The Basis of Inference

D. A. S. FRASER, F.R.S.C.*

To a mathematician inference suggests deduction. But to mathematical statisticians, inference means a deductive system, the system he believes in, or hopes for, or lives with by act of faith, a system to describe the process by which observation, measurement, experimentation become science or knowledge. There is, however, an embarrassment of riches—there are many different systems for inference. And the embarrassment goes deeper: the systems lead to different results for the same problem, and more seriously they lead to answers where no answer is appropriate—a moderate literature in the social sciences now shows that many “proven results” are, in fact, not true.

The origins of mathematical statistics are the origins of probability theory—French mathematicians calculating the odds for French gamblers. Development was slow until the second quarter of this century; then the pace became intense, the literature voluminous, and the results dissatisfying. Many statisticians have now withdrawn to the mother discipline of probability theory, where the uncertainty is described and is not in the description: the Annals of Mathematical Statistics is now largely a journal of probability theory.

The contemporary systems of inference are all based on the classical model of statistics: a variable \( x \) designating the response being examined, a parameter \( \theta \) designating possible states for the entity being examined, and a density function \( f(x; \theta) \), giving the relative frequency of different values \( x \) given a state \( \theta \). This is a special kind of model. It is equivalent to a box with input variable \( \theta \), output variable \( x \), and behaviour function \( f(x; \theta) \); no account is taken of mechanisms in the entity being examined and, correspondingly, no specification is made of the contents for the box. The box is conceptually empty; it is a black box.

One system of inference uses the classical model alone. Other systems of inference add external structure to the classical model—i.e., they add external structure to the black box. One prominent system is a refashioning of a very old system due to the Reverend Thomas Bayes some two hundred years ago, who chose not to publish it during his lifetime. The Bayes system treats the unknown state \( \theta \) as if it had occurred randomly with a known density \( p(\theta) \), the prior density. In effect, the black box of the classical model is coupled to a second box with no input variable, with output variable \( \theta \),

*Department of Mathematics, University of Toronto.
and with behaviour function $p(\theta)$. The composite box has no input variable, but has output variable $x$, an intervening variable $\theta$, and linking behaviour functions $p(\theta), f(x: \theta)$. In the typical application, the prior density $p(\theta)$ is taken, by desperate assumption, to represent uniform ignorance, or diffuse knowledge, or the non-informative background, on the insight of a crystal ball. As Bayes is currently popular, this is a fruitful area for mathematical analyses.

The inference systems add external structure to the classical model—with great ingenuity and sophistication, with high degrees of mathematical abstraction, with a wealth of notation, and perhaps too frequently with an open disregard for the entity being examined. The basic ingredient, however, is the black box.

But should the box be black? Should the basic ingredient of the statistical model be a box that is conceptually empty? Other areas of science are analytic, and incisive, and are concerned with the pursuit of the smallest elements of structure. But statistics has been developed by mathematicians, and mathematicians start with the given and add definitions and axioms; the black box was the given—it was not questioned. The theoretical work of R. A. Fisher is close to an exception; Fisher, a mathematician, but also a geneticist and biologist, developed large areas of theory in the context of applications, but unfortunately he did not formalize any internal structure for the black box.

The statistical model has not always been a black box. Early statistics treated the error of measurement as an entity in its own right. This cannot be embraced by the classical model.

And it need not be a black box now. Consider a model for a measuring instrument and for its use to measure a quantity $\theta$. Suppose the measuring instrument has been investigated and its error of measurement admits description by the error distribution $f(e)de$ on $R^1$. And let $x_1, \ldots, x_n$ be a sequence of $n$ measurements on the quantity $\theta$. The model is the simple measurement model;

$$\prod_1^n f(e_i) \prod_1^n de_i, \quad x = [\theta, 1]e.$$ 

The model has an error distribution $\Pi f(e_i) \Pi de_i$ on $R^n$ which describes the multiple operation of the instrument; $e = (e_1, \ldots, e_n)$ is a variable. And the model has a structural equation $x = [\theta, 1]e$ in which a realized error vector $e$ has been translated $[\theta, 1]$ to give the observation vector $x$; the vector $e$ is an unknown constant, the transformation $[\theta, 1]$ is an unknown translation, and the vector $x$ is a known constant. The notation $[a, c]$ designates the affine transformation $[a, c]x = a1 + cx$.

The measurement model is a special case of a much more general model, the structural model. Let $f(E)de$ be an error distribution on a space $\mathcal{X}$, an open subset of a Euclidean manifold. Let $\theta$ be a quantity, an element of a group of differentiable transformations that is unitary on $\mathcal{X}$ ($G$ is unitary: $g'E = g''E$ implies $g' = g''$ for $g', g''$ in $G$). The structural model is
\[ f(E) dE, \quad X = \theta E. \]

The model has an error distribution \( f(E) dE \) describing the internal operation of a physical system \((E)\) is a variable). And the model has a structural equation \( X = \theta E \) in which a realized value \( E \) from the error distribution has been transformed by \( \theta \) to give the observation \( X \) (\( \theta, E \) are unknown constants; \( X \) is a known constant).

The group \( G \) generates orbits on \( \mathfrak{X} \):

\[ GX = \{ gX : g \in G \}. \]

Let \([X]\) be a transformation variable mapping \( \mathfrak{X} \) into a duplicate \( G^* \) of the group \( G \):

\[ [gX] = g[X], \quad \text{for every} \ g, X. \]

A transformation variable \([X]\) can be used to give the position of a point \( X \) on its orbit relative to a reference point \( D(X) = [X]^{-1}X \):

\[ X = [X]D(X). \]

The orbits \( GX \) can be indexed by the reference points \( D(X) \).

The structural equation allows one to see into the system being examined and to discover characteristics of the realized error: the orbit of \( E \) is the orbit of \( X \),

\[ D(E) = D(X). \]

The structural equation, however, gives no further information concerning \( E \), and the structural equation

\[ X = \theta E \]

reduces to

\[ [X] = \theta[E]. \]

All the conditions are fulfilled for making probability statements concerning an unknown constant:\(^1\) the constant arose as a realized value from a known probability mechanism; the only information concerning the value of the constant is in the form of an event for the probability mechanism. For the structural model, the unknown \( E \) satisfies \( D(E) = D(X) \), which has the form of an event based on the variable \( D(E) \). There is no additional information concerning the position \([E]\) because the orbit is a homogeneous space. The structural equation by its own information context then gives the reduced structural model

\[ g([E]; D(X))d[E], \quad [X] = \theta[E]. \]

The reduced model has an error probability distribution \( g([E]; D(X))d[E] \), which provides probability statements concerning the unknown error location \([E]\). And the model has a structural equation \([X] = \theta[E] \) in which the unknown \([E]\) determines the link between the unknown \( \theta \) and the known \( X \).

The derivation of the error probability distribution \( g([E]; D)d[E] \) is

\(^1\)The conditions that allow card players to make probability statements concerning concealed cards after card-shuffling and after card-dealing.
simple. Let \(J_N(X)\) be the Jacobian of the transformation from the reference point \(D = D(X)\) to the point \(X\); the compensated differential
\[
dm(X) = dX / J_N(X)
\]
is then the invariant differential that agrees with Euclidean volume at the reference point. Let \(J_L(g)\) designate the Jacobian of the left transformation from the identity \(i\) to the point \(g\); the adjusted differential
\[
d\mu(g) = dg / J_L(g)
\]
is then the unique left Haar differential that agrees with Euclidean volume at the identity.

The change of variables from \(E\) to position \([E]\), given the orbit \(D(E) = D\), can now be made. The probability element is
\[
f(E)dE = f(E)J_N(E)dm(E) = f([E];D)J_N(E)d\mu([E]),
\]
where \(\delta\) is a proportionality constant. The normalized probability element along the orbit \(D(E) = D\) is then
\[
g([E];D)d[E] = k(D)f([E];D)J_N(E)d[E] / J_L([E]).
\]

The reduced model produces probability statements concerning the unknown \([E]\); for example,
\[
Pr([E] \in A^*) = \int_{A^*} g([E]; D(X))d[E].
\]
The orbit is, of course, known: \(D(E) = D(X)\).

A probability statement concerning the unknown \([E]\) in the structural equation
\[
[X] = \theta [E] \quad (X \text{ known})
\]
is ipso facto a probability statement concerning the unknown \(\theta\):
\[
Pr[\theta \in A] = Pr([E] \in A^{-1}[X]) = \int_{A^*} g([E]; D(X))d[E],
\]
where
\[
A^* = A^{-1}[X] = \{\theta^{-1}[X]: \theta \in A\}.
\]
Such a probability statement concerning \(\theta\) is called a structural probability statement. Structural probability statements can be calculated directly by the substitution \([E] = \theta^{-1}[X]\) for the error probability distribution:
\[
g^*(\theta; X)d\theta = g(\theta^{-1}[X]; D(X))d\theta^{-1}[X]
\]
\[
= k(D)f(\theta^{-1}X)J_N(\theta^{-1}X)d\mu(\theta^{-1}[X])
\]
\[
= k(D)f(\theta^{-1}X)J_N(\theta^{-1}X)\Delta(\theta^{-1}[X])d\mu([X]^{-1}\theta)
\]
\[
= k(D)f(\theta^{-1}X)J_N(\theta^{-1}X)J_L^*(\theta^{-1}[X])d\theta / J_L(\theta^{-1}[X] J_L(\theta).
The derivation involves the Jacobian $J_L^*(g)$ of a right transformation from $i$ to $g$, the unique right Haar differential
\[ dv(g) = dg / J_L^*(g) \]
coinciding with Euclidean volume at $i$, and the modular function $\Delta(g)$ relating the left and right differentials:
\[ \Delta(g) = J^*(g) / J(g), \]
\[ d\mu(g) = \Delta(g) dv(g) = \Delta(g) d\mu(g^{-1}), \]
\[ dv(g) = \Delta^{-1}(g) d\mu(g) = \Delta^{-1}(g) dv(g^{-1}). \]

The distribution $g^*(\theta; X) d\theta$, which provides probability statements concerning $\theta$, is called the structural distribution for the quantity $\theta$.

The black box has been replaced by an internally structured box—the classical model has been replaced by a more comprehensive model that describes additional properties of a system. In the more comprehensive model, probability statements can be made concerning the quantity $\theta$; these are probability statements in the sense of classical probability (the gambling house doesn't lose). No external structure or theory of interference is needed.