IEOR E4602: Quantitative Risk Management Spring 2016

Multivariate Distributions

We will study multivariate distributions in these notes, focusing[1] in particular on multivariate normal, normal-mixture, spherical and elliptical distributions. In addition to studying their properties, we will also discuss techniques for simulating and, very briefly, estimating these distributions. Familiarity with these important classes of multivariate distributions is important for many aspects of risk management. We will defer the study of *copulas* until later in the course.

1 Preliminary Definitions

Let $\mathbf{X} = (X_1, \ldots X_n)$ be an n-dimensional vector of random variables. We have the following definitions and statements.

Definition 1 (Joint CDF) *For all $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, the joint cumulative distribution function (CDF) of $\mathbf{X}$ satisfies*

$$F_{\mathbf{X}}(\mathbf{x}) \;=\; F_{\mathbf{X}}(x_1, \ldots, x_n) \;=\; P(X_1 \le x_1, \ldots, X_n \le x_n).$$

Definition 2 (Marginal CDF) *For a fixed i, the marginal CDF of X_i satisfies*

$$F_{X_i}(x_i) \;=\; F_{\mathbf{X}}(\infty, \ldots, \infty, x_i, \infty, \ldots \infty).$$

It is straightforward to generalize the previous definition to *joint* marginal distributions. For example, the joint marginal distribution of X_i and X_j satisfies $F_{ij}(x_i, x_j) \;=\; F_{\mathbf{X}}(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty, x_j, \infty, \ldots \infty)$. If the joint CDF is *absolutely continuous*, then it has an associated probability density function (PDF) so that

$$F_{\mathbf{X}}(x_1, \ldots, x_n) \;=\; \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n)\, du_1 \ldots du_n.$$

Similar statements also apply to the marginal CDF's. A collection of random variables is independent if the joint CDF (or PDF if it exists) can be factored into the product of the marginal CDFs (or PDFs). If $\mathbf{X_1} = (X_1, \ldots, X_k)^\top$ and $\mathbf{X_2} = (X_{k+1}, \ldots, X_n)^\top$ is a partition of $\mathbf{X}$ then the *conditional* CDF satisfies

$$F_{\mathbf{X_2}|\mathbf{X_1}}(\mathbf{x_2}|\mathbf{x_1}) \;=\; P(\mathbf{X_2} \le \mathbf{x_2} | \mathbf{X_1} = \mathbf{x_1}).$$

If $\mathbf{X}$ has a PDF, $f(\cdot)$, then it satisfies

$$F_{\mathbf{X_2}|\mathbf{X_1}}(\mathbf{x_2}|\mathbf{x_1}) \;=\; \int_{-\infty}^{x_{k+1}} \cdots \int_{-\infty}^{x_n} \frac{f(x_1, \ldots, x_k, u_{k+1}, \ldots, u_n)}{f_{\mathbf{X_1}}(\mathbf{x_1})}\, du_{k+1} \ldots du_n$$

where $f_{\mathbf{X_1}}(\cdot)$ is the joint marginal PDF of $\mathbf{X_1}$. Assuming it exists, the *mean* vector of $\mathbf{X}$ is given by

$$\mathrm{E}[\mathbf{X}] \;:=\; (\mathrm{E}[X_1], \ldots, \mathrm{E}[X_n])^\top$$

whereas, again assuming it exists, the *covariance* matrix of $\mathbf{X}$ satisfies

$$\mathrm{Cov}(\mathbf{X}) \;:=\; \mathbf{\Sigma} \;:=\; \mathrm{E}\left[(\mathbf{X} - \mathrm{E}[\mathbf{X}])\,(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\top\right]$$

[1]We will not study copulas in these notes as well defer this topic until later in the course.

so that the $(i,j)^{th}$ element of Σ is simply the covariance of X_i and X_j. Note that the covariance matrix is symmetric so that $\mathbf{\Sigma}^\top = \mathbf{\Sigma}$, its diagonal elements satisfy $\mathbf{\Sigma}_{i,i} \geq 0$, and it is *positive semi-definite* so that $x^\top \mathbf{\Sigma} x \geq 0$ for all $x \in \mathbf{R}^n$. The *correlation* matrix, $\rho(\mathbf{X})$ has as its $(i,j)^{th}$ element $\rho_{ij} := \mathrm{Corr}(X_i, X_j)$. It is also symmetric, positive semi-definite and has 1's along the diagonal. For any matrix $\mathbf{A} \in \mathbb{R}^{k \times n}$ and vector $\mathbf{a} \in \mathbb{R}^k$ we have

$$\mathrm{E}[\mathbf{AX} + \mathbf{a}] = \mathbf{A}\mathrm{E}[\mathbf{X}] + \mathbf{a} \quad (1)$$

$$\mathrm{Cov}(\mathbf{AX} + \mathbf{a}) = \mathbf{A}\ \mathrm{Cov}(\mathbf{X})\ \mathbf{A}^\top. \quad (2)$$

Finally, the *characteristic function* of $\mathbf{X}$ is given by

$$\phi_{\mathbf{X}}(s) := \mathrm{E}\left[e^{is^\top \mathbf{X}}\right] \quad \text{for} \quad s \in \mathbb{R}^n \quad (3)$$

and, if it exists, the *moment-generating function* (MGF) is given by (3) with s replaced by $-i\, s$.

---

2 The Multivariate Normal Distribution

If the n-dimensional vector $\mathbf{X}$ is multivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ then we write

$$\mathbf{X} \sim \mathsf{MN}_n(\boldsymbol{\mu}, \mathbf{\Sigma}).$$

The standard multivariate normal has $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{I_n}$, the $n \times n$ identity matrix. The PDF of $\mathbf{X}$ is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\, \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (4)$$

where $|\cdot|$ denotes the determinant, and its characteristic function satisfies

$$\phi_{\mathbf{X}}(\mathbf{s}) = \mathrm{E}\left[e^{i\mathbf{s}^\top \mathbf{X}}\right] = e^{i\mathbf{s}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{s}^\top \mathbf{\Sigma} s}. \quad (5)$$

Recall again our partition of $\mathbf{X}$ into $\mathbf{X_1} = (X_1, \ldots, X_k)^\top$ and $\mathbf{X_2} = (X_{k+1}, \ldots, X_n)^\top$. If we extend this notation naturally so that

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{\mathbf{1}} \\ \mu_{\mathbf{2}} \end{pmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{21}} & \mathbf{\Sigma_{22}} \end{pmatrix}.$$

then we obtain the following results regarding the marginal and conditional distributions of $\mathbf{X}$.

Marginal Distribution

The marginal distribution of a multivariate normal random vector is itself multivariate normal. In particular, $\mathbf{X_i} \sim \mathsf{MN}(\mu_{\mathbf{i}}, \mathbf{\Sigma_{ii}})$, for $i = 1, 2$.

Conditional Distribution

Assuming $\mathbf{\Sigma}$ is positive definite, the conditional distribution of a multivariate normal distribution is also a multivariate normal distribution. In particular,

$$\mathbf{X_2} \mid \mathbf{X_1} = \mathbf{x_1} \sim \mathsf{MN}(\mu_{\mathbf{2.1}}, \mathbf{\Sigma_{2.1}})$$

where $\mu_{2.1} = \mu_{\mathbf{2}} + \mathbf{\Sigma_{21}}\ \mathbf{\Sigma_{11}^{-1}}\ (\mathbf{x_1} - \mu_{\mathbf{1}})$ and $\mathbf{\Sigma_{2.1}} = \mathbf{\Sigma_{22}} - \mathbf{\Sigma_{21}}\mathbf{\Sigma_{11}^{-1}}\mathbf{\Sigma_{12}}$.

Linear Combinations

Linear combinations of multivariate normal random vectors remain normally distributed with mean vector and covariance matrix given by (1) and (2), respectively.

Estimation of Multivariate Normal Distributions

The simplest and most common method of estimating a multivariate normal distribution is to take the sample mean vector and sample covariance matrix as our estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, respectively. It is easy to justify this choice since they are the *maximum likelihood* estimators. It is also common to take $n/(n-1)$ times the sample covariance matrix as an estimator of $\mathbf{\Sigma}$ as this estimator is known to be *unbiased*.

Testing Normality and Multivariate Normality

There are many tests that can be employed for testing normality of random variables and vectors. These include standard univariate tests and tests based on *QQplots*, as well *omnibus moment tests* based on whether the skewness and kurtosis of the data are consistent with a multivariate normal distribution. Section $3.1.4$ of *MFE* should be consulted for details on these tests.

2.1 Generating Multivariate Normally Distributed Random Vectors

Suppose we wish to generate $\mathbf{X} = (X_1, \ldots, X_n)$ where $\mathbf{X} \sim \mathsf{MN}_n(\mathbf{0}, \mathbf{\Sigma})$. Note that it is then easy to handle the case where $\mathrm{E}[\mathbf{X}] \neq \mathbf{0}$. Let $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ where the Z_i's are IID $\mathsf{N}(0,1)$ for $i = 1, \ldots, n$. If $\mathbf{C}$ is an $(n \times m)$ matrix then it follows that

$$\mathbf{C}^\top \mathbf{Z} \sim \mathsf{MN}(0, \mathbf{C}^\top \mathbf{C}).$$

Our problem therefore reduces to finding $\mathbf{C}$ such that $\mathbf{C}^\top \mathbf{C} = \mathbf{\Sigma}$. We can use the *Cholesky decomposition* of $\mathbf{\Sigma}$ to find such a matrix, $\mathbf{C}$.

The Cholesky Decomposition of a Symmetric Positive-Definite Matrix

A well known fact from linear algebra is that any symmetric positive-definite matrix, $\mathbf{M}$, may be written as

$$\mathbf{M} = \mathbf{U}^\top \mathbf{D} \mathbf{U}$$

where $\mathbf{U}$ is an upper triangular matrix and $\mathbf{D}$ is a diagonal matrix with positive diagonal elements. Since $\mathbf{\Sigma}$ is symmetric positive-definite, we can therefore write

$$\mathbf{\Sigma} \;=\; \mathbf{U}^\top \mathbf{D} \mathbf{U} \;=\; (\mathbf{U}^\top \sqrt{\mathbf{D}})(\sqrt{\mathbf{D}}\mathbf{U}) \;=\; (\sqrt{\mathbf{D}}\mathbf{U})^\top (\sqrt{\mathbf{D}}\mathbf{U}).$$

The matrix $\mathbf{C} = \sqrt{\mathbf{D}}\mathbf{U}$ therefore satisfies $\mathbf{C}^\top \mathbf{C} = \mathbf{\Sigma}$. It is called the Cholesky Decomposition of $\mathbf{\Sigma}$.

The Cholesky Decomposition in *Matlab* and *R*

It is easy to compute the Cholesky decomposition of a symmetric positive-definite matrix in *Matlab* and *R* using the *chol* command and so it is also easy to simulate multivariate normal random vectors. As before, let $\mathbf{\Sigma}$ be an $(n \times n)$ variance-covariance matrix and let $\mathbf{C}$ be its Cholesky decomposition. If $\mathbf{X} \sim \mathsf{MN}(\mathbf{0}, \mathbf{\Sigma})$ then we can generate random samples of $\mathbf{X}$ in Matlab as follows:

Sample Matlab Code

```
>> Sigma = [1.0 0.5 0.5;
            0.5 2.0 0.3;
            0.5 0.3 1.5];
>> C = chol(Sigma);
>> Z = randn(3,1000000);
>> X = C'*Z;
>> cov(X')

ans =
    0.9972    0.4969    0.4988
    0.4969    1.9999    0.2998
    0.4988    0.2998    1.4971
```

We must be very careful[2] in *Matlab*[3] and *R* to pre-multiply Z by $\mathbf{C}^\top$ and not $\mathbf{C}$. We have the following algorithm for generating multivariate random vectors, $\mathbf{X}$.

Generating Correlated Normal Random Variables

> **generate** $\mathbf{Z} \sim \mathsf{MN}(\mathbf{0}, \mathbf{I})$
> /∗ Now compute the Cholesky Decomposition ∗/
> **compute** $\mathbf{C}$ such that $\mathbf{C}^\top \mathbf{C} = \mathbf{\Sigma}$
> **set** $\mathbf{X} = \mathbf{C}^\top \mathbf{Z}$

---

3 Normal-Mixture Models

Normal-mixture models are a class of models generated by introducing randomness into the covariance matrix and / or the mean vector. Following the development of *MFE* we have the following definition of a *normal variance mixture*:

Definition 3 *The random vector* $\mathbf{X}$ *has a normal variance mixture if*

$$\mathbf{X} \sim \boldsymbol{\mu} \ + \ \sqrt{W}\ \mathbf{A}\mathbf{Z}$$

where

(i) $\mathbf{Z} \sim \mathit{MN}_k(\mathbf{0}, \mathbf{I}_k)$

(ii) $W \geq 0$ *is a scalar random variable independent of* $\mathbf{Z}$ *and*

(iii) $\mathbf{A} \in \mathbb{R}^{n \times k}$ *and* $\boldsymbol{\mu} \in \mathbb{R}^n$ *are a matrix and vector of constants, respectively.*

Note that if we condition on W, then $\mathbf{X}$ is multivariate normally distributed. This observation also leads to an obvious simulation algorithm for generating samples of $\mathbf{X}$: first simulate a value of W and then simulate $\mathbf{X}$ conditional on the generated value of W. We are typically interested in the case where $\text{rank}(\mathbf{A}) = n \leq k$ and $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^\top$ is a full-rank positive definite matrix. In this case we obtain a non-singular normal variance mixture. Assuming W is integrable[4], we immediately see that

$$\mathrm{E}[\mathbf{X}] \ = \ \boldsymbol{\mu} \quad \text{and} \quad \mathrm{Cov}(\mathbf{X}) \ = \ \mathrm{E}[W]\ \mathbf{\Sigma}.$$

---

[2] We must also be careful that $\mathbf{\Sigma}$ is indeed a genuine variance-covariance matrix.

[3] Unfortunately, some languages take $\mathbf{C}^\top$ to be the Cholesky Decomposition rather $\mathbf{C}$. You must therefore always be aware of exactly what convention your programming language / package is using.

[4] That is, W is integrable if $\mathrm{E}[W] < \infty$.

We refer to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as the *location* vector and *dispersion* matrix of the distribution. It is also clear that the correlation matrices of $\mathbf{X}$ and $\mathbf{AZ}$ are the same as long as W is integrable. This means that if $\mathbf{A} = \mathbf{I}_k$ then the components of $\mathbf{X}$ are uncorrelated, though they are not in general independent. The following result[5] emphasizes this point.

Lemma 1 *Let* $\mathbf{X} = (X_1, X_2)$ *have a normal mixture distribution with* $\mathbf{A} = \mathbf{I_2}$*,* $\boldsymbol{\mu} = \mathbf{0}$ *and* $\mathrm{E}[W] < \infty$ *so that* $\mathrm{Cov}(X_1, X_2) = 0$*. Then* X_1 *and* X_2 *are independent if and only if* W *is a constant with probability* 1*. (If* W *is constant then* X_1 *and* X_2 *are IID N*$(0, W)$*.)*

Proof: If W is a constant then it immediately follows from the independence of Z_1 and Z_2 that X_1 and X_2 are also independent. Suppose now that X_1 and X_2 are independent. Note that

$$\begin{aligned} \mathrm{E}[|X_1|\,|X_2|] &= \mathrm{E}[W\,|Z_1|\,|Z_2|] \;=\; \mathrm{E}[W]\,\mathrm{E}[|Z_1|\,|Z_2|] \\ &\geq \left(\mathrm{E}[\sqrt{W}]\right)^2\,\mathrm{E}[|Z_1|\,|Z_2|] \;=\; \mathrm{E}[|X_1|]\,\mathrm{E}[|X_2|] \end{aligned}$$

with equality only if W is a constant. But the independence of X_1 and X_2 implies that we must have equality and so W is indeed constant almost surely. ■

Example 1 (The Multivariate Two-Point Normal Mixture Model)

Perhaps the simplest example of the normal-variance mixture is obtained when W is a discrete random variable. If W is binary and takes on two values, w_1 and w_2 with probabilities p and $1-p$, respectively, then we obtain the two-point normal mixture model. We can create a two *regime* model by setting w_2 large relative to w_1 and choosing p large. Then $W = w_1$ can correspond to an *ordinary* regime whereas $W = w_2$ corresponds to a *stress* regime. ▮

Example 2 (The Multivariate t Distribution)

The multivariate t distribution with ν degrees-of-freedom (dof) is obtained when we take W to have an *inverse gamma* distribution or equivalently, if $\nu/W \sim \chi^2_\nu$ and this is the more familiar description of the t distribution. We write $X \sim t_n(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and note that $\mathrm{Cov}(\mathbf{X}) \;=\; \nu\boldsymbol{\Sigma}/(\nu-2)$ but this is only defined when $\nu > 2$. As we can easily simulate chi-squared random variables, it is clearly also easy to simulate multivariate t random vectors. The multivariate t distribution plays an important role in risk management as it often provides a very good fit to asset return distributions.

▮

We can easily calculate the characteristic function of a normal variance mixture. Using (5), we obtain

$$\begin{aligned} \phi_{\mathbf{X}}(s) \;=\; \mathrm{E}\left[e^{is^\top \mathbf{X}}\right] &= \mathrm{E}\left[\mathrm{E}\left[e^{is^\top \mathbf{X}}\,|W\right]\right] \\ &= \mathrm{E}\left[e^{is^\top \boldsymbol{\mu} - \frac{1}{2}Ws^\top \boldsymbol{\Sigma} s}\right] \\ &= e^{is^\top \boldsymbol{\mu}}\,\widehat{W}\left(\frac{1}{2}s^\top \boldsymbol{\Sigma} s\right) \end{aligned} \tag{6}$$

where $\widehat{W}(\cdot)$ is the *Laplace transform* of W. As a result, we sometimes use the notation $\mathbf{X} \sim \mathrm{M}_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{W}\right)$ for normal variance mixtures. We have the following proposition[6] showing that affine transformations of normal variance mixtures remain normal variance mixtures.

Proposition 1 *If* $\mathbf{X} \sim M_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{W}\right)$ *and* $\mathbf{Y} = \mathbf{BX} + \mathbf{b}$ *for* $\mathbf{B} \in \mathbb{R}^{k\times n}$ *and* $\mathbf{b} \in \mathbb{R}^k$ *then* $\mathbf{Y} \sim M_k\left(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top, \widehat{W}\right)$.

The proof is straightforward using (6). This result is useful in the following setting: suppose a collection of risk factors has a normal variance mixture distribution. Then the usual linear approximation to the loss distribution will also have a (1-dimensional) normal variance mixture distribution.

[5] This result is Lemma 3.5 in *MFE*.
[6] This is Proposition 3.9 in *MFE*.

Normal Mean-Variance Mixtures

We could also define normal mixture distributions where the mean vector, $\boldsymbol{\mu}$, is also a function of the scalar random variable, W, so that $\boldsymbol{\mu} = m(W)$, say. We would still obtain that $\mathbf{X}$ is multivariate normal, conditional on W. An important class of normal mean-variance mixtures is given by the so-called *generalized hyperbolic distributions*. They, as well as the normal variance mixtures, are closed under addition, are easy to simulate and can be calibrated using standard statistical techniques. We will not study these normal mean-variance mixtures in this course but *MFE* should be consulted if further details are required.

---

4 Spherical and Elliptical Distributions

We now provide a very brief introduction to spherical and elliptical distributions. Spherical distributions generalize *uncorrelated* multivariate normal and t distributions. In addition to having uncorrelated[7] components, they have identical and symmetric marginal distributions. The elliptical distributions can be obtained as *affine* transformations of the spherical distributions. They include, for example, general multivariate normal and t distributions as well as all spherical distributions. Elliptical distributions are an important class of distributions: they inherit much of the normal distribution's tractability yet they are sufficiently rich to include empirically plausible distributions such as, for example, many heavy-tailed distributions.

Our introduction to these distributions will be very brief for several reasons. First, there is a large body of literature associated with spherical and elliptical distributions and we simply don't have time to study this literature in any great detail. Second, we are already familiar with the (heavy-tailed) multivariate t distribution which often provides an excellent fit to financial return data. Hence, the need to study other multivariate empirically plausible distributions is not quite so pressing. And finally, it is often the case that our ultimate goal is to study the loss[8] distribution. This is a univariate distribution, however, and it is often more convenient to take a *reduced form* approach and to directly estimate this distribution rather than estimating the multivariate distribution of the underlying risk factors. We may see an example of this approach later in the course when we use *extreme value theory* (EVT) to estimate the VaR of a portfolio.

4.1 Spherical Distributions

We first define spherical distributions. In order to do so, recall that a linear transformation $\mathbf{U} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$ is *orthogonal* if $\mathbf{U}\mathbf{U}^\top = \mathbf{U}^\top\mathbf{U} = \mathbf{I}_n$.

Definition 4 *A random vector $\mathbf{X} = (X_1, \ldots, X_n)$ has a spherical distribution if*

$$\mathbf{U}\mathbf{X} \sim \mathbf{X} \tag{7}$$

for every orthogonal linear transformation, $\mathbf{U} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}}$.

In particular, (7) implies that the distribution of $\mathbf{X}$ is invariant under rotations. A better understanding of spherical distributions may be obtained from the following[9] theorem.

Theorem 2 *The following are equivalent:*

1. *$\mathbf{X}$ is spherical.*

2. *There exists a function $\psi(\cdot)$ such that for all $\mathbf{s} \in \mathbb{R}^n$,*

$$\phi_{\mathbf{X}}(\mathbf{s}) = \psi(\mathbf{s}^\top\mathbf{s}) = \psi(s_1^2 + \cdots + s_n^2). \tag{8}$$

[7]The only spherical distribution that has *independent* components is the standard multivariate normal distribution.

[8]Assuming of course that we have already decided it is a good idea to estimate the loss distribution. This will not be the case when we have too little historical data, or the market is too "crowded" or otherwise different to how it behaved historically.

[9]This is Theorem 3.19 in *MFE*.

3. *For all* $\mathbf{a} \in \mathbb{R}^n$

$$\mathbf{a}^\top X \;\sim\; ||\mathbf{a}||\, X_1$$

where $||\mathbf{a}||^2 = \mathbf{a}^\top \mathbf{a} = a_1^2 + \cdots + a_n^2$.

Proof: The proof is straightforward but see Section 3.3 of *MFE* for details. ■

Part (2) of Theorem 2 shows that the characteristic function of a spherical distribution is completely determined by a function of a scalar variable. This function, $\psi(\cdot)$, is known as the *generator* of the distribution and it is common to write $\mathbf{X} \sim \mathsf{S}_n(\psi)$.

Example 3 (Multivariate Normal)

We know the characteristic function of the standard multivariate normal, i.e. $\mathbf{X} \sim \mathsf{MN}_n(\mathbf{0}, \mathbf{I}_n)$, satisfies

$$\phi_{\mathbf{X}}(\mathbf{s}) \;=\; e^{-\frac{1}{2}\mathbf{s}^\top \mathbf{s}}$$

and so it follows from (8) that $\mathbf{X}$ is spherical with generator $\psi(s) = \exp(-\frac{1}{2}s)$. ■

Example 4 (Normal Variance Mixtures)

Suppose $\mathbf{X}$ has a standardized, uncorrelated normal variance mixture so that $\mathbf{X} \sim \mathsf{M}_n\left(\mathbf{0}, \mathbf{I}_n, \widehat{W}\right)$. Then (6) and part (2) of Theorem 2 imply that $\mathbf{X}$ is spherical with $\psi(s) = \widehat{W}(s/2)$. ■

It is worth noting that there are also spherical distributions that are *not* normal variance mixture distributions. Another important and insightful result regarding spherical distributions is given in the following theorem. A proof may be found in Section 3.4 of *MFE*.

Theorem 3 *The random vector* $\mathbf{X} = (X_1, \ldots, X_n)$ *has a spherical distribution if and only if it has the representation*

$$\mathbf{X} \;\sim\; R\,\mathbf{S}$$

where $\mathbf{S}$ *is uniformly distributed on the unit sphere* $\mathcal{S}^{n-1} := \{\mathbf{s} \in \mathbb{R}^n \;:\; \mathbf{s}^\top \mathbf{s} = 1\}$ *and* $R \geq 0$ *is a random variable independent of* $\mathbf{S}$.

4.2 Elliptical Distributions

Definition 5 *The random vector* $\mathbf{X} = (X_1, \ldots X_n)$ *has an elliptical distribution if*

$$\mathbf{X} \;\sim\; \boldsymbol{\mu} \;+\; \mathbf{A}\,\mathbf{Y}$$

where $\mathbf{Y} \sim S_k(\psi)$ *and* $\mathbf{A} \in \mathbb{R}^{n \times k}$ *and* $\boldsymbol{\mu} \in \mathbb{R}^n$ *are a matrix and vector of constants, respectively.*

We therefore see that elliptical distributions are obtained via multivariate affine transformations of spherical distributions. It is easy to calculate the characteristic function of an elliptical distribution. We obtain

$$\begin{aligned}
\phi_{\mathbf{X}}(\mathbf{s}) &= \mathrm{E}\left[e^{i\mathbf{s}^\top(\boldsymbol{\mu} \,+\, \mathbf{A}\,\mathbf{Y})}\right] \\
&= e^{i\mathbf{s}^\top \boldsymbol{\mu}}\,\mathrm{E}\left[e^{i(\mathbf{A}^\top \mathbf{s})^\top \mathbf{Y}}\right] \\
&= e^{i\mathbf{s}^\top \boldsymbol{\mu}}\,\psi\left(\mathbf{s}^\top \boldsymbol{\Sigma} \mathbf{s}\right)
\end{aligned}$$

where as before $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^\top$. It is common to write $\mathbf{X} \sim \mathsf{E}_n(\boldsymbol{\mu}, \mathbf{\Sigma}, \psi)$ and we refer to $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ as the *location* vector and *dispersion* matrix, respectively. It is worth mentioning, however, that $\mathbf{\Sigma}$ and ψ are only uniquely determined up to a positive constant.

As mentioned earlier, the elliptical distributions form a rich class of distributions, including both heavy- and light-tailed distributions. Their importance is due to this richness as well as to their general tractability. For example, elliptical distributions are closed under linear operations. Moreover, the marginal and conditional distributions of elliptical distributions are elliptical distributions. They may be estimated using maximum likelihood methods such as the *EM* algorithm or other iterative techniques. Additional information and references may be found in *MFE* but we note that software applications such as `R` or `Matlab` will often fit these distributions for you.