ON FISHER'S EXAMPLE OF VARIETY OF LOGICAL TYPES OF INFERENCE

BY

PETER TAN

TECHNICAL REPORT NO. 7
DECEMBER 5, 1969

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-15909

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON FISHER'S EXAMPLE OF VARIETY OF LOGICAL TYPES OF INFERENCE

By

PETER TAN
Stanford University and
Carleton University

TECHNICAL REPORT NO. 7
DECEMBER 5, 1969

PREPARED UNDER THE AUSPICES
OF

National Science Foundation Grant GP-15909

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
On Fisher's Example of Variety of Logical Types of Inference

Peter Tan, Stanford University *

Summary: Fisher uses a bivariate normal distribution to illustrate two different types of inference possible, one based on a sufficient statistic while the other on an ancillary statistic, which arise from different mathematical specifications of the problem. This note shows that for this example a method of inference called Structural Inference by Fraser [1], with classical probabilities describing the unknown parameters, is possible for both cases and the search for sufficient statistics and ancillary statistics are both extraneous to the problem and needless if we look into the internal structure of the comprehensive models [2]. It is further pointed out that structural inference is typically different from a method of inference based on the likelihood function alone.

* Now at Department of Mathematics, Carleton University, Ottawa, Canada
1. **The Problem and Fisher's Solutions.**

In his discussion on "variety of logical types" of inductive inference, the late Sir R. A. Fisher states the following:

"It is a noteworthy peculiarity of inductive inference that comparatively slight differences in the mathematical specification of a problem may have logically important effects on the inference possible." ([3], p. 133)

He goes on to point out the existence of the following three types of problems.

a) There exists a sufficient statistic which has the same dimension as the parameter. Inference about the parameter can then be made from the reduced model of the sufficient statistic.

b) No sufficient statistic as prescribed in (a) exists, but there exists an ancillary statistic whose probability distribution is independent of the parameter so that "the most likely estimate could be made exhaustive by means of the ancillary values". ([3], p. 138) For this type of problems, inference should be made conditional on the realized value of the ancillary statistic.

c) Neither a sufficient statistic nor an ancillary statistic exists. In such cases Fisher suggests for rational inference the calculation of the mathematical likelihood for each plausible value of the unknown parameter.
The example Fisher uses is the following:

Suppose the $x$ and $y$ are two observed quantities, known each to be normally, and independently distributed with unit variance, $x$ about an unknown value $\theta$ and $y$ about $\phi$. It will be required to draw inferences about the pair of values $(\theta, \phi)$ which may be represented as an unknown point, $H$, on a plane, on which $(x, y)$ may be represented by an observed point, $O$.

Using geometric arguments, Fisher shows that the above three types of problems are illustrated by the following three cases:

a) $H$ is known to lie on a given straight line.

The maximum likelihood estimate, $M$ on the given line, which gives the shortest distance from the observed point, $O$, to the line, is a sufficient estimate of $H$ in the sense of Fisher.

b) $H$ is known to lie on a circle.

No sufficient statistic in the sense of Fisher exists; but the distance, $a$, between the center, $C$, of the circle and the observed point, $O$, is an ancillary statistic. Conditional on the realized value $a$ of this ancillary statistic, the position $M$ on the circle, where the line segment $OC$ or its extension meets the circle, becomes conditionally sufficient. It is also the maximum likelihood estimate of the parametric point $H$. Conditional (fiducial) probability statements about $H$ on the circle can then be drawn from the observed position $M$, given the ancillary value $a$. 
c) The given functional relationship between \( \theta \) and \( \phi \) does not confine the point \( H \) either to a straight line, or to a circle, but to some other plane curve.

In this case Fisher states that "it is not to be expected either that a sufficient statistic should exist, or that the most likely estimate could be made exhaustive by means of ancillary values".

This bivariate normal example had actually been studied before Fisher analytically by A.R.G. Owen ([4], 1948) in the latter's search for ancillary statistics. This reference, however, was not given by Fisher.

In Section 3 we shall obtain structural probability distributions for the parameter \( \theta \) when the pair \((\theta, \phi)\) lies either on a straight line or on a circle by the same structural method developed by Fraser [1], without explicit use of either sufficient statistics or ancillary statistics. This will show that the disparity in methods of inference possible for the two types of problems examined by Fisher can be replaced by a unified method when the group structure of the specification is utilized in statistical inference.

In Section 4 we shall further show by this example that inference from the likelihood function without utilization of the inherent group structure of the problem when it exists is typically different from structural inference.
2. **General Theory for a Structural Model.**

Let $\mathcal{X}$ be an open set in Euclidean space $\mathbb{R}^n$; $G$ is a unitary group of one-to-one transformations of $\mathcal{X}$ onto $\mathcal{X}$ such that

$$g_1^x = g_2^x, \quad g_1 \in G, \quad g_2 \in G$$

for any $x \in \mathcal{X}$ implies $g_1 = g_2$; $E$ is an unobservable error variable with a known probability distribution on the space $\mathcal{X}$; $X$ is an observable random variable on $\mathcal{X}$ which is related to $E$ by a structural equation

$$X = \theta E, \quad \theta \in G.$$  

The transformations $g$ in $G$ carry a point $X$ into the orbit of $X$:

$$GX = \{gX: \quad g \in G\} = GE,$$

while the orbits form a partition of $\mathcal{X}$.

In the case where $G$ is an open set in the real line, the position of each point $E$ in $GX$ can be described by a group element $g$ such that $E = gD$ where $D$ is a chosen reference point on the orbit $GX$, or by a real-valued continuous transformation variable $r(E) = r(gD)$ whose values have a one-to-one correspondence with $G$, and we may write

$$r(gD) = g^* r(D), \quad g \in G.$$  

$G^* = \{g^*\}$ is easily seen to be a group of transformations of the
range space of $r(E)$ onto itself with one-to-one correspondence with $G$.

The conditional distribution of $r(E)$ given the orbit $GX$ indexed by $D(X)$, obtained either by a routine method or by a method in [1] using invariant differentials, is specified by the probability element

$$g(r; D)dr.$$ 

From the structural equation

$$E = \theta^{-1}X$$

and the conditional distribution of $r(E)$, and the one-to-one correspondence of $G$ and the range space of $r(E)$, we can derive a structural probability distribution of $\theta$ conditional on the orbit $GX$.

The probability distribution of $GX = GE$ is derived from the known distribution of $E$; it is independent of $\theta$. The orbital statistic represented by $D(X)$ or its equivalent is therefore an ancillary statistic. The conditional distribution of $r(E)$ given the orbit $GX$ is, in general, dependent on the orbit. When it is independent of the orbit, $r(X)$ is a sufficient statistic.
3. **Solution of Fisher's Problem by the Structural Method.**

Let \((x, y)\) be a bivariate normal variable with probability density

\[
f(x, y|\theta, \phi) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}[(x - \theta)^2 + (y - \phi)^2]\right\}.
\]

**Case (a).** \(\phi = a\theta + b\), where \(a, b\) are real constants, \(-\infty < \theta < \infty\).

Let \(E = (E_1, E_2)\) be a bivariate vector variable, where

\[
E_i = (e_{i1}, e_{i2}, \ldots, e_{in})', \quad i = 1, 2, \text{ with the known probability density}
\]

\[
f(E) = \left(\frac{1}{2\pi}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} e_{ii}^2 + \sum_{i \neq j}^{n} (e_{ij} - b)^2\right\}.
\]

Let \(X = (x, y)\), where \(x = (x_1, x_2, \ldots, x_n)'\) and \(y = (y_1, y_2, \ldots, y_n)'\), be a sample of \(n\) observations of the bivariate normal variable.

\(G\) is a group of transformations of \(R^n \times R^n\) onto itself defined by

\[
G = \{g_t : -\infty < t < \infty\}
\]

such that

\[
g_tX = g_t(x, y) = (x + tl, y + ta),
\]

where \(l = (1, 1, \ldots, 1)'\) and \(a = (a, a, \ldots, a)'\). \(X\) and \(E\) are related by the structural equation.
\[ X = g_0 E \text{ or } (x_1, y_1) = (e_{11} + \theta, e_{21} + a\theta). \]

The orbit of \( X \) under \( G \) is

\[ GX = \{ g_t X = (x, y) + t(1, a), -\infty < t < \infty \}. \]

The position of a point \( E \) in \( GX \) can be described by the value of the real-valued variable

\[ r(E) = (\bar{e}_1 + a\bar{e}_2 - ab)/1 + a^2, \quad \bar{e}_i = \frac{1}{n} (e_{i1} + e_{i2} + \ldots + e_{in}), \quad i = 1, 2. \]

Let \( E \) in \( GX \) be the unique point where \( D = (D_1, D_2), \)

\[ D_i = (d_{i1}, \ldots, d_{in}), \quad i = 1, 2, \text{ and } r(D) = 0. \]

Then \( E = g_r(E)D \), and

\[ e_{1i} = d_{i1} + r(E), \quad e_{2i} = d_{2i} + ar(E), \quad i = 1, 2, \ldots, n. \]

The conditional distribution of \( r(E) \), given the orbit \( GX \) indexed by \( D \) has the probability element

\[
g(r: D)dr = k(D)\exp\left\{ -\frac{1}{2} \left[ \sum_{i=1}^{n} (d_{i1} + r)^2 + \sum_{i=1}^{n} (d_{2i} + ar - b)^2 \right] \right\} dr
\]

\[ = k'(D)\exp\left\{-\frac{n(l+a^2)}{2} r^2\right\} dr, \]

where \( k(D) \) and \( k'(D) \) are constants dependent at most on \( D \) only.

\( k'(D) \) is, in fact, independent of \( D \) since it is a normalizing factor of a normal variate \( r \) with zero mean and known variance.

Since \( r(E) = r(X) - \theta \), where \( r(X) \) is the observed value of a statistic, each value of \( r(E) \) determines a unique value of \( \theta \).
The distribution of \( r(E) \) leads to a structural probability distribution of \( \theta \) whose probability element is

\[
g(\theta)d\theta = \left[ \frac{1}{2\pi} n(1 + a^2) \right]^{\frac{1}{2}} \exp\left\{ -\frac{n}{2} (1 + a^2)(\theta - r(X))^2 \right\} d\theta
\]

where

\[
r(X) = \frac{\bar{x} + ay}{1 + a^2}.
\]

The conditional distribution of \( r \) given the orbit \( GX \) is actually independent of the orbit. \( r(X) \) is therefore a sufficient statistic. One can easily see that \( r(X) \) is the unbiased, maximum likelihood estimate of \( \theta \). It is, as Fisher has observed correctly, the \( \theta \)-coordinate in the \((\theta, \phi)\)-plane of the point \( M \) on the line \( \phi = a\theta + b \) which is nearest the observation \((\bar{x}, \bar{y})\).

The orbit \( GX \) can be represented by \( S^2 \), the square of the distance from \((\bar{x}, \bar{y})\) to \( M \), given by

\[
S^2 = \left( -ax + \bar{y} - b \right)^2 / (1 + a^2).
\]

\( S^2 \) is an ancillary statistic; \( t \) and \( S^2 \) are statistically independent.

**Case (b)**: \((\theta - a)^2 + (\phi - b)^2 = r^2\).

The error variable \( E = (E_1, E_2) \) and the observation \( X = (x, y) \) remain defined as in case (a) except that the known probability distribution of \( E \) is now taken to possess the density
f(E) = \left( \frac{1}{2\pi} \right)^n \exp \left\{ - \frac{1}{2} \sum_{i=1}^{n} (e_{1i} - a - r)^2 + \sum_{i=1}^{n} (e_{2i} - b)^2 \right\}.

There exists a natural group of rotations

\[ G = \left\{ g_t = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \ 0 \leq t < 2\pi \right\} \]

of the sample space such that

\[ g_t x = g_t (x, y) = (x', y') = x' \]

where

\[ x' = (x'_1, x'_2, \ldots, x'_n) \]
and \[ y' = (y'_1, y'_2, \ldots, y'_n) \]
are defined by the following matrix equation:

\[
\begin{pmatrix}
\cos t & -\sin t \\
\sin t & \cos t
\end{pmatrix}
\begin{pmatrix}
x'_i - a \\
y'_i - b
\end{pmatrix}
= \begin{pmatrix}
x'_i - a \\
y'_i - b
\end{pmatrix}, \quad i = 1, 2, \ldots, n.
\]

E and X are related by the structural equation

\[ X = g_t E, \quad \tau = \tan^{-1}(\phi - b/\theta - a). \]

The orbit of X under G is

\[ GX = \{ E = g_t X: \ 0 \leq t < 2\pi \}. \]
The position of each point $E$ on the orbit can be described by the group element

$$\beta(E) = \tan^{-1}\left[\frac{(\bar{e}_2 - b)/(\bar{e}_1 - a)}{(\bar{e}_1 - a)^2 + (\bar{e}_2 - b)^2}\right],$$

where

$$\bar{e}_i = \frac{1}{n}(e_{1i} + e_{2i} + \ldots + e_{ni}), \quad i = 1, 2; (\bar{e}_1 - a)^2 + (\bar{e}_2 - b)^2 = R^2,$$

and the orbit can be represented by the reference point

$$D = (D_1, D_2), \quad D_i = (d_{1i}, d_{2i}, \ldots, d_{ni}), \quad i = 1, 2, \text{ with } \beta(D) = 0.$$

It follows that

$$(e_{1i} - a)^2 + (e_{2i} - b)^2 = (d_{1i} - a)^2 + (d_{2i} - b)^2, \quad i = 1, 2, \ldots, n,$$

for all $E$ in $GX = GD$.

The conditional distribution of $\beta(E)$ given the orbit with reference point $D$ has the probability element

$$g(\beta: D)d\beta = k(D)\exp\{nrR \cos \beta\}d\beta,$$

where $k(D)$ is a constant depending at most on $D$.

The structural equation $X = g_t E$ leads to the relations

$$\beta(X) = \beta(E) + \tau,$$

$$\beta(X) = \tan^{-1}\left[\frac{(\bar{y} - b)/(\bar{x} - a)}{(\bar{x} - a)^2 + (\bar{y} - b)^2}\right].$$
As $\beta(X)$ is an observed value of a statistic, the conditional distribution of $\beta$ gives rise to the structural probability distribution of the parametric variable $\tau$ whose probability element is

$$ g(\tau; D)d\tau = k(D)\exp\{nrR \cos(\tau - \beta)\}d\tau $$

where the normalizing constant $k(D)$ is

$$ k(D) = 1/2\pi I_0(nrR), $$

$I_0$ being a Bessel function.

It is to be noted that the conditional distributions of $\beta$ and $\tau$ given the orbital value $R$ is dependent on $R$.

The maximum likelihood estimate of $\theta$ is

$$ \hat{\theta} = a + \tau \cos\beta(X) $$

and $(\hat{\theta}, \hat{R}^2)$ or $(\beta, R)$ or $(\bar{x} - a, \bar{y} - b)$ forms a minimal sufficient statistic for $\theta$. Given the ancillary value $R$, $\beta$ or $\hat{\theta}$ becomes conditionally sufficient. $\hat{\theta}$ is, in the words of Fisher, "made exhaustive by means of the ancillary values".

4. Remarks On The Unity In Statistical Inference Achieved By The Structural Method.

The structural probability distributions for the parameter $\theta$ for the two different specifications are obtained by the same method of inference. The group structure in each case reveals the existence of an ancillary statistic, conditional on which inference about $\theta$ is made from a conditionally sufficient statistic. The same method
can be readily applied to an example investigated by Sprott [5,6] without the need of first searching for an ancillary statistic.

We shall further show that the structural inference about \( \theta \) in the previous section is in fact an inference from the likelihood function conditional on the realized value of the orbital (ancillary) statistic. It is an inference from the likelihood function alone when and only when the conditional distribution of the likelihood function as a position statistic on a given orbit is independent of the orbit.

The logarithm of the likelihood function for a sample \( X = (x, y) \) of size \( n \) from the bivariate normal distribution is

\[
\ell(\theta, \phi | X) = -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ (x_i - \theta)^2 + (y_i - \phi)^2 - (x_i - \theta_o)^2 - (y_i - \phi_o)^2 \right] \right\}
\]

where \( (\theta_o, \phi_o) \) is a fixed parameter point.

When \( \phi = a\theta + b \), the function \( \ell(\theta, \phi | X) \) can be put in the form

\[
\ell(\theta | X) = -\frac{n}{2} (1 + a^2) \left\{ -2r(X) (\theta - \theta_o) + (\theta^2 - \theta_o^2) \right\}
\]

\[
= c_o + c_1 r(X) \theta + c_2 \theta^2
\]

where \( c_o, c_1, c_2 \) are constants independent of \( X \) and \( \theta \). The function \( \ell \) considered as a statistic is equivalent to the statistic \( r(X) \) and can be used to index the position of each point \( E \) on the orbit \( GX \). Structural inference about \( \theta \) based on the conditional
distribution of r(X) given the orbit is therefore equivalent to an
inference from the conditional distribution of the likelihood function
λ given the orbit. In this case it is equivalent to inference from
the likelihood function alone due to independence of λ(·|X) and the
orbit.

When \((\theta-a)^2 + (\phi-b)^2 = r^2\), \(\lambda(\theta,\phi|X)\) can be put in the form
\[\lambda(\theta|X) = n(x - a)[(\theta-a) - (\theta_o-a)] + n(y-b)[(\phi-b) - (\phi_o-b)]\]

\[= [n_r \cos\beta(X)] \cos\tau(\theta) + [n_r \sin\beta(X)] \sin\tau(\theta) + K(s,\beta(X)),\]

where \(K\) is independent of \(\theta\). For a fixed value of \(S\), there is
one-to-one correspondence between the set of values \(0 < \beta < 2\pi\)
and the class \(L(S)\) of functions \(\lambda:\)
\[\{\lambda(\theta|X) = \lambda(\tau;\beta) = K_1(\beta) \cos \tau + K_2(\beta) \sin \tau + K_3(\beta):\]

\[-n_r < K_1, \ K_2 < n_r}\}.

The structural probability distribution of \(\tau(\theta)\) obtained from the
conditional distribution of \(\beta(X)\) given the orbit \(GX\) is equivalent
to that obtained from the conditional distribution of the likelihood
function \(\lambda\) considered as a statistic given the orbit \(GX\) represented
by the realized value of \(S\). Since the class \(L(S)\) of likelihood
functions varies with \(S\), statistical inference conditional on the
ancillary statistic \(S(X)\) is distinctly different from inference made

13
from the likelihood function alone.

Acknowledgement. Part of the material in this paper is a revised version of a section in the author's Ph.D. Thesis [7] submitted to the University of Toronto.
References


