STATISTICAL MODELS AND INVARIANCE

By D. A. S. Fraser

University of Toronto

0. Summary. Brillinger [2] gives necessary and sufficient conditions for a model to be invariant under a Lie group of transformations. The problems that can be handled by his conditions are surveyed, and found effectively to be restricted to one-dimensional problems amendable to Lindley's [8] method and to problems connected with conflicts between Bayes' and fiducial theory.

The problem of finding the general model invariant under a given group is proposed. Brillinger's theorem produces differential equations for the model. A general solution can be obtained by direct methods.

1. Introduction. Statistical models invariant under a transformation group enter prominently into many areas of statistical theory: in hypothesis testing and decision theory, Lehmann [7] and Blackwell and Girshick [2]; in estimation theory, Pitman [9]; in fiducial theory, Fraser [3] and Hora and Buehler [6]; in Bayesian theory, Stone [10]; in structural theory, Fraser [5]. The problem of determining whether a model is invariant under a transformation group seems then of general importance. Brillinger [2] examines this problem and notes that it has received negligible attention in the literature. He then develops necessary and sufficient conditions for invariance under *some* Lie group of transformations.

Brillinger's necessary condition can be applied with a *given* Lie group of transformations on sample space and on parameter space. For this it operates typically by contradiction if the condition is not fulfilled then the model is not invariant with respect to the given Lie group.

Brillinger expresses the necessary condition in terms of differential operators, the infinitesimal generators that can generate a Lie transformation group. The use of differential operators in this context is notationally rather elaborate. As a simplification the condition can be expressed in terms of Jacobian matrices, the basic ingredient of any differential analysis of transformations. In this alternative form the condition can be applied more generally to a differentiable class of transformations that need not be a group. And in this alternative form if used with a *given* connected differentiable class having an identity transformation, it becomes both necessary and sufficient. See Section 2.

In the special case that the class of transformations is a group, the Jacobian matrices in the condition simplify and are independent of the parameters of the group. See Section 3.

Brillinger's necessary condition can also be used to investigate whether a given model is invariant with respect to *some* Lie group of transformations. For this also, it operates by contradiction: if the model is such that the condition cannot

Received 28 November 1966.

be fulfilled using matrices of the simplified form involving no parameter, then the model is not invariant with respect to any Lie group of transformations. Brillinger's Examples 2, 3 illustrate this kind of application.

Brillinger's sufficient condition for a model to be invariant under some Lie group is of the following two-fold nature: the necessary condition must be satisfied by some matrices of the simplified form involving no parameter; and these matrices must integrate to form a Lie group on the sample space and on the parameter space. Brillinger notes that the sufficient condition is somewhat tautological. In effect it says the model is invariant with respect to *some* Lie group if a Lie group can be found satisfying the earlier mentioned necessary condition (group connectedness needed).

The essential content of Brillinger's theorem thus relates to a *given* group of transformations. It is given in this restricted form in Section 3, only rephrased in terms of matrices. The tautological nature of the sufficient condition is then avoided. And the range of applications is in no way restricted—rather the exploratory aspect of finding matrices of the simplified form is removed from the theorem and attached to the application, where in essence it belongs.

There is however one context in which the general version of the theorem has additional substance. This context involves a Lie group on a one-dimensional sample space and one-dimensional parameter space. Group theory shows that such a group produces translations on the two spaces appropriately expressed. Brillinger isolates this case as a corollary to the theorem. He notes that the corollary is inherent in Lindley [8]. The corollary, however, can be proved directly with elementary calculus: see Lindley [8] and Fraser [4].

Brillinger gives one example to illustrate his sufficiency condition; it is however an example with the one-dimensional form covered by the corollary. Let $F(x \mid \theta)$ be the Pareto distribution $F(x \mid \theta) = 1 - x^{-\theta}$ with $x \geqq 1, \theta > 0$. The distribution satisfies the condition in the corollary:

$$F_x/F_\theta = \theta/x \ln x$$

a function of x times a function of θ. Brillinger uses infinitesimal operators to generate the group, but the elementary calculus in Lindley [8] as abstracted in Fraser [4] gives the group directly: for fixed F the condition becomes

$$\partial\theta/\partial x = \theta/x \ln x, \qquad \ln x = C\theta;$$

accordingly, the distribution function can be written

$$F(x \mid \theta) = G(\ln x/\theta),$$

giving the multiplicative group on $\ln x$ and on θ.

Brillinger's material has been surveyed here with intent to delineate the kinds of problem that can be handled by his theorem. These are indicated by the following:

(i) a criterion by which some models can be shown *not* to be invariant with respect to any group.

(ii) a method for finding a group with respect to which a given model is invariant.

(iii) the one-dimensional problems covered by Lindley's result.

Brillinger's examples illustrate the first and the third. The second, if not of the one-dimensional form covered by the third, seems relatively inaccessible.

The first and second kinds of problem are concerned with whether a given model is invariant with respect to *some* group—or freely, can a model be structured by a group of transformations? Brillinger expresses interest in this as it bears on the relationship between fiducial theory and Bayes' theory, in effect on the pathologies existing in an area of overlap of two theories. But is there more positive interest in the question of invariance with respect to *some* Lie group? I feel that the question is in the wrong direction. Rather, in many applications the transformation structure is primary and the model derives from it; for example, in regression analysis a response vector is obtained from a vector of errors by relocation with respect to given structural vectors. This alternative view is basic to structural theory: *the transformations are primary and of physical significance, the frequency model* $f(x \mid \theta)$ *is derivative.*

It is natural then to draw back to the question of invariance with respect to a *given* group of transformations; in fact, to questions directly treated by the alternative form of Brillinger's theorem as given in Section 3;

First: Is a model $f(x \mid \theta)$ invariant with respect to a group G? In essence, this is not a differential question. For example, if $\mathbf{y}$ is multivariate normal with unrestricted mean and covariance matrix and g is the affine group, then one can check directly: $g\mathbf{y}$ is multivariate normal; does it have parameter given by $g\theta$? In unusual circumstances one might check by differential methods in the same way that one might investigate whether $h(x) = h(0)$ by checking whether $h'(x) = 0$ along connected paths.

Second: What is the general form of a model invariant under a specified group? The alternative form of Brillinger's theorem seems directly applicable: differential equations can be obtained and, subject to difficulties of integration, the general model can be found by solution. But again the question is essentially not a differential question. In fact, the general solution can be obtained explicitly by direct analysis. This is examined in Section 4.

2. Invariance and a class of transformations. Let $\mathbf{x}$ be a vector variable with open sample space $X \subset R^p$, $\boldsymbol{\theta}$ be a vector parameter with open parameter space $\Omega \subset R^k$, and $f(\mathbf{x} \mid \boldsymbol{\theta})$ be a statistical model giving probability density with respect to Euclidean volume. Let $T = \{(g)\}$ be a class of transformations applicable on X and on Ω with indexing parameter $\mathbf{g}$ taking values in an open arc-wise connected space $G \subset R^r$. And suppose the maps $\mathbf{x} \rightarrow (\mathbf{g})\mathbf{x}$ and $\boldsymbol{\theta} \rightarrow (\mathbf{g})\boldsymbol{\theta}$ are 1–1 with non-singular Jacobian matrices

$$\partial(\mathbf{g})\mathbf{x}/\partial\mathbf{x}', \qquad \partial(\mathbf{g})\boldsymbol{\theta}/\partial\boldsymbol{\theta}'$$

that have derivatives with respect to $\mathbf{g}$ continuous in $\mathbf{x}$ and $\mathbf{g}$.

A statistical model is invariant with respect to the class of transformations T if a variable $\mathbf{x}$ with distribution $\boldsymbol{\theta}$ transforms to a variable $(\mathbf{g})\mathbf{x}$ with distribution $(\mathbf{g})\boldsymbol{\theta}$ for all $\mathbf{g}$ and $\boldsymbol{\theta}$. In terms of the density function, this invariance is described by

$$f((g)x\,|(g)\theta)|\partial(\mathbf{g})\mathbf{x}/\partial\mathbf{x}'| = f(\mathbf{x}\,|\,\boldsymbol{\theta}) \quad \forall \mathbf{g}, \mathbf{x}, \boldsymbol{\theta}. \tag{1}$$

Both the definition and the density description provide direct methods for checking for invariance.

Consider now the use of differential methods. For this suppose that $\ln f(\mathbf{x}\,|\,\boldsymbol{\theta})$ has first derivatives with respect to $\mathbf{x}$ and $\boldsymbol{\theta}$:

$$\partial \ln f(\mathbf{x}\,|\,\boldsymbol{\theta})/\partial\mathbf{x}', \qquad \partial \ln f(\mathbf{x}\,|\,\boldsymbol{\theta})/\partial\boldsymbol{\theta}'.$$

The value of a derivative at a point $(\mathbf{g})x$ will be designated in the manner

$$\partial \ln f(\mathbf{y}\,|\,\theta)/\partial\mathbf{y}'\,|_{\mathbf{y}=(\mathbf{g})\mathbf{x}} = \partial \ln f((\mathbf{g})\mathbf{x}\,|\,\boldsymbol{\theta})/\partial(\mathbf{g})\mathbf{x}'. \tag{2}$$

This notational form has convenience for calculation of the derivative of a Jacobian determinant:

$$(\partial/\partial\mathbf{h}')|\partial(\mathbf{h})\mathbf{x}/\partial(\mathbf{g})\mathbf{x}'|_{\mathbf{h}=\mathbf{g}} = (\partial/\partial\mathbf{h}')\textstyle\sum_i |\partial(\mathbf{h})\mathbf{x}/\partial(\mathbf{g})\mathbf{x}'|^{(i)}\,|_{\mathbf{h}=\mathbf{g}},$$

where the superscript (i) is to denote that the derivative is to be applied only to the ith row;

$$\begin{aligned}(\partial/\partial\mathbf{h}')|\partial(\mathbf{h})\mathbf{x}/\partial(\mathbf{g})\mathbf{x}'|_{\mathbf{h}=\mathbf{g}} &= (\partial/\partial\mathbf{h}')\textstyle\sum_i \partial(\mathbf{h})x_i/\partial(\mathbf{g})x_i\,|_{\mathbf{h}=\mathbf{g}} \\ &= (\partial/\partial(\mathbf{g})\mathbf{x}')(\partial(\mathbf{g})\mathbf{x}/\partial\mathbf{g}')\end{aligned} \tag{3}$$

where the row vector of partial derivatives is applied to the Jacobian matrix by matrix multiplication.

Consider now the logarithmic derivative of equation (1) with respect to the parameter $\mathbf{g}$:

$$\begin{aligned}&(\partial/\partial\mathbf{g}')\ln f((\mathbf{g})\mathbf{x}\,|\,(\mathbf{g})\boldsymbol{\theta}) \\ &\qquad + (\partial/\partial\mathbf{h}')\ln\{|\partial(\mathbf{h})\mathbf{x}/\partial(\mathbf{g})\mathbf{x}'|\cdot|\partial(\mathbf{g})\mathbf{x}/\partial\mathbf{x}'|_{\mathbf{h}=\mathbf{g}}\} \equiv 0,\end{aligned}$$

$$\begin{aligned}&(\partial/\partial(\mathbf{g})\mathbf{x}')\ln f((\mathbf{g})\mathbf{x}\,|(\mathbf{g})\boldsymbol{\theta})(\partial(\mathbf{g})\mathbf{x}/\partial\mathbf{g}') \\ &\qquad + (\partial/\partial(\mathbf{g})\boldsymbol{\theta}')\ln f(\mathbf{g})\mathbf{x})|(\mathbf{g})\boldsymbol{\theta})(\partial(\mathbf{g})\boldsymbol{\theta}/\partial\mathbf{g}') \\ &\qquad + (\partial/\partial(\mathbf{g})\mathbf{x}')(\partial(\mathbf{g})\mathbf{x}/\partial\mathbf{g}') \equiv 0 \qquad \forall\,\mathbf{g}, \mathbf{x}, \boldsymbol{\theta}.\end{aligned} \tag{4}$$

The derivative being zero along connected g-paths is a necessary and sufficient condition for equation (1) to hold:

THEOREM 1. *The statistical model* $f(\mathbf{x}\,|\,\boldsymbol{\theta})$ *with first logarithmic derivatives with respect to* $\mathbf{x}$ *and* $\boldsymbol{\theta}$ *is invariant with respect to the class of transformations* T *if and only if* (4) *holds.*

Equation (4) has simple form; it is a vector equation that is linear in the logarithmic derivatives of f with respect to variable and parameter.

The extension to differentiable manifolds containing X, Ω, G is immediate: derivatives and Jacobian matrices are in terms of local coordinates.

3. Invariance and a group of transformations. Suppose now that the transformations T form a group, and that the coordinates of $\mathbf{g}$ are essential. The group of transformations is then conveniently designated by $G = \{\mathbf{g}\}$; let $\mathbf{g}, \mathbf{h}, \mathbf{a}$ be elements. Some Jacobian matrices at the identity $\mathbf{e}$ are

$$(5)\qquad \partial\mathbf{ag}/\partial\mathbf{a}'\,|_{\mathbf{a}=\mathbf{e}} = A(\mathbf{g}),$$
$$\partial\mathbf{ax}/\partial\mathbf{a}'\,|_{\mathbf{a}=\mathbf{e}} = B(\mathbf{x}); \qquad \partial\mathbf{a}\boldsymbol{\theta}/\partial\mathbf{a}'\,|_{\mathbf{a}=\mathbf{e}} = C(\boldsymbol{\theta}).$$

Some general Jacobian matrices can then be written

$$\partial\mathbf{gh}/\partial\mathbf{g}' = A(\mathbf{gh})A^{-1}(\mathbf{g});$$
$$\partial\mathbf{gx}/\partial\mathbf{g} \;= B(\mathbf{gx})A^{-1}(\mathbf{g}); \qquad \partial\mathbf{g}\boldsymbol{\theta}/\partial\mathbf{g}' = C(\mathbf{g}\boldsymbol{\theta})A^{-1}(\mathbf{g});$$

and the necessary and sufficient condition (4) for invariance can be multiplied on the right by $A(\mathbf{g})$ giving

$$(\partial/\partial\mathbf{gx}')\ln f(\mathbf{gx}\mid\mathbf{g}\boldsymbol{\theta})\cdot B(\mathbf{gx}) + (\partial/\partial\mathbf{g}\boldsymbol{\theta}')\ln f(\mathbf{gx}\mid\mathbf{g}\boldsymbol{\theta})\cdot C(\mathbf{g}\boldsymbol{\theta})$$
$$+ (\partial/\partial\mathbf{gx}')\cdot B(\mathbf{gx}) \equiv 0 \qquad \forall\, \mathbf{g}, \mathbf{x}, \boldsymbol{\theta}.$$

This can be rewritten as

$$(6)\qquad (\partial/\partial\mathbf{x}')\ln f(\mathbf{x}\mid\boldsymbol{\theta})\cdot B(\mathbf{x}) + (\partial/\partial\boldsymbol{\theta}')\ln f(\mathbf{x}\mid\boldsymbol{\theta})\cdot C(\boldsymbol{\theta})$$
$$+ (\partial/\partial\mathbf{x}')\cdot B(\mathbf{x}) \equiv 0, \qquad \forall\, x, \theta.$$

THEOREM 2 (Brillinger). *The statistical model $f(\mathbf{x}\mid\boldsymbol{\theta})$ with first logarithmic derivatives with respect to $\mathbf{x}$ and $\boldsymbol{\theta}$ is invariant with respect to the connected group $G = T$ if and only if* (6) *holds.*

Again the theorem extends to differentiable manifolds by expressing derivatives and Jacobians with respect to local coordinates.

4. The invariant models of a group. Consider a group G that gives transformations on a sample space X and transformations on a parameter space Ω. The general invariant model for such a group can be given explicit form.

Let $T(\theta)$ be the orbit of θ under the group G.

$$\mathrm{T}(\theta) = \{g\theta \mid g \,\varepsilon\, G\}.$$

The orbits form a partition of Ω.

On each orbit $T(\theta)$ choose a parameter value $\theta_0(\theta)$ as a reference value. The orbits and the reference values are then in 1–1 correspondence $T(\theta) \leftrightarrow \theta_0(\theta)$.

For any reference value θ_0 let $H(\theta_0)$ be the stabilizer subgroup: $H(\theta_0) = \{h \mid h\theta_0 = \theta_0,\ h \,\varepsilon\, G\}$. And for a general point θ on the orbit $T(\theta_0)$ let $[\theta] = \{g \mid g\theta_0 = \theta,\ g \,\varepsilon\, G\}$; the sets $[\theta]$ are the left cosets $gH(\theta_0)$ of the stabilizer subgroup. And let $[\theta]_*$ be an arbitrary element in the coset $[\theta]$.

Let $x(\theta_0)$ be *any* variable with distribution invariant under the group $H(\theta_0)$;

i.e. $x(\theta_0)$ and $hx(\theta_0)$ have the same distribution, $h \varepsilon H(\theta_0)$. Then the general invariant model for the group G has the generic representation

$$x = [\theta]_* x(\theta_0(\theta)). \tag{7}$$

The general invariant model has, for each θ_0, an arbitrary distribution $x(\theta_0)$ that is symmetric under $H(\theta_0)$; the general distribution is then generated by action of the group.

Now, let $\mathbf{x}$ be a variable with sample space R^p, let θ be a parameter with parameter space Ω, and let G be a group such that the transformations on x are differentiable and the stabilizer subgroups $H(\theta_0)$ have invariant differentials

$$m(\mathbf{x} \mid \theta_0)\, d\mathbf{x} \equiv m(h\mathbf{x} \mid \theta_0)\, dh\mathbf{x}, \qquad h \varepsilon H(\theta_0) \tag{8}$$

Let $f(x \mid \theta_0)$ be any density with respect to (8) that has the symmetry property:

$$f(\mathbf{x} \mid \theta_0) = f(h\mathbf{x} \mid \theta_0), \qquad h \varepsilon H(\theta_0).$$

The general invariant model for the group G has then the form

$$f([\theta]_*^{-1}\mathbf{x} \mid \theta_0(\theta))\, m([\theta]_*^{-1}\mathbf{x} \mid \theta_0(\theta))\, |\partial[\theta]_*^{-1}\mathbf{x}/\partial\mathbf{x}|\, d\mathbf{x}. \tag{9}$$

EXAMPLE. Consider a variable $\mathbf{x}$ with sample space $X = R^p$ and let the group G be the positive affine group on p-space. Suppose *the application of G to X gives the positive affine transformations*; these can be represented in matrix form by using augmented variables and matrices:

$$\mathbf{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{bmatrix} \quad \mathbf{g} = \begin{bmatrix} 1 & 0 & & 0 \\ a_1 & a_{11} & \cdots & a_{lp} \\ \vdots & \vdots & & \vdots \\ x_p & a_{pl} & \cdots & a_{pp} \end{bmatrix}, \qquad |\,\mathbf{g}\,| > 0.$$

Let $\theta = (\tau, \lambda)$ where τ is a positive definite symmetric matrix having the form

$$\tau = \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \tau_{11} & \cdots & \tau_{lp} \\ \vdots & \vdots & & \vdots \\ \mu_p & \tau_{pl} & \cdots & \tau_{pp} \end{bmatrix}$$

and suppose *the application of G to Ω* has the form

$$g\theta = g(\tau, \lambda) = (g\tau g', \lambda).$$

The group G is transitive on the matrices τ. Thus the orbits are indexed by λ; a convenient reference point on the λ orbit is $\theta_0 = (I, \lambda)$. The stabilizer subgroup for θ_0 is the group 0 of positive rotations:

$$h = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0_{11} & & 0_{1p} \\ \vdots & & & \\ 0 & 0_{p1} & & 0_{pp} \end{pmatrix}, \qquad |h| = +1.$$

The sets $[\theta]$ are the left cosets of 0 in the positive affine group.

Let $f(\mathbf{x} \mid \lambda)$ be any rotationally symmetric density:

$$f(h\mathbf{x} \mid \lambda) = f(\mathbf{x} \mid \lambda) \qquad h \varepsilon 0.$$

The general invariant model has then the form $f([\theta]_*^{-1}\mathbf{x} \mid \lambda)\, d\mathbf{x}$.

For a sufficiently large sample from the variable $\mathbf{x}$, the group G is unitary on the product sample space. Structural inference for this case has been submitted to the *Journal of the Royal Statistical Society* by D. A. S. Fraser and S. Haq. Structural inference when the group is not unitary has been examined by J. Bondar.

REFERENCES

[1] Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[2] Brillinger, D. R. (1963). Necessary and sufficient conditions for a statistical model to be invariant under a Lie group. *Ann. Math. Statist.* **34** 492–500.

[3] Fraser, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.

[4] Fraser, D. A. S. (1962). On the consistency of the fiducial method. *J. Roy. Statist. Soc. B* **24** 425–434.

[5] Fraser, D. A. S. (1966). Structural probability and a generalization. *Biometrika* **53** 1–9.

[6] Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37** 643–656.

[7] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[8] Lindley, D. V. (1959). Fiducial distributions and Bayes' theorem. *J. Roy. Statist. Soc. B* **21** 102–107.

[9] Pitman, E. J. G. (1939). Location and scale parameters. *Biometrika* **30** 391–421.

[10] Stone, M. (1965). Right Haar measure for convergence in probability to quasi posterior distributions. *Ann. Math. Statist.* **36** 440–453.