AN ANALYTIC-ALGEBRAIC APPROACH TO
STATISTICAL MODELS AND INFERENCE

David Brenner and D.A.S. Fraser, F.R.S.C.

This paper reports on current research in which an analytical-physical approach is taken to the definition of the statistical model and to the content of the inference base, given by the model and data together. The paper also presents some integrating results that follow from this.

The primitive statistical model with virtually no background information (except the existence of probabilities for events $A$ on the sample space $S$) may be expressed $M = \{ (S,A,P) : P \in P \}$ where the knowledge about the system being investigated is summarized by the collection $P$ of possible candidates for the true probability measure $P_\ast$. With this we have recourse only to the knowledge that the empirical relative frequency of any event $A$ in $A$ converges to the true probability $P_\ast(A)$ of that event. The preceding formalizes the minimum context referred to as a random system.

Likelihood. As an example of the analytical approach, consider the following (as in Fraser [3, pp. 70, 98]). Let $(S,A,\mu)$ be the sample space $S$ (an open subset of $\mathbb{R}^k$), the
Borel class $\mathcal{A}$ of events, and a support measure $\mu$ (Borel measure or an equivalent). With this the conventional statistical model is $M = \{f(\cdot | \theta) : \theta \in \Omega \}$, a class of continuous density functions on $S$. The inference problem is to assess what the model $M$ and the data $y$ say concerning the unknown $\theta_*$ in $\Omega$.

For a data point $y$ the model specifies the probability $f(y | \cdot)\mu(dy)$ of its occurrence, a function on $\Omega$. As probability at a point is zero and the size of reference neighbourhood is arbitrary, the model provides only $L(y) = \{cf(y | \cdot) : c \in \mathbb{R}^+\}$, a ray in $\mathbb{R}^\Omega$; this is the likelihood function. The preimage of the mapping $L(\cdot)$ from $S$ to $\mathbb{R}^\Omega / \mathbb{R}^+$ is a partition of $S$ into sets of points that are equivalent relevant to the model. This reduction by eliminating the arbitrary is the same as that obtained in a weaker sense by invoking the sufficiency principle.

Inference Base. The analytic approach focuses initially on the requirements for the statistical model. Criteria for the model $M$ are organized in Fraser [3, p. 3]: the purpose is to describe unknowns; the model contains real components relative to the investigation (objective); and components of the investigation are included in the model (comprehensive). This leads to the inference base $T = (M, y)$, and the problem is to determine what can be said concerning the unknowns in the base $T$ without the inclusion of arbitrary elements.
Event Structure. Clear examples do exist where the statistical model has a distinct component, an objective probability space \((S_0, A_0, P)\) or \((S_0, A_0, \mu, p)\) where \(P\) denotes a measure and \(p\) a density with respect to a support measure \(\mu\). The use of probability and conditional probability focuses on the nature of the information available from a data point \(y\) concerning the realized value on the probability space. We have examined the following special case in detail: \(M = (S, A, \mu, p; \Phi)\) where \(\Phi = \{\phi\}\) is a class of bijections \(S \rightarrow S\). The response variable satisfies \(y = \phi z\) where \(\phi\) is the unknown presentation in \(\Phi\) and \(z\) is the realized value on the probability space. The information from a data point \(y\) concerning a realized \(z\) can be formalized as the \(B\)-orbit of an information display: let \(B\) be the group of bijections on the response space \(S\), and \(D(\phi, y) = \{(\phi, \phi^{-1} y) : \phi \in \Phi\}\) be the set of preimage values for \(z\) labelled by the possible presentations \(\phi\); then \(I(\phi, y) = BD(\phi, y) = \{D(s\phi, sy) : s \in B\}\) gives this information display.

Proposition. The preimage partition of the information function is the orbit space \(S/G\), where \(G\) is the invariant group \(\{s : s\phi = \phi, s \in B\}\).

The information obtained from a value \(z\) is given by \(I(\phi, \phi z)\) where \(\phi\) is the unknown presentation. This is a function \(S + \text{im} I(\phi, \cdot)\) if and only if \(S/G = T\) where \(T\) consists of sets \(\cup \phi^{-1} y\). For this to hold under repeated
sampling with an effective point we have that $\hat{\phi} = G\phi_0$ or with a relabelled $\phi_0$ we have that the new $\phi = G$ is a group. This is then a structural model as examined in Fraser [1,3].

**Identified Form.** The preceding involved an objective probability space as a component of the model. Some examples of this have been discussed in Fraser [1,3] and some risks inherent in the lack of objectiveness of the probability space [2]. We report now on some current research on the identification of distribution form. The starting point is the background information concerning a system under investigation and the examination of this by a class of bijective functions $T = \{t\}$ which allow each of the possible distribution forms to be examined on a space $\forall e$. Three different definitions are given for a platform of functions. In each case the requirement that $T$ be a platform is a necessary and sufficient condition for a nominal distribution form to be objective.

The preceding discussion gives the grounds for an objective probability space to be a component of a statistical model. This should not be confused with another direction on statistical inference involving the addition of a pivotal quantity as originally promoted in the writings of Fisher. The approach of this article is that such an addition is arbitrary. The objective distribution form discussed above contains functions that are formally inverses of pivotal quantities but the resemblance is only formal. The group closure property was fundamental and obvious examples show that without this
closure contradictions can occur.

Unifications. The research currently in progress is based on the preceding results and involves a direct analytical examination of the inference base and the separation of the types of information available - categorical, frequency and diffuse. Some preliminaries on the separation of categorical information may be found in Fraser [3, p. 49]. In a second direction the event information concerning realized values leads to the automatic conditioning found with identified form and structural models. The present indication is that the four necessary reduction methods [3, pp. 49, 68] will evolve as a simple consistent categorical separation of the inference base.

References


University of Toronto,
Toronto, Canada M5S 1A1.

Received March 28, 1980