Multivariate Distributions and Dimension Reduction Techniques

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. We will also study factor models and principal components analysis (PCA) as examples of *dimension reduction* techniques. We will generally follow the notation of Chapter 3 of *Quantitative Risk Management* by *McNeil, Frey and Embrechts*[2]. This chapter contains detailed discussions of these topics and should be consulted if further details are required.

1 Multivariate Distributions

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

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

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 $\mathbf{X_1} = (X_1, \ldots, X_k)^T$ and $\mathbf{X_2} = (X_{k+1}, \ldots, X_n)^T$ is a partition of $\mathbf{X}$ then the *conditional* CDF satisfies

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

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

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

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

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

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

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 $\mathbf{\Sigma}^T = \mathbf{\Sigma}$, its diagonal elements satisfy $\mathbf{\Sigma}_{i,i} \geq 0$, and it is *positive semi-definite* so that $x^T\mathbf{\Sigma}x \geq 0$ 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

$$\mathrm{E}\left[\mathbf{AX}+\mathbf{a}\right] \;=\; \mathbf{A}\mathrm{E}\left[\mathbf{X}\right] \,+\, \mathbf{a} \tag{1}$$

$$\mathrm{Cov}(\mathbf{AX}+\mathbf{a}) \;=\; \mathbf{A}\;\mathrm{Cov}(\mathbf{X})\;\mathbf{A}^T. \tag{2}$$

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

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

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

1.1 The Multivariate Normal Distribution

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

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

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

$$f(\mathbf{x}) \;=\; \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}}\; e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\;\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)} \tag{4}$$

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

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

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

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

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 $\mathbf{\Sigma_{2.1}} = \mathbf{\Sigma_{22}} - \mathbf{\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}$.

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 μ 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.

1.2 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 $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$ where $Z_i \sim \mathsf{N}(0,1)$ and IID for $i = 1, \ldots, n$. If $\mathbf{C}$ is an $(n \times m)$ matrix then it follows that

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

Our problem therefore reduces to finding $\mathbf{C}$ such that $\mathbf{C}^T\mathbf{C} = \mathbf{\Sigma}$. Finding such a matrix, $\mathbf{C}$, requires us to compute the *Cholesky decomposition* of $\mathbf{\Sigma}$.

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

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

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

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

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

The Cholesky Decomposition in Matlab

It is easy to compute the Cholesky decomposition of a symmetric positive-definite matrix in Matlab 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[3] in Matlab[4] to pre-multiply Z by $\mathbf{C}^T$ and not $\mathbf{C}$. We have the following algorithm for generating multivariate random vectors, $\mathbf{X}$.

Generating Correlated Normal Random Variables

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

1.3 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*

$$\mathbf{X} \sim \mu + \sqrt{W}\,\mathbf{AZ}$$

where

(i) $\mathbf{Z} \sim MN_k(\mathbf{0}, \mathbf{I_k})$

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

(iii) $\mathbf{A} \in \mathbb{R}^{n \times k}$ *and* $\mu \in \mathbb{R}^n$ *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 when $\text{rank}(A) = n \leq k$ and $\mathbf{\Sigma}$ is a full-rank positive definite matrix. In this case we obtain a non-singular normal variance mixture. Assuming W is integrable[5], we immediately see that

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

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

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

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

where $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$. We refer to μ and $\mathbf{\Sigma}$ as the *location* vector and *dispersion* matrix of the distribution. It is also clear that the correlation matrices of $\mathbf{X}$ and $\mathbf{Z}$ are the same as long as W is integrable. This means that if $\mathbf{Z}$ is a vector of independent normal random variables, then the components of $\mathbf{X}$ are uncorrelated, though they are not in general independent. The following result[6] emphasizes this point.

Lemma 1 *Let $\mathbf{X} = (X_1, X_2)$ have a normal mixture distribution with $\mathbf{A} = \mathbf{I_2}$, $\mu = \mathbf{0}$ 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$(0, 1)$.)*

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 $X \sim t_n(\nu, \mu, \mathbf{\Sigma})$ and note that $\mathrm{Cov}(\mathbf{X}) = \nu/(\nu - 2)\mathbf{\Sigma}$ 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.

Sample R Code Using QRMlib for Visualizing, Simulating and Estimating Multivariate t Data

```
> BiDensPlot(func=dmt,xpts=c(-4,4),ypts=c(-4,4),mu=c(0,0),Sigma=equicorr(2,-0.7),nu=4)
> tdata <- rmt(2000,df=4,rho=0.7,d=3)
> fit.mst(tdata)
```

▮

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

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

where $\widehat{W}(\cdot)$ is the *Laplace transform* of W. As a result, we sometimes use the notation $\mathbf{X} \sim \mathrm{M}_n\left(\mu, \mathbf{\Sigma}, \widehat{W}\right)$ for normal variance mixtures. We have the following proposition[7] showing that affine transformations of normal variance mixtures remain normal variance mixtures.

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

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

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.

Normal Mean-Variance Mixtures

We could also define normal mixture distributions where the mean vector, μ, is also a function of the scalar random variable, W, so that $\mu = m(W)$, 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.

1.4 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[8] 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 is 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[9] 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 will see an example of this approach later in the course when we use *extreme value theory* (EVT) to estimate the VaR of a portfolio.

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 $\mathbf{UU}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_n$.

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

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

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[10] theorem.

Theorem 2 *The following are equivalent:*

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

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

[9] 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.

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

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

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

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

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

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

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

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

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* $\mathcal{S}^{n-1} := \{\mathbf{s} \in \mathbb{R}^n \;:\; \mathbf{s}^T\mathbf{s} = 1\}$ *and* $R \geq 0$ *is a random variable independent of* $\mathbf{S}$.

Elliptical Distributions

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

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

where $\mathbf{Y} \sim S_k(\psi)$ *and* $\mathbf{A} \in \mathbb{R}^{n\times k}$ *and* $\mu \in \mathbb{R}^n$ *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

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

where as before $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$. It is common to write $\mathbf{X} \sim \mathsf{E}_n(\mu, \mathbf{\Sigma}, \psi)$ and we refer to μ 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 as they do 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. For the reasons stated earlier, however, we will not study them any further in this course. Additional information and references may be found in *MFE*.

---

2 Dimension Reduction Techniques

We now consider some dimension reduction techniques that are often useful for identifying the key components that drive the risk or randomness in a system. We begin with principal components analysis (*PCA*).

2.1 Principal Components Analysis

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ denote an n-dimensional random vector with variance-covariance matrix, Σ. In the context of risk management, we take this vector to represent the (normalized) changes, over some appropriately chosen time horizon, of an n-dimensional vector of risk factors. These risk factors could represent security price returns, returns on futures contracts of varying maturities, or changes in spot interest rates, again of varying maturities. The goal of PCA is to construct linear combinations

$$P_i \;=\; \sum_{j=1}^{n} w_{ij}\, Y_j \quad \text{for} \quad i = 1, \ldots, n$$

in such a way that:

(1) the P_i's are *orthogonal* so that $\mathrm{E}[P_i\, P_j] = 0$ for $i \neq j$ and

(2) the P_i's are ordered so that: (i) P_1 explains the largest percentage of the total variability in the system and (ii) each P_i explains the largest percentage of the total variability in the system that has *not* already been explained by $P_1, \ldots, P_{i-1}$.

In practice it is common to apply PCA to *normalized*[11] random variables that satisfy $\mathrm{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = 1$. This is achieved by subtracting the means from the original random variables and dividing by their standard deviations. This is done to ensure that no one component of $\mathbf{Y}$ can influence the analysis by virtue of that component's measurement units. We will therefore assume that the Y_i's have already been normalized. The key tool of PCA is the *Spectral Decomposition* of linear algebra which states that any symmetric matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be written as

$$\mathbf{A} \;=\; \mathbf{\Gamma}\, \mathbf{\Delta}\, \mathbf{\Gamma}^T \tag{9}$$

where:

(i) $\mathbf{\Delta}$ is a diagonal matrix, $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, of the *eigen-values* of $\mathbf{A}$ which, without loss of generality, are ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and

(ii) $\mathbf{\Gamma}$ is an orthogonal matrix with the i^{th} column of $\mathbf{\Gamma}$ containing the i^{th} standardized[12] eigen-vector, γ_i, of $\mathbf{A}$. The orthogonality of $\mathbf{A}$ implies $\mathbf{\Gamma}\, \mathbf{\Gamma}^T = \mathbf{\Gamma}^T\, \mathbf{\Gamma} = \mathbf{I}_n$.

[11] Working with normalized random variables is equivalent to working with the correlation matrix of the un-normalized variables.

[12] By standardized we mean $\gamma_i^T \gamma_i = 1$.

Since $\mathbf{\Sigma}$ is symmetric we can take $\mathbf{A} = \mathbf{\Sigma}$ in (9) and the positive semi-definiteness of $\mathbf{\Sigma}$ implies $\lambda_i \geq 0$ for all $i = 1, \ldots, n$. The *principal components* of $\mathbf{Y}$ are then given by $\mathbf{P} = (P_1, \ldots, P_n)$ satisfying

$$\mathbf{P} \;=\; \mathbf{\Gamma}^T \, \mathbf{Y}. \tag{10}$$

Note that:

(a) $\mathrm{E}[P] \;=\; \mathbf{0}$ since $\mathrm{E}[Y] \;=\; \mathbf{0}$ and

(b) $\mathrm{Cov}(\mathbf{P}) \;=\; \mathbf{\Gamma}^T \, \mathbf{\Sigma} \, \mathbf{\Gamma} \;=\; \mathbf{\Gamma}^T \, (\mathbf{\Gamma} \, \mathbf{\Delta} \, \mathbf{\Gamma}^T) \, \mathbf{\Gamma} \;=\; \Delta$ so that the components of $\mathbf{P}$ are uncorrelated and $\mathrm{Var}(P_i) = \lambda_i$. This is consistent with (1) above.

The matrix $\mathbf{\Gamma}^T$ is called the matrix of factor *loadings*. Note that we can invert (10) to obtain

$$\mathbf{Y} \;=\; \mathbf{\Gamma} \, \mathbf{P}. \tag{11}$$

We can measure the ability of the first few principal components to explain the total variability in the system. We see from (b) that

$$\sum_{i=1}^{n} \mathrm{Var}(P_i) \;=\; \sum_{i=1}^{n} \lambda_i \;=\; \mathrm{trace}(\Sigma) \;=\; \sum_{i=1}^{n} \mathrm{Var}(Y_i) \tag{12}$$

where we have used the fact that the *trace* of a matrix, i.e. the sum of its diagonal elements, is also equal to the sum of its eigen-values. If we take $\sum_{i=1}^{n} \mathrm{Var}(P_i) = \sum_{i=1}^{n} \mathrm{Var}(Y_i)$ to measure the total variability, then by (13) we may interpret the ratio

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$$

as measuring the percentage of the total variability that is explained by the first k principal components. This is consistent with (2) above since the λ_i's are non-increasing. In particular, it is possible to show that the first principal component, $P_1 = \gamma_1^T \mathbf{Y}$, satisfies

$$\mathrm{Var}(\gamma_1^T \mathbf{Y}) \;=\; \max \left\{ \mathrm{Var}(\mathbf{a}^T \mathbf{Y}) \;:\; \mathbf{a}^T \mathbf{a} \;=\; 1 \right\}.$$

Moreover, it is also possible to show that each successive principal component, $P_i = \gamma_i^T \mathbf{Y}$, satisfies the same optimization problem but with the added constraint that it be orthogonal, i.e. uncorrelated, to $P_1, \ldots, P_{i-1}$.

In financial applications, it is often the case that just two or three principal components are sufficient to explain anywhere from 60% to 95% or more of the total variability. Moreover, it is often possible to interpret the first two or three components. For example, if $\mathbf{Y}$ represents (normalized) changes in the spot interest rate for n different maturities, then the first principal component can usually be interpreted as an (approximate) parallel shift in the spot rate curve, whereas the second component represents a flattening or steepening of the curve. In equity applications, the first component often represents a systematic market factor that impacts all of the stocks whereas the second (and possibly other) components may be identified with industry specific factors.

Empirical PCA

In practice we do not know the true variance-covariance matrix but it may be estimated using historical data. Suppose then that we have the multivariate observations, $\mathbf{X}_1, \ldots \mathbf{X}_m$, where $\mathbf{X}_t = (X_{t1}, \ldots, X_{tn})^T$, represents the date t sample observation. It is important that these observations represent a *stationary* time series such as asset returns or yield changes. However $\mathbf{X}_t$ should not represent a vector of price levels, for example, as the latter generally constitute *non-stationary* time series. If μ_j and σ_j for $j = 1, \ldots, n$, are the sample mean and standard deviation, respectively, of $\{X_{tj} \;:\; t = 1, \ldots, m\}$, then we can normalize the data by setting

$$Y_{tj} \;=\; \frac{X_{tj} - \mu_j}{\sigma_j} \quad \text{for} \;\; t = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n.$$

Let $\mathbf{\Sigma}$ be the sample[13] variance-covariance matrix and let us assume that the $\mathbf{Y}_i$'s come from some stationary time-series. Using the same notation as before, we see that the sample covariance matrix of $\mathbf{Y}$ is then given by

$$\mathbf{\Sigma} \;=\; \frac{1}{m}\,\sum_{t=1}^{m}\mathbf{Y}_t\,\mathbf{Y}_t^T.$$

The principal components are then computed using this covariance matrix. From (11), we see that the original data is obtained from the principal components as

$$\begin{aligned}\mathbf{X}_t &= \text{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{Y}_t \;+\; \mu \\ &= \text{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{\Gamma}\,\mathbf{P}_t \;+\; \mu \end{aligned} \tag{13}$$

where $\mathbf{P}_t \;:=\; (P_{t1},\ldots,P_{tn}) \;=\; \mathbf{\Gamma}^T\mathbf{Y}_t$ is the t^{th} sample principal component vector.

Applications of PCA in Finance

There are many applications of PCA in finance and risk management. They include:

1. **Scenario Generation**

   It is easy to generate scenarios using PCA. Suppose today is date t and we want to generate scenarios over the period $[t, t+1]$. Then (13) evaluated at date $t+1$ states

   $$\mathbf{X}_{t+1} \;=\; \text{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{\Gamma}\,\mathbf{P}_{t+1} \;+\; \mu.$$

   We can then apply stresses to the first few principal components, either singly or jointly, to generate loss scenarios. Moreover, we know that $\text{Var}(P_i) = \lambda_i$ and so we can easily control the severity of the stresses.

2. **Building Factor Models**

   If we believe the first k principal components explain a sufficiently large amount of the total variability then we may partition the $n \times n$ matrix $\mathbf{\Gamma}$ according to $\mathbf{\Gamma} \;=\; [\mathbf{\Gamma}_1\;\mathbf{\Gamma}_2]$ where $\mathbf{\Gamma}_1$ is $n \times k$ and $\mathbf{\Gamma}_2$ is $n \times (n-k)$. Similarly we can write $\mathbf{P}_t = [\mathbf{P}_t^{(1)}\;\mathbf{P}_t^{(2)}]^T$ where $\mathbf{P}_t^{(1)}$ is $k \times 1$ and $\mathbf{P}_t^{(2)}$ is $(n-k) \times 1$. We may then use (13) to write

   $$\mathbf{X}_{t+1} \;=\; \mu \;+\; \text{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{\Gamma}_1\,\mathbf{P}_{t+1}^{(1)} \;+\; \epsilon_{t+1} \tag{14}$$

   where $\epsilon_{t+1} \;:=\; \text{diag}(\sigma_1,\ldots,\sigma_n)\,\mathbf{\Gamma}_2\,\mathbf{P}_{t+1}^{(2)}$ now represents an *error* term. We can interpret (14) as a k-factor model for the changes in risk factors, $\mathbf{X}_{t+1}$. Note, however, that the components of ϵ_{t+1} are not independent which would be the case in a typical factor model.

3. **Estimating VaR and CVaR**

   We can use the model (14) and Monte-Carlo to simulate portfolio returns. This could be done by: (i) ignoring the error term, ϵ_{t+1}, which we know is small relative to the uncertainty in $\mathbf{P}_{t+1}^{(1)}$ and (ii) estimating the joint distribution of the first k principal components. Since they are uncorrelated by construction and we know their variances we could, for example, assume $\mathbf{P}_{t+1}^{(1)} \sim \text{MN}_k(\mathbf{0},\,\text{diag}(\lambda_1,\ldots,\lambda_k))$. Otherwise, we could use standard statistical techniques to estimate the distribution of $\mathbf{P}_{t+1}^{(1)}$.

[13] Usually we would write $\widehat{\mathbf{\Sigma}}$ for a sample covariance but we will stick with $\mathbf{\Sigma}$ here.

4. **Portfolio Immunization**

It is also possible to hedge or immunize a portfolio against moves in the principal components. For example, suppose we wish to hedge the value of a portfolio against movements in the first k principal components. Let V_t be the time t value of the portfolio and assume that our hedge will consist of positions, ϕ_i, in the securities with time t prices, S_{ti}, for $i = 1, \ldots, k$. As usual let $Z_{(t+1)j}$ be the date $t+1$ level of the j^{th} risk factor so that $\Delta Z_{(t+1)j} = X_{(t+1)j}$. If the change in value of the *hedged* portfolio between dates t and $t+1$ is denoted by ΔV^*_{t+1}, then we have

$$\begin{aligned}
\Delta V^*_{t+1} &\approx \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_i \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \Delta Z_{(t+1)j} \\
&= \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_i \frac{\partial S_{ti}}{\partial Z_{tj}} \right) X_{(t+1)j} \qquad (15) \\
&\approx \sum_{j=1}^{n} \left(\left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_i \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \left(\mu_j + \sigma_j \sum_{l=1}^{k} \mathbf{\Gamma_{jl}} P_l \right) \right) \\
&= \sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_i \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \mu_j \qquad (16) \\
&\quad + \sum_{l=1}^{k} \left(\sum_{j=1}^{n} \left(\frac{\partial V_t}{\partial Z_{tj}} + \sum_{i=1}^{k} \phi_i \frac{\partial S_{ti}}{\partial Z_{tj}} \right) \sigma_j \mathbf{\Gamma_{jl}} \right) P_l \qquad (17)
\end{aligned}$$

where, ignoring ϵ, we used the factor model representation in (14) in going from (15) to (16). We can now use (17) to hedge the risk associated with the first k principal components: we simply solve for the ϕ_i's so that the coefficients of the P_l's in (17) are zero. This is a system of k linear equations in k unknowns and so it is easily solved. If we include an additional hedging asset then we could also, for example, ensure that the total value of the hedged portfolio is equal to the value of the original un-hedged portfolio.

2.2 Factor Models

Factor models play an important[14] role in finance, particularly in the equity space where they are often used to build low-dimensional models of stock returns. In this context they are often used for both portfolio construction and portfolio hedging. Our discussion is very brief[15] and We begin with the definition[16] of a k-factor model:

Definition 6 *We say the random vector* $\mathbf{X} = (X_1, \ldots, X_n)^T$ *follows a* **linear k-factor model** *if it satisfies*

$$\mathbf{X} = \mathbf{a} + \mathbf{B}\,\mathbf{F} + \epsilon \qquad (18)$$

where

(i) $\mathbf{F} = (F_1, \ldots, F_k)^T$ *is a random vector of* common factors *with* $k < n$ *and with a positive-definite covariance matrix;*

(ii) $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ *is a random vector of idiosyncratic error terms which are uncorrelated and have mean zero;*

(iii) $\mathbf{B}$ *is an* $n \times k$ *constant matrix of* factor loadings, *and* $\mathbf{a}$ *is an* $n \times 1$ *vector of constants;*

(iv) $\mathrm{Cov}(F_i, \epsilon_j) = 0$ *for all* i, j.

[14] There are many vendors of factor models in the financial services industry. It is arguable as to how much value these models actually provide when it comes to portfolio construction.

[15] See Section 3.4 of *MFE* for further details and in particular, references to more complete sources.

[16] This is Definition 3.3 in *MFE*.

If $\mathbf{X}$ is multivariate normally distributed and follows (18) then it is possible to find a version of the model where $\mathbf{F}$ and ϵ are also multivariate normally distributed. In this case the error terms, ϵ, are independent. If $\mathbf{\Omega}$ is the covariance matrix of $\mathbf{F}$ then the covariance matrix, $\mathbf{\Sigma}$, of $\mathbf{X}$ must satisfy (why?)

$$\mathbf{\Sigma} \;=\; \mathbf{B}\,\mathbf{\Omega}\,\mathbf{B}^T \;+\; \mathbf{\Upsilon}$$

where $\mathbf{\Upsilon}$ is a diagonal matrix of the variances of ϵ.

Exercise 1 *Show that if (18) holds then there is also a representation*

$$\mathbf{X} \;=\; \mu \;+\; \mathbf{B}^*\,\mathbf{F}^* \;+\; \epsilon \tag{19}$$

where $\mathrm{E}[\mathbf{X}] = \mu$ *and* $\mathrm{Cov}(\mathbf{F}^*) = \mathbf{I}_k$ *so that* $\mathbf{\Sigma} \;=\; \mathbf{B}^*\,(\mathbf{B}^*)^T \;+\; \mathbf{\Upsilon}$.

Example 5 (Factor Models Based on Principal Components)

The factor model of (14) may be interpreted as a k-factor model with $\mathbf{F} = \mathbf{P}^{(1)}$ and $\mathbf{B} = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)\,\mathbf{\Gamma}_1$. Note that as constructed, the covariance of the error term, ϵ, in (14) is not diagonal and so it does not satisfy part (ii) of our definition above. Nonetheless, it is quite common to construct factor models in this manner and to then make the assumption that ϵ is indeed a vector of uncorrelated error terms. ∎

Calibration Approaches

There are two different types of factor models and the calibration approach depends on the type:

1. **Observable Factor Models** are models where the factors have been identified in advance and are observable. The factors in these models typically have a fundamental economic interpretation. A classic example would be a 1-factor model where the market index plays the role of the single factor. Other potential factors include macro-economic and other financial variables. These models are usually calibrated and tested for goodness-of-fit using *multivariate*[17] regression techniques.

2. **Latent Factor Models** are models where the factors have *not* been identified in advance. They therefore need to be estimated as part of the overall statistical analysis. Two standard methods for building such models are *factor analysis* and principal components analysis, as we have already seen.

Factor Models in Risk Management

It is straightforward to use a factor model such as (18) to manage risk. For a given portfolio composition and fixed matrix, $\mathbf{B}$, of factor loadings, the sensitivity of the total portfolio value to each factor, F_i for $i = 1, \ldots, k$, is easily computed. The portfolio composition can then be adjusted if necessary in order to achieve the desired overall factor sensitivity. Obviously this process is easier to understand and justify when the factors are easy to interpret. When this is not the case then the model is purely statistical. This tends to occur when statistical methods such as factor analysis or PCA are used to build the factor model. Of course, as we have seen in the case of PCA, it is still possible even then for the identified factors to have an economic interpretation.

[17] A multivariate regression refers to a regression where there is more than one *dependent* variable.