Structural Probability and Prediction for the Multivariate Model

By D. A. S. Fraser and M. Safiul Haq†

University of Toronto

[Received October 1966. Final revision November 1968]

SUMMARY

The multivariate normal distribution can be used to describe the response variable of a system. A more comprehensive multivariate model is described in this paper: it has a distribution describing an error variable internal to a system, with a known multivariate distribution; and it has a positive affine transformation, the physical quantity, which generates a response vector from an error vector. This more comprehensive model is a structural model and it provides structural probability statements concerning the physical quantity.

Error and structural distributions are derived for the multivariate model. The structural distribution for a quantity can be used to generate structural prediction distributions: various prediction distributions are obtained for the multivariate model. The results are specialized to cover the case of the multivariate normal structural model.

The classical multivariate normal has been analysed by Bayesian methods (Geisser and Cornfield, 1963). The more comprehensive multivariate structural model does not need the use of subjective methods.

1. INTRODUCTION

CONSIDER a system operating under stable conditions. Suppose that experience has led to the identification of internal sources of variation and to the form of this variation as it affects the response variable of the system. Let E be an error variable describing the internal sources of variation and having a known distribution on a space \( \mathcal{X} \).

Suppose the general characteristics of the system are given by a quantity \( \theta \), a transformation belonging to a group \( G \) of transformations unitary on \( \mathcal{X} \) (\( G \) is unitary on \( \mathcal{X} \) if \( g_1 X = g_2 X \) for some \( X \) in \( \mathcal{X} \) and \( g_1 \) and \( g_2 \) in \( G \) implies \( g_1 = g_2 \)). The quantity \( \theta \) applied to a value \( E \) from the error distribution gives a value \( X \) for the response \( X = \theta E \).

Now suppose: \( \mathcal{X} \) is an open set in \( \mathbb{R}^N \); \( G \) is an open set in \( \mathbb{R}^L \); the transformations \( g = hg \), \( X = hgX \) are continuously differentiable with respect to \( g \), \( h \) and \( X \); and \( [X] \) is a continuous transformation variable (a transformation variable is a function from \( \mathcal{X} \) to \( G \) such that \( [gX] = g[X] \) for all \( g \) and \( X \)). Also suppose that the error variable \( E \) has a density function \( f \) with respect to an invariant measure element \( m \) on \( \mathcal{X} \):

\[
f(E) \, dm(E).
\]

Let \( \mu \) and \( \nu \) be the left and right invariant error measures on \( G \) and \( \Delta \) be the corresponding modular function:

\[
d\mu(g) = \Delta(g) \, d\nu(g) = d\mu(g^{-1}).
\]

† Present address: University of Western Ontario.
The description of the system gives the following structural model:

\[ f(E) \, dm(E), \quad X = \theta E. \]  

(3)

The model has two parts: an error distribution on \( \mathcal{F} \) (\( E \) is a variable); and a structural equation \( X = \theta E \) in which a realized value \( E \) from the error distribution gives the relation between the known observation \( X \) and the unknown quantity \( \theta \) in \( G \) (\( E \) is constant; \( X \) is known, \( \theta \) and \( E \) are unknown).

Consider the information in the structural equation concerning \( E \). The orbit of \( E \) can be calculated from the known \( X \):

\[ GE = \{ gE: g \in G \} = \{ g \theta^{-1} X: g \in G \} = GX. \]  

(4)

But the position of \( E \) on its orbit cannot be calculated. Position can be described by the transformation \([X]\),

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

where \( D(X) \) is a fixed reference point on the orbit \( GX \). The structural equation gives \([E] = \theta^{-1}[X]\);

it represents \([E]\) as an unknown transformation \( g = \theta^{-1}[X] \) in \( G \). The structural equation gives no information concerning the error position \([E]\); for with a different value of \([X]\) the error position \([E]\) is still represented as an unknown transformation in \( G \).

Sufficient conditions for classical probability statements concerning unknown constants are: (i) the constants were generated as realized values from a random process with known probability characteristics; (ii) the only other information concerning the unknown constants has the form of an event for the random process that generated the constants.

Consider further the structural model and suppose there is no outside information concerning the quantity \( \theta \). This can occur minimally if the structural model is being examined in isolation to see what information it alone supplies concerning \( \theta \). The information concerning the error value \( E \) has the form of an event:

\[ GE = GX. \]

The conditions are fulfilled for making probability statements concerning the unknown error value \( E \); the probability statements are based on the conditional distribution of the error variable \( E \) given the event \( GE = GX \).

The conditional distribution of \( E \) given the orbit \( GE \) can be derived from invariance properties. For notation it is convenient to label the orbits by means of the reference points: \( GX \mapsto D(X) \). The conditional distribution of \( E \) given \( D(E) = D \) has probability element

\[ k(D)f([E] \, D) \, d\mu([E]) \]  

(5)

in terms of position \([E]\) on the orbit.

The information in the structural model gives the following reduced model

\[ k(D(X))f([E] \, D(X)) \, d\mu([E]), \quad [X] = \theta[E]. \]  

(6)

The reduced model has two parts: an error probability distribution \((|E|) \) is a variable); and a structural equation \([X] = \theta[E]\) in which \([X]\) is a known constant and \( \theta \) and \([E]\)
are unknown constants. The error probability distribution provides probability statements concerning the unknown constant \([E]\) in the structural equation.

A probability statement concerning the unknown \([E]\) is *ipso facto* a probability statement concerning the unknown \(\theta\).

\[
[X] = \theta[E],
\]

where \([X]\) is fixed and known; and in general the error probability distribution describing the unknown \([E]\) is *ipso facto* a probability distribution describing the unknown \(\theta\), the structural distribution:

\[
k(D(X)) f(\theta^{-1} X) \Delta(\theta^{-1}[X]) \, d\mu(\theta). \tag{7}
\]

Consider now the prediction of a future variable \(y\) with distribution

\[
p(y; \theta) \, dn(y). \tag{8}
\]

The joint probability element for \(\theta\) and \(y\) is

\[
k(D) f(\theta^{-1} X) \Delta(\theta^{-1}[X]) p(y; \theta) \, d\mu(\theta) \, dn(y) \tag{9}
\]

and the marginal structural distribution for \(y\) is

\[
k(D) \Delta([X]) \int_G f(h^{-1} X) p(y; h) \Delta(h^{-1}) \, d\mu(h) \, dn(y). \tag{10}
\]

Now consider a future variable \(y\) generated by the transformation \(\theta\) applied to an error variable \(E^*\) with distribution

\[
p(E^*) \, dn(E^*) \tag{11}
\]

relative to an invariant measure \(n\). The future variable can be expressed directly in terms of the error variables \([E]\) and \(E^*\):

\[
y = [X][E]^{-1} E^*. \tag{12}
\]

The structural distribution of \(y\) can then be obtained by integration. The joint element for \([E]\) and \(E^*\) is

\[
k(D) f([E] D) p(E^*) \, d\mu([E] E^*) \, dn(E^*); \tag{13}
\]

the joint element for \([E]\) and \(y\) is

\[
k(D) f([E] D) p([E] [X]^{-1} y) \, d\mu([E] E^*) \, dn(y); \tag{14}
\]

the joint element for \(h = [X][E]^{-1}\) and \(y\) is

\[
k(D) f(h^{-1} X) p(h^{-1} y) \Delta(h^{-1}[X]) \, d\mu(h) \, dn(y). \tag{15}
\]

The marginal distribution of \(y\) is obtained by integrating with respect to \(h\); it agrees with the expression (10) obtained directly.

For probability statements about future variables Fisher (1959, p. 114) used pivotal quantities involving the future variable and the statistic in hand. For a sample of size \(n\) from a normal distribution he used \(t = (y - \bar{x})/s(1 + n^{-1})^{1/2}\) as the pivotal quantity (\(\bar{x}\) is the sample mean, \(s^2\) is the unbiased estimate of the population variance and \(y\) is the future observation).
2. THE MULTIVARIATE ERROR MODEL

2.1. The Model

Let \((e_1, ..., e_k)\) be an error variable with probability element

\[ h(e_1, ..., e_k) de_1 ... de_k: \]

let

\[
\theta = \begin{pmatrix}
1 & 0 & 0 \\
\mu_1 & \gamma_{11} & \cdots & \gamma_{1k} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_k & \gamma_{k1} & \cdots & \gamma_{kk}
\end{pmatrix}
\]  

be a transformation matrix with positive determinant; and suppose a response value \((x_1, ..., x_n)\) is obtained by the transformation \(\theta\) applied to an error value:

\[
\begin{pmatrix}
x_1 \\
\vdots \\
x_k
\end{pmatrix} = \theta
\begin{pmatrix}
e_1 \\
\vdots \\
e_k
\end{pmatrix}.
\]

(16)

For \(n\) observations let

\[
E = \begin{pmatrix}
1 & \ldots & 1 \\
e_{11} & \ldots & e_{1n} \\
\vdots & \ddots & \vdots \\
e_{k1} & \ldots & e_{kn}
\end{pmatrix} = \begin{pmatrix}
1' \\
e_1' \\
\vdots \\
e_k'
\end{pmatrix}
\]

be the composite error variable with probability element

\[
f(E) dE = \prod_{j=1}^{n} h(e_{1j}, ..., e_{kj}) \prod de_{ij},
\]

and let

\[
X = \begin{pmatrix}
1 & \ldots & 1 \\
x_{11} & \ldots & x_{1n} \\
\vdots & \ddots & \vdots \\
x_{k1} & \ldots & x_{kn}
\end{pmatrix} = \begin{pmatrix}
1' \\
x_1' \\
\vdots \\
x_k'
\end{pmatrix}
\]

be the composite response observation. The structural model for the sample is then

\[
f(E) dE, \quad X = \theta E,
\]

(17)

the multivariate error model.
2.2. Orbits

As a step towards defining orbits let $V(X)$ be the $k+1$ dimensional linear subspace in $R^n$ generated by the row vectors,

$$1' = (1, \ldots, 1),$$
$$x_1' = (x_{11}, \ldots, x_{1n}),$$
$$\vdots$$
$$x_k' = (x_{k1}, \ldots, x_{kn})$$

of the matrix $X$. The case with linear dependence is ignored as a degenerate set of points. Let $D(X)$ be a matrix of row vectors,

$$D(X) = \begin{pmatrix} 1 & \ldots & 1 \\ a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{k1} & \ldots & a_{kn} \end{pmatrix} = \begin{pmatrix} 1' \\ a_1' \\ \vdots \\ a_k' \end{pmatrix}$$

(18)

with the properties:

(i) the row vectors in $D(X)$ generate $V(X)$ and have the same orientation as the row vectors in $X$ (connected by a transformation with positive determinant), but are otherwise independent of $X$;

(ii) the row vectors of $D(X)$ are orthogonal;

(iii) the row vectors $a'_1, \ldots, a'_k$ have unit length.

A matrix with these properties can be obtained by the following procedure:

(i) project $(1,0,\ldots,0)$ into the subspace $V(X)$, orthogonalize to $1' = (1, \ldots, 1)'$, normalize, and obtain $(a_{11}, \ldots, a_{1n})$;

(ii) project $(0,1,0,\ldots,0)$ into the subspace $V(X)$, orthogonalize to $a_1$, normalize, and obtain $(a_{21}, \ldots, a_{2n})$;

and so on.

Now let $O_X$ be the orthogonal rotation (positive determinant) in $R^{k+1}$,

$$O_X = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & o_{11} & \ldots & o_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & o_{k1} & \ldots & o_{kk} \end{pmatrix}$$

(19)

which, as a linear operator on the row vectors in $D(X)$, carries:

(i) $a_1$ into the unit residual vector for $x_1$ regressed on $1'$;

(ii) $a_2$ into the unit residual vector for $x_2$ regressed on $1, a_1$;

(iii) $a_k$ into the unit residual vector for $x_k$ regressed on $1, a_1, a_2$;

and so on. Let $T_X$ be the positive lower triangular matrix,

$$T_X = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \tilde{x}_1 & s_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_k & b_{k1} & \ldots & s_k \end{pmatrix},$$
whose rows contain the regression coefficients and residual lengths for the regressions in (i), (ii), .... The position transformation \([X]\) can then be given as

\[
[X] = T_X O_X.
\] (20)

The definition of \(T\) and \(O\) shows that \(X = [X]D(X)\) and \([X]^{-1}X = D(X)\); and it shows that \([X]\) is unique for given reference point \(D(X)\).

### 2.3. The Invariant Differential

The structural parameter \(\theta\) takes values in the positive affine group

\[
G = \left\{ g = \begin{pmatrix}
1 & 0 & \ldots & 0 \\
a_1 & c_{11} & \ldots & c_{1k} \\
\vdots & \vdots & \ddots & \vdots \\
a_k & c_{k1} & \ldots & c_{kk}
\end{pmatrix} : |g| > 0 \right\}.
\]

A transformation on the sample space has the form

\[
X^* = gX
\]

and is diagonal column by column. Accordingly,

\[
\left| \frac{\partial gX}{\partial X} \right| = |g|^n,
\]

and an invariant differential on the sample space \(R^{kn}\) is

\[
\frac{dX}{[X]^n}.
\] (21)

A left transformation on the group is also diagonal by columns; accordingly

\[
\left| \frac{\partial h}{\partial g} \right| = |h|^{k+1},
\]

and the left invariant differential is

\[
d\mu(g) = \frac{dg}{|g|^{k+1}}.
\] (22)

A right transformation on the group is diagonal by rows; for example,

\[
(a_1^*, c_{i1}^*, \ldots, c_{ik}^*) = (a_i, c_{i1}, \ldots, c_{ik})h
\]

with Jacobian \(|h|\). Accordingly

\[
\left| \frac{\partial h}{\partial g} \right| = |h|^k,
\]

and the right invariant differential is

\[
d\nu(g) = \frac{dg}{|g|^k}.
\] (23)

The modular function is then

\[
\Delta(g) = \frac{\partial \mu}{\partial \nu} = |g|^{-1}.
\] (24)
2.4. The Probability Elements

The probability element for the error variable can be put in invariant form:

\[ f(E) \, dE = f(E) \left| [E] \right|^n \times \frac{dE}{\left| [E] \right|^n}. \]  

(25)

The probability element for the response variable is then

\[ f(\theta^{-1} X) \frac{\left| [X] \right|^n}{\left| \theta \right|^n} \times \frac{dX}{\left| [X] \right|^n}. \]  

(26)

The conditional probability element for the position \([X]\) given \(D(X)\) is

\[ k(D) f(\theta^{-1} [X] D) \frac{\left| [X] \right|^n}{\left| \theta \right|^n} \times \frac{d[X]}{\left| [X] \right|^k}. \]  

(27)

and the structural probability element for \(\theta\) is

\[ k(D) f(\theta^{-1} X) \frac{\left| [X] \right|^n}{\left| \theta \right|^n} \times \frac{\left| \theta \right|^n}{\left| [X] \right|^n} \times \frac{d\theta}{\left| \theta \right|^k}. \]  

(28)

The structural probability element for a future variable \(y\) with model \(p(y; \theta)\) is

\[ k(D) \int_G f(h^{-1} X) p(y; h) \frac{\left| [X] \right|^{n-1}}{h} \times d(h^{-1}) \, dn(y). \]  

(29)

If the \(X^*\) stands for a future observation matrix based on \(n'\) observations from the same multivariate model, the prediction probability element for \([X^*]\) can be written as

\[ k_n(D) k_{n'}(D^*) \int_{h \in G} f_n(h^{-1} X) f_{n'}(h^{-1} X^*) \frac{\left| [X] \right|^{n-1}}{h} \times \frac{\left| [X^*] \right|^{n'}}{h} \times \frac{d[h]}{h} \, d[X^*]. \]  

(30)

where \(D^*\) is the orbit for the future observations.

3. Rotationally Symmetric Error

Consider the multivariate error model described in the preceding section and suppose that the latent error distribution is rotationally symmetric

\[ f(E) = f(O E). \]

for all orthogonal rotations of the form (19). The transformation \(\theta\) can be factored as in (20): \(\theta = T_\theta O_\theta\). In this form the parameter component \(O_\theta\) rotates the error variable; it does not alter the distribution. Accordingly it is redundant.

3.1. The Left Invariant Differential

Let the transformation \([X]\) be developed uniquely as a product,

\[ [X] = T_x O_x. \]

The same factorization is available for a general group element

\[ g = T_g O_g. \]

Consider a differential element in the neighbourhood of the identity element on \(G\). Under a transformation \(O\) the differential element \(dO\) is invariant.
The subsequent transformation $T$ is a lower triangular matrix with positive diagonal elements, and 1 in the top left corner. These matrices form a group. The Jacobian of a left transformation using $T$ is

$$t_{11} \ldots t_{kk+1} = \left| T \right|_\Delta = t_{10}^2 t_{21}^2 \ldots t_{kk+1}^2,$$

the increasing determinant; and the left invariant differential is

$$\frac{dT}{\left| T \right|_\Delta}.$$  \hspace{1cm} (31)

Accordingly

$$\frac{dT dO}{\left| T \right|_\Delta},$$  \hspace{1cm} (32)

is invariant for the affine group $G$. The invariant measure is essentially unique; hence

$$\frac{dg}{\left| g \right|^{k+1}} = \frac{dT dO}{\left| T \right|_\Delta},$$  \hspace{1cm} (33)

(the constant of proportionality is 1).

### 3.2. The Structural Distribution

The structural distribution for the group element $\theta$ can be integrated to obtain the marginal distribution for the essential parameter $T_\theta$:

$$\int k(D)f(O_\theta^{-1}T_\theta^{-1}X) \frac{\left| X \right|^n}{\left| T_\theta \right|^n} \times \frac{\left| T_\theta \right|}{\left| X \right|} \frac{dT_\theta dO_\theta}{\left| T_\theta \right|_\Delta}$$

$$= k(D)f(T_\theta^{-1}X) \frac{\left| X \right|^n}{\left| T_\theta \right|^n} \times \frac{\left| T_\theta \right|}{\left| X \right|} \frac{dT_\theta}{\left| T_\theta \right|_\Delta} \int dO_\theta$$

$$= \prod_{t=2}^k A_t k(D)f(T_\theta^{-1}X) \frac{\left| X \right|^n}{\left| T_\theta \right|^n} \times \frac{\left| T_\theta \right|}{\left| X \right|} \frac{dT_\theta}{\left| T_\theta \right|_\Delta}.$$  \hspace{1cm} (34)

In this formula the notation

$$A_t = \frac{2\pi^{l/2}}{\Gamma(l/2)}$$

is used for the area of the unit sphere in $R^l$. The distribution (34) arises as a posterior distribution in Bayesian Theory; the corresponding prior is $\left| T_\theta \right|/\left| T_\theta \right|_\Delta dT_\theta$.

For the structural model generated from a rotationally symmetric error distribution, a convenient parameter is

$$\Sigma_\theta = \theta \theta' = T_\theta O_\theta O_\theta' T_\theta'$$

$$= T_\theta T_\theta';$$  \hspace{1cm} (35)

there is a 1–1 correspondence between $\Sigma_\theta$ and $T_\theta$ ($T_\theta$ is the positive lower triangular square root of $\Sigma_\theta$). If the error distribution has zero means, unit variances and zero covariances, then $\Sigma_\theta$ is the second-moment matrix. The transformation from $T_\theta$ to $\Sigma_\theta$ is straightforward:

$$\frac{\partial \Sigma_\theta}{\partial \theta} = 2^k \left| T_\theta \right|_{\nabla},$$  \hspace{1cm} (36)
where \( |T|_v \) is the decreasing determinant of the lower triangular matrix; the decreasing determinant has general form

\[
|T|_v = t_{00}^{k+1} t_{11}^k \ldots t_{kk}^1.
\]

The marginal structural distribution for \( \Sigma_0 \) is

\[
\frac{1}{2^k} \prod_{i=2}^k A_i k(D_i) f(T_{0}^{-1}X) \left| \begin{array}{c} \Sigma_0^{1/2} \\ X \end{array} \right|^{-1/2} \times \left| \begin{array}{c} T_{X^*}^{n-1} \\ T_{0}^{n-1} \end{array} \right| dT_0 dT_{X^*} dO_{X^*}.
\]

this is one of a spectrum of Bayesian posterior distributions and corresponds to the prior density \( |\Sigma_0|^{-(k+1)/2} \).

The corresponding prediction probability element for \( [X^*] \) based on a future observation matrix \( X^* \) from the same model is obtained as

\[
\frac{1}{2^k} \prod_{i=2}^k A_i k_n(D_i) k_{n^*}(D_{*i}) f_n(T_{0}^{-1}X) f_n(T_{0}^{-1}X^*)
\]

\[
\times \left| \begin{array}{c} T_{X^*}^{n-1} \\ T_{0}^{n-1} \end{array} \right| dT_0 dT_{X^*} dO_{X^*},
\]

where \( n' \) is the sample size for the future observations and \( D^* \) stands for the orbit based on future observations. This distribution is one of a spectrum of Bayesian prediction distributions; it corresponds to the prior \( |T_0|^{-1/2} \).

4. The Multivariate Normal

4.1. The Multivariate Normal

The multivariate normal can be obtained by specializing the rotationally symmetric error model:

\[
f(E) = (2\pi)^{-kn/2} \exp\{-\frac{1}{2} \sum e_{ij}^2\} = (2\pi)^{-kn/2} \exp\{-\frac{1}{2} [tr EE' - n] \}.
\]

The conditional distribution of \( [E] \) given \( D(E) \) is obtained from formula (27):

\[
\frac{k(D)}{(2\pi)^{kn/2}} \exp\{-\frac{1}{2} [tr [E] D D' [E] - n] \} \left| [E] \right|^n d[E] \left| [E] \right|^{k+1}.
\]

The transformation \( [X] \) can be factored into triangular and orthogonal components as

\[
[X] = T_X O_X,
\]

and the invariant differential factored as

\[
\frac{d[X]}{|X|^{k+1}} = \frac{dT_X dO_X}{|T_X| \Delta}.
\]

The probability element for \( [E] \) given \( D(E) \) is then

\[
\frac{k(D)}{(2\pi)^{kn/2}} \exp\{-\frac{1}{2} [tr [E] D_n [E] - n] \} \left| [E] \right|^n \frac{dT_E dO_E}{|T_E| \Delta},
\]

(40)
where

\[ D_n = DD' = \begin{pmatrix} n & 0 \\ 1 & \ddots \\ 0 & \ddots & 1 \end{pmatrix}. \]

This can be further simplified by the adjustment

\[ \mathbf{T}_\mathbf{X} = \begin{pmatrix} n^k \bar{x}_1 & 0 \\ n^k \bar{x}_1 & s_1 \\ \vdots & \vdots & \ddots \\ n^k \bar{x}_k & b_{k1} & \cdots & s_k \end{pmatrix} = \mathbf{T}_\mathbf{X} \mathbf{D}_n^k, \]

\[ [\bar{X}] = \mathbf{T}_\mathbf{X} \mathbf{O}_\mathbf{X} = [X] \mathbf{D}_n^k. \]  \hspace{1cm} (41)

The probability element then becomes

\[ \frac{k(D)}{(2\pi)^{kn/2}} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{T}_\mathbf{E} \mathbf{T}_\mathbf{E}' - n \right\} \left| \mathbf{T}_\mathbf{E} \right|^n \frac{d\mathbf{T}_\mathbf{E}}{n^{k/2}} \left| \mathbf{T}_\mathbf{E} \right|^{\Delta} d\mathbf{O}_\mathbf{E}. \]

This expression factors and shows that: the diagonal components of \( \mathbf{T}_\mathbf{E} \) are \( \chi \) variables; the off-diagonal components of \( \mathbf{T}_\mathbf{E} \) are standard normal; and the distribution of \( \mathbf{O}_\mathbf{E} \) is uniform. This determines the constant and the probability element becomes

\[ \frac{1}{\prod_{i=1}^{k} A_{n-i}} \frac{k(D)}{(2\pi)^{kn/2}} \exp \left\{ -\frac{1}{2} \text{tr} \mathbf{T}_\mathbf{E} \mathbf{T}_\mathbf{E}' - n \right\} \left| \mathbf{T}_\mathbf{E} \right|^n \frac{d\mathbf{T}_\mathbf{E}}{n^{k/2}} \left| \mathbf{T}_\mathbf{E} \right|^{\Delta} d\mathbf{O}_\mathbf{E} \times \prod_{i=2}^{k} A_i \]  \hspace{1cm} (42)

the normalizing constant is

\[ k(D) = \frac{n^{k/2}}{\prod_{i=1}^{k} A_{n-i}} \frac{\prod_{i=1}^{k} A_{n-i}}{\prod_{i=2}^{k} A_i}. \]

The distribution for the inner product matrix

\[ \mathbf{S}_\mathbf{E} = \mathbf{E}\mathbf{E}' = \mathbf{T}_\mathbf{E} \mathbf{T}_\mathbf{E}' \]

conditional on \( \mathbf{D}(\mathbf{E}) \) is easily obtained: the change in differential is

\[ d\mathbf{S}_\mathbf{E} = 2^k n^{k/2} \left| \mathbf{T}_\mathbf{E} \right|^{\frac{n}{2}} d\mathbf{T}_\mathbf{E}; \]  \hspace{1cm} (43)

the probability element is

\[ \frac{1}{2^k n^{k/2} (2\pi)^{kn/2}} \prod_{i=1}^{k} A_{n-i} \exp \left\{ -\frac{1}{2} [\text{tr} \mathbf{S}_\mathbf{E} - n] \right\} \frac{\left| \mathbf{T}_\mathbf{E} \right|^n}{\left| \mathbf{T}_\mathbf{E} \right|^{k/2}} d\mathbf{S}_\mathbf{E}. \]  \hspace{1cm} (44)
Now consider the structural distribution for the parameter $T_\theta$. From formula (34) the probability element is

$$
\frac{n^{k/2} \prod_{i=1}^k A_{n-i}}{(2\pi)^{kn/2}} \exp\left\{-\frac{1}{2}[\text{tr} T_\theta^{-1} T_X T_X' T_\theta'^{-1} - n]\right\} \frac{|T_X|^{n-1} dT_\theta}{|T_\theta|^{n-1} |T_\theta'|_\Delta}. \tag{45}
$$

The structural probability element for $\Sigma_\theta$ is obtained from (37):

$$
\frac{n^{k/2} \prod_{i=1}^k A_{n-i}}{2^k (2\pi)^{kn/2}} \exp\left\{-\frac{1}{2}[\text{tr} \Sigma_\theta^{-1} S_X - n]\right\} \frac{|T_X|^{n-1} d\Sigma_\theta}{|\Sigma_\theta|^{(n-1)/2} |\Sigma_\theta'|_{(k+1)/2}} \tag{46}
$$

where

$$S_X = T_X T_X' = XX'$$

is the inner product matrix of the rows of $X$. This can also be obtained by Bayesian analysis using the prior $|\Sigma_\theta|^{-(k+1)/2}$.

4.2. The Marginal Structural Distributions

The structural distributions of the parameter $\Sigma_\theta$ of the multivariate normal is given by the expression (46). The parameter $\Sigma_\theta$ records the second moments of the variables $1, x_1, \ldots, x_k$:

$$\Sigma_\theta = \begin{pmatrix}
1 & \mu_1 & \cdots & \mu_k \\
\mu_1 & \mu_{11} & \cdots & \mu_{1k} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_k & \mu_{k1} & \cdots & \mu_{kk}
\end{pmatrix}.$$

The more common parameters of the multivariate normal are the means and covariances:

$$\mu = \begin{pmatrix}
\mu_1 \\
\vdots \\
\mu_k
\end{pmatrix}, \quad \Sigma = \begin{pmatrix}
\sigma_{11} & \cdots & \sigma_{1k} \\
\vdots & \ddots & \vdots \\
\sigma_{k1} & \cdots & \sigma_{kk}
\end{pmatrix}.$$

The two forms of parameter can be related in matrix notation by defining

$$M_\theta = \begin{pmatrix}
1 & 0 & 0 \\
\mu_1 & 1 & 0 \\
\vdots & \ddots & \ddots \\
\mu_k & 0 & 1
\end{pmatrix},$$

then

$$M_\theta^{-1} \Sigma_\theta = \begin{pmatrix}
1 & \mu' \\
0 & \Sigma
\end{pmatrix}, \quad M_\theta^{-1} \Sigma_\theta M_\theta^{-1} = \begin{pmatrix}
1 & 0' \\
0 & \Sigma
\end{pmatrix}. \tag{47}$$
The statistic $S_X$ can be similarly analysed. Let

$$M_X = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \bar{x}_1 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ \bar{x}_k & 0 & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0' \\ \bar{x} & I \end{pmatrix};$$

and $S$ be the Wishart inner product matrix,

$$S = \begin{bmatrix} \Sigma(x_{ij}-\bar{x}_i)^2 & \ldots & \Sigma(x_{ij}-\bar{x}_i)(x_{kj}-\bar{x}_k) \\ \vdots & \ddots & \vdots \\ \Sigma(x_{kj}-\bar{x}_k)(x_{ij}-\bar{x}_i) & \ldots & \Sigma(x_{kj}-\bar{x}_k)^2 \end{bmatrix};$$

then

$$M_X^{-1}S_X = \begin{pmatrix} n & x' \\ 0 & S \end{pmatrix}, \quad S_X = M_X \begin{pmatrix} n & 0' \\ 0 & S \end{pmatrix} M_X'.$$

(48)

Consider the triangular matrices $T_X$ and $T_\theta$ using the notation $T_{(X)}, T_{(\theta)}$ for submatrices:

$$T_X = \begin{pmatrix} n^i & 0' \\ n^i \bar{x} & T_{(X)} \end{pmatrix}, \quad T_\theta = \begin{pmatrix} 1 & 0' \\ \mu & T_{(\theta)} \end{pmatrix}.$$

The expression in the exponent of formulas (45) and (46) can then be simplified:

$$T_\theta^{-1}T_X = \begin{pmatrix} 1 & 0' \\ -T_{(\theta)}^{-1} \mu & T_{(\theta)}^{-1} \end{pmatrix} \begin{pmatrix} n^i & 0' \\ n^i \bar{x} & T_{(X)} \end{pmatrix} = \begin{pmatrix} n^i & 0' \\ n^i T_{(\theta)}^{-1} (x-\mu) & T_{(X)} \end{pmatrix},$$

and

$$T_\theta^{-1}T_X T_\theta^{-1} = \begin{pmatrix} n & 0' \\ 0 & n T_{(\theta)}^{-1} (x-\mu)(x-\mu)' T_{(\theta)}^{-1} + T_{(\theta)}^{-1} T_{(X)} T_{(X)} T_{(\theta)}^{-1} \end{pmatrix},$$

so that

$$\text{tr} T_\theta^{-1}T_X T_\theta^{-1} - n = n(\bar{x}-\mu)' \Sigma^{-1}(\bar{x}-\mu) + \text{tr} \Sigma^{-1} S.$$

The change of variable from $\Sigma_\theta$ to $(\mu, \Sigma)$ is straightforward:

$$\frac{\partial \Sigma_\theta}{\partial \mu \partial \Sigma} = \begin{pmatrix} I & * \\ 0 & I \end{pmatrix}, \quad d\Sigma_\theta = d\mu d\Sigma;$$

(49)

the first identity submatrix derives from the identity transformation of $\mu$ to $\mu$, and the second corresponds to the transformation from covariances to the full second moments.
The structural probability element (46) can then be transformed:
\[
\frac{n^{k/2}}{2^{k} (2\pi)^{kn/2}} \prod_{i=1}^{k} A_{n-i} \exp \left\{ -\frac{1}{2} [n(\bar{x} - \mu)'] \Sigma^{-1} (\bar{x} - \mu) + \text{tr} \Sigma^{-1} S \right\} \left| \Sigma \right|^{(n-1)/2} d\Sigma \left| \Sigma \right|^{(k+1)/2} d\mu \\
= \frac{n^{k/2}}{(2\pi)^{k/2}} \prod_{i=1}^{k} A_{n-i} \exp \left\{ -\frac{1}{2} n(\bar{x} - \mu)\Sigma^{-1}(\bar{x} - \mu) \right\} d\mu \\
\times \frac{n^{k/2}}{2^{k} (2\pi)^{k(n-1)/2}} \prod_{i=1}^{k} A_{n-i} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right\} \left| \Sigma \right|^{(n-1)/2} \left| \Sigma \right|^{(k+1)/2} d\Sigma \\
= f_2(\mu | \Sigma, x)f_2(\Sigma | S).
\] (50)

Thus the marginal structural distribution for \( \Sigma \) is
\[
\frac{n^{k/2}}{2^{k} (2\pi)^{k(n-1)/2}} \prod_{i=1}^{k} A_{n-i} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right\} \left| \Sigma \right|^{(n-1)/2} \left| \Sigma \right|^{(n+k)/2} d\Sigma.
\] (51)

This distribution has appeared as an invariant fiducial distribution (Hastings, 1962).

The marginal distribution of \( \mu \) can be obtained by using an integration relationship implicit in the probability distribution (51):
\[
\int \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right\} \left| \Sigma \right|^{-(n+k)/2} d\Sigma = \frac{2^{k} (2\pi)^{k(n-1)/2}}{\prod_{i=1}^{k} A_{n-i}} \left| S \right|^{-(n-1)/2}.
\] (52)

The marginal structural element for \( \mu \) is
\[
\int \frac{n^{k/2}}{2^{k} (2\pi)^{kn/2}} \prod_{i=1}^{k} A_{n-i} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} [n(\bar{x} - \mu)'(\bar{x} - \mu) + S] \right\} \left| \Sigma \right|^{-(n+k+1)/2} d\Sigma \left| S \right|^{(n-1)/2} d\mu \\
= \frac{n^{k/2}}{\pi^{k/2}} \frac{\Gamma(n/2)}{\Gamma((n-k)/2)} \left[ 1 + n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \right]^{-n/2} d\mu \left| S \right|^{(k+1)/2}.
\] (53)

This structural distribution for \( \mu \) is consistent with Hotelling’s distribution for the quantity
\[
T^2 = n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) = \frac{k}{n-k} F,
\]
where \( F \) has an \( F \) distribution on \( k \) over \( n-k \) degrees of freedom. Some commentary on fiducial distributions for \( \mu \) may be found in Geisser and Cornfield (1963) and Fraser (1964). Geisser and Cornfield (1963), Tiao and Zellner (1964) and Geisser (1965) include the distributions (50), (51), (53) among their Bayesian posterior distributions; the corresponding prior is \( \left| \Sigma \right|^{-1/2} \).
4.3. Prediction of a Set of Variables

The structural probability element for $\mu$ and $\Sigma$ is given by (50). Consider a future observation $x^*$ from the multivariate model with response distribution

$$
\int |S|^{(n-1)/2} \exp \left\{ -\frac{1}{2} (x^* - \mu) \Sigma^{-1} (x^* - \mu) \right\} dS = (2\pi)^{k/2} n^{k/2} \exp \left\{ -\frac{1}{2} (x^* - \mu)' \Sigma^{-1} (x^* - \mu) \right\} dx^*.
$$

The joint probability element for $x^*$, $\mu$ and $\Sigma$ is

$$
n^{k/2} \frac{|\Sigma|^{(n-1)/2}}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} (x^* - \mu)' \Sigma^{-1} (x^* - \mu) - \frac{1}{2} n (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu) \right\}
\times \frac{1}{2^{k(2\pi)^{k/2}}} \frac{1}{\Sigma^{(k+1)/2}}
\exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right\} ds = \frac{1}{2^{k} (2\pi)^{k/2}} n^{k/2} \int \frac{1}{h^{(n-1)/2}} dS d\Sigma d\mu dx^*. 
$$

The exponent (55) involving $\mu$ can be written as

$$
n + \frac{1}{2} \left( \left( x^* + \frac{n \bar{x}}{n+1} \right)' \Sigma^{-1} \left( x^* + \frac{n \bar{x}}{n+1} \right) - \frac{n}{2(n+1)} (x^* - \bar{x})' \Sigma^{-1} (x^* - \bar{x}) \right).$$

The probability element for $\Sigma$ and $x^*$ is obtained by integrating over $\mu$:

$$
\left( \frac{n}{n+1} \right)^{k/2} \frac{1}{2^{k(2\pi)^{k/2}}} \frac{1}{\Sigma^{(n+k+1)/2}} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right\} ds = \frac{1}{2^{k} (2\pi)^{k/2}} n^{k/2} \frac{\Gamma(n/2)}{\Gamma((n-k)/2)} \int \frac{1}{\Sigma^{(n+k)/2}} d\Sigma dx^*.
$$

The predictive probability element for $x^*$ is then obtained by the further integration over $\Sigma$ using (52):

$$
\left( \frac{n}{n+1} \right)^{k/2} \frac{1}{2^{k(2\pi)^{k/2}}} \frac{1}{\Sigma^{(n+k+1)/2}} \frac{\Gamma(n/2)}{\Gamma((n-k)/2)} \left\{ 1 + \frac{n}{n+1} (x^* - \bar{x})' \Sigma^{-1} (x^* - \bar{x}) \right\}^{-n/2} dS^* dx^*.
$$

The prediction distribution for the future observations is of the same form as the marginal structural distribution of the means obtained in (53).

4.4. Prediction of the Mean Vector and the Wishart Matrix

Consider a future sample of size $n'$ from the same multivariate normal distribution, and let $\bar{x}^*$ and $S^*$ be the corresponding mean vector and Wishart's matrix respectively. The probability element for $\bar{x}^*$ and $S^*$ can be written from (44) and (49):

$$
n' \frac{1}{2^{k(2\pi)^{k/2}}} \frac{1}{\Sigma^{n'/2}} \exp \left\{ -\frac{1}{2} (\bar{x}^* - \mu)' \Sigma^{-1} (\bar{x}^* - \mu) - \frac{1}{2} \text{tr} \Sigma^{-1} S^* \right\} \frac{n'}{n' + n} \left( \frac{n'}{n' + n} \right)^{n'/2} dS^* dx^*.
$$

The distribution of $\bar{x}^*$, $S^*$ and $\Sigma$ is obtained by integrating the joint distribution of $\mu$, $\Sigma$, $\bar{x}^*$, $S^*$ over $\mu$:

$$
\left( \frac{mn'}{n+n'} \right)^{k/2} \frac{1}{2^{k(2\pi)^{k/2}}} \frac{1}{\Sigma^{(n+k)/2}} \frac{1}{n'+n} \frac{\Gamma(n'/2)}{\Gamma((n'-k)/2)} \left\{ 1 + \frac{n'}{n' + n} (\bar{x}^* - \bar{x})' \Sigma^{-1} (\bar{x}^* - \bar{x}) \right\}^{-n'/2} dS^* d\bar{x}^* d\Sigma.
$$
The probability element for \( S^* \) and \( \bar{x}^* \) is then obtained by the further integration with respect to \( \Sigma \) using (52):

\[
\left( \frac{n'}{n+n'} \right)^{k/2} \frac{k}{\prod_{i=1}^{k} A_{n-i} \prod_{i=1}^{k} A_{n'-i}} |S|^{(n-1)/2} |S^*|^{(n-k-2)/2} \frac{2^k}{\prod_{i=1}^{k} A_{n+n'-i}} \times |S+S^* + \frac{nn'}{n+n'} (\bar{x}^* - \bar{x})(\bar{x}^* - \bar{x})'|^{-(n+n'-1)/2} dS^* d\bar{x}^* \tag{60}
\]

\[
= \frac{k}{\prod_{i=1}^{k} A_{n-i} \prod_{i=1}^{k} A_{n'-i}} \frac{k}{\prod_{i=1}^{k} A_{n+n'-1-i}} |S|^{(n-1)/2} |S^*|^{(n'-k-2)/2} dS^* \left( \frac{nn'}{n+n'} \right)^{k/2} \frac{2^k}{\prod_{i=1}^{k} A_{n+n'-i}} \times \left( 1 + \frac{nn'}{n+n'} (\bar{x}^* - \bar{x})(S+S^*)^{-1}(\bar{x}^* - \bar{x}) \right)^{-(n+n'-1)/2} \frac{d\bar{x}^*}{|S+S^*|^{1/2}}.
\]

This is a generalized \( F \) distribution for \( S^* \) and a conditional multivariate \( t \)-distribution for \( \bar{x}^* \). Geisser (1965) considers Bayesian prediction with the multilinear model; his insightful choice of prior leads to posterior and prediction distributions that agree with those just given.

REFERENCES