Exact Distributions of Scaled Multivariate Normal Residuals

Gordon M. Kaufman*

February 1998

Sloan Working Paper #4000

Abstract

Densities of scaled multivariate normal vector and matrix residuals are shown to possess simple, spherically symmetric functional forms.

1 Introduction

Quiroz and Dudley (1991) [1] develop a class of tests for multivariate normality whose properties are determined by a sample statistic defined as follows: let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, be a realization of $n(p \times 1)$ random vectors (rvs) $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ with common range $\Re^P$. Define the sample mean to be $\overline{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ and the *unscaled* $(p \times p)$ sample covariance matrix.

$$\mathbf{S} = \sum_{j=1}^{n}[\mathbf{x}_j - \overline{\mathbf{x}}][\mathbf{x}_j - \overline{\mathbf{x}}]^t. \tag{1.1}$$

With $(p \times n)\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $(p \times n)\overline{\mathbf{X}} = [\overline{\mathbf{x}}, \ldots, \overline{\mathbf{x}}]$,

$$\mathbf{S} = [\mathbf{X} - \overline{\mathbf{X}}][\mathbf{X} - \overline{\mathbf{X}}]^t. \tag{1.2}$$

*Professor of Operations Research and Management, Sloan School of Management, Massachusetts Institute of Technology. Professor Kaufman's work was supported by the Air Force Office of Scientific Research, Contract #AFOSR-F49620-1-0307.

One statistic of interest to Quiroz and Dudley is a sample standardized version of a generic $\mathbf{x}_j$:

$$\mathbf{x}_j \longrightarrow \mathbf{y}_j \equiv \sqrt{n-1}\mathbf{S}^{-\frac{1}{2}}[\mathbf{x}_j - \overline{\mathbf{x}}]. \tag{1.3}$$

They show that if $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ and iid $\mathrm{N}(\mu, \mathbf{\Sigma}), \mathbf{\Sigma} > 0$, then the right-hand side of (1.3) is representable as $\tilde{\mathbf{Q}}\tilde{\mathbf{S}}_U^{-\frac{1}{2}}[\tilde{\mathbf{u}}_j - \tilde{\overline{\mathbf{u}}}]$; here $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n$ are iid $\mathrm{N}(\mathbf{0}, \mathbf{I}), \tilde{\overline{\mathbf{u}}} = \frac{1}{n}\sum_{j=1}^{n} \tilde{\mathbf{u}}_j, \tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n], \tilde{\overline{\mathbf{U}}} = [\tilde{\overline{\mathbf{u}}}, \ldots, \tilde{\overline{\mathbf{u}}}], \tilde{\mathbf{S}}_U \equiv \frac{1}{n-1}[\tilde{\mathbf{U}} - \tilde{\overline{\mathbf{U}}}][\tilde{\mathbf{U}} - \tilde{\overline{\mathbf{U}}}]^t$ and $\tilde{\mathbf{Q}}$ is a random orthogonal matrix depending on $\mathbf{\Sigma}$ and $\tilde{\mathbf{X}}$. Several authors [Small (1978), Quiroz and Dudley (1991)] [2, 1] call the statistic (1.3) *scaled residuals* and use this statistic as a springboard for some new tests of multivariate normality. On page 544 (op. cit.) they say, "By the way, we do not know whether the distribution of $\{\mathbf{S}^{\frac{1}{2}}(\mathbf{x}_i - \overline{\mathbf{x}}), i = 1, \ldots, n\}$ depends on $\mathbf{\Sigma}$ (it clearly doesn't depend on μ." [The answer is: this distribution does not depend on $\mathbf{\Sigma}$.] To our knowledge, the exact distribution of (1.3) has not appeared in the literature.

Our aim is to derive the exact density of a standardized version of a generic $\tilde{\mathbf{x}}_j$ like (1.3) and its natural generalizations when $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ are iid $\mathrm{N}(\mu, \mathbf{\Sigma}), \mathbf{\Sigma} > 0$ so that test statistics can be easily computed. Throughout $(p \times p)\mathbf{\Sigma} > 0$ means either $\mathbf{\Sigma}$ is PDS or $\{\mathbf{\Sigma} \mid (p \times p)\mathbf{\Sigma} > \mathbf{0}\}$ and $\mathbf{I} > \mathbf{Z} > \mathbf{0}$ means that both $\mathbf{I} - \mathbf{Z} > \mathbf{0}$ and $\mathbf{Z} > \mathbf{0}$ and $\mathbf{\Sigma} \otimes \mathbf{B}$ denotes the Kroenecker product of $\mathbf{\Sigma}$ and $\mathbf{B}$.

The standardized versions of $\tilde{\mathbf{x}}_j$ and of $[\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_q], 2 \leq q < n$ we study possess spherically symmetric densities (Propositions 1, 2, and 3).

2 Densities

Proposition 1 *If $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ are iid $N(\mu, \mathbf{\Sigma}), \mathbf{\Sigma} > 0$ then with $\gamma \equiv \frac{1}{2}(p+1)$, when $n > p+1$ the rv $\tilde{\mathbf{y}}_1 \equiv \tilde{\mathbf{S}}^{-\frac{1}{2}} \left(\frac{n}{n-1}\right)^{\frac{1}{2}} \left[\tilde{\mathbf{x}}_1 - \tilde{\overline{\mathbf{x}}}\right]$ has density*

$$\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\pi^{\frac{1}{2}p}\Gamma\left(\frac{1}{2}(n-p-1)\right)} \quad (1-\mathbf{y}^t\mathbf{y})^{\frac{1}{2}(n-2)-\gamma} \tag{2.1}$$

with support $\{\mathbf{y} \mid \mathbf{y} \in \Re^p, \mathbf{y}^t\mathbf{y} \leq 1\}$.

The vector $\tilde{\mathbf{s}} = (\tilde{s}_1, \ldots, \tilde{s}_p)$ with squared values $\tilde{s}_1 = \tilde{y}_1^2, \ldots, \tilde{s}_p = \tilde{y}_p^2$ of elements of $\tilde{\mathbf{y}}_1$ possesses density

$$\frac{\Gamma\left(\frac{1}{2}(n-1)\right)}{\pi^{\frac{1}{2}p}\Gamma\left(\frac{1}{2}(n-p-1)\right)} \quad s_1^{-\frac{1}{2}} \ldots s_p^{-\frac{1}{2}}(1-(s_1+\cdots+s_p))^{\frac{1}{2}(n-2)-\gamma} \tag{2.2}$$

with support $\{\mathbf{s} \mid 0 \leq s_1+\cdots+s_p \leq 1\}$. The density (2.2) is a p-dimensional Dirichlet distribution indexed by parameter $\left(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}(n-p)\right)$. By convention $D_p\left(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}(n-p)\right)$ will denote density (2.2).

If we partition $\mathbf{Y}= [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ into $(p \times q)\mathbf{Y}_1 = [\mathbf{y}_1, \ldots, \mathbf{y}_q]$ and $(p \times (n-q))\mathbf{Y}_2 = [\mathbf{y}_{q+1}, \ldots, \mathbf{y}_n]$, both the marginal distribution of $\tilde{\mathbf{Y}}_1$ for $2 \leq q < n$ and the conditional distribution of $\tilde{\mathbf{Y}}_1$ given $\tilde{\mathbf{Y}}_2 = \mathbf{Y}_2$ can be derived by the same method used to calculate the marginal density of $\tilde{\mathbf{y}}_j$, a column of $\mathbf{Y}$. Successive conditioning of the joint distribution of $\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_q$ yields q mutually independent but not identically distributed rvs, each with a density of the form (2.1) [Propositions 5 and 6].

Consider the joint density of $(p \times n)\tilde{\mathbf{X}}$ written as

$$(2\pi)^{-\frac{1}{2}pn} \mid \mathbf{\Sigma} \mid^{-\frac{1}{2}n} e^{-\frac{1}{2}tr\mathbf{\Sigma}^{-1}[\mathbf{S}+n(\overline{\mathbf{x}}-\mu)(\overline{\mathbf{x}}-\mu)^t]} \tag{2.3}$$

with $\mathbf{S}$ and $\overline{\mathbf{x}}$ as defined at the outset. Then

1. $\tilde{\mathbf{S}}$ is Wishart with $(n-1)df$,

2. $\tilde{\overline{\mathbf{x}}}$ is Normal,

3. $\tilde{\mathbf{S}}$ and $\tilde{\overline{\mathbf{x}}}$ are independent.

However, $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{x}}_1 - \tilde{\overline{\mathbf{x}}}$ are not independent and it is this lack of independence that makes exact computation of the density of $\tilde{\mathbf{S}}^{-\frac{1}{2}}[\tilde{\mathbf{x}}_1 - \tilde{\overline{\mathbf{x}}}]$ interesting.

To prove Proposition 1 we begin by splitting $\mathbf{S}$ into two pieces; one composed of $(\mathbf{x}_1 - \overline{\mathbf{x}})(\mathbf{x}_1 - \overline{\mathbf{x}})^t$ and the other composed of a centered on the mean $\overline{\mathbf{x}}_{.1} \equiv \dfrac{1}{n-1}\sum_{j=2}^{n} \mathbf{x}_j$ version of $\mathbf{S}$ with $\mathbf{x}_1 - \overline{\mathbf{x}}$ deleted. The rv $\tilde{\mathbf{x}}_1 - \tilde{\overline{\mathbf{x}}}$ is then probabilistically independent of $\tilde{\mathbf{S}}_{.1} \equiv \sum_{j=2}^{n} \left[\tilde{\mathbf{x}}_j - \tilde{\overline{\mathbf{x}}}_{.1}\right]\left[\tilde{\mathbf{x}}_j - \tilde{\overline{\mathbf{x}}}_{.1}\right]^t$.

The factorization of $\mathbf{S}$ we have in mind is

$$\mathbf{S} = \mathbf{S}_{.1} + \frac{n}{n-1}(\mathbf{x}_1 - \overline{\mathbf{x}})(\mathbf{x}_1 - \overline{\mathbf{x}})^t \tag{2.4}$$

with

$$\mathbf{S}_{.1} = \sum_{j=2}^{n}(\mathbf{x}_j - \overline{\mathbf{x}}_{.1})(\mathbf{x}_j - \overline{\mathbf{x}}_{.1})^t. \tag{2.5}$$

To see that (2.4) obtains, add and subtract $\overline{\mathbf{x}}_{.1}$, in $(\mathbf{x}_1 - \overline{\mathbf{x}}), j = 2, \ldots, n$. The cross product terms vanish and as

$$\overline{\mathbf{x}}_{.1} - \overline{\mathbf{x}} = -\frac{1}{n-1}(\mathbf{x}_1 - \overline{\mathbf{x}}),$$

(2.4) follows. Consequently, with $N = n-1$, we can write the joint density of $\tilde{\mathbf{X}}$ in terms of $\tilde{\mathbf{S}}_{.1}, \tilde{\overline{\mathbf{x}}}$ and $\mathbf{u} \equiv \left(\frac{n}{n-1}\right)^{\frac{1}{2}}(\mathbf{x}_1 - \overline{\mathbf{x}})$ as

$$(2\pi)^{-\frac{1}{2}pN}|\mathbf{\Sigma}|^{-\frac{1}{2}N}e^{-\frac{1}{2}tr\mathbf{\Sigma}^{-1}[\mathbf{S}_{.1}+\mathbf{u}\mathbf{u}^t]} \times (2\pi)^{-\frac{1}{2}p}|\mathbf{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}n(\overline{\mathbf{x}}-\mu)^t\mathbf{\Sigma}^{-1}(\overline{\mathbf{x}}-\mu)}. \tag{2.6}$$

In turn,

1. $\tilde{\mathbf{S}}_{.1}$ is Wishart with $N - Idf$,

2. $\tilde{\mathbf{u}}$ is Normal mean $\mathbf{0}$ and variance $\mathbf{\Sigma}$,

3. $\tilde{\bar{\mathbf{x}}}$ is Normal mean μ and variance $\frac{1}{n}\mathbf{\Sigma}$,

and all three rvs are independent. The joint density of $\tilde{\mathbf{S}}_{.1}$ and $\tilde{\mathbf{u}}$ is

$$(2\pi)^{-\frac{1}{2}p}c_p(N-1)|\mathbf{\Sigma}|^{-\frac{1}{2}N}e^{-\frac{1}{2}tr\mathbf{\Sigma}^{-1}[\mathbf{S}_{.1}+\mathbf{u}\mathbf{u}^t]}|\mathbf{S}_{.1}|^{\frac{1}{2}(N-1)\gamma} \tag{2.7}$$

and the constant

$$1/c_p(N-1) = 2^{\frac{1}{2}p(N-1)}\pi^{\frac{1}{4}p(p-1)}\prod_{j=1}^{p}\Gamma\left(\frac{1}{2}(N-j)\right). \tag{2.8}$$

With $\mathbf{y}_1 \equiv \mathbf{S}^{-\frac{1}{2}}\mathbf{u}$ and $\mathbf{S} = \mathbf{S}_{.1} + \mathbf{u}\mathbf{u}^t$, the Jacobian $J((\mathbf{S}_{.1}, \mathbf{u}) \longrightarrow (\mathbf{S}, \mathbf{y}))$ of the transformation from $\mathbf{S}_{.1}$ and $\mathbf{u}$ to $\mathbf{S}$ and $\mathbf{y}$ is $|\mathbf{S}|^{\frac{1}{2}}$ and

$$|\mathbf{S}_{.1}| = |\mathbf{S} - \mathbf{u}\mathbf{u}^t| = |\mathbf{S}| \cdot |\mathbf{I} - \mathbf{y}\mathbf{y}^t|. \tag{2.9}$$

Consequently, $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{y}}$ have joint density

$$(2\pi)^{-\frac{1}{2}p}c_p(N-1)|\mathbf{\Sigma}|^{-\frac{1}{2}N}e^{-\frac{1}{2}tr\mathbf{\Sigma}^{-1}\mathbf{S}}|\mathbf{S}|^{\frac{1}{2}N-\gamma} \times |\mathbf{I} - \mathbf{y}\mathbf{y}^t|^{\frac{1}{2}(N-1)-\gamma}. \tag{2.10}$$

As $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{y}}$ are independent, upon integrating over $\mathbf{S} > \mathbf{0}$, $\tilde{\mathbf{y}}$ has marginal density

$$(2\pi)^{-\frac{1}{2}p}\frac{c_p(N-1)}{c_p(N)}|\mathbf{I} - \mathbf{y}\mathbf{y}^t|^{\frac{1}{2}(n-2)-\gamma} \tag{2.11}$$

with range $\{\mathbf{y} \mid \mathbf{y} \in \Re^p \text{ and } \mathbf{y}^t\mathbf{y} \leq 1\}$. As $c_p(N-1)/c_p(N) = 2^{\frac{1}{2}p}\prod_{j=1}^{p}\Gamma\left(\frac{1}{2}(N-j+1)\right)/\Gamma\left(\frac{1}{2}(N-j)\right) = 2^{\frac{1}{2}p}\Gamma\left(\frac{1}{2}N\right)/\Gamma\left(\frac{1}{2}(N-p)\right)$ and $|\mathbf{I} - \mathbf{y}\mathbf{y}^t| = (1 - \mathbf{y}^t\mathbf{y})$ we have proven (2.1).

To see that the range of $\tilde{\mathbf{y}}$ is restricted to the set of $\mathbf{y} \in \Re^p$ such that $\mathbf{y}\mathbf{y}^t \leq 1$ first write $\mathbf{y}^t\mathbf{y} = \mathbf{u}^t[\mathbf{S}_{.1} + \mathbf{u}\mathbf{u}^t]^{-1}\mathbf{u}$, then diagonalize $\mathbf{S}_{.1}$. Because $\mathbf{S}_{.1} > \mathbf{0}$ there is a $(p \times p)$ orthogonal $\mathbf{Q}$ such that $\mathbf{Q}^t\mathbf{\Lambda}\mathbf{Q} = \mathbf{S}_{.1}$, $(p \times p)\mathbf{\Lambda} > \mathbf{0}$ and diagonal. With $\mathbf{z} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}\mathbf{u}, \mathbf{y}^t\mathbf{y} = \mathbf{z}^t[\mathbf{I} + \mathbf{z}\mathbf{z}^t]^{-1}\mathbf{z}$. For any $\mathbf{z} \in \Re^p$ there is an orthogonal $\mathbf{P}(\mathbf{z})$ and a scalar $\alpha(\mathbf{z})$ such that $\mathbf{P}(\mathbf{z})\mathbf{z} = (\alpha(\mathbf{z}), 0, \ldots, 0)^t$. Hence, $\mathbf{y}^t\ \mathbf{y} = \frac{\alpha^2(\mathbf{z})}{(1+\alpha^2(\mathbf{z}))} \leq 1$. This finishes proof of Proposition 1.

For $p \leq n$, let $V_{p,n}$ be the collection of all p-tuples of $(n \times 1)$ orthonormal vectors in $\Re^p$ – a Stieffel manifold. For each $(p \times n)\mathbf{Q}^t \in V_{p,n}$, $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}_p$. In turn, if $(p \times n)\mathbf{X} \in \Re^{pn}$, the Jacobian of the transformation from $\mathbf{X}$ to $(\mathbf{Q}, \mathbf{R})$ such that $\mathbf{X} = \mathbf{R}^{\frac{1}{2}}\mathbf{Q}^t$ for $(p \times p)\mathbf{R} > \mathbf{0}$ is $2^{-p}|\mathbf{R}|^{\frac{1}{2}(n-p-1)}$ and

$$\int\limits_{V_{p,n}} d\mathbf{Q}^t = \frac{2^p \pi^{\frac{1}{2}(p-n)}}{\pi^{\frac{1}{4}p(p-1)} \prod_{j=1}^{p} \Gamma\left(\frac{1}{2}(n-j+1)\right)} \equiv vol(V_{p,n}). \tag{2.12}$$

These facts play a role in the computation of generalizations of (2.11). We can use (2.12) to check the normalizing constant for the density (2.1). As y is $(p \times 1)$ and

$$vol(V_{1,p}) = \frac{2\pi^{\frac{1}{2}p}}{\Gamma\left(\frac{1}{2}p\right)} \tag{2.13}$$

upon transforming from $\mathbf{y}$ to $(\mathbf{q}, r)$ with $\mathbf{y} = r^{\frac{1}{2}}\mathbf{q}$ and $\mathbf{q} \in V_{1,p}$,

$$\begin{aligned} &\int\limits_{\{\mathbf{y}|\mathbf{y}\in\Re^p, \mathbf{y}^t\mathbf{y}\leq 1\}} (1-\mathbf{y}\mathbf{y}^t)^{\frac{1}{2}(n-2)-\gamma} d\mathbf{y} \\ &= \frac{1}{2}\int\limits_0^1 (1-r)^{\frac{1}{2}(n-p-1)-1} r^{\frac{1}{2}p-1} dr \;\cdot\; vol(V_{1,p}) = \frac{\pi^{\frac{1}{2}p}\Gamma\left(\frac{1}{2}(N-p)\right)}{\Gamma\left(\frac{1}{2}N\right)}. \end{aligned} \tag{2.14}$$

Proposition 2 *For $1 \leq q < n-p$ the density of $(p \times q)\tilde{\mathbf{Y}} \equiv \mathbf{S}^{-\frac{1}{2}}[(\tilde{\mathbf{x}}_1 - \tilde{\bar{\mathbf{x}}}) \cdots (\tilde{\mathbf{x}}_q - \tilde{\bar{\mathbf{x}}})]$ is*

$$(2\pi)^{-\frac{1}{2}pq} \frac{c_p(N-q)}{c_p(N)} |\mathbf{A}|^{\frac{1}{2}p} \;\cdot\; |\mathbf{I} - \mathbf{Y}\mathbf{A}\mathbf{Y}^t|^{\frac{1}{2}(N-q)-\gamma} \tag{2.15}$$

with $(q \times q)$

$$\mathbf{A} = \mathbf{I} + \frac{1}{n-q}\mathbf{1}\mathbf{1}^t \tag{2.16}$$

or

$$(2\pi)^{-\frac{1}{2}pq} \frac{c_p(N-q)}{c_p(N)} |\mathbf{A}|^{\frac{1}{2}(N-q-1)} \;\cdot\; |\mathbf{A}^{-1} - \mathbf{Y}^t\mathbf{Y}|^{\frac{1}{2}(N-q)-\gamma} \tag{2.17}$$

with

$$\mathbf{A}^{-1} = \mathbf{I} + \frac{1}{n}\mathbf{1}\mathbf{1}^t \tag{2.18}$$

The range of $\tilde{\mathbf{Y}}$ is $\{\mathbf{Y}|\mathbf{Y} \in \Re^{pq}$ and $\mathbf{I} > \mathbf{Y}\mathbf{A}\mathbf{Y}^t > \mathbf{0}\}$ if $p \leq q$ and is $\{\mathbf{Y}|\mathbf{Y} \in \Re^{pq}$ and $\mathbf{A}^{-1} > \mathbf{Y}^t\mathbf{Y} > \mathbf{0}\}$ if $p > q$.

The scale factor $\left(\frac{n}{n-1}\right)^{\frac{1}{2}}$ appearing in the definition of $\mathbf{y}_1 = \mathbf{S}^{-\frac{1}{2}}\left(\frac{n}{n-1}\right)^{\frac{1}{2}}[\mathbf{x}_1 - \overline{\mathbf{x}}]$ in Proposition 1 appears in $\mathbf{A}$ for $q = 1$. We omit it from the definitions of $(p \times q)\mathbf{Y}$ in Proposition 2. It reappears when the marginal density of the first column of $\tilde{\mathbf{Y}}$ is calculated from (2.15). For $2 \leq q < n - p$ the marginal densities (2.15) and (2.17) can be computed in the same fashion as (2.1). Factor $\mathbf{S}$ into

$$\sum_{j=1}^{q}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^t + \sum_{k=q+1}^{n}(\mathbf{x}_k - \overline{\mathbf{x}})(\mathbf{x}_k - \overline{\mathbf{x}})^t.$$

Define $\mathbf{x}_{.q} = \frac{1}{n-q}\sum_{k=q+1}^{n}\mathbf{x}_k$ and set

$$\mathbf{S}_{.q} = \sum_{k=q+1}^{n}(\mathbf{x}_k - \overline{\mathbf{x}}_{.q})(\mathbf{x}_k - \overline{\mathbf{x}}_{.q})^t. \tag{2.19}$$

In terms of $\mathbf{x}_{.q}$ and $\mathbf{S}_{.q}$

$$\begin{aligned} \mathbf{S} &= \sum_{j=1}^{q}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^t + \mathbf{S}_{.q} \\ &+ (n-q)(\overline{\mathbf{x}}_{.q} - \overline{\mathbf{x}})(\overline{\mathbf{x}}_{.q} - \overline{\mathbf{x}})^t. \end{aligned}$$

With $(q \times 1)\mathbf{1} = (1 \ldots 1)^t$ and $(p \times q)\overline{\mathbf{X}}_q = [\overline{\mathbf{x}} \ldots \overline{\mathbf{x}}]$

$$\begin{aligned} \overline{\mathbf{x}}_{.q} - \overline{\mathbf{x}} &= \frac{1}{n-q}(q\overline{\mathbf{x}} - [\mathbf{x}_1 + \cdots + \mathbf{x}_q]) \\ &= -\frac{1}{n-q}[(\mathbf{x}_1 - \overline{\mathbf{x}}) + \cdots + (\mathbf{x}_q - \overline{\mathbf{x}})] \\ &= -\frac{1}{n-q}[\mathbf{X}_q - -\overline{\mathbf{X}}_q]\mathbf{1} \end{aligned}$$

so that

$$\mathbf{S} = \mathbf{S}_{.q} + [\mathbf{X}_q - \overline{\mathbf{X}}_q]\mathbf{A}[\mathbf{X}_q - \overline{\mathbf{X}}_q]^t. \tag{2.20}$$

Equipped with (2.20), the joint density of $\overline{\mathbf{X}}$ in terms of $\mathbf{S}_{.q}, \overline{\mathbf{x}}$ and $\mathbf{U} = [\mathbf{X}_q - \overline{\mathbf{X}}_q]$ may be written as

$$\begin{aligned} &(2\pi)^{-\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}e^{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}[\mathbf{S}_{.q}+\mathbf{UAU}^t]} \\ \times\quad &(2\pi)^{-\frac{1}{2}p}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\overline{\mathbf{x}}-\mu)^t\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\mu)}. \end{aligned} \tag{2.21}$$

For $2 \leq q < n - p$,

1. $\tilde{\mathbf{S}}_{.q}$ is Wishart with $N - q$ d.f.,

2. $\tilde{\mathbf{U}}$ has a Normal distribution which, when expressed in terms of the $(pq \times 1)$ vector $(\tilde{\mathbf{u}}_1^t, \ldots, \tilde{\mathbf{u}}_q^t)^t$ of columns of $\tilde{\mathbf{U}}$, has mean $\mathbf{0}$ and variance $\mathbf{A}^{-1} \otimes \boldsymbol{\Sigma}$ and

3. $\tilde{\overline{\mathbf{x}}}$ is Normal mean μ and variance $\boldsymbol{\Sigma}$, and

4. $\tilde{\mathbf{S}}_{.q}, \mathbf{U}$ and $\tilde{\overline{\mathbf{x}}}$ are independent.

The joint density of $\tilde{\mathbf{S}}_{.q}$ and $\tilde{\mathbf{U}}$ is

$$(2\pi)^{-\frac{1}{2}pq}c_p(N-q)|\boldsymbol{\Sigma}|^{-\frac{1}{2}(N-q)} \cdot |\mathbf{A}^{-1} \otimes \boldsymbol{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}[\mathbf{S}_{.q}+\mathbf{UAU}^t]}|\mathbf{S}_{.q}|^{\frac{1}{2}(N-q)-\gamma}. \tag{2.22}$$

With $\mathbf{Y} \equiv \mathbf{S}^{-\frac{1}{2}}\mathbf{U}$ and $\mathbf{S} = \mathbf{S}_{.q} + \mathbf{UAU}^t$, the Jacobian $J((\mathbf{S}_{.q}, \mathbf{U}) \longrightarrow (\mathbf{S}, \mathbf{Y}_1)) = |\mathbf{S}|^{\frac{1}{2}q}$ and

$$|\mathbf{S}_{.q}| = |\mathbf{S} - \mathbf{UAU}^t| = |\mathbf{S}| \cdot |\mathbf{I} - \mathbf{YAY}^t|$$

so that $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{Y}}$ have joint density

$$(2\pi)^{-\frac{1}{2}pq}c_p(N-q)|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}e^{-\frac{1}{2}tr\boldsymbol{\Sigma}^{-1}\mathbf{S}}|\mathbf{S}|^{\frac{1}{2}N-\gamma} \times |\mathbf{A}|^{\frac{1}{2}p} \cdot |\mathbf{I} - \mathbf{YAY}^t|^{\frac{1}{2}(N-q)-\gamma}. \tag{2.23}$$

As $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{Y}}$ are independent, upon integrating over $\tilde{\mathbf{S}} > \mathbf{0}$ we find that $\tilde{\mathbf{Y}}$ has marginal density

$$(2\pi)^{-\frac{1}{2}pq}\frac{c_p(N-q)}{c_p(N)}|\mathbf{A}|^{\frac{1}{2}p} \cdot |\mathbf{I}_p - \mathbf{YAY}^t|^{\frac{1}{2}(N-q)-\gamma}. \tag{2.24}$$

By use of the determinantal identity

$$|\mathbf{I} - \mathbf{YAY}^t| = |\mathbf{A}| \cdot |\mathbf{A}^{-1} - \mathbf{Y}^t\mathbf{Y}| \tag{2.25}$$

we can write (2.24) as shown in Proposition 2, (2.17).

The joint density (2.23) possesses a nice factorization into a marginal density (2.1) for $\tilde{\mathbf{y}}_1$ and a conditional on $\tilde{\mathbf{y}}_1 = \mathbf{y}_1$ density for $\tilde{\mathbf{Y}}_{12} \equiv [\tilde{\mathbf{y}}_2 \ldots \tilde{\mathbf{y}}_q]$ that leads to

Proposition 3 *Partition* $(p \times q)\mathbf{Y} = [\mathbf{y}_1 \mathbf{Y}_{12}]$ *with* $\mathbf{y}_1(p \times 1)$ *and*

$$(q \times q)\mathbf{A} = \begin{bmatrix} a_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

with a_{11} *scalar. Define* $a_{11.2} = a_{11} - \mathbf{a}_{12}\mathbf{A}_{22}^{-1}\mathbf{a}_{21}$ *and*

$$p \times (q-1)\mathbf{Z} = [\mathbf{I} - \mathbf{y}_1 a_{11.2} \mathbf{y}_1^t]^{-\frac{1}{2}}[\mathbf{Y}_{12} + \mathbf{y}_1 \mathbf{a}_{12} \mathbf{A}_{22}^{-1}]. \tag{2.26}$$

Then via the definition (2.16) of $\mathbf{A}$,

1. $a_{11.2} = \frac{n}{n-1}$,
2. $\mathbf{a}_{12}\mathbf{A}_{22}^{-1} = \frac{1}{n-1}\mathbf{1}^t$ *with* $(1 \times (q-1))\mathbf{1}^t$,
3. *the rvs* $\tilde{\mathbf{y}}_1$ *and* $\tilde{\mathbf{Z}}$ *are independent,*
4. $\tilde{\mathbf{y}}_1$ *has marginal density (2.1), and*
5. $\tilde{\mathbf{Z}}$ *has density*

$$(2\pi)^{-\frac{1}{2}p(q-1)} \frac{c_p(N-q)}{c_p(N-1)} |\mathbf{A}_{22}|^{\frac{1}{2}p} \, |\mathbf{I} - \mathbf{Z}\mathbf{A}_{22}\mathbf{Z}^t|^{\frac{1}{2}(N-q-1)-\gamma}. \tag{2.27}$$

To establish Proposition 3, begin by partitioning $(p \times q)\mathbf{Y} = [\mathbf{y}_1 \mathbf{Y}_{12}]$ with $\mathbf{y}_1(p \times 1)$ and partitioning $(q \times q)$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

with $a_{11}(1 \times 1)$. Then

$$\mathbf{YAY}^t = a_{11}\mathbf{y}_1\mathbf{y}_1^t + \mathbf{Y}_{12}\mathbf{a}_{21}\mathbf{y}_1^t + \mathbf{y}_1\mathbf{a}_{12}\mathbf{Y}_{12}^t + \mathbf{Y}_{12}\mathbf{A}_{22}\mathbf{Y}_{12}^t.$$

Complete the square in $\mathbf{Y}_{12}$:

$$\mathbf{YAY}^t = [\mathbf{Y}_{12} + \mathbf{y}_1\mathbf{a}_{12}\mathbf{A}_{22}^{-1}]\ \mathbf{A}_{22}\ [\mathbf{Y}_{12} + \mathbf{y}_1\mathbf{a}_{12}\mathbf{A}_{22}^{-1}]^t - \mathbf{y}_1\mathbf{a}_{12}\mathbf{A}_{22}^{-1}\mathbf{a}_{21}\mathbf{y}_1^t. \tag{2.28a}$$

Then,

$$\mathbf{I} - \mathbf{YAY}^t = \mathbf{I} - \mathbf{y}_1 a_{11.2}\mathbf{y}_1^t - [\mathbf{Y}_{12} + \mathbf{y}_1\mathbf{a}_{12}\mathbf{A}_{22}^{-1}]\ \mathbf{A}_{22}^t\ [\mathbf{Y}_{12} + \mathbf{y}_1\mathbf{a}_{12}\mathbf{A}_{22}^{-1}]^t \tag{2.28b}$$

so that

$$|\mathbf{I} - \mathbf{YAY}^t| = |\mathbf{I} - \mathbf{y}_1 a_{11.2}\mathbf{y}_1^t|\ \cdot\ |\mathbf{I} - \mathbf{ZA}_{22}\mathbf{Z}^t|. \tag{2.29}$$

The Jacobian of the transformation from $(\mathbf{y}_1, \mathbf{Y}_{12})$ to $(\mathbf{y}_1, \mathbf{Z})$ is $|\mathbf{I} - \mathbf{y}_1 a_{11.2}\mathbf{y}_1^t|^{\frac{1}{2}(q-1)}$ so that the joint density of $(p \times 1)\tilde{\mathbf{y}}_1$ and $(p \times (q-1))\tilde{\mathbf{Z}}$ is proportional to

$$|\mathbf{I} - \mathbf{y}_1 a_{11.2}\mathbf{y}_1^t|^{\frac{1}{2}(N-1)-\gamma}\ \cdot\ |\mathbf{I} - \mathbf{ZA}_{22}\mathbf{Z}^t|^{\frac{1}{2}(N-q)-\gamma}. \tag{2.30}$$

As $a_{11} = \frac{n-q+1}{n-q}, \mathbf{a}_{12} = \mathbf{a}_{21}^t = \frac{1}{n-q}\mathbf{1}_{q-1}^t, \mathbf{A}_{22} = \mathbf{I}_{q-1} + \frac{1}{n-q}\mathbf{1}_{q-1}\mathbf{1}_{q-1}^t$, and $\mathbf{A}_{22}^{-1} = \mathbf{I}_{q-1} - \frac{1}{n-1}\mathbf{1}_{q-1}\mathbf{1}_{q-1}^t$, after some algebra we find that $a_{11.2} = \frac{n}{n-1}$ and $\mathbf{a}_{12}\mathbf{A}_{22}^{-1} = \frac{1}{n-1}\mathbf{1}_{q-1}^t$.

The normalizing constant for the kernel of the density of $\tilde{\mathbf{y}}_1$, shown in (2.30) follows directly from (2.1). The scale factor all $a_{11.2}^{\frac{1}{2}} = \left(\frac{n}{n-1}\right)^{\frac{1}{2}}$ is exactly that used to define $\mathbf{y}$ in (2.1) and the marginal density of $\tilde{\mathbf{Z}}$ is that of $\tilde{\mathbf{Y}}$ as displayed in (2.24) with $q \longrightarrow q-1$ and $\mathbf{A} \longrightarrow \mathbf{A}_{22}$.

The normalizing constant for $|\mathbf{I} - \mathbf{ZA}_{22}\mathbf{Z}^t|^{\frac{1}{2}(N-q)-\gamma}$ can be found by integration as follows: treat $q < p+1$ and $q \geq p+1$ separately. Begin with $q \geq p+1$ and let $\mathbf{W} = \mathbf{ZA}_{22}^{\frac{1}{2}} = \mathbf{R}^{\frac{1}{2}}\mathbf{Q}, (p \times p)\mathbf{R} > \mathbf{0}$ and $p \times (q-1)\mathbf{Q} \in V_{p,q-1}$. Then $J(\mathbf{Z} \longrightarrow \mathbf{W}) = |\mathbf{A}_{22}|^{-\frac{1}{2}p}, J(\mathbf{W} \longrightarrow (\mathbf{R}, \mathbf{Q})) = 2^{-p}|\mathbf{R}|^{\frac{1}{2}(q-p-2)}$ and $J(\mathbf{Z} \longrightarrow (\mathbf{R}, \mathbf{Q})) = |\mathbf{A}_{22}|^{\frac{1}{2}p} \cdot 2^{-p}|\mathbf{R}|^{\frac{1}{2}(q-p-2)}$ so that

$$\begin{aligned} &\int_{\mathbf{Z} \in \Re^{p(q-1)}} |\mathbf{I} - \mathbf{ZA}_{22}\mathbf{Z}^t|^{\frac{1}{2}(N-1)-\gamma} d\mathbf{Z} \\ &= 2^{-p}|\mathbf{A}_{22}|^{-\frac{1}{2}p} \int_{\mathbf{R}>\mathbf{0}} |\mathbf{I} - \mathbf{R}|^{\frac{1}{2}(N-q)-\gamma}|\mathbf{R}|^{\frac{1}{2}(q-1)-\gamma} d\mathbf{R} \times vol(V_{p,q-1}). \end{aligned} \tag{2.31}$$

As

$$vol(V_{p,q-1}) = 2^{\frac{1}{2}pq}\pi^{\frac{1}{2}p(q-1)}c_p(q-1)$$

and

$$\int\limits_{\mathbf{R}>\mathbf{0}} |\mathbf{I}-\mathbf{R}|^{\frac{1}{2}(N-q)-\gamma}|\mathbf{R}|^{\frac{1}{2}(q-1)-\gamma}d\mathbf{R} = \frac{c_p(N-1)}{c_p(N-q)c_p(q-1)}, \tag{2.32}$$

$$\int\limits_{\mathbf{Z}\in\Re^{p(q-1)}} |\mathbf{I}-\mathbf{Z}\mathbf{A}_{22}\mathbf{Z}^t|^{\frac{1}{2}(N-q)-\gamma}d\mathbf{Z} = \frac{(2\pi)^{\frac{1}{2}p(q-1)}c_p(N-1)}{c_p(N-q)}|\mathbf{A}_{22}|^{\frac{1}{2}p}, \tag{2.33}$$

the reciprocal of the normalizing constant shown in (2.27).

Linear combinations of columns of $\tilde{\mathbf{Y}}$ possess a density of the form (2.1). Let

$$(q\times q)\mathbf{C} = \begin{bmatrix} c_1 & \mathbf{0} \\ \mathbf{c}_2 & \mathbf{I}_{q-1} \end{bmatrix} \tag{2.34}$$

and transform from $(p\times q)\mathbf{Y}$ to $[\mathbf{m}\mathbf{Y}_{12}] \equiv \mathbf{Y}\mathbf{C}, \mathbf{m} \equiv \sum_{k=1}^{q} c_k\mathbf{y}_k \equiv \mathbf{Y}\mathbf{c}$. As the $J(\mathbf{Y} \longrightarrow (\mathbf{m},\mathbf{Y}_{12})) = |\mathbf{C}|^{-\frac{1}{2}p}$, $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{Y}}_{12}$ have joint density proportional to

$$|\mathbf{I}-[\mathbf{m}\ \mathbf{Y}_{12}]\mathbf{C}^{-1}\mathbf{A}(\mathbf{C}^{-1})^t[\mathbf{m}\ \mathbf{Y}_{12}]^t|^{\frac{1}{2}(N-q)-\gamma}. \tag{2.35}$$

Define

$$\mathbf{G} = \mathbf{C}^{-1}\mathbf{A}(\mathbf{C}^{-1})^t = \begin{bmatrix} g_{11} & \mathbf{g}_{12} \\ \mathbf{g}_{21} & \mathbf{G}_{22} \end{bmatrix} \tag{2.36}$$

with g_{11} scalar and, as in the development of Proposition 3, factor (2.35) into two independent kernels, one of which is

$$|\mathbf{I}-\mathbf{m}g_{11.2}\mathbf{m}^t|^{\frac{1}{2}(N-1)-\gamma} = (1-\mathbf{m}^t g_{11.2}\mathbf{m})^{\frac{1}{2}(N-1)-\gamma}. \tag{2.37}$$

Here $g_{11.2} = g_{11} - \mathbf{g}_{12}\mathbf{G}_{22}^{-1}\mathbf{g}_{21}$.

The normalizing constant of (2.37) is $\left(\frac{g_{11.2}}{\pi}\right)^{\frac{1}{2}p}\frac{\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}(N-p))}$ so we have

Proposition 4 *A linear combination* $\tilde{\mathbf{m}} = \tilde{\mathbf{Y}}\mathbf{c}$ *of columns of* $\tilde{\mathbf{Y}}$ *possesses density*

$$\left(\frac{g_{11.2}}{\pi}\right)^{\frac{1}{2}p} \frac{\Gamma\left(\frac{1}{2}N\right)}{\Gamma\left(\frac{1}{2}(N-p)\right)} \left(1 - \mathbf{m}^t g_{11.2}^{-1}\mathbf{m}\right)^{\frac{1}{2}(N-1)-\gamma} \tag{2.38}$$

with G *as defined in (2.36).*[1]

We established the independence of $(p \times 1)\tilde{\mathbf{y}}_1$ and $(p \times (q-1))\tilde{\mathbf{Z}}$ as defined in (2.26), and, in addition, that the densities of $\tilde{\mathbf{y}}_1$ and of $\tilde{\mathbf{Z}}$ are proportional to $|1 - \mathbf{y}_1\mathbf{a}_{11.2}\mathbf{y}_1^t|^{\frac{1}{2}(N-1)-\gamma}$ and $|\mathbf{I} - \mathbf{Z}\mathbf{A}_{22}\mathbf{Z}^t|^{\frac{1}{2}(N-q)-\gamma}$ respectively. By $q-2$ additional successive factorings of $\mathbf{I} - \mathbf{Z}\mathbf{A}_{22}\mathbf{Z}^t$ as $\mathbf{I} - \mathbf{Y}\mathbf{A}\mathbf{Y}^t$ is factored in (2.28b) we generate q densities of q mutually independent $(p \times 1)$ rvs which we shall call $\tilde{\mathbf{y}}_1, \tilde{\mathbf{w}}_2, \ldots, \tilde{\mathbf{w}}_q$. Here

$$\tilde{\mathbf{w}}_2 = \tilde{\mathbf{z}}_1 = \left[\mathbf{I}_q - \left(\frac{n}{n-1}\right)\tilde{\mathbf{y}}_1\tilde{\mathbf{y}}_1^t\right]^{-\frac{1}{2}} \left[\tilde{\mathbf{y}}_2 + \frac{1}{n-1}\tilde{\mathbf{y}}_1\right] \tag{2.39}$$

and so on. The density of $\tilde{\mathbf{w}}_k, 1 < k \leq q-1$, the kth successively conditioned rv, is

$$\left(\frac{n-k}{n-k+1}\right)^{\frac{1}{2}p} \frac{\Gamma(\frac{1}{2}(N-k))}{\pi^{\frac{1}{2}p}\Gamma(\frac{1}{2}(N-p))} \left(1 - \left(\frac{n-k}{n-k+1}\right)\mathbf{w}_k^t\mathbf{w}_k\right)^{\frac{1}{2}(N-k)-\gamma} \tag{2.40}$$

and that of $\left(\frac{n-k}{n-k+1}\right)\tilde{\mathbf{w}}_k^t\tilde{\mathbf{w}}_k$ is Beta with parameter $(\frac{1}{2}(N-k), \frac{1}{2}p)$ and expectation $\mathrm{E}(\tilde{s}_k) = \frac{p}{(N-k-p)}$. This is summarized as

Proposition 5 $\tilde{\mathbf{y}}_1$ *and successively conditioned rvs* $\tilde{\mathbf{w}}_2, \ldots, \tilde{\mathbf{w}}_q$ *are mutually independent and* $\tilde{\mathbf{w}}_k, 2 \leq k \leq q$ *has density (1.38). In addition, the joint distribution of squares of the* k *elements of* $\tilde{\mathbf{w}}_k$ *is* $D_p(\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}(n-p-k))$.

Reverse the abuse of notation adopted early on and redefine $\mathbf{y}_1$ as shown in (2.39) to be $\left(\frac{n}{n-1}\right)\mathbf{y}_1$ so that we can write the kernel of the density of $\tilde{\mathbf{y}}_1$ as $(1 - \mathbf{y}_1^t\mathbf{y})^{\frac{1}{2}(N-1)-\gamma}$. This kernel can be successively factored in a fashion that yields p mutually independent scalar

[1][N.B. $N = n-1$ so $\tilde{\mathbf{m}}\sqrt{g_{11.2}}$ has density (2.1).

rvs:

$$\mathbf{y} \longrightarrow \begin{pmatrix} y_1 \\ \\ \mathbf{v}_{2.1} \end{pmatrix} \text{ with } \mathbf{v}_{2.1} = (1 - y_1^2)^{-\frac{1}{2}} \begin{pmatrix} y_2 \\ \vdots \\ y_p \end{pmatrix} \equiv \begin{pmatrix} v_2 \\ \vdots \\ v_p \end{pmatrix}$$

$$\mathbf{v}_{2.1} \longrightarrow \begin{pmatrix} v_2 \\ \\ \mathbf{v}_{3.2} \end{pmatrix} \text{ with } \mathbf{v}_{3.2} = (1 - v_2^2)^{-\frac{1}{2}} \begin{pmatrix} v_3 \\ \vdots \\ v_p \end{pmatrix} \equiv \begin{pmatrix} \xi_2 \\ \vdots \\ \xi_p \end{pmatrix}$$

and so on. Repeating this p times, yields p mutually independent rvs $\tilde{y}_1, \tilde{v}_2, \tilde{\xi}_3, \ldots, \tilde{\eta}_p$, each of whose square is Beta distributed: $\tilde{y}_1$ is Beta with a parameter $(\frac{1}{2}N, 1), \tilde{v}_2$ is Beta with parameter $(\frac{1}{2}(N-1), 1)$ and so on. Finally, if we relabel, we have

Proposition 6 *By successive (linear) conditioning, the* $(p \times q)$ *rv* $\tilde{\mathbf{Y}}$ *can be mapped into* $(p \times q)\tilde{\mathbf{V}}$ *such that*

1. $\{\tilde{v}_{ij}, i = 1, \ldots, p; j = 1, \ldots, q\}$ *are mutually independent.*
2. *Each* $\tilde{v}_{ij}^2$ *is Beta distributed.*

References

[1] Quiroz, A. and Dudley, R. M., Some new tests for multivariate normality, *Probab. Th. Rel. Fields*, **87**, 521–546 (1991).

[2] Small, N. J. H., Plotting squared radii, *Biometrika*, **65**, 657–658 (1978).