CHAPTER 3

Generalized cumulants

3.1 Introduction and definitions

In Chapter 2 we examined in some detail how cumulant tensors transform under affine transformation of X. Cumulants of order two and higher transform like Cartesian tensors, but the first-order cumulant does not. In this chapter, we show how cumulant tensors transform under non-linear or non-affine transformation of X. The algorithm that we describe relies heavily on the use of index notation and is easy to implement with the assistance of suitable tables. Applications are numerous. For example, the maximum likelihood estimator and the maximized likelihood ratio statistic can be expressed as functions of the log likelihood derivatives at the true but unknown parameter point, θ. The distribution of these derivatives at the true θ is known as a function of θ. With the methods developed here, we may compute moments or cumulants of any derived statistic, typically as an asymptotic approximation, to any required order of approximation.

In general, it is a good deal more convenient to work with polynomial functions rather than, say, exponential or logarithmic functions of X. The first step in most calculations is therefore to expand the function of interest as a polynomial in X and to truncate at an appropriate point. The essential ingredient when working with polynomial functions is to develop a notation capable of coping with generalized cumulants of the type

$$\kappa^{i,jk} = \mathrm{cov}(X^i, X^j X^k).$$

It seems obvious and entirely natural to denote this quantity by $\kappa^{i,jk}$, thereby indexing the set of generalized cumulants by partitions of the indices. In other words, to each partition there corresponds a unique cumulant and to each cumulant there corresponds a unique partition. For example,

$$\begin{aligned} \kappa^{i,jkl} &= \mathrm{cov}(X^i, X^j X^k X^l) \\ \kappa^{ij,kl} &= \mathrm{cov}(X^i X^j, X^k X^l) \\ \kappa^{i,j,kl} &= \mathrm{cum}(X^i, X^j, X^k X^l). \end{aligned} \tag{3.1}$$

Thus $\kappa^{i,j,kl}$, the third-order cumulant of the three variables X^i, X^j and the product $X^k X^l$, is said to be of order $b = 3$ and degree $p = 4$. The order is the number of blocks of the partition and the degree is the number of indices. *Ordinary* cumulants with $b = p$ and *ordinary* moments with $b = 1$ are special cases of generalized cumulants. The order of the blocks and of the indices within blocks in (3.1) is immaterial provided only that the partition is preserved. In this way, the notion of symmetry under index permutation is carried over to generalized cumulants.

3.2 The fundamental identity for generalized cumulants

Just as moments can be expressed as combinations of ordinary cumulants according to (2.6), so too generalized cumulants can be expressed in a similar way. First, we give the expressions for the four generalized cumulants listed above and then the general formula is described. The following four formulae may be derived from first principles using (2.6) and (2.7).

$$\begin{aligned}
\kappa^{i,jk} &= \kappa^{ijk} - \kappa^{i}\kappa^{jk} \\
&= \kappa^{i,j,k} + \kappa^{j}\kappa^{i,k} + \kappa^{k}\kappa^{i,j} \\
\kappa^{i,jkl} &= \kappa^{ijkl} - \kappa^{i}\kappa^{jkl} \\
&= \kappa^{i,j,k,l} + \kappa^{j}\kappa^{i,k,l}[3] + \kappa^{i,j}\kappa^{k,l}[3] + \kappa^{i,j}\kappa^{k}\kappa^{l}[3] \\
\kappa^{ij,kl} &= \kappa^{ijkl} - \kappa^{ij}\kappa^{kl} \\
&= \kappa^{i,j,k,l} + \kappa^{i}\kappa^{j,k,l}[2] + \kappa^{k}\kappa^{i,j,l}[2] + \kappa^{i,k}\kappa^{j,l}[2] \\
&\qquad + \kappa^{i}\kappa^{k}\kappa^{j,l}[4] \\
\kappa^{i,j,kl} &= \kappa^{ijkl} - \kappa^{i}\kappa^{jkl} - \kappa^{j}\kappa^{ikl} - \kappa^{ij}\kappa^{kl} + 2\kappa^{i}\kappa^{j}\kappa^{kl} \\
&= \kappa^{i,j,k,l} + \kappa^{k}\kappa^{i,j,l}[2] + \kappa^{i,k}\kappa^{j,l}[2].
\end{aligned} \tag{3.2}$$

Again, the bracket notation has been employed, but now the interpretation of groups of terms depends on the context. Thus, for example, in the final expression above, $\kappa^{i,k}\kappa^{j,l}[2] = \kappa^{i,k}\kappa^{j,l} + \kappa^{i,l}\kappa^{j,k}$ must be interpreted in the context of the partition on the left, namely $i|j|kl$. The omitted partition of the same type is $ij|kl$, corresponding to the cumulant product $\kappa^{i,j}\kappa^{k,l}$. Occasionally, this shorthand notation may lead to ambiguity and if so, it becomes necessary to list the individual partitions explicitly. However, the reader quickly becomes overwhelmed by the sheer number of terms that complete lists involve. For this reason we make every effort to avoid explicit complete lists.

An alternative notation, useful in order to avoid the kinds of ambiguity alluded to above, is to write $\kappa^{i,k}\kappa^{j,l}[2]_{ij}$ for the final term in (3.2). However, this notation conflicts with the summation convention and, less seriously, $\kappa^{i,k}\kappa^{j,l}[2]_{ij}$ is the same as $\kappa^{i,k}\kappa^{j,l}[2]_{kl}$. For these reasons, the unadorned bracket notation will be employed where there is no risk of ambiguity.

From the above examples, it is possible to discern, at least qualitatively, the rule that expresses generalized cumulants in terms of ordinary cumulants. An arbitrary cumulant of order b involving p random variables may be written as $\kappa(\Upsilon^*)$ where $\Upsilon^* = \{\upsilon_1^*|\ldots|\upsilon_b^*\}$ is a partition of p indices into b non-empty blocks. Rather conveniently, every partition that appears on the right in (3.2) has coefficient $+1$ and, in fact, the general expression may be written

$$\kappa(\Upsilon^*) = \sum_{\Upsilon \vee \Upsilon^* = 1} \kappa(\upsilon_1)\cdots\kappa(\upsilon_\nu), \tag{3.3}$$

where the sum is over all $\Upsilon = \{\upsilon_1|\ldots|\upsilon_\nu\}$ such that Υ and Υ^* are not both sub-partitions of any partition other than the full set, $\Upsilon_1 = \{(1, 2, \ldots, p)\}$ containing one block. Partitions satisfying this condition are said to be *complementary* to Υ^*. The notation and terminology used here are borrowed from lattice theory where $\Upsilon \vee \Upsilon^*$, also equal to $\Upsilon^* \vee \Upsilon$, is the *least upper bound* of Υ and Υ^*. A proof of this result is given in Section 3.6, using properties of the lattice of set partitions. In practice, the following graph-theoretical description of the condition $\Upsilon \vee \Upsilon^* = 1$ seems preferable because it is easier to visualize.

Any partition of p items or indices, say $\Upsilon^* = \{\upsilon_1^*|\ldots|\upsilon_b^*\}$, can be represented as a graph on p vertices. The edges of the graph consist of all pairs (i, j) that are in the same block of Υ^*. Thus the graph of Υ^* is the union of b disconnected complete graphs and we use the notation Υ^* interchangeably for the graph and for the partition. Since Υ and Υ^* are two graphs sharing the same vertices, we may define the edge sum graph $\Upsilon \oplus \Upsilon^*$, whose edges are the union of the edges of Υ and Υ^*. The condition that $\Upsilon \oplus \Upsilon^*$ be connected is identical to the condition $\Upsilon \vee \Upsilon^* = 1$

in (3.3). For this reason, we use the terms *connecting partition* and *complementary partition* interchangeably. In fact, this graph-theoretical device provides a simple way of determining the least upper bound of two or more partitions: the blocks of $\Upsilon \vee \Upsilon^*$ are just the connected components of the graph $\Upsilon \oplus \Upsilon^*$. The connections in this case need not be direct. In other words, the blocks of $\Upsilon \vee \Upsilon^*$ do not, in general, correspond to cliques of $\Upsilon \oplus \Upsilon^*$.

Consider, for example, the (4,6) cumulant $\kappa^{12,34,5,6}$ with $\Upsilon^* = 12|34|5|6$. Each block of the partition $\Upsilon = 13|24|56$ joins two blocks of Υ^*, but Υ and Υ^* are both sub-partitions of $1234|56$, so condition (3.3) is not satisfied. In the graph-theoretical representation, the first four vertices of $\Upsilon \oplus \Upsilon^*$ are connected and also the last two, but the vertices fall into two disconnected sets. By contrast, $\Upsilon = 1|23|456$ satisfies the required condition $\Upsilon \vee \Upsilon^* = 1$, giving rise to the contribution $\kappa^1\kappa^{2,3}\kappa^{4,5,6}$.

Expression (3.3) gives generalized cumulants in terms of ordinary cumulants. The analogous expression for generalized cumulants in terms of ordinary moments is

$$\kappa(\Upsilon^*) = \sum_{\Upsilon \geq \Upsilon^*} (-1)^{\nu-1}(\nu-1)!\, \mu(\upsilon_1)\cdots\mu(\upsilon_\nu) \tag{3.4}$$

where $\mu(\upsilon_j)$ is an ordinary moment, and ν is the number of blocks of Υ. The sum extends over all partitions Υ such that Υ^* is a sub-partition of Υ. This expression follows from the development in Section 2.4 and can be regarded as effectively equivalent to (2.9). See also Section 3.6.

It is not difficult to see that (2.8) and (2.9) are special cases of (3.3) and (3.4). If we take $\Upsilon^* = \Upsilon_1$, the unpartitioned set, then every partition Υ satisfies the condition required in (3.3), giving (2.8). On the other hand, if we take $\Upsilon^* = \Upsilon_p$, the fully partitioned set, then Υ^* is a sub-partition of every partition. Thus every partition contributes to (3.4) in this case, giving (2.9).

3.3 Cumulants of homogeneous polynomials

For definiteness, consider two homogeneous polynomials

$$P_2 = a_{ij}X^iX^j \quad \text{and} \quad P_3 = a_{ijk}X^iX^jX^k$$

of degree 2 and 3 respectively. In many ways, it is best to think of P_2 and P_3, not as quadratic and cubic forms, but as linear forms in pairs of variables and triples of variables respectively. From this point of view, we can see immediately that

$$\begin{aligned}
E(P_2) &= a_{ij}\kappa^{ij} = a_{ij}\{\kappa^{i,j} + \kappa^i\kappa^j\} \\
E(P_3) &= a_{ijk}\kappa^{ijk} = a_{ijk}\{\kappa^{i,j,k} + \kappa^i\kappa^{j,k}[3] + \kappa^i\kappa^j\kappa^k\} \\
\operatorname{var}(P_2) &= a_{ij}a_{kl}\kappa^{ij,kl} \\
&= a_{ij}a_{kl}\{\kappa^{i,j,k,l} + \kappa^i\kappa^{j,k,l}[2] + \kappa^k\kappa^{i,j,l}[2] \\
&\qquad + \kappa^{i,k}\kappa^{j,l}[2] + \kappa^i\kappa^k\kappa^{j,l}[4]\} \\
&= a_{ij}a_{kl}\{\kappa^{i,j,k,l} + 4\kappa^i\kappa^{j,k,l} + 2\kappa^{i,k}\kappa^{j,l} + 4\kappa^i\kappa^k\kappa^{j,l}\}
\end{aligned}$$

$$\begin{aligned}
\operatorname{cov}(P_2, P_3) &= a_{ij}a_{klm}\kappa^{ij,klm} \\
&= a_{ij}a_{klm}\{\kappa^{i,j,k,l,m} + \kappa^i\kappa^{j,k,l,m}[2] + \kappa^k\kappa^{i,j,l,m}[3] \\
&\quad + \kappa^{i,k}\kappa^{j,l,m}[6] + \kappa^{k,l}\kappa^{i,j,m}[3] + \kappa^i\kappa^k\kappa^{j,l,m}[6] + \kappa^k\kappa^l\kappa^{i,j,m}[3] \\
&\quad + \kappa^{i,k}\kappa^{j,l}\kappa^m[6] + \kappa^{i,k}\kappa^{l,m}\kappa^j[6] + \kappa^{i,k}\kappa^j\kappa^l\kappa^m[6]\}
\end{aligned}$$

For the final expression above where $\Upsilon^* = ij|klm$, the list of 42 complementary partitions can be found in Table 1 of the Appendix, where Υ^* is coded numerically as $123|45$. Since the arrays of coefficients a_{ij} and a_{ijk} are symmetrical, the permutation factors in [.] can be changed into ordinary arithmetic factors. Thus the 42 complementary partitions contribute only 10 distinct terms. These cannot be condensed further except in special cases. In the expression for $\mathrm{var}(P_2)$, on the other hand, the two classes of partitions $i|jkl[2]$ and $k|ijl[2]$ make equal contributions and further condensation is then possible as shown above.

In specific applications, it is often the case that the arrays a_{ij} and a_{ijk} have some special structure that can be exploited in order to simplify expressions such as those listed above. Alternatively, it may be that the cumulants have simple structure characteristic of independence, exchangeability, or identical distributions. In such cases, the joint cumulants listed above can be condensed further using power notation.

By way of illustration, we suppose that a_{ij} is a residual projection matrix of rank r, most commonly written as $\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ in the notation of linear models where $\mathbf{X}$ is a model matrix of known constants. We suppose in addition that the random variables $Y^i - \kappa^i$ are independent and identically distributed so that $\kappa^{i,j} = \kappa_2\delta^{ij}$, and the fourth-order cumulants are $\kappa_4\delta^{ijkl}$. On the assumption that $E(Y)$ does indeed lie in the column space of $\mathbf{X}$, it follows that $a_{ij}\kappa^j = 0$, and hence

$$\begin{aligned} E(P_2) &= a_{ij}\kappa^{i,j} = \kappa_2 \sum a_{ii} = r\kappa_2 \\ \mathrm{var}(P_2) &= a_{ij}a_{kl}\{\kappa^{i,j,k,l} + 2\kappa^{i,k}\kappa^{j,l}\} \\ &= \kappa_4 \sum a_{ii}^2 + 2r\kappa_2^2 \end{aligned}$$

In this example, P_2 is just the residual sum of squares on r degrees of freedom after linear regression on $\mathbf{X}$, where the theoretical errors are independent and identically distributed but not necessarily normal.

3.4 Polynomial transformations

It is not difficult now to develop the transformation law for cumulants under arbitrary non-linear transformation to new variables Y. The formulae developed in Section 3.2 refer to the particular polynomial transformation

$$Y^1 = \prod_{j\in\upsilon_1^*} X^j, \quad Y^2 = \prod_{j\in\upsilon_2^*} X^j, \quad \ldots, \quad Y^b = \prod_{j\in\upsilon_b^*} X^j.$$

To the extent that any continuous function can be approximated with arbitrary accuracy by means of a polynomial, there is little loss of generality in considering polynomial transformations

$$Y^r = a^r + a_i^r X^i + a_{ij}^r X^i X^j + a_{ijk}^r X^i X^j X^k + \cdots . \tag{3.5}$$

It is necessary at this stage to insist that the infinite expansion (3.5) be convergent, in principle for all X. In practice, in asymptotic calculations, the expansion must be convergent for all X for which the probability is appreciable.

To state the transformation law of cumulants in a concise way, it is helpful to abbreviate (3.5) by using 'matrix' notation as follows:

$$Y^r = (A_0^r + A_1^r + A_2^r + A_3^r + \cdots)X \tag{3.6}$$

where, for example, $A_2^r X$ is understood to represent a vector whose components are quadratic or bilinear forms in X. We abbreviate (3.6) further by introducing the operators $P^r = A_0^r + A_1^r + A_2^r + \cdots$ and writing

$$Y^r = P^r X. \tag{3.7}$$

The cumulant generating function of $Y = Y^1, \ldots, Y^q$ may now be written purely formally as

$$K_Y(\xi) = \exp(\xi_r P^r)\kappa_X \tag{3.8}$$

where $\exp(\xi_r P^r)$ is an operator acting on the cumulants of X as follows.

$$\begin{aligned} K_Y(\xi) = \{1 &+ \xi_r P^r + \xi_r\xi_s P^r P^s/2! \\ &+ \xi_r\xi_s\xi_t P^r P^s P^t/3! + \cdots\}\kappa_X. \end{aligned} \tag{3.9}$$

We define $1\kappa_X = 0$ and

$$P^r\kappa_X = a^r + a^r_i\kappa^i + a^r_{ij}\kappa^{ij} + a^r_{ijk}\kappa^{ijk} + \cdots .$$

Compound operators P^rP^s and $P^rP^sP^t$ acting on κ_X produce generalized cumulants of order $b = 2$ and $b = 3$ respectively. For example, $A^r_1A^s_2\kappa_X = a^r_i a^s_{jk}\kappa^{i,jk}$, which is not to be confused with $A^r_2A^s_1\kappa_X = a^r_{ij}a^s_k\kappa^{ij,k}$. In like manner, third-order compound operators produce terms such as

$$\begin{aligned} A^r_1A^s_1A^t_1\kappa_X &= a^r_i a^s_j a^t_k\kappa^{i,j,k} \\ A^r_1A^s_2A^t_1\kappa_X &= a^r_i a^s_{jk} a^t_l\kappa^{i,jk,l}. \end{aligned}$$

Compound operators involving A_0 produce terms such as

$$\begin{aligned} A^r_0A^s_1\kappa_X &= a^r a^s_i\kappa^{,i} = 0 \\ A^r_1A^s_0A^t_2\kappa_X &= a^r_i a^s a^t_{kl}\kappa^{i,,kl} = 0 \end{aligned}$$

and these are zero because they are mixed cumulants involving, in effect, one variable that is a constant.

Expansion (3.9) for the first four cumulants gives

$$\begin{aligned} K_Y(\xi) &= \xi_r\{a^r + a^r_i\kappa^i + a^r_{ij}\kappa^{ij} + a^r_{ijk}\kappa^{ijk} + \cdots\} \\ &+ \xi_r\xi_s\{a^r_i a^s_j\kappa^{i,j} + a^r_i a^s_{jk}\kappa^{i,jk}[2] + a^r_{ij}a^s_{kl}\kappa^{ij,kl} + \cdots\}/2! \\ &+ \xi_r\xi_s\xi_t\{a^r_i a^s_j a^t_k\kappa^{i,j,k} + a^r_i a^s_j a^t_{kl}\kappa^{i,j,kl}[3] + a^r_i a^s_{jk} a^t_{lm}\kappa^{i,jk,lm}[3] + \cdots\}/3! \\ &+ \xi_r\xi_s\xi_t\xi_u\{a^r_i a^s_j a^t_k a^u_l\kappa^{i,j,k,l} + a^r_i a^s_j a^t_k a^u_{lm}\kappa^{i,j,k,lm}[4] + \cdots\}/4! \\ &+ \cdots \end{aligned} \tag{3.10}$$

The leading terms in the above expansions are the same as (2.11), the law governing affine transformation.

The proof of (3.8) is entirely elementary and follows from the definition of generalized cumulants, together with the results of Section 2.4. A similar, though less useful formal expression can be developed for the moment generating function $M_X(\xi)$, which may be written

$$M_Y(\xi) = \exp(\xi_r P^r) * \kappa_X \tag{3.11}$$

where $1 * \kappa_X = 1$, $P^r * \kappa_X = P^r\kappa_X$ as before, and commas are omitted in the application of compound operators giving

$$\begin{aligned} A^r_0A^s_1 * \kappa_X &= a^r a^s_i\kappa^i \\ A^r_1A^s_1 * \kappa_X &= a^r_i a^s_j\kappa^{ij} \\ A^r_1A^s_0A^t_2 * \kappa_X &= a^r_i a^s a^t_{jk}\kappa^{ijk} \end{aligned}$$

and so on. Once again, the proof follows directly from the definition of moments.

In practice, in order to make use of (3.10), it is usually necessary to re-express all generalized cumulants in terms of ordinary cumulants. This exercise involves numerous applications of (3.3) and the formulae become considerably longer as a result.

3.5 Classifying complementary partitions

In order to use the fundamental identity (3.3) for the generalized cumulant $\kappa(\Upsilon^*)$ we need to list all partitions complementary to Υ^*. If the number of elements of Υ^* is more than, say six to eight, the number of complementary partitions can be very large indeed. If the listing is done by hand, it is difficult to be sure that no complementary partitions have been omitted. On the other hand, programming a computer to produce the required list is not an easy task. Further, it is helpful to group the partitions into a small number of equivalence classes in such a way that all members of a given class make the same contribution to the terms that occur in (3.10).

Suppose, for example, that we require the covariance of the two scalars

$$a_{ij}X^iX^j \quad \text{and} \quad b_{ij}X^iX^j$$

both of which are quadratic in X. We find

$$\text{cov}\{a_{ij}X^iX^j,\ b_{ij}X^iX^j\} = a_{ij}b_{kl}\kappa^{ij,kl}.$$

To simplify the following expressions, we take $\kappa^i = 0$, giving

$$a_{ij}b_{kl}\{\kappa^{i,j,k,l} + \kappa^{i,k}\kappa^{j,l} + \kappa^{i,l}\kappa^{j,k}\}.$$

Since, by assumption, $a_{ij} = a_{ji}$ and $b_{ij} = b_{ji}$, this expression reduces to

$$a_{ij}b_{kl}\{\kappa^{i,j,k,l} + 2\kappa^{i,k}\kappa^{j,l}\}.$$

Thus the two partitions $ik|jl$ and $il|jk$ make identical contributions to the covariance and therefore they belong to the same equivalence class. This classification explains the grouping of terms in (3.2).

The classification of complementary partitions is best described in terms of the intersection matrix $M = \Upsilon^* \cap \Upsilon$, where m_{ij} is the number of elements in $\upsilon_i^* \cap \upsilon_j$. Since the order of the blocks is immaterial, this matrix is defined only up to independent row and column permutations. Two partitions, Υ_1 and Υ_2, whose intersection matrices are M_1 and M_2, are regarded as equivalent if $M_1 = M_2$ after suitably permuting the blocks of Υ_1 and Υ_2 or the columns of M_1 and M_2. It is essential in this comparison that the ith rows of M_1 and M_2 refer to the same block of Υ^*.

To take a simple example, suppose that $\Upsilon^* = 12|34|5$, $\Upsilon_1 = 135|24$, $\Upsilon_2 = 123|45$ and $\Upsilon_3 = 134|25$ with intersection matrices

$$M_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}; \qquad M_2 = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}; \qquad M_3 = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

These matrices are all distinct. However, the partitions 145|23, 235|14 and 245|13 have M_1 as intersection matrix, the partition 124|35 has M_2 as intersection matrix and 234|15 has M_3 as intersection matrix. Thus these eight complementary partitions are written as

$$135|24[4] \ \cup\ 123|45[2] \ \cup\ 134|25[2].$$

If Υ_1 and Υ_2 are equivalent partitions in the sense just described, they must have the same number of blocks and also identical block sizes. Further, when permuting columns, we need only consider blocks of equal size: by convention, the blocks are usually arranged in decreasing size.

Tables 1 and 2 in the Appendix give lists of complementary partitions classified according to the above scheme. A typical element of each equivalence class is given and the number of elements in that class follows in [.].

3.6 Elementary lattice theory

3.6.1 *Generalities*

A lattice is a finite partially ordered set $\mathcal{L}$ having the additional property that for every pair of elements $a, b \in \mathcal{L}$ there is defined a unique greatest lower bound $c = a \wedge b$ and a unique least upper bound $d = a \vee b$ where $c, d \in \mathcal{L}$. These additional properties should not be taken lightly and, in fact, some commonly occurring partially ordered sets are not lattices because not every pair has a unique least upper bound or greatest lower bound. One such example is described in Exercise 3.34.

Lattices of various types arise rather frequently in statistics and probability theory. For example, in statistics, the class of factorial models, which can be described using the operators $+$ and $*$ on factors A, B, $C, \ldots$, forms a lattice known as the *free distributive lattice*. Typical elements of this lattice are $a = A + B * C$ and $b = A{*}B + B{*}C + C{*}A$, each corresponding to a factorial model. In the literature on discrete data, these models are also called *hierarchical*, but this usage conflicts with standard terminology in the analysis of variance where hierarchical refers to the presence of several variance components. In this particular example, $a < b$ because a is a sub-model of b, and the partial order has a useful statistical interpretation.

In probability theory or in set theory where $A_1, A_2, \ldots$ are subsets of Ω, it is sometimes useful to consider the lattice with elements Ω, A_i, $A_i \cap A_j$, $A_i \cap A_j \cap A_k$ and so on. We say that $a < b$ if $a \subset b$; the lattice so formed is known as the *binary lattice*. The celebrated inclusion-exclusion principle for calculating $\mathrm{pr}(A_1 \cup A_2 \cup \cdots)$ is a particular instance of (3.12) below (Rota, 1964). Exercise 2.22 provides a third example relevant both to statistics and to probability theory.

In this book, however, we are concerned principally with the lattice of set partitions where $a < b$ if a is a sub-partition of b. The least upper bound, $a \vee b$ was described in Section 3.2 as the partition whose blocks are the connected vertices of the graph $a \oplus b$: the greatest lower bound $a \wedge b$ is the partition whose blocks are the non-empty intersections of the blocks of a and b. Figure 3.2 gives the Hasse diagrams of the partition lattices of sets of up to four items. Each partition in the ith row of one of these diagrams contains i blocks. Partitions in the same row are unrelated in the partial order: partitions in different rows may or may not be related in the partial order. Notice that $123|4$ has three immediate descendants whereas $13|24$ has only two.

Let $\mathcal{L}$ be an arbitrary lattice, and let $f(\cdot)$ be a real-valued function defined on $\mathcal{L}$. We define the new function, $F(.)$ on $\mathcal{L}$ by

$$F(a) = \sum_{b \leq a} f(b), \tag{3.11a}$$

analogous to integrating over the interval $[0, a]$. The formal inverse operation analogous to differentiation may be written

$$f(b) = \sum_{c \leq b} m(c, b) F(c) \tag{3.12}$$

where $m(a, b)$ is known as the Möbius function for the lattice. See Exercise 2.?? for a justification of this analogy with differential and integral calculus.

It may be helpful by way of comparison to consider the matrix equation $F = Lf$, where f and F are vectors, and L_{ab} is the indicator function for $b \leq a$. The inverse relation is $f = MF$ where the Möbius matrix $M = L^{-1}$ is also a lower triangular matrix. Note also that M^T is the matrix inverse of L^T. Consequently, if we define $\bar{F}(a) = \sum_{b \geq a} f(b)$, then the inverse relation involves the transposed Möbius matrix $f(b) = \sum_{c \geq b} m(b, c) \bar{F}(c)$.

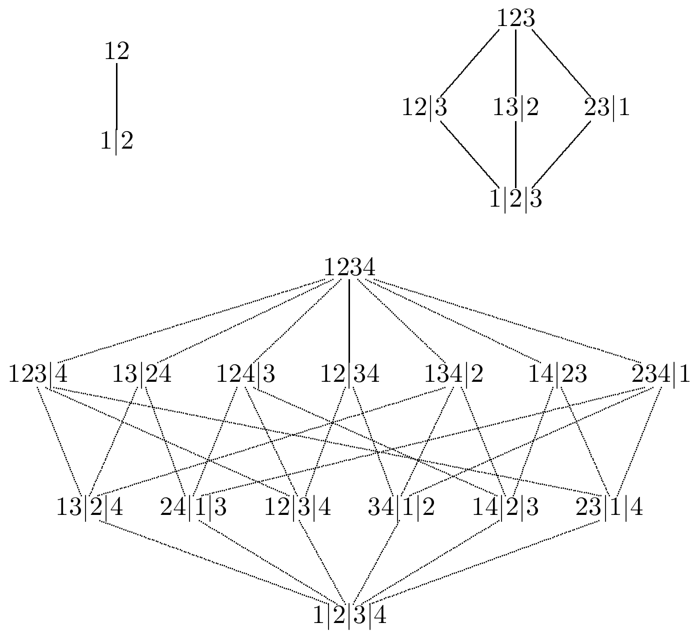

Figure 3.1: *Hasse diagrams for smaller partition lattices.*

Substitution of (3.12) into (3.11a) or (3.11a) into (3.12) gives

$$\begin{aligned} F(a) &= \sum_{b\le a}\sum_{c\le b} m(c,b)F(c) \\ &= \sum_{c}\sum_{c\le b\le a} m(c,b)F(c) \\ &= \sum_{c}\delta_{ac}F(c). \end{aligned} \qquad \begin{aligned} f(b) &= \sum_{a\le b}\sum_{c\le a} m(a,b)f(c) \\ &= \sum_{c}\sum_{c\le a\le b} m(a,b)f(c) \\ &= \sum_{c}\delta_{bc}f(c). \end{aligned}$$

The relation is required to hold for all functions F or f, which justifies the final line above. It follows, for fixed $a \le c$, that the sums over the lattice interval $[a, c]$,

$$\sum_{a\le b\le c} m(a,b) = \sum_{a\le b\le c} m(b,c) = \delta(a,c)$$

are zero unless $a = c$. In this sense, the Möbius function can be thought of as a series of contrasts or difference operators on all lattice intervals. The Möbius function may therefore be obtained recursively by $m(a, a) = 1$ followed by either

$$m(a,c) = -\sum_{a\le b<c} m(a,b), \quad \text{or} \quad m(a,c) = -\sum_{a<b\le c} m(b,c),$$

for $a < c$, and $m(a, c) = 0$ otherwise.

3.6.2 *Möbius function for the partition lattice*

We now show by induction that the Möbius function satisfies

$$m(a,1) = (-1)^{|a|-1}(|a|-1)! \qquad (3.13)$$

where $|a|$ is the number of blocks of the partition. We use the property established above, that

$$m(a, 1) = -\sum_{a<b\leq 1} m(b, 1).$$

The Stirling number of the second kind, S_n^m, is the number of ways of partitioning a set of n elements into m non-empty subsets. A new element can be placed into one of the existing blocks or into a new block. Consequently, S_n^m satisfies the recurrence relation

$$S_{n+1}^m = mS_n^m + S_n^{m-1}.$$

Let a be a partition having $n+1$ blocks, and suppose that (3.13) is satisfied for all partitions having up to n blocks. Application of the formula for the Möbius function gives

$$-m(a, 1) = \sum_{m=1}^{n} S_{n+1}^m (-1)^{m-1}(m-1)!.$$

The recurrence relation for Stirling numbers gives

$$\begin{aligned} -m(a, 1) &= \sum_{m=1}^{n} S_n^m (-1)^{m-1} m! + \sum_{m=1}^{n} S_n^{m-1} (-1)^{m-1}(m-1)! \\ &= S_n^n (-1)^{n-1} n!. \end{aligned}$$

Since $S_n^n = 1$, the result (3.13) follows.

More generally, $m(\Upsilon, \Upsilon^*)$ is the product of terms like (3.13). Suppose that $\Upsilon < \Upsilon^*$ and that $\Upsilon^* = \{\upsilon_1^*|\ldots|\upsilon_\nu^*\}$ is a partition into ν blocks. In the partition Υ, each block of Υ^* is partitioned into smaller subsets, υ_j^* being partitioned into b_j blocks in the finer partition. Then

$$m(\Upsilon, \Upsilon^*) = \prod_j (-1)^{b_j - 1}\,(b_j - 1)!.$$

For example, $m(1|2|3|4, 12|34)$ is equal to the product of $m(1|2, 12)$ and m(3—4, 34), $whichis (-1)^2$ while $m(1|2|34|56, 1234|56)$ is equal to $m(1|2|34, 1234) m(56, 56)$. The pairs 34 and 56 may be regarded as single indexes, so the Möbius function is equal to 2×1.

3.6.3 *Inclusion-exclusion and the binary lattice*

Let $A_1, \ldots, A_k$ be subsets of the sample space Ω. The binary lattice, $\mathcal{B}$, on these k generators, is the set whose 2^k elements are Ω, the sets A_i, the pairwise intersections $A_i \cap A_j$ with $(i < j)$, the triple intersections, and so on up to $A_1 \cap \cdots \cap A_k$. The partial order is defined by set inclusion, $a < b$ if a is a proper subset of b. The Hasse diagram of this lattice is illustrated for $k = 3$ in the first diagram in Fig. 3.2.

It is convenient to define, in association with $\mathcal{B}$, a new isomorphic lattice $\mathcal{B}'$ whose elements are the 2^k non-overlapping sets

$$\{A_1, \bar{A}_1\} \cap \{A_2, \bar{A}_2\} \cap \cdots \cap \{A_k, \bar{A}_k\}.$$

Elements in $\mathcal{B}'$ are associated with elements of $\mathcal{B}$ by deleting all the $\bar{A}$s, or equivalently by replacing all the $\bar{A}$s by Ω. The new lattice inherits its partial order from $\mathcal{B}$, so it is isomorphic with $\mathcal{B}$. Figure 3.1 illustrates the two lattices for $k = 3$.

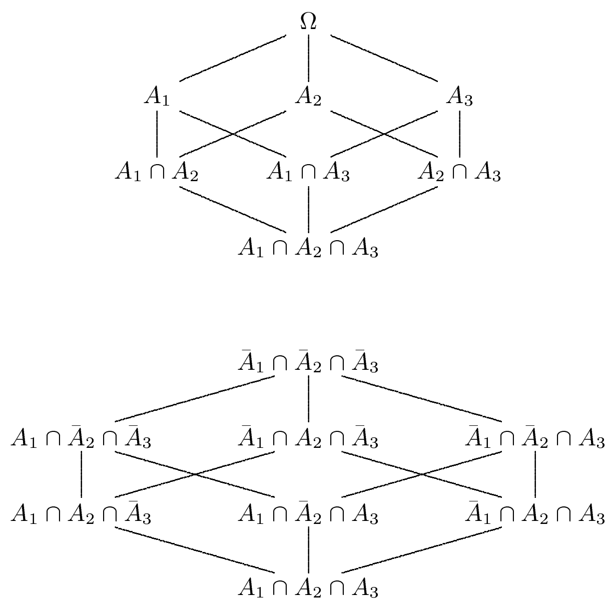

Figure 3.2: *Hasse diagram for the binary lattice on three generators (top), with the associated lattice of non-overlapping sets (bottom).*

The important relation between $\mathcal{B}$ and $\mathcal{B}'$ is the following. Let b and b' be corresponding elements in the two lattices. Then

$$b = \bigcup_{a' \le b'} a',$$

primes being used to denote elements of $\mathcal{B}'$. In other words, b is the union of b' and its descendants in the associated lattice. Since the elements of $\mathcal{B}'$ are disjoint, the probability of the union is the sum of the probabilities,

$$\mathrm{pr}(b) = \sum_{a' \le b'} \mathrm{pr}(a'),$$

analogous to integration over the lattice. The inverse relation is

$$\mathrm{pr}(b') = \sum_{a \le b} m(a, b)\, \mathrm{pr}(b).$$

As is shown in Exercise 3.??, the Möbius function for the binary lattice satisfies $m(a, \Omega) = (-1)^{|a|}$, where $|a|$ is the number of sets intersecting in a. In particular, taking $b' = \bar{A}_1 \cap \cdots \cap \bar{A}_k$ and $b = \Omega$, we obtain the celebrated inclusion-exclusion rule

$$\begin{aligned}
\mathrm{pr}(A_1 \cup \cdots \cup A_k) &= 1 - \mathrm{pr}(\bar{A}_1 \cap \cdots \cap \bar{A}_k) \\
&= 1 - \sum m(b, \Omega)\ \mathrm{pr}(b) \\
&= 1 - \mathrm{pr}(\Omega) + \sum \mathrm{pr}(A_j) - \sum \mathrm{pr}(A_i \cap A_j) + \cdots \\
&= \sum \mathrm{pr}(A_j) - \sum \mathrm{pr}(A_i \cap A_j) + \cdots + (-1)^{k-1}\, \mathrm{pr}(A_1 \cap \cdots \cap A_k).
\end{aligned}$$

3.6.4 *Cumulants and the partition lattice*

Let $\mathcal{L}$ be the partition lattice on p labels or indices. Each element of $\mathcal{L}$ is a partition $\Upsilon = \{\upsilon_1|\ldots|\upsilon_b\}$ into a number of non-empty sets called blocks. We now define the following functions on $\mathcal{L}$.

$$\begin{aligned} f(\Upsilon) &= \kappa(\upsilon_1)\cdots\kappa(\upsilon_b) && \text{cumulant product} \\ F(\Upsilon) &= \mu(\upsilon_1)\cdots\mu(\upsilon_b) && \text{moment product} \\ g(\Upsilon) &= \kappa(\Upsilon) && \text{generalized cumulant} \end{aligned}$$

It is helpful notationally to denote by 0 and 1 the minimal and maximal elements of $\mathcal{L}$. In particular, the element 0 has p blocks of size one each; the element 1 has a single block of size p.

We begin with the established result that any moment is expressible as a sum of cumulant products, summed over the relevant partition lattice. In the current notation,

$$g(1) = F(1) = \sum_{0\le a\le 1} f(a) = \sum_{0\le\Upsilon\le 1} \kappa(\upsilon_1)\cdots\kappa(\upsilon_b).$$

Let Υ^* be an arbitrary partition with blocks $\upsilon_1^*|\ldots|\upsilon_b^*$. Now the moment $\mu(\upsilon_1^*)$ is a sum of cumulant products, summed over the partitions of υ_1^*. On multiplying together b such expressions, one for each block of Υ^*, we obtain the following:

$$\begin{aligned} F(\Upsilon^*) &= \mu(\upsilon_1^*)\cdots\mu(\upsilon_b^*) \\ &= \sum_{\Upsilon\le\Upsilon^*} \kappa(\upsilon_1)\cdots\kappa(\upsilon_\nu) \\ &= \sum_{\Upsilon\le\Upsilon^*} f(\Upsilon) \end{aligned} \tag{3.14}$$

The lattice interval $[0,\Upsilon^*]$ is the direct product of complete sub-lattices $[0,\upsilon_1^*]\times\cdots\times[0,\upsilon_b^*]$, which explains the second line above. The relation between F and f justifies our choice of notation.

Application of the Möbius inversion formula gives

$$f(\Upsilon^*) = \sum_{\Upsilon\le\Upsilon^*} m(\Upsilon,\Upsilon^*)\, F(\Upsilon). \tag{3.15}$$

In particular, if $\Upsilon^* = 1$, we may use the fact that $m(\Upsilon,1) = (-1)^{b-1}(b-1)!$, to obtain the expression for ordinary cumulants in terms of moment products:

$$f(1) = \sum_{\Upsilon}(-1)^{b-1}(b-1)!\,\mu(\upsilon_1)\cdots\mu(\upsilon_b). \tag{3.16}$$

Now suppose that each index in the preceding expression is in fact a compound index representing a product of variables, so that the expanded partition becomes Υ^*. Then $f(1)$ becomes the generalized cumulant $\kappa(\Upsilon^*)$ given by

$$\begin{aligned} \kappa(\Upsilon^*) &= \sum_{\Upsilon\ge\Upsilon^*} m(\Upsilon,1)\,\mu(\upsilon_1)\cdots\mu(\upsilon_b) \\ &= \sum_{\Upsilon\ge\Upsilon^*} m(\Upsilon,1)\,F(\Upsilon) \end{aligned}$$

Substitution of expression (3.14) for $F(\Upsilon)$ gives

$$\kappa(\Upsilon^*) = \sum_{\Upsilon\ge\Upsilon^*} m(\Upsilon,1) \sum_{\Pi\le\Upsilon} f(\Pi).$$

Finally, reversal of the order of summation gives

$$\begin{aligned}\kappa(\Upsilon^*) = g(\Upsilon^*) &= \sum_{\Pi} \sum_{\Upsilon \geq \Upsilon^* \vee \Pi} m(\Upsilon, 1)\, f(\Pi) \\ &= \sum_{\Pi \vee \Upsilon^* = 1} \kappa(\pi_1) \cdots \kappa(\pi_\sigma) \qquad (3.17)\end{aligned}$$

where $\pi_1|\ldots|\pi_\sigma$ are the blocks of Π. This completes the proof of the fundamental identity (3.3).

In matrix notation, for the special case of four variables, we may write this equation as shown in Table 3.1.

Table 3.1 *Generalized cumulants expressed in terms of ordinary cumulants*

κ^{rstu}		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	$\kappa^{r,s,t,u}$
$\kappa^{rsu,t}$		1	1		1	1	1	1	1		1		1		1		$\kappa^{r,s,u}\kappa^{t}$
$\kappa^{rtu,s}$		1	1	1		1	1	1	1	1			1	1			$\kappa^{r,t,u}\kappa^{s}$
$\kappa^{stu,r}$		1	1	1	1		1	1	1	1	1	1					$\kappa^{s,t,u}\kappa^{r}$
$\kappa^{rs,tu}$		1	1	1	1	1		1	1		1	1	1	1			$\kappa^{r,s}\kappa^{t,u}$
$\kappa^{rt,su}$		1	1	1	1	1	1		1	1		1	1		1		$\kappa^{r,t}\kappa^{s,u}$
$\kappa^{ru,st}$	=	1	1	1	1	1	1	1		1	1			1	1		$\kappa^{r,u}\kappa^{s,t}$
$\kappa^{rs,t,u}$		1			1	1		1	1								$\kappa^{r,s}\kappa^{t}\kappa^{u}$
$\kappa^{rt,s,u}$		1		1		1	1		1								$\kappa^{r,t}\kappa^{s}\kappa^{u}$
$\kappa^{ru,s,t}$		1	1			1	1	1									$\kappa^{r,u}\kappa^{s}\kappa^{t}$
$\kappa^{st,r,u}$		1		1	1		1	1									$\kappa^{s,t}\kappa^{r}\kappa^{u}$
$\kappa^{su,r,t}$		1	1		1		1		1								$\kappa^{s,u}\kappa^{r}\kappa^{t}$
$\kappa^{tu,r,s}$		1	1	1				1	1								$\kappa^{t,u}\kappa^{r}\kappa^{s}$
$\kappa^{r,s,t,u}$		1															$\kappa^{r}\kappa^{s}\kappa^{t}\kappa^{u}$

3.6.5 *Further relationships among cumulants*

The three functions described at the beginning of the previous section are defined in general as follows:

$$F(b) = \sum_{a \leq b} f(a) \qquad\qquad f(a) = \sum_{b \leq a} m(b, a) F(b)$$

$$K(a) = \sum_{b \geq a} m(b, 1) F(b) \qquad\qquad K(a) = \sum_{a \vee b = 1} f(b)$$

Ignoring, for the moment, the statistical interpretation of these functions, the first three expressions are simply a matter of definition, and could be applied to any lattice. The final expression for K in terms of f is a consequence of these definitions, and also applies to any lattice.

In order to obtain an expression for f in terms of K, it is helpful to use matrix notation since the relationships are linear and invertible. In this notation, the various expressions may be written

$$F = Lf, \quad f = L^{-1}F = MF, \quad K = L^T W F = L^T W L f,$$

where $W = \operatorname{diag}\{m(a, 1)\}$. Matrix notation makes it obvious that the transformation from f to K involves a symmetric matrix. What is less obvious is that the elements of $L^T W L$ are all zero or one, and that $(L^T W L)_{ab} = I(a \vee b = 1)$. In any event, the inverse relation is $f = M W^{-1} M^T K = \Sigma K$, where where Σ is a symmetric matrix with elements

$$\sigma(b, c) = \sum_{a} \frac{m(a, b)\, m(a, c)}{m(a, 1)},$$

and summation may be restricted to $a \leq b \wedge c$.

Explicitly, if $p = 3$, we may write the five partitions in order as $a_1 = 123$, $a_2 = 1|23$, $a_3 = 2|13$, $a_4 = 3|12$ and $a_5 = 1|2|3$. The matrix giving generalized cumulants in terms of cumulant products in (3.3) is

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where a unit entry in position (i, j) corresponds to the criterion $a_i \vee a_j = 1$. The matrix inverse corresponding to (3.18) above is

$$\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & 2 \\ 0 & -1 & 1 & 1 & -1 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & 1 & -1 & -1 \\ 2 & -1 & -1 & -1 & 1 \end{pmatrix}$$

Application of (3.18), using the second row of the above matrix, gives

$$\kappa^i \kappa^{j,k} = \{\kappa^{j,ik} + \kappa^{k,ij} - \kappa^{i,jk} - \kappa^{i,j,k}\}/2$$

and this particular expression can be verified directly.

For $p = 4$, the matrix $\sigma(.,.)$ in (3.18) is shown in Table 3.2.

Table 3.2 *Matrix giving cumulant products in terms of generalized cumulants*

	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	0	−1	−1	−1	−1	1	1	1	−1	−1	2	−1	2	2	−2
	0	−1	−1	−1	−1	1	1	1	−1	2	−1	2	−1	2	−2
	0	−1	−1	−1	−1	1	1	1	2	−1	−1	2	2	−1	−2
	0	−1	−1	−1	−1	1	1	1	2	2	2	−1	−1	−1	−2
	0	1	1	1	1	−1	−1	−1	−2	1	1	1	1	−2	−1
1	0	1	1	1	1	−1	−1	−1	1	−2	1	1	−2	1	−1
–	0	1	1	1	1	−1	−1	−1	1	1	−2	−2	1	1	−1
6	0	−1	−1	2	2	−2	1	1	2	−1	−1	−1	−1	−1	1
	0	−1	2	−1	2	1	−2	1	−1	2	−1	−1	−1	−1	1
	0	2	−1	−1	2	1	1	−2	−1	−1	2	−1	−1	−1	1
	0	−1	2	2	−1	1	1	−2	−1	−1	−1	2	−1	−1	1
	0	2	−1	2	−1	1	−2	1	−1	−1	−1	−1	2	−1	1
	0	2	2	−1	−1	−2	1	1	−1	−1	−1	−1	−1	2	1
	6	−2	−2	−2	−2	−1	−1	−1	1	1	1	1	1	1	−1

The most important of these identities, or at least the ones that occur most frequently in the remainder of this book, are (3.15) and its inverse and the expression involving connecting partitions for generalized cumulants in terms of cumulant products. The inverse of the latter expression seems not to arise often. It is given here in order to emphasize that the relationship is invertible.

Finally, we note from (3.15) that any polynomial in the moments, homogeneous in the sense that every term is of degree 1 in each of p variables, can be expressed as a similarly homogeneous polynomial in the cumulants, and vice-versa for cumulants in terms of moments. In addition, from (3.16) we see that every generalized cumulant of degree p is uniquely expressible as a homogeneous polynomial in the moments. Inversion of (3.16) shows that every homogeneous polynomial in the moments is expressible as a *linear* function of generalized cumulants, each of degree p. Similarly, every homogeneous polynomial in the cumulants is expressible as a linear function of generalized cumulants. It follows that any polynomial in the cumulants or in the moments, homogeneous or not, is expressible as a linear function of generalized cumulants. Furthermore, this linear representation is unique because the generalized cumulants are linearly independent (Section 3.8). In fact, we have previously made extensive use of this important property of generalized cumulants. Expansion (3.10) for the cumulant generating function of the polynomial (3.5) is a linear function of generalized cumulants.

3.7 Some examples involving linear models

The notation used in this section is adapted to conform to the conventions of linear models, where $y = y^1, \ldots, y^n$ is the observed value of the random vector $Y = Y^1, \ldots, Y^n$ whose cumulants are κ^i, $\kappa^{i,j}$, $\kappa^{i,j,k}$ and so on. The common case where the observations are independent is especially important but, for the moment, the cumulant arrays are taken as arbitrary with no special structure. Now, Y and y lie in R^n, but the usual linear model specifies that the mean vector with components κ^i lies in the p-dimensional subspace, S_p of R^n spanned by the vectors $x_1, \ldots, x_p$ with components x^i_r. Thus we may write $E(Y) = x_r\beta^r$, or using components,

$$E(Y^i) = \kappa^i = x^i_r\beta^r$$

where β is a p-dimensional vector of parameters to be estimated.

It is important at this stage, to ensure that the notation distinguish between, on the one hand, vectors such as y^i, κ^i, $\kappa^{i,j}$ in R^n and its tensor products, and on the other hand, vectors such as β^r, $\beta^{r,s}$ in S_p and its tensor products. To do so, we use the letters $i, j, k, \ldots$ to indicate components of vectors in R^n, and $r, s, t, \ldots$ to indicate components of vectors in S_p.

Assume now that $\kappa^{i,j}$ is a known matrix of full rank, and that its matrix inverse is $\kappa_{i,j}$. The cumulants of $Y_i = \kappa_{i,j}Y^j$ may be written with subscripts in the form κ_i, $\kappa_{i,j}$, $\kappa_{i,j,k}$, and so on, and y_i is a quantity that can be computed. The least squares estimate of β is a vector b with components b^r satisfying the orthogonality condition $\langle y - x_r b^r,\, x_s\rangle = 0$ for each s. In other words, the residual vector $y - x_r b^r$ is required to be orthogonal to each of the basis vectors of S_p. This condition evidently requires a metric tensor to determine orthogonality in R^n, the natural one in this case being the inverse covariance matrix $\kappa_{i,j}$. Thus, we arrive at the condition $x^i_s\kappa_{i,j}(y^j - x^j_r b^r) = 0$, or

$$(x^i_r x^j_s \kappa_{i,j})b^s = x^i_r\kappa_{i,j}Y^j = x^i_r Y_i. \tag{3.19}$$

Since the information matrix $x^i_r x^j_s\kappa_{i,j}$ arises rather frequently, we denote it by $\beta_{r,s}$ and its inverse by $\beta^{r,s}$. The reason for this unusual choice of notation is that β^r, $\beta^{r,s}$ and $\beta_{r,s}$ play exactly the same roles in S_p as κ^i, $\kappa^{i,j}$ and $\kappa_{i,j}$ in R^n. In the terminology used in differential geometry and metric spaces, $\beta_{r,s}$ are the components of the induced metric tensor on S_p. Thus (3.19) becomes

$$\beta_{r,s}b^s = b_r = x^i_r Y_i$$

so that b^r and b_r play exactly the same role in S_p as Y^i and Y_i in R^n.

The cumulants of b_r are

$$\begin{aligned} x^i_r\kappa_{i,j}\kappa^j &= x^i_r\kappa_{i,j}x^j_s\beta^s = \beta_{r,s}\beta^s = \beta_r, \\ x^i_r x^j_s\kappa_{i,j} &= \beta_{r,s}, \qquad x^i_r x^j_s x^k_t\kappa_{i,j,k} = \beta_{r,s,t}, \\ x^i_r x^j_s x^k_t x^l_u\kappa_{i,j,k,l} &= \beta_{r,s,t,u} \end{aligned}$$

and so on. The cumulants of the least squares estimate b^r are obtained by raising indices giving β^r, $\beta^{r,s}$,

$$\beta^{r,s,t} = \beta^{r,u}\beta^{s,v}\beta^{t,w}\beta_{u,v,w}$$

and so on.

When we use the term tensor in connection with y^i, κ^i, $\kappa^{i,j}$, $\kappa_{i,j}$ and so on, we allude to the possibility of a change of basis vectors in R^n. In most applications, the transformations that would normally be contemplated in this context are rather limited, but the general theory permits arbitrary linear transformation. In connection with vectors b^r, β^r, $\beta^{r,s}$ and so on, the term tensor refers to the possibility of reparameterization by choosing an alternative set of vectors spanning S_p. In this sense, x_r^i are the components of a tensor in $S_p^* \otimes R^n$, the space of linear transformations from $S_p to\ R^n$.

The residual sum of squares may be written as the difference between the total sum of squares and the regression sum of squares as follows.

$$S^2 = y^i y^j \kappa_{i,j} - b^r b^s \beta_{r,s} = y^i y^j (\kappa_{i,j} - \lambda_{i,j}),$$

where $\lambda^{i,j} = x_r^i x_s^j \beta^{r,s}$ is the covariance matrix of the fitted values, and $\lambda_{i,j} = \kappa_{i,k}\kappa_{j,l}\lambda^{k,l}$ is one choice of annihilating generalized inverse. Note also that

$$\lambda_{i,j} x_r^j = \kappa_{i,k}\lambda^{k,l}\kappa_{l,j} x_r^j = \kappa_{i,j} x_r^j$$

and

$$\kappa^{i,j}\lambda_{i,j} = p.$$

These identities are more easily demonstrated using matrix notation. It follows that the cumulants of S and b are given by

$$\begin{aligned}
E(S^2) &= \kappa^{i,j}(\kappa_{i,j} - \lambda_{i,j}) = n - p \\
\operatorname{var}(S^2) &= (\kappa_{i,j} - \lambda_{i,j})(\kappa_{k,l} - \lambda_{k,l})\kappa^{ij,kl} \\
&= (\kappa_{i,j} - \lambda_{i,j})(\kappa_{k,l} - \lambda_{k,l})\{\kappa^{i,j,k,l} + \kappa^{i,k}\kappa^{j,l}[2]\} \\
&= 2(n-p) + \kappa^{i,j,k,l}(\kappa_{i,j} - \lambda_{i,j})(\kappa_{k,l} - \lambda_{k,l}) \\
\operatorname{cov}(b^r, S^2) &= \beta^{r,s} x_s^i \kappa_{i,j}(\kappa_{k,l} - \lambda_{k,l})\kappa^{j,kl}.
\end{aligned}$$

In the important special case where the observations are independent, we may write, in an obvious notation,

$$\operatorname{var}(S^2) = 2(n-p) + \sum_i \rho_4^i (1 - h_{ii})^2$$

where ρ_4^i is the standardized fourth cumulant of Y^i, $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$ is the projection matrix producing fitted values. Also,

$$\operatorname{cov}(\mathbf{b},\, S^2) = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{C}_3$$

where $c_3^i = \kappa_3^i(1 - h_{ii})/\kappa_2^i$. Alternatively, we may write

$$\operatorname{cov}(\hat{\mu},\, S^2) = \mathbf{H}\mathbf{C}_3$$

where $\hat{\mu} = \mathbf{X}\mathbf{b}$ is the vector of fitted values.

3.8 Cumulant spaces

We examine here the possibility of constructing distributions in R^p having specified cumulants or moments up to a given finite order, n. Such constructions are simpler in the univariate case and the discussion is easier in terms of moments than in terms of cumulants. Thus, we consider the question of whether or not there exists a distribution on the real line having moments $\mu'_1, \mu'_2, \ldots, \mu'_n$, the higher-order moments being left unspecified. If such a distribution exists, it follows that

$$\int (a_0 + a_1 x + a_2 x^2 + \cdots)^2 \, dF(x) dx \geq 0 \tag{3.20}$$

for any polynomial, with equality only if the density is concentrated at the roots of the polynomial. The implication is that for each $r = 1, 2, \ldots, [n/2]$, the matrix of order $r + 1 \times r + 1$

$$M'_r = \begin{pmatrix} 1 & \mu'_1 & \mu'_2 & \cdots & \mu'_r \\ \mu'_1 & \mu'_2 & \mu'_3 & \cdots & \mu'_{r+1} \\ \mu'_2 & \mu'_3 & \mu'_4 & \cdots & \mu'_{r+2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \mu'_r & \mu'_{r+1} & \mu'_{r+2} & \cdots & \mu'_{2r} \end{pmatrix}$$

must be non-negative definite. Equivalently, we may eliminate the first row and column and work with the reduced matrix whose (r, s) element is $\mu_{r,s} = \text{cov}(X^r, X^s)$, where indices here denote powers. The first two inequalities in this sequence are $\kappa_2 \geq 0$ and $\rho_4 \geq \rho_3^2 - 2$.

The above sequence of conditions is necessary but not sufficient to establish that a distribution having the required moments exists. By way of illustration, if we take $\mu'_1 = \mu'_2 = 1$, it follows that $\kappa_2 = 0$ giving $f_X(x) = \delta(x - 1)$, where $\delta(\cdot)$ is Dirac's delta function giving unit probability mass to the origin. In other words, all higher-order moments are determined by these particular values of μ'_1 and μ'_2. Similarly, if $\mu'_1 = 0$, $\mu'_2 = 1$, $\mu'_3 = 0$, $\mu'_4 = 1$, giving $\rho_4 = -2$, we must have

$$f_X(x) = \tfrac{1}{2}\delta(x - 1) + \tfrac{1}{2}\delta(x + 1)$$

because M'_2 has a zero eigenvalue with eigenvector $(-1, 0, 1)$, implying that F must be concentrated at the roots of the polynomial $x^2 - 1$. Since $\mu'_1 = 0$, the weights must be $\frac{1}{2}$ at $x = \pm 1$. Again, all higher-order moments and cumulants are determined by this particular sequence of four moments. It follows that there is no distribution whose moments are $0, 1, 0, 1, 0, 2, \ldots$ even though the corresponding M'_3 is positive semi-definite.

More generally, if for some $k \geq 1$, the matrices $M'_1, \ldots, M'_{k-1}$ are positive definite and $|M'_k| = 0$, then F is concentrated on exactly k points. The k points are the roots of the polynomial whose coefficients are given by the eigenvector of M'_k whose eigenvalue is zero. The probability mass at each point is determined by the first $k - 1$ moments. Since the value of any function at k distinct points can be expressed as a linear combination of the first $k - 1$ polynomials at those points, it follows that

$$\text{rank}(M'_r) = \min(k, r). \tag{3.21}$$

Equivalently, the above rank condition may be deduced from the fact that the integral (3.20) is a sum over k points. Thus M'_r is a convex combination of k matrices each of rank one. The positive definiteness condition together with (3.21) is sufficient to ensure the existence of $f(\cdot)$.

In the case of multivariate distributions, similar criteria may be used to determine whether any particular sequence of arrays is a sequence of moments from some distribution. The elements of the multivariate version of M'_r are arrays suitably arranged. For example, the $(2, 2)$ component of M'_r is a square array of second moments whereas the $(1, 3)$ component is the same array arranged as a row vector. The existence of a distribution is guaranteed if, for each $r = 1, 2, \ldots,$ M'_r is non-negative definite and the rank is maximal, namely

$$\operatorname{rank}(M'_r) = \binom{p+r}{r}.$$

See Exercise 3.31. In the case of rank degeneracy, additional conditions along the lines of (3.21) are required to ensure consistency.

The moment space, $\mathcal{M}_n$, is the set of vectors $(\mu'_1, \ldots, \mu'_n)$ having n components that can be realized as moments of some distribution. Since the moments of the degenerate distribution $\delta(x-t)$ are t^r, any vector of the form $(t, t^2, \ldots, t^n)$ must lie in $\mathcal{M}_n$. In addition, since the moments of a mixture of distributions are mixtures of the moments, $\mathcal{M}_n$ is convex. Consequently, all convex combinations of polynomial vectors, $(t, t^2, \ldots, t^n)$, lie in $\mathcal{M}_n$. Finally, $\mathcal{M}_n$ has dimension n because there are n linearly independent vectors of the above polynomial type. Hence, the moments are functionally independent in the sense that no non-trivial function of any finite set of moments is identically zero for all distributions. Discontinuous functions such as $1 - H(\kappa_2)$ that are identically zero on $\mathcal{M}_n$ are regarded as trivial. The argument just given applies also to the multivariate case where the basis vectors of $\mathcal{M}_n^{(p)}$ are $(t^i, t^j t^k, \ldots)$, superscripts now denoting components. It follows immediately that if F has a finite number of support points, say k, then $\operatorname{rank}(M'_r) \le k$ and that this limit is attained for sufficiently large r.

The cumulant space, $\mathcal{K}_n$, is defined in the obvious way as the set of all vectors that can be realized as cumulants of some distribution. The transformation from $\mathcal{M}_n$ to $\mathcal{K}_n$ is nonlinear and the convexity property of $\mathcal{M}_n$ is lost in the transformation. To see that $\mathcal{K}_n$ is not convex for $n \ge 4$, it is sufficient to observe that the set $\kappa_4 \ge -2\kappa_2^2$ is not convex in the (κ_2, κ_4) plane. However, it is true that if t_1 and t_2 are vectors in $\mathcal{K}_n$, then $t_1 + t_2$ is also in $\mathcal{K}_n$ (Exercise 3.29). As a consequence, if t is in $\mathcal{K}_n$ then λt is also in $\mathcal{K}_n$ for any positive integer $\lambda \ge 1$. This property invites one to suppose that λt lies in $\mathcal{K}_n$ for all $\lambda \ge 1$, not necessarily an integer, but Exercise 3.30 demonstrates that this is not so for $n \ge 6$. The claim is true by definition for infinitely divisible distributions and the counterexample is related to the fact that not all distributions are infinitely divisible. Finally, the transformation from $\mathcal{M}_n$ to $\mathcal{K}_n$ is one to one and continuous, implying that $\mathcal{K}_n$ has dimension n. As a consequence, there are no non-trivial functions of the cumulants that are zero for all distributions. A similar result with obvious modifications applies in the multivariate case.

Identity (3.3) shows that the generalized cumulants are functionally dependent. However, they are *linearly* independent as the following argument shows. Without loss of generality, we may consider an arbitrary linear combination of generalized cumulants, each of the same degree, with coefficients $c(\Upsilon)$. The linear combination, $\sum_{\Upsilon^*} c(\Upsilon^*)\kappa(\Upsilon^*)$ may be written as

$$\begin{aligned} &\sum_{\Upsilon^*} c(\Upsilon^*) \sum_{\Upsilon \ge \Upsilon^*} (-1)^{\nu-1}(\nu-1)!\, \mu(\upsilon_1)\cdots\mu(\upsilon_\nu) \\ &= \sum_{\Upsilon} (-1)^{\nu-1}(\nu-1)!\, \mu(\upsilon_1)\cdots\mu(\upsilon_\nu) C(\Upsilon) \end{aligned} \tag{3.22}$$

where $C(\Upsilon) = \sum_{\Upsilon^* \le \Upsilon} c(\Upsilon^*)$ is an invertible linear function of the original coefficients. The implication is that $\sum c(\Upsilon)\kappa(\Upsilon) = 0$ with $c \ne 0$ implies a syzygy in the moments. From the previous discussion, this is known to be impossible because the moments are functionally independent. A simple extension of this argument covers the case where the linear combination involves generalized cumulants of unequal degree.

3.9 Bibliographic notes

There is some difficulty in tracing the origins of the fundamental identity (3.3). Certainly, it is not stated in Fisher's (1929) paper on k-statistics but Fisher must have known the result in some form in order to derive his rules for determining the joint cumulants of k-statistics. In fact, Fisher's procedure was based on the manipulation of differential operators (Exercise 3.11) and involved an expression for $M_X(\xi)$ essentially the same as (3.11) above. His subsequent calculations for joint cumulants were specific to the k-statistics for which many of the partitions satisfying $\Upsilon \vee \Upsilon^* = 1$ vanish on account of the orthogonality of the k-statistics. Rather surprisingly, despite the large number of papers on k-statistics that appeared during the following decades, the first explicit references to the identity (3.3) did not appear until the papers by James (1958), Leonov & Shiryaev (1959) and James & Mayne (1962). The statement of the result in these papers is not in terms of lattices or graphs. James (1958) uses the term *dissectable* for intersection matrices that do not satisfy the condition in (3.3). He gives a series of rules for determining the moments or cumulants of any homogeneous polynomial symmetric function although his primary interest is in k-statistics. He notes that for k-statistics, only the pattern of non-zero elements of the intersection matrix is relevant, but that in general, the numerical values are required: see Chapter 4. Leonov & Shiryaev (1959) use the term *indecomposability* defined essentially as a connectivity condition on the matrix $\Upsilon^* \cap \Upsilon$: see Exercises 3.4 and 3.5.

Rota's (1964) paper is the source of the lattice-theory notation and terminology. The notation and the derivation of (3.3) are taken from McCullagh (1984b). Alternative derivations and alternative statements of the result can be found in Speed (1983). The earliest statement of the result in a form equivalent to (3.3), though in a different notation, appears to be in Malyshev (1980) under the colourful title 'vacuum cluster expansions'.

There is also a more specialized literature concerned with variances of products: see, for example, Barnett (1955), or Goodman (1960, 1962).

For a thorough discussion of moment spaces, the reader is referred to Karlin & Studden (1966).

3.10 Further results and exercises 3

3.1 Let $X^1, \ldots, X^n$ be independent and identically distributed. By expressing $\bar{X}$ as a linear form and the sample variance, s^2 as a quadratic form, show that $\operatorname{cov}(\bar{X}, s^2) = \kappa_3/n$. Hence show that $\operatorname{corr}(\bar{X}, s^2) \to \rho_3/(2+\rho_4)^{1/2}$ as $n \to \infty$. Show also that $\operatorname{cov}(\bar{X}^2, s^2) = \kappa_4/n^2 + 2\kappa_1\kappa_3/n$ and show that the limiting correlation is, in most cases, non-zero for non-normal variables.

3.2 Let $Y^1, \ldots, Y^n$ be independent and identically distributed random variables with zero mean, variance κ_2, and higher-order cumulants κ_3, κ_4 and so on. Consider the following statistics.

$$k_2 = \sum (Y^i - \bar{Y})^2/(n-1) \qquad l_2 = \sum (Y^i)^2/n$$

$$k_3 = n\sum (Y^i - \bar{Y})^3/\big((n-1)(n-2)\big) \qquad l_3 = \sum (Y^i)^3/n$$

Express k_2 and k_3 as homogeneous polynomials in the form $\phi_{ij}Y^iY^j$ and $\phi_{ijk}y^iY^jY^k$. Show that the coefficients ϕ take the values $1/n$, $-1/(n(n-1))$ or $2/(n(n-1)(n-2))$ depending on the number of distinct indices. Show that k_2 and l_2 are unbiased for κ_2, and that k_3 and l_3 are unbiased for κ_3. Find the variances of all four statistics. Discuss briefly the efficiency of each statistic under the assumption that κ_3 and all higher-order cumulants can be neglected.

3.3 Let M be the intersection matrix $\Upsilon^* \cap \Upsilon$. Columns j_1 and j_2 are said to *hook* if, for some i, $m_{ij_1} > 0$ and $m_{ij_2} > 0$. The set of columns is said to *communicate* if there exists a sequence $j_1, j_2, \ldots, j_\nu$ such that columns j_l and j_{l+1} hook. The matrix M is said to be *indecomposable* if its columns communicate. Show that M is indecomposable if and only if M^T is indecomposable.

3.4 Show, in the terminology of Exercise 3.4, that M is indecomposable if and only if $\Upsilon \vee \Upsilon^* = 1$. (Leonov & Shiryaev, 1959; Brillinger, 1975, Section 2.3.)

3.5 If Υ^* is a 2^k partition (a partition of $2k$ elements into k blocks of 2 elements each), show that the number of 2^k partitions Υ satisfying $\Upsilon \vee \Upsilon^* = 1$, is $2^{k-1}(k-1)!$.

3.6 If $X = X^1, \ldots, X^p$ are jointly normal with zero mean and covariance matrix $\kappa^{i,j}$ of full rank, show that the rth order cumulant of $Y = \kappa_{i,j} X^i X^j$ is $2^{r-1} p(r-1)!$. Hence show that the cumulant generating function of Y is $-\frac{1}{2} p \log(1-2\xi)$ and therefore that Y has the χ^2 distribution on p degrees of freedom.

3.7 Show that $\kappa^{r,s}\kappa^{t,u}\xi_{rt}\xi_{su} = \mathrm{tr}\{(\kappa\xi)^2\}$ where $\kappa\xi$ is the usual matrix product.

3.8 For any positive definite matrix $\mathbf{A}$, define the matrix $\mathbf{B} = \log(\mathbf{A})$ by

$$\mathbf{A} = \exp(\mathbf{B}) = \mathbf{I} + \mathbf{B} + \mathbf{B}^2/2! + \cdots + \mathbf{B}^r/r! + \cdots.$$

By inverting this series, or otherwise, show that

$$\log|\mathbf{A}| = \mathrm{tr}\log(\mathbf{A}),$$

where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$.

3.9 If $X = X^1, \ldots, X^p$ are jointly normal with zero mean, show that the joint cumulant generating function, $\log E\exp(\xi_{ij}Y^{ij})$ of $Y^{ij} = X^i X^j$ is $K_Y(\xi) = -\frac{1}{2}\log|\mathbf{I} - 2\xi\kappa|$. Hence derive the cumulant generating function for the Wishart distribution.

3.10 Let X be a scalar random variable with moments μ_r and cumulants κ_r in the notation of Section 2.5. Show that the rth moment of the polynomial

$$Y = P(X) = a_0 + a_1 X + a_2 X^2 + \cdots$$

is given formally by $P(d)M_X(\xi)|_{\xi=0}$ where $d = d/d\xi$. Hence show that the moment generating function of Y is

$$M_Y(\zeta) = \exp\{\zeta P(d)\} M_X(\xi)|_{\xi=0}$$

in which the operator is supposed to be expanded in powers before attacking the operand (Fisher, 1929, Section 10).

3.11 Show that any polynomial expression in $X^1, \ldots, X^p$, say

$$Q_4 = a_0 + a_i X^i + a_{ij} X^i X^j/2! + a_{ijk} X^i X^j X^k/3! \\ + a_{ijkl} X^i X^j X^k X^l/4!,$$

can be expressed as a homogeneous polynomial in $X^0, X^1, \ldots, X^p$

$$Q_4 = \sum_{ijkl=0}^{p} b_{ijkl} X^i X^j X^k X^l/4!\,,$$

where $X^0 = 1$. Give expressions for the coefficients b in terms of the as. Find $E(Q_4)$ and $\mathrm{var}(Q_4)$ and express the results in terms of the as.

3.12 Consider the first-order autoregressive process $Y_0 = \epsilon_0 = 0$, $Y_j = \beta Y_{j-1} + \epsilon_j$, $j = 1, \ldots, n$, where $|\beta| < 1$ and ϵ_j are independent $N(0,1)$ random variables. Show that the log likelihood function has first derivative $U = \partial l/\partial\beta = T_1 - \beta T_2$ where $T_1 = \Sigma Y_j Y_{j-1}$ and $T_2 = \Sigma Y_{j-1}^2$ with summation from 1 to n. By expressing U as a quadratic form in ϵ, show that the first three cumulants of U are $E(U) = 0$,

$$\begin{aligned}\kappa_2(U) &= \sum_{1\le i<j\le n} \beta^{2j-2i-2} \\ &= (1-\beta^2)^{-1}\{n - (1-\beta^{2n})/(1-\beta^2)\} = E(T_2) \\ \kappa_3(U) &= 6 \sum_{1\le i<j<k\le n} \beta^{2k-2i-3} \\ &= \frac{6\{n\beta(1-\beta^2) - 2\beta + n\beta^{2n-1}(1-\beta^2) + 2\beta^{2n+1}\}}{(1-\beta^2)^3}.\end{aligned}$$

3.13 Let $Y = Y^1, \ldots, Y^p$ have zero mean and covariance matrix $\kappa^{i,j}$. Show that the 'total variance', $\sigma^2 = E(Y^i Y^j \delta_{ij})$, is invariant under orthonormal transformation of Y. For any given direction, ϵ, define

$$\begin{aligned}\sigma_\epsilon^2 &= \mathrm{var}(\epsilon_i Y^i) = \epsilon_i\epsilon_j\kappa^{i,j} \\ \tau_\epsilon^2 &= \epsilon_i\epsilon_j\{\sigma^2\delta^{ij} - \kappa^{i,j}\} = \epsilon_i\epsilon_j I^{ij}.\end{aligned}$$

Give an interpretation of σ_ϵ^2 and τ_ϵ^2 as regression and residual variances respectively. Show also that, in mechanics, τ_ϵ^2 is the moment of inertia of a rigid body of unit mass about the axis ϵ. Hence, interpret I^{ij} as the *inertia tensor* (Synge & Griffith, 1949, Section 11.3; Jeffreys & Jeffreys, 1956, Section 3.08).

3.14 *MacCullagh's formula*: Using the notation of the previous exercise, let $X^i = Y^i + \kappa^i$, where κ^i are the components of a vector of length $\rho > 0$ in the direction ϵ. Show that

$$\begin{aligned}E\left(\frac{1}{|X|}\right) &= \frac{1}{\rho} + \frac{1}{2\rho^3}\left(-\sigma^2 + 3\sigma_\epsilon^2\right) + O(\rho^{-4}) \\ &= \frac{1}{\rho} + \frac{1}{2\rho^3}\left(2\sigma^2 - 3\tau_\epsilon^2\right) + O(\rho^{-4}),\end{aligned}$$

(MacCullagh, 1855; Jeffreys & Jeffreys, 1956, Section 18.09). [The above expression gives the potential experienced at an external point (the origin) due to a unit mass or charge distributed as $f_X(x)$. The correction term is sometimes called the *gravitational quadrupole* (Kibble, 1985, Chapter 6).]

3.15 Show that the sum

$$\sum_{\Upsilon\ge\Upsilon^*} (-1)^{\nu-1}(\nu-1)! = \delta(\Upsilon^*, 1)$$

is zero unless $\Upsilon^* = 1$.

3.16 By reversing the order of summation in (3.17), show that the generalized cumulant may be written

$$\kappa(\Upsilon^*) = \sum_{\Pi} \kappa(\pi_1)\cdots\kappa(\pi_\sigma) \sum_{\substack{\Upsilon\ge\Upsilon^* \\ \Upsilon\ge\Pi}} (-1)^{\nu-1}(\nu-1)!.$$

Hence, using the result of the previous exercise, prove the fundamental identity (3.3).

3.17 Let X be a scalar random variable whose distribution is Poisson with mean 1. Show that the cumulants of X of all orders are equal to 1. Hence show that the rth moment is

$$\mu_r' = E(X^r) = B_r$$

where B_r, the rth Bell number, is the number of partitions of a set of r elements. Hence derive a generating function for the Bell numbers.

3.18 Let $\Upsilon^* = \{\upsilon_1^*, \ldots, \upsilon_b^*\}$ be a partition of a set of p elements into b non-empty blocks of sizes $|\upsilon_1^*|, \ldots, |\upsilon_b^*|$. By using (3.16) or otherwise, show that the number of partitions complementary to Υ^* is

$$\sum_{\substack{\Upsilon \geq \Upsilon^* \\ \Upsilon = \{\upsilon_1, \ldots, \upsilon_\nu\}}} (-1)^{\nu-1}(\nu-1)!\, B_{|\upsilon_1|} \cdots B_{|\upsilon_\nu|}.$$

3.19 Interpret the expression in Exercise 3.19 as the cumulant of order b

$$\text{cum}\left(X^{|\upsilon_1^*|}, \ldots, X^{|\upsilon_b^*|}\right)$$

where X is defined in Exercise 3.18 and the superscripts denote powers. Simplify this result in the special case where Υ^* is a 2^k partition and compare with Exercise 3.6.

3.20 Show that the number of sub-partitions of Υ^* is given by

$$\sum_{\Upsilon \leq \Upsilon^*} 1 = B_{|\upsilon_1^*|} \cdots B_{|\upsilon_b^*|}.$$

3.21 Using the result given in Exercise 3.19, show that the total number of ordered pairs of partitions (Υ_1, Υ_2) satisfying $\Upsilon_1 \vee \Upsilon_2 = 1$ is

$$C_p^{(2)} = \sum_{\Upsilon} (-1)^{\nu-1}(\nu-1)!\, B^2_{|\upsilon_1|} \cdots B^2_{|\upsilon_\nu|}$$

where the partitions contain p elements and B_r^2 is the square of the rth Bell number. Deduce also that $C_p^{(2)}$ is the pth cumulant of $Y = X_1X_2$ where the Xs are independent Poisson random variables with unit mean.

3.22 Show that the number of ordered pairs (Υ_1, Υ_2) satisfying $\Upsilon_1 \vee \Upsilon_2 = \Upsilon^*$ for some fixed partition Υ^*, is

$$\sum_{\Upsilon \leq \Upsilon^*} m(\Upsilon, \Upsilon^*) B^2_{|\upsilon_1|} \cdots B^2_{|\upsilon_\nu|},$$

where $m(\Upsilon, \Upsilon^*)$ is the Möbius function for the partition lattice, defined below (3.12). Hence prove that the total number of ordered triplets $(\Upsilon_1, \Upsilon_2, \Upsilon_3)$ satisfying $\Upsilon_1 \vee \Upsilon_2 \vee \Upsilon_3 = 1$ is

$$C_p^{(3)} = \sum_{\Upsilon} (-1)^{\nu-1}(\nu-1)!\, B^3_{|\upsilon_1|} \cdots B^3_{|\upsilon_\nu|}.$$

Show also, in the notation of Exercise 3.22, that $C_p^{(3)}$ is the pth cumulant of the triple product $Y = X_1X_2X_3$.

3.23 Generalize the result of the previous exercise to ordered k-tuplets of partitions. Give a simple explanation for this result.

3.24 An alternative way of representing a partition by means of a graph is to use p labelled edges instead of labelled vertices. In this form, the graph of $\Upsilon^* = \{\upsilon_1, \ldots, \upsilon_\nu\}$ comprises p labelled edges emanating from b unlabelled vertices, giving a total of $b + p$ vertices of which p are 'free'. If $\Upsilon = \{\upsilon_1, \ldots, \upsilon_\nu\}$ is another partition of the same indices represented as a graph in the same way, we define the graph $\Upsilon \otimes \Upsilon^*$ by connecting corresponding free vertices of the two graphs. This gives a graph with $b + \nu$ unlabelled vertices and p labelled edges, parallel edges being permitted. Show that the graph $\Upsilon \otimes \Upsilon^*$ is connected if and only if $\Upsilon \vee \Upsilon^* = 1$.

3.25 In the notation of the previous exercise, prove that all cycles in the graph $\Upsilon \otimes \Upsilon^*$ have even length. (A cycle is a path beginning and ending at the same vertex.) Such a graph is said to be even. Show that all even connected graphs have a unique representation as $\Upsilon \otimes \Upsilon^*$. Hence prove that the number of connected even graphs having p labelled edges is $(C_p^{(2)} + 1)/2$ where $C_p^{(2)}$ is defined in Exercise 3.22 (Gilbert, 1956).

3.26 In the terminology of the previous two exercises, what does $C_p^{(3)}$ in Exercise 3.23 correspond to?

3.27 By considering the mixture density, $pf_1(x) + (1 - p)f_2(x)$, show that the moment space, $\mathcal{M}_n$, is convex.

3.28 By considering the distribution of the sum of two independent random variables, show that the cumulant space, $\mathcal{K}_n$ is closed under vector addition.

3.29 Show that there exists a unique distribution whose odd cumulants are zero and whose even cumulants are $\kappa_2 = 1$, $\kappa_4 = -2$, $\kappa_6 = 16$, $\kappa_8 = -272$, $\kappa_{10} = 7936, \ldots$. Let $M_r'(\lambda)$ be the moment matrix described in Section 3.8, corresponding to the cumulant sequence, $\lambda\kappa_1, \ldots, \lambda\kappa_{2r}$. Show that for the particular cumulant sequence above, the determinant of $M_r'(\lambda)$ is

$$|M_r'| = 1!\,2!\,...r!\,\lambda^r(\lambda - 1)^{r-1} \cdots (\lambda - r + 1),$$

for $r = 1, 2, 3, 4$. Hence prove that there is no distribution whose cumulants are $\{\lambda\kappa_r\}$ for non-integer $\lambda < 3$. Find the unique distribution whose cumulants are $\{\lambda\kappa_r\}$ for $\lambda = 1, 2, 3$.

3.30 By counting the number of distinguishable ways of placing r identical objects in $p + 1$ labelled boxes, show that, in p dimensions,

$$\operatorname{rank}(M_r') \leq \binom{p + r}{r},$$

where M_r' is defined in Section 3.8.

3.31 Show that achievement of the limit $\bar{\rho}_4 = \bar{\rho}_{23}^2 - p - 1$ implies the following constraint on the third cumulants

$$\bar{\rho}_{23}^2 - \bar{\rho}_{13}^2 - p + 1 = 0.$$

3.32 Show that, if complex-valued random variables are permitted, there are no restrictions on the moment spaces or on the cumulant spaces such as those discussed in Section 3.8.

3.33 Describe the usual partial order on the set of partitions of the *number* k. Explain why this set, together with the usual partial order, does *not* form a lattice for $k \geq 5$. Verify by direct inspection that the structure is a lattice for each $k \leq 4$.