The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 593–605

THE CONDITIONAL WISHART: NORMAL AND NONNORMAL

By D. A. S. Fraser

University of Toronto

0. Summary. The response variable of a general multivariate model can be constructed as a positive *affine* transformation of a vector error variable. In the case of an error variable that is rotationally symmetric, the multivariate model has parameters that can be expressed as the mean vector and the variance matrix. In the case of an error variable that is standard normal, it becomes the ordinary multivariate normal.

For the general multivariate model with rotationally-symmetric error the sample inner-product matrix is a conventional statistic for inference. The distribution of this statistics is derived: the distribution is a conditional distribution given observable characteristics of the error variable.

The response variable of a more specialized multivariable model can be constructed as a positive *linear* transformation of a vector error variable. In the case of an error variable that is rotationally symmetric, the model has parameters that can be expressed in terms of the variance matrix, the mean vector being zero. In the case of an error variable that is standard normal; it becomes the ordinary central multivariate normal.

For the multivariate model with rotationally-symmetric error, the distribution of the Wishart matrix is derived; the distribution is a conditional distribution given observable characteristics of the error variable. And for the multivariate model with standard normal error, the noncentral distribution of the Wishart matrix is also derived, again as the appropriate conditional distribution.

1. The affine multivariate model. Consider a process or system with a sequence of p response variables $x_1, \cdots, x_p$. Suppose the operation of the process or related processes has been investigated and the internal error of the process, as it affects the response variable, can be described by a sequence of error variables $e_1, \cdots, e_p$ with an identified error distribution $f(e_1, \cdots, e_p)\, de_1 \cdots de_p$ on R^p. Let $\mu_1, \cdots, \mu_p$ be the general levels for the p response variables. And suppose the error variables $e_1, \cdots, e_p$ affect the general response levels in the form of a linear distribution: let $\gamma_{jj'}$ be the coefficient for the j'th error as it affects the jth response. A realized sequence of errors and the corresponding sequence of response values are then connected by the matrix equation

$$\begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{bmatrix} = \theta \begin{bmatrix} 1 \\ e_1 \\ \vdots \\ e_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \mu_1 & \gamma_{11} & & \gamma_{1p} \\ \vdots & \vdots & & \vdots \\ \mu_p & \gamma_{11} & \cdots & \gamma_{pp} \end{bmatrix} \begin{bmatrix} 1 \\ e_1 \\ \vdots \\ e_p \end{bmatrix}.$$

Suppose that the matrix quantity θ is an unknown of the process.

Received 23 January 1967; revised 20 November 1967.

For a sequence of n observations on the system, let $\mathbf{x}_1 = (x_{11}, \cdots, x_{1n})'$ be the observations for the first response, ..., and $\mathbf{x}_p = (x_{p1}, \cdots, x_{pn})'$ be the observations for the pth response. The system and the n observations can then be described by the *affine multivariate model*

$$\prod_1^n f(e_{1i}, \cdots, e_{pi}) \prod de_{11} \cdots de_{pi}$$

$$\begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix} = \theta \begin{bmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{bmatrix},$$

or in matrix notation

$$f(E)\, dE$$

$$X = \theta E$$

with

$$X = \begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix}, \qquad E = \begin{bmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{bmatrix},$$

$$\theta = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline \mu_1 & \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & \vdots & & \vdots \\ \mu_p & \gamma_{p1} & \cdots & \gamma_{pp} \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline \boldsymbol{\mu} & \Gamma \end{array}\right].$$

The model has an *error distribution* $f(E)\, dE$ describing the internal error of the process; *E is a variable.* And it has a *structural equation* $X = \theta E$ in which a realized error E has determined the relationship between the *known* observation X and the *unknown* θ; *E is a constant.*

The translations $\boldsymbol{\mu}$ and the distortions Γ combine to form the positive affine group G for R^p:

$$G = \left\{ g = \begin{pmatrix} 1 & 0 \\ \mathbf{a} & C \end{pmatrix} : \begin{matrix} -\infty < a_j < \infty \\ -\infty < c_{jj'} < \infty \end{matrix}, \; |C| > 0 \right\}.$$

The transformations g applied to X in R^{pn} form a unitary transformation group provided $n \geqq p + 1$ and certain trivial points are excluded (G is *unitary* on $\mathfrak{X} = \{x\}$ if $g'x = g''x$ implies $g' = g''$).

Let $[X]$ be a *transformation variable* mapping R^{pn} into G^*, a duplicate of the group G:

$$[gX] = g[X] \qquad \forall b, X.$$

A transformation variable can be defined in two stages that are convenient for the subsequent Wishart analysis. Consider the point X as a sequence of p points $\mathbf{x}_1, \cdots, \mathbf{x}_p$ in R^n. A transformation g applied to these p points with appended one-vector,

$$g \begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_p' \end{bmatrix},$$

carries them into p points in the linear subspace $L(\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_p)$ again with appended one-vector. The transformation has matrix C with positive determinant; accordingly the p new points have the same orientation positive or negative, relative to the one-vector as have the p original points.

Now suppose that $n \geqq p + 1$ and that trivial points X with $\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_p$ linearly dependent are excluded. Let $L^+(\mathbf{1}; \mathbf{x}_1, \cdots, \mathbf{x}_p)$ be the $p + 1$ dimensional subspace $L(\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_p)$ together with an orientation, positive or negative—the orientation of the p vectors $\mathbf{x}_1, \cdots, \mathbf{x}_p$ relative to the one-vector. A transformation g carries p vectors of a nontrivial X into p new vectors, subspace and orientation preserved.

Let $\bar{x}_1$ and $s_{(1)}(X)$ be the regression coefficient and residual length of $\mathbf{x}_1$ regressed on $\mathbf{1}$; and let $\mathbf{d}_1^*(X)$ be the unit residual vector. Let $\bar{x}_2$, $t_{21}(X)$ and $s_{(2)}$ be the regression coefficients and residual length of $\mathbf{x}_2$ regressed on $\mathbf{1}, \mathbf{d}_1^*$; and let $\mathbf{d}_2^*(X)$ be the unit residual vector. Finally let $\bar{x}_p$, $t_{p1}(X), \cdots, t_{pp-1}(X)$ and $s_{(p)}(X)$ be the regression coefficients and residual length of $\mathbf{x}_p$ regressed on $\mathbf{1}, \mathbf{d}_1^*(X), \cdots, \mathbf{d}_{p-1}^*(X)$; and let $\mathbf{d}_p^*(X)$ be the unit residual vector. Then

$$X = {}_T[X]D^*(X)$$

$$= \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ \bar{x}_1 & s_{(1)}(X) & & & 0 \\ \bar{x}_2 & t_{21}(X) & s_{(2)}(X) & & \\ \vdots & \vdots & & & \vdots \\ \bar{x}_p & t_{p1}(X) & \cdots & t_{pp-1}(X) & s_{(p)} \end{bmatrix} \begin{bmatrix} \mathbf{1}' \\ \mathbf{d}_1^{*\prime}(X) \\ \\ \vdots \\ \mathbf{d}_p^{*\prime}(X) \end{bmatrix}.$$

Now let $\mathbf{x}_1^0, \cdots, \mathbf{x}_p^0$ be the projections of the first p coordinate vectors into $L(\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_p)$, the sign of the last vector being chosen so that $L^+(\mathbf{1}; \mathbf{x}_1^0, \cdots, \mathbf{x}_p^0) = L^+(\mathbf{1}; \mathbf{x}_1, \cdots, \mathbf{x}_p)$. And let

$$X^0 = \begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1^0 \\ \vdots \\ \mathbf{x}_p^0 \end{bmatrix}, \qquad D(X) = \begin{bmatrix} \mathbf{1}' \\ \mathbf{d}_1'(X) \\ \vdots \\ \mathbf{d}_p'(X) \end{bmatrix} = D^*(X^0).$$

The orthonormal vectors in $D^*(X)$ and $D(X) = D^*(X^0)$ are related by an orthogonal matrix:

$$\begin{aligned} D^*(X) &= [X]_O D(X) \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & o_{11}(X) & \cdots & o_{1p}(X) \\ \vdots & \vdots & & \\ 0 & o_{p1}(X) & \cdots & o_{pp}(X) \end{bmatrix} D(X) \\ &= \begin{bmatrix} 1 & 0 \\ 0 & O(X) \end{bmatrix} D(X) \end{aligned}$$

where $O(X)$ is a $p \times p$ rotation matrix (positive determinant).

A general point X can now be expressed as a transformation ${}_T[X][X]_O$ from a *reference point* $D(X)$, a point with vectors $\mathbf{d}_1(X), \cdots, \mathbf{d}_p(X)$ in $L^+(\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_p)$, a point that depends on the oriented subspace and not otherwise on X.

$$\begin{aligned} X &= {}_T[X][X]_O D(X) = [X]D(x) \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \bar{x}_1 & c_{11}(X) & \cdots & c_{1p}(X) \\ \vdots & \vdots & & \vdots \\ \bar{x}_p & c_{p1}(X) & \cdots & c_{pp}(X) \end{bmatrix} D(X) \\ &= \begin{bmatrix} 1 & 0 \\ \bar{\mathbf{x}} & C(X) \end{bmatrix} D(X). \end{aligned}$$

The variable $[X] = {}_T[X][X]_O$ is a transformation variable—a consequence of $D(X)$ being a fixed point relative to the oriented subspace.

The Jacobian of a transformation $\tilde{X} = gX$ is $|g|^n$; the Jacobian from $D(X)$ to X is $|[X]|^n$; the compensated differential

$$dm(X) = dX/|[X]|^n$$

is the unique invariant differential relative to G that agrees with Euclidean volume at $D(X)$.

Now consider invariant differentials on the group

$$\begin{bmatrix} 1 & 0 \\ \tilde{\mathbf{a}} & \tilde{C} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \mathbf{a} & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \mathbf{a}^* & C^* \end{bmatrix}.$$

The left transformation operates column-by-column. For any given column, the Jacobian is $|C|$; the Jacobian of the left transformation is $|g|^n$; the Jacobian from the identity to g is $|g|^{p+1}$; the compensated differential

$$d\mu(g) = dg/|g|^{p+1}$$

is then the left invariant differential on the group that agrees with Euclidean volume at the identity.

2. Distributions of the affine multivariate model. Consider further the affine multivariate model

$$f(E)\,dE$$
$$X = \theta E.$$

The structural equation gives access to characteristics of the realized error E,

$$D(E) = [E]^{-1}E = [X]^{-1}X = D(X).$$

But it gives no information concerning the position of E relative to the reference point $D(E) = D(X)$:

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

The affine multivariate model by its own information content gives therefore the reduced model

$$g([E]: D(X))\, d[E]$$
$$[X] = \theta[E].$$

The reduced model contains an error probability distribution $g([E]: D(X))\, d[E]$ which provides probability statements concerning the unknown error position $[E]$. And it has a structural equation in which the unknown error position $[E]$ is the link between the known $[X]$ and the unknown θ. The error probability distribution $g([E]: D)\, d[E]$ is the conditional distribution of $[E]$ given $D(E) = D$.

The error probability distribution can be derived in a direct manner. The probability element can be expressed in terms of $[E]$,

$$\begin{aligned} f(E)\, dE &= f([E]D)\,|[E]|^{n}\, dm(E) \\ &= f([E]D)\,|[E]|^{n}\delta\, d\mu([E]) \\ &= f([E]D)\,|[E]|^{n-p-1}\delta\, d[E], \end{aligned}$$

and then normalized,

$$g([E]: D)\, d[E] = k(D) f([E]D)\,|[E]|^{n-p-1}\, d[E].$$

The classical model, that corresponds to the affine multivariate model, describes a response *variable* X. The classical model is

$$\begin{aligned} f(X: \theta)\, dX &= f(\theta^{-1}X)\, d\theta^{-1}X \\ &= f(\theta^{-1}X)\,|\theta^{-1}[X]|^{n}\, dm(\theta^{-1}X) \\ &= f(\theta^{-1}X)\,|\theta^{-1}[X]|^{n}\, dm(X) \\ &= f(\theta^{-1}X)\,|\theta|^{-n}\, dX. \end{aligned}$$

The classical model conditional on the characteristic $D(X) = D$ is

$$\begin{aligned} f([X]: D, \theta)\, d[X] &= g(\theta^{-1}[X]: D)\, d\theta^{-1}[X] \\ &= g(\theta^{-1}[X]: D)|\theta|^{-(p+1)}\, d[X] \\ &= k(D) f(\theta^{-1}[X]D)|\theta^{-1}[X]|^{n}\, d[X]/|[X]|^{p+1}. \end{aligned}$$

The conventional statistic in multivariate analysis is the sample inner product or Wishart matrix:

$$\begin{aligned} S(X) = XX' &= \begin{bmatrix} \sum n & \sum x_1 & \cdots & \sum x_p \\ \sum x_1 & \sum x_1^2 & \cdots & \sum x_1 x_p \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_p & \sum x_p x_1 & \cdots & \sum x_p^2 \end{bmatrix} \\ &= {}_T[X]\,[X]_O D(X) D'(X) [X]_O{}'\, {}_T[X]' \\ &= {}_T[X] J_n\, {}_T[X]' \end{aligned}$$

where J_n is a trivial diagonal matrix

$$J_n = \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The triangular matrices ${}_T[X]$ form a positive lower triangular group

$$G_T = \left\{ k = \begin{pmatrix} 1 & 0 & & \cdots & 0 \\ a_1 & s_{(1)} & & \cdots & 0 \\ a_2 & t_{21} & s(2) & & \vdots \\ \vdots & \vdots & & & \\ a_p & t_{p1} & \cdots & t_{pp-1} & s_{(p)} \end{pmatrix} : \begin{array}{c} -\infty < a_j < \infty \\ -\infty < t_{jj} < \infty \\ 0 < s_{(j)} < \infty \end{array} \right\}.$$

A left transformation operates column-by-column; the Jacobian from i to k is

$$(s_{(1)} \cdots s_{(p)})(s_{(1)} \cdots s_{(p)})(s_{(2)} \cdots s_{(p)}) \cdots (s_{(p)});$$

the compensated differential

$$d\mu_T(k) = dk/|k|_\Delta, \qquad |k|_\Delta = 1^1 s_{(1)}^2 \cdots s_{(p)}^{p+1}$$

is then the left invariant differential that agrees with Euclidean volume at the identity.

The orthogonal matrices $[X]_O$ form a rotation group

$$G_O = \left\{ h = \begin{pmatrix} 1 & 0 \\ 0 & O \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & o_{11} & \cdots & o_{1p} \\ \vdots & \vdots & & \vdots \\ 0 & o_{p1} & \cdots & o_{pp} \end{pmatrix} : \begin{array}{c} O'O = I \\ |O| > 0 \end{array} \right\}.$$

The Euclidean differential dO is the unique invariant differential that agrees with Euclidean volume at the identity. (The differential dO is the Euclidean differential subject to the constraints $O'O = I$.)

The invariant differentials on G_T, G_O combine to form the invariant differential for $g = kh$ in G:

$$dg/|g|^{p+1} = dk\, dh/|k|_\Delta .$$

The transformation from ${}_T[X]$ to $S(X)$ is one-one; it is diagonal in blocks and operates row-by-row on ${}_T[X]$:

$$\partial S(X)/\partial_T[X] = 2^p n^p |{}_T[X]|_\nabla$$

where

$$|k|_\nabla = 1^{p+1} s_{(1)}^{p} \cdots s_{(p)}^{1} .$$

The distribution of the conventional variable $S(X)$, conditional on the error characteristic $D(X) = D$, is obtained from $g([X]: D, \theta)\, d[X]$ by change of variable and integration:

$$k(D) \int_{O_X} f(\theta^{-1}{}_T[X][X]_O D)|\theta^{-1}{}_T[X]|^n\, d_T[X]\, d[X]_O/|{}_T[X]|_\Delta$$
$$= k(D) \textstyle\prod_2^p A_j \bar{f}(\theta^{-1}{}_T[X], D)|\theta|^{-n}|{}_T[X]|^n (2^p n^p |{}_T[X]|^{p+2})^{-1}\, dS(X)$$

where

$$\bar{f}(\theta^{-1}k, D) = \int_{G_O} f(\theta^{-1}khD) \left(\textstyle\prod_2^p A_j\right)^{-1} dh$$

is the average of f over the rotations h in G_O and where A_j is the area of a unit sphere in j-space ($A_j = 2\pi^{j/2}/\Gamma(j/2)$).

3. The affine model with error symmetry. Now suppose that the error distribution in the affine multivariate model is rotationally symmetric:

$$f(h(1, e_1, \cdots, e_k)') = f((1, e_1, \cdots, e_k)'),$$
$$f(hE) = f(E)$$

for all h in the rotation group G_O.

The quantity θ can be factored into a triangular component ${}_T[\theta]$ in G_T and a rotation component $[\theta]_O$ in G_O in the pattern in Section 1:

$$\theta = {}_T[\theta][\theta]_O .$$

And if the error pattern is scaled so that the variances are unity then,

$$\Sigma = \theta\theta' = {}_T[\theta][\theta]_T$$

records the second moments of the variables $1, x_1, \cdots, x_p$ and hence determines the first and second moments of $x_1, \cdots, x_p$. The distribution of the Wishart matrix $S(X)$ in the preceding section can then be simplified:

$$k(E) \textstyle\prod_2^p A_j \bar{f}({}_T[\theta]^{-1}{}_T[X], D)|\theta|^{-n}|{}_T[X]|^n\, dS(X)/2^p n^p |{}_T[X]|^{p+2}$$

where

$$\bar{f}(k, D) = \int_{G_O} f(khD)\, dh / \prod_2^p A_j$$

is the average of f over the rotations h is G_O.

Now suppose that the error distribution is standard normal:

$$f(E) = (2\pi)^{-pn/2} \exp\{-\tfrac{1}{2}(\text{tr } EE' - n)\}.$$

The distribution of ${}_T[E]$ and $[E]_O$ conditional on the observed characteristic $D(E) = D = D(X)$ is

$$k(D) \prod_2^p A_j (2\pi)^{-pn/2} \exp\{-\tfrac{1}{2}(\text{tr } {}_T[E] J_n\, {}_T[E]' - n)\}\; d_T[E]\, d[E]_O / |{}_T[E]|^n |{}_T[E]|_\Delta .$$

This is a uniform distribution for $[E]_O$ on G_O and a combination of independent normal and chi variables for ${}_T[E]$. The normal and chi densities determine the constant

$$k(E) = n^{p/2} \prod_2^p A_{n-j} / \prod_2^p A_j .$$

The constant can be substituted and the distribution of the Wishart matrix $S(X)$ simplified:

$$\prod_1^p A_{n-j} (2^p (2\pi)^{pn/2})^{-1} \exp\{-\tfrac{1}{2}(\text{tr } \Sigma^{-1} S(X) - n)\}$$
$$\cdot\, |n^{-1} S(X)|^{(n-p-2)/2} |\Sigma|^{-n/2} dS(X) / n^{p/2} .$$

4. The linear multivariate model. Consider a process with response variables $x_1, \cdots, x_p$. Suppose the response effect of the internal error has been identified and can be described by $e_1, \cdots, e_p$ with error distribution $f(e_1, \cdots, e_p)$ on R^p. Suppose the linearity properties of the process show that a response vector is a linear distortion of a realized error vector:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} = \theta \begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & & \vdots \\ \gamma_{p1} & \cdots & \gamma_{pp} \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_p \end{pmatrix}.$$

For a sequence of n observations on the system, let $\mathbf{x}_j = (x_{11}, \cdots, x_{1n})'$ be the sequence of observation on the jth response. The system and the p observations can then be described by the *linear multivariate model*

$$\prod f(e_{1i}, \cdots, e_{pi}) \prod de_{1i} \cdots de_{pi}$$
$$\begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_p' \end{pmatrix} = \theta \begin{pmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}$$

or

$$f(E)\, dE$$
$$X = \theta E$$

where

$$X = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_p' \end{pmatrix}, \qquad E = \begin{pmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}.$$

The model has an *error distribution* $f(E)\,dE$ describing the internal operation of the process (E is a variable). And the model has a structural equation $X = \theta E$ in which a realized error E has determined the relationship between the known observation X and the unknown θ (E is a constant).

The linear distortions θ form the positive linear group

$$G = \left\{ g = \begin{pmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & & \vdots \\ c_{p1} & \cdots & c_{pp} \end{pmatrix} : \begin{matrix} -\infty < c_{jj} < \infty \\ |C| > 0 \end{matrix} \right\}.$$

For $n \geqq p$ the group is a unitary transformation group on R^{pn} (certain trivial points deleted). Let ${}_T[X]$, $D^*(X)$, $D(X)$, $[X]_O$ be defined as in Section 1 but with the one-vector deleted. Then

$$X = [X]D(X) = {}_T[X][X]_O D(X)$$

and $D(X)$ is a fixed point in the oriented subspace $L^+(\mathbf{x}_1, \cdots, \mathbf{x}_p)$.

The invariant differential on R^{pn} that agrees with Euclidean volume at $D(X)$ is

$$dm(X) = dV/|[X]|^n.$$

And the invariant differential on G that agrees with Euclidean volume at the identity is $dg/|g|^p$.

The affine multivariate model admits identification of the error characteristic

$$D(E) = [E]^{-1}E = [X]^{-1}X = D(X),$$

but gives no information concerning $[E]$ other than its origin from the error distribution. The linear model by its own information content gives the reduced model

$$g([E]: D(X))\, d[E]$$
$$[X] = \theta[E]$$

where

$$g([E]: D)\, d[E] = k(D) f([E]D)|[E]|^n\, d[E]/|[E]|^k.$$

The classical model, that corresponds to the linear multivariable model, describes a response *variable* X. The classical model is

$$f(X: \theta)\, dX = f(\theta^{-1}X)|\theta|^{-n}\, dX.$$

The classical model conditional on the characteristic $D(X) = D$ is

$$g([X]: D, \theta)\, d[X] = g(\theta^{-1}[X]: D)|\theta|^{-p}\, d[X]$$
$$= k(D) f(\theta^{-1}X)|\theta^{-1}[X]|^{n}\, d[X]/|[X]|^{p}.$$

The conventional statistic of multivariate analysis is the sample inner product or Wishart matrix:

$$S(X) = XX' = \begin{pmatrix} \sum x_1^2 & \cdots & \sum x_1 x_p \\ \vdots & & \vdots \\ \sum x_p x_1 & \cdots & \sum x_p^2 \end{pmatrix}$$
$$= {}_T[X][X]_O D(X) D'(X)[X]_O' {}_T[X]' = {}_T[X]\, {}_T[X]'.$$

The triangular matrices ${}_T[X]$ form a positive lower triangular group

$$G_T = \left\{ k = \begin{bmatrix} s_{(1)} & & \cdots & 0 \\ t_{21} & s_{(2)} & \cdots & 0 \\ \vdots & & & \vdots \\ t_{p1} & \cdots & t_{pp-1} & s_{(p)} \end{bmatrix} : \begin{matrix} -\infty < t_{jj'} < \infty \\ 0 < s_{(p)} < \infty \end{matrix} \right\}.$$

And the left invariant differential is

$$d\mu_T(k) = dk/|k|_\Delta.$$

The orthogonal matrices $[X]_O$ form a rotation group

$$G_0 = \left\{ h = \begin{pmatrix} o_{11} & \cdots & o_{1p} \\ \vdots & & \vdots \\ o_{p1} & \cdots & o_{pp} \end{pmatrix} : \begin{matrix} O'O = I \\ |O| > 0 \end{matrix} \right\}.$$

The Euclidean differential dO with O satisfying the constraints is the left invariant differential on G_O. The differentials combine

$$dg/|g|^{p} = dk/|k|_\Delta\, dh$$

where $g = kh$ is the triangular-orthogonal factorization.

The distribution of the inner product matrix $S(X)$ conditional on $D(X) = D$ is

$$k(D) \prod_2^p A_j \bar{f}(\theta^{-1}{}_T[X], D)|\theta|^{-n}|{}_T[X]|^{n}\, dS(X)/2^{p}|[X]|^{p+1}$$

where

$$\bar{f}(\theta^{-1}k, D) = \int_{G_O} f(\theta^{-1}khD)\left(\prod_2^p A_j\right)^{-1} dh$$

is the average of f over the rotation group G_O.

5. The linear model with error symmetry. Now suppose that the error distribution for the linear multivariate model is rotationally symmetric

$$f(h(e_1, \cdots, e_p)') = f((e_1, \cdots, e_p)')$$
$$f(hE) = f(E)$$

for all h in the rotation group G_O. The quantity θ can be factored into triangular and orthogonal components

$$\theta = {}_T[\theta][\theta]_O .$$

And if the error distribution is scaled to give unit variance then the quantity

$$\Sigma = \theta\theta' = {}_T[\theta]\ {}_T[\theta]'$$

is the variance matrix for the response variable $(x_1, \cdots, x_p)$. The distribution of the Wishart matrix $S(X)$ can then be simplified

$$k(D)\prod_2^p A_j \bar{f}({}_T[\theta]^{-1}{}_T[X], D)|\theta|^{-n}|{}_T[X]|^{-n}\ dS(X)/2^p|{}_T[X]|^{(p+1)}$$

where

$$\bar{f}(k, D) = \int_{G_O} f(khD)\ dh/\prod_2^p A_j$$

is the average over the rotations h in G_O.

Now suppose that the error distribution is standard normal:

$$f(E) = (2\pi)^{-pn/2} \exp\{-\tfrac{1}{2} \operatorname{tr} EE'\}.$$

The distribution of ${}_T[E]$ and $[E]_O$ conditional on the observed characteristic $D(E) = D = D(X)$ is

$$k(D)\prod_2^p A_j(2\pi)^{-pn/2} \exp\{-\tfrac{1}{2} \operatorname{tr} {}_T[E]\ {}_T[E]'\}\ |{}_T[E]|^n\ d_T[E]\ d[E]_O/|{}_T[E]|_\Delta$$
$$= \prod_1^p A_{n-j}2^{-p}(2\pi)^{-pn/2} \exp\{-\tfrac{1}{2} \operatorname{tr} {}_T[E]\ {}_T[E]'\}\ |{}_T[E]|^n\ d_T[E]\ d[E]_O/|{}_T[E]|_\Delta .$$

This is a uniform distribution for $[E]_O$ and a combination of normal and chi variables for ${}_T[E]$; hence the value of the constant is determined.

The distribution of the Wishart matrix $S(X)$ conditional on the characteristic $D(X) = D$ is then

$$\prod_1^p A_{n-j}2^{-p}(2\pi)^{-pn/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}S(X)\}\ |S(X)|^{(n-p-1)/2}|\Sigma|^{-n/2}\ dS(X)$$

6. The noncentral Wishart distribution. Consider the linear multivariate model with normal error as developed in the preceding two sections. In some contexts the distribution of the Wishart matrix may be used to calculate a critical value for a test. A power function calculation for such a test would then need the distribution of the Wishart matrix as generated by a noncentral normal distribution. The appropriate distribution would again be a conditional distribution given the observed characteristic $D(X) = D$. This noncentral conditional Wishart distribution is derived in this section.

The distribution for a sample from the multivariable normal with mean $\mu = (\mu_1, \cdots, \mu_k)'$ and variance matrix Σ has element

$$f^*(X)\ dX = (2\pi)^{-pn/2} \exp\{-\tfrac{1}{2} \operatorname{tr} (X - \boldsymbol{\mu}\mathbf{1}')'\Sigma^{-1}(X - \boldsymbol{\mu}\mathbf{1}')\}\ dX$$
$$= (2\pi)^{-pn/2} \exp\{-\tfrac{1}{2} \operatorname{tr} (\Sigma^{-1}S(X) + n\Sigma^{-1}\boldsymbol{\mu}\boldsymbol{\mu}')\} \exp\{\operatorname{tr} \mathbf{1}\boldsymbol{\mu}'\Sigma^{-1}X\}\ dX.$$

Let θ be a transformation in G such that $\theta\theta' = \Sigma$. The preceding distribution for X can be obtained by a transformation θ applied to a substitute error distribution with mean $\theta^{-1}\mathfrak{u}$ and variance matrix I. It follows that the noncentral Wishart distribution for $S(X) = XX'$ can then be obtained from the Wishart formula in Section 4 by using the substitute error distribution; the noncentral Wishart distribution is

$$k^*(D) \exp\{-\tfrac{1}{2} \operatorname{tr} (\Sigma^{-1}S(X) + n\Sigma^{-1}\mathfrak{u}\mathfrak{u}')|S(X)|^{(n-p-1)/2}|\Sigma|^{-n/2} \cdot \int_{G_O} \exp\{\operatorname{tr} \mathbf{1}\mathfrak{u}'\Sigma^{-1}\, {}_T[X]hD\}\, dh/\textstyle\prod_2^p A_j \cdot dS(X)$$

where the average is applicable if $p \geqq 2$. An average over orthogonal rotations is available for the noncentral chi-square distribution; the general evaluation of the average of an expression $\exp\{\operatorname{tr} Wh\}$ for h in an orthogonal group is treated by James (1955). The average is

$$\int_{G_O} \exp\{\operatorname{tr} \mathbf{1}\mathfrak{u}'\Sigma^{-1}\, {}_T[X]hD\}\, dh/\textstyle\prod_2^p A_j = \int_{G_O} \exp\{\operatorname{tr} D\mathbf{1}\mathfrak{u}'\Sigma^{-1}\, {}_T[X]h\}\, dh/\prod_2^p A_j$$
$$= \int_{G_O} \exp\{\operatorname{tr} \mathbf{t}\mathbf{u}h\}\, dh/\textstyle\prod_2^p A_j$$

where

$$\mathbf{t} = D\mathbf{1}, \qquad \mathbf{u} = {}_T[X]'\Sigma^{-1}\mathfrak{u}.$$

Let h_1 be a rotation with first column $\mathbf{t}/|\mathbf{t}|$; let h_2 be a rotation with first row $\mathbf{u}'/|\mathbf{u}|$; let

$$\delta = \delta(\mathfrak{u}, \Sigma, D) = |\mathbf{u}|\, |\mathbf{t}|;$$

and let I_{11} be the matrix

$$I_{11} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & & \cdots & 0 \end{pmatrix}.$$

Then the average is

$$\int_{G_O} \exp\{\delta \operatorname{tr} h_1 I_{11} h_2 h\} dh/\textstyle\prod_2^p A_j = \int_{G_O} \exp\{\delta \operatorname{tr} I_{11}h^*\}\, dh^*/\prod_2^p A_j$$
$$= \int \exp\{o_{11}\}\, d\mathbf{o}/A_p$$

where $\mathbf{o}$ is a vector uniformly distributed over the unit sphere in R^p. The average is

$$\textstyle\sum_{s=0}^{\infty} (4^s s!)^{-1}\Gamma(p/2)(\Gamma(p/2 + s))^{-1}\delta^{2s}$$

as extracted from the noncentral chi distribution. Thus the probability element of the noncentral Wishart, conditional on the characteristic $D(X) = D$, is

$$k^* \textstyle\sum_{s=0}^{\infty} \Gamma(r/2)(\Gamma(r/2 + s))^{-1}\delta^{2s}(4^s s!)^{-1} \exp\{-\tfrac{1}{2} \operatorname{tr} (\Sigma^{-1}S(X) + n\Sigma^{-1}\mathfrak{u}\mathfrak{u}'\} \cdot |S(X)|^{(n-p-1)/2}|\Sigma|^{-n/2}\, dS(X)$$

for $p \geqq 2$ and

$$\delta = |D(X)\mathbf{1}|\, |{}_T[X]'\Sigma^{-1}\mathfrak{u}|.$$

This is an ordinary noncentral Wishart distribution but the noncentrality depends on the observed characteristic $D(X)$.

The probability element for the real variable $S(X) = s_{(1)}^2(X)$ $(p = 1)$ conditional on $D(X) = D$ needs no average; it is

$$k^* \exp\{\textstyle\sum_1^n d_i(x)\mu\sigma^{-2}S^{\frac{1}{2}}(\mathbf{x})\} \exp\{-\tfrac{1}{2}(S(X)\sigma^{-2} + n\mu^2\sigma^{-2}\} S^{(n-2)/2}(X)\sigma^{-n}\, dS(X).$$

NOTE. The affine multivariate model and the linear multivariate model are examples of structural models, some results for structural models may be found in Fraser (1966), (1967).

REFERENCES

FRASER, D. A. S. (1966). Structural probability and a generalization. *Biometrika* **53** 1–9.

FRASER, D. A. S. (1967). Data transformations and the linear model. *Ann. Math. Statist.* **38** 1456–1465.

JAMES, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

JAMES, A. T. (1955). A generating function for averages over the orthogonal group. *Proc. Royal Soc. London A* **229** 367–375.