CHAPTER 5

Edgeworth series

5.1 Introduction

While the lower-order cumulants of X are useful for describing, in both qualitative and quantitative terms, the shape of the joint distribution, it often happens that we require either approximations for the density function itself or, more commonly, approximations for tail probabilities or conditional tail probabilities. For example, if X^1 is a goodness-of-fit statistic and $X^2, \ldots, X^p$ are estimates of unknown parameters, it would often be appropriate to assess the quality of the fit by computing the tail probability

$$\mathrm{pr}(X^1 \geq x^1 | X^2 = x^2, \ldots, X^p = x^p),$$

small values being taken as evidence of a poor fit. The cumulants themselves do not provide such estimates directly and it is necessary to proceed via an intermediate step where the density, or equivalently, the cumulative distribution function, is approximated by a series expansion. Series expansions of the Edgeworth type all involve an initial first-order approximation multiplied by a sum of correction terms whose coefficients are simple combinations of the cumulants of X. In the case of the Edgeworth expansion itself, the first-order approximation is based on the normal density having the same mean vector and covariance matrix as X. The correction terms then involve cumulants and cumulant products as coefficients of the Hermite polynomials. More generally, however, in order to minimize the number of correction terms, it may be advantageous to use a first-order approximation other than the normal. The correction terms then involve derivatives of the approximating density and are not necessarily polynomials.

In Section 5.2 we discuss the nature of such approximations for arbitrary initial approximating densities. These series take on different forms depending on whether we work with the density function, the log density function, the cumulative distribution function or some transformation of the cumulative distribution function. For theoretical calculations the log density is often the most convenient: for the computation of significance levels, the distribution function may be preferred because it is more directly useful.

The remainder of the chapter is devoted to the Edgeworth expansion itself and to derived expansions for conditional distributions.

5.2 A formal series expansion

5.2.1 *Approximation for the density*

Let $f_X(x;\kappa)$ be the joint density function of the random variables $X^1, \ldots, X^p$. The notation is chosen to emphasize the dependence of the density on the cumulants of X. Suppose that the initial approximating density, $f_0(x) \equiv f_0(x;\lambda)$ has cumulants λ^i, $\lambda^{i,j}$, $\lambda^{i,j,k}, \ldots$, which differ from the cumulants of X by

$$\eta^i = \kappa^i - \lambda^i, \quad \eta^{i,j} = \kappa^{i,j} - \lambda^{i,j}, \quad \eta^{i,j,k} = \kappa^{i,j,k} - \lambda^{i,j,k}$$

and so on. Often, it is good policy to choose $f_0(x;\lambda)$ so that $\eta^i = 0$ and $\eta^{i,j} = 0$, but it would be a tactical error to make this assumption at an early stage in the algebra, in the hope that substantial simplification would follow. While it is true that this choice makes many terms vanish, it does so at the cost of ruining the symmetry and structure of the algebraic formulae.

The cumulant generating functions for $f_X(x;\kappa)$ and $f_0(x;\lambda)$ are assumed to have their usual expansions

$$K_X(\xi) = \xi_i\kappa^i + \xi_i\xi_j\kappa^{i,j}/2! + \xi_i\xi_j\xi_k\kappa^{i,j,k}/3! + \cdots$$

and

$$K_0(\xi) = \xi_i\lambda^i + \xi_i\xi_j\lambda^{i,j}/2! + \xi_i\xi_j\xi_k\lambda^{i,j,k}/3! + \cdots.$$

Subtraction gives

$$K_X(\xi) = K_0(\xi) + \xi_i\eta^i + \xi_i\xi_j\eta^{i,j}/2! + \xi_i\xi_j\xi_k\eta^{i,j,k}/3! + \cdots$$

and exponentiation gives

$$M_X(\xi) = M_0(\xi)\{1 + \xi_i\eta^i + \xi_i\xi_j\eta^{ij}/2! + \xi_i\xi_j\xi_k\eta^{ijk}/3! + \cdots\} \tag{5.1}$$

where

$$\begin{aligned}\eta^{ij} &= \eta^{i,j} + \eta^i\eta^j,\\ \eta^{ijk} &= \eta^{i,j,k} + \eta^i\eta^{j,k}[3] + \eta^i\eta^j\eta^k\\ \eta^{ijkl} &= \eta^{i,j,k,l} + \eta^i\eta^{j,k,l}[4] + \eta^{i,j}\eta^{k,l}[3] + \eta^i\eta^j\eta^{k,l}[6] + \eta^i\eta^j\eta^k\eta^l\end{aligned}$$

and so on, are the formal 'moments' obtained by treating η^i, $\eta^{i,j}$, $\eta^{i,j,k}, \ldots$ as formal 'cumulants'. It is important here to emphasize that $\eta^{ij} \neq \kappa^{ij} - \lambda^{ij}$, the difference between second moments of X and those of $f_0(x)$. Furthermore, the ηs are not, in general, the cumulants of any real random variable. For instance, $\eta^{i,j}$ need not be positive definite.

To obtain a series approximation for $f_X(x;\kappa)$, we invert the approximate integral transform, (5.1), term by term. By construction, the leading term, $M_0(\xi)$, transforms to $f_0(x)$. The second term, $\xi_i M_0(\xi)$, transforms to $f_i(x) = -\partial f_0(x)/\partial x^i$ as can be seen from the following argument. Integration by parts with respect to x^r gives

$$\int_{-\infty}^{\infty} \exp(\xi_i x^i)\frac{\partial f_0(x)}{\partial x^r}\,dx^r = \exp(\xi_i x^i) f_0(x) \Big|_{x^r=-\infty}^{x^r=\infty} - \xi_r \int_{-\infty}^{\infty} \exp(\xi_i x^i) f_0(x)\,dx^r.$$

For imaginary ξ, $\exp(\xi_i x^i)$ is bounded, implying that the first term on the right is zero. Further integration with respect to the remaining $p-1$ variables gives

$$\int_{R^p} \exp(\xi_i x^i) f_r(x) dx = \xi_r M_0(\xi).$$

The critical assumption here is that $f_0(x)$ should have continuous partial derivatives everywhere in R^p. For example, if $f_0(x)$ had discontinuities, these would give rise to additional terms in the above integration.

By the same argument, if $f_0(x)$ has continuous second-order partial derivatives, it may be shown that $\xi_r\xi_s M_X(\xi)$ is the integral transform of $f_{rs}(x) = (-1)^2\partial^2 f_0(x)/\partial x^r\partial x^s$ and so on. In other words, (5.1) is the integral transform of

$$\begin{aligned} f_X(x) = f_0(x) + \eta^i f_i(x) + \eta^{ij} f_{ij}(x)/2!\\ + \eta^{ijk} f_{ijk}(x)/3! + \cdots \end{aligned} \tag{5.2}$$

where the alternating signs on the derivatives have been incorporated into the notation.

Approximation (5.2) looks initially very much like a Taylor expansion in the sense that it involves derivatives of a function divided by the appropriate factorial. In fact, this close resemblance can be exploited, at least formally, by writing

$$X = Y + Z$$

where Y has density $f_0(y)$ and Z is a pseudo-variable independent of Y with 'cumulants' η^i, $\eta^{i,j}$, $\eta^{i,j,k}, \ldots$ and 'moments' η^i, η^{ij}, $\eta^{ijk}, \ldots$. Conditionally on $Z = z$, X has density $f_0(x - z)$ and hence, the marginal density of X is, formally at least,

$$f_X(x) = E_Z\{f_0(x - Z)\}.$$

Taylor expansion about $Z = 0$ and averaging immediately yields (5.2). This formal construction is due to Davis (1976).

To express (5.2) as a multiplicative correction to $f_0(x)$, we write

$$f_X(x;\kappa) = f_0(x)\{1 + \eta^i h_i(x) + \eta^{ij} h_{ij}(x)/2! \\ + \eta^{ijk} h_{ijk}(x)/3! + \cdots\} \tag{5.3}$$

where

$$h_i(x) = h_i(x;\lambda) = f_i(x)/f_0(x), \quad h_{ij}(x) = h_{ij}(x;\lambda) = f_{ij}(x)/f_0(x)$$

and so on. In many instances, $h_i(x)$, $h_{ij}(x), \ldots$ are simple functions with pleasant mathematical properties. For example, if $f_0(x) = \phi(x)$, the standard normal density, then $h_i(x)$, $h_{ij}(x), \ldots$ are the standard Hermite tensors, or polynomials in the univariate case. See Sections 5.3 and 5.4.

5.2.2 *Approximation for the log density*

For a number of reasons, both theoretical and applied, it is often better to consider series approximations for the log density $\log f_X(x;\kappa)$ rather than for the density itself. One obvious advantage, particularly where polynomial approximation is involved, is that $f_X(x;\kappa) \geq 0$, whereas any polynomial approximation is liable to become negative for certain values of x. In addition, even if the infinite series (5.3) could be guaranteed positive, there is no similar guarantee for the truncated series that would actually be used in practice. For these reasons, better approximations may be obtained by approximating $\log f_X(x;\kappa)$ and exponentiating.

Expansion of the logarithm of (5.3) gives

$$\begin{aligned} \log f_X(x;\kappa) \simeq{} & \log f_0(x) + \eta^i h_i(x) \\ & + \{\eta^{i,j} h_{ij}(x) + \eta^i \eta^j h_{i,j}(x)\}/2! \\ & + \{\eta^{i,j,k} h_{ijk}(x) + \eta^i \eta^{j,k} h_{i,jk}(x)[3] + \eta^i \eta^j \eta^k h_{i,j,k}(x)\}/3! \\ & + \{\eta^{i,j,k,l} h_{ijkl} + \eta^i \eta^{j,k,l} h_{i,jkl}[4] + \eta^{i,j} \eta^{k,l} h_{ij,kl}[3] \\ & \quad + \eta^i \eta^j \eta^{k,l} h_{i,j,kl}[6] + \eta^i \eta^j \eta^k \eta^l h_{i,j,k,l}\}/4! \end{aligned} \tag{5.4}$$

where

$$\begin{aligned} h_{i,j} &= h_{ij} - h_i h_j \\ h_{i,j,k} &= h_{ijk} - h_i h_{jk}[3] + 2h_i h_j h_k \\ h_{i,jk} &= h_{ijk} - h_i h_{jk} \\ h_{ij,kl} &= h_{ijkl} - h_{ij} h_{kl} \end{aligned}$$

and so on. Compare equations (2.7) and (3.2). Note that, apart from sign, the hs with fully partitioned indices are derivatives of the log density

$$\begin{aligned} h_i(x) &= -\partial \log f_0(x)/\partial x^i \\ h_{i,j}(x) &= \partial^2 \log f_0(x)/\partial x^i \partial x^j \\ h_{i,j,k}(x) &= -\partial^3 \log f_0(x)/\partial x^i \partial x^j \partial x^k \end{aligned}$$

and so on.

The h-functions with indices partially partitioned are related to each other in exactly the same way as generalized cumulants. Thus, for example, from (3.2) we find

$$h_{i,jk} = h_{i,j,k} + h_j h_{i,k} + h_k h_{i,j}$$

and

$$h_{i,jkl} = h_{i,j,k,l} + h_j h_{i,k,l}[3] + h_{i,j} h_{k,l}[3] + h_{i,j} h_k h_l[3],$$

using the conventions of Chapter 3.

Often it is possible to choose $f_0(x)$ so that its first-order and second-order moments agree with those of X. In this special case, (5.4) becomes

$$\begin{aligned} \log f_X(x) - \log f_0(x) &= \eta^{i,j,k} h_{ijk}(x)/3! + \eta^{i,j,k,l} h_{ijkl}(x)/4! \\ &+ \eta^{i,j,k,l,m} h_{ijklm}(x)/5! + \eta^{i,j,k,l,m,n} h_{ijklmn}(x)/6! \\ &+ \eta^{i,j,k} \eta^{l,m,n} h_{ijk,lmn}(x)[10]/6! + \cdots . \end{aligned} \tag{5.5}$$

This simplification greatly reduces the number of terms required but at the same time it manages to conceal the essential simplicity of the terms in (5.4). The main statistical reason for keeping to the more general expansion (5.4) is that, while it is generally possible to chose $f_0(x)$ to match the first two moments of X under some null hypothesis H_0, it is generally inconvenient to do so under general alternatives to H_0, as would be required in calculations related to the power of a test.

5.2.3 *Approximation for the cumulative distribution function*

One of the attractive features of expansions of the Edgeworth type is that the cumulative distribution function

$$F_X(x;\kappa) = \mathrm{pr}(X^1 \le x^1, \ldots, X^p \le x^p; \kappa)$$

can easily be approximated by integrating (5.2) term by term. If $F_0(x)$ is the cumulative distribution function corresponding to $f_0(x)$, this gives

$$\begin{aligned} F_X(x;\kappa) \simeq F_0(x) + \eta^i F_i(x) + \eta^{ij} F_{ij}(x)/2! \\ + \eta^{ijk} F_{ijk}(x)/3! + \cdots \end{aligned} \tag{5.6}$$

where

$$F_i(x) = (-1)\partial F_0(x)/\partial x^i; \quad F_{ij}(x) = (-1)^2 \partial^2 F_0(x)/\partial x^i \partial x^j$$

and so on with signs alternating.

Of course, we might prefer to work with the multivariate survival probability

$$S_X(x) = \mathrm{pr}(X^1 \ge x^1, X^2 \ge x^2, \ldots, X^p \ge x^p; \kappa)$$

and the corresponding probability, $S_0(x)$, derived from $f_0(x)$. Integration of (5.2) term by term gives

$$S_X(x;\kappa) \simeq S_0(x) + \eta^i S_i(x) + \eta^{ij} S_{ij}(x)/2! + \eta^{ijk} S_{ijk}(x)/3! + \cdots$$

where

$$S_i(x) = -\partial S_0(x)/\partial x^i; \qquad S_{ij}(x) = (-1)^2 \partial^2 S_0(x)/\partial x^i \partial x^j$$

and so on with signs alternating as before.

In the univariate case, $p = 1$, but not otherwise, $F_X(x;\kappa) + S_X(x;\kappa) = 1$ if the distribution is continuous, and the correction terms in the two expansions sum to zero giving $F_0(x) + S_0(x) = 1$.

Other series expansions can be found for related probabilities. We consider here only one simple example, namely

$$\bar{S}_X(x;\kappa) = \mathrm{pr}(X^1 \le x^1 \text{ or } X^2 \le x^2 \text{ or } \cdots \text{ or } X^p \le x^p; \kappa)$$

as opposed to $F_X(x;\kappa)$, whose definition involves replacing 'or' with 'and' in the expression above. By definition, in the continuous case, $\bar{S}_X(x;\kappa) = 1 - S_X(x;\kappa)$. It follows that

$$\bar{S}_X(x;\kappa) \simeq \bar{S}_0(x) + \eta^i \bar{S}_i(x) + \eta^{ij}\bar{S}_{ij}(x)/2! + \eta^{ijk}\bar{S}_{ijk}(x)/3! + \cdots$$

where

$$\bar{S}_i(x) = -\partial \bar{S}_0(x)/\partial x^i; \quad \bar{S}_{ij}(x) = (-1)^2 \partial^2 \bar{S}_0(x)/\partial x^i \partial x^j$$

and so on, again with alternating signs on the derivatives.

5.3 Expansions based on the normal density

5.3.1 *Multivariate case*

Suppose now, as a special case of (5.2) or (5.3) that the initial approximating density is chosen to be $f_0(x) = \phi(x;\lambda)$, the normal density with mean vector λ^i and covariance matrix $\lambda^{i,j}$. The density may be written

$$\phi(x;\lambda) = (2\pi)^{-p/2}|\lambda^{i,j}|^{-1/2}\exp\{-\tfrac{1}{2}(x^i - \lambda^i)(x^j - \lambda^j)\lambda_{i,j}\}$$

where $\lambda_{i,j}$ is the matrix inverse of $\lambda^{i,j}$ and $|\lambda^{i,j}|$ is the determinant of the covariance matrix. The functions $h_i(x) = h_i(x;\lambda)$, $h_{ij}(x) = h_{ij}(x;\lambda), \ldots$, obtained by differentiating $\phi(x;\lambda)$, are known as the Hermite tensors. The first six are given below

$$\begin{aligned}
h_i &= \lambda_{i,j}(x^j - \lambda^j) \\
h_{ij} &= h_i h_j - \lambda_{i,j} \\
h_{ijk} &= h_i h_j h_k - h_i \lambda_{j,k}[3] \\
h_{ijkl} &= h_i h_j h_k h_l - h_i h_j \lambda_{k,l}[6] + \lambda_{i,j}\lambda_{k,l}[3] \\
h_{ijklm} &= h_i h_j h_k h_l h_m - h_i h_j h_k \lambda_{l,m}[10] + h_i \lambda_{j,k}\lambda_{l,m}[15] \\
h_{ijklmn} &= h_i \cdots h_n - h_i h_j h_k h_l \lambda_{m,n}[15] + h_i h_j \lambda_{k,l}\lambda_{m,n}[45] \\
&\qquad - \lambda_{i,j}\lambda_{k,l}\lambda_{m,n}[15]
\end{aligned} \tag{5.7}$$

The general pattern is not difficult to describe: it involves summation over all partitions of the indices, unit blocks being associated with h_i, double blocks with $-\lambda_{i,j}$ and partitions having blocks of three or more elements being ignored.

In the univariate case, these reduce to the Hermite polynomials, which form an orthogonal basis with respect to $\phi(x;\lambda)$ as weight function, for the space of functions continuous over $(-\infty, \infty)$. The Hermite tensors are the polynomials that form a similar orthogonal basis for functions continuous over R^p. Their properties are discussed in detail in Section 5.4.

The common practice is to chose $\lambda^i = \kappa^i$, $\lambda^{i,j} = \kappa^{i,j}$, so that

$$\eta^i = 0, \quad \eta^{i,j} = 0, \quad \eta^{i,j,k} = \kappa^{i,j,k}, \quad \eta^{i,j,k,l} = \kappa^{i,j,k,l}$$
$$\eta^{ijk} = \kappa^{i,j,k}, \quad \eta^{ijkl} = \kappa^{i,j,k,l}, \quad \eta^{ijklm} = \kappa^{i,j,k,l,m},$$
$$\eta^{ijklmn} = \kappa^{i,j,k,l,m,n} + \kappa^{i,j,k}\kappa^{l,m,n}[10]$$

and so on, summing over all partitions ignoring those that have blocks of size 1 or 2. Thus (5.3) gives

$$\begin{aligned} f_X(x;\kappa) \simeq \phi(x;\kappa)\{1 &+ \kappa^{i,j,k}h_{ijk}(x)/3! + \kappa^{i,j,k,l}h_{ijkl}(x)/4! \\ &+ \kappa^{i,j,k,l,m}h_{ijklm}(x)/5! \\ &+ (\kappa^{i,j,k,l,m,n} + \kappa^{i,j,k}\kappa^{l,m,n}[10])h_{ijklmn}(x)/6! + \cdots\}. \end{aligned} \tag{5.8}$$

This form of the approximation is sometimes known as the Gram-Charlier series. To put it in the more familiar and useful Edgeworth form, we note that if X is a standardized sum of n independent random variables, then $\kappa^{i,j,k} = O(n^{-1/2})$, $\kappa^{i,j,k,l} = O(n^{-1})$ and so on, decreasing in power of $n^{-1/2}$. Thus, the successive correction terms in the Gram-Charlier series are of orders $O(n^{-1/2})$, $O(n^{-1})$, $O(n^{-3/2})$ and $O(n^{-1})$ and these are not monotonely decreasing in n. The re-grouped series, formed by collecting together terms that are of equal order in n,

$$\begin{aligned} \phi(x;\kappa)\Big[1 &+ \kappa^{i,j,k}h_{ijk}(x)/3! \\ &+ \{\kappa^{i,j,k,l}h_{ijkl}(x)/4! + \kappa^{i,j,k}\kappa^{l,m,n}h_{ijklmn}(x)[10]/6!\} \\ &+ \{\kappa^{i,j,k,l,m}h_{ijklm}(x)/5! + \kappa^{i,j,k}\kappa^{l,m,n,r}h_{i\cdots r}(x)[35]/7! \\ &\quad + \kappa^{i,j,k}\kappa^{l,m,n}\kappa^{r,s,t}h_{i\cdots t}(x)[280]/9!\} + \cdots\Big] \end{aligned} \tag{5.9}$$

is called the Edgeworth series and is often preferred for statistical calculations. The infinite versions of the two series are formally identical and the main difference is that truncation of (5.8) after a fixed number of terms gives a different answer than truncation of (5.9) after a similar number of terms.

At the origin of x, the approximating density takes the value

$$\phi(0;\kappa)\left[1 + 3\rho_4/4! - \{9\rho_{13}^2 + 6\rho_{23}^2\}/72 + O(n^{-2})\right] \tag{5.10}$$

where ρ_4, ρ_{13}^2 and ρ_{23}^2 are the invariant standardized cumulants of X. Successive terms in this series decrease in whole powers of n as opposed to the half-powers in (5.9).

5.3.2 *Univariate case*

The univariate case is particularly important because the notion of a tail area, as used in significance testing, depends on the test statistic being one-dimensional. In the univariate case it is convenient to resort to the conventional power notation by writing

$$\begin{aligned} h_1(x;\kappa) &= \kappa_2^{-1}(x-\kappa_1) \\ h_2(x;\kappa) &= h_1^2 - \kappa_2^{-1} \\ h_3(x;\kappa) &= h_1^3 - 3\kappa_2^{-1}h_1 \end{aligned}$$

and so on for the Hermite polynomials. These are derived in a straightforward way from the Hermite tensors (5.7). The standard Hermite polynomials, obtained by putting $\kappa_1 = 0$, $\kappa_2 = 1$ are

$$\begin{aligned} h_1(z) &= z \\ h_2(z) &= z^2 - 1 \\ h_3(z) &= z^3 - 3z \\ h_4(z) &= z^4 - 6z^2 + 3 \\ h_5(z) &= z^5 - 10z^3 + 15z \\ h_6(z) &= z^6 - 15z^4 + 45z^2 - 15. \end{aligned} \tag{5.11}$$

The first correction term in (5.9) is

$$\begin{aligned}\kappa_3 h_3(x;\kappa)/6 &= \kappa_3\{\kappa_2^{-3}(x-\kappa_1)^3 - 3\kappa_2^{-2}(x-\kappa_1)\}/6 \\ &= \rho_3 h_3(z)/6 = \rho_3(z^3-3z)/6\end{aligned}$$

where $z = \kappa_2^{-1/2}(x-\kappa_1)$ and $Z = \kappa_2^{-1/2}(X-\kappa_1)$ is the standardized version of X. Similarly, the $O(n^{-1})$ correction term becomes

$$\rho_4 h_4(z)/4! + 10\rho_3^2 h_6(z)/6!$$

Even in theoretical calculations it is rarely necessary to use more than two correction terms from the Edgeworth series. For that reason, we content ourselves with corrections up to order $O(n^{-1})$, leaving an error that is of order $O(n^{-3/2})$.

For significance testing, the one-sided tail probability may be approximated by

$$\begin{aligned}\mathrm{pr}(Z \ge z) \simeq 1 - \Phi(z) \\ + \phi(z)\{\rho_3 h_2(z)/3! + \rho_4 h_3(z)/4! + \rho_3^2 h_5(z)/72\},\end{aligned} \tag{5.12}$$

which involves one correction term of order $O(n^{-1/2})$ and two of order $O(n^{-1})$. The two-sided tail probability is, for $z > 0$,

$$\begin{aligned}\mathrm{pr}(|Z| \ge z) = 2\{1 - \Phi(z)\} \\ + 2\phi(z)\{\rho_4 h_3(z)/4! + \rho_3^2 h_5(z)/72\}\end{aligned} \tag{5.13}$$

and this involves only correction terms of order $O(n^{-1})$. In essence, what has happened here is that the $O(n^{-1/2})$ corrections are equal in magnitude but of different sign in the two tails and they cancel when the two tails are combined.

5.3.3 *Regularity conditions*

So far, we have treated the series expansions (5.8) to (5.13) in a purely formal way. No attempt has been made to state precisely the way in which these series expansions are supposed to approximate either the density or the cumulative distribution function. In this section, we illustrate in an informal way, some of the difficulties encountered in making this notion precise.

First, it is not difficult to see that the infinite series expansion (5.3) or (5.8) for the density is, in general, divergent. For example, if the cumulant differences, η^i, $\eta^{i,j}$, $\eta^{i,j,k}, \ldots$ are all equal to 1, then $\eta^{ij} = 2$, $\eta^{ijk} = 5$, $\eta^{ijkl} = 15, \ldots$, otherwise known as the Bell numbers. These increase exponentially fast so that (5.3) fails to converge. For this reason we need a justification of an entirely different kind for using the truncated expansion, truncated after some fixed number of terms.

To this end, we introduce an auxiliary quantity, n, assumed known, and we consider the formal mathematical limit $n \to \infty$. The idea now is to approximate $K_X(\xi)$ by a truncated series such that after $r+2$ terms the error is $O(n^{-r/2})$, say. Formal inversion gives a truncated series expansion for the density or distribution function which, under suitable smoothness conditions, has an error also of order $O(n^{-r/2})$. In all the applications we have in mind, X is a standardized sum and $f_0(x)$ is the normal density having the same mean and variance as X. Then $\eta^i = 0$, $\eta^{i,j} = 0$, $\eta^{i,j,k} = O(n^{-1/2})$, $\eta^{i,j,k,l} = O(n^{-1})$ and so on. Thus the truncated expansion for $K_X(\xi)$, including terms up to degree 4 in ξ, has error $O(n^{-3/2})$. To obtain similar accuracy in the approximations for $M_X(\xi)$ or $f_X(x)$ it is necessary to go as far as terms of degree 6.

This rather informal description ignores one fairly obvious point. The truncated series approximation (5.5) for $F_X(x;\kappa)$ is continuous. If X has a discrete distribution, $F_X(x;\kappa)$ will have discontinuities with associated probabilities typically of order $O(n^{-1/2})$. Any continuous approximation for $F_X(x)$ will therefore have an error of order $O(n^{-1/2})$, however many correction terms are employed.

In the univariate case, if X is the standardized sum of n independent and identically distributed random variables, Feller (1971, Chapter XVI) gives conditions ensuring the validity of expansions (5.9) for the density and (5.12) for the cumulative distribution function. The main condition, that

$$\limsup_{\zeta\to\infty} |M_X(i\zeta)| < 1,$$

excludes lattice distributions but ensures that the error in (5.12) is $o(n^{-1})$ uniformly in z. With weaker conditions, but again excluding lattice distributions, the error is known to be $o(n^{-1/2})$ uniformly in z (Esseen, 1945; Bahadur & Ranga-Rao, 1960). In the lattice case, (5.12) has error $o(n^{-1/2})$ provided that it is used with the usual correction for continuity. An asymptotic expansion given by Esseen (1945) extends the Edgeworth expansion to the lattice case: in particular, it produces correction terms for (5.12) so that the renaining error is $O(n^{-3/2})$ uniformly in z.

Similar conditions dealing with the multivariate case are given by Barndorff-Nielsen & Cox (1979).

The case where X is a standardized sum of non-independent or non-identically distributed random variables is considerably more complicated. It is necessary, for example, to find a suitably strengthened version of the Lindeberg condition involving, perhaps, the higher-order cumulants. In addition, if X is a sum of discrete and continuous random variables, it would be necessary to know whether the discrete or the continuous components dominated in the overall sum. No attempt will be made to state such conditions here.

For details of regularity conditions, see Bhattacharya & Rao (1976) or Skovgaard, (1981a,b, 1986b).

5.4 Some properties of Hermite tensors

5.4.1 *Tensor properties*

We now investigate the extent to which the arrays of polynomials $h_i(x;\lambda)$, $h_{ij}(x;\lambda)$, $h_{ijk}(x;\lambda), \ldots$, defined apart from choice of sign, as the coefficient of $\phi(x;\lambda)$ in the partial derivatives of $\phi(x;\lambda)$, deserve to be called tensors. To do so, we must examine the effect on the polynomials of transforming from x to new variables $\bar{x}$. In doing so, we must also take into account the induced transformation on the λs, which are the formal cumulants of x.

Consider the class of non-singular affine transformations on x

$$\bar{x}^i = a^i + a^i_r x^r \tag{5.14}$$

together with the induced transformation on the λs

$$\bar{\lambda}^i = a^i + a^i_r \lambda^r \quad \text{and} \quad \bar{\lambda}^{i,j} = a^i_r a^j_s \lambda^{r,s}.$$

It is easily checked that

$$(x^i - \lambda^i)(x^i - \lambda^j)\lambda_{i,j} = (\bar{x}^i - \bar{\lambda}^i)(\bar{x}^j - \bar{\lambda}^j)\bar{\lambda}_{i,j} \tag{5.15}$$

so that this quadratic form is invariant under nonsingular affine transformation. If a^i_r does not have full rank, this argument breaks down. Apart from the determinantal factor $|\lambda^{i,j}|^{-1/2}$, which has no effect on the definition of the Hermite tensors, $\phi(x;\lambda)$ is invariant under the transformation

(5.14) because it is a function of the invariant quadratic form (5.15). It follows therefore that the derivatives with respect to $\bar{x}$ are

$$\frac{\partial\phi}{\partial x^r}\frac{\partial x^r}{\partial \bar{x}^i}, \qquad \frac{\partial^2\phi}{\partial x^r \partial x^s}\frac{\partial x^r}{\partial \bar{x}^i}\frac{\partial x^s}{\partial \bar{x}^j} + \frac{\partial\phi}{\partial x^r}\frac{\partial^2 x^r}{\partial \bar{x}^i \partial \bar{x}^j}$$

and so on. Since $\partial^2 x^r/\partial\bar{x}^i\partial\bar{x}^j \equiv 0$, these derivatives may be written in the form

$$b_i^r\phi_r, \quad b_i^r b_j^s \phi_{rs}, \quad b_i^r b_j^s b_k^t \phi_{rst}$$

and so on where b_i^r is the matrix inverse of a_r^i. Thus the derivatives of ϕ are Cartesian tensors and they transform in the covariant manner. It follows immediately that $h_i(x;\lambda)$, $h_{ij}(x;\lambda)$ and so on are also Cartesian tensors, justifying the terminology of Section 5.3.

The above result applies equally to the more general case discussed in Section 5.2 where the form of the initial approximating density $f_0(x;\lambda)$ is left unspecified. See Exercise 5.5.

5.4.2 *Orthogonality*

The principal advantage of index notation is its total transparency: the principal disadvantage, nowhere more evident than in Chapter 1 where tensors were defined, is that general expressions for tensors of arbitrary order are difficult to write down without introducing unsightly subscripted indices. For this reason, we have avoided writing down general expressions for Hermite tensors, even though the general pattern is clear and is easily described in a few words (Section 5.3.1). To prove orthogonality it is necessary either to devise a suitable general notation or to use a less direct method of proof.

Our method of proof uses the generating function

$$\begin{aligned}\exp\{\xi_i(x^i-\lambda^i) - \tfrac{1}{2}\xi_i\xi_j\lambda^{i,j}\} = 1 + \xi_i h^i + \xi_i\xi_j h^{ij}/2! \\ + \xi_i\xi_j\xi_k h^{ijk}/3! + \cdots\end{aligned} \tag{5.16}$$

where

$$h^i = \lambda^{i,r}h_r(x;\lambda), \qquad h^{ij} = \lambda^{i,r}\lambda^{j,s}h_{rs}(x;\lambda),$$
$$h^{ijk} = \lambda^{i,r}\lambda^{j,s}\lambda^{k,t}h_{rst}(x;\lambda)$$

and so on, are the contravariant expressions of the Hermite tensors. To show that (5.16) is indeed a generating function for the hs, we observe that

$$\phi(x^i - \lambda^{i,j}\xi_j;\lambda) = \phi(x;\lambda)\exp\{\xi_i(x^i-\lambda^i) - \xi_i\xi_j\lambda^{i,j}/2\}.$$

Taylor expansion about $\xi = 0$ of $\phi(x^i - \lambda^{i,j}\xi_j;\lambda)$ gives

$$\phi(x;\lambda)\{1 + \xi_i h^i + \xi_i\xi_j h^{ij}/2! + \xi_i\xi_j\xi_k h^{ijk}/3! + \cdots\},$$

which we have simplified using the definition of Hermite tensors. This completes the proof that $\exp\{\xi_i(x^i-\lambda^i) - \xi_i\xi_j\lambda^{i,j}/2\}$ is the generating function for the Hermite tensors.

To prove orthogonality, consider the product

$$\begin{aligned}&\phi(x^i - \lambda^{i,j}\zeta_j;\lambda)\exp\{\xi_i(x^i-\lambda^i) - \xi_i\xi_j\lambda^{i,j}/2\} \\ &= \phi(x;\lambda)\{1 + \zeta_i h^i + \zeta_i\zeta_j h^{ij}/2! + \cdots\}\{1 + \xi_i h^i + \xi_i\xi_j h^{ij}/2! + \cdots\}.\end{aligned}$$

Simplification of the exponent gives

$$\phi(x^i - \lambda^{i,j}\xi_j - \lambda^{i,j}\zeta_j;\lambda)\exp(\xi_i\zeta_j\lambda^{i,j})$$

and integration with respect to x over R^p gives $\exp(\xi_i \zeta_j \lambda^{i,j})$. Orthogonality follows because $\exp(\xi_i \zeta_j \lambda^{i,j})$ involves only terms of equal degree in ξ and ζ. Moreover, from the expansion

$$\exp(\xi_i \zeta_j \lambda^{i,j}) = 1 + \xi_i \zeta_j \lambda^{i,j} + \xi_i \xi_j \zeta_k \zeta_l \lambda^{i,k} \lambda^{j,l}/2! + \cdots,$$

it follows that the inner products over R^p are

$$\begin{aligned} \int h^i h^j \phi \, dx &= \lambda^{i,j} \\ \int h^{ij} h^{kl} \phi \, dx &= \lambda^{i,k} \lambda^{j,l} + \lambda^{i,l} \lambda^{j,k} \\ \int h^{ijk} h^{lmn} \phi \, dx &= \lambda^{i,l} \lambda^{j,m} \lambda^{k,m} [3!] \end{aligned} \tag{5.17}$$

and so on for tensors of higher order.

In the univariate case where the hs are standard Hermite polynomials, the orthogonality relations are usually written in the form $\int h_r(x) h_s(x) \phi(x) \, dx = r! \delta_{rs}$.

The extension of the above to scalar products of three or more Hermite polynomials or tensors is given in Exercises 5.9–5.14.

5.4.3 *Generalized Hermite tensors*

From any sequence of arrays h^i, h^{ij}, $h^{ijk}, \ldots$, each indexed by an unordered set of indices, there may be derived a new sequence h^i, $h^{i,j}$, $h^{i,j,k}, \ldots$, each indexed by a fully partitioned set of indices. If $M_h(\xi)$ is the generating function for the first set, then $K_h(\xi) = \log M_h(\xi)$ is the generating function for the new set and the relationship between the two sequences is identical to the relationship between moments and cumulants. In the case of Hermite tensors, we find on taking logs in (5.16) that

$$K_h(\xi) = \xi_i (x^i - \lambda^i) - \xi_i \xi_j \lambda^{i,j}/2 \tag{5.18}$$

so that $h^i = x^i - \lambda^i$, $h^{i,j} = -\lambda^{i,j}$ and all other arrays in the new sequence are identically zero. In fact, apart from choice of sign, the arrays in this new sequence are the partial derivatives of $\log \phi(x; \lambda)$ with respect to the components of x.

Just as ordinary moments and ordinary cumulants are conveniently considered as special cases of generalized cumulants, so too the sequences h^i, h^{ij}, h^{ijk}, $\cdots$ and h^i, $h^{i,j}$, $h^{i,j,k}$ are special cases of generalized Hermite tensors. Some examples with indices fully partitioned are given below.

$$\begin{aligned} h^{i,jk} &= h^{ijk} - h^i h^{jk} \\ &= h^{i,j,k} + h^j h^{i,k} + h^k h^{i,j} \\ &= -(x^j - \lambda^j) \lambda^{i,k} - (x^k - \lambda^k) \lambda^{i,j} \\ h^{i,jkl} &= h^{ijkl} - h^i h^{jkl} \\ &= h^{i,j,k,l} + h^j h^{i,k,l}[3] + h^{i,j} h^{k,l}[3] + h^{i,j} h^k h^l [3] \\ &= -h^k h^l \lambda^{i,j}[3] + \lambda^{i,j} \lambda^{k,l}[3] \\ h^{ij,kl} &= h^{ijkl} - h^{ij} h^{kl} \\ &= h^{i,j,k,l} + h^i h^{j,k,l}[4] + h^{i,k} h^{j,l}[2] + h^{i,k} h^j h^l [4] \\ &= -h^j h^l \lambda^{i,k}[4] + \lambda^{i,k} \lambda^{j,l}[2] \\ h^{i,j,kl} &= h^{ijkl} - h^i h^{jkl}[2] - h^{ij} h^{kl} + 2 h^i h^j h^{kl} \\ &= h^{i,j,k,l} + h^k h^{i,j,l}[2] + h^{i,k} h^{j,l}[2] \\ &= \lambda^{i,k} \lambda^{j,l}[2] \end{aligned}$$

In these examples, the first line corresponds to the expressions (3.2) for generalized cumulants in terms of moments. The second line corresponds to the fundamental identity (3.3) for generalized cumulants in terms of ordinary cumulants and the final line is obtained on substituting the coefficients in (5.18).

A generalized Hermite tensor involving β indices partitioned into α blocks is of degree $\beta-2\alpha+2$ in x provided that $\beta-2\alpha+2\geq 0$. Otherwise the array is identically zero.

The importance of these generalized Hermite tensors lies in the formal series expansion (5.4) for the log density, which is used in computing approximate conditional cumulants.

5.4.4 *Factorization of Hermite tensors*

Suppose that X is partitioned into two vector components $X^{(1)}$ and $X^{(2)}$ of dimensions q and $p-q$ respectively. We suppose also that the normal density $\phi(x;\lambda)$, from which the Hermite tensors are derived, is expressible as the product of two normal densities, one for $X^{(1)}$ and one for $X^{(2)}$. Using the indices $i,j,k,\ldots$ to refer to components of $X^{(1)}$ and $r,s,t,\ldots$ to refer to components of $X^{(2)}$, we may write $\lambda^{i,r}=0$ and $\lambda_{i,r}=0$. Since the two components in the approximating density are uncorrelated, there is no ambiguity in writing $\lambda_{i,j}$ and $\lambda_{r,s}$ for the matrix inverses of $\lambda^{i,j}$ and $\lambda^{r,s}$ respectively. More generally, if the two components were correlated, this notation would be ambiguous because the leading $q\times q$ sub-matrix of

$$\begin{bmatrix}\lambda^{i,j} & \lambda^{i,s}\\ \lambda^{r,j} & \lambda^{r,s}\end{bmatrix}^{-1},$$

namely $\{\lambda^{i,j}-\lambda^{i,r}\{\lambda^{r,s}\}^{-1}\lambda^{s,j}\}^{-1}$, is not the same as the matrix inverse of $\lambda^{i,j}$.

Since $h^{i,r}=-\lambda^{i,r}=0$ and all higher-order generalized Hermite tensors, with indices fully partitioned, are zero, it follows immediately from (3.3) that

$$\begin{aligned}h^{i,rs}=0;\quad h^{ij,r}=0;\quad h^{i,rst}=0;\quad h^{ij,rs}=0\\ h^{ijk,r}=0;\quad h^{i,j,rs}=0\end{aligned}$$

and so on for any sub-partition of $\{(i,j,k,\ldots),\ (r,s,t,\ldots)\}$. This is entirely analogous to the statement that mixed cumulants involving two independent random variables and no others are zero. The corresponding statement for moments involves multiplication or factorization so that

$$\begin{aligned}&h^{irs}=h^ih^{rs};\qquad h^{ijr}=h^{ij}h^r;\qquad h^{ijrs}=h^{ij}h^{rs};\\ &h^{i,jrs}=h^{i,j}h^{rs};\qquad h^{i,r,js}=h^{i,j}h^{r,s};\\ &h^{ir,js}=h^{i,j}h^{r,s}+h^ih^jh^{r,s}+h^{i,j}h^rh^s=h^{ij}h^{r,s}+h^{i,j}h^rh^s\end{aligned}$$

and so on.

5.5 Linear regression and conditional cumulants

5.5.1 *Covariant representation of cumulants*

For calculations involving conditional distributions or conditional cumulants, it is often more convenient to work not with the cumulants of X^i directly but rather with the cumulants of $X_i=\kappa_{i,j}X^j$, which we denote by $\kappa_i,\ \kappa_{i,j},\ \kappa_{i,j,k},\ldots$. We refer to these as the covariant representation of X and the covariant representation of its cumulants. This transformation may, at first sight, seem inconsequential but it should be pointed out that, while the notation X^i is unambiguous, the same cannot be said of X_i because the value of X_i depends on the entire set of variables $X^1,\ldots,X^p$. Deletion of X^p, say, leaves the remaining components $X^1,\ldots,X^{p-1}$ and their joint cumulants unaffected but the same is not true of $X_1,\ldots,X_{p-1}$.

5.5.2 *Orthogonalized variables*

Suppose now that X is partitioned into two vector components $X^{(1)}$ and $X^{(2)}$ of dimensions p and $q-p$ respectively and that we require the conditional cumulants of $X^{(1)}$ after linear regression on $X^{(2)}$. This is a simpler task than finding the conditional cumulants of $X^{(1)}$ given $X^{(2)} = x^{(2)}$, because it is assumed implicitly that only the conditional mean of $X^{(1)}$ and none of the higher-order conditional cumulants depends on the value of $x^{(2)}$. Further, each component of the conditional mean is assumed to depend linearly on $x^{(2)}$.

We first make a non-singular linear transformation from the original $(X^{(1)}, X^{(2)})$ to new variables $Y = (Y^{(1)}, Y^{(2)})$ in such a way that $Y^{(2)} = X^{(2)}$ and $Y^{(1)}$ is uncorrelated with $Y^{(2)}$. Extending the convention established in Section 5.4.4, we let the indices $i, j, k, \ldots$ refer to components of $X^{(1)}$ or $Y^{(1)}$, indices $r, s, t, \ldots$ refer to components of $X^{(2)}$ or $Y^{(2)}$ and indices $\alpha, \beta, \gamma, \ldots$ refer to components of the joint variable X or Y. The cumulants of X are κ^α partitioned into κ^i and κ^r, $\kappa^{\alpha,\beta}$ partitioned into $\kappa^{i,j}$ $\kappa^{i,r}$ and $\kappa^{r,s}$ and so on. The same convention applies to the covariant representation, κ_i, κ_r, $\kappa_{i,j}$, $\kappa_{i,r}$ and $\kappa_{r,s}$.

At this stage, it is necessary to distinguish between $\kappa_{r,s}$, the $(p-q) \times (p-q)$ sub-matrix of $[\kappa^{\alpha,\beta}]^{-1}$ and the $(p-q) \times (p-q)$ matrix inverse of $\kappa^{r,s}$, the covariance matrix of $Y^{(2)}$. The usual way of doing this is to distinguish the cumulants of Y from those of X by means of an overbar. Thus $\bar\kappa^{r,s} = \kappa^{r,s}$ is the covariance matrix of $Y^{(2)}$. In addition, $\bar\kappa^{i,j}$ is the covariance matrix of $Y^{(1)}$ and $\bar\kappa^{i,r} = 0$ by construction. The inverse matrix has elements $\bar\kappa_{r,s}$, $\bar\kappa_{i,j}$, $\bar\kappa_{i,r} = 0$, where $\bar\kappa^{r,s}$ $\bar\kappa_{s,t} = \delta^r_t$.

The linear transformation to orthogonalized variables $Y^{(1)}$ and $Y^{(2)}$ may be written

$$Y^i = X^i - \beta^i_r X^r; \qquad Y^r = X^r \tag{5.19}$$

where $\beta^i_r = \kappa^{i,s}\bar\kappa_{r,s}$ is the regression coefficient of X^i on X^r. This is the contravariant representation of the linear transformation but, for our present purposes, it will be shown that the covariant representation is the more convenient to work with. The covariance matrix of $Y^{(1)}$ is

$$\bar\kappa^{i,j} = \kappa^{i,j} - \kappa^{i,r}\kappa^{j,s}\bar\kappa_{r,s} = \{\kappa_{i,j}\}^{-1}. \tag{5.20}$$

Further, using formulae for the inverse of a partitioned matrix, we find

$$\begin{aligned}\kappa_{i,j} &= \bar\kappa_{i,j}\\ \kappa_{i,r} &= -\kappa_{i,j}\beta^j_r\\ \kappa_{r,s} &= \bar\kappa_{r,s} + \beta^i_r\beta^j_s\kappa_{i,j}.\end{aligned}$$

Hence the expressions for Y_i and Y_r are

$$\begin{aligned}Y_i &= \bar\kappa_{i,j}Y^j = \kappa_{i,j}Y^j = \kappa_{i,j}(X^j - \beta^j_r X^r)\\ &= \kappa_{i\alpha}X^\alpha = X_i\end{aligned}$$

and

$$\begin{aligned}Y_r &= \bar\kappa_{r,s}Y^s = \bar\kappa_{r,s}X^s = \bar\kappa_{r,s}\kappa^{s,\alpha}X_\alpha\\ &= X_r + \beta^i_r X_i.\end{aligned}$$

Thus, the covariant representation of the linear transformation (5.19) is

$$Y_i = X_i; \quad Y_r = X_r + \beta^i_r X_i. \tag{5.21}$$

It follows that the joint cumulants of $(Y^{(1)}, Y^{(2)})$ expressed in covariant form in terms of the cumulants of $(X^{(1)}, X^{(2)})$, are

$$\begin{aligned}&\bar\kappa_i = \kappa_i, && \bar\kappa_{i,j} = \kappa_{i,j}, && \bar\kappa_{i,j,k} = \kappa_{i,j,k}, \ldots\\ &\bar\kappa_{i,r} = 0, && \bar\kappa_{r,s} = \kappa_{r,s} + \beta^i_r\kappa_{i,s}[2] + \beta^i_r\beta^j_s\\ &\kappa_{i,j} = \kappa_{r,s} - \beta^i_r\beta^j_s\kappa_{i,j}\end{aligned}$$

$$\begin{aligned}
\bar{\kappa}_{i,j,r} &= \kappa_{i,j,r} + \beta_r^k \kappa_{i,j,k} \\
\bar{\kappa}_{i,r,s} &= \kappa_{i,r,s} + \beta_r^j \kappa_{i,j,s}[2] + \beta_r^j \beta_s^k \kappa_{i,j,k} \\
\bar{\kappa}_{r,s,t} &= \kappa_{r,s,t} + \beta_r^i \kappa_{i,s,t}[3] + \beta_r^i \beta_s^j \kappa_{i,j,t}[3] + \beta_r^i \beta_s^j \beta_r^k \kappa_{i,j,k}.
\end{aligned} \tag{5.22}$$

The main point that requires emphasis here is that, in covariant form, the cumulants of $Y^{(1)}$ are the same as those of $X^{(1)}$ and they are unaffected by the Gram-Schmidt orthogonalization (5.19).

The contravariant expressions for the cumulants of $(Y^{(1)}, Y^{(2)})$ are

$$\begin{aligned}
\bar{\kappa}^i &= \kappa^i - \beta_r^i \kappa^r \\
\bar{\kappa}^{i,j} &= \kappa^{i,j} - \beta_r^i \kappa^{r,j}[2] + \beta_r^i \beta_s^j \kappa^{r,s} = \kappa^{i,j} - \beta_r^i \beta_s^j \kappa^{r,s} \\
\bar{\kappa}^{i,j,k} &= \kappa^{i,j,k} - \beta_r^i \kappa^{r,j,h}[3] + \beta_r^i \beta_s^j \kappa^{r,s,k}[3] - \beta_r^i \beta_s^j \beta_r^k \kappa^{r,s,t} \\
\bar{\kappa}^{i,r} &= 0, \qquad \bar{\kappa}^{r,s} = \kappa^{r,s} \\
\bar{\kappa}^{i,j,r} &= \kappa^{i,j,r} - \beta_s^j \kappa^{i,r,s}[2] + \beta_s^i \beta_r^j \kappa^{r,s,t} \\
\bar{\kappa}^{i,r,s} &= \kappa^{i,r,s} - \beta_r^i \kappa^{r,s,t} \\
\bar{\kappa}^{r,s,t} &= \kappa^{r,s,t}
\end{aligned} \tag{5.23}$$

It is important to emphasize at this stage that $Y^{(1)}$ and $Y^{(2)}$ are not independent unless the strong assumptions stated in the first paragraph of this section apply. Among the cumulants given above, exactly two of the mixed third-order expressions are non-zero, namely $\kappa^{i,r,s}$ and $\kappa^{i,j,r}$. A non-zero value of $\kappa^{i,r,s}$ means that, although $E(Y^i|X^{(2)})$ has no linear dependence on $X^{(2)}$, there is some quadratic dependence on products $X^r X^s$. There is a close connection here with Tukey's (1949) test for non-additivity. Similarly, a non-zero value of $\kappa^{i,j,r}$ implies heterogeneity of covariance, namely that $\text{cov}(Y^i, Y^j|X^{(2)})$ depends linearly or approximately linearly on $X^{(2)}$. See (5.28).

5.5.3 *Linear regression*

Suppose now, in the notation of the previous section, that the orthogonalized variables $Y^{(1)}$ and $Y^{(2)}$ are independent. It follows that the second- and higher-order conditional cumulants of $X^{(1)}$ given $X^{(2)}$ are the same as the unconditional cumulants of $Y^{(1)}$. These are given in covariant form in the first set of equations (5.22) and in contravariant form in (5.23). To complete the picture, we require only the conditional mean of $X^{(1)}$. We find from (5.19) that

$$E\{X^i|X^{(2)}\} = \kappa^i + \beta_r^i(x^r - \kappa^r) = \kappa^i + \kappa^{i,r} h_r(x^{(2)}; \bar{\kappa})$$

where $h_r(x^{(2)}; \bar{\kappa}) = \bar{\kappa}_{r,s}(x^s - \kappa^s)$ is the first Hermite polynomial in the components of $X^{(2)}$.

The above expression may be written in covariant form as

$$E(X^i|X^{(2)}) = \bar{\kappa}^{i,j}\{\kappa_j - \kappa_{j,r} x^r\}$$

so that $\kappa_j - \kappa_{j,r} x^r$ is the covariant representation of the conditional mean of $X^{(1)}$.

These results may be summarized by a simple recipe for computing the conditional cumulants of $X^{(1)}$ after linear regression on $X^{(1)}$

(i) First compute the cumulants of $\kappa_{\alpha,\beta} X^\beta$, giving $\kappa_i\ \kappa_r$; $\kappa_{i,j}$, $\kappa_{i,r}$, $\kappa_{r,s}$; $\kappa_{i,j,k}, \ldots$.

(ii) Compute $\bar{\kappa}^{i,j}$ the $p \times p$ matrix inverse of $\kappa_{i,j}$.

(iii) Replace κ_i by $\kappa_i - \kappa_{i,r} x^r$.

(iv) Raise all indices by multiplying by $\bar{\kappa}^{i,j}$ as often as necessary, giving

$$\begin{aligned}
&\bar{\kappa}^{i,j}\{\kappa_j - \kappa_{j,r} x^r\} \\
&\bar{\kappa}^{i,k} \bar{\kappa}^{j,l} \kappa_{k,l} = \bar{\kappa}^{i,j} \\
&\bar{\kappa}^{i,j,k} = \bar{\kappa}^{i,l} \bar{\kappa}^{j,m} \bar{\kappa}^{k,n} \kappa_{l,m,n}
\end{aligned}$$

and so on for the transformed cumulants.

5.6 Conditional cumulants

5.6.1 *Formal expansions*

In the previous section we considered the simple case where only the conditional mean of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ and none of the higher-order cumulants depends on the value of $x^{(2)}$. In this special case it was possible to compute exact expressions for the conditional cumulants. More generally, all of the conditional cumulants may depend, to some extent, on the value of $x^{(2)}$. Our aim in this section is to use the series expansion (5.4), together with the results of the previous three sections, to find formal series expansions for at least the first four conditional cumulants and, in principle at least, for all of the conditional cumulants.

To simplify matters, we assume that the initial approximating density for $(X^{(1)}, X^{(2)})$ is the product of two normal densities, $\phi_1(x^{(1)}; \lambda^{(1)})\phi_2(x^{(2)}; \lambda^{(2)})$. The advantage derived from this choice is that the Hermite tensors factor into products as described in Section 5.4.4. In practice, it is often sensible to choose the argument $\lambda^{(1)}$ of ϕ_1 to be the mean vector and covariance matrix of $X^{(1)}$ but, for reasons given in Section 5.2.1, we shall not do so for the moment.

Expansion (5.4) for the logarithm of the joint density of $X^{(1)}$ and $X^{(2)}$ may be written

$$\begin{aligned}
&\log\phi_1(x^{(1)};\lambda^{(1)}) + \log\phi_2(x^{(2)};\lambda^{(2)}) + \eta^i h_i + \eta^r h_r \\
&\quad + \bigl\{\eta^{i,j}h_{ij} + \eta^{i,r}h_i h_r[2] + \eta^{r,s}h_{rs} + \eta^i\eta^j h_{i,j} + \eta^r\eta^s h_{r,s}\bigr\}/2! \\
&\quad + \bigl\{\bigl(\eta^{i,j,k}h_{ijk} + \eta^{i,j,r}h_{ij}h_r[3] + \eta^{i,r,s}h_i h_{rs}[3] + \eta^{r,s,t}h_{rst}\bigr) \\
&\quad\;\; + \bigl(\eta^i\eta^{j,k}h_{i,jk} + \eta^i\eta^{j,r}h_{i,j}h_r[2] + \eta^r\eta^{i,s}h_i h_{r,s}[2] + \eta^r\eta^{s,t}h_{r,st}\bigr)[3]\bigr\}/3! \\
&+ \bigl\{\bigl(\eta^{i,j,k,l}h_{ijkl} + \eta^{i,j,k,r}h_{ijk}h_r[4] + \eta^{i,j,r,s}h_{ij}h_{rs}[6] + \eta^{i,r,s,t}h_i h_{rst}[4] \\
&\qquad\qquad + \eta^{r,s,t,u}h_{rstu}\bigr) \\
&\quad + \bigl(\eta^i\eta^{j,k,l}h_{i,jkl} + \eta^i\eta^{j,k,r}h_{i,jk}h_r[3] + \eta^i\eta^{j,r,s}h_{i,j}h_{rs}[3] \\
&\qquad + \eta^r\eta^{i,j,s}h_{ij}h_{r,s}[3] + \eta^r\eta^{i,s,t}h_i h_{r,st}[3] + \eta^r\eta^{s,t,u}h_{r,stu}\bigr)[4] \\
&\quad + \bigl(\eta^{i,j}\eta^{k,l}h_{ij,kl} + \eta^{i,j}\eta^{k,r}h_{ij,k}h_r[4] + \cdots + \eta^{r,s}\eta^{t,u}h_{rs,tu}\bigr)[3] \\
&\quad + \bigl(\eta^i\eta^j\eta^{k,l}h_{i,j,kl} + \eta^i\eta^r\eta^{j,s}h_{i,j}h_{r,s}[4] + \eta^r\eta^s\eta^{t,u}h_{r,s,tu}\bigr)[6]\bigr\}/4! \\
&+ \cdots .
\end{aligned} \tag{5.24}$$

This series may look a little complicated but in fact it has a very simple structure and is easy to extrapolate to higher-order terms. Such terms will be required later. Essentially every possible combination of terms appears, except that we have made use of the identities $h_{i,j,k} = 0$, $h_{i,j,k,l} = 0$ and so on.

On subtracting the logarithm of the marginal distribution of $X^{(2)}$, namely

$$\log\phi_2(x^{(2)};\lambda^{(2)}) + \eta^r h_r + \{\eta^{r,s}h_{rs} + \eta^r\eta^s h_{r,s}\}/2! + \cdots$$

we find after collecting terms that the conditional log density of $X^{(1)}$ given $X^{(2)}$ has a formal Edgeworth expansion in which the first four cumulants are

$$\begin{aligned}
E(X^i|X^{(2)}) &= \kappa^i \\
&\quad + \eta^{i,r}h_r + \eta^{i,r,s}h_{rs}/2! + \eta^{i,r,s,t}h_{rst}/3! + \eta^{i,r,s,t,u}h_{rstu}/4! + \cdots \\
&\quad + \eta^r\eta^{i,s}h_{r,s} + \eta^r\eta^{i,s,t}h_{r,st}/2! + \eta^r\eta^{i,s,t,u}h_{r,stu}/3! + \cdots \\
&\quad + \eta^{i,r}\eta^{s,t}h_{r,st}/2! + \eta^{i,r}\eta^{s,t,u}h_{r,stu}/3! + \eta^{r,s}\eta^{i,t,u}h_{rs,tu}/(2!\,2!) + \cdots \\
&\quad + \eta^r\eta^{s,t}\eta^{i,u}h_{r,st,u}/2! + \cdots
\end{aligned}$$

$$\begin{aligned}\operatorname{cov}(X^i, X^j|X^{(2)}) &= \kappa^{i,j} + \eta^{i,j,r}h_r + \eta^{i,j,r,s}h_{rs}/2! + \eta^{i,j,r,s,t}h_{rst}/3! + \cdots \\ &+ \eta^r\eta^{i,j,s}h_{r,s} + \eta^r\eta^{i,j,s,t}h_{r,st}/2! + \cdots \\ &+ \eta^{i,r}\eta^{j,s}h_{r,s} + \eta^{i,r}\eta^{j,st}[2]h_{r,st}/2! + \eta^{r,s}\eta^{i,j,t}h_{rs,t}/2! + \cdots\end{aligned}$$

$$\begin{aligned}\operatorname{cum}(X^i, X^j, X^k|X^{(2)}) &= \kappa^{i,j,k} + \eta^{i,j,k,r}h_r + \eta^{i,j,k,r,s}h_{rs}/2! + \cdots \\ &+ \eta^r\eta^{i,j,k,s}h_{r,s} + \eta^r\eta^{i,j,k,s,t}h_{r,st}/2! + \cdots \\ &+ \eta^{i,r}\eta^{j,k,s}[3]h_{r,s} + \eta^{i,r}\eta^{j,k,s,t}[3]h_{r,st}/2! + \eta^{r,s}\eta^{i,j,k,t}h_{rs,t}/2! + \cdots \\ &+ \eta^{r,s,i}\eta^{j,k,t}[3]h_{rs,t}/2! + \cdots\end{aligned}$$

$$\begin{aligned}\operatorname{cum}(X^i, X^j, X^k, X^l|X^{(2)}) &= \kappa^{i,j,k,l} + \eta^{i,j,k,l,r}h_r + \eta^{i,j,k,l,r,s}h_{rs}/2! + \cdots \\ &+ \eta^r\eta^{i,j,k,l,s}h_{r,s} + \eta^{i,r}\eta^{j,k,l,s}[4]h_{r,s} + \eta^{i,j,r}\eta^{k,l,s}[3]h_{r,s} + \cdots \\ &+ \cdots . \end{aligned} \tag{5.25}$$

Of course, $\eta^{r,s,t} = \kappa^{r,s,t}$, $\eta^{r,s,i} = \kappa^{r,s,i}$, $\eta^{r,s} = \kappa^{r,s} - \lambda^{r,s}$ and so on, but the ηs have been retained in the above expansions in order to emphasize the essential simplicity of the formal series. We could, in fact, choose $\lambda^r = \kappa^r$, $\lambda^{r,s} = \kappa^{r,s}$ giving $\eta^r = 0$, $\eta^{r,s} = 0$. This choice eliminates many terms and greatly simplifies computations but it has the effect of destroying the essential simplicity of the pattern of terms in (5.25). Details are given in the following section.

5.6.2 *Asymptotic expansions*

If X is a standardized sum of n independent random variables, we may write the cumulants of X as 0, $\kappa^{\alpha,\beta}$, $n^{-1/2}\kappa^{\alpha,\beta,\gamma}$, $n^{-1}\kappa^{\alpha,\beta,\gamma,\delta}$ and so on. Suppose for simplicity that the components $X^{(1)}$ and $X^{(2)}$ are uncorrelated so that $\kappa^{i,r} = 0$. Then, from (5.25), the conditional cumulants of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ have the following expansions up to terms of order $O(n^{-1})$.

$$\begin{aligned}E(X^i|X^{(2)}) &= \kappa^i + n^{-1/2}\kappa^{i,r,s}h_{rs}/2! \\ &\qquad + n^{-1}\{\kappa^{i,r,s}\kappa^{t,u,v}h_{rs,tuv}/(3!\,2!) + \kappa^{i,r,s,t}h_{rst}/3!\}\end{aligned}$$

$$\begin{aligned}\operatorname{cov}(X^i, X^j|X^{(2)}) &= \kappa^{i,j} + n^{-1/2}\kappa^{i,j,r}h_r + n^{-1}\big\{\kappa^{i,j,r,s}h_{rs}/2! \\ &\qquad + \kappa^{i,j,r}\kappa^{s,t,u}h_{r,stu}/3! + \kappa^{i,r,s}\kappa^{j,t,u}h_{rs,tu}/(2!\,2!)\big\}\end{aligned}$$

$$\begin{aligned}\operatorname{cum}(X^i, X^j, X^k|X^{(2)}) &= n^{-1/2}\kappa^{i,j,k} \\ &\qquad + n^{-1}\{\kappa^{i,j,k,r}h_r + \kappa^{i,j,r}\kappa^{k,s,t}[3]h_{r,st}/2!\}\end{aligned}$$

$$\operatorname{cum}(X^i, X^j, X^k, X^l|X^{(2)}) = n^{-1}\{\kappa^{i,j,k,l} + \kappa^{i,j,r}\kappa^{k,l,s}[3]h_{r,s}\}.$$

In the above expansions, the Hermite tensors are to be calculated using the exact mean vector κ^r and covariance matrix $\kappa^{r,s}$ of $X^{(2)}$. Note that, to the order given, the conditional mean is a cubic function of $X^{(2)}$, the conditional covariances are quadratic, the conditional skewnesses are linear and the conditional kurtosis is constant, though not the same as the unconditional kurtosis. All higher-order cumulants are $O(n^{-3/2})$ or smaller.

5.7 Normalizing transformation

In the multivariate case, there is an infinite number of smooth transformations, $g(.)$, that make the distribution of $Y = g(X)$ normal to a high order of approximation. Here, in order to ensure a unique solution, at least up to choice of signs for the components, we ask that the transformation be *triangular*. In other words, Y^1 is required to be a function of X^1 alone, Y^2 is required to be a function of the pair X^1, X^2, and so on. Algebraically, this condition may be expressed by writing

$$\begin{aligned} Y^1 &= g_1(X^1) \\ Y^2 &= g_2(X^1, X^2) \\ Y^r &= g_r(X^1, \ldots, X^r) \qquad (r = 1, \ldots, p). \end{aligned}$$

To keep the algebra as simple as possible without making the construction trivial, it is assumed that X is a standardized random variable with cumulants

$$0, \qquad \delta^{ij}, \qquad n^{-1/2}\kappa^{i,j,k}, \qquad n^{-1}\kappa^{i,j,k,l}$$

and so on, decreasing in powers of $n^{1/2}$. Thus, X is standard normal to first order: the transformed variable, Y is required to be standard normal with error $O(n^{-3/2})$.

The derivation of the normalizing transformation is unusually tedious and rather unenlightening. For that reason, we content ourselves with a statement of the result, which looks as follows.

$$\begin{aligned} Y^i = X^i &- n^{-1/2}\left\{3\kappa^{i,r,s}h_{rs} + 3\kappa^{i,i,r}h_i h_r + \kappa^{i,i,i}h_{ii}\right\}/3! \\ &- n^{-1}\left\{4\kappa^{i,r,s,t}h_{rst} + 6\kappa^{i,i,r,s}h_{rs}h_i + 4\kappa^{i,i,i,r}h_{ii}h_r + \kappa^{i,i,i,i}h_{iii}\right\}/4! \\ &+ n^{-1}\left\{(36\kappa^{\alpha,i,r}\kappa^{\alpha,s,t} + 18\kappa^{i,i,r}\kappa^{i,s,t})h_{rst}\right. \\ &\quad + (18\kappa^{\alpha,i,i}\kappa^{\alpha,r,s} + 12\kappa^{i,i,i}\kappa^{i,r,s} + 36\kappa^{\alpha,i,r}\kappa^{\alpha,i,s} + 27\kappa^{i,i,r}\kappa^{i,i,s})h_{rs}h_i \\ &\quad + (36\kappa^{\alpha,i,i}\kappa^{\alpha,i,r} + 30\kappa^{i,i,i}\kappa^{i,i,r})h_{ii}h_r \\ &\quad \left. + (9\kappa^{\alpha,i,i}\kappa^{\alpha,i,i} + 8\kappa^{i,i,i}\kappa^{i,i,i})h_{iii}\right\}/72 \\ &+ n^{-1}\left\{(36\kappa^{\alpha,\beta,i}\kappa^{\alpha,\beta,r} + 36\kappa^{\alpha,i,i}\kappa^{\alpha,i,r} + 12\kappa^{i,i,i}\kappa^{i,i,r})h_r\right. \\ &\quad \left. + (18\kappa^{\alpha,\beta,i}\kappa^{\alpha,\beta,i} + 27\kappa^{\alpha,i,i}\kappa^{\alpha,i,i} + 10\kappa^{i,i,i}\kappa^{i,i,i})h_i\right\}/72 \end{aligned} \tag{5.26}$$

In the above expression, all sums run from 1 to $i-1$. Greek letters repeated as superscripts are summed but Roman letters are not. In other words, $\kappa^{\alpha,\beta,i}\kappa^{\alpha,\beta,i}$ is a shorthand notation for

$$\sum_{\alpha=1}^{i-1}\sum_{\beta=1}^{i-1}\kappa^{\alpha,\beta,i}\kappa^{\alpha,\beta,i}.$$

In addition, the index i is regarded as a fixed number so that

$$\kappa^{i,i,i,r}h_{ii}h_r = \sum_{r=1}^{i-1}\kappa^{i,i,i,r}h_{ii}h_r.$$

Fortunately, the above polynomial transformation simplifies considerably in the univariate case because most of the terms are null. Reverting now to power notation, we find

$$Y = X - \rho_3(X^2-1)/6 - \rho_4(X^3-3X)/24 + \rho_3^2(4X^3-7X)/36$$

to be the polynomial transformation to normality. In this expression, we have inserted explicit expressions for the Hermite polynomials. In particular, $4X^3 - 7X$ occurs as the combination $4h_3 + 5h_1$.

5.8 Bibliographic notes

The terms 'Edgeworth series' and 'Edgeworth expansion' stem from the paper by Edgeworth (1905). Similar series had previously been investigated by Chebyshev, Charlier, and Thiele (1897): Edgeworth's innovation was to group the series inversely by powers of the sample size rather than by the degree of the Hermite polynomial.

For a historical perspective on Edgeworth's contribution to Statistics, see the discussion paper by Stigler (1978).

Jeffreys (1966, Section 2.68) derives the univariate Edgeworth expansion using techniques similar to those used here.

Wallace (1958) gives a useful discussion in the univariate case, of Edgeworth series for the density and Cornish-Fisher series for the percentage points. See also Cornish & Fisher (1937).

Proofs of the validity of Edgeworth series can be found in the books by Cramér (1937) and Feller (1971). Esseen (1945) and Bhattacharya & Ranga-Rao (1976) give extensions to the lattice case. See also Chambers (1967) or Bhattacharya & Ghosh (1978).

Skovgaard (1981a) discusses the conditions under which a transformed random variable has a density that can be approximated by an Edgeworth series.

Michel (1979) discusses regularity conditions required for the validity of Edgeworth expansions to conditional distributions.

The notation used here is essentially the same as that used by Amari & Kumon (1983): see also Amari (1985). Skovgaard (1986) prefers to use coordinate-concealing notation for conceptual reasons and to deal with the case where the eigenvalues of the covariance matrix may not tend to infinity at equal rates.

5.9 Further results and exercises 5

5.1 Show, under conditions to be stated, that if

$$f_X(x;\kappa) = f_0(x) + \eta^i f_i(x) + \eta^{ij} f_{ij}(x)/2! + \eta^{ijk} f_{ijk}(x)/3! + \cdots$$

then the moment generating function of $f_X(x;\kappa)$ is

$$M_0(\xi)\{1 + \xi_i\eta^i + \xi_i\xi_j\eta^{ij}/2! + \xi_i\xi_j\xi_k\eta^{ijk}/3! + \cdots\}$$

where $M_0(\xi)$ is the moment generating function of $f_0(x)$.

5.2 Using expansion (5.2) for the density, derive expansion (5.4) for the log density.

5.3 Give a heuristic explanation for the formal similarity of expansions (5.2) and those in Section 5.2.3.

5.4 Show that any generalized Hermite tensor involving β indices partitioned into α blocks, is of degree $\beta - 2\alpha - 2$ in x or is identically zero if $\beta - 2\alpha - 2 < 0$.

5.5 If $f_X(x)$ is the density function of $X^1, \ldots, X^p$, show that the density of

$$Y^r = a^r + a^r_i X^i$$

is

$$f_Y(y) = J f_X\{b^i_r(y^r - a^r)\},$$

where $b^i_r a^r_j = \delta^i_j$ and J is the determinant of b^i_r. Hence show that the partial derivatives of $\log f_X(x)$ are Cartesian tensors.

5.6 Show that the mode of a density that can be approximated by an Edgeworth series occurs at

$$\hat{x}^i = -\kappa^{i,j,k}\kappa_{j,k}/2 + O(n^{-3/2}).$$

5.7 Show that the median of a univariate density that can be approximated by an Edgeworth series occurs approximately at the point

$$\hat{x} = \frac{-\kappa_3}{6\kappa_2}.$$

Hence show that, to the same order of approximation, in the univariate case,

$$\frac{(\text{mean} - \text{median})}{(\text{mean} - \text{mode})} = \frac{1}{3}$$

(Haldane, 1942). See also Haldane (1948) for a discussion of medians of multivariate distributions.

5.8 Let X be a normal random variable with mean vector λ^r and covariance matrix $\lambda^{r,s}$. Define

$$h^r = h^r(x;\lambda), \quad h^{rs}(x;\lambda), \ldots$$

to be the Hermite tensors based on the same normal distribution, i.e.,

$$h^r = x^r - \lambda^r$$
$$h^{rs} = h^r h^s - \lambda^{r,s}$$

and so on as in (5.7). Show that the random variables

$$h^r(X), \quad h^{rs}(X), \quad h^{rst}(X), \ldots$$

have zero mean and are uncorrelated.

5.9 Using the notation established in the previous exercise, show that

$$\mathrm{cum}\left(h^{rs}(X), h^{tu}(X), h^{vw}(X)\right) = \lambda^{r,r}\lambda^{s,v}\lambda^{t,w}[8]$$
$$\mathrm{cum}\left(h^{r}(X), h^{s}(X), h^{tu}(X)\right) = \lambda^{r,t}\lambda^{s,u}[2]$$
$$\mathrm{cum}\left(h^{r}(X), h^{st}(X), h^{uvw}(X)\right) = \lambda^{r,u}\lambda^{s,v}\lambda^{t,w}[6].$$

Give an expression for the cumulant corresponding to an arbitrary partition of the indices.

5.10 Suppose now that X has cumulants κ^r, $\kappa^{r,s}$, $\kappa^{r,s,t}, \ldots$, and that the Hermite tensors are based on the normal density with mean λ^r and covariance matrix $\lambda^{r,s}$. Show that

$$E\{h^r(X)\} = \eta^r$$
$$E\{h^{rs}(X)\} = \eta^{rs}$$
$$E\{h^{rst}(X)\} = \eta^{rst}$$

and so on, where the ηs are defined in Section 5.2.1.

5.11 Using the notation established in the previous exercise, show that

$$\mathrm{cov}\left(h^r(X), h^s(X)\right) = \kappa^{r,s}$$
$$\mathrm{cov}\left(h^r(X), h^{st}(X)\right) = \kappa^{r,s,t} + \eta^s\kappa^{r,t}[2]$$
$$\mathrm{cov}\left(h^{rs}(X), h^{tu}(X)\right) = \kappa^{r,s,t,u} + \eta^r\kappa^{s,t,u}[4] + \kappa^{r,t}\kappa^{s,u}[2] + \eta^r\eta^t\kappa^{s,u}[4]$$
$$\mathrm{cov}\left(h^r(X), h^{stu}(X)\right) = \kappa^{r,s,t,u} + \eta^s\kappa^{r,t,u}[3] + \kappa^{r,s}\eta^{t,u}[3] + \eta^s\eta^t\kappa^{r,u}[3].$$

5.12 Generalize the result of the previous exercise by showing that the joint cumulant corresponding to an arbitrary set of Hermite tensors involves a sum over connecting partitions. Describe the rule that determines the contribution of each connecting partition.

5.13 Show that

$$\int h_1(x)h_2(x)h_3(x)\phi(x)\,dx = 6$$

$$\int h_1(x)h_2(x)h_3(x)h_4(x)\phi(x)\,dx = 264$$

where $h_r(x)$ is the standard univariate Hermite polynomial of degree r and $\phi(x)$ is the standard normal density. [Hint: use the tables of connecting partitions.]

5.14 More generally, using the notation of the previous exercise, show that, for $i > j > k$,

$$\int h_i(x)h_j(x)h_k(x)\phi(x)\,dx = \frac{i!\,j!\,k!}{\{\frac{1}{2}(j+k-i)\}!\,\{\frac{1}{2}(i+k-j)\}!\,\{\frac{1}{2}(i+j-k)\}!}$$

when $j + k - i$ is even and non-negative, and zero otherwise, (Jarrett, 1973, p. 26).

5.15 Using (5.26) or otherwise, show that in the univariate case, where X is a standardized sum with mean zero, unit variance and so on, then

$$Y^* = X - \rho_3 X^2/6 - \rho_4 X^3/24 + \rho_3^2 X^3/9$$

has mean $-\rho_3/6$ and standard deviation

$$1 - \rho_4/8 + 7\rho_3^2/36$$

when terms of order $O(n^{-3/2})$ are ignored. Show also that

$$\frac{Y^* + \rho_3/6}{1 - \rho_4/8 + 7\rho_3^2/36} \sim N(0,1) + O(n^{-3/2})$$

5.16 Taking the definition of Y^* as given in the previous exercise, show that

$$W/2 = (Y^*)^2 = X^2/2 - \rho_3 X^3/3! - \{\rho_4 - 3\rho_3^2\}X^4/4!$$

has mean given by

$$E(W) = 1 + (5\rho_3^2 - 3\rho_4)/12 = 1 + b/n.$$

Deduce that

$$\frac{W}{1 + b/n} \sim \chi_1^2 + O(n^{-3/2}).$$

5.17 Using the equation following (5.26), show by reversal of series, that X may be expressed as the following polynomial in the normally distributed random variable Y

$$X = Y + \rho_3(Y^2 - 1)/6 + \rho_4(Y^3 - 3Y)/24 - \rho_3^2(2Y^3 - 5Y)/36.$$

Hence, express the approximate percentage points of X in terms of standard normal percentage points (Cornish & Fisher, 1937).

5.18 Let $X = Y + Z$ where Y has density function $f_0(y)$ and Z is independent of Y with moments η^i, η^{ij}, $\eta^{ijk}, \ldots$. Show formally, that the density of X is given by

$$f_X(x) = E_Z\{f_0(x - Z)\}.$$

Hence derive the series (5.2) by Taylor expansion of $f_0(x)$. By taking $\eta^i = 0$, $\eta^{i,j} = 0$, and $f_0(x) = \phi(x;\kappa)$, derive the usual Edgeworth expansion for the density of X, taking care to group terms in the appropriate manner. (Davis, 1976).