Fiducial inference for location and scale parameters

By D. A. S. Fraser

University of Toronto

Summary

Some recent papers on fiducial inference for location and scale parameters are examined from the transformation-parameter viewpoint developed in Fraser (1961a, b). For the multivariate normal case with a covariance matrix of known form, an apparent conflict between two fiducial distributions is found to be due to the inappropriate use of one of the distributions. For the multivariate normal case with an unknown covariance matrix, the Fisher–Cornish fiducial distribution for the means is apparently invalid as its derivation violates accepted rules for obtaining fiducial distributions. A fiducial distribution can, however, be developed from the transformation-parameter model and it utilizes additional structure in the form of regressions among the co-ordinates of the basic variable. The alternative fiducial distribution in Statistical Methods and Scientific Inference (Fisher, 1956) is in conflict with this transformation-parameter fiducial distribution.

1. Introduction

Some fiducial distributions for the bivariate and multivariate normal will be examined and compared with fiducial distributions developed from the transformation-parameter viewpoint (Fraser, 1961a, b). An advantage of the transformation-parameter fiducial distributions is their frequency interpretation; a description of this may be found in abbreviated form in Fraser (1961a, b) or in more transparent form for a simple example in Fraser (1963). The interpretation requires an additional ingredient in the statistical model—it requires a position function that describes where the observation is relative to the parameter or conversely where the parameter value is with respect to the observation; and further the position function must be transformation-invariant with respect to the group that defines the transformation-parameter model. In such a model there is a well-defined frequency distribution for parameter position relative to corresponding observation. If this distribution is located at a particular observation, it yields the fiducial distribution of possible parameter values for that observation.

2. With Covariance Matrix of Known Form

A multivariate generalization of the ordinary t-distribution was developed by Cornish (1954). For the standard form of this distribution let \((x_1, \ldots, x_p)\) be a sample from a normal distribution of zero mean and variance \(\sigma^2\) and let \(s^2\) be an independent estimate of \(\sigma^2\) based on \(n\) degrees of freedom. Then the standard multivariate t-distribution is the distribution of \((t_1, \ldots, t_p)\) where

\[ t_i = x_i / s \]

and it has probability element

\[ \frac{\Gamma\left[\frac{1}{2}(n + p)\right]}{\Gamma\left[\frac{1}{2}n\right]} \left(\frac{\sum x^2}{n}\right)^{-\frac{1}{2}(n+p)} \Pi(dt_i / \sqrt{n}). \]  

(2.1)
The more general multivariate t-distribution is simply related to this standard distribution by linear transformations. Let \( y = (y_1, \ldots, y_p)' \) have a multivariate normal distribution with zero means and non-singular covariance matrix \( \sigma^2 W = \sigma^2 \Lambda \Lambda' \). This admits representation in terms of the preceding variables

\[
y = Ax, \quad x = A^{-1}y.
\]

The distribution of the new t-statistics \( u = (u_1, \ldots, u_p), \)

\[
u_i = y_i / s,
\]
is obtained from the transformation \( u = At \)

\[
\frac{\Gamma\left[\frac{1}{2}(n + p)\right]}{\Gamma\left[\frac{1}{2}n\right] \pi^{1p}} \left(1 + \frac{u'W^{-1}u}{n}\right)^{-\frac{1}{2}(n+p)} |W|^{-\frac{1}{2}p} \Pi(d_{ui}/\sqrt{n}). \tag{2-2}
\]

For an application, this could give the distribution of the t-variables for the regression coefficients in a regression model with normally distributed errors.

The multivariate t-distribution provides a basis for deriving a fiducial distribution (Cornish, 1960). Let \( x_1, \ldots, x_p \) be independent normal variables with means \( \mu_1, \ldots, \mu_p \) and common variance \( \sigma^2 \), and let \( s^2 \) be an independent estimate of variance based on \( n \) degrees of freedom. The variables \( (t_1, \ldots, t_p) \)

\[
t_i = (x_i - \mu_i) / s
\]
have a multivariate t-distribution. With \( (t_1, \ldots, t_p) \) as the pivotal quantity the standard fiducial inversion expresses \( (\mu_1, \ldots, \mu_p) \) as a relocated multivariate t-distribution

\[
\frac{\Gamma\left[\frac{1}{2}(n + p)\right]}{\Gamma\left[\frac{1}{2}n\right] n^{1p}} \left(1 + \frac{\Sigma(x_i - \mu_i)^2}{s^2n}\right)^{-\frac{1}{2}p} s^{-p} \Pi(d_{ui}/\sqrt{n}). \tag{2-3}
\]
The more general case in which the variables \( (y_1, \ldots, y_n) \), have a covariance matrix

\[
\sigma^2 W = \sigma^2 \Lambda \Lambda' \]
can be treated by reference to the standardized variables \( x = A^{-1}y \) and the resulting distribution is analogous in form to that above but is in terms of the more general multivariate t-distribution.

The transformation-parameter method is applicable to this statistical model. The appropriate group produces changes in unit of measurement for \( x_1, \ldots, x_p \), \( s \) and changes in location for each of \( x_1, \ldots, x_p \)

\[
x_i^* = a_i + bx_i, \quad s^* = bs,
\]
or equivalently

\[
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 0 \\
0 & \cdots & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^* \\
\vdots \\
x_p^* \\
\end{bmatrix} =
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 0 \\
0 & \cdots & 1 \\
\end{bmatrix}
\begin{bmatrix}
a_1 \\
\vdots \\
a_p \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
\vdots \\
x_p \\
s \\
\end{bmatrix}
\]

The lower triangular matrices with positive \( b \)'s and real \( a \)'s form a group.
Let \( x_1, \ldots, x_p, s \) designate the general variables with parameter \( \mu_1, \ldots, \mu_p, \sigma \). A transformation using \( \mu_1, \ldots, \mu_p, \sigma \) can generate such variables from the special case with means equal to zero and standard deviation equal to one

\[
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
x_1 & \cdots & x_p
\end{bmatrix}
= \begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
\mu_1 & \cdots & \mu_p
\end{bmatrix}
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
\sigma & \cdots & \sigma
\end{bmatrix}
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
z_1 & \cdots & z_p
\end{bmatrix}
\chi/\sqrt{n},
\]

\((z_1, \ldots, z_p)\) is a standard normal sample and \( \chi \) is chi on \( n \) degrees of freedom. There is a fixed frequency distribution for parameter-variable position and it can be described explicitly by solving the preceding equation for the inverse of the right pivotal matrix. The fiducial distribution can be obtained by relocating at a particular outcome \((x_1, \ldots, x_p, s)\) this fixed distribution for parameter-variable position

\[
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
\mu_1 & \cdots & \mu_p
\end{bmatrix}
= \begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
x_1 & \cdots & x_p
\end{bmatrix}
\begin{bmatrix}
1 & \cdots & 0 \\
0 & \ddots & 1 \\
\sqrt{n}z_1/\chi & \cdots & -\sqrt{n}z_p/\chi
\end{bmatrix}
\sqrt{n/\chi},
\]

The fiducial probability element, obtained by calculating the Jacobian of the above transformation, is

\[
\prod_{i=1}^{p} \frac{\left(\frac{x_i - \mu_i}{\sigma}\right)}{\sigma} \chi \left(\frac{\sqrt{n}s}{\sigma}\right)^{1/2} d\mu_i \, d\sigma,
\]

where \( n(z) \) and \( \chi(\chi) \) designate the normal and chi density functions. The more general fiducial distribution for the regression model may be found in Fraser (1961a).

The marginal distribution for the \( \mu \)'s can be obtained by integration. The representation

\[
\mu_i = x_i - \sqrt{n}z_i/\chi
\]

shows, however, that they have the relocated and rescaled multivariate \( t \)-distribution obtained by Cornish. A frequency interpretation is thus available for Cornish's fiducial distribution.

Cornish (1960) gives a numerical example involving two means. An elliptical fiducial region is calculated with boundary having constant fiducial density. Also an approximate confidence region is given which is based on a method by Dunnett. It seems of interest to note that both regions are simultaneously fiducial regions and confidence regions. From the fiducial viewpoint there is no need to restrict attention to the natural elliptical region and any meaningful region having the desired probability can be chosen.

Cornish (1962) gives a similar example involving two location parameters and records two regions which are referred to as fiducial regions. The example involves a covariance matrix of known form; the essentials, however, can be discussed in terms of canonical variables. Let \( x_1, \ldots, x_p \) be independent normal variables with means \( \mu_1, \ldots, \mu_p \) and common variance \( \sigma^2 \), and let \( s^2 \) be an independent estimate of variance based on \( n \) degrees of freedom. The equations

\[
\frac{\mu_i - x_i}{s} = t_i = -\frac{\sqrt{n}z_i}{\chi}
\]

provide a representation for the fiducial distribution of \((\mu_1, \ldots, \mu_p)\). From this it follows that the probability for

\[
\frac{\sum(\mu_i - x_i)^2}{ns^2} = \frac{\chi^2}{\chi^2} \leq \frac{p}{n} F
\]
is given by $F$-tables with $p$ over $n$ degrees of freedom. This inequality describes the first region given by Cornish.

A second region is obtained by reference to Hotelling’s $T^2$ distribution. The probability that

$$T^2 = \frac{\sum (\mu_i - x_i)^2}{n \sigma^2} \leq \frac{p}{n - p + 1} F$$

is obtained from $F$-tables with $p$ over $n - p + 1$ degrees of freedom. The resulting elliptical region contains the first region. This second region appears to be based on an inappropriate use of Hotelling’s distribution. Hotelling’s variable and the derivation of the associated distribution involves a sample covariance matrix having a Wishart distribution. Only in this case is there a reduction in the denominator degrees of freedom by $p - 1$. This second region thus seems to be invalid.

3. Inference for regression and scale parameters

In a recent paper, Verhagen (1961) has extended the proper regions and fiducial functions of Pitman (1938) to a regression model with a fixed error distribution involving a single scale parameter. Such a model can be put into transformation-parameter form using a seemingly natural group, and the resulting fiducial distribution is that developed by Verhagen. This alternative framework however provides a frequency interpretation for Verhagen’s distribution, and it makes the restriction to proper regions irrelevant—unless a confidence interpretation is wanted.

Let $(y_1, \ldots, y_n)$ be the response-variable vector with a distribution located at the vector point $\sum_{i=1}^{p} \beta_i (x_{i1}, \ldots, x_{in})$ and with an additive error pattern given by the vector $\sigma (u_1, \ldots, u_n)$ where the underlying error variable $(u_1, \ldots, u_n)$ has a fixed distribution with density $f(.)$ on Euclidean $n$-space. The group of transformations can be expressed easily in matrix notation by using a somewhat redundant notation for the variable $(y_1, \ldots, y_n)$. Designate this response variable by the matrix

$$Y = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pn} \\ y_1 & \cdots & y_n \end{bmatrix}$$

and consider transformations of the form

$$GY = \begin{bmatrix} 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \\ \alpha_1 & \cdots & \alpha_p & c \end{bmatrix} Y$$

with $c$ positive and the $\alpha$’s real. These transformations rescale the variables and alter them in a pattern based on the structural vectors in the model. The lower triangular matrices $G$ form a group under matrix multiplication.

The general distribution for $Y$ can be expressed uniquely in terms of a transformation from the standard case with $\beta_i = 0$ and $\sigma = 1$. Let

$$\theta = \begin{bmatrix} 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \\ \beta_1 & \cdots & \beta_p & \sigma \end{bmatrix}, \quad U = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pn} \\ u_1 & \cdots & u_n \end{bmatrix}.$$

Then $Y = \theta U$ describes the distribution for $Y$. This is a transformation-parameter model.
Let \( b_1, \ldots, b_p \) be the regression coefficients and error standard deviation as obtained by a least-squares analysis on \( \mathbf{Y} \), and let \( \mathbf{T}_Y \) be defined by

\[
\mathbf{T}_Y = \begin{bmatrix}
1 & 0 & 0 \\
0 & \ddots & 0 \\
b_1 & \ldots & b_p & s
\end{bmatrix},
\]

when necessary, dependence on the vector \( \mathbf{Y} \) can be indicated by \( b_j = b_j(\mathbf{Y}) \) and \( s = s(\mathbf{Y}) \). The statistic \( \mathbf{A}_Y = \mathbf{T}_Y^{-1}\mathbf{Y} \) defines the orbits under the group; it is an ancillary statistic. The statistic \( \mathbf{T}_Y \) is conditionally sufficient given the ancillary \( \mathbf{A}_Y \).

The probability element for the response variable \( \mathbf{Y} \) is

\[
f((\theta^{-1}\mathbf{Y})) \sigma^{-n} d\mathbf{Y} = f((\theta^{-1}\mathbf{Y})) (s_\mathbf{Y}/\sigma)^n d\mathbf{Y}/s_\mathbf{Y}^n.
\]

In the rearranged form the measure element \( d\mathbf{Y}/s_\mathbf{Y}^n \) is invariant under the transformation group.

The conditional relative frequency on an orbit is obtained by using the invariant density function \( f((\theta^{-1}\mathbf{Y})) (s_\mathbf{Y}/\sigma)^n \) in conjunction with the invariant measure on the orbit; it has probability element

\[
w((\mathbf{T}_Y^{-1}\mathbf{Y})) f((\theta^{-1}\mathbf{Y})) (s_\mathbf{Y}/\sigma)^n d\mathbf{T}_Y/s_\mathbf{Y}^n,
\]

where the measure element \( d\mathbf{T}_Y/s_\mathbf{Y}^n \) is the left invariant measure on the group, and \( w((\mathbf{T}^{-1}\mathbf{Y})) \) is a normalizing constant that depends on the ancillary statistic. The fiducial density element is then obtained from the formula in Fraser (1961b, §8):

\[
w((\mathbf{T}_Y^{-1}\mathbf{Y})) f((\theta^{-1}\mathbf{Y})) (s_\mathbf{Y}/\sigma)^n (s_\mathbf{Y})^{-p} d\mathbf{T}_Y/\sigma^{n+1}
\]

\[
= s_\mathbf{Y}^{-p} w((\mathbf{T}^{-1}\mathbf{Y})) f((\theta^{-1}\mathbf{Y})) \sigma^{-(n+1)} d\theta
\]

\[
= s_\mathbf{Y}^{-p} w((\mathbf{T}^{-1}\mathbf{Y})) \sum_i \left( \frac{y_i - \sum_j x_{ij} \beta_j}{\sigma} \right) \Pi d\beta_i d\sigma.
\]

This is the formula (10) in Verhagen.

4. WITH UNKNOWN COVARIANCE MATRIX

Let \( (x_1, \ldots, x_p) \) have a multivariate normal distribution with means \( (\mu_1, \ldots, \mu_p) \) and covariance matrix \( \mathbf{S} \) and let \( \mathbf{S} \) be an inner product matrix having a Wishart distribution for \( \mathbf{S} \) based on \( n \) degrees of freedom. For the case \( p = 2 \), Fisher (1954) presents a fiducial distribution for \( (\mu_1, \mu_2) \)

\[
\frac{d\mu_1 d\mu_2}{2\pi |\mathbf{S}|^{\frac{1}{2}} [1 + \Sigma(x_i - \mu_i) (x_j - \mu_j) S^{ij}]^{\frac{k}{2(n+2)}}},
\]

where \( S^{ij} = S^{-1} \). This result is based on the probability statements

\[
\Pr \left\{ \lambda(x_1 - \mu_1) + \mu(x_2 - \mu_2) < t \left[ \frac{\lambda^2 S_{11} + 2\lambda \mu S_{12} + \mu^2 S_{22}}{n} \right] \right\} = P(t),
\]

where \( P(t) \) is the distribution function for the \( t \)-distribution with \( n \) degrees of freedom.

Cornish (1961) derives the generalization of this for \( p > 2 \)

\[
\frac{\Gamma(\frac{1}{2}(n+p))}{\Gamma(\frac{1}{2}n)^{\frac{1}{2}(p+1)}} \left[ 1 + (\mu - \mathbf{x})' S^{-1}(\mu - \mathbf{x}) \right]^{-\frac{k}{2(n+p)}} \left[ \frac{S}{n} \right]^{-\frac{1}{2}} \Pi (d\mu_i/\sqrt{n}). \tag{4.1}
\]
This is a relocated multivariate t-distribution, and is based on the probability statements

\[
\Pr \left( \lambda' (\mu - x) < t (\frac{\lambda' S \lambda}{n})^{\frac{1}{2}} \right) = P(t).
\]

(4.2)

Consider the following justification of the above result. The fiducial inversion is performed by finding a joint distribution for \((\mu_1, \ldots, \mu_n)\) which agrees with the formula (4.2) as interpreted with \((\mu_1, \ldots, \mu_n)\) as a variable. This is equivalent to treating

\[
\frac{\lambda' (\mu - x) n^{\frac{1}{2}}}{(\lambda' S \lambda)^{\frac{1}{2}}}
\]

as a t-variable derived from a distribution for \(\mu\) or to treating \(\lambda' (\mu - x)\) as a t-variable derived from a distribution for \(\mu\) and rescaled by the fixed factor \((n/\lambda' S \lambda)^{\frac{1}{2}}\). But the marginal distribution of the linear compound \(\lambda' (\mu - x)\) as derived from the multivariate t-distribution for \((\mu - x)\) with fixed matrix \(W = S n^{-1}\) is also a t-distribution rescaled by the factor \((n/\lambda' S \lambda)^{\frac{1}{2}}\). From probability theory it is known that a joint distribution is determined when the marginal distributions of all linear compounds are fixed. Thus assuming that \(\mu - x\) has the preceding multivariate t-distribution or that \(\mu\) has a relocation of that distribution as given by formula (4.1) gives results consistent with the probability statements (4.2). It follows that if there exists a fiducial distribution for \(\mu\) then formulas (4.2) require it to be the multivariate t-distribution just described.

It is of interest to note that the preceding argument can be used verbatim to derive the fiducial distribution (2.3) in the case of a covariance matrix of known form. The initiating statement treats \(\lambda' (\mu - x)\) as a t-variable rescaled by the factor \((1/\lambda' S \lambda)^{\frac{1}{2}}\). But clearly the two cases have essential differences in their structure and it is perhaps suspicious that they lead to the same fiducial distributions.

Consider now some related aspects of this derivation. The inequalities for the original variables can be examined in pairs

\[
\begin{align*}
&\lambda' (\mu - x) \leq t_1 (\lambda' S \lambda/n)^{\frac{1}{2}} \quad \lambda' S \lambda \text{ are statistically independent;} \\
&\tau' (\mu - x) \leq t_2 (\tau' S \tau/n)^{\frac{1}{2}} \quad \tau' S \tau \text{ are statistically independent.}
\end{align*}
\]

(4.3)

The probabilities for such pairs of inequalities can be calculated from the normal and Wishart distributions for \(x\) and \(S\). However, certain of these probabilities can be calculated more simply. Let \(\lambda\) and \(\tau\) be chosen so that \(\lambda' (\mu - x)\) and \(\tau' (\mu - x)\) are statistically independent; then \(\lambda' S \lambda\) and \(\tau' S \tau\) are statistically independent. In this case the joint probability can be calculated on the basis of independent t-variables. Thus certain pairs of t-variables have a joint distribution that is based on independent component variables.

Consider these probability statements when \((\mu_1, \ldots, \mu_p)\) is treated as the variable. From this point of view \(\lambda' \mu\) and \(\tau' \mu\) are statistically independent. But an examination of the multivariate t-distribution discloses that it does not admit factorization so that two linear combinations of the variables are statistically independent. Thus the multivariate t-distribution does not conform to the probability statements on pairs of inequalities (4.3). It follows then that there does not exist a joint distribution for \((\mu_1, \ldots, \mu_n)\) that conforms to the probability statements on pairs of inequalities (4.3).

It thus seems that the Fisher–Cornish distribution is not valid—that it is inconsistent with basic rules for deriving fiducial distributions.
Cornish (1962) considers elliptical regions based on the distribution (4.2)

\[(\mu - x)' S^{-1}(\mu - x) \leq \frac{p}{n} F\]

with \(F\) based on \(p\) over \(n\) degrees of freedom. Cornish also considers elliptical confidence regions based on Hotelling’s distribution

\[(\mu - x') S^{-1}(\mu - x) \leq \frac{p}{n-p+1} F'\]

with \(F'\) based on \(p\) over \(n-p+1\) degrees of freedom. For the same probability level the first ellipsoid is contained within the second. The relationship between \(\mu\) and \(x\) is a position relationship; the discrepancy between the two ellipsoids is a major conflict. However, the use of Hotelling’s distribution has direct justification based on \(T^2\) as pivotal quantity; there would thus seem to be good grounds here for suspicion of the first kind of region and the multivariate \(t\)-distribution on which it is based.

These observations seem to discredit the Cornish–Fisher distribution and remove the theoretical grounds for the concluding remarks in Cornish (1962): ‘Similar remarks apply to other situations in which Hotelling’s distribution has hitherto provided the test of significance; in all cases Hotelling’s test is now superseded by the test based on the simultaneous fiducial distribution . . . .’

Fiducial distributions for the parameter \(\mu, \Sigma\) can be derived within the transformation-parameter framework. For this a position relationship between variable and parameter is required. There are, however, a variety of such relationships, a few of which have a certain naturalness in terms of the identity of the co-ordinate variables (Fraser, 1961a). As an example a transformation group could produce location and scale changes on the first co-ordinate, location and scale changes on the second co-ordinate coupled with a regression change on the first co-ordinate, and so on. This can be represented in terms of triangular matrices and leads to fiducial distributions of the form considered in Mauldon (1955). The resulting fiducial distribution is asymmetric in the variances of the different co-ordinates. Alternatively, if the co-ordinates were structured in a different order by a regression group a different fiducial distribution would be obtained—based essentially on a permutation of co-ordinates. Rather than leading to a paradox as indicated by Mauldon, this would seem to emphasize the dependence of the fiducial argument on the additional structure that provides the position relationship between variable and parameter.

In Statistical Methods and Scientific Inference Fisher develops a fiducial distribution for \(\mu, \Sigma\) in the bivariate case. The distribution is symmetric in the variances of the two co-ordinates. Consequently it is different from the transformation-parameter distributions just mentioned. In a situation where the regression structure is a meaningful part of the model and the transformation-parameter framework is acceptable, the distribution given by Fisher would seem to be inappropriate and lacking in validity.
REFERENCES


