CHAPTER 1

Index notation

1.1 Introduction

It is a fact not widely acknowledged that, with appropriate choice of notation, many multivariate statistical calculations can be made simpler and more transparent than the corresponding univariate calculations. This simplicity is achieved through the systematic use of index notation and special arrays called tensors. For reasons that are given in the following sections, matrix notation, a reliable workhorse for many second-order calculations, is totally unsuitable for more complicated calculations involving either non-linear functions or higher-order moments. The aim of this book is to explain how index notation simplifies many statistical calculations, particularly those involving moments or cumulants of non-linear functions. Other applications where index notation greatly simplifies matters include k-statistics, Edgeworth and conditional Edgeworth approximations, saddlepoint and Laplace approximations, calculations involving conditional cumulants, moments of maximum likelihood estimators, likelihood ratio statistics and the construction of ancillary statistics. These topics are the subject matter of later chapters.

In some ways, the most obvious and, at least initially, most disconcerting aspect of index notation is that the components of the vector of primary interest, usually a parameter, θ, or a random variable, X, are indexed using superscripts. Thus, θ^2, the second component of the vector θ, is not to be confused with the square of any component. For this reason, powers are best avoided unless the context leaves no room for ambiguity, and the square of θ^2 is written simply as $\theta^2\theta^2$. In view of the considerable advantages achieved, this is a very modest premium to pay.

1.2 The summation convention

Index notation is a convention for the manipulation of multi-dimensional arrays. The elements of these arrays are called either *components* or *coefficients* depending on the context. In the context of parametric inference and in manipulations associated with likelihood functions, it is appropriate to take the unknown parameter as the vector of interest: see the first example in Section 2.4. Here, however, we take as our vector of interest the p-dimensional random variable X with components $X^1, \ldots, X^p$. In this context, arrays of constants used in the formation of linear combinations are called *coefficients*. This terminology is merely a matter of convention but it appears to be useful and the notation does emphasize it. Thus, for example, $\kappa^i = E(X^i)$ is a one-dimensional array whose components are the means of the components of X and $\kappa^{ij} = E(X^i X^j)$ is a two-dimensional array whose components are functions of the joint distributions of pairs of variables.

Probably the most convenient aspect of index notation is the implied summation over any index repeated once as a superscript and once as a subscript. The range of summation is not stated explicitly but is implied by the positions of the repeated index and by conventions regarding the range of the index. Thus,

$$a_i X^i \equiv \sum_{i=1}^{p} a_i X^i \tag{1.1}$$

specifies a linear combination of the Xs with coefficients $a_1, \ldots, a_p$. Quadratic and cubic forms in

X with coefficients a_{ij} and a_{ijk} are written in the form

$$a_{ij}X^iX^j \quad \text{and} \quad a_{ijk}X^iX^jX^k \tag{1.2}$$

and the extension to homogeneous polynomials of arbitrary degree is immediate.

For the sake of simplicity, and with no loss of generality, we take all multiply-indexed arrays to be symmetric under index permutation but, of course, subscripts may not be interchanged with superscripts. The value of this convention is clearly apparent when we deal with scalars such as $a_{ij}a_{kl}\omega^{ijkl}$, which, by convention only, is the same as $a_{ik}a_{jl}\omega^{ijkl}$ and $a_{il}a_{jk}\omega^{ijkl}$. For instance, if $p = 2$ and $a_{ij} = \delta_{ij} = 1$ if $i = j$ and 0 otherwise, then, without the convention,

$$a_{ij}a_{kl}\omega^{ijkl} - a_{ik}a_{jl}\omega^{ijkl} = \omega^{1122} + \omega^{2211} - \omega^{1212} - \omega^{2121}$$

and this is not zero unless ω^{ijkl} is symmetric under index permutation.

Expressions (2.1) and (2.2) produce one-dimensional or scalar quantities, in this case scalar random variables. Suppose instead, we wish to construct a vector random variable Y with components $Y^1, \ldots, Y^q$, each of which is linear in X, we may write

$$Y^r = a_i^r X^i \tag{1.3}$$

and $r = 1, \ldots, q$ is known as a *free* index. Similarly, if the components of Y are homogeneous quadratic forms in X, we may write

$$Y^r = a_{ij}^r X^i X^j. \tag{1.4}$$

Non-homogeneous quadratic polynomials in X may be written in the form

$$Y^r = a^r + a_i^r X^i + a_{ij}^r X^i X^j.$$

Where two sets of indices are required, as in (2.3) and (2.4), one referring to the components of X and the other to the components of Y, we use the sets of indices $i, j, k, \ldots$ and $r, s, t, \ldots$. Occasionally it will be necessary to introduce a third set, $\alpha, \beta, \gamma, \ldots$ but this usage will be kept to a minimum.

All of the above expressions could, with varying degrees of difficulty, be written using matrix notation. For example, (2.1) is typically written as $\mathbf{a}^T\mathbf{X}$ where $\mathbf{a}$ and $\mathbf{X}$ are column vectors; the quadratic expression in (2.2) is written $\mathbf{X}^T\mathbf{A}\mathbf{X}$ where $\mathbf{A}$ is symmetric, and (2.3) becomes $\mathbf{Y} = \mathbf{A}^*\mathbf{X}$ where $\mathbf{A}^*$ is of order $q \times p$. From these examples, it is evident that there is a relationship of sorts between column vectors and the use of superscripts, but the notation $\mathbf{X}^T\mathbf{A}\mathbf{X}$ for $a_{ij}X^iX^j$ violates the relationship. The most useful distinction is not in fact between rows and columns but between coefficients and components and it is for this reason that index notation is preferred here.

1.3 Tensors

The term *tensor* is used in this book in a well-defined sense, similar in spirit to its meaning in differential geometry but with minor differences in detail. It is not used as a synonym for array, index notation or the summation convention. A cumulant tensor, for example, is a symmetric array whose elements are functions of the joint distribution of components of the random variable of interest, X say. The values of these elements in any one coordinate system are real numbers but, when we describe the array as a tensor, we mean that the values in one coordinate system, Y say, can be obtained from those in any other system, X say, by the application of a particular transformation formula. The nature of this transformation is the subject of Sections 3.4 and 4.5, and in fact, we consider not just changes of basis, but also non-invertible transformations.

When we use the adjectives *covariant* and *contravariant* in reference to tensors, we refer to the way in which the arrays transform under a change of variables from the original x to new variables y. In statistical calculations connected with likelihood functions, x and y are typically parameter vectors but in Chapters 2 and 3, x and y refer to random variables. To define the adjectives covariant and contravariant more precisely, we suppose that ω is a d-dimensional array whose elements are functions of the components of x, taken d at a time. We write $\omega = \omega^{i_1 i_2 \cdots i_d}$ where the d components need not be distinct. Consider the transformation $y = g(x)$ from x to new variables $y = y^1, \ldots, y^p$ and let $a_i^r \equiv a_i^r(x) = \partial y^r / \partial x^i$ have full rank for all x. If $\bar{\omega}$, the value of ω for the transformed variables, satisfies

$$\bar{\omega}^{r_1 r_2 \cdots r_d} = a_{i_1}^{r_1} a_{i_2}^{r_2} \ldots a_{i_d}^{r_d} \omega^{i_1 i_2 \cdots i_d} \tag{1.5}$$

then ω is said to be a contravariant tensor. On the other hand, if ω is a covariant tensor, we write $\omega = \omega_{i_1 i_2 \cdots i_d}$ and the transformation law for covariant tensors is

$$\bar{\omega}_{r_1 r_2 \cdots r_d} = b_{r_1}^{i_1} b_{r_2}^{i_2} \ldots b_{r_d}^{i_d} \omega_{i_1 i_2 \cdots i_d} \tag{1.6}$$

where $b_r^i = \partial x^i / \partial y^r$, the matrix inverse of a_i^r, satisfies $a_i^r b_r^j = \delta_i^j = a_r^j b_i^r$.

The function $g(.)$ is assumed to be an element of some group, either specified explicitly or, more commonly, to be inferred from the statistical context. For example, when dealing with transformations of random variables or their cumulants, we usually work with the general linear group (2.3) or the general affine group (2.8). Occasionally, we also work with the smaller orthogonal group, but when we do so, the group will be stated explicitly so that the conclusions can be contrasted with those for the general linear or affine groups. On the other hand, when dealing with possible transformations of a vector of parameters, it is natural to consider non-linear but invertible transformations and $g(.)$ is then assumed to be a member of this much larger group. In other words, when we say that an array of functions is a tensor, the statement has a well defined meaning only when the group of transformations is specified or understood.

It is possible to define hybrid tensors having both subscripts and superscripts that transform in the covariant and contravariant manner respectively. For example, if ω^{ij} and ω_{ijk} are both tensors, then the product $\gamma_{klm}^{ij} = \omega^{ij} \omega_{klm}$ is a tensor of covariant order 3 and contravariant order 2. Furthermore, we may sum over pairs of indices, a process known as *contraction*, giving

$$\gamma_{kl}^{i} = \gamma_{klj}^{ij} = \omega^{ij} \omega_{klj}.$$

A straightforward calculation shows that γ_{kl}^i is a tensor because, under transformation of variables, the transformed value is

$$\bar{\gamma}_{klm}^{ij} = \gamma_{tuv}^{rs} a_r^i a_s^j b_k^t b_l^u b_m^v$$

and hence, summation over $m = j$ gives

$$\bar{\gamma}_{kl}^{i} = \gamma_{tu}^{r} a_r^i b_k^t b_l^u.$$

Thus, the tensor transformation property is preserved under multiplication and under contraction. An important consequence of this property is that scalars formed by contraction of tensors must be invariants. In effect, they must satisfy the transformation law of zero order tensors. See Section 2.6.

One of the problems associated with tensor notation is that it is difficult to find a satisfactory notation for tensors of arbitrary order. The usual device is to use subscripted indices as in (2.5) and (2.6), but this notation is aesthetically unpleasant and is not particularly easy to read. For these reasons, subscripted indices will be avoided in the remainder of this book. Usually we give explicit expressions involving up to three or four indices. The reader is then expected to infer the necessary generalization, which is of the type (2.5), (2.6) if we work with tensors but is usually more complicated if we work with arbitrary arrays.

1.4 Examples

In this and in the following chapter, X and Y are random variables but when we work with log likelihood derivatives, it is more appropriate to contemplate transformation of the parameter vector and the terms covariant and contravariant then refer to parameter transformations and not to data transformations. To take a simple example, relevant to statistical theory, let $l(\theta; Z) = \log f_Z(Z; \theta)$ be the log likelihood function for $\theta = \theta^1, \ldots, \theta^p$ based on observations Z. The partial derivatives of l with respect to the components of θ may be written

$$U_r(\theta) = \partial l(\theta; Z)/\partial \theta^r$$
$$U_{rs}(\theta) = \partial^2 l(\theta; Z)/\partial \theta^r \partial \theta^s$$

and so on. The maximum likelihood estimate of θ satisfies $U_r(\hat{\theta}) = 0$ and the observed information for θ is $I_{rs} = -U_{rs}(\hat{\theta})$, with matrix inverse I^{rs}. Suppose now that we were to re-parameterize in terms of $\phi = \phi^1, \ldots, \phi^p$. If we denote by an asterisk derivatives with respect to ϕ, we have

$$\begin{aligned} U_r^* &= \theta_r^i U_i, & U_{rs}^* &= \theta_r^i \theta_s^j U_{ij} + \theta_{rs}^i U_i, \\ I_{rs}^* &= \theta_r^i \theta_s^j I_{ij}, & I^{*rs} &= \phi_i^r \phi_j^s I^{ij} \end{aligned} \tag{1.7}$$

where

$$\theta_r^i = \partial \theta^i / \partial \phi^r, \qquad \theta_{rs}^i = \partial^2 \theta^i / \partial \phi^r \partial \phi^s$$

and θ_r^i is assumed to have full rank with matrix inverse $\phi_i^r = \partial \phi^r / \partial \theta^i$. Arrays that transform like U_r, I_{rs} and I^{rs} are tensors, the first two being covariant of orders 1 and 2 respectively and the third being contravariant of order 2. The second derivative, $U_{rs}(\theta)$, is not a tensor on account of the presence of second derivatives with respect to θ in the above transformation law. Note also that the array U_{rs}^* cannot be obtained by transforming the array U_{rs} alone: it is necessary also to know the value of the array U_r. However $E\{U_{rs}(\theta); \theta\}$, the Fisher information for θ, is a tensor because the second term in U_{rs}^* has mean zero at the true θ.

To take a second example, closer in spirit to the material in the following two chapters, let $X = X^1, \ldots, X^p$ have mean vector $\kappa^i = E(X^i)$ and covariance matrix

$$\kappa^{i,j} = \operatorname{cov}(X^i, X^j) = E(X^i X^j) - E(X^i)E(X^j).$$

Suppose we make an affine transformation to new variables $Y = Y^1, \ldots, Y^q$, where

$$Y^r = a^r + a_i^r X^i. \tag{1.8}$$

The mean vector and covariance matrix of Y are easily seen to be

$$a^r + a_i^r \kappa^i \quad \text{and} \quad a_i^r a_j^s \kappa^{i,j}$$

where $a_i^r = \partial Y^r / \partial X^i$. Thus, even though the transformation may not be invertible, the covariance array transforms like a contravariant tensor. Arrays that transform in this manner, but only under linear or affine transformation of X, are sometimes called *Cartesian tensors* (Jeffreys, 1952). Such transformations are of special interest because $a_i^r = \partial Y^r / \partial X^i$ does not depend on X. It will be shown that cumulants of order two or more are not tensors in the sense usually understood in differential geometry, but they do behave as tensors under the general affine group (2.8). Under non-linear transformation of X, the cumulants transform in a more complicated way as discussed in Section 4.4.

Tensors whose components are unaffected by coordinate transformation are called *isotropic.* This terminology is most commonly used in Mechanics and in the physics of fluids, where all three coordinate axes are measured in the same units. In these contexts, the two groups of most relevance are the orthogonal group, O, and the orthogonal group with positive determinant, O^+. In either case, δ^i_j, δ_{ij} and δ^{ij} are isotropic tensors. There is exactly one isotropic third-order tensor under O^+ (Exercise 1.22). However, this tensor, called the *alternating tensor*, is anti-symmetrical and does not occur in the remainder of this book. All fourth-order isotropic tensors are functions of the three second-order isotropic tensors (Jeffreys, 1952, Chapter 7). The only symmetrical isotropic fourth-order tensors are

$$\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}, \quad \delta^{ij}\delta^{kl} + \delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk},$$
$$\delta^{ij}\delta_{kl} \quad \text{and} \quad \delta^i_k\delta^j_l + \delta^i_l\delta^j_k,$$

(Thomas, 1965, Section 7). Isotropic tensors play an important role in physics (see Exercise 1.21) but only a minor role in statistics.

1.5 Elementary matrix theory

For later use, we state here without detailed proof, some elementary matrix-theory results using tensor notation and terminology. Our main interest lies in matrix inverses and spectral decompositions or eigenvalue decompositions of real symmetric matrices. We consider first the tensorial properties of generalized inverse matrices.

1.5.1 *Generalized inverse matrices*

Let ω_{ij} be a symmetric covariant tensor written as Ω using matrix notation. A generalized inverse of Ω is any matrix Ω^- satisfying

$$\Omega\Omega^-\Omega = \Omega, \tag{1.9}$$

implying that $\Omega\Omega^-$ acts as the identity on the range of Ω. It follows that $\text{rank}(\Omega\Omega^-) \geq \text{rank}(\Omega)$, and hence that $\text{rank}(\Omega^-) \geq \text{rank}(\Omega)$. Post-multiplication of (2.9) by Ω^- gives $(\Omega\Omega^-)(\Omega\Omega^-) = \Omega\Omega^-$. In other words, $\Omega\Omega^-$ is idempotent and hence

$$\text{tr}(\Omega\Omega^-) = \text{rank}(\Omega\Omega^-) = \text{rank}(\Omega). \tag{1.10}$$

This conclusion is independent of the choice of generalized inverse. In addition, if Ω^- is a generalized inverse of Ω and if $\mathbf{A}$ is any full-rank matrix, then $\mathbf{A}\Omega^-\mathbf{A}$ is a generalized inverse of $\mathbf{A}^{-1}\Omega\mathbf{A}^{-1}$. The last result follows directly from the definition (2.9).

Reverting now to tensor notation, we write the generalized matrix inverse of ω_{ij} as ω^{ij}. The inverse can in fact always be chosen to be symmetric, so that this notation does not conflict with our convention regarding symmetry. It follows from (2.10) that, whatever the choice of inverse,

$$\text{rank}(\omega_{ij}) = \omega_{ij}\omega^{ij}.$$

In addition, if ω^{ij} is any generalized inverse of ω_{ij}, then $a^r_i a^s_j \omega^{ij}$ is a generalized inverse of $b^i_r b^j_s \omega_{ij}$, where a^i_r is a full rank matrix with inverse b^r_i. In other words, ω^{ij} is a contravariant tensor.

These arguments are entirely independent of the choice of generalized inverse. Occasionally, however, it is convenient to choose a generalized inverse with the additional property that

$$\Omega^-\Omega\Omega^- = \Omega^- \tag{1.11}$$

implying that $\text{rank}(\Omega^-) = \text{rank}(\Omega)$. In other words, Ω is a generalized inverse of Ω^-. The symmetry of the tensor formulae is greatly enhanced if the generalized inverse matrix is chosen to have the property (2.11) for then we need not distinguish between ω^{ij} as a generalized inverse matrix of ω_{ij} and ω_{ij} as a generalized inverse of ω^{ij}.

In fact, it is always possible to choose a generalized inverse with the properties (2.9) and (2.11) and having the additional symmetry property that

$$\Omega^{-}\Omega = (\Omega^{-}\Omega)^T, \quad \Omega\Omega^{-} = (\Omega\Omega^{-})^T.$$

Such a generalized inverse is unique and is known as the Moore-Penrose inverse (Rao, 1973, Section 1b). See also Exercise 1.10.

Conditions (2.9), (2.11) are entirely natural whether Ω is the matrix representation of a symmetric covariant tensor, a symmetric contravariant tensor or an asymmetric (1,1) tensor. On the other hand, the symmetry conditions, as stated above, appear to be quite unnatural if Ω is the matrix representation of an asymmetric (1,1) tensor, but otherwise, the conditions seem sensible. The symmetry condition arises naturally if the usual Euclidean inner product with equal weights is used to determine orthogonality: see Kruskal (1975, Section 6). On the other hand, if a weighted inner product is appropriate, as it often is in statistical applications, then the symmetry condition would seem to be inappropriate. See, for example, Exercise (2.11).

1.5.2 *Spectral decomposition*

Any real symmetric covariant tensor ω_{ij} may be written in the form

$$\omega_{ij} = \sigma_i^r \sigma_j^s \Lambda_{rs} \tag{1.12}$$

where $\Lambda_{rr} = \lambda_r$, a real number, $\Lambda_{rs} = 0$ for $r \neq s$ and σ_i^r is a real orthogonal matrix satisfying

$$\sigma_i^r \sigma_j^s \delta_{rs} = \delta_{ij}.$$

The values $\lambda_1, \ldots, \lambda_p$ are known as the eigenvalues of ω_{ij} and (2.12) is known as the eigenvalue decomposition or spectral decomposition of ω. This decomposition implies that the quadratic form $Q = \omega_{ij}x^i x^j$ may be written as $Q = \sum_1^p \lambda_r (y^r)^2$ where $y^r = \sigma_i^r x^i$ is an orthogonal transformation of x. The set of eigenvalues is unique but evidently the representation (2.12) is not unique because we may at least permute the components of y. Further, if some of the eigenvalues are equal, say $\lambda_1 = \lambda_2$, any orthogonal transformation of the components (y^1, y^2) satisfies (2.12).

Under orthogonal transformation of x, ω_{ij} transforms to $\bar{\omega}_{ij} = a_i^k a_j^l \omega_{kl}$ where a_i^k is orthogonal. The spectral decomposition (2.12) then becomes

$$\bar{\omega}_{ij} = (a_i^k \sigma_k^r)(a_j^k \sigma_l^s)\Lambda_{rs} \tag{1.13}$$

where $a_i^k \sigma_k^r$ is an orthogonal matrix. On comparing (2.12) with (2.13) we see that the set of eigenvalues of ω_{ij} is invariant under orthogonal transformation of coordinates. The eigenvalues are not invariant under arbitrary nonsingular transformation because $a_i^k \sigma_k^r$ is not, in general, orthogonal unless a_i^k is orthogonal.

Consider now the alternative decomposition

$$\omega_{ij} = \tau_i^r \tau_j^s \epsilon_{rs} \tag{1.14}$$

where $\epsilon_{rr} = \pm 1$ or zero, $\epsilon_{rs} = 0$ for $r \neq s$ and no constraints are imposed on τ_i^r other than that it should be real and have full rank. The existence of such a decomposition follows from (2.12). Again, the representation (2.14) is not unique because, if we write $y^r = \tau_i^r x^i$, then $Q = \omega_{ij}x^i x^j$ becomes

$$Q = \sum\nolimits^{+} (y^r)^2 - \sum\nolimits^{-} (y^r)^2 \tag{1.15}$$

where the first sum is over those y^r for which $\epsilon_{rr} = +1$ and the second sum is over those components for which $\epsilon_{rr} = -1$. Two orthogonal transformations, one for the components y^r

for which $\epsilon_{rr} = +1$ and one for the components for which $\epsilon_{rr} = -1$, leave (2.15) unaffected. Furthermore, the components for which $\epsilon_{rr} = 0$ may be transformed linearly and all components may be permuted without affecting the values of ϵ_{rs} in (2.14). At most, the order of the diagonal elements, ϵ_{rr} can be changed by the transformations listed above.

Under linear transformation of x, (2.14) becomes

$$\bar{\omega}_{ij} = (a_i^k \tau_k^r)(a_j^l \tau_l^s)\epsilon_{rs},$$

so that the matrix rank

$$\omega_{ij}\omega^{ij} = \epsilon_{rs}\epsilon^{rs}$$

and signature,

$$\delta^{rs}\epsilon_{rs} = \sum \epsilon_{rr}$$

are invariant functions of the covariant tensor ω_{ij}. This result is known as Sylvester's law of inertia (Gantmacher, 1960, Chapter X; Cartan, 1981, Section 4).

The geometrical interpretation of Sylvester's law is that the equation

$$x^0 = Q = \omega_{ij}x^i x^j$$

describes a hypersurface of dimension p in R^{p+1} and the qualitative aspects of the shape of this surface that are invariant under linear transformation of $x^1, \ldots, x^p$ are the numbers of positive and negative principal curvatures. This makes good sense because the effect of such a transformation is to rotate the coordinates and to re-define distance on the surface. The surface, in a sense, remains intact. If we were to change the sign of x^0, the positive and negative curvatures would be reversed.

In the particular case where ω_{ij} is positive definite of full rank, the matrix τ_i^r in (2.14) is known as a matrix square root of ω. For this case, if σ_r^s is an orthogonal matrix, then $\tau_i^r\sigma_r^s$ is also a matrix square root of ω_{ij}. Subject to this choice of orthogonal transformation, the matrix square root is unique.

1.6 Invariants

An invariant is a function whose value is unaffected by transformations within a specified class or group. To take a simple example, let ω_i^r be a mixed tensor whose value under linear transformation becomes

$$\bar{\omega}_i^r = a_s^r \omega_j^s b_i^j,$$

where $a_s^r b_i^s = \delta_i^r$. In matrix notation, $\bar{\Omega} = \mathbf{A}\Omega\mathbf{A}^{-1}$ is known as a *similarity* transformation or *unitary* transformation, (Dirac, 1958, Chapter 26). In Cartan's terminology (Cartan, 1981, Section 41), the matrices Ω and $\bar{\Omega}$ are said to be *equivalent*. It is an elementary exercise to show that $\mathrm{tr}(\bar{\Omega}) = \mathrm{tr}(\Omega)$, so that the sum of the eigenvalues of a real (1,1) tensor is (a) real and (b) invariant. The same is true of any symmetric polynomial function of the eigenvalues. In particular, the determinant is invariant. For an interpretation, see Exercise 1.13.

The second example is closer in spirit to the material in the following chapters in the sense that it involves random variables in an explicit way. Let $\kappa^{i,j}$ be the covariance matrix of the components of X and consider the effect on $\kappa^{i,j}$ of making an orthogonal transformation of X. By (2.13), the set of eigenvalues is unaffected. Thus the set of eigenvalues is an invariant of a symmetric contravariant tensor, but only within the orthogonal group. Only the rank and signature are invariant under nonsingular linear or affine transformation. Other examples of invariant functions of the cumulants are given in Section 3.8.

These examples pinpoint one serious weakness of matrix notation, namely that no notational distinction is made between a (1,1) tensor whose eigenvalues are invariant under the full linear group, and a (2,0) tensor, invariably symmetric, wwhose eigenvalues are invariant only under the smaller orthogonal group.

The log likelihood function itself is invariant under arbitrary smooth parameter transformation, not necessarily linear. Under nonsingular transformation of the data, the log likelihood is not invariant but transforms to $l(\theta; z) + c(z)$ where $c(z)$ is the log determinant of the Jacobian of the transformation. However $l(\hat{\theta}; z) - l(\theta; z)$, the maximized log likelihood ratio statistic, is invariant under transformation of the data. Some authors define the log likelihood as the equivalence class of all functions that differ from $l(\theta; z)$ by a function of z and, in this sense, the log likelihood function is an invariant.

To test the simple hypothesis that θ takes on some specified value, say θ_0, it is desirable in principle to use an invariant test statistic because consistency of the observed data with the hypothesized value θ_0 is independent of the coordinate system used to describe the null hypothesis value. Examples of invariant test statistics include the likelihood ratio statistic, $l(\hat{\theta}; z) - l(\theta_0; z)$, and the quadratic score statistic, $U_r U_s i^{rs}$, where i^{rs} is the matrix inverse of $-E\{U_{rs}; \theta\}$.

One of the main reasons for working with tensors, as opposed to arbitrary arrays of functions, is that it is easy to recognize and construct invariants. For example, any scalar derived from tensors by the process of contraction is automatically an invariant. This is a consequence of the tensor transformation properties (2.5) and (2.6). If the arrays are tensors only in some restricted sense, say under linear transformation only, then any derived scalars are invariants in the same restricted sense.

By way of illustration, consider the group of orthogonal transformations and suppose that the array ω^{ij} satisfies the transformation laws of a tensor under orthogonal transformation of X. Since we are dealing with orthogonal transformations only, and not arbitrary linear transformations, it follows that δ^{ij} and δ_{ij} are tensors. This follows from

$$a_i^r a_j^s \delta^{ij} = \delta^{rs}$$

where a_i^r is any orthogonal matrix. Note, however, that δ^{ijk} and δ_{ijk} are not tensors (Exercise 1.19). Hence the scalars

$$\omega^{ij}\delta_{ij}, \quad \omega^{ij}\omega^{kl}\delta_{ik}\delta_{jl}, \ldots$$

are invariants. In terms of the eigenvalues of ω^{ij}, these scalars may be written as power sums,

$$\sum_j \lambda_j, \quad \sum_j \lambda_j^2, \ldots.$$

Another function invariant under nonsingular linear transformation is the matrix rank, which may be written $\omega^{ij}\omega_{ij}$, where ω_{ij} is any generalized inverse of ω^{ij}. However not every invariant function can easily be derived by means of tensor-like manipulations of the type described here. For example, Sylvester's law of inertia states that the numbers of positive and negative eigenvalues of a real symmetric matrix are invariant under nonsingular linear transformation. In other words, if a_i^r is a nonsingular matrix then the sign pattern of the eigenvalues of ω^{ij} is the same as the sign pattern of the eigenvalues of $a_i^r a_j^s \omega^{ij}$. There appears to be no simple way to deduce this result from tensor-like manipulations alone. See, however, Exercise 1.9 where the signature is derived by introducing an auxiliary tensor ω_{ij}^+ and forming the invariant $\omega^{ij}\omega_{ij}^+$.

1.7 Direct product spaces

1.7.1 *Kronecker product*

Suppose that $Y^1, \ldots, Y^n$ are independent and identically distributed vector-valued random variables, each having p components. Should we so wish, we can regard $\mathbf{Y} = (Y^1, \ldots, Y^n)$ as a point in R^{np}, but usually it is preferable to consider $\mathbf{Y}$ explicitly as an element of the direct product space $R^n \times R^p$. One advantage of this construction is that we can then require derived statistics to be invariant under one group of transformations acting on R^n and to be tensors under a different group acting on R^p. For example, if $\kappa^{r,s}$ is the $p \times p$ covariance matrix of Y^1, then $\kappa^{r,s}$ is a tensor under the action of the affine group on R^p. Any estimate $k^{r,s}$, say, ought to have the same property. Furthermore, since the joint distribution of $Y^1, \ldots, Y^n$ is unaffected by permuting the n vectors, $k^{r,s}$ ought to be invariant under the action of the symmetric group (of permutations) on R^n.

Using tensor notation, the covariance matrix of $\mathbf{Y}$ may be written as $\delta^{ij}\kappa^{r,s}$, where indices i and j run from 1 to n while indices r and s run from 1 to p. Both δ^{ij} and $\kappa^{r,s}$ are tensors, but under different groups acting on different spaces. The Kronecker product, $\delta^{ij}\kappa^{r,s}$, is a tensor under the direct product group acting on $R^n \times R^p$.

More generally, if $\omega^{ij}\kappa^{r,s}$ is a tensor under the direct product of two groups acting on R^n and R^p, it may be necessary to compute the matrix inverse or generalized inverse in terms of ω_{ij} and $\kappa_{r,s}$, the generalized inverses on R^n and R^p respectively. It is immediately apparent that $\omega_{ij}\kappa_{r,s}$ is the required generalized inverse and that the inverse is covariant in the sense implied by the positions of the indices.

In matrix notation, the symbol $\otimes$ is usually employed to denote the Kronecker product. However, $\mathbf{A} \otimes \mathbf{B}$ is not the same matrix as $\mathbf{B} \otimes \mathbf{A}$ on account of the different arrangement of terms. No such difficulties arise with index notation because multiplication of real or complex numbers is a commutative operation.

Other kinds of direct products such as the Hadamard product (Rao, 1973, p. 30) do not arise in tensor calculations.

1.7.2 *Factorial design and Yates's algorithm*

Suppose that one observation is taken at each combination of the levels of factors A, B and C. Denote by Y^{ijk}, the yield or response recorded with A at level i, B at level j and C at level k. We make no claim that Y is, in any useful sense, a tensor. In fact, we occasionally write Y_{ijk} in place of Y^{ijk} where convenient to do so. This is the conventional notation for factorial models, where the indices are ordered according to the factors, and the array is not symmetric under index permutation. Further, unless the factors have equal numbers of levels, the indices have unequal ranges. This application is included here more to illustrate the value of index notation and the summation convention than as an example of a tensor.

Corresponding to each factor, we introduce *contrast matrices*, a_r^i, b_s^j, c_t^k, where the letters i,j,k refer to factor levels and the letters r,s,t refer to factor contrasts. The term 'contrast' is misused here because usually the first column of the contrast matrix is not a contrast at all, but a column of unit values. The remaining columns typically sum to zero and are therefore contrasts in the usual sense. The contrast matrices have full rank and are usually chosen to be orthogonal in the sense that

$$\begin{aligned} a_r^i a_{r'}^{i'} \delta_{ii'} &= a_{rr'} = 0 \quad \text{if } r \neq r' \\ b_s^j b_{s'}^{j'} \delta_{jj'} &= b_{ss'} = 0 \quad \text{if } s \neq s' \\ c_t^k c_{t'}^{k'} \delta_{kk'} &= c_{tt'} = 0 \quad \text{if } t \neq t' \end{aligned}$$

For example, if A has two levels, B has three ordered levels and C has four ordered levels, it is customary to make use of the orthogonal polynomial contrasts

$$a_r^i = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad b_s^j = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \qquad c_t^k = \begin{pmatrix} 1 & -3 & 1 & -1 \\ 1 & -1 & -1 & 3 \\ 1 & 1 & -1 & -3 \\ 1 & 3 & 1 & 1 \end{pmatrix}$$

Hence, the inner products give

$$a_{rr'} = \mathrm{diag}\{2, 2\}, \quad b_{ss'} = \mathrm{diag}\{3, 2, 6\}, \quad c_{tt'} = \mathrm{diag}\{4, 20, 4, 20\}.$$

By convention, contrast matrices are arranged so that rows refer to factor levels and columns to factor contrasts.

Whatever the contrasts chosen, the design matrix $\mathbf{X}$ corresponding to the factorial model $A{*}B{*}C$ is just the Kronecker product

$$x_{rst}^{ijk} = a_r^i b_s^j c_t^k.$$

In other symbols, the saturated factorial model is just

$$E(Y^{ijk}) = x_{rst}^{ijk} \beta^{rst} = a_r^i b_s^j c_t^k \beta^{rst},$$

where β^{rst} is the 'interaction' of contrast r of factor A with contrast s of factor B and contrast t of factor C. For instance, with the contrast matrices given above, β^{111} is just the mean and β^{132} is written conventionally as $B_Q C_L$. In other words, β^{132} is a measure of the change in the quadratic effect of B per unit increase in the level of C.

The so-called 'raw' or unstandardized contrasts that are produced by the three steps of Yates's algorithm, are given by

$$b_{rst} = a_r^i b_s^j c_t^k Y_{ijk}. \tag{1.16}$$

The linear combinations that are implied by the above expression are exactly those that arise when Yates's algorithm is performed in the conventional way.

To derive the least squares estimates of the parameters, we raise the indices of b, using the expression

$$\hat{\beta}^{rst} = a^{rr'} b^{ss'} c^{tt'} b_{r's't'}.$$

If the contrast matrices are each orthogonal, this expression reduces to

$$\hat{\beta}^{rst} = b_{rst}/(a_{rr} b_{ss} c_{tt}),$$

with variance $\sigma^2/(a_{rr} b_{ss} c_{tt})$, no summation intended.

The extension to an arbitrary number of factors, each having an arbitrary number of levels, is immediate. For further discussion, see Good (1958, 1960) or Takemura (1983).

From the numerical analytic point of view, the number of computations involved in (2.16) is considerably fewer than what would be required to solve n linear equations if the factorial structure of $\mathbf{X}$ were not utilized. For example, with k factors each at two levels, giving $n = 2^k$ observations, (2.16) requires nk additions and subtractions as opposed to $O(n^2)$ operations if the factorial structure were ignored. Herein lies the appeal of Yates's algorithm and also the fast Fourier transform, which uses the same device.

1.8 Bibliographic notes

Tensor notation is used widely in applied mathematics, mathematical physics and differential geometry. Definitions and notation vary to some extent with the context. For example, Jeffreys (1952) and Jeffreys & Jeffreys (1956) are concerned only with the effect on equations of motion of rotating the frame of reference or axes. Consequently, their definition of what they call a *Cartesian tensor* refers only to the orthogonal group and not, in general to arbitrary linear or non-linear transformation of coordinates. Their notation differs from that used here, most noticeably through the absence of superscripts.

Other useful references, again with a bias towards applications to physics, include McConnell (1931) and Lawden (1968). Thomas (1965, Section 6) emphasises the importance of transformation groups in the definition of a tensor.

For more recent work, again connected mainly with mathematical physics, see Richtmyer (1981).

In the theory of differential geometry, which is concerned with describing the local behaviour of curves and surfaces in space, notions of curvature and torsion are required that are independent of the choice of coordinate system on the surface. This requirement leads naturally to the notion of a tensor under the group of arbitrary invertible parameterizations of the surface. Gaussian and Riemannian curvature as well as mean curvature are invariants derived from such tensors. This work has a long history going back to Levi-Cività, Ricci, Riemann and Gauss's celebrated *theorem egregium*. Details can be found in the books by Eisenhart (1926), Weatherburn (1950), Sokolnikoff (1951) and Stoker (1969). For a more recent treatment of Riemannian geometry, see Willmore (1982) or Spivak (1970).

For a discussion of the geometry of generalized inverses, see Kruskal (1975) and the references therein.

The notion of a *spinor* is connected with rotations in Euclidean space and has applications in the theory of special relativity. Inevitably, there are strong similarities with quaternions, which are also useful for studying rotations. Tensors arise naturally in the study of such objects. See, for example, Cartan (1981).

Despite the widespread use in statistics of multiply-indexed arrays, for example, in the study of factorial, fractional factorial and other designs, explicit use of tensor methods is rare, at least up until the past few years. For an exception, see Takemura (1983). The reasons for this neglect are unclear: matrix notation abounds and is extraordinarily convenient provided that we do not venture far beyond linear models and second-moment assumptions. In any case, the defects and shortcomings of matrix notation become clearly apparent as soon as we depart from linear models or need to study moments beyond the second. For example, many of the quantities that arise in later chapters of this book cannot be expressed using matrix notation. This is the realm of the tensor, and it is our aim in the remainder of this book to demonstrate to the reader that great simplification can result from judicious choice of notation.

1.9 Further results and exercises 1

1.1 Derive the transformation laws (2.7) for log likelihood derivatives.

1.2 Show that if ω^{ijk} is a contravariant tensor and ω_{ijk} is a covariant tensor, then $\omega^{ijk}\omega_{ijk}$ is an invariant.

1.3 Show directly, using the notation in (2.7), that $U_r U_s I^{rs}$ is invariant under the group of invertible transformations acting on the parameter space.

1.4 Let $i_{rs} = -E\{U_{rs};\theta\}$ and let i^{rs} be the matrix inverse. Under which group of transformations is $U_r U_s i^{rs}$ an invariant?

1.5 Show, using the notation in (2.7), that

$$V_{ij} = U_{ij} - \kappa_{ij,k} i^{kl} U_l$$

is a covariant tensor, where $\kappa_{rs,t}$ is the covariance of U_{rs} and U_t.

1.6 Let a_j^i be the elements of a square matrix, not necessarily symmetrical, and let its inverse, b_i^j satisfy $a_j^i b_k^j = \delta_k^i = a_k^j b_j^i$. Show that the derivatives satisfy

$$\frac{\partial b_i^j}{\partial a_s^r} = -b_r^j b_i^s$$
$$\frac{\partial a_i^j}{\partial b_s^r} = -a_r^j a_i^s$$

1.7 Show that the spectral decomposition of the the symmetric matrix $\mathbf{A}$

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T, \quad \mathbf{Q}\mathbf{Q}^T = \mathbf{I}, \quad \Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$$

is unique up to permutations of the columns of $\mathbf{Q}$ and the elements of Λ if the eigenvalues of $\mathbf{A}$ are distinct.

1.8 Show that there exists a linear transformation $Y^r = a_i^r X^i$ from X to Y such that the quadratic form $\omega_{ij} X^i X^j$ may be written as $\epsilon_{rs} Y^r Y^s$ where $\epsilon_{ii} = \pm 1$ and $\epsilon_{ij} = 0$ if $i \neq j$.

1.9 Consider the decomposition $\omega_{ij} = \tau_i^r \tau_j^s \epsilon_{rs}$ of the symmetric covariant tensor ω_{ij}, where the notation is that used in (2.14). Define

$$\omega_{ij}^{+} = \tau_i^r \tau_j^s |\epsilon_{rs}|.$$

Show that ω_{ij}^{+} is a covariant tensor and that the scalar

$$s = \omega^{ij} \omega_{ij}^{+}$$

is independent of the choice of τ_i^r and also of the choice of generalized inverse ω^{ij}. Show that s is the signature of ω_{ij}.

1.10 In the notation of the previous exercise, let

$$\omega^{ij} = \gamma_r^i \gamma_s^j \epsilon^{rs},$$

where $\epsilon^{rs} = \epsilon_{rs}$ and $\gamma_r^i \tau_j^r = \delta_j^i$. Show that ω^{ij} is the Moore-Penrose inverse of ω_{ij}.

1.11 Show that the identity matrix is a generalized inverse of any projection matrix. Show also that a projection matrix, not necessarily symmetric, is its own generalized inverse satisfying (2.9) and (2.11). Under what conditions is a projection matrix self-inverse in the Moore-Penrose sense?

1.12 Let ω^{ij}, with inverse ω_{ij}, be the components of a $p \times p$ symmetric matrix of rank p. Show that

$$\gamma^{rs,ij} = \omega^{ri}\omega^{sj} + \omega^{rj}\omega^{si},$$

regarded as a $p^2 \times p^2$ matrix with rows indexed by (r, s) and columns by (i, j), is symmetric with rank $p(p + 1)/2$. Show also that $\omega_{ri}\omega_{sj}/2$ is a generalized inverse. Find the Moore-Penrose generalized inverse.

1.13 Consider the linear mapping from R^p to itself given by

$$\bar{X}^r = \omega_i^r X^i$$

where ω_i^r is nonsingular. Show that, under simultaneous change of coordinates

$$Y^r = a_i^r X^i, \quad \bar{Y}^r = a_i^r \bar{X}^i,$$

ω_i^r transforms as a mixed tensor. By comparing the volume of a set, B say, in the X coordinate system with the volume of the transformed set, $\bar{B}$, interpret the determinant of ω_i^r as an invariant. Give similar interpretations of the remaining $p-1$ invariants, e.g. in terms of surface area and so on.

1.14 Let $\pi_1, \ldots, \pi_k$ be positive numbers adding to unity and define the multinomial covariance matrix

$$\omega_{ij} = \begin{cases} \pi_i(1-\pi_i) & i = j \\ -\pi_i\pi_j & i \neq j \end{cases}.$$

Show that ω_{ij} has rank $k-1$ and that

$$\omega^{ij} = \begin{cases} 1/\pi_i & i = j \\ 0 & \text{otherwise} \end{cases}$$

is a generalized inverse of rank k. Find the Moore-Penrose generalized inverse.

1.15 Let the $p \times q$ matrix $\mathbf{A}$, with components a_i^r, be considered as defining a linear transformation from the domain, R^q, to the range in R^p. Interpret the *singular values* of $\mathbf{A}$ as invariants under independent orthogonal transformation of the domain and range spaces. For the definition of singular values and their application in numerical linear algebra, see Chambers (1977, Section 5.e).

1.16 Let $\mathbf{A}$, $\mathbf{A}^{-1}$ and $\mathbf{X}$ be symmetric matrices with components a_{ij}, a^{ij} and x_{ij} respectively. Show that the Taylor expansion for the log determinant of $\mathbf{A}+\mathbf{X}$ about the origin may be written

$$\begin{aligned}\log\det(\mathbf{A}+\mathbf{X}) = \log\det(\mathbf{A}) &+ x_{ij}a^{ij} - x_{ij}x_{kl}a^{ik}a^{jl}/2 \\ &+ x_{ij}x_{kl}x_{mn}a^{ik}a^{jm}a^{ln}/3 + \cdots\end{aligned}$$

Describe the form of the general term in this expansion. Compare with Exercise 1.6 and generalize to asymmetric matrices.

1.17 Justify the claim that Kronecker's delta, δ_i^j, is a tensor.

1.18 Show that δ_{ij} is a tensor under the orthogonal group but not under any larger group.

1.19 Show that δ_{ijk}, $\delta_{ijkl}, \ldots$ are tensors under the *symmetric* group but not under any larger group.
[The symmetric group, which is most conveniently represented by the set of permutation matrices, arises naturally in the study of sample moments based on simple random samples, where the order in which the observations are recorded is assumed to be irrelevant. See Chapter 4.]

1.20 Show that the symmetric group is a subgroup of the orthogonal group, which, in turn, is a subgroup of the general linear group.

1.21 *Hooke's Law*: In the mechanics of deformable solids, the components of the *stress tensor*, p_{ij}, measure force per unit area in the following sense. Let $\mathbf{e}_i$ be the unit vector in the ith coordinate direction and let $\bar{\mathbf{e}}_i$ be the orthogonal plane. Then p_{ii} is the force per unit area normal to the plane, also called the *normal stress*, and p_{ij}, $j \neq i$ are the *shear stresses* acting in the plane. The components of the *strain tensor*, q_{ij}, which are dimensionless, measure percentage deformation or percentage change in length. Both arrays are symmetric tensors under the orthogonal group.

In the case of *elastic* deformation of an *isotropic* material, Hooke's law in its most general form states that the relationship between stress and strain is linear. Thus,

$$p_{rs} = b_{rs}^{ij} q_{ij},$$

where b_{rs}^{ij} is an *isotropic* fourth-order tensor given by

$$b_{rs}^{ij} = \lambda \delta^{ij} \delta_{rs} + 2\mu \delta_r^i \delta_s^j,$$

for constants λ, μ that are characteristic of the material.

Show that the inverse relationship giving the strains in terms of the stresses may be written in the form

$$q_{ij} = \left(\lambda' \delta^{rs} \delta_{ij} + 2\mu' \delta_i^r \delta_j^s\right) p_{rs},$$

where the new constants are given by

$$\mu' = \frac{1}{4\mu}, \qquad \lambda' + 2\mu' = \frac{\lambda + \mu}{\mu(3\lambda + 2\mu)} = E^{-1}.$$

In the terminology used in Mechanics, E is known as *Young's modulus* or *modulus of elasticity*, μ is called the *rigidity* or *shear modulus* and $\sigma = \lambda/\{2(\lambda + \mu)\}$ is *Poisson's ratio*. Note that $E = 2(1 + \sigma)\mu$, implying that two independent constants entirely determine the three-dimensional elastic properties of the material. (Murnaghan, 1951, Chapters 3,4; Jeffreys & Jeffreys 1956, Section 3.10; Drucker, 1967, Chapter 12).

1.22 The array ϵ_{ijk} of order $3 \times 3 \times 3$ defined by

$$\begin{aligned} \epsilon_{123} &= \epsilon_{231} = \epsilon_{312} = 1 \\ \epsilon_{213} &= \epsilon_{132} = \epsilon_{321} = -1 \\ \epsilon_{ijk} &= 0 \quad \text{otherwise,} \end{aligned}$$

is known as the *alternating tensor* (Ames & Murnaghan, 1929, p. 440). For any 3×3 matrix a_r^i, show that

$$\epsilon_{ijk} a_r^i a_s^j a_t^k = \epsilon_{rst} \det(\mathbf{A}).$$

Hence show that ϵ_{ijk} is an isotropic tensor under O^+, the orthogonal group with positive determinant (Jeffreys & Jeffreys, 1956, Sections 2.07, 3.03).

Write down the generalization of the alternating tensor appropriate for a $p \times p \times p$ array.