CHAPTER 1

Introduction

This monograph deals with problems concerning distributions in statistical models in which there is a group of invariance transformations. The methods to be presented make use of mathematical tools that involve interplay between groups and integration. The purpose of this monograph is not only to demonstrate by examples the statistical usefulness of the methods, but also to present a systematic account of the mathematical background.

One of the most important problems in a statistical model with an invariance group is to obtain the distribution of a maximal invariant (precise definitions of this and other notions will be made in subsequent chapters). However, other statistics that are not invariant but equivariant may be of interest too (for instance, maximum likelihood estimators). Often these distributional problems can be handled without group methods on an ad hoc basis, but the full use of a known invariant measure on the group provides a more systematic treatment that usually also is easier to carry out. Invariant measures are guaranteed for locally compact groups, and it is to those groups that this monograph will be confined. That includes all such "nice" groups as translations and matrix groups, but not large groups such as all continuous and strictly monotonic transformations of the real line. Thus, the statistical problems to be considered are typically of the parametric rather than nonparametric type.

As a justification for the use of group methods in a simple example consider the derivation of the χ^2-distribution, or equivalently, the χ-distribution. Let $X = (X_1, \dots, X_n)$ be a random vector in n-dimensional Euclidean space and put $R^2 = \sum_1^n X_i^2$. In the simplest case, leading to the central χ^2-distribution, the X_i are independent and identically distributed (iid) $N(0,1)$ so that the distribution of X is

$$(2\pi)^{-\frac{n}{2}} \exp\left(-\tfrac{1}{2}r^2\right) \lambda(dx),$$

in which $r^2 = \sum_1^n x_i^2$ and $\lambda(dx) = dx_1 \dots dx_n$ is n-dimensional Lebesgue measure. At first glance, sophisticated methods for deriving the distribution of R^2 would seem totally out of place. Indeed, introduce polar coordinates with radial coordinate r and a set of $n-1$ additional angular coordinates, for brevity denoted by the single symbol θ and its random version by Θ. Then X is in 1-1 correspondence with (R, Θ) and $\lambda(dx) = r^{n-1} dr \mu(d\theta)$ for some measure μ that does not depend on r. Now integrate over θ to obtain the distribution of R. Since the integrand does not depend on θ, the result of the integration is a constant A_n, which is the area of the unit sphere $\sum_1^n x_i^2 = 1$ (the value of A_n is listed in (7.7.9)). Hence, the distribution of R is

$$(2\pi)^{-\frac{n}{2}} A_n \exp\left(-\tfrac{1}{2}r^2\right) r^{n-1} dr,$$

which is the familiar χ-distribution with n degrees of freedom. Note that in this derivation it is not necessary to make an explicit choice for θ.

The limitation of the above elementary derivation is evident when the density of X with respect to λ depends not only on r but also on θ. For instance, the X_i could be as before except that not all means are 0 (so that R^2 is noncentral χ^2). More generally, X could have an arbitrary density with respect to λ that fails to be spherically symmetric. In order to perform the integration over θ it would then be necessary to make a specific choice for θ and to compute the Jacobian of the transformation from x to (r, θ). Any such choice is highly arbitrary, and, furthermore, the Jacobian computation tends to be rather messy. It is here that group methods offer a more systematic and convenient—therefore more attractive—approach. The relevant group G in the present problem is $O(n)$, the group of all $n \times n$ orthogonal matrices. Now regard X as an $n \times 1$ column vector and consider the transformation $X \to \Gamma X$ with an arbitrary $\Gamma \in G$. This defines a so-called **action** of G on n-space. Since $R^2 = X'X$, R^2 is invariant under this transformation for every $\Gamma \in G$, and so is R of course; even maximal invariant, for it is easy to show that any invariant function of X depends on X only through R. Define the function t on n-space into the nonnegative half-line by $t(x) = r$, so that $t(X) = R$, and suppose that the distribution of X is $p(x)\lambda(dx)$ with some probability density p. Then the methods in this monograph will show that the density of R at r is $\int p(\Gamma x)\mu_G(d\Gamma)$ with respect to an explicitly described measure $\nu(dr)$, where x is any point for which $t(x) = r$, and μ_G is the unique invariant probability measure on G (invariant here means that the measure of a set does not change if the set is orthogonally transformed).

The above example will now be generalized to X being an $n \times p$ random matrix. Put $S = X'X$ ($'$ denotes transpose), then S has the standard Wishart distribution with n degrees of freedom if the elements of X are iid $N(0,1)$. When $p = 1$, $S = R^2$ of the previous example and the χ^2-distribution re-emerges. There are many derivations of the Wishart distribution that do not use group methods; see, e.g., Anderson (1984), Section 7.2. None of these derivations is elementary, unlike the elementary derivation of the χ^2-distribution when $p = 1$. However, making full use of group methods provides a very attractive alternative derivation. Furthermore, the method does not depend on the particular density of X. The starting point is a Gram-Schmidt decomposition $X = UT$, in which $U : n \times p$ has orthonormal columns and T is $p \times p$ upper triangular with positive diagonal elements. Then $S = T'T$ (Cholesky decomposition) and S is in 1-1 correspondence with T so that the distribution of T is just as useful as that of S. The invariance group G in this example is again $O(n)$ with action $X \to \Gamma X$, $\Gamma \in G$. Obviously, S remains invariant under this action (in fact, maximal invariant). The group methods in this monograph lead to factorization of Lebesgue measure λ on

np-dimensional space into a product measure on the product space in which (U, T) takes its values. Then any density of X with respect to λ translates into a joint density of (U, T) with respect to the product measure. In particular, in the standard Wishart case, U and T turn out to be independent, leading immediately to the marginal distributions of U and T (therefore, of S). For arbitrary density of X, an integral expression for the density of T is obtained; this is useful, for instance, for the noncentral Wishart distribution. The details are in Section 9.2. Incidentally, U above is a so-called *equivariant* statistic: when X is changed to ΓX, U transforms into ΓU; see Example 2.1.13.

If the group G is compact, as is the case with $O(n)$ in the previous examples, then it is often possible to obtain a partial result with help of more modest means. Suppose a random variable X takes its values in $\mathfrak{X}$ and has a density $p(x)$ with respect to a measure $\lambda(dx)$ that is invariant under the transformations of a compact group G that acts on $\mathfrak{X}$ (e.g., Lebesgue measure is invariant under orthogonal transformations). Let $t : \mathfrak{X} \to \mathfrak{Y}$ be a maximal invariant function and let $\nu = \lambda t^{-1}$ be the measure that λ induces on $\mathfrak{Y}$ via the function t. Then the density of $T = t(X)$ with respect to ν at $y \in \mathfrak{Y}$ is $\int p(gx)\mu_G(dg)$, where x is any point in $\mathfrak{X}$ for which $t(x) = y$, and μ_G is the unique invariant probability measure on G. This follows from elementary integration properties and is proved in Eaton (1983), Proposition 7.15. However, this relatively easy result obscures the fact that it does not come to grips with the measure ν, except for its definition. In contrast, the method of global cross section employed in this monograph enables a more complete answer, including an explicit expression for the measure ν. In problems involving density ratios, the explicit knowledge of ν may not be needed and then the above simple derivation is useful. This will be elaborated in Chapter 13.

When the group G is not compact several complications arise. First, the invariant measure on G is no longer finite and therefore cannot be chosen to be a probability measure. Consequently, the proof alluded to in the previous paragraph does not go through. Second, there is a priori no guarantee that the induced measure ν is really a measure, i.e., finite on compacta, nor that $\int p(gx)\mu_G(dg)$ (or whatever takes its place) is finite. Third, and perhaps most importantly, the action of G on $\mathcal{X}$ may be bad even though both G and $\mathcal{X}$ are very regular. In particular, the orbit space may fail to be a Hausdorff space. No specific statistical example of this phenomenon is known to this writer, but in mathematics the classical example of such unpleasant action is the **irrational flow on the torus** (see, e.g., Hewitt and Ross, 1979, Section 6.17, Example (e); or Greup, Halperin, and Vanstone, 1972, Section 3.10). The two-dimensional torus can be pictured as a ring (the geometric, not algebraic, object) in three-dimensional space, but is analytically more conveniently represented by a unit square $0 \leq x, y \leq 1$ in which the parallel edges are identified; i.e., the points $(0, y)$ and $(1, y)$ coincide for all $0 \leq y \leq 1$, and so do the points $(x, 0)$ and $(x, 1)$ for all $0 \leq x \leq 1$. The group G is the real line R under addition. Take any irrational number α, then the action of G on the torus is defined by

$$(x, y) \to (x + t \pmod 1,\ y + \alpha t \pmod 1), \quad t \in R.$$

The **orbit** of the point (x, y) results if t runs from $-\infty$ to ∞. Then every orbit is dense in the torus and it is not possible to enclose two distinct orbits in disjoint open invariant neighborhoods.

The above discussion points to the necessity of additional assumptions on the action of G. This has led Bourbaki (1966b) to introduce the notion of a **proper action**. For locally compact group and space this amounts to the following: if A and B are compact subsets of $\mathcal{X}$, then the set of $g \in G$ for which gA and B have nonempty intersection should have compact closure. This rules out, for instance, the irrational flow on the torus since any orbit starting from a point in A passes through B for arbitrarily large values of t.

In Chapters 8 through 11 where useful results are derived we shall not explicitly assume the action of G on $\mathcal{X}$ to be proper. Instead, stronger assumptions will be made that among other things imply properness of the action. The reason for imposing stronger conditions is to be able to achieve as complete a result as possible. Our method endeavors to find a maximal invariant Z and an equivariant Y such that there is a 1-1 correspondence between X and (Y, Z). Let (Y, Z) take its values in $\mathcal{Y} \times \mathcal{Z}$. Then under some conditions (to be specified in Chapter 8) a measure λ on $\mathcal{X}$ transforms into a product measure, say $\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Z}}$, on $\mathcal{Y} \times \mathcal{Z}$. Here $\mu_{\mathcal{Y}}$ can be obtained directly from the invariant measure on G (Sections 7.3 and 7.4); the question how to obtain $\mu_{\mathcal{Z}}$ will be dealt with in Chapter 8. If p is the density of X with respect to λ and if $p(x) = p_0(y, z)$, then p_0 is the density of (Y, Z) with respect to $\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Z}}$. Then the marginal distributions of Y and Z may be obtained by integration over the other variable. If p_0 is of the special form $p_0(y, z) = p_1(y)p_2(z)$, then Y and Z are seen to be independent with distributions $p_1(y)\mu_{\mathcal{Y}}(dy)$ and $p_2(z)\mu_{\mathcal{Z}}(dz)$, respectively. This is the case in the Wishart example if the rows of X are iid $N(0, \Sigma)$, where Y and Z are U and T, respectively, in the example.

It should be clear from the above discussion that the most essential step in the derivation of distributions that result from the action of a group G consists in relating a given measure λ on the sample space $\mathcal{X}$ to a product measure on a product space $\mathcal{Y} \times \mathcal{Z}$ in which $\mathcal{Z}$ is any space that is in one-one correspondence with the orbits in $\mathcal{X}$. The other space $\mathcal{Y}$ could be G itself, or some other space on which G acts transitively. The two spaces, $\mathcal{X}$ and $\mathcal{Y} \times \mathcal{Z}$, may but need not be in 1-1 correspondence. This point will be discussed in Chapter 12. But in any case the process of deriving the product measure $\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Z}}$ on $\mathcal{Y} \times \mathcal{Z}$ from the measure λ on $\mathcal{X}$ will be called loosely **factorization** of λ. Examples of such factorizations were first given by Stein (1956a, b). More systematic treatments can be found in Schwartz (1966), Wijsman (1967b), Koehn (1970), Bondar (1976), Farrell (1976, 1985), Woteki and Mayer (1976), Andersson (1982), Andersson, Brøns, and Jensen (1983), Eaton (1983), and Wijsman (1986). In Eaton (1989) there is an expository account based on Andersson (1982). In some of the above references the space $\mathcal{Z}$ is taken to be a **global cross section**, often simply called a **cross section**, i.e., a subset of $\mathcal{X}$ that meets each orbit in exactly one point. These various treatments differ in their underlying assumptions. For instance, the cross section $\mathcal{Z}$ is assumed to be differentiable in Wijsman (1967b, 1986), but only measurable in Schwartz (1966), Bondar (1976), Farrell (1976, 1985). A more detailed account may be found in Wijsman (1986), Section 2. Of course, as is to be expected, stronger conclusions can be drawn from stronger assumptions. Thus, the existence of a differentiable cross section leads to an explicit expression for $\mu_{\mathcal{Z}}$ rather than merely its existence. Therefore, in this monograph we shall pursue the method of differentiable cross section since it is capable of providing more complete factorization results. Each such factorization leads to a joint distribution of (Y, Z) but generally study of this joint distribution has not been pursued in Chapters 9–11. The emphasis will be on the marginal distribution of Z, expressed as an integral. There are also applications, not included in this monograph, of cross sections to statistical problems other than factorization of a measure. The reader is referred to Wijsman (1986), Section 4.

The methods of Andersson (1982) and Andersson, Brøns, and Jensen (1983) do not involve cross sections but, instead, assume the action of G to be proper. Within the use of proper action Andersson (1982) bases his treatment on Bourbaki's *quotient measure* whereas Andersson, Brøns, and Jensen (1983) rely on Bourbaki's theory of integration as applied to product measure. These notions will be discussed in Section 7.3 and 6.5, respectively. We shall not use the approach of these authors, but a comparison of their assumptions and results with ours will be made in Chapter 12.

The basic method for factoring a measure, to be developed in Chapter 8, involves computing a Jacobian at each point of a cross section. This can be avoided if the model has additional structure, consisting of the presence of another group H such that GH is a group that acts transitively on $\mathcal{X}$ and such that a few additional requirements are satisfied. This will be called **special group structure**. If it is present, then any H-orbit can be taken as a cross section and the product measure $\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Z}}$ follows directly from the Haar measures on G and H without Jacobian computation, except for a positive multiplicative constant. The latter can be found by computing a Jacobian at one point or by integrating an "easy" density.

The conditions for special group structure and the theorems that can be derived from it are treated in Chapter 8. The statistical problems to which special group structure applies will be called **Type I**, and examples are given in Chapter 9. In contrast, **Type II** problems have no special group structure but there is a differentiable cross section so that the basic factorization method applies. Examples appear in Chapter 10. Finally, in **Type III** problems there is an extra group H such that GH is a group acting transitively on $\mathfrak{X}$, but not all other conditions for special group structure are met and consequently an H-orbit is no longer a cross section. However, under certain conditions an H-orbit is a **global slice** and can still be useful in reducing the original problem to a smaller one. This is presented in Chapter 11. In several of the Type I and Type II problems it is shown how to obtain a solution in steps by applying in succession the solutions of simpler problems that have been solved already.

In some statistical problems what is needed is not the distribution of a maximal invariant, but rather the density ratio of two such distributions. This is easier to obtain than the distributions themselves since in the product measure $\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Z}}$ it is not necessary to know an expression for $\mu_{\mathcal{Z}}$, only its existence. This kind of problem can be handled by **local cross sections**, or by using quotient measure, and will be treated in Chapter 13.

The purpose of this monograph is mainly twofold. First, to present an exposition of the method of differentiable cross section combined with invariant measures on groups and enough examples to demonstrate its usefulness; second, to provide the mathematical background behind the method. In addition, there is in Chapter 12 a comparison between the cross section method and the factorization method of Andersson, Brøns, and Jensen (1983). This made it necessary to include in Chapters 2 and 7 material on proper action. There are other reasons for this inclusion as well. For instance, proper action is a natural requirement for the construction of an invariant or relatively invariant measure on a homogeneous space (Sections 7.3, 7.4). From the point of view of statistical application, the most important chapters are 8–10 and 13. Some readers may decide to start with Chapter 8 and go back to earlier chapters only as the need arises.

A major portion of this monograph, Chapters 2–7, is occupied by the mathematical background. Even though most of this material is available in the mathematical literature, it is scattered throughout several sources. It seemed to this writer that it would be useful to have in one place those mathematical concepts and theorems that are pertinent to groups acting on spaces and measures on these groups and spaces, especially measures that enjoy some sort of invariance under the group action. The material in Chapters 2–7 is built up in a more or less logical way. First some generalities about topological groups and group action in Chapter 2. Then an introduction to differentiable manifolds in Chapters 3 and 4, concentrating on tangent spaces, differential of a mapping, vector fields, and differential forms. This is followed in Chapter 5 by the special but important case of Lie groups, which makes it possible to define invariant vector fields (Lie algebra) and invariant differential forms. The latter (those of highest degree) are going to be used as invariant measures on groups, but before that can be defined, a certain amount of Bourbaki-type integration theory has to be developed (Chapter 6). The various concepts come together in Chapter 7, with the construction of invariant and relatively invariant measures on locally compact groups and coset spaces. Some or all of these chapters could be read for their own sake as well, rather than as background for Chapters 8–13. For instance, Chapter 3 could be read as an introduction to differential geometry in statistical parameter spaces, as treated, e.g., in Amari, Barndorff-Nielsen, Kass, Lauritzen, and Rao (1987) (where references to other relevant publications can be found). Chapter 5 may be useful as an introduction to Lie groups; the concrete examples that appear in that chapter may be helpful to the understanding of notions such as invariant vector fields, etc.. The introduction to the Bourbaki theory of integration in Chapter 6 contains some comparison with the classical and Daniell integration theories, which may be of interest. Not all of the contents of Chapter 2–7 can be found in the literature. At least to the best of this writer's knowledge, the material in Section 5.9 and Proposition 7.7.6 are new.

For the most part, definitions and results are stated carefully. Many of the proofs are included, but for some proofs (especially the longer ones) the reader is referred to the literature. As a result, this monograph is self-contained to a considerable extent, but not completely. For those readers who would like more information on some of the topics touched upon in Chapters 2–7 here follows a short list of books that have been especially helpful to this writer. For general topology: Kelley (1955) and Bourbaki (1966b). The latter is an excellent English translation of the original French edition. It is unequaled in its treatment of groups acting on spaces, including proper action. The basics of invariant and relatively invariant measures on locally compact groups can be found in Bourbaki (1963), and the foundations of integration on locally compact spaces in Bourbaki (1965). Further integration theory appears in Bourbaki (1967). None of these Bourbaki volumes have been translated into English; however, even readers with only elementary knowledge of French will have no difficulty with it. For classical measure and integration: Halmos (1950). For Lie groups: Cohn (1957) and Chevalley (1946), the former being the more elementary of the two. For Haar measure on locally compact groups: Nachbin (1976), and for the Daniell integral (which is essentially Bourbaki's method): Taylor (1965, 1985). Finally, for manifolds, tangent spaces, differential of a mapping, etc.: Bishop and Crittenden (1964).

To conclude this chapter, we indicate here briefly how this monograph relates to several well-known books that deal with the concept of invariance in statistics. Lehmann (1959, 1986, Chapter 6) treats a general theory of invariance in hypothesis testing, first developed by G. Hunt and C. Stein (unpublished). This deals with the derivation and properties of invariant tests, not with measures on the invariance groups. The same can be said about the treatment of equivariant estimators in Ferguson (1967, Chapter 4) and Lehmann (1983, Chapter 3). There is some overlap of this monograph with that of Barndorff-Nielsen, Blæsild, and Eriksen (1989). The latter has many examples of invariant or relatively invariant measures, and of factorization of measures. Our monograph is also in spirit close to Farrell (1976, 1985), Muirhead (1982), and Eaton (1983, 1989), all of which deal with invariant measures on groups and related manifolds. In contrast, there is no overlap to speak of with the monograph of Diaconis (1988), which discusses applications of group representation in statistics, especially of the permutation group on n symbols.