,

$$\int\ldots\int H_{k_1\ldots k_p}(x)\, H_{j_1\ldots j_p}(x)\varphi(x)dx_1\ldots dx_p$$

$$= \begin{cases} k!\ /\ (k_1!\ldots k_p!) & \text{if } (k_1,\ldots,k_p) = (j_1,\ldots,j_p) \\ 0 & \text{otherwise} \end{cases} \qquad (2.8)$$

The equivalence between (2.6) and (2.8) is easily established.

Proof. of Theorem 2.2. We shall prove the version (2.8). Suppose $k_i\neq j_i$ for some i, say $k_1>j_1$. Then, by integration by parts

$$\int\left(H_{k_1\ldots k_p}(x)\varphi(x)\right) H_{j_1\ldots j_p}(x)\, dx_1$$
$$= \left[-H_{(k_1-1)\ldots k_p}(x)\varphi(x)\, H_{j_1\ldots j_p}(x)\right]_{-\infty}^{\infty}$$
$$+ \int\left(H_{(k_1-1)\ldots k_p}(x)\varphi(x)\right) \frac{d}{dx_1} H_{j_1\ldots j_p}(x)\, dx_1$$

The square bracket term is zero, and if $k_1>1$, we may repeat the proces until we end up with

$$\int H_{(k_1-j_1-1)\ldots k_p}(x)\varphi(x) \left(\frac{d}{dx_1}\right)^{j_1+1} H_{j_1\ldots j_p}(x)dx_1 = 0 \ ,$$

since the highest degree of x_1 in $H_{j_1\ldots j_p}(x)$ is j_1. If $(k_1,\ldots k_p) = (j_1,\ldots,j_p)$, the same computation, after k_1 integrations by parts, yields

$$\int H_{0k_2\ldots k_p}(x)\varphi(x) \left(\frac{d}{dx_1}\right)^{k_1} H_{k_1\ldots k_p}(x)dx_1 \ .$$

In fact, since $H_{k_1\ldots k_p}(x) = H_{k_1}(x_1)\ldots H_{k_p}(x_p)$ (cf. (2.2)), and the coefficient of $x_1^{k_1}$ in $H_{k_1}(x_1)$ is one, this integral equals

$$k_1! \left(H_{k_2\ldots k_p}(x)\right)^2 \varphi(x)$$

Integration with respect to the other coordinates in the same way, yields the result.

□

When using Hermite polynomials, it is usually practical to work directly with their definition (2.2) - (2.4) or using some of their properties, e.g. (2.5) or (2.6). We shall, however, use the Hermite polynomials to construct certain approximations, and to do this one needs the polynomials themselves. These may be constructed recursively or by use of the following formula, which is easily obtained from Theorem B.5.

$$H_{k,\Delta}(y^{\otimes k}, x) = \begin{cases} \sum_{j=0}^{k/2} (-\tfrac{1}{2})^{k/2-j}\, k!/\left[(2j)!\,(k/2-j)!\right]\, \Delta(y,x)^{2j}\, \Delta(y,y)^{k/2-j} & \text{if } k \text{ is even} \\ \sum_{j=0}^{(k-1)/2} (-\tfrac{1}{2})^{(k-1)/2-j}\, k!\,/\left[(2j+1)!\,((k-1)/2-j)!\right]\, \Delta(y,x)^{2j+1}\, \Delta(y,y)^{(k-1)/2-j} & \text{if } k \text{ is odd} \end{cases} \quad (2.9)$$

Now, let P be a given probability measure on E, with characteristic function $\psi: E^* \to C$ and the first s cumulants $\chi_1,\ldots,\chi_s$ existing ; $\chi_k \in E^{\otimes k}$ being a symmetric tensor. We want to construct an approximation to P, depending only on these cumulants.

Assume, that $\chi_1=0$ (equivalently, choose χ_1 as origin in E), and that χ_2 is regular, and consider

$$\log \psi(t) \sim -\tfrac{1}{2}\chi_2(t^{\otimes 2}) + \ldots + \frac{i^s}{s!}\chi_s(t^{\otimes s}) \quad \text{as } t\to 0 \text{ , } t\in E^* \tag{2.10}$$

The approximation, we shall construct, is motivated as an approximation for sums of independent random variables. Therefore consider the function $t \to n\psi(t/\sqrt{n})$, which is the characteristic function of a normalized sum of n independent random variables with distribution P. Letting $\tau=1/\sqrt{n}$ we get

$$\log \tau^{-2}\psi(\tau t) \sim -\tfrac{1}{2}\chi_2(t^{\otimes 2}) + \ldots + \tau^{s-2}\frac{i^s}{s!}\chi_s(t^{\otimes s}) \quad \text{as } \tau\to 0 \text{ , } t \text{ fixed} \tag{2.11}$$

Taking the exponential of both sides and expanding around $\tau=0$ we obtain

$$\tau^{-2}\psi(\tau t) \sim \exp\{-\tfrac{1}{2}\chi_2(t^{\otimes 2})\}\left[1 + \sum_{k=1}^{s-2} \tau^k \tilde{P}_k(it:\{\chi_j\})\right] \tag{2.12}$$

where $\tilde{P}_k(it : \{\chi_j\})$ is the k'th <u>Cramér-Edgeworth polynomial</u> in (it) with coefficients depending on $\chi_1,\dots,\chi_{k+2}$. The formula for $\tilde{P}_k$ is

$$\tilde{P}_k(it : \{\chi_j\})$$

$$= k!\,\frac{d^k}{d\tau^k}\left[\exp\left\{\tau\,\frac{i^3}{6}\chi_3(t^{\otimes 3}) + \dots + \tau^{s-2}\,\frac{i^s}{s!}\chi_s(t^{\otimes s})\right\}\right]\Big|_{\tau=0}$$

$$= \sum_{\nu\in T(k)} \prod_{j=1}^{k} \frac{1}{\nu_j!}\left[\chi_{j+2}((it)^{\otimes j+2})/(j+2)!\right]^{\nu_j} \qquad (2.13)$$

obtained from Theorem B.5.

Here, we have extended the (χ_j's) by linearity to admit complex arguments.

Letting $\tau=1$ in (2.12) we obtain the approximation

$$\psi(t) \approx \psi_s(t:\{\chi_j\})$$

$$= \exp\{-\tfrac{1}{2}\chi_2(t^{\otimes 2})\}\left[1+\sum_{k=1}^{s-2}\tilde{P}_k(it:\{\chi_j\})\right] \qquad (2.14)$$

which has the advantage, compared to (2.10), of being easily invertible, as we shall see below.

With $\Delta = \chi_2^{-1}\epsilon(E^{\otimes 2})$, we see that $\exp\{-\tfrac{1}{2}\chi_2(t,t)\}$ is the characteristic function of Φ_Δ, the normal distribution with center 0 and variance χ_2. The inversion of (2.14) follows from Lemma 2.1 or directly by using the fact, that multiplying the characteristic function by (-it) corresponds to differentiation of the density in the direction t. Thus, the measure with characteristic function $\psi_s(t:\{\chi_j\})$ has density

$$f_s(x:\{\chi_j\}) = \left[1+\sum_{k=1}^{s-2} P_k(-D:\{\chi_j\})\right]\varphi_\Delta(x) \qquad (2.15)$$

with respect to λ_Δ , where $P_k(-D:\{\chi_j\})$ is the operator obtained from $\tilde{P}_k(it:\{\chi_j\})$ by substituting minus the differential operator D for (it) ; e.g. a term

$$\chi_j((it)^{\otimes j})\chi_k((it)^{\otimes k})\exp\{-\tfrac{1}{2}\chi_2(t^{\otimes 2})\}$$

$$= <\chi_j \otimes \chi_k \ , \ (it)^{\otimes j+k}> \exp\{-\tfrac{1}{2}\chi_2(t^{\otimes 2})\}$$

becomes

$$(-1)^{j+k} <D^{j+k}\varphi_\Delta(x) \ , \ \chi_j \otimes \chi_k>$$

$$= H_{j+k,\Delta} \ (\chi_j \otimes \chi_k \ , \ x)\varphi_\Delta(x)$$

The measure $Q_s(\cdot \ , \ \{\chi_j\})$ with density $f_s(x:\{\chi_j\})$ with respect to λ_Δ is called the <u>Edgeworth approximation</u> of order s-1. Often, we shall just write ψ_s , f_s and Q_s for the Edgeworth approximations to the characteristic function, the density and the measure. Notice, that the first order Edgeworth approximation is the usual normal approximation.

From (2.13) , (2.14) and (2.5) we deduce, that

$$f_s(x:\{\chi_j\})$$

$$= \sum_{m=3}^{3(s-2)} \sum_{\nu\in T(m)}^{*} H_{m,\Delta}(\chi_3^{\otimes \nu_3} \otimes \cdots \otimes \chi_s^{\otimes \nu_s} \ , \ x)\varphi_\Delta(x) \ / \ \prod_{j=3}^{s} \nu_j! j!^{\nu_j} \qquad (2.16)$$

where $\sum_{\nu\in T(m)}^{*}$ is the sum over all partitions $\nu = (\nu_1,\ldots,\nu_m)\in T(m) = \{\sum j\nu_j = m\}$ for which $\nu_1 = \nu_2 = 0$. This formula is a multivariate analogue of Petrov (1975) , VI . (1.9).

An important question is, of course, how well the Edgeworth approximations work as approximations. This will be discussed in the next section; here we shall only give a theorem on approximation of moments. We have not assumed moments of P of higher order than s , but we shall define the formal cumulants of P (of any order) by

$$\tilde{\chi}_\alpha = \begin{cases} \chi_\alpha & \text{if } 1\leq\alpha\leq s \\ 0 & \text{if } \alpha>s \end{cases} \tag{2.17}$$

and the formal moments of P by the well-known relations between moments and cumulants, i.e. by

$$\tilde{\mu}_\alpha(t^{\otimes\alpha}) = \sum_{\nu\in T(\alpha)} \alpha! \prod_{j=1}^{\alpha} \frac{1}{\nu_j!}\left[\tilde{\chi}_j(t^{\otimes j})/j!\right]^{\nu_j}, \quad t\in E^*, \quad \alpha\geq 1 \tag{2.18}$$

Also, we define for each "direction" $t\in E^*$,

$$\rho_s(t) = \sup\{(|\chi_j(t^{\otimes j})|/\chi_2(t^{\otimes 2})^{j/2})^{\frac{1}{j-2}} \mid 3\leq j\leq s\} \tag{2.19}$$

which is invariant under multiplication of t by a real (non zero) constant. Then, we have the following similarity between the formal moments and the moments of Q_s .

Theorem 2.3. With notation as above, the Edgeworth approximation Q_s has moments (m_α) of all orders, satisfying

$$m_\alpha = \tilde{\mu}_\alpha \qquad \text{if } 1\leq\alpha\leq s+2$$

$$|(m_\alpha-\tilde{\mu}_\alpha)(t^{\otimes\alpha})|\leq c_1(s,\alpha)\chi_2(t)^{\alpha/2}\rho_s(t)^{s-1}\max\{1,\rho_s(t)^{\alpha-s-3}\}$$

$$\alpha>s+2, \quad t\in E^* \tag{2.20}$$

where $c_1(s,\alpha)$ is a constant depending only on s and α (see (2.25)).

Proof. That P_s has moments of all orders, is obvious from the form of its density. In proving (2.20) we may restrict attention to one-dimensional ($E=\mathbb{R}$) distributions, since it is essentially an assertion on the moments of $\langle t,X\rangle$, X having distribution P respectively Q_s. With $E=\mathbb{R}$ moments and cumulants are real numbers. The moments of Q_s are obtained by differentiation of the characteristic function ψ_s. First, using (2.13), we get

$$\frac{d^\alpha}{dt^\alpha}\tilde{P}_k(it:\{\chi_j\})\Big|_{t=0} = \sum_{\nu\in T(k)}^{(\alpha)} i^\alpha \alpha! \prod_{j=1}^{k} (\chi_{j+2}/(j+2)!)^{\nu_j}/\nu_j! ,$$

$\sum^{(\alpha)}$ being the sum over all $\nu=(\nu_1,\dots,\nu_k)$ for which $\sum(j+2)\nu_j=\alpha$. Next, if $\alpha\geq 1$ we have

$$m_\alpha = i^{-\alpha}\frac{d^\alpha}{dt^\alpha}\Big[\exp\{-\tfrac{1}{2}\chi_2 t^2\}\ (1+\sum_{k=1}^{s-2}\tilde{P}_k(it:\{\chi_j\}))\Big]\Big|_{t=0}$$

$$= \sum_{\beta=0}^{\alpha}\binom{\alpha}{\beta} i^{-(\alpha-\beta)} H_{\alpha-\beta,\,\chi_2}(0) \sum_{k=0}^{s-2}\beta! \sum_{\nu\in T(k)}^{(\beta)} \prod_{j=1}^{k} (\chi_{j+2}/(j+2)!)^{\nu_j}/\nu_j! \qquad (2.21)$$

where the second factor $\sum_{k=0}^{s-2}(\dots)$ is taken to be one if $k=\beta=0$ and zero if $k=0$, $\beta>0$. Using (2.9), (2.21) equals

$$\sum_{\beta=0}^{\alpha}\frac{\alpha!\chi_2^{(\alpha-\beta)/2}}{((\alpha-\beta)/2)!2^{(\alpha-\beta)/2}}\sum_{\nu\in T(\beta)}^{*}\prod_{j=3}^{s}(\chi_j/j!)^{\nu_j}/\nu_j! \qquad (2.22)$$

where the first sum is restricted to $\alpha-\beta$ being even, and $\sum^*$ is restricted to the $\nu=(\nu_1,\dots,\nu_s)$ for which $\nu_1=\nu_2=0$ and $\sum(j-2)\nu_j\leq s-2$. Letting $(\alpha-\beta)/2$ play the role of ν_2, we see, that $2(\alpha-\beta)/2+\sum_3^s j\nu_j=\alpha$, such that (2.22) equals

$$\sum_{\nu\in T(\alpha)}^{**} \alpha! \prod_{j=2}^{\alpha} \frac{1}{\nu_j!} (\tilde{\chi}_j/j!)^{\nu_j} \tag{2.23}$$

where $\sum^{**}$ is restricted to the ν's satisfying $\sum(j-2)\nu_j \leq s-2$.

For $\tilde{\mu}_\alpha$ we have the formula (2.18), i.e.

$$\tilde{\mu}_\alpha = \sum_{\nu\in T(\alpha)} \alpha! \prod_{j=1}^{\alpha} (\tilde{\chi}_j/j!)^{\nu_j}/\nu_j! \tag{2.24}$$

which compared to (2.23), on noting that if $\nu\in T(\alpha)$ and $\nu_1=0$, we have $\sum(j-2)\nu_j = \alpha-2\sum\nu_j \leq \alpha-4$, shows that $\tilde{\mu}_\alpha = m_\alpha$ if $\alpha \leq s+2$, and otherwise

$$|\tilde{\mu}_\alpha - m_\alpha| \leq \sum_{\nu\in T(\alpha)}^{***} \alpha! \prod_{j=2}^{\alpha} (\rho_s^{\,j-2} \chi_2^{\,j/2}/j!)^{\nu_j}/\nu_j!$$

$$\leq c_1(s,\alpha)\, \chi_2^{\alpha/2} \rho_s^{\,s-1} \max\{1, \rho_s^{\,\alpha-s-3}\} ,$$

where $\sum^{***}$ is restricted to ν's for which $\nu_1=0$ and $\sum(j-2)\nu_j > s-2$, ρ_s is short for $\rho_s(1)$ and

$$c_1(s,\alpha) = \sum_{\nu\in T(\alpha)}^{***} \alpha! \Big/ \prod_{j=2}^{\alpha} j!^{\nu_j} \nu_j! \tag{2.25}$$

□

Section 3. Edgeworth expansions.

In this section we shall investigate how well an Edgeworth approximation behaves as an approximation of the distribution or density of a random variable X on a vectorspace E, $\dim E = p < \infty$. We shall first obtain some bounds for the approximation and later show how these behave for a sum of independent random variables as the number n of terms tends to infinity. It turns out, that the Edgeworth approximations appear as asymptotic expansions in this setting.

The scheme of proof is first to approximate the characteristic function as in (2.12) in a neighbourhood of the origin. The next step is to invert the approximation to yield an approximation to the density. To do this, one has to prove, that the contribution from the tail of the characteristic function is not of too great importance. The inversion takes a different form for absolutely continuous and for discrete distributions. In any case the local expansion (of the density) may be integrated, either with respect to the Lebesgue or the counting measure, to yield an approximation to the distribution (as a set function). In this respect the discrete (lattice) case is much less tractable.

The method follows closely that of Bhattacharya & Rao (1976), chapter 9. Besides the notational differences, there are some important changes. We don't expand the derivatives of the characteristic function, which in Bhattacharya & Rao are used to allow integration of the local expansion. Instead, we use a simple method, which either requires the existence of one more moment to obtain the same expansion or contributes a logarithmic factor to the last term of the expansion. The advantage of this approach, besides its simplicity, is, that it allows us to handle cases in which the eigenvalues of the variance grow at different rates in the sequence of distributions considered.

Recall, that

$$\rho_s(t) = \sup\{(|\chi_j(t^{\otimes j})|/\chi_2(t^{\otimes 2})^{j/2})^{1/(j-2)} | 3 \le j \le s\} \tag{3.1}$$

depends only on t via its "direction". The expansions of the characteristic function $\psi: E^* \to \mathbb{C}$ are also constructed considering each direction separately. Thus, the method is essentially the same as in the one-dimensional case.

The variance χ_2 defines an inner product on E^*, and we shall denote the norm of $t \in E^*$ by

$$||t|| = \chi_2(t^{\otimes 2})^{\frac{1}{2}} \tag{3.2}$$

Lemma 3.1. _If_ $t \in E^*$ _satisfies_ $||t||\rho_s(t) \le \delta$ _for some_ $\delta > 0$, _then_

$$|\exp\{\sum_{k=2}^{s} i^k \chi_k(t^{\otimes k})/k!\} - \exp\{-\tfrac{1}{2}||t||^2\} \sum_{k=0}^{s-2} \tilde{P}_k(it:\{\chi_j\})|$$

$$\le \rho_s(t)^{s-1} c_3(s,\delta) \max\{||t||^{s+1}, ||t||^{3(s-1)}\} \exp\{c_5(s,\delta)||t||^2\} \tag{3.3}$$

where the constants c_3 _and_ c_5 _depend on_ s _and_ δ _only, and_

$$c_5(s,\delta) = \sum_{k=1}^{s-2} \delta^k/(k+2)! - \tfrac{1}{2} \tag{3.4}$$

Proof. Consider the functions

$$g(u,t) = \sum_{k=3}^{s} i^k \chi_k(t^{\otimes k}) u^{k-2}/k! \quad , \quad u \in \mathbb{R}$$

$$f(u,t) = \exp\{g(u,t)\} \quad , \quad u \in \mathbb{R}$$

$$h(u,t) = f(u,t) - \sum_{k=0}^{s-2} \tilde{P}_k(it:\{\chi_j\}) \quad , \quad u\in\mathbb{R}$$

By the definition of the Cramér-Edgeworth polynomials $(\tilde{P}_k)$ we have

$$\frac{d^j}{du^j} h(u,t)\Big|_{u=0} = 0 \quad \text{if} \quad 0\leq j\leq s-2$$

$$\frac{d^{s-1}}{du^{s-1}} h(u,t)\Big|_{u=0} = \frac{d^{s-1}}{du^{s-1}} f(u,t) \ . \tag{3.5}$$

Also, if $0\leq u\leq 1$, $1\leq k\leq s-1$ and $||t||\rho_s(t)\leq\delta$, then

$$\left|\frac{d^k}{du^k} g(u,t)\right| \leq \sum_{j=k}^{s-2} |\chi_{j+2}(t^{\otimes j+2})|\, j!/\ [(j-k)!(j+2)!]$$

$$\leq \rho_s(t)^k c_2(s,k,\delta) \ ,$$

where

$$c_2(s,k,\delta) = \sum_{j=k}^{s-2} \delta^{j-k} j!/[(j-k)!(j+2)!] \tag{3.6}$$

Using this, we obtain

$$\left|\frac{d^{s-1}}{du^{s-1}} f(u,t)\right| \leq |f(u,t)| \sum_{\nu\in T(s-1)} (s-1)! \prod_{k=1}^{s-1} \left[\left|\frac{d^k}{du^k} g(u,t)\right|/k!\right]^{\nu_k}/\nu_k!$$

$$\leq |f(u,t)|\ \rho_s(t)^{s-1} \max\{||t||^{s+1}\ ,\ ||t||^{3(s-1)}\} c_3(s,\delta)(s-1)! \tag{3.7}$$

c_3 depending on s and δ only

Since also

$$|g(u,t)| \leq ||t||^2 c_4(s,\delta) \quad , \tag{3.8}$$

where

$$c_4(s,\delta) = \sum_{k=1}^{s-2} \delta^k/(k+2)! \tag{3.9}$$

it follows by (3.5) and Taylor's formula, that

$$|h(1,t)| \leq \exp\{c_4(s,\delta)||t||^2\} c_3(s,\delta) \; \rho_s(t)^{s-1} \max\{||t||^{s+1}, \; ||t||^{3(s-1)}\} \tag{3.10}$$

Since $|\exp\{-\frac{1}{2}||t||^2\} \; h(1,t)|$ is the quantity to be estimated, the lemma follows.

□

Theorem 3.2. *Let* t *be such, that* $|1-\psi(ut)| < 1$ *if* $0 \leq u \leq 1$, *and suppose that*

$$M(t) = \sup \{|(D^{s+1}\log \psi(ut)) \; (t^{\otimes s+1})|/||t||^{s+1} | 0<u<1\} \tag{3.11}$$

is finite. If also $||t||\rho_s(t) \leq \delta$, *then*

$$|\psi(t) - \psi_s(t)|$$

$$\leq \exp\{c_5(s,\delta)||t||^2 + M(t)||t||^{s+1}/(s+1)!\}$$

$$\times \; [M(t)/(s+1)! + \rho_s(t)^{s-1} c_3(s,\delta)] \max\{||t||^{s+1} \; , \; ||t||^{3(s-1)}\} \tag{3.12}$$

Proof. Define

$$h_1(t) = \sum_{k=2}^{s} i^k \chi_k(t^{\otimes k})/k!$$

$$h_2(t) = \log \psi(t) - h_1(t) .$$

Then

$$|\psi(t) - h_1(t)| = |(\exp\{h_2(t)\}-1)\exp\{h_1(t)\}| . \tag{3.13}$$

Since the first s derivatives of h_2 at zero vanish, the definition of $M(t)$ and Taylors formula gives

$$|h_2(t)| \leq M(t)||t||^{s+1}/(s+1)! \tag{3.14}$$

Also, by (3.8), $|h_1(t)| \leq c_5(s,\delta)||t||^2$, and by use of the inequality $|e^x-1| \leq |x|e^{|x|}$ for $x \in \mathbb{C}$, (3.13)- (3.14) implies

$$\begin{aligned} &|\psi(t) - \exp\{h_1(t)\}| \\ &\leq |h_2(t)|\exp\{|h_1(t)| + |h_2(t)|\} \\ &\leq (M(t)||t||^{s+1}/(s+1)!) \exp\{c_5(s,\delta)||t||^2 + M(t)||t||^{s+1}/(s+1)!\} \quad , \end{aligned}$$

which, combined with Lemma 3.1, gives the result.

□

Theorem 3.2 provides our final approximation to the characteristic function and corresponds to Bhattacharya & Rao (1976), Theorem 9.9. The use of this approximation relies on the fact, that if P is the distribution of n i.i.d. random variables, having $s+1$ finite moments, then $\rho_s(t)$ is of order

$1/\sqrt{n}$ and $M(t)$ of order $n^{-(s-1)/2}$ if $||t||\rho_s(t)\leq\delta$, such that the integral of the bound (3.12) over this set of t's is $O(n^{-(s-1)/2})$ as n tends to infinity.

More generally, consider a sequence (P_n) of distributions with characteristic functions ψ_n , cumulants $\chi_{j,n}$; $j=1,\dots,s$; $n = 1,2,\dots$; etc. We shall be concerned with expansions of (f_n) , the sequence of densities as $n\to\infty$; the density being either with respect to the Lebesgue measure or with respect to the counting measure on a lattice.

In <u>the continuous case</u>, the density f_n is obtained by inversion of the characteristic function,

$$f_n(x) = (2\pi)^{-p} \int_{E^*} \psi_n(t)e^{-i<t,x>}d\lambda_{\chi_{2,n}}(t) \tag{3.15}$$

is the density of P_n with respect to λ_{Δ_n} , if $|\psi_n(t)|$ is integrable. Using Theorem 3.2 and (3.15) we reach the following theorem. Note, that $||t||_n = \sqrt{\chi_{2,n}(t^{\otimes 2})}$.

<u>Theorem 3.3.</u> <u>Suppose that a sequence</u> (ε_n) <u>of positive numbers exists, and that for each</u> $t\in E^*$, <u>a sequence</u> $(a_n(t))$ <u>of positive numbers exists, such that</u> $a_n(t)$ <u>depends only on the direction</u> $t/||t||_n$ <u>of</u> t , <u>and such that for</u> $n\to\infty$ <u>we have</u>

$$\rho_{s,n}(t) \leq a_n(t)^{-1} \text{ for all } t\in E^* \tag{3.16}$$

$$a_n(t)^{-(s-1)} = O(\varepsilon_n) = o(1) \text{ uniformly in } t\in E^* \tag{3.17}$$

<u>and assume also, that for any</u> $\delta>0$ <u>we have</u>

$$\int_{||t||_n \geq \delta a_n(t)}^{\infty} |\psi_n(t)| d\lambda_{\chi_{2,n}}(t) = O(\varepsilon_n) \tag{3.18}$$

<u>and for some</u> $\delta > o$,

$$M(\delta a_n(t)t/||t||_n) = O(a_n(t)^{-(s-1)}) \text{ uniformly in } t \in E^* \tag{3.19}$$

<u>where</u> M <u>is given by</u> (3.11) . <u>Then for sufficiently large</u> n , <u>the density</u> f_n <u>of</u> P_n <u>with respect to</u> λ_{Δ_n} <u>exists, and</u>

$$|f_n(x) - f_{s,n}(x)| = O(\varepsilon_n) \text{ uniformly in } x \in E , \tag{3.20}$$

<u>where</u> $f_{s,n}$ <u>is the Edgeworth approximation of order</u> (s-1) <u>to</u> f_n ; <u>see (2.15)</u> .

<u>Proof</u>. By (3.15) we have

$$|f_n(x) - f_{s,n}(x)|$$

$$= (2\pi)^{-p} \left| \int (\psi_n(t) - \psi_{s,n}(t)) e^{-i<t,x>} d\lambda_{\chi_{2,n}}(t) \right|$$

$$\leq (2\pi)^{-p} \int |\psi_n(t) - \psi_{s,n}(t)| d\lambda_{\chi_{2,n}}(t) \tag{3.21}$$

Now, for any integrable function h on E^* , we may decompose the integral, writing $t = u(t/||t||_n)$, $0 \leq u < \infty$, to obtain

$$\int h(t) d\lambda(t) = \oint \int_0^\infty u^{p-1} h(ut) du \, d\tilde{\lambda}(t) , \tag{3.22}$$

where λ is the Lebesgue measure corresponding to the inner product on E^* , and $\oint \ldots d\tilde{\lambda}(t)$ is the integral over the unit sphere, $||t||=1$, with respect to the geometric surface measure $\tilde{\lambda}$ induced by λ . Thus, we only need to bound the integral over the sphere of

$$I(t) = \int_0^\infty u^{p-1}|\psi_n(ut)-\psi_{s,n}(ut)|du \ , \ t\in E^* \ , \ ||t||_n=1 \qquad (3.23)$$

by $O(\varepsilon_n)$. Given $\delta>0$ we estimate $I(t)$ by

$$I(t)\leq I_1(t)+I_2(t)+I_3(t)$$

where

$$I_1(t) = \int_0^{\delta a_n(t)} u^{p-1}|\psi_n(ut)-\psi_{s,n}(ut)|du$$

$$I_2(t) = \int_{\delta a_n(t)}^\infty u^{p-1}|\psi_n(ut)|du$$

$$I_3(t) = \int_{\delta a_n(t)}^\infty u^{p-1}|\psi_{s,n}(ut)|du$$

By (3.18), $\oint I_2(t)d\tilde{\lambda}(t)=O(\varepsilon_n)$, and by (3.16),(3.17) and Skovgaard (1981a), Lemma 4.1 it follows that $I_3(t)=O(\varepsilon_n)$, since $\exp\{-\frac{1}{2}a_n(t)^2\delta^2\}a_n(t)^k$ = $O(\varepsilon_n)$ for any $k\in N$. To estimate $I_1(t)$ we use Theorem 3.2 and (3.16),(3 17), (3.19) . First notice, that if $u\leq\delta a_n(t)$, then if δ is sufficiently small

$$M(ut)\leq M(\delta a_n(t)t) = O(a_n(t)^{-(s-1)}) = O(\varepsilon_n). \qquad (3.24)$$

Next, for the exponent of the bound (3.12) we have

$$c_5(s,\delta)u^2+M(ut)u^{s+1}/(s+1)!$$

$$\leq u^2\left[-\tfrac{1}{2}+\sum_{k=1}^{s-2}\delta^k/(k+2)!+O(a_n(t)^{-(s-1)})a_n(t)^{s-1}\delta^{s+1}\right]$$

$$\leq -u^2/4 \qquad (3.25)$$

if n is sufficiently large and δ sufficiently small.

That $I_1(t)$ and hence $I(t)$ is $O(\varepsilon_n)$ is now a trivial consequence of (3.24) and (3.25), and since all the bounds are uniform in t, the theorem is proved.

□

The formulation of the theorem, stating the conditions for each direction separately may seem to make these hard to verify. In fact, compared to the theorems of Bhattacharya & Rao (1976) or to Skovgaard (1981a), Th.5.3, the "directional" approach makes the theorem both easier to apply and more flexible in the sense, that cases, where the rate of convergence depends on the direction, are also covered. However, the resulting expansion of the distribution gives a slightly larger error rate compared to the other approaches, but this seems to be of minor importance for our purpose. We shall give an example to illustrate the use of Theorem 3.3 at the end of this section.

Before turning to the lattice case, we shall give a lemma showing how (3.18) may be verified in the case of sums of independent variables. Therefore, let $X_1, X_2, \ldots$ be mutually independent random variables on E ; X_j having mean zero, k'th cumulant $\kappa_{k,j}$ for $1 \leq k \leq s$ and characteristic function $g_j : E^* \to \mathbb{C}$. Let

$$U_n = X_1 + \ldots + X_n \tag{3.26}$$

have distribution P_n , cumulants $\chi_{k,n} = \kappa_{k,1} + \ldots + \kappa_{k,n}$, and characteristic function given by

$$\psi_n(t) = \prod_{j=1}^{n} g_j(t) .$$

Let E^* be equipped with a fixed Euclidean norm denoted $||\cdot||$, (which norm is chosen, doesn't matter), whereas the norm given by $\chi_{2,n}$ will be denoted by $||\cdot||_n$, i.e. $||t||_n = \chi_{2,n}(t^{\otimes 2})$.

The following lemma is a "directional" version of Skovgaard (1981a), Lemma 5.5.

Lemma 3.4. Let P_n be the distribution of U_n as above, and let the sequences $(a_n(t))$ and (ε_n) be given. Then under the following conditions, (3.18) holds:

I. A finite set $S \subset \mathbb{N}$ exists, such that

$$\int \prod_{j \in S} |g_j(t)| \, d\lambda(t) < \infty \tag{3.27}$$

where λ is the Lebesgue measure corresponding to the fixed inner product.

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

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

satisfies

$$\gamma_n(t)/[(1+||t||_n^2/a_n(t)^2) \log (||t||_n/\varepsilon_n)] \to \infty \tag{3.28}$$

uniformly in t in $\{||t|| = 1\}$.

Proof. Almost identical to the first part of the proof of Lemma 5.5 in Skovgaard (1981a). □

Notice, that there is no normalization in the definition (3.26) of U_n. This is taken care of by letting the variance $\chi_{2,n}$ define the metric and hence the canonical Lebesgue measure on E , with respect to which the density is taken. This is a more satisfactory method of normalization, since it avoids the arbitraryness in the choice of a square-root of the variance.

Let us now turn to the lattice case. If dim $E=p$, a lattice in E denotes a set of the form

$$L = \{z_1\xi_1+\dots+z_p\xi_p \; ; \; z_j \in \mathbb{Z} \; ; \; j=1,\dots,p\} \tag{3.29}$$

where $\xi_1,\dots,\xi_p \in E$ are linearly independent. A random variable $X \in E$ is called a lattice random variable (on L), if it takes values in x_o+L with probability one for some $x_o \in E$, and its variance χ_2 is regular. L is called the minimal lattice for X, if also there is no sublattice of L, which contains the support of $X-x_o$ for some $x_o \in E$. For a more thorough account on probability measures on lattices, see Bhattacharya & Rao (1976), Chapter 5, from where we have adopted the notation.

Next, let $\eta_1,\dots,\eta_p \in E^*$ be a dual base to $\xi_1,\dots,\xi_p \in E$, i.e.

$$\langle \xi_i, \eta_j \rangle = \delta_{ij} ,$$

where δ_{ij} is the Kronecker-delta. Define the set $F^* \leq E^*$ associated with L by

$$F^* = \{t_1\eta_1+\dots+t_p\eta_p \; ; \; |t_j|<\pi \; ; \; j=1,\dots,p\} \tag{3.30}$$

The <u>inversion formula for a lattice random variable</u> X takes the following form. Denote the characteristic function $\psi: E^* \to C$, and let λ be any Lebesgue measure on E^*. Then, if X is concentrated on $x_o \in L$, we have

$$P\{X=x\} = \lambda(F^*)^{-1} \int_{F^*} \psi(t) e^{-i<t,x>} \, d\lambda(t), \; x \in x_o + L, \tag{3.31}$$

where $\lambda(F^*)$ is the Lebesguemeasure λ of F^*.

To prove, that the point probabilities (3.31) may be approximated by values of the normal density function, we need a condition, that ensures, that the lattice is sufficiently dense compared to the variance of X. Thus, besides the norm $||t|| = \chi_2(t,t)^{\frac{1}{2}}$ on E^*, we define the <u>lattice norm</u> of $t = t_1\eta_1 + \ldots + t_p\eta_p$ as

$$d(t) = (t_1^2 + \ldots + t_p^2)^{\frac{1}{2}}, \tag{3.32}$$

obtained by defining $(\eta_1, \ldots, \eta_p)$ as an orthonormal base on E^*. Thus, $d(t)$ measures t in "lattice units". The analogue of Theorem 3.3 now takes the following form. Consider a sequence (X_n) of lattice random variables on $x_n + L_n \in E$. The approximations to the point probabilities become

$$p_{s,n}(x) = f_{s,n}(x) \, (2\pi)^p / \lambda_{\chi_{2,n}}(F_n^*) \tag{3.33}$$

with obvious notation; $f_{s,n}$ is given by (2.15) and is the $(s-1)$'th order density approximation in the continuous case. In the sequel, let $\lambda_n = \lambda_{\chi_{2,n}}$ denote the Lebesgue measure on E^* induced by the variance of X_n.

Theorem 3.5. Let (X_n) be a sequence of lattice random variables as described above. Let (ε_n) and $(a_n(t))$ be sequences of positive numbers, $a_n(t)$ depending only on the direction $t/||t||_n$ of t, and suppose the following conditions are fulfilled as $n\to\infty$:

(a) $\rho_{s,n}(t) \leq a_n(t)^{-1}$ for all $t \in E^*$ (3.34)

(b) $a_n(t)^{-(s-1)} = O(\varepsilon_n) = o(1)$ uniformly in $t \in E^*$ (3.35)

(c) For any $\delta > 0$,

$$\int_{F_n^* \cap \{||t||_n \geq \delta a_n(t)\}} |\psi_n(t)| \, d\lambda_n(t) = O(\varepsilon_n) \tag{3.36}$$

(d) For some $\delta > 0$,

$$M(\delta a_n(t)/||t||_n) = O(a_n(t)^{-(s-1)}) \text{ uniformly in } t \in E^* \tag{3.37}$$

(e) A constant C exists, such that

$$||t||_n^2/d_n(t)^2 + C \geq \frac{2}{\pi^2} \log \varepsilon_n^{-1}, \quad t \in E^*, \quad n=1,2,\ldots \tag{3.38}$$

Then

$$\sup \{|P\{X_n=x\} - p_{s,n}(x)| \; ; \; x \in x_n + L_n\} = O(\varepsilon_n) \, \lambda_n(F_n^*)^{-1} \tag{3.39}$$

Proof. Exactly as in the proof of Theorem 3.3 it is shown, that

$$|P\{X_n=x\} - \lambda_n(F_n^*)^{-1} \int_{F_n^*} \psi_n(t) e^{-i<t,x>} d\lambda_n(t)|$$

$= O(\varepsilon_n) \, \lambda_n(F_n^*)^{-1}$ uniformly on the support of X_n.

Hence, it only remains to prove, that

$$\int_{E^* \setminus F_n^*} |\psi_{s,n}(t)| \, d\lambda_n(t) = O(\varepsilon_n) \,. \tag{3.40}$$

Since $t \notin F_n^*$ implies $d_n(t) \geq \pi$, (3.38) shows that

$$E^* \setminus F_n^* \subseteq \{ ||t||_n^2 \geq 2 \log \varepsilon_n^{-1} - \pi^2 C \} .$$

Estimating the left hand side of (3.40) by use of Lemma 4.1 of Skovgaard (1981a), the result follows.

□

Remark 3.6. Note, that the infimum of the ratio $||t||_n^2/d_n(t)^2$ in (3.38) is just the smallest eigenvalue of $\chi_{2,n}$ in terms of the inner product generated by the lattice. Thus, (3.38) is the extra condition ensuring, that the variance grows sufficiently quickly compared to the distances between lattice points. That the bound of (3.39) becomes $O(\varepsilon_n)\lambda_n(F_n^*)^{-1}$ instead of just $O(\varepsilon_n)$ is necessary to get a useful result, since $\lambda_n(F_n^*)^{-1}$ is approximately proportional to the point probabilities around the "center" of the distribution.

As in the continuous case, it is the condition (3.36), that may be hard to verify. For sums of independent lattice random variables the analogue of Lemma 3.4 only differs in that the Condition I disappears, because there are no problems with the integrability of ψ_n. In this sense the lattice case is simpler than the continuous one.

In the case of sums of i.i.d. random variables on the integer lattice, $d_n(t)$ will be independent of n , while $||t||_n$ will be proportional to $\sqrt{n}$. For an $(s-1)$'th order Edgeworth expansion ε_n will be $\sqrt{n}^{-(s-1)}$. $a_n(t)$ will be of order $\sqrt{n}$, such that (3.34), (3.35) and (3.38) are trivially valid. Assuming the existence of moments of order $s+1$, (3.37) follows from the continuity of M around zero. (3.36) follows from (3.28), since $\gamma_n(t)$ will increase proportionally to n. Thus the expansions of i.i.d. random variables follow easily from these results, and the resulting expansion will be a power series expansion in $\sqrt{n}^{-1}$.

Suppose now, that local expansion of the form (3.20) or (3.39) has been established, what can we say about the corresponding approximation of the measure as a set function? A direct integration over an unbounded set results in general in an unbounded error term, but we can take advantage of the form of the approximation to prove that we can still integrate the approximation without losing control of the error. The idea is as follows. Suppose the approximating measure is normal and the error of the local expansion is $\varepsilon<1$. For some fixed (small) $\delta>0$ define the set

$$A = \{x\in E \mid \|x\|^2 \leq 2(1+\delta)\log \varepsilon^{-1}\} \tag{3.41}$$

Then for any measurable set B we may bound the error $|P(B)-\Phi(B)|$ by the error within A plus $P(A^c)+\Phi(A^c)$. Since the volume of A is $O((\log \varepsilon^{-1})^p)$ we get

$$|P(A\cap B)-\Phi(A\cap B)| = O(\varepsilon[\log \varepsilon^{-1}]^p) \quad \text{as} \quad \varepsilon\to 0 . \tag{3.42}$$

By Skovgaard (1981a), Lemma 4.1, $\Phi(A^c)$ is $O(\varepsilon)$, while

$$P(A^c) = 1-P(A) \leq 1-Q(A)+|P(A)-Q(A)| ,$$

such that also this may be bounded by the "within A error". A precise version of this argument is given in Skovgaard (1981c). Lemma 7.2. for the Edgeworth expansions, using the set A in (3.41) the final error term for the measure becomes

$$O(\varepsilon_n[\log \varepsilon_n^{-1}]^p) \quad \text{as} \quad n\to\infty \quad (\varepsilon_n\to 0) \tag{3.43}$$

For the usual case of a power series expansion, this logarithmic factor is of no importance, since the error will still be of smaller order than the last term of the expansion. In particular, if the expansion with one extra term is valid, this term determines the error, and the logarithmic term then disappears.

While the integration of $f_{s,n}(x)$ in the continuous case is simple, because the distribution function is a linear combination of derivatives of the normal distribution function, the summation causes great problems in the discrete case. Using a multivariate version of the Euler-MacLaurin summation formula, Bhattacharya & Rao (1976) derives an expansion of the distribution function. Summation of the point probabilities within ellipsoids or even spheres is extremely complicated and relates to problems of analytic number theory, such as expanding the number of integer lattice points within a ball of radius R as $R\to\infty$. Some work on the problem has been done by Esseen (1945).

An example. Consider the example of Section 6 in Skovgaard (1981a), i.e. $X_1, X_2, \ldots$ independent, $(\alpha+\beta t_i)X_i$ being distributed as gamma with shape parameter $\omega_i>0$, and $\alpha>0$, $\beta>0$, $t_i>0$. With the "directional" approach of this section it is possible to show the validity of the Edgeworth expansion of any order of the sufficient statistic $(T_1,T_2) = (X_i, t_iX_i)$, without the condition $(\log n)/m_n\to 0$ where m_n is the smallest eigenvalue of the covariance matrix of (T_1,T_2). As in the paper it then follows, that the distribution of the maximum likelihood estimator may be expanded to any order solely under the conditions

$$\sum\omega_i(t_i-\bar{t})^2\to\infty \; , \quad \sum\omega_i/t_i^2\to\infty \quad ,$$

where $\bar{t} = \sum\omega_i t_i / \sum\omega_i$. The error term of the (s-1)'th order expansion becomes $O(m_n^{-(s-1)/2})$. We shall not go through the calculations in details, since they are similar to those given in the paper, but only give a few remarks on the differences.

Instead of (6.3) we obtain (in the same way) with $u^2+v^2 = 1$

$$|K_\nu((u,v)^{\otimes\nu})| \leq (\nu-1)!\; K_2((u,v)^{\otimes 2}) \; , \tag{3.44}$$

such that $\rho_{s,n}((u,v)) \leq s\; K_2(u,v)^{-\frac{1}{2}} \sqrt{u^2+v^2}$. We may take $a_n((u,v)) = \rho_{s,n}((u,v))^{-1}$ and $\varepsilon_n = m_n^{-\frac{1}{2}}$. Then (3.16) and (3.17) are obvious, and (3.19) follows from the analyticity of the cumulant generating function and (3.44). (3.18) is proved by Lemma 3.4 as in the paper, but again with a directional approach to establish (3.28).

Appendix A. Tensor Products.

We shall give a brief summary of some basic concepts related to tensor products of vectorspaces. Our use of these concepts is mainly notational and we only use some of the simplest results of this theory; first of all some isomorphisms between different spaces. By use of this appendix, it should be possible for the reader to understand our notation used in Section 2 and 3 or to "translate" it to more familiar expressions, while the reader, who is familiar with multivariate algebra will appreciate the simplicity with which multivariate calculations can be handled in this language. An introduction to the theory of multivariate algebra may be found e.g. in Greub (1967).

Let $E, F, G, E_1, E_2, \ldots$ be finite-dimensional real vectorspaces, and $E^*, F^*, \ldots$ their duals, i.e. the spaces of linear mappings into $\mathbb{R}$.

Theorem A.1. Given $E_1, \ldots, E_k$ there exists a unique (up to isomorphisms) pair $(E_1 \otimes \ldots \otimes E_k, \mu)$ of a real vectorspace and a map μ: $E_1 \times \ldots \times E_k \to E_1 \otimes \ldots \otimes E_k$ with the following properties:

(a) μ is k-linear (linear in each component).

(b) For each k-linear map $A : E_1 \times \ldots \times E_k \to F$, there is a unique linear map $g : E_1 \otimes \ldots \otimes E_k \to F$, such that $A = g \circ \mu$.

Proof. Omitted.

The pair $(E_1 \otimes \ldots \otimes E_k, \mu)$ is called the tensor product of $E_1, \ldots, E_k$. Sometimes $E_1 \otimes \ldots \otimes E_k$ will be referred to as the tensor product. If $x_1 \in E_1, \ldots, x_k \in E_k$, then $\mu(x_1, \ldots, x_k)$ is called the tensor product of $x_1, \ldots, x_k$ and is denoted $x_1 \otimes \ldots \otimes x_k$. Such an element of $E_1 \otimes \ldots \otimes E_k$ is called an elementary tensor. Any tensor $a \in E_1 \otimes \ldots \otimes E_k$ can be written as a finite sum of elementary tensors. The dimension of $E_1 \otimes \ldots \otimes E_k$ is the product of the dimensions of the E_i's. The k-fold tensor product $E\otimes\ldots\otimes E$ of E with itself is denoted $E^{\otimes k}$.

If $A\in E^*$ and $x\in E$ we let $\langle A,x\rangle = A(x)$ denote their inner product to emphasize the symmetry arising from the fact, that E is isomorphic to the dual of E^*.

Theorem A.2. The spaces $(E_1\otimes\ldots\otimes E_k)^*$ and $E_1^*\otimes\ldots\otimes E_k^*$ are isomorphic, the isomorphism being given by

$$\langle A_1\otimes\ldots\otimes A_k\ ,\ x_1\otimes\ldots\otimes x_k\rangle = A_1(x_1)\ldots A_k(x_k) \tag{A.1}$$

when $A_i\in E_i^*$, $x_i\in E_i$.

Proof. Omitted.

In fact since $\mathbb{R}\otimes\ldots\otimes\mathbb{R}$ is isomorphic to $\mathbb{R}$, Theorem A.2 follows from the more general construction of tensorproducts of linear mappings. If $f_i:E_i\to F_i$ are linear mappings, then there is a unique linear map

$$f_1\otimes\ldots\otimes f_k\ :\ E_1\otimes\ldots\otimes E_k\to F_1\otimes\ldots\otimes F_k \tag{A.2}$$

such that $(f_1\otimes\ldots\otimes f_k)\ (x_1\otimes\ldots\otimes x_k) = f_1(x_1)\ \ldots\ f_k(x_k)$ for all $x_1\in E_1,\ldots,x_k\in E_k$. This map is called the tensor product of $f_1,\ldots,f_k$.

We shall need the concept of a symmetric tensor in $E^{\otimes k}$, defined below in two equivalent ways.

Definition A.3. A tensor $a\in E^{\otimes k}$ is called symmetric, if it can be written as a finite sum of elementary tensors of the form $x^{\otimes k}$, $x\in E$, or equivalently if the following condition holds: For any $A_1\in E^*,\ldots,A_k\in E^*$ and any permutation σ on $\{1,\ldots,k\}$ we have

$$<A_1\otimes\ldots\otimes A_k \; , \; a> \; = \; <A_{\sigma(1)}\otimes\ldots\otimes A_{\sigma(k)} \; , \; a> \tag{A.3}$$

□

Each tensor $a\in E^{\otimes k}$ has a unique symmetric equivalent under the equivalence relation $\sim$ given by

$$a_1 \sim a_2 \iff <A,a_1> \; = \; <A,a_2> \text{ for all symmetric } A\in(E^*)^{\otimes k} \; .$$

In particular we denote <u>the symmetric equivalent of an elementary tensor</u> $x_1\otimes\ldots\otimes x_k$ by $x_1\odot\ldots\odot x_k$ and call $\odot$ the symmetrical tensor product; a purely formal concept. We have

$$x_1\odot\ldots\odot x_k \; = \; \frac{1}{k!}\sum x_{\sigma(1)}\otimes\ldots\otimes x_{\sigma(k)} \; , \tag{A.4}$$

where the sum is over all permutations σ on $\{1,\ldots,k\}$. It should be noted, that the $\odot$ product as defined here is not standard, and differs from that in Federer (1969), Ch.1, first of all by a constant factor and secondly by taking values in $E^{\otimes k}$ instead of in the space of equivalence classes. The simplicity of the concept as used here and our rare use of it compensates for this difference.

A matrix-like representation of the tensor product can be given as follows. Let $x = (x_1,\ldots,x_k)\in E$ and $y = (y_1,\ldots,y_m)\in F$ be coordinate versions of two vectors, then $x\otimes y\in E\otimes F$ may be represented by the $k\times m$ matrix (x_iy_j), and addition of tensors corresponds to addition of the elements of their matrices. If $A = (a_{ij})$ is the $k\times m$ matrix of a bilinear map, then it is seen, that since

$$x'Ay \; = \; \sum_i \sum_j a_{ij}(x_iy_j) \; ,$$

A may be regarded as a linear mapping of the tensor product $E \otimes F$, represented by $k \times m$ matrices, into $\mathbb{R}$. In the same way tensor products of k vectorspaces may be represented by k-dimensional arrays of numbers, etc.

<u>Moments and cumulants</u> of a random variable $X \in E$ will be considered as tensors in the following way. If $A_1,\dots,A_k \in E^*$, then the joint k'th moment of $\langle A_1,X\rangle, \dots, \langle A_k,X\rangle$, i.e. the expectation of $\langle A_1,X\rangle \dots \langle A_k,X\rangle$ may be written

$$\mu_k(A_1,\dots,A_k) = E\{\langle A_1,X\rangle \dots \langle A_k,X\rangle\} , \tag{A.5}$$

such that the k'th moment μ_k is a k-linear map of $E^* \times \dots \times E^*$ into $\mathbb{R}$, or equivalently

$$\mu_k \in (E^* \otimes \dots \otimes E^*)^* \simeq E^{\otimes k} . \tag{A.6}$$

Similarly for the k'th cumulant $\chi_k \in E^{\otimes k}$. Particular attention should be given to the variance

$$\chi_2 = \mu_2 - \mu_1 \otimes \mu_1 \in E^{\otimes 2} .$$

As a bilinear form on E^* it defines an <u>inner product</u> and a norm $||A||^2 = \chi_2(A^{\otimes 2})$ on E^* if it is regular, i.e. if no linear function of X is degenerate. Thus E^* becomes a <u>Euclidean space</u>. Also χ_2 defines an isomorphism between E and E^* by letting $\chi_2(A) \in E$ be given by $\langle \chi_2(A),B\rangle = \chi_2(A \otimes B)$, $A,B \in E^*$.

As such $\chi_2 \in \mathrm{Lin}(E^*,E)$, the space of linear mappings of E^* into E, and the inverse mapping $\chi_2^{-1} \in \mathrm{Lin}(E,E^*)$ may in turn be regarded as an element of $(E^*)^{\otimes 2}$, thus defining an inner product on E . χ_2 and χ_2^{-1} are then <u>isometric mappings</u> between the Euclidean spaces E and E^* .

The norms on E and E^* define norms on tensor products of these, e.g. if $A \in (E^*)^{\otimes k} \simeq (E^{\otimes k})^*$, then we define

$$||A|| = \sup\{|A(x^{\otimes k})| \mid x \in E , ||x|| \leq 1\} \quad \text{(A.7)}$$

Let us finally define the <u>polynomials</u> on a vectorspace. Let $B_k(E,F)$ denote <u>the k-linear mappings</u> of $E\times\ldots\times E$ into F , $k \in \mathbb{N}$, while $B_0(E,F)$ refers to F . By the <u>polynomials on E taking values in F</u> we shall mean the set Pol(E,F) of mapping $P:E \to F$ of the form

$$P(x) = A_0 + \sum_{k=1}^{n} A_k(x,\ldots,x) , n \in \mathbb{N} , A_k \in B_k(E,F) \quad \text{(A.8)}$$

or equivalently, since $B_k(E,F)$ is isomorphic to $\mathrm{Lin}(E^{\otimes k},F)$, we may write

$$P(x) = A_0 + \sum_{k=1}^{n} A_k(x^{\otimes k}) , n \in \mathbb{N} , A_k \in \mathrm{Lin}(E^{\otimes k}, F) . \quad \text{(A.9)}$$

The smallest possible n is called the <u>degree</u> of P , and the A's are called the <u>coefficients</u> of P .

Appendix B. Differentiability.

We shall give some notation and some not quite standard results on differentiability; especially on higher differentials of various kinds of mappings. Most of the material may be found in standard references on mathematical analysis; some of the results on higher differentials are briefly, but in precise and selfcontained form given in Federer (1969), 3.1.11.

We shall be concerned only with mappings between finite-dimensional normed vectorspaces E,F,G, etc. Which norm is chosen is immaterial. In this appendix we shall deal with real vectorspaces only, but in a few other places mappings between complex vectorspaces occur, and we use the results of this appendix without further remarks, since there are no essential differences. Also, since we are not trying to minimize regularity conditions, we shall mainly deal with continuously differentiability to avoid unnecessary technicalities.

Let $f:B\to F$, be a mapping defined on an open set $B\subseteq E$ and let $x_0\in B$.

Definition B.1. f is said to be differentiable at x_0, if there exists a linear mapping $Df(x_0): E\to F$, such that

$$f(x)\sim f(x_0) + Df(x_0)(x-x_0) + o(x-x_0) \text{ as } x\to x_0 \qquad (B.1)$$

If f is differentiable at x_0 in B, then $Df(x_0)\in Lin(E,F)$ is called the differential of f at x_0. If f is differentiable on a neighbourhood U of x_o (i.e. differentiable at any point in U), and the differential

$$Df : U\to Lin(E,F)$$

$$[x\to Df(x)]$$

is continuous at x_o , then f is said to be continuously differentiable at x_o .

Remark. Note, that $Lin(E,F)$ is a real vectorspace. We equip this space with the norm

$$||A|| = \sup \{ ||A(x)|| \;\big|\; ||x|| = 1 \} , \; A \in Lin(E,F) \qquad (B.2)$$

Some obvious notations are being used, e.g. $f:B \to F$ is differentiable, if f is differentiable on B , etc. We let $C^1(B,F)$ denote the class of continuously differentiable functions.

In matrix-notation, if $x = (x_1,\ldots,x_m)$ and $f(x) = (f_1(x),\ldots,f_n(x))$, then if f is differentiable at x ,

$$\left(\frac{df_i}{dx_j}(x) \right) = DF(x)_{ij}$$

is the (i,j)'th coordinate of the matrix of $Df(x)$.

Since the differential $Df : B \to Lin(E,F)$ also maps B into a finite-dimensional real vector space, we may consider the differentiability. Thus, we say that f is twice (continuously) differentiable at x_o , if DF is defined and (continuously) differentiable at x_o. In that case, the second differential, i.e. the differential of Df , is denoted D^2f , and $D^2f(x_o)$ maps E into $Lin(E,F)$. Thus, the mapping

$$D^2f : B \to Lin(E,Lin(E,F)) \qquad (B.3)$$

Similarly, we define $C^p(B,F)$, the class of p times continuously differentiable mappings of B into F , recursively, and also the p'th differential of $f \in C^p(B,F)$ by

$$D^p f = D(D^{p-1}f) : B \to \mathrm{Lin}(E,\ldots,\mathrm{Lin}(E,F)) \qquad (B.4)$$

The class $C^\infty(B,F)$ of infinitely often differentiable functions of B into F is the intersection of the classes $C^p(B,F)$, $p \in N$.

If $f \in C^2(B,F)$, then $D^2f(x_o)$, $x_o \in B$, may be regarded as a bilinear mapping of $E \times E$ into F , by the definition $(D^2f(x_o)(x_1))(x_2) = D^2f(x_o)(x_1,x_2)$, since it is linear in both of the arguments $x_1,x_2 \in E$. One of the main theorems on differentiability is the following:

<u>Theorem B.2. If</u> $f \in C^2(B,F)$ <u>and</u> $x \in B$, <u>then</u> $D^2F(x)$ <u>belongs to the class</u> $B_2(E,F)$ <u>of symmetric, bilinear mappings of</u> $E \times E$ <u>into</u> F .

<u>Proof</u>. Omitted.

Analogously, if $f \in C^p(B,F)$, $D^pf(x)$ may be regarded as a p-linear mapping of $E\times\ldots\times E$ into F , and it is a consequence of Theorem B.2, that it is symmetric.

As a p-linear mapping , $D^pf(x)$ factorizes through the <u>tensor product</u> $E^{\otimes p}$, such that $D^pf(x)$ may be regarded as a linear mapping of $E^{\otimes p}$ into F. We shall use the same notation for this mapping, writing

$D^pf(x)\colon E\times\ldots\times E \to F$ or $D^pf(x)\colon E^{\otimes p} \to F$ and if $x_1,\ldots,x_p \in E$, then

$$D^pf(x)(x_1,\ldots,x_p) = D^pf(x)(x_1 \otimes \ldots \otimes x_p) \qquad (B.5)$$

are two ways of writing the same thing. The coordinates of (B.5) are, if $x_i = (x_{i1},\dots,x_{im})$ and $f(x) = (f_1(x),\dots,f_n(x))$,

$$[D^p f(x)(x_1,\dots,x_p)]_j$$

$$= \sum_{k_1=1}^{m} \cdots \sum_{k_p=1}^{m} \left(\frac{d^p}{dx_{k_1}\cdots dx_{k_p}} f_j(x) \right) x_{1k_1}\cdots x_{pk_p} \tag{B.6}$$

such that the "matrix" of $D^p f(x)$ is the $m\times\dots\times m\times n$ array of partial derivatives at x .

Since $D^p f(x)$ is symmetric, it is given by its values on the "diagonal", i.e.

$$D^p f(x)(y^{\otimes p}) = \frac{d^p}{du^p} f(x+uy)\Big|_{u=0} \ , \ y\in E \ , \ u\in\mathbb{R} \ , \tag{B.7}$$

which may be denoted the p'th <u>directional derivative</u> in the direction y .

In sequel, we shall shortly review some of the theorems on higher differentials, that we shall need.

<u>Theorem B.4.</u> (<u>Taylor's formula</u>). If $f\in C^p(B,F)$, $x,x_o\in B$, then

$$\text{(a)}\ f(x) = f(x_o) + Df(x_o)(x-x_o)+\dots+\frac{1}{p!}D^p f(x_o)\left((x-x_o)^{\otimes p}\right) + o(||x-x_o||^p) \ \text{as} \ x\to x_o \tag{B.8}$$

$$\text{(b)}\ f_j(x) = f_j(x_o)+Df_j(x_o)(x-x_o)+\dots+\frac{1}{(p-1)!}D^{p-1}f_j(x_o)\left((x-x_o)^{\otimes p-1}\right) + \frac{1}{p!}D^p f_j(x^*)\left((x-x_o)^{\otimes p}\right) \tag{B.9}$$

where $x^* = x_o+\theta(x-x_o)$, $0<\theta<1$.

Proof. Omitted.

In some of the following formulae, we shall use the set $T(p)$ of partitions of an integer p. We define

$$T(p) = \{\nu = (\nu_1,\ldots,\nu_p) \mid \sum_{j=1}^{p} j\nu_j = p \; , \; \nu_i \geq 0\} \; , \qquad (B.10)$$

the ν_i's being required to be integers.

Theorem B.5. (The chain rule). Let $f \in C^p(B,F)$ and $g \in C^p(D,G)$, where $f(B) \subseteq D \subseteq F$. Then $g \circ f \in C^p(B,G)$ and for $x \in B$, $z = f(x)$ we have

$$D^p(g \circ f)(x)(y^{\otimes p})$$

$$= \sum_{\nu \in T(p)} p! \; D^{\Sigma \nu_j} g(z)[(Df(x)(y))^{\otimes \nu_1} \otimes \ldots \otimes (D^p f(x)(y^{\otimes p}))^{\otimes \nu_p}]$$

$$/ \prod_{j=1}^{p} \nu_j! \; j!^{\nu_j} \qquad (B.11)$$

Proof. See Federer (1969), 3.1.11.

□

Theorem B.6. Let $A \in \mathrm{Lin}(E^{\otimes s}, F)$ and define $f : E \to F$ by $f(x) = A(x^{\otimes s})$. Then, $f \in C^\infty(E,F)$ and

$$D^p f(x_o)(x_1 \otimes \ldots \otimes x_p) = \begin{cases} 0 & \text{if} \quad p > s \\ A\left(x_1 \odot \ldots \odot x_p \odot x_o^{\odot (s-p)}\right) & \text{if} \quad p \leq s \end{cases} \qquad (B.12)$$

where the $\odot$ product is the symmetrized version of the tensor product; see (A.4).

Proof. Omitted.

Notice, that Theorem B.6 tells us how to differentiate a polynomial, since (B.12) is the derivative of one of its terms. If A is symmetric, then the $\odot$ product may be replaced by the $\otimes$ product in (B.12).

Now, let F,G and H be equipped with a product $F\times G\to H$ $[(\phi,\gamma)\to\phi\gamma]$, which satisfies

(I) $||\phi\gamma||\leq||\phi||\ ||\gamma||$, $(\phi,\gamma)\in F\times G$

(II) $(\lambda_1\phi_1+\lambda_2\phi_2)\gamma = \lambda_1\phi_1\gamma+\lambda_2\phi_2\gamma$, $\lambda_1,\lambda_2\in\mathbb{R}$

(III) $\phi(\lambda_1\gamma_1+\lambda_2\gamma_2) = \lambda_1\phi\gamma_1+\lambda_2\phi\gamma_2$, $\lambda_1,\lambda_2\in\mathbb{R}$

An example of a product of this kind is composition of functions.

Theorem B.7. (The product rule). Let $f\in C^p(B,F)$ and $g\in C^p(B,G)$, then the function $fg : B\to H$ given by $(fg)(x) = f(x)g(x)$ belongs to $C^p(B,H)$, and

$$(D^p(fg)(x))(y^{\otimes p}) = \sum_{j=0}^{p} \binom{p}{j}(D^j f(x)(y^{\otimes j}))(D^{p-j}g(x)(y^{\otimes p-j})) \qquad (B.13)$$

Proof. See Federer (1969),3.1.11.

□

Theorem B.7 tells how to differentiate a product of two real (or complex) functions; in this setting it is known as Leibnitz' formula. It also tells how to differentiate a composition of two functions with respect to a parameter, a case of frequent occurrence in expansions in statistics. It should not be confused with the chain rule concerned with differentiation with respect to the (only) argument of the function.

Analogously, there are two rules for differentiation of the inverse of a function; one in the usual sense of expressing the differentials of a functions inverse (with respect to composition) in terms of the differentials of the function itself; the other in the sense of a product structure, e.g. differentiation of one divided by a real function or differentiation of the inverse of a function with respect to a parameter.

The first of these rules is exactly Lemma 4.3 of Skovgaard (1981a) and we shall not repeat it here. Note, that it does not give a formula for the higher differentials in general only the structure of these. In terms of coordinates a formula is known even for implicit functions; see Bolotov & Yuzhakov (1978).

For the other inverse-rule we shall give two versions, the first in an algebraic setting, the second adapted to our purpose.

Theorem B.8. Let A be a normed algebra, and let a^{-1} be a right-inverse of $a \in A$, i.e. $aa^{-1} = e$. Denote this inverse mapping by $R: a \to a^{-1}$, and let B be the domain of R. Then if $a \in \mathrm{int} B$, R is infinitely often differentiable at a, and

$$D^k R(a)(x^{\otimes k}) = (-1)^k k! a^{-1} x a^{-1} x \ldots a^{-1} x a^{-1} \quad , \quad k \in N \tag{B.14}$$

where k x's appear in the product.

Proof. Omitted.

Theorem B.9. Let $E = \mathrm{Lin}(G,H)$ and $F = \mathrm{Lin}(H,G)$, where $\dim G = \dim H$, and denote elements of E by $f,g,h,\ldots$ Let R be the mapping $f \to f^{-1}$ defined on the subset of E of bijective mappings. Then R is infinitely

<u>often differentiable, and</u>

$$D^kR(f)(h^{\otimes k}) = (-1)^k k!f^{-1}hf^{-1}h \ldots hf^{-1} \quad , k\in N,\ f,h\in E \tag{B.15}$$

<u>if</u> f <u>is bijective</u>.

<u>Proof.</u> Omitted.

A more familier way of writing (B.15) is

$$(f+h)^{-1} \sim f^{-1} - f^{-1}hf^{-1} + f^{-1}hf^{-1}hf^{-1} - \ldots . \tag{B.16}$$

as $h\to 0$.

Note, that Theorem B.9 is not a special case of Theorem B.8, since f and f^{-1} are elements of different spaces.

References

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

Bhattacharya, R.N. & Rao, R.R. (1976). Normal approximation and asymptotic expansions. Wiley, New York.

Bolotov, V.A. & Yuzhakov, A.P. (1978). A generalization of the inversion formulas of systems of power series in systems of implicit functions. (Russian). Mat. Zametki 23, 47-54.

Chambers, J.M. (1967). On methods of asymptotic approximation for multivariate distributions. Biometrika 54, 367-383.

Chandra, T.K. & Ghosh, J.K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. Sankhya A 41, 22-47.

Chibisov, D.M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. Theor. Probability Appl. 17, 620-630.

Chibisov, D.M. (1973a). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. Theor. Probability Appl. 18, 295-303.

Chibisov, 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.

Esseen, C.G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta Math. 77 1-125.

Federer, H. (1969). Geometric measure theory. Springer, New York.

Greub, W.H. (1967). Multilinear algebra. Springer, Berlin.

Ivanov, A.V. (1976). An asymptotic expansion for the distribution of the least squares estimator of the non-linear regression parameter. Theor. Probability Appl. 21, 557-570.

Michel, R. (1979). Asymptotic expansions for conditional distributions. J.Multivariate Anal. 9. 393-400.

Petrov, V.V. (1975). Sums of independent random variables. Springer, Berlin.

Pfanzagl, J. (1973). Asymptotic expansions related to minimum contrast estimators. Ann. Statist. 1, 993-1026

Skovgaard, I.M. (1981a). Transformation of an Edgeworth expansion by a sequence of smooth functions. Scand. J. Statist. To appear.

Skovgaard (1981b). Edgeworth expansions of the distributions of maximum likelihood estimators in the general (non i.i.d.) case. Scand. J. Statist. To appear.

Skovgaard (1981c). A second order investigation of asymptotic ancillarity. Preprint 7, Univ. of Copenhagen, Inst. Math. Stat.

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.

No. 3 Asmussen, Søren: Equilibrium Properties of the M/G/1 Queue.

No. 4 Johansen, Søren and Keiding, Susanne: A Family of Models for the Elimination of Substrate in the Liver.

No. 5 Skovgaard, Ib M.: Edgeworth Expansions of the Distributions of Maximum Likelihood Estimators.

No. 6 Tjur, Tue: A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model.

No. 7 Lauritzen, Steffen L. : Examples of Extreme Point Models in Statistics.

PREPRINTS 1981

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

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Søren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.

No. 3 Lauritzen, Steffen.: Time Series Analysis in 1880. A Discussion of Contributions made by T.N. Thiele.

No. 4 Asmussen, Søren: Conditioned Limit Theorems Relating a Random Walk to its Associate, with Applications to Risk Reserve Processes and the GI/G/1 Queue.

No. 5 Johansen, Søren: The Statistical Analysis of a Markov Branching Process.

No. 6 Asmussen, Søren: Time - Dependent Approximations in some Queueing Systems with Imbedded Markov Chains Related to Random Walks.

No. 7 Skovgaard, Ib M.: A Second - Order Investigation of Asymptotic Ancillarity.

No. 8 Rootzén, Holger: The Rate of Extremes of Stationary Normal Sequences.

No. 9 Skovgaard, Ib M.: Large Deviation Approximations for Maximum Likelihood Estimators.

No. 10 Jensen, Ulla Funck: A Stochastic Projection Model with Implications for Multistate Demography and Manpower Analysis.

No. 11 Johansen, Søren: An Extension of Cox's Regression Model.

No. 12 Skovgaard, Ib M.: Edgeworth Expansions in Statistics.