THE CONSTRUCTION AND PROPERTIES OF ELLIPSOIDAL PROBABILITY DENSITY FUNCTIONS

GRAHAM L. GILLER

ABSTRACT. A recipe for the construction of a multivariate probability density function from a normalized symmetric univariate function using a distance metric methodology is developed. A method to perform multivariate integration in a polar coordinate system in an arbitary number of dimensions is described. This method is then used to compute a constant that normalizes the multivariate p.d.f. constructed using the recipe specified here. The use of the Kolmogorov test, as applied to the univariate distribution of the square of the metric distance, for distributional identification is described. A summary of the general properties of the constructed p.d.f. is given including the computation of: the principal moments; Mardia's multivariate kurtosis measure; and, the characteristic and moment generatic functions. Some elements of maximum likelihood estimation are explored.

1. THE DISTANCE METRIC FORMALISM FOR MULTIVARIATE PROBABILITY DENSITY FUNCTIONS

1.1. **The Uninormal and Multinormal Distributions.** The mathematical form of the Normal probability density function has a number of remarkably useful properties. In particular, it has a very natural generalization from a univariate form to a multivariate form. If $f(x|\mu,\sigma)$ represents the univariate p.d.f.

$$f(x|\mu,\sigma)\,df = \frac{1}{\sqrt{2\pi}\sigma}\exp -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 dx, \tag{1}$$

we find that the joint probability for n variables has the form[17]

$$f(\boldsymbol{x}|\boldsymbol{\mu},\Sigma)\,df = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp -\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T\Sigma^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\,d^n\boldsymbol{x}. \tag{2}$$

It is notable that the functional form of Equation 2 is the same as that of Equation 1, with the univariate "standardized distance" measure $\frac{x-\mu}{\sigma}$ replaced by multivariate metric $\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T\Sigma^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)$ and the normalization scale σ replaced by the square root of the determinant of the covariance matrix $|\Sigma|^{\frac{1}{2}}$. This form of Equation 2 arises from directly constructing the multivariate p.d.f.from the combination of univariate functions, according to the rules of joint probability[14].

1.2. **A Method for Generalization from Other Univariate Distributions.** Unfortunately, the majority of univariate densities do not posess this appealing property of the Normal distribution[1]. In fact, the choice of "natural" multivariate distributions available to the analyst is quite constrained. In order to generate a workable family of multivariate distributions, we therefore choose to follow the suggestion of Equation 1 and Equation 2 by writing the

---

Date: October 22, 2007. *Giller Investments Research Note*: 20030320/1.

[1]That the joint distribution of multiple variables has the same functional form as the individual conditional distributions from which it is built.

argument of a univariate p.d.f.in terms of a standarized distance measure and then replacing that measure itself with a multivariate metric.

At this point we will limit our analysis to symmetrical distributions. i.e. Functions for which we can make the transformation:

$$f(x^2)\,dx \to f\{(\boldsymbol{x}-\boldsymbol{\mu})^T\,\Sigma^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\}\,d^n\boldsymbol{x}. \tag{3}$$

Furthermore, for clarity of exposition in the following section, we shall put the multivariate location term, $\boldsymbol{\mu}$, to $\mathbf{0}$. (This term is trivial to reintroduce by translation.)

Therefore, our procedure to generate a multivariate p.d.f.suitable for study is to take a parent univariate p.d.f., $f(x^2)$, and transform it to the multivariate p.d.f.$f(\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x})$. An additional normalization term, $\mathcal{A}$, is introduced to restore the normalization of the p.d.f.to unity[2]. i.e.

$$f(x^2)\,dx \to \mathcal{A}f(\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x})\,d^n\boldsymbol{x}. \tag{4}$$

To establish the value of $\mathcal{A}$, we must perform the integral

$$\int\cdots\int_{\boldsymbol{\Omega}} f(\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x})\,d^n\boldsymbol{x} \tag{5}$$

over the region $\boldsymbol{\Omega}$ for which the p.d.f. is defined. From experience with the multinormal distribution, we know that for an infinite domain this integral can be done in polar coordinates[18]. The correlated variables $\boldsymbol{x}$ can easily be transformed into uncorrelated variables $\boldsymbol{f}$ (sometimes called *factors* or *principal components*) via an orthogonal transformation $\boldsymbol{f} = R\boldsymbol{x}$. (The required matrix, R, is a rotation into a coordinate system in which the matrix $R\Sigma R^T$ is diagonal — i.e. it is the matrix of the eigenvectors of Σ.) The inverse of such a matrix is trivially equal to the matrix with each diagonal term replaced by its reciprocal. The distance metric is then reduced to a scaled sum-of-squares.

$$\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x} \to \boldsymbol{f}^T\boldsymbol{\lambda}^{-1}\boldsymbol{f} = \sum_{i=1}^{n}\frac{f_i^2}{\sigma_i^2}. \tag{6}$$

The matrix $\boldsymbol{\lambda}$ is the diagonal matrix $(\sigma_1^2, \sigma_2^2 \ldots \sigma_n^2)$ of the eigenvalues of Σ. For a multinormal model, the term σ_i^2 represents the variance of factor i.

Transforming into an n-dimensional elliptical polar coordinate system (i.e. one with one length, r, and $n-1$ hyper-angles, θ_i) allows the n-dimensional hyper-volume integral $\int\cdots\int_\Omega d^n\boldsymbol{x}$ to be replaced with a single distance integral and $n-1$ hyper-angle integrals. As the scaled sum-of-squares has no angular dependence in this coordinate system, the hyper-angular integrals are trivial and we have reduced the multiple integral of Equation 5 to a single integral.

The following section will discuss the general form of n-dimensional spherical polar coordinates[3], to be used in this discussion.

2. A Generalization of Spherical Polar Coordinates to an Arbitary Number of Dimensions

At a graduate level, the analyst should be very familiar with operations involving integration in one, two and three dimensions in a number of coordinate systems in addition to simple cartesian coordinates. However, less of us have any intuition or experience with generalizing those operations into four, five, or more dimensions in any coordinate system

[2]The normalization of the original p.d.f.is disrupted by replacing single integrals with multiple integrals in n-dimensional space.

[3]The generalization to elliptical polar is trivial.

other than the cartesian system. In this section we illustrate how to extend "polar" systems into higher dimensions, and then present a general method for the construction of a "polar" system in an arbitary number of dimensions.

2.1. **The Lower Dimensions.** Let us consider the coordinate of a point on a hyper-sphere in each of 1, 2, 3, 4 and 5 dimensions. A sphere is defined to be the locus of all points at a given distance from the origin. We will define "distance," r, to be the square root of the sum of the squares of the coordinates (i.e. $r^2 = \sum_{i=1}^{n} x_i^2$). Therefore, a "1d" sphere is the pair of points $\{+r, -r\}$; a "2d" sphere is the circle with radius r; a "3d" sphere is a normal sphere of radius r; and so forth.

In one dimension, there is no difference between the cartesian coordinate (x) and the polar coordinate (r). In two dimensions, we take the "x" axis to be primary, and define the angle θ to be the angle between that axis and the line from the origin to a point on the circle of radius r. Thus $(x, y) \to r(\cos\theta, \sin\theta)$ for $\theta \in [0, 2\pi]$. Similarly, in three dimensions, we take the "z" axis to be principal and project a point on the sphere onto the axis and onto the $x - y$ plane. In the plane, we reapply the method for two dimensions. This leads to the familiar expression $(x, y, z) \to r(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ with $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$.

In each higher dimension we follow the same procedure: first project a point on the sphere onto a primary axis (this gives a "$r\cos\theta$" coordinate with $\theta \in [0, \pi]$); then project the point into the "plane" spanned by the other coordinates (leading to a "$r\sin\theta$" factor in all the other coordinates); then, we apply the result for the $n - 1$*th.* dimension and iterate until the plane we are projecting into is actually a two dimensional plane. This is the sole angle which does not have a redundancy[4] and so this final angle is defined on $[0, 2\pi]$.

The results for the first five dimensions are summarized in Table 1, below. In the table,

n	Cartesian	Polar
1	(x)	(r)
2	(y, x)	$r(\sin\alpha, \cos\alpha)$
3	(y, x, z)	$r(\sin\alpha\sin\beta, \sin\alpha\cos\beta, \cos\alpha)$
4	(y, x, z, u)	$r(\sin\alpha\sin\beta\sin\gamma, \sin\alpha\sin\beta\cos\gamma, \sin\alpha\cos\beta, \cos\alpha)$
5	(y, x, z, u, v)	$r(\sin\alpha\sin\beta\sin\gamma\sin\delta, \sin\alpha\sin\beta\sin\gamma\cos\delta,$ $\sin\alpha\sin\beta\cos\gamma, \sin\alpha\cos\beta, \cos\alpha)$

TABLE 1. Cartesian and Polar Coordinates for 1 to 5 Dimensions

angles are represented by Greek letters and distances by Roman. The ordering of x and y are reversed to help emphasize the pattern produced by iterative projection into lower dimensions[5]. It can be readily verified that the sum of the squares of the coordinates is r^2 in each case.

2.2. **A General Scheme.** From Table 1 we can infer a general scheme for n dimensions. Let the cartesian coordinates be $\{x_i\}_{i=1}^{n}$. Let the polar distance coordinate be r and the

[4]Angles in $[\pi, 2\pi]$ can be mapped into angles in $[0, \pi]$ by a rotation in the sub-plane about the axis we are projecting onto.

[5]From this table, it would appear that our conventional assignment of x and y are the wrong way around!

$n-1$ angles be $\{\theta_i\}_{i=1}^{n-1}$. Choosing the *primary* axis arbitarily to be x_n, we write[6]

$$x_1 = r \prod_{i=1}^{n-1} \sin\theta_i. \tag{7}$$

For the remaining coordinates, x_i where $i > 1$, we have

$$x_i = r\cos\theta_{n-i+1} \prod_{j=1}^{n-i} \sin\theta_j. \tag{8}$$

The domains of the polar coordinates are:

$$r \in [0,\infty];\ \theta_i \in [0,\pi]\,\forall\, 1 < i < n-1;\ \theta_{n-1} \in [0, 2\pi]. \tag{9}$$

2.3. **Transformation of the Volume Integral.** The transformation of the hyper-volume integral[7], given the coordinate transformation specified by Equation 7 and Equation 8, is given by Equation 10

$$\int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} d^n\boldsymbol{x} = \int_{r=0}^{\infty} \int_{\theta_1=0}^{\pi} \int_{\theta_2=0}^{\pi} \cdots \int_{\theta_{n-1}=0}^{2\pi} |J_n| dr \prod_{i=1}^{n-1} d\theta_i, \tag{10}$$

where J_n represents the *Jacobian* of the transformation. J_n is defined to be the determinant given by Equation 11.

$$J_n = \frac{\partial(x_1, x_2 \ldots x_n)}{\partial(r, \theta_1, \theta_2, \ldots \theta_{n-1})} = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta_1} & \cdots & \frac{\partial x_1}{\partial \theta_{n-1}} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta_1} & \cdots & \frac{\partial x_2}{\partial \theta_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial r} & \frac{\partial x_n}{\partial \theta_1} & \cdots & \frac{\partial x_n}{\partial \theta_{n-1}} \end{vmatrix} \tag{11}$$

From our definition of the coordinate transformations, we find:

$$\frac{\partial x_1}{\partial r} = \prod_{i=1}^{n-1} \sin\theta_i = \frac{x_1}{r}; \tag{12}$$

and,

$$\frac{\partial x_1}{\partial \theta_i} = r\cos\theta_i \prod_{\substack{j=1 \\ j\neq i}}^{n-1} \sin\theta_j = x_1 \cot\theta_i; \tag{13}$$

for the first row of the determinant, and:

$$\frac{\partial x_i}{\partial r} = \cos\theta_{n-i+1} \prod_{j=1}^{n-i} \sin\theta_j = \frac{x_i}{r}; \tag{14}$$

$$\left.\frac{\partial x_i}{\partial \theta_j}\right|_{\substack{i>1 \\ j<n-i+1}} = r\cos\theta_{n-i+1}\cos\theta_j \prod_{\substack{k=1 \\ k\neq j}}^{n-i} \sin\theta_k = x_i \cot\theta_j; \tag{15}$$

$$\left.\frac{\partial x_i}{\partial \theta_{n-i+1}}\right|_{i>1} = -r\sin\theta_{n-i+1} \prod_{j=1}^{n-i} \sin\theta_j = -x_i \tan\theta_{n-i+1}; \tag{16}$$

---

[6]When the product term $\prod_{i=j}^{k}$ appears for $k < j$ we take it's value to be unity.

[7]The argument followed here is well known in the literature. See reference [18], for example.

and,

$$\left.\frac{\partial x_i}{\partial \theta_j}\right|_{\substack{i>1\\ j>n-i+1}} = 0; \tag{17}$$

for the remaining rows. The Jacobian therefore has the value

$$J_n = \begin{vmatrix} \frac{x_1}{r} & x_1 \cot\theta_1 & x_1 \cot\theta_2 & \cdots & x_1 \cot\theta_{n-2} & x_1 \cot\theta_{n-1} \\ \frac{x_2}{r} & x_2 \cot\theta_1 & x_2 \cot\theta_2 & \cdots & x_2 \cot\theta_{n-2} & -x_2 \tan\theta_{n-1} \\ \frac{x_3}{r} & x_3 \cot\theta_1 & x_3 \cot\theta_2 & \cdots & -x_3 \tan\theta_{n-2} & 0 \\ \frac{x_4}{r} & x_4 \cot\theta_1 & x_4 \cot\theta_2 & \cdots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{x_n}{r} & -x_n \tan\theta_1 & 0 & \cdots & 0 & 0 \end{vmatrix}. \tag{18}$$

We may take out common factors in each row or column, giving

$$J_n = \frac{\prod_{i=1}^{n} x_i}{r} \begin{vmatrix} 1 & \cot\theta_1 & \cot\theta_2 & \cdots & \cot\theta_{n-2} & \cot\theta_{n-1} \\ 1 & \cot\theta_1 & \cot\theta_2 & \cdots & \cot\theta_{n-2} & -\tan\theta_{n-1} \\ 1 & \cot\theta_1 & \cot\theta_2 & \cdots & -\tan\theta_{n-2} & 0 \\ 1 & \cot\theta_1 & \cot\theta_2 & \cdots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & -\tan\theta_1 & 0 & \cdots & 0 & 0 \end{vmatrix}. \tag{19}$$

Substuting Equation 7 and Equation 8 into the scale factor of Equation 19, we see that this term is equal to

$$r^{n-1} \prod_{i=1}^{n-1} \sin^{n-i}\theta_i \prod_{i=1}^{n-1} \cos\theta_i. \tag{20}$$

We may now reduce the determinant by using the property of determinants that adding any multiple of one row to another row leaves the value of the determinant unchanged. In Equation 19, we successively subtract row k from row $k-1$ to reduce the determinant as follows:

$$\text{row } 1 - \text{row } 2 \quad \Rightarrow \quad \text{row } 1 \to (0, 0, 0 \ldots 0, \cot\theta_{n-1} + \tan\theta_{n-1}) \tag{21}$$

$$\text{row } 2 - \text{row } 3 \quad \Rightarrow \quad \text{row } 2 \to (0, 0, 0 \ldots \cot\theta_{n-2} + \tan\theta_{n-2}, 0) \tag{22}$$

$$\vdots$$

$$\text{row}(n-2) - \text{row}(n-1) \quad \Rightarrow \quad \text{row}(n-2) \to (0, \cot\theta_1 + \tan\theta_1, 0 \ldots, 0) \tag{23}$$

Applying the results of Equation 21 to the determinant of Equation 19 gives the determinant of Equation 24.

$$\begin{vmatrix} 0 & 0 & 0 & \cdots & 0 & \cot\theta_{n-1}+\tan\theta_{n-1} \\ 0 & 0 & 0 & \cdots & \cot\theta_{n-2}+\tan\theta_{n-2} & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & \cot\theta_1+\tan\theta_1 & 0 & \cdots & \vdots & \vdots \\ 1 & -\tan\theta_1 & 0 & \cdots & 0 & 0 \end{vmatrix} \tag{24}$$

Equation 24 evaluates to

$$\prod_{i=1}^{n-1} \frac{1}{\cos\theta_i \sin\theta_i}. \tag{25}$$

Substituting Equation 20 and Equation 25 into Equation 18, we have the final expression for the Jacobian of this transformation.

$$J_n = r^{n-1} \prod_{i=1}^{n-2} \sin^{n-i-1}\theta_i. \tag{26}$$

We can easily verify that this expression gives the correct result for the well known lower dimensions: $dx \to dr$; $dx\,dy \to r\,dr\,d\theta$; and, $dx\,dy\,dz \to r^2 \sin\theta\,dr\,d\theta\,d\phi$. Note that in Equation 26, the sine of angles θ_i is only ever taken for $1 < i < n-2$ which is the set of angles for which $\theta_i \in [0,\pi]$. Therefore, $\sin\theta_i$ is never negative, and the absolute value $|J_n|$ from Equation 10 can be replaced with just J_n itself.

If we are integrating a function, $f(r)$, which depends solely on the radial distance, r, in two or more dimensions ($n > 1$) we see that the angular integrals in Equation 11 are trivial.

$$\int\cdots\int_{\mathbf{\Omega}} f(r)\,d^n\boldsymbol{x} = \int_0^\infty f(r)\,r^{n-1}\,dr \prod_{i=1}^{n-2} \int_{\theta_i=0}^{\pi} \sin^{n-i-1}\theta_i\,d\theta_i \int_{\theta_{n-1}=0}^{2\pi} d\theta_{n-1} \tag{27}$$

The Beta function has the integral representation of Equation 28, below[4].

$$\int_0^{\pi/2} \sin^{\mu-1} x \cos^{\nu-1} x\,dx = \frac{1}{2} B\Big(\frac{\mu}{2}, \frac{\nu}{2}\Big) \text{ where } \Re\mu > 0, \Re\nu > 0. \tag{28}$$

Therefore, the integrals containing sine terms in Equation 27 may be written

$$\int_{\theta_i=0}^{\pi} \sin^{n-i-1}\theta_i\,d\theta_i = B\Big(\frac{n-i}{2}, \frac{1}{2}\Big) = \pi^{\frac{1}{2}} \frac{\Gamma\big(\frac{n-i}{2}\big)}{\Gamma\big(\frac{n-i+1}{2}\big)}. \tag{29}$$

Substituting these results into Equation 27, we have

$$\int\cdots\int_{\mathbf{\Omega}} f(r^2)\,d^n\boldsymbol{x} = 2\pi^{\frac{n}{2}} \prod_{i=1}^{n-2} \frac{\Gamma\big(\frac{n-i}{2}\big)}{\Gamma\big(\frac{n-i+1}{2}\big)} \int_{r=0}^{\infty} f(r^2)\,r^{n-1}\,dr. \tag{30}$$

3. The Normalization of a Multivariate Distribution based on a Distance Metric Transformation of a Univariate Distribution

In this section, the results of Sections 1 and 2 are combined to yield a full expression for the normalization constant introduced when we construct a multivariate distribution using the method presented here.

3.1. **Solution for a General Distribution.** We consider a symmetric parent univariate distribution, normalized over an infinite domain.

$$\int_{-\infty}^{\infty} f(x^2)\, dx = 1. \tag{31}$$

Following Equation 3, we construct a multivariate distribution using the distance metric and must evaluate the new normalization constant, $\mathcal{A}$.

$$\frac{1}{\mathcal{A}} = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} f\{(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\}\, d^n\boldsymbol{x}. \tag{32}$$

The first step in performing this integral is the simple translation $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\mu}$. Since the domain of integration is infinite, and writing $\mathcal{N}$ for $1/\mathcal{A}$, we have

$$\mathcal{N} = \int_{y_1=-\infty}^{\infty} \cdots \int_{y_n=-\infty}^{\infty} f(\boldsymbol{y}^T \Sigma^{-1} \boldsymbol{y})\, d^n\boldsymbol{y}. \tag{33}$$

We know from the analysis sketch of Section 1 that this expression should be further reduced by diagonalizing the coordinate system through an orthogonal rotation. This leaves an elliptical polar system which may be transformed into a spherical polar system via simple scaling operations. Such operations lead to a normalization constant that is proportional to the square root of the product of the eigenvalues of the matrix Σ.

We require that Σ be a symmetric positive definite matrix. (This is consistent with the parallel we are drawing with the multinormal distribution, for which Σ is equal to the covariance matrix of $\boldsymbol{x}$ and is s.p.d. by definition.) The matrix of eigenvalues of such a matrix may be factored as $\boldsymbol{\lambda} = D^2 = DD^T = D^T D$, where D is a diagonal matrix with positive definite elements σ_i. Furthermore, we know that the inverse $\boldsymbol{\lambda}^{-1} = (D^2)^{-1} = (D^{-1})^2$ is trivial.

If R is the matrix of eigenvectors of Σ, then we may define the matrix D by

$$R^T \Sigma R = D^2. \tag{34}$$

From Equation 34 and the definitions of D and R, it follows that

$$R^T \Sigma^{-1} R = (D^{-1})^2. \tag{35}$$

Therefore, rather than the transformation $\boldsymbol{f} = R\boldsymbol{y}$ (suggested in Section 1), we shall make the transformation

$$\boldsymbol{g} = D^{-1} R \boldsymbol{y}. \tag{36}$$

Equation 36 implies that $\boldsymbol{y} = R^T D\boldsymbol{g}$ and $\boldsymbol{y}^T = \boldsymbol{g}^T DR$. The distance metric may now be written

$$\begin{aligned} \boldsymbol{y}^T \Sigma^{-1} \boldsymbol{y} &= \boldsymbol{g}^T DR\Sigma^{-1} R^T D\boldsymbol{g} \\ &= \boldsymbol{g}^T D(R^T \Sigma R)^{-1} D\boldsymbol{g} \\ &= \boldsymbol{g}^T D(D^2)^{-1} D\boldsymbol{g} \\ &= g^2. \end{aligned} \tag{37}$$

Using the transformation of Equation 36, Equation 33 becomes

$$\mathcal{N} = \int_{g_1=-\infty}^{\infty} \cdots \int_{g_n=-\infty}^{\infty} f(g^2)\, |K_n|\, d^n\boldsymbol{g}. \tag{38}$$

where K_n is the Jacobian

$$K_n = \frac{\partial(y_1 \dots y_n)}{\partial(g_1 \dots g_n)} = \begin{vmatrix} \frac{\partial y_1}{\partial g_1} & \cdots & \frac{\partial y_1}{\partial g_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial g_1} & \cdots & \frac{\partial y_n}{\partial g_n} \end{vmatrix}. \tag{39}$$

From Equation 36 we have $\frac{\partial y_i}{\partial g_j} = (R^T D)_{ij}$, therefore $K_n = |R^T D| = |D|$. Since $|D^2| = |\Sigma|$ (from Equation 35), we have $|D| = |\Sigma|^{\frac{1}{2}}$. Furthermore, since Σ is s.p.d. by construction, the determinant of Σ is positive definite and we may replace $|K_n|$ by K_n in Equation 38. We may now make the spherical polar transformation of Section 2, to give

$$\mathcal{N} = 2\pi^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}} \prod_{i=1}^{n-2} \frac{\Gamma\left(\frac{n-i}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)} \int_{g=0}^{\infty} f(g^2)\, g^{n-1}\, dg. \tag{40}$$

3.2. **Evaluation of the Product Factor.** We have shown above that the normalization factor $\mathcal{A}$ is proportional to $|\Sigma|^{-\frac{1}{2}}$. The full expression, Equation 40, also contains a term dependent on the precise p.d.f. in use and a term that is the product of a series of ratios of gamma functions. In this section we will use the known result of Equation 2 to derive a more compact expression for this factor.

Let us represent this gamma function factor as P_n. i.e.

$$P_n = \prod_{i=1}^{n-2} \frac{\Gamma\left(\frac{n-i}{2}\right)}{\Gamma\left(\frac{n-i+1}{2}\right)} \tag{41}$$

Equation 40 now becomes

$$\mathcal{N} = 2\pi^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}} P_n \int_{g=0}^{\infty} f(g^2)\, g^{n-1}\, dg. \tag{42}$$

For the standardized distribution $N(0,1)$, we have

$$f(x^2) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}. \tag{43}$$

Substituting this into Equation 42 gives

$$\mathcal{N} = \sqrt{2\pi^{n-1}|\Sigma|} P_n \int_0^{\infty} e^{-\frac{1}{2}g^2} g^{n-1}\, dg. \tag{44}$$

The integral in this expression is recognized as a representation of the gamma function[5]. Substituting this result into Equation 44 gives an expression for $\mathcal{N}$ in terms of $|\Sigma|$ and P_n.

$$\mathcal{N} = \sqrt{(2\pi)^{n-1} |\Sigma|} P_n \Gamma\left(\frac{n}{2}\right). \tag{45}$$

Comparing Equation 2 and Equation 45, we must have

$$P_n = \frac{1}{\Gamma\left(\frac{n}{2}\right)}. \tag{46}$$

Therefore

$$\mathcal{N} = \frac{2\pi^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} f(g^2)\, g^{n-1}\, dg. \tag{47}$$

The normalized univariate p.d.f. $f(x^2)$ is renormalized by dividing by this factor when the distance metric transformation is used to generate a multivariate p.d.f.

4. A TEST STATISTIC FOR THE IDENTIFICATION OF MULTIVARIATE DISTRIBUTIONS

4.1. **Definition.** In this section we propose a statistic that may be used to aid testing the hypothesis that a given dataset is drawn from a multivariate distribution such as one derived using the methodology outlined in this document.

For univariate distributions, powerful tests for distribution identification can be constructed using either the Kolmogorov statistic or a statistics from the Smirnov-Cramér-von Mises Group[1]. Both of these methods rely on the comparison of the empirical distribution function (or "order statistics") with the actual distribution function. This comparison is well defined in one dimension, but it is problematic to generalize to arbitary numbers of dimensions. The problems arise from defining the direction in n-space through which one travels to define the order statistic. Another way of stating this is that no simple analog of the transformation to convert an arbitary univariate distribution into a uniform distribution may be defined for a multivariate distribution. Some authors[3] have suggested examining every possible ordering of the transformation to uniformity; however, for n dimensions there are $n!$ such orderings and the procedure clearly can only be performed for n of order 2 to 4. Problems involving 20 or more dimensions are out of the question!

Therefore, we propose using the Kolmogorov test on the distance metric itself. This test has the advantages that it is easy to define, understand, and execute. It has the drawback that it is not the most powerful test we could define (for example, it is possible that the conditional distribution of variable x_i has excess kurtosis and that the conditional distribution of variable x_j has insufficient kurtosis in such a way that when they are combined in the distance metric this effect is exactly cancelled and the test proposed has zero power to identify such a defect).

Define g^2 as the value of the distance metric, and the density function and distribution function of g^2 as $f'(g^2)$ and $F'(g^2)$ respectively. The probability that $g^2 < G^2$, where G^2 is a particular sample value of the distance metric, is equal to the probability that $\boldsymbol{x}$ lies within the region Ω_G defined to be the hypervolume enclosed by the hyperellipsoidal shell parameterized by particular value of the distance metric. i.e.

$$F'(G^2) = \Pr(g^2 < G^2) = \Pr(\boldsymbol{x} \in \Omega_G) \text{ where } G^2 = (\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \tag{48}$$

Therefore

$$F'(G^2) = \int \cdots \int_{\Omega_G} \mathcal{A} f\{(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\}\, d^n\boldsymbol{x}. \tag{49}$$

Applying the results of Section 3, we have

$$F'(G^2) = \frac{\int_{g=0}^{G} f(g^2)\, g^{n-1}\, dg}{\int_{g=0}^{\infty} f(g^2)\, g^{n-1}\, dg}. \tag{50}$$

4.2. **An Example: The Multinormal Statistic.** To illustrate the use of the statistic, we will evaluate its p.d.f. for the case of the multinormal distribution. Using Equation 43 to define the univariate p.d.f, we have

$$\int_{g=0}^{G} e^{-\frac{1}{2}g^2} g^{n-1}\, dg = 2^{\frac{n}{2}-1} \int_{u=0}^{\frac{1}{2}G^2} e^{-u} u^{\frac{n}{2}-1}\, du. \tag{51}$$

This integral is the "lower" incomplete gamma function[6], $\gamma(\cdot)$. Therefore

$$F'(G^2) = \frac{\gamma(\frac{n}{2}, \frac{1}{2}G^2)}{\Gamma(\frac{n}{2})}. \tag{52}$$

For an integral argument, n, the incomplete gamma function has the simple series expansion[6]:

$$\gamma(n,x) = \Gamma(n)\left(1 - e^{-x}\sum_{k=0}^{n-1}\frac{x^k}{k!}\right). \tag{53}$$

For an even number of dimensions, $n \in \{2, 4, 6\dots\}$, we have

$$F'(G^2) = 1 - e^{-\frac{1}{2}G^2}\sum_{k=0}^{\frac{n}{2}-1}\frac{G^{2k}}{2^k k!}. \tag{54}$$

The density $f'(G^2)$ is given by

$$f'(G^2) = \frac{dF'(G^2)}{dG^2} = \frac{1}{\Gamma(\frac{n}{2})}\frac{\partial\gamma(\frac{n}{2}, \frac{1}{2}G^2)}{\partial G^2}. \tag{55}$$

The derivative of $\gamma(\alpha, x)$ w.r.t. x is $x^{\alpha-1}e^{-x}$[7]. Using this result, Equation 55 becomes

$$f'(z) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}z^{\frac{n}{2}-1}e^{-\frac{1}{2}z} \tag{56}$$

(where we have written z for G^2). This is, of course, the well known χ^2 distribution for n degrees of freedom[19]. It has the mean n and variance $2n$.

5. Measures of Location, Dispersion and Shape for Multivariate Distributions with Ellipsoidal Symmetry

5.1. The Population Mean, Mode and Median.

The population mean and mode are the simplest measures of location to define. The mean is given by

$$E\boldsymbol{x} = \mathcal{A}\int\cdots\int_\Omega \boldsymbol{x}f\{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}\,d^n\boldsymbol{x}. \tag{57}$$

Using the transformations discussed above, we may write Equation 57 as

$$E\boldsymbol{x} = \boldsymbol{\mu} + \mathcal{A}R^T D|D|\int\cdots\int_\Omega \boldsymbol{g}f(g^2)\,d^n\boldsymbol{g}. \tag{58}$$

The integral on the r.h.s. of Equation 58 clearly vanishes by virtue of our definition of $f(g^2)$, leaving the simple result $E\boldsymbol{x} = \boldsymbol{\mu}$.

The population mode is the most likely value of the coordinate vector $\boldsymbol{x}$. i.e.

$$\boldsymbol{m} = \arg\max_{\boldsymbol{x}\in\Omega} f\{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}. \tag{59}$$

If the p.d.f. is differentiable everywhere, then $\boldsymbol{m}$ is a solution of

$$\boldsymbol{\nabla}f\{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\} = \frac{df}{dz}\boldsymbol{\nabla}z = \boldsymbol{0} \text{ where } z = (\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \tag{60}$$

z is clearly minimized at $\boldsymbol{x} = \boldsymbol{\mu}$. If f is a monotonic decreasing function of z, then the p.d.f. will be unimodal with $\boldsymbol{m} = \boldsymbol{\mu}$. If f is a monotonic increasing function of z, for a finite domain Ω, then $\boldsymbol{m}$ must be a region of the surface S bounding Ω. If f is not a monotonic function then $\boldsymbol{m}$ is any point on the elliptical shell $z = z_0$ where z_0 is the global maximum of $f(z)$ within Ω.

The median, M, of a continuous univariate p.d.f. may be defined to be the solution of the equation

$$\int_{-\infty}^{M} f(x)\,dx = \int_{M}^{\infty} f(x)\,dx. \tag{61}$$

For a multivariate p.d.f, we may similarly define the median to be the surface S_M that divides the domain into two distinct sets of regions $\Omega_1 = \cup_i \Omega_{1i}$ and $\Omega_2 = \cup_i \Omega_{2i}$. (Note that S_M is not necessarily simply connected.) S_M must satisfy

$$\int \cdots \int_{\Omega_1} f(g^2)\, d^n\boldsymbol{g} = \int \cdots \int_{\Omega_2} f(g^2)\, d^n\boldsymbol{g}. \tag{62}$$

There is clearly no unique solution to this equation in terms of S_M when $n > 1$; however, a choice that makes sense for unimodal distributions is the ellipsoidal shell S_M enclosing the region Ω_M centered on $\boldsymbol{\mu}$. M^2 is defined by Equation 48 and its particular value is chosen to satisfy Equation 62 with $\Omega_1 = \Omega_M$ (and Ω_2 its complement).

Using the ellipsoidal symmetry of a multivariate p.d.f. constructed using the methodology presented here, Equation 62 then becomes

$$\int_{g=0}^{M} f(g^2)\, g^{n-1}\, dg = \int_{g=M}^{\infty} f(g^2)\, g^{n-1}\, dg. \tag{63}$$

5.2. **The Population Covariance Matrix.** Let V be the population covariance matrix for our constructed multivariate p.d.f.Using the notation developed above, we may write the ijth. element of this matrix as

$$V_{ij} = \mathcal{A} \int \cdots \int_{\Omega} y_i y_j f(\boldsymbol{y}^T \Sigma^{-1} \boldsymbol{y})\, d^n\boldsymbol{y}. \tag{64}$$

Using the transformation of Equation 36, Equation 64 may be written

$$V_{ij} = \mathcal{A}|D| \sum_{klpq} R_{ki} D_{kl} R_{pj} D_{pq} \int_{g_1=-\infty}^{\infty} \cdots \int_{g_n=-\infty}^{\infty} g_l g_q f(g^2)\, d^n\boldsymbol{g}. \tag{65}$$

Due to the symmetry of $f(g^2)$, the integral in Equation 65 may be written

$$\begin{aligned} \int_{g_1=-\infty}^{\infty} \cdots \int_{g_n=-\infty}^{\infty} g_l g_q f(g^2)\, d^n\boldsymbol{g} &= \delta_{lq} \int_{g_1=-\infty}^{\infty} \cdots \int_{g_n=-\infty}^{\infty} g_q^2 f(g^2)\, d^n\boldsymbol{g} \\ &= \frac{\delta_{lq}}{n} \int_{g_1=-\infty}^{\infty} \cdots \int_{g_n=-\infty}^{\infty} g^2 f(g^2)\, d^n\boldsymbol{g} \\ &= \frac{2\pi^{\frac{n}{2}} \delta_{lq}}{n\Gamma(\frac{n}{2})} \int_{g=0}^{\infty} f(g^2)\, g^{n+1}\, dg. \end{aligned} \tag{66}$$

Here δ_{lq} is the Kronecker delta. (Note that the final step in Equation 66 is a transformation to a polar integral. This introduces the $\pi^{n/2}$ and gamma function terms.) Substituting this expression into Equation 65 and summing over the index q, gives

$$V_{ij} = \frac{2\pi^{\frac{n}{2}} \mathcal{A}|D|}{n\Gamma(\frac{n}{2})} \sum_{klp} R_{ki} D_{kl} R_{pj} D_{pl} \int_{g=0}^{\infty} f(g^2)\, g^{n+1}\, dg. \tag{67}$$

From Equation 34, we see that the summation in Equation 67 is equal to Σ_{ij}. After Substituting the value of $\mathcal{A}$ from Equation 47, we have

$$V_{ij} = \frac{\Sigma_{ij}}{n} \frac{\int_{g=0}^{\infty} f(g^2)\, g^{n+1}\, dg}{\int_{g=0}^{\infty} f(g^2)\, g^{n-1}\, dg}. \tag{68}$$

The covariance matrix for the constructed distribution is seen to be proportional to the matrix used to define the metric distance. Obviously, this factor can be eliminated by rescaling the matrix Σ. When a metric distance is defined with the covariance matrix itself, then this distance is referred to as the Mahalanobis distance[8][11], $\Delta_V^2(\boldsymbol{x}, \boldsymbol{\mu})$. Equation 68 shows that the metric distance defined here is always proportional to the Mahalanobis distance between the random vector $\boldsymbol{x}$ and its expectation.

For the multinormal distribution, the ratio of integrals in Equation 68 is equal to n and so, for this distribution, we have the required result that $V = \Sigma$.

5.3. **Measures of Distributional Shape.** In additional to the location and width of a distribution, we are interested in characterising the dispersion of probability density between distributions after standardizing with respect to these two factors. For univariate distributions this is commonly done using the skew and kurtosis parameters, which are computed from the third and fourth central moments of the p.d.f, respectively.

Pearson proposed the ratio of the distance from the mean to the mode in units of the population standard deviation[15] as a measure of the asymmetry of a p.d.f.A multivariate extension of this concept would be a vector proportional to the Mahanalobis distance between the mean and the mode and directed parallel to that vector. e.g.

$$\boldsymbol{s} = \frac{\Delta_V(\boldsymbol{\mu}, \boldsymbol{m})}{\Delta_I(\boldsymbol{\mu}, \boldsymbol{m})}(\boldsymbol{\mu} - \boldsymbol{m}). \tag{69}$$

(In the case of a multimodal p.d.f, it would make sense to average this quantity over all of the modes.) Mardia[12] defines an alternate measure based upon multivariate generalizations of the β_1 and β_2 parameters that arise in discussion of the Pearson System of distributions[16]. Mardia's skewness is

$$\beta_{1,n} = E\{(\boldsymbol{x} - \boldsymbol{\mu})^T V^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\}^3, \tag{70}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are i.i.d. with covariance matrix V. Both of these measures vanish in the case of ellipsoidal symmetry.

The role of the kurtosis measure is to parameterize the way in which the probability density is spread within the domain of definition of the p.d.f.The β_2 parameter, defined to be the ratio of the fourth central moment to the square of the second central moment[15], is generally taken to be indicative of this spread (often "standardized" to relative to the Normal by subtracting 3). Many univariate distributions with positive kurtosis exhibit a functional form with a sharper peak at the mean and with heavy tails. Those with a negative kurtosis exhibit a broad flat shape near the mean and light tails[9]. Mardia defines the measure

$$\beta_{2,n} = E\{(\boldsymbol{x} - \boldsymbol{\mu})^T V^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\}^2, \tag{71}$$

as a multivariate generalization of the kurtosis parameter. If we write the constant of proportionality from Equation 68 as $1/\xi$, then we have $\beta_{2,n} = \xi^2 E g^4$.

In Section 4.1 we obtained an expression for the c.d.f. of the metric distance. Using the z variable defined above, Equation 50 can be written

$$F'(z) = \frac{1}{2}\frac{\int_0^z f(u)\, u^{\frac{n}{2}-1}\, du}{\int_0^\infty f(g^2)\, g^{n-1}\, dg}. \tag{72}$$

Differentiating w.r.t. z gives the p.d.f. of z:

$$f'(z) = \frac{1}{2}\frac{f(z)\, z^{\frac{n}{2}-1}}{\int_0^\infty f(g^2)\, g^{n-1}\, dg}. \tag{73}$$

[8]The subscript V indicates that the distance is calculated using the matrix V^{-1}.

[9]Although there are many counter examples to this generalization within the literature, it is a useful description of these functional shapes.

Evaluating $Ez^2 = \int_0^\infty z^2 f'(z)\,dz$, and transforming back to the g variable, gives

$$Eg^4 = \frac{\int_0^\infty f(g^2)\,g^{n+3}\,dg}{\int_0^\infty f(g^2)\,g^{n-1}\,dg} \Rightarrow \beta_{2,n} = n^2 \frac{\int_0^\infty f(g^2)\,g^{n+3}\,dg \int_0^\infty f(g^2)\,g^{n-1}\,dg}{\left\{\int_0^\infty f(g^2)\,g^{n+1}\,dg\right\}^2}. \tag{74}$$

As in Section 5.2, this expression has a particularly simple form for multinormal distribution. In this case, we have $\beta_{2,n}^N = n(n+2)$. (Note that, for the univariate case, $\beta_{2,1}^N = 3$). We may also define a multivariate generalization of the *excess* kurtosis as $\beta_{2,n} - \beta_{2,n}^N$.

5.4. **Invariance Under Affine Transformations.** An Affine transformation is one of the form $\boldsymbol{x} \to A\boldsymbol{x} + \boldsymbol{b}$ where A is a non-singular matrix and $\boldsymbol{b}$ is a vector. The Mahanalobis distance is invariant under an Affine transformation, i.e.

$$(\boldsymbol{x}' - \boldsymbol{\mu}')^T \Sigma'^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}') = (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \text{ where } \begin{cases} \boldsymbol{x}' &= A\boldsymbol{x} + \boldsymbol{b} \\ \boldsymbol{\mu}' &= A\boldsymbol{\mu} + \boldsymbol{b} \\ \Sigma' &= A\Sigma A^T \end{cases}. \tag{75}$$

The normalized elemental volume is also invariant under the associated transformation of the matrix Σ, i.e. $d^n\boldsymbol{x}/|\Sigma|^{1/2} = d^n\boldsymbol{x}'/|\Sigma'|^{1/2}$.

Any p.d.f. constructed in the manner described above[10] may be written in the form: $f\{\Delta_\Sigma(\boldsymbol{x}, \boldsymbol{\mu})\}\,d^n\boldsymbol{x}/|\Sigma|^{1/2}$. It follows that any such distribution is invariant under an Affine transformation. i.e. $dF(\boldsymbol{x}'|\boldsymbol{\mu}', \Sigma') = dF(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma)$, meaning that if $\boldsymbol{x}$ is drawn from some particular p.d.f, then $\boldsymbol{x}'$ will also be drawn from that same p.d.f. (albeit with transformed parameters). Furthermore, it is also clearly true that any statistic which is a function of the Manalobis distance is also invariant under an Affine transformation.

6. The Characteristic Function and the Moment Generating Function

6.1. **The Characteristic Function.** A function derived from a multivariate density that has many useful applications is the characteristic function, $\phi(\boldsymbol{k}) = Ee^{i\boldsymbol{k}\cdot\boldsymbol{x}}$. For a general p.d.f, defined as above, we have

$$\phi(\boldsymbol{k}) = \mathcal{A} \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} f\{(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\}\,d^n\boldsymbol{x}. \tag{76}$$

Making same change of variables as previously, we may write Equation 76 as

$$\phi(\boldsymbol{k}) = \mathcal{A}|\Sigma|^{\frac{1}{2}} e^{i\boldsymbol{k}\cdot\boldsymbol{\mu}} \int_{g_1=-\infty}^{\infty} \cdots \int_{g_n=-\infty}^{\infty} e^{i\boldsymbol{k}\cdot R^T D\boldsymbol{g}} f(g^2)\,d^n\boldsymbol{g}. \tag{77}$$

It is a well known result that the n-dimensional Fourier Transform of a radial function is equivalent to a one dimensional Hankel Transform[10]. For space-vector $\boldsymbol{x}$ and wave-vector $\boldsymbol{y}$, the result is

$$\mathcal{F}[u](\boldsymbol{y}) = \frac{1}{2\pi^{\frac{n}{2}}} \int_{x_1=0}^{\infty} \cdots \int_{x_n=0}^{\infty} e^{i\boldsymbol{x}\cdot\boldsymbol{y}} u(x)\,d^n\boldsymbol{x} = y^{1-\frac{n}{2}} \int_{x=0}^{\infty} u(x) J_{\frac{n}{2}-1}(xy)\,x^{\frac{n}{2}}\,dx, \tag{78}$$

where $J_\nu(x)$ is the Bessel function of the first kind, $\mathcal{F}[u](\cdot)$ is the Fourier Transform of $u(\cdot)$, and $x = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$ and $y = \sqrt{\boldsymbol{y}^T\boldsymbol{y}}$. Clearly

$$\phi(\boldsymbol{k}) = (2\pi)^{\frac{n}{2}} \mathcal{A}|\Sigma|^{\frac{1}{2}} e^{i\boldsymbol{k}\cdot\boldsymbol{\mu}} \mathcal{F}[f](DR\boldsymbol{k}). \tag{79}$$

[10] In fact, any p.d.f. with ellipsoidal symmetry.

Let $\boldsymbol{s} = DR\boldsymbol{k} \Rightarrow s^2 = \boldsymbol{s}^T\boldsymbol{s} = \boldsymbol{k}^T R^T D^2 R\boldsymbol{k} = \Delta^2_{\Sigma^{-1}}(\boldsymbol{k}, \boldsymbol{0})$. Substituting this value into Equation 79 gives

$$\phi(\boldsymbol{k}) = e^{i\boldsymbol{k}\cdot\boldsymbol{\mu}}\Gamma\left(\frac{n}{2}\right)\left\{\frac{2}{\Delta_{\Sigma^{-1}}(\boldsymbol{k},\boldsymbol{0})}\right\}^{\frac{n}{2}-1} \frac{\int_0^\infty f(g^2)J_{\frac{n}{2}-1}\{g\Delta_{\Sigma^{-1}}(\boldsymbol{k},\boldsymbol{0})\}\, g^{\frac{n}{2}}\,dg}{\int_0^\infty f(g^2)\, g^{n-1}\,dg}. \tag{80}$$

6.2. **The Moment Generating Function.** Another useful function is the moment generating function, defined to be $Ee^{-\boldsymbol{k}\cdot\boldsymbol{x}}$. This is clearly given by the characteristic function evaluated for a purely imaginary argument, $i\boldsymbol{k}$. We recognize it as the Bilateral Laplace Transformation. Making the substitution $\boldsymbol{k} \to i\boldsymbol{k}$ in Equation 80, and using $\Delta_M(i\boldsymbol{x}, i\boldsymbol{y}) = i\Delta_M(\boldsymbol{x}, \boldsymbol{y})$, gives

$$\phi(i\boldsymbol{k}) = e^{-\boldsymbol{k}\cdot\boldsymbol{\mu}}\Gamma\left(\frac{n}{2}\right)\left\{\frac{2}{\Delta_{\Sigma^{-1}}(\boldsymbol{k},\boldsymbol{0})}\right\}^{\frac{n}{2}-1} \frac{\int_0^\infty f(g^2)I_{\frac{n}{2}-1}\{g\Delta_{\Sigma^{-1}}(\boldsymbol{k},\boldsymbol{0})\}\, g^{\frac{n}{2}}\,dg}{\int_0^\infty f(g^2)\, g^{n-1}\,dg}. \tag{81}$$

where $I_\nu(x) = i^{-\nu}J_\nu(ix)$ is the modified Bessel function of the first kind.

In making the transformation above we have glossed over the convergence criteria for the upper integral. In the direction $-\boldsymbol{\mu}$ the transform kernel is diverging exponentially and so the p.d.f. must converge faster than this and in the direction $\boldsymbol{\mu}$ the kernal is converging exponentially and so the p.d.f. cannot diverge faster than this. If these criteria are not met, the m.g.f. will not exist, although the c.f. may.

However, in view of the formula above, we can directly specify the convergence criteria in terms of the g variable alone. From the series expansion of $I_\nu(x)$[8] we see that all the functions of half integral order converge at the origin, except for the function $I_{-1/2}(x)$ which diverges as $x^{-1/2}$ and this divergence is cancelled by the $x^{1/2}$ term in the corresponding integrand. For large x the all of the functions converge to $e^x/\sqrt{2\pi x}$ and therefore they diverge slower than exponentially.

6.3. **The Roots of the Gradient of the Moment Generating Function.** In some circumstances (e.g. asset allocation problems) it is useful to evaluate the roots of the gradient of the m.g.f. Removing all of the extraneous scale factors and writing $\Delta(\boldsymbol{k})$ for $\Delta_{\Sigma^{-1}}(\boldsymbol{k}, \boldsymbol{0})$, we may define a modified m.g.f. as

$$\psi_{\frac{n}{2}}(\boldsymbol{k}) = e^{-\boldsymbol{k}\cdot\boldsymbol{\mu}}\{\Delta(\boldsymbol{k})\}^{1-\frac{n}{2}} \int_0^\infty f(g^2)I_{\frac{n}{2}-1}\{g\Delta(\boldsymbol{k})\}\, g^{\frac{n}{2}}\,dg. \tag{82}$$

Let $\boldsymbol{\nabla_k}$ be the vector with components $\partial/\partial k_i$. When taking the gradient of Equation 82, we note that the symmetry of Σ means that $\boldsymbol{\nabla_k}\Delta(\boldsymbol{k}) = \Sigma\boldsymbol{k}/\Delta(\boldsymbol{k})$.

Solving for the roots of the gradient of $\psi(\cdot)$, we find that the solution may be written

$$\boldsymbol{k}\Psi_{\frac{n}{2}}\{\Delta(\boldsymbol{k})\} = \Sigma^{-1}\boldsymbol{\mu}, \tag{83}$$

where $\Psi_\nu(x)$ is a scalar function. This shows that, when $\Psi(\cdot)$ is convergent, the root is always in the direction $\Sigma^{-1}\boldsymbol{k}$. When $\Psi(\cdot)$ diverges for finite $\boldsymbol{k}$ the root is at the origin. The function $\Psi_\nu(x)$ from Equation 83 is defined to be

$$\Psi_\nu(x) = \frac{1}{x}\frac{\int_0^\infty f(g^2)I_\nu(gx)\, g^{\nu+1}\,dg}{\int_0^\infty f(g^2)I_{\nu-1}(gx)\, g^{\nu}\,dg} \tag{84}$$

for non-negative x and $\nu \geq 1/2$. If Equation 83 has a non-trivial root, it is simple to show that the value of $\Delta(\boldsymbol{k})$ at the root is the solution to

$$x\Psi_{\frac{n}{2}}(x) = \sqrt{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}. \tag{85}$$

6.4. **Example: The Multinormal Solution.** For the multinormal distribution the integral of Equation 82 is well known[11] and is independent of n.

$$\psi(\boldsymbol{k}) = \frac{1}{\sqrt{2\pi}} e^{-\boldsymbol{k}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{k}^T\Sigma\boldsymbol{k}}. \tag{86}$$

For this p.d.f, the $\Psi(\cdot)$ function is identically equal to unity and the root of the gradient of $\psi(\cdot)$ is always $\boldsymbol{k} = \Sigma^{-1}\boldsymbol{\mu}$. Therefore the function $\Psi(\cdot)$ can be thought of a factor scaling the solution for the normal distribution in a manner controlled by the actual distribution in use.

7. Maximum Likelihood Estimation

In this section we will derive an expression for the maximum likelihood estimator of $\boldsymbol{\mu}$, written $\hat{\boldsymbol{\mu}}$, and that of Σ, written $\hat{\Sigma}$. Let the multivariate p.d.f. under study be parameterized by the set $\{\boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}\}$. The vector $\boldsymbol{\theta}$ represents all distributional parameters not specified by $\boldsymbol{\mu}$ and Σ and is not to be interpreted as a vector within the coordinate space that defines $\boldsymbol{\mu}$ and Σ.

Let $\{\boldsymbol{x}_i\}_{i=1}^N$ be a set of N observations drawn from a p.d.f. with ellipsoidal symmetry. The probability density associated with observation i may be written

$$dF(\boldsymbol{x}_i) = \mathcal{A}(\Sigma, \boldsymbol{\theta}) f(g_i^2, \boldsymbol{\theta})\, d^n\boldsymbol{x}_i \text{ where } g_i^2 = (\boldsymbol{x}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}). \tag{87}$$

Therefore, the log-likelihood for the entire sample, $\{\boldsymbol{x}_i\}$, is

$$L(\boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}) = N \ln \mathcal{A}(\Sigma, \boldsymbol{\theta}) + \sum_{i=1}^N \ln f(g_i^2, \boldsymbol{\theta}). \tag{88}$$

7.1. **The Maximum Likelihood Estimator of the Population Mean.** The maximum likelihood estimator of $\boldsymbol{\mu}$ is defined to be the value of $\boldsymbol{\mu}$ that maximizes Equation 88. If the first derivative of the log-likelihood is continuous everywhere within the region where the p.d.f. is defined, then this is equal to the root of $\boldsymbol{\nabla}_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \Sigma, \boldsymbol{\theta})$. ($\boldsymbol{\nabla}_{\boldsymbol{\mu}}$ is the vector with components $\partial/\partial\mu_i$.) The gradient of the log-likelihood is

$$\boldsymbol{\nabla}_{\boldsymbol{\mu}} L = \sum_{i=1}^N \boldsymbol{\nabla}_{\boldsymbol{\mu}} \ln f(g_i^2, \boldsymbol{\theta}) = \sum_{i=1}^N \frac{f'(g_i^2, \boldsymbol{\theta})}{f(g_i^2, \boldsymbol{\theta})} \boldsymbol{\nabla}_{\boldsymbol{\mu}} g_i^2, \tag{89}$$

where $f'(\cdot)$ is the first derivative of $f(\cdot)$ w.r.t. g_i^2. Since Σ is symmetric by definition, we have $\boldsymbol{\nabla}_{\boldsymbol{\mu}} g_i^2 = 2\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{x}_i)$. Substituing this result into Equation 89 gives

$$\boldsymbol{\nabla}_{\boldsymbol{\mu}} L = 2\Sigma^{-1}\left\{\boldsymbol{\mu}\sum_{i=1}^N \frac{f'(g_i^2, \boldsymbol{\theta})}{f(g_i^2, \boldsymbol{\theta})} - \sum_{i=1}^N \frac{f'(g_i^2, \boldsymbol{\theta})}{f(g_i^2, \boldsymbol{\theta})}\boldsymbol{x}_i\right\}. \tag{90}$$

Therefore, the maximum likelihood estimator $\hat{\boldsymbol{\mu}}$ is the solution to

$$\boldsymbol{\mu}\sum_{i=1}^N \frac{f'(g_i^2, \boldsymbol{\theta})}{f(g_i^2, \boldsymbol{\theta})} = \sum_{i=1}^N \frac{f'(g_i^2, \boldsymbol{\theta})}{f(g_i^2, \boldsymbol{\theta})}\boldsymbol{x}_i. \tag{91}$$

This expression clearly shows that the maximum likelihood estimator of the population mean is a "weighted mean" of the sample vectors.

---

[11] In fact, for this case, the Fourier Transform is straightforward in Cartesian coordinates.

7.2. **The Sample Mean Differential Equation.** If the function $f(\cdot)$ satisfies the partial differential equation:

$$\frac{\partial f(z,\boldsymbol{\theta})}{\partial z} = f(z,\boldsymbol{\theta})h(\boldsymbol{\theta}), \tag{92}$$

where $h(\boldsymbol{\theta})$ is an unknown function that is not dependent on $\boldsymbol{x}_i$, $\boldsymbol{\mu}$ or Σ, then the ratio $f'(\cdot)/f(\cdot)$ is invariant under the sums of Equation 91. Under these conditions, the root of Equation 91 is

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i. \tag{93}$$

So for a p.d.f. that satisfies Equation 92, the maximum likelihood estimator of the population mean is the sample mean.

7.3. **A General Solution to the Sample Mean Differential Equation.** Let $f(z,\boldsymbol{\theta})$ have the form $a(\boldsymbol{\theta})e^{b(z,\boldsymbol{\theta})}$. Then differentiating w.r.t. z gives

$$\frac{\partial f(z,\boldsymbol{\theta})}{\partial z} = f(z,\boldsymbol{\theta})\frac{\partial b(z,\boldsymbol{\theta})}{\partial z}. \tag{94}$$

Comparing this expression with Equation 92 shows that the function $b(\cdot)$ must be linear in z. i.e. that $b(z,\boldsymbol{\theta}) = h(\boldsymbol{\theta})z + c(\boldsymbol{\theta})$. Absorbing the constant of partial integration, $c(\boldsymbol{\theta})$, into the definition of the function $a(\boldsymbol{\theta})$, leads to the following general solution to the p.d.e. of Equation 92:

$$f(z,\boldsymbol{\theta}) = a(\boldsymbol{\theta})e^{h(\boldsymbol{\theta})z}. \tag{95}$$

Our definition of $f(x^2,\boldsymbol{\theta})$ as a univariate probability density function imposes the constraint that

$$\int_{x=-\infty}^{\infty} f(x^2,\boldsymbol{\theta})\,dx = 1 \Rightarrow f(z,\boldsymbol{\theta}) = \sqrt{\frac{-h(\boldsymbol{\theta})}{\pi}}e^{h(\boldsymbol{\theta})z}. \tag{96}$$

For convergence of the normalization integral in Equation 96, $h(\boldsymbol{\theta})$ must be everywhere real and negative. Clearly a particular solution to Equation 92 is

$$h(\boldsymbol{\theta}) = -\frac{1}{2} \Rightarrow f(x^2) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}. \tag{97}$$

7.4. **The Maximum Likelihood Estimator of the Covariance Parameter.** As in Section 7.1, the maximum likelihood estimator of the covariance parameter, $\hat{\Sigma}$ is defined to be the root of $\nabla_\Sigma L(\boldsymbol{\mu},\Sigma,\boldsymbol{\theta})$, where ∇_Σ is the matrix with elements $\partial/\partial\Sigma_{ij}$.

Using Equation 47, we see that Equation 88 may be written

$$L(\boldsymbol{\mu},\Sigma,\boldsymbol{\theta}) = N\ln\mathcal{B}(\boldsymbol{\theta}) - \frac{N}{2}\ln|\Sigma| + \sum_{i=1}^{N}\ln f(g_i^2,\boldsymbol{\theta}). \tag{98}$$

Therefore

$$\nabla_\Sigma L = \sum_{i=1}^{N}\frac{f'(g_i^2,\boldsymbol{\theta})}{f(g_i^2,\boldsymbol{\theta})}\nabla_\Sigma g_i^2 - \frac{1}{2}N\nabla_\Sigma\ln|\Sigma|. \tag{99}$$

Now, Jacobi's formula for the derivative of the determinant of an invertible matrix is $d|A| = |A|\operatorname{tr}(A^{-1}dA)$, so for a symmetric invertible matrix we have the remarkably compact result $\nabla_A \ln|A| = A^{-1}$. Additionally, $dA^{-1} = -A^{-1}\, dA\, A^{-1}$, which gives the result $\nabla_A(\boldsymbol{a}^T A^{-1}\boldsymbol{b}) = -A^{-1}\boldsymbol{a}\boldsymbol{b}^T A^{-1}$. Using these expressions in Equation 99, gives

$$\hat{\Sigma} = -\frac{2}{N}\sum_{i=1}^{N}\frac{f'(g_i^2,\boldsymbol{\theta})}{f(g_i^2,\boldsymbol{\theta})}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T. \tag{100}$$

For the special case of the multinormal distribution, the we see that this becomes the familiar result

$$\hat{V} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T. \tag{101}$$

7.5. **The Maximum Likelihood Estimator of the Distributional Parameters.** Finally, we define the maximum likelihood estimator of the additional distributional parameter set, $\hat{\boldsymbol{\theta}}$, as the root of $\boldsymbol{\nabla_\theta} L(\boldsymbol{\mu},\Sigma,\boldsymbol{\theta})$. Differentiating Equation 98, we see that $\hat{\boldsymbol{\theta}}$ is the solution of

$$\frac{\boldsymbol{\nabla_\theta}\mathcal{B}(\boldsymbol{\theta})}{\mathcal{B}(\boldsymbol{\theta})} = -\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{\nabla_\theta} f(g_i^2,\boldsymbol{\theta})}{f(g_i^2,\boldsymbol{\theta})}. \tag{102}$$

8. The Simulation of Multivariate Distributions

Analysis of the results of Monte-Carlo simulation of data drawn from a given p.d.f. is a powerful technique in applied probability and statistics. In this section we will briefly consider a method that may be used to generate a distribution with ellipsoidal symmetry.

8.1. **General Methodology.** For the ellipsoidal distributions discussed here, this simulation is a relatively straightforward procedure. We may follow the procedure of Marsaglia[13] to pick points uniformly distributed on a hypersphere in n-space. Remarkably, this procedure is to simply pick a vector $\hat{\boldsymbol{f}} \sim N(\boldsymbol{0},\boldsymbol{1}_n)$ and then normalize the vector to unity i.e. $\hat{\boldsymbol{g}} = \hat{\boldsymbol{f}}/|\hat{\boldsymbol{f}}|$ is uniformly distributed on the surface of an $(n-1)$-sphere. To generate a vector with the appropriate covariance matrix we reverse the transformation of Equation 36 to generate the hyperellipsoid $\hat{\boldsymbol{y}} = R^T D\hat{\boldsymbol{g}}$. Since the radial and angular motions are independent for an ellipsoidal variate, a vector drawn from the full p.d.f. may then be computed as

$$\boldsymbol{x} = \boldsymbol{\mu} + R^T D g\hat{\boldsymbol{g}}, \tag{103}$$

where g is a scale factor drawn from the p.d.f. of the Mahanalobis distance $\Delta_{\boldsymbol{1}_n}$.

The distribution of $F''(G) = \Pr(g < G)$ represents the probability that a random vector $\boldsymbol{g}$ lies within the volume enclosed by a hyperspherical shell, Ω_G, with radius G. Since the radius of a sphere is non-negative by definition we see that $\Pr(g < G) = \Pr(g^2 < G^2)$. This latter expression has already been evaluated in general in connection with the Kolmogorov statistic of Section 4.1. Hence, we have

$$F''(G) = \frac{\int_{g=0}^{G} f(g^2)\, g^{n-1}\, dg}{\int_{g=0}^{\infty} f(g^2)\, g^{n-1}\, dg} \text{ and } f''(G) = \frac{f(G^2)\, G^{n-1}}{\int_{g=0}^{\infty} f(g^2)\, g^{n-1}\, dg}. \tag{104}$$

(The expression for the density has been obtained by differentiating the p.d.f. w.r.t. the upper limit of the integral in the numerator.) Once the univariate distribution is known it is a simple matter to generate a variate with the associated univariate density by the standard technique of drawing u from the uniform distribution $U(0,1)$ and solving $F''(g) = u$ for g.

8.2. **An Example: Generation of Multinormal Variates.** To give a verifiable example of the method we will evaluate the p.d.f. $f''(G)$ for the multinormal case[12]. Substituting $f(g^2) = e^{-\frac{1}{2}g^2}/\sqrt{2\pi}$ into Equation 104, it is straightforward to demonstrate that

$$f''(g) = \frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} g^{n-1} e^{-\frac{1}{2}g^2}. \tag{105}$$

This is recognizable as the density for the χ_n distribution, which is the distribution of the square root of a χ_n^2 variate[2]. This is clearly the definition of the normalization factor removed in Section 8.1, on the preceding page. For the univariate case it is trivial to verify that this identical to the p.d.f. of $|x|$ where $x \sim N(0, 1)$.

References

[1] Eadie, W.T., Drijard, D., James, F.E., Roos, M. & Sadoulet, B., "Statistical Methods in Experimental Physics," pp. 268–269, North-Holland/American Elsevier, 1971.
[2] Evans, M., Hastings, N., & Peacock, B., "Statistical Distributions," 3rd. Edn., p. 57 John Wiley & Sons, Inc., 2000.
[3] Justel, Ana, Peña, Daniel, & Zamar, Rubén, "A Multivariate Kolmogorov-Smirnov Test of Goodness of Fit," *Statistics & Probability Letters*, Vol. **35** (1997), pp. 251–259
[4] Gradshteyn, I.S. & Ryzhik, I.M., "Table of Integrals, Series, and Products," 6th. Edn., p. 389, Academic Press, 2000.
[5] Gradshteyn, I.S. & Ryzhik, I.M., "Table of Integrals, Series, and Products," 6th. Edn., p. 884, Academic Press, 2000.
[6] Gradshteyn, I.S. & Ryzhik, I.M., "Table of Integrals, Series, and Products," 6th. Edn., p. 890, Academic Press, 2000.
[7] Gradshteyn, I.S. & Ryzhik, I.M., "Table of Integrals, Series, and Products," 6th. Edn., p. 891, Academic Press, 2000.
[8] Gradshteyn, I.S. & Ryzhik, I.M., "Table of Integrals, Series, and Products," 6th. Edn., p. 909, Academic Press, 2000.
[9] Gradshteyn, I.S. & Ryzhik, I.M., "Table of Integrals, Series, and Products," 6th. Edn., p. 910, Academic Press, 2000.
[10] Ito, K. [Editor], "The Encyclopedic Dictionary of Mathematics," 2nd. (English) Edn., p. 1763 Mathematical Society of Japan, 1986.
[11] Mardia, K.V., Kent, J.T., & Bibby, J.M., "Multivariate Analysis," p. 31, Academic Press, 1979.
[12] Mardia, K.V., "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika*, Vol. **57** (1970), pp. 519–530
[13] Marsaglia, G., "Sampling from the Surface of a Sphere," *Ann. Math. Stat.*, Vol. **43** (1972), p. ?
[14] Stuart, Alan & Ord, Keith, "Kendall's Advanced Theory of Statistics," Vol. I, 6th. Edn., p. 20, Edward Arnold, 1994.
[15] Stuart, Alan & Ord, Keith, "Kendall's Advanced Theory of Statistics," Vol. I, 6th. Edn., p. 108, Edward Arnold, 1994.
[16] Stuart, Alan & Ord, Keith, "Kendall's Advanced Theory of Statistics," Vol. I, 6th. Edn., pp. 215–240, Edward Arnold, 1994.
[17] Stuart, Alan & Ord, Keith, "Kendall's Advanced Theory of Statistics," Vol. I, 6th. Edn., p. 267, Edward Arnold, 1994.
[18] Stuart, Alan & Ord, Keith, "Kendall's Advanced Theory of Statistics," Vol. I, 6th. Edn., pp. 374–375, Edward Arnold, 1994.
[19] Stuart, Alan & Ord, Keith, "Kendall's Advanced Theory of Statistics," Vol. I, 6th. Edn., p. 376, Edward Arnold, 1994.

[12] Of course, there are more efficient methods available to generate variates drawn from the multinormal distribution should that be the p.d.f. the analyst wishes to simulate.