1

Univariate statistics

In this chapter we review the basics of univariate statistics. For more on this subject see Mood, Graybill, and Boes (1974) and Casella and Berger (2001).

In Section 1.1 we introduce the definition of random variable and the concept of distribution, as well as four equivalent ways to represent a distribution: the most intuitive, i.e. the probability density function, and three equivalent representations, namely the cumulative distribution function, the characteristic function and the quantile. Depending on the applications, all of the above representations prove useful.

In Section 1.2 we discuss the parameters that summarize the main features of a distribution, such as the location, the dispersion, the degree of symmetry and the thickness of the tails. Then we present the graphical representation of these properties.

In Section 1.3 we introduce a few distributions that are useful to model and solve asset allocation problems.

1.1 Building blocks

A *random variable* X is the number that corresponds to a measurement that has yet to take place. The measurement can assume a range of values on the real axis $\mathbb{R}$, each with a specific probability.

For example, consider a stock that trades today on the exchange at the following price (e.g. in dollars):

$$\widetilde{x} \equiv 100. \tag{1.1}$$

Tomorrow's price X for this stock is a random variable. Something about this measurement is known: for example we might argue that tomorrow's measurement is more likely to be in the neighborhood of today's value (1.1) than in the neighborhood of, say, $x \equiv 10$.

The stochastic features of the different possible measurements of a random variable X can be described in terms of a *distribution*. A distribution is characterized by a *space of events* $\mathfrak{E}$ and a *probability* $\mathbb{P}$.

The unknown outcome x of the measurement of X corresponds to one specific event $\mathfrak{e}$ among many that can take place in the space of events $\mathfrak{E}$. Therefore, a random variable is a function from the space of events to the range of measurements on the real line $\mathbb{R}$: if a specific event $\mathfrak{e}$ takes place, the measurement will take on the value $x \equiv X(\mathfrak{e})$. In a different universe, a different event $\mathfrak{e}'$ might have taken place and thus the measurement would have been a different value $x' \equiv X(\mathfrak{e}')$.

The likelihood of different possible events is described by a probability $\mathbb{P}$, which is a measure on the space of events. The following notation stands for the probability of all the events $\mathfrak{e}$ in the space of events $\mathfrak{E}$ that give rise to a measurement of X in a given interval $[\underline{x}, \overline{x}]$:

$$\mathbb{P}\{X \in [\underline{x}, \overline{x}]\} \equiv \mathbb{P}\{\mathfrak{e} \in \mathfrak{E} \text{ such that } X(\mathfrak{e}) \in [\underline{x}, \overline{x}]\}. \tag{1.2}$$

A distribution can be represented in three equivalent ways.

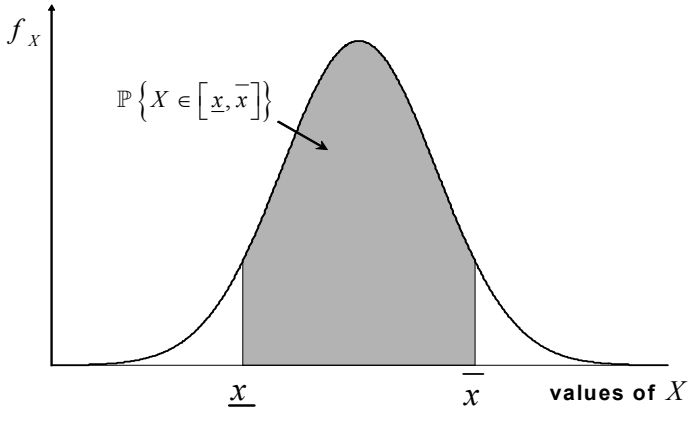

Fig. 1.1. Probability density function

The most intuitive way to represent the distribution of the random variable X is by means of the *probability density function (pdf)* f_X. Intuitively, the pdf shows a peak where the outcome of the measurement of X is more likely to occur. More formally, the probability density function is defined in such a way that the probability $\mathbb{P}$ that a measurement takes place in a generic interval $[\underline{x}, \overline{x}]$ is the area comprised the interval and the density, see Figure 1.1:

$$\mathbb{P}\{X \in [\underline{x}, \overline{x}]\} \equiv \int_{\underline{x}}^{\overline{x}} f_X(x)\, dx. \tag{1.3}$$

In particular, we notice that, since a probability is non-negative, the probability density function is non-negative:

$$f_X(x) \geq 0. \tag{1.4}$$

Furthermore, since the measurement of X must assume a value on the real axis, the following normalization must hold:

$$\int_{-\infty}^{+\infty} f_X(x)\, dx = 1. \tag{1.5}$$

For example the function

$$f_X(x) \equiv \frac{1}{\sqrt{\pi}} e^{-(x-\tilde{x})^2}, \tag{1.6}$$

which we plot in Figure 1.1, has a bell shape which is peaked around the current price (1.1). We show in a more general context in Section 1.3.2 that (1.6) satisfies (1.4) and (1.5). Therefore it is a probability density function which could model tomorrow's price for the stock.

To introduce the second equivalent way to describe a distribution we notice from (1.3) that, in order to compute probabilities, we always need to integrate the probability density function f_X over some interval. The *cumulative distribution function (cdf)* F_X is defined as the probability that the measurement be less than a generic value x, see Figure 1.2. In formulas:

$$F_X(x) \equiv \mathbb{P}\{X \leq x\} = \int_{-\infty}^{x} f_X(u)\, du. \tag{1.7}$$

In other words, the cumulative distribution function is obtained from the probability density function by applying $(B.27)$, the integration operator:

$$F_X = \mathcal{I}[f_X]. \tag{1.8}$$

This means that the probability density function can be recovered from the cumulative distribution function by applying the derivative operator $(B.25)$, which is the inverse of the integration operator:

$$f_X = \mathcal{D}[F_X]. \tag{1.9}$$

Therefore the two representations are equivalent.

Given the properties (1.4) and (1.5) of the probability density function, it is easy to check that the cumulative distribution function is non-decreasing and satisfies the following normalization conditions:

$$F_X(-\infty) = 0, \quad F_X(+\infty) = 1. \tag{1.10}$$

On the other hand, any function with the above properties defines a cumulative distribution function.

We plot in Figure 1.2 the cumulative distribution function that corresponds to the density (1.6). This cumulative distribution function can be expressed in terms of the error function ($B.75$) as follows:

$$F_X(x) = \frac{1}{2}\left(1 + \text{erf}\left(x - \widetilde{x}\right)\right), \tag{1.11}$$

where $\widetilde{x} = 100$ is today's price (1.1) of the stock. This is a specific instance of a more general result, see Section 1.3.2.

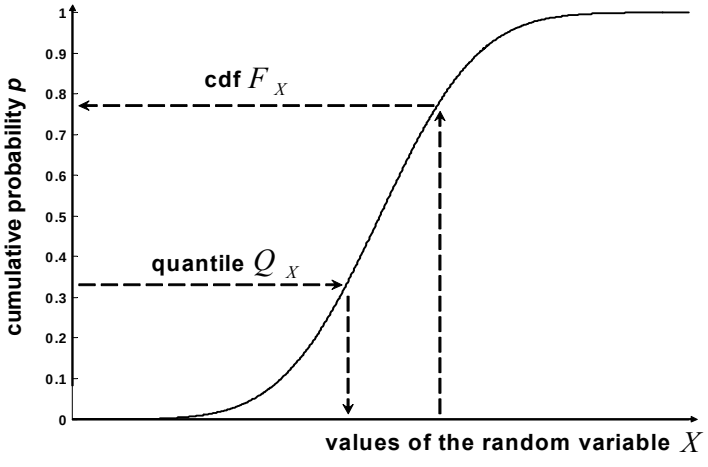

Fig. 1.2. Cumulative distribution function and quantile

A third way to describe the properties of a distribution is through the *characteristic function (cf)* ϕ_X, defined in terms of the expectation operator ($B.56$) as follows:

$$\phi_X(\omega) \equiv \mathrm{E}\left\{e^{i\omega X}\right\}, \tag{1.12}$$

where $i \equiv \sqrt{-1}$ is the imaginary unit. The characteristic function can assume values in the complex plane.

It is not straightforward to determine the properties of a generic characteristic function implied by the properties (1.4) and (1.5) of the probability density function. Nevertheless, a set of sufficient conditions is provided by

Polya's theorem, which states that function ϕ is a characteristic function of a distribution if it is real-valued, even, convex on the positive real axis, and if it satisfies:

$$\phi(0) \equiv 1, \quad \lim_{\omega \to \infty} \phi(\omega) \equiv 0, \tag{1.13}$$

see Cuppens (1975).

A comparison of (1.12) with $(B.34)$ and $(B.56)$ shows that the characteristic function is the Fourier transform of the probability density function:

$$\phi_X = \mathcal{F}[f_X]. \tag{1.14}$$

Therefore the probability density function can be recovered from the characteristic function by means of $(B.40)$, i.e. the inverse Fourier transform:

$$f_X = \mathcal{F}^{-1}[\phi_X]. \tag{1.15}$$

At times, the characteristic function proves to be the easiest way to describe a distribution.

The characteristic function of the distribution in the example (1.6) reads:

$$\phi_X(\omega) = e^{i\widetilde{x}\omega - \frac{1}{4}\omega^2}, \tag{1.16}$$

where $\widetilde{x} = 100$ is today's price (1.1) of the stock. This is a specific instance of a more general result, see Section 1.3.2.

We stress that the probability density function f_X, the cumulative distribution function F_X and the characteristic function ϕ_X are three equivalent ways to represent the distribution of the random variable X. We summarize in Figure 1.3 the mutual relationships among these representations.

We also discuss a fourth, fully equivalent way to describe all the properties of a random variable which is very important in financial applications, see Section 5.5. The *quantile* Q_X of the random variable X is the inverse of the cumulative distribution function:

$$Q_X(p) \equiv F_X^{-1}(p), \tag{1.17}$$

where $p \in [0,1]$ denotes a specific value of cumulative probability, see Figure 1.2. By definition, the quantile associates with a cumulative probability p the number x such that the probability that X be less than x is p. In other words, the quantile is defined implicitly by the following equation:

$$\mathbb{P}\{X \leq Q_X(p)\} = p. \tag{1.18}$$

Since the quantile is equivalent to the cumulative distribution function, it is equivalent to any of the above representations of the distribution of X.

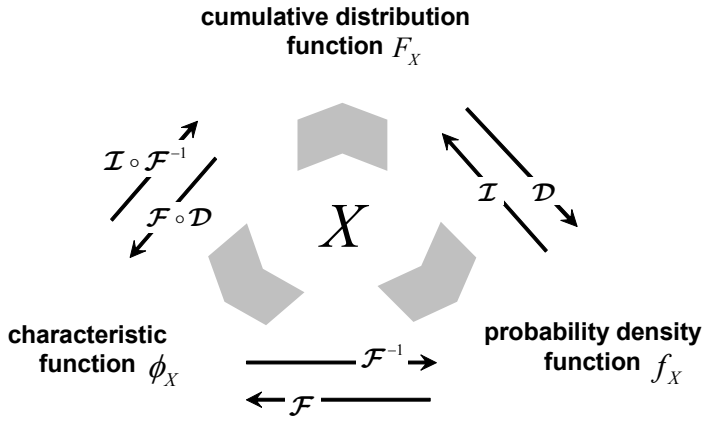

Fig. 1.3. Equivalent representations of a univariate distribution

The quantile of the distribution of our example (1.6) reads in terms of the inverse of the error function ($B.75$) as follows:

$$Q_X(p) = \widetilde{x} + \text{erf}^{-1}(2p-1), \tag{1.19}$$

where $\widetilde{x} = 100$ is today's price (1.1) of the stock. This is a specific instance of a more general result, see Section 1.3.2.

In the above discussion we have made the implicit assumption that the probability density function f_X is smooth and positive. This is not always the case.

For instance, the definition of quantile provided in (1.17) only makes sense if the cumulative distribution function is *strictly* increasing, because only in this case with each point on the vertical axis of the cumulative function is associated one and only one point on the horizontal axis, see Figure 1.2. In order for the cumulative distribution function to be strictly increasing, the probability density function must be strictly positive. Indeed, the cumulative distribution function is flat in those regions where the probability density function is null.

To handle situations such as the above example we have two options: either we build a more sophisticated mathematical framework that does not rely on the assumptions of smoothness and positivity for probability density function,

or we make the above hypotheses legitimate by regularizing the probability density function as in Appendix B.4. We choose throughout the book the second approach, for practical as well as "philosophical" reasons, see $(B.54)$ and comments thereafter.

To handle the above example, Since the regularized probability density function $f_{X;\epsilon}$ obtained with $(B.54)$ is strictly positive, the respective regularized cumulative distribution function $F_{X;\epsilon}$ is strictly increasing and thus invertible. Therefore we can properly define the regularized quantile as in (1.17) as the inverse of the cumulative distribution function:

$$Q_{X;\epsilon} \equiv F^{-1}_{X;\epsilon}. \tag{1.20}$$

The exact quantile is recovered as the limit of the regularized quantile when the bandwidth ϵ tends to zero, if this limit exists. Otherwise, we simply work with the approximate quantile.

1.2 Summary statistics

In this section we discuss a few parameters that summarize the most information about the properties of a distribution.

1.2.1 Location

Suppose that we need to summarize all the information regarding the random variable X in only one number, the one value that best represents the whole range of possible outcomes. We are looking for a *location* parameter $\text{Loc}\{X\}$ that provides a fair indication of where on the real axis the random variable X will end up taking its value.

A location parameter should enjoy a few intuitive features. In the first place, if the distribution is peaked around a specific value, the location parameter should be close to that peak. In particular, a constant a can be seen as an infinitely peaked random variable, see $(B.22)$ and comments thereafter. Thus the location of a constant should be the constant itself:

$$\text{Loc}\{a\} = a. \tag{1.21}$$

More in general, the location parameter should track any affine transformation of the random variable:

$$\text{Loc}\{a + bX\} = a + b\,\text{Loc}\{X\}, \tag{1.22}$$

where a and $b > 0$ are the constants that define the affine transformation.

Property (1.22) is called the *affine equivariance* of the location parameter.

To understand this property, imagine that the variable X is the price of a stock in cents and that we are interested in the value of our portfolio, which consists of that stock and an extra dollar in cash. Assume that we believe that tomorrow the stock price will be located in a neighborhood of, say, the following value in cents:

$$\text{Loc}\{X\} = 298c. \tag{1.23}$$

Then the whole portfolio should be located around the following value in dollars:

$$\text{Loc}\left\{1 + \frac{X}{100}\right\} = 3.98\$ = 1 + \frac{\text{Loc}\{X\}}{100}. \tag{1.24}$$

An immediate choice for the location parameter is the center of mass of the distribution, i.e. the weighted average of each possible outcome, where the weight of each outcome is provided by its respective probability. This corresponds to computing the *expected value* ($B.56$) of the random variable:

$$\text{E}\{X\} \equiv \int_{-\infty}^{+\infty} x f_X(x)\, dx. \tag{1.25}$$

As we prove in Appendix www.1.4, the expected value is affine equivariant, i.e. it satisfies (1.22). Therefore the expected value of a random variable is a sensible parameter of location, when the integral that defines it converges.

Whenever the characteristic function (1.12) of X is known and analytical, i.e. it can be recovered entirely from its Taylor series expansion, computing the expected value is easy, as we show in Appendix www.1.6.

An alternative choice for the location parameter is the *median*, which is the quantile (1.17) relative to the specific cumulative probability $p \equiv 1/2$:

$$\text{Med}\{X\} \equiv Q_X\left(\frac{1}{2}\right). \tag{1.26}$$

From (1.18), the median is defined equivalently by the following implicit equation:

$$\int_{-\infty}^{\text{Med}\{X\}} f_X(x)\, dx = \frac{1}{2}. \tag{1.27}$$

As we prove in Appendix www.1.4, the median is affine equivariant, i.e. it satisfies (1.22). Therefore the median of a random variable is also a sensible parameter of location.

Consider a distribution that is *symmetrical* around some value $\widetilde{x}$, i.e. a distribution such that the probability density function f_X satisfies:

$$(\text{Refl} \circ \text{Shift}_{\widetilde{x}})[f_X] = \text{Shift}_{\widetilde{x}}[f_X], \tag{1.28}$$

where the reflection and shift operators are defined in ($B.32$) and ($B.33$) respectively. In this case it is intuitive to assume that the symmetry point is a

good parameter of location. Indeed, we prove in Appendix www.1.5 that the symmetry point coincides with both the median and the expected value:

$$\text{Med}\{X\} = \text{E}\{X\} = \widetilde{x}. \tag{1.29}$$

A third parameter of location is the *mode*, which refers to the shape of the probability density function f_X. Indeed, the mode is defined as the point that corresponds to the highest peak of the density function:

$$\text{Mod}\{X\} \equiv \underset{x\in\mathbb{R}}{\text{argmax}}\{f_X(x)\}. \tag{1.30}$$

By construction, the mode is peaked around the most likely outcomes. In Appendix www.1.4 we show that the mode is affine equivariant, i.e. it satisfies (1.22): therefore the mode of a random variable is also a sensible parameter of location. Nevertheless, there might exist two or more equally high global maxima, in which case the mode is not defined.

In the example (1.6) it is easy to see that the above three parameters of location, namely expected value, median and mode, coincide:

$$\text{E}\{X\} = \text{Med}\{X\} = \text{Mod}\{X\} = \widetilde{x}, \tag{1.31}$$

where $\widetilde{x} = 100$ is today's price (1.1) of the stock. This is a specific instance of a more general result, see Section 1.3.2.

We remark that the expected value summarizes "global" features of the distribution, in that the whole density f_X contributes to the result, see (1.25); the median only involves "half" of the distribution, see (1.27); the mode provides a "local" picture, in that only a specific value matters, see (1.30).

1.2.2 Dispersion

In this section we summarize in one number the degree of dispersion of the random variable X. In other words, we are looking for a *dispersion* parameter $\text{Dis}\{X\}$ that yields an indication of the extent to which the location parameter might be wrong in guessing the outcome of X.

As in the case of the location parameter, we require that the dispersion parameter display an intuitive property:

$$\text{Dis}\{a + bX\} = |b|\,\text{Dis}\{X\}, \tag{1.32}$$

where a and b are constants. Property (1.32) is called the *affine equivariance* of the dispersion parameter.

To understand the affine equivariance property of the dispersion parameter, imagine that the variable X is tomorrow's price of a stock in cents and that we assess a dispersion of, say 10 cents. Then the dispersion in dollars of the stock price should be 0.1 dollars:

$$\text{Dis}\left\{\frac{X}{100}\right\} = 0.10\$ = \frac{\text{Dis}\{X\}}{100}. \tag{1.33}$$

Furthermore, the dispersion of a portfolio made of that stock and a given amount m of cents in cash should be the same as the dispersion of the stock alone:

$$\text{Dis}\{X\} = 10c = \text{Dis}\{X + m\}. \tag{1.34}$$

In view of multivariate generalizations it is useful to reformulate (1.32) the affine equivariance property in a different way. First we define the *z-score* of the random variable X, which is a normalized version of X located in zero and with unitary dispersion:

$$Z_X \equiv \frac{X - \text{Loc}\{X\}}{\text{Dis}\{X\}}. \tag{1.35}$$

The affine equivariance property of the location parameter (1.22) and of the dispersion parameter (1.32) are equivalent to the condition that the squared z-score remain unaffected by affine transformations:

$$Z^2_{a+bX} = Z^2_X. \tag{1.36}$$

A popular dispersion parameter is the *interquantile range*, defined as the difference of two arbitrary quantiles

$$\text{Ran}\{X\} \equiv Q_X(\overline{p}) - Q_X(\underline{p}), \tag{1.37}$$

where $\overline{p} > \underline{p}$. The standard choice is $\overline{p} \equiv 3/4$, which corresponds to the *upper quartile,* and $\underline{p} \equiv 1/4$, which corresponds to the *lower quartile*. We prove in Appendix www.1.4 that the range is affine equivariant, i.e. it satisfies (1.32).

To introduce another dispersion parameter, consider the *modal dispersion*:

$$\text{MDis}\{X\} \equiv -\left.\frac{1}{\frac{d^2 \ln f_X}{dx^2}}\right|_{x=\text{Mod}\{X\}}, \tag{1.38}$$

see O'Hagan (1994). As we prove in a more general multivariate setting in Appendix www.2.5, the square root of the modal dispersion is affine equivariant and thus it is a suitable dispersion parameter. To see the rationale of this definition, consider a second-order Taylor approximation of the probability density function of X in a neighborhood of the mode:

$$f_X(x) \approx f_X(\text{Mod}\{X\}) + \left.\frac{d^2 f_X}{dx^2}\right|_{x=\text{Mod}\{X\}} (x - \text{Mod}\{X\})^2. \tag{1.39}$$

The larger in absolute value the second derivative, which is negative around a maximum, the thinner the the probability density function around the mode, and thus the less the dispersion of X. Considering the logarithm of the pdf in the definition (1.38) and taking the square root of the result makes the ensuing parameter affine equivariant.

To define more dispersion parameters we notice that intuitively the dispersion of X is a sort of distance between X and its location parameter. We recall that the space L_X^p of functions of X is a vector space with the norm $\|\cdot\|_{X;p}$, see $(B.57)$ and $(B.58)$. Therefore we can define a dispersion parameter in a natural way as the distance between the random variable and its location parameter:

$$\text{Dis}\{X\} \equiv \|X - \text{Loc}\{X\}\|_{X;p}. \tag{1.40}$$

The general properties $(A.7)$ of a norm imply that this definition of dispersion is affine equivariant, i.e. it satisfies (1.32).

In particular, if we set $p \equiv 1$ in (1.40) and we define the location parameter as the expected value (1.25), we obtain the *mean absolute deviation (MAD)*:

$$\begin{aligned}\text{MAD}\{X\} &\equiv \text{E}\{|X - \text{E}\{X\}|\} \\ &= \int_{\mathbb{R}} |x - \text{E}\{X\}| f_X(x)\, dx.\end{aligned} \tag{1.41}$$

On the other hand, if we set $p \equiv 2$ in (1.40) and again we define the location parameter as the expected value (1.25) we obtain the *standard deviation*:

$$\begin{aligned}\text{Sd}\{X\} &\equiv \left(\text{E}\left\{(X - \text{E}\{X\})^2\right\}\right)^{\frac{1}{2}} \\ &= \sqrt{\int_{\mathbb{R}} (x - \text{E}\{X\})^2 f_X(x)\, dx}.\end{aligned} \tag{1.42}$$

When the integral in (1.42) converges, the standard deviation is the benchmark dispersion parameter. The square of the standard deviation, which is very important in applications, is called the *variance*:

$$\text{Var}\{X\} \equiv (\text{Sd}\{X\})^2 = \int_{\mathbb{R}} (x - \text{E}\{X\})^2 f_X(x)\, dx. \tag{1.43}$$

Whenever the characteristic function (1.12) of X is known and it is analytical, i.e. it can be recovered entirely from its Taylor series expansion, computing the variance is straightforward, see Appendix www.1.6.

In our example (1.6) the range reads:

$$\text{Ran}\{X\} = \text{erf}^{-1}\left(\frac{1}{2}\right) - \text{erf}^{-1}\left(-\frac{1}{2}\right) \approx 0.95; \tag{1.44}$$

the mean absolute deviation reads:

$$\text{MAD}\{X\} = \frac{1}{\sqrt{\pi}} \approx 0.56; \tag{1.45}$$

and the standard deviation reads:

$$\text{Sd}\{X\} = \frac{1}{\sqrt{2}} \approx 0.71. \tag{1.46}$$

These are specific instances of more general results, see Section 1.3.2.

We remark that, similarly to the expected value, the standard deviation and the mean absolute deviation summarize global features of the distribution, in that the whole density f_X contributes to the result. On the other hand, similarly to the median, the range involves parts of the distribution. Finally, similarly to the mode, the modal dispersion provides a local picture, in that only a small neighborhood of a specific value matters.

1.2.3 Higher-order statistics

By means of the expectation operator ($B.56$) we can introduce the moments, summary statistics that provide more insight into the features of a distribution.

The *k-th raw moment* of a random variable X is the expectation of the k-th power of the random variable:

$$\text{RM}_k^X \equiv \text{E}\left\{X^k\right\}. \tag{1.47}$$

The *k-th central moment* of a random variable is a location-independent version of the respective raw moment:

$$\text{CM}_k^X \equiv \text{E}\left\{\left(X - \text{E}\{X\}\right)^k\right\}. \tag{1.48}$$

We already discussed the first raw moment of a random variable X, which is the expected value (1.25); we also discussed the second central moment, which is the variance (1.43).

The third central moment provides a measure of the degree of symmetry of the distribution of X. The standard measure of symmetry of a distribution is the *skewness*, which is the third central moment normalized by the standard deviation, in such a way to make it scale-independent:

$$\text{Sk}\{X\} \equiv \frac{\text{CM}_3^X}{\left(\text{Sd}\{X\}\right)^3}. \tag{1.49}$$

In particular, a distribution whose probability density function is symmetric around its expected value has null skewness. If the skewness is positive (negative), occurrences larger than the expected value are more (less) likely than occurrences smaller than the expected value.

In our example (1.6) we have:

$$\mathrm{Sk}\{X\} = 0. \tag{1.50}$$

This is a specific instance of a more general result, see Section 1.3.2. The result (1.50) is consistent with the symmetry of the probability density function (1.6).

The fourth moment provides a measure of the relative weight of the tails with respect to the central body of a distribution. The standard quantity to evaluate this balance is the *kurtosis*, defined as the normalized fourth central moment:

$$\mathrm{Ku}\{X\} \equiv \frac{\mathrm{CM}_4^X}{(\mathrm{Sd}\{X\})^4}. \tag{1.51}$$

The kurtosis gives an indication of how likely it is to observe a measurement far in the tails of the distribution: a large kurtosis implies that the distribution displays "*fat tails*".

In our example (1.6) we have:

$$\mathrm{Ku}\{X\} = 3. \tag{1.52}$$

This is a specific instance of a more general result, see Section 1.3.2.

We remark that all the above moments and summary statistics involve in general integrations. If the integral that defines the expectation operator (B.56) does not converge, the respective moment is not defined. Nevertheless, whenever the characteristic function of the distribution is known and analytical, i.e. it can be recovered entirely from its Taylor series expansion, we can compute these quantities by means of simple differentiation and some algebra, as we show in Appendix www.1.6.

1.2.4 Graphical representations

To obtain an immediate idea of the properties of location and dispersion of a random variable X it is useful to represent them graphically.

One way to do this is by means of a *box plot*, which is the plot of the first, second and third quartile: the box plot summarizes the location of the given distribution, in this case the median, and its dispersion, in this case the interquartile range. More in general, the plot of a few key quantiles gives an idea of the main features of the probability density function f_X, and thus of the distribution of X, see Figure 1.4. Furthermore, the box plot gives an idea of the degree of symmetry of the distribution: if the distance between lower quartile and median exceeds the distance between median and upper quartile the distribution is more spread below the median than it is above the median.

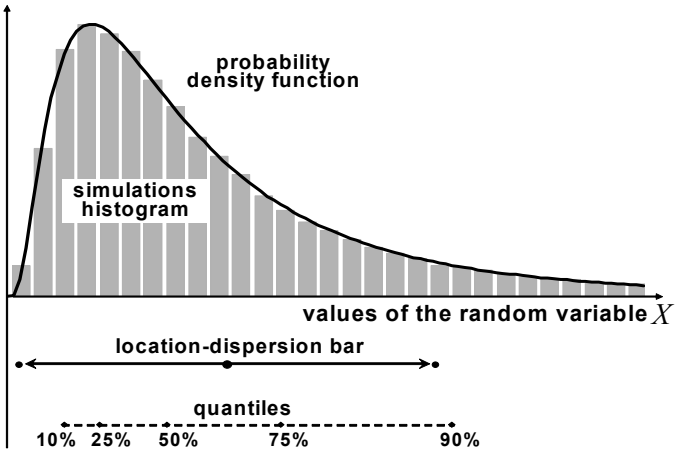

Fig. 1.4. Summary statistics of univariate distributions

Another way to summarize the main features of a distribution is by means of the location-dispersion bar, namely the set of points x which are not any farther from the location parameter of X than one dispersion:

$$\text{Loc}\{X\} - \text{Dis}\{X\} \leq x \leq \text{Loc}\{X\} + \text{Dis}\{X\}. \tag{1.53}$$

The location-dispersion bar is an interval centered on the location parameter and wide twice the dispersion parameter, see Figure 1.4. The dispersion bar becomes particularly useful in its generalization to a multivariate setting, see Section 2.4.3.

1.3 Taxonomy of distributions

In this section we discuss a few distributions that are useful in asset allocation applications. All the distribution introduced are special univariate cases of the more general distributions introduced in Section 2.6.

1.3.1 Uniform distribution

The *uniform distribution* models the situation where the realization of the random variable X is bound to take place on an interval $[a, b]$ and all the values within that interval are equally likely outcomes of the measurement of X.

We use the following notation to indicate that X is uniformly distributed on the interval $[a, b]$:

$$X \sim \text{U}([a, b]). \tag{1.54}$$

Equivalent representations

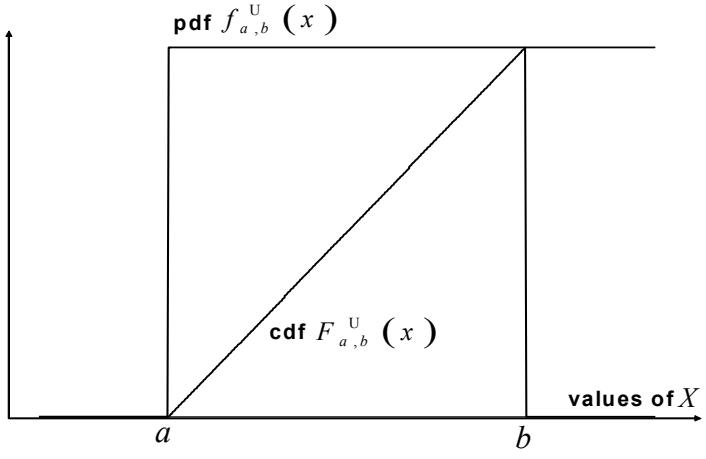

Fig. 1.5. Uniform distribution: pdf and cdf

The probability density function of the uniform distribution reads:

$$f^{\mathrm{U}}_{a,b}(x) = \frac{1}{b-a}\mathbb{I}_{[a,b]}(x), \tag{1.55}$$

where $\mathbb{I}$ is the indicator function ($B.72$), see Figure 1.5.

The cumulative distribution function of the uniform distribution reads:

$$F^{\mathrm{U}}_{a,b}(x) = \frac{x-a}{b-a}\mathbb{I}_{[a,b]}(x) + H^{(b)}(x), \tag{1.56}$$

where H is the Heaviside step function ($B.73$), see Figure 1.5.

The characteristic function of the uniform distribution reads:

$$\phi^{\mathrm{U}}_{a,b}(\omega) = \frac{1}{\omega}\frac{2}{b-a}\sin\left(\frac{b-a}{2}\omega\right)e^{i\frac{a+b}{2}\omega}, \tag{1.57}$$

see Abramowitz and Stegun (1974).

Inverting (1.56) we obtain the quantile of the uniform distribution:

$$Q^{\mathrm{U}}_{a,b}(p) = a + (b-a)\,p. \tag{1.58}$$

Summary statistics

The standard parameters that summarize the properties of the uniform distribution, namely expected value, standard deviation, skewness and kurtosis, read respectively:

$$\mathrm{E}\{X\} = a + \frac{1}{2}(b-a) \tag{1.59}$$

$$\mathrm{Sd}\{X\} = \frac{1}{\sqrt{12}}(b-a) \tag{1.60}$$

$$\mathrm{Sk}\{X\} = 0 \tag{1.61}$$

$$\mathrm{Ku}\{X\} = \frac{9}{5}, \tag{1.62}$$

see Abramowitz and Stegun (1974).

It is possible to compute explicitly also other parameters of location and dispersion. Since the uniform distribution is symmetrical, from (1.29) the median is equal to the expected value:

$$\mathrm{Med}\{X\} = a + \frac{1}{2}(b-a). \tag{1.63}$$

The mode is not defined. An integration yields the mean absolute deviation:

$$\mathrm{MAD}\{X\} = \frac{1}{8}(b-a). \tag{1.64}$$

The interquartile range is easily obtained from (1.58) and reads:

$$\mathrm{Ran}\{X\} = \frac{1}{2}(b-a). \tag{1.65}$$

1.3.2 Normal distribution

The *normal distribution* is by far the most used and studied distribution. Its bell-shaped profile and its analytical tractability make it the benchmark choice to describe random variables that are peaked around a given value but can take on values on the whole real axis. The normal distribution depends on two parameters μ and σ^2. The parameter μ is a location parameter that turns out to be the expected value and the parameter $|\sigma|$ is a dispersion parameter that turns out to be the standard deviation.

We use the following notation to indicate that X is normally distributed according to those parameters:

$$X \sim \mathrm{N}\left(\mu, \sigma^2\right). \tag{1.66}$$

The case $\mu \equiv 0$ and $\sigma^2 \equiv 1$ defines the *standard normal distribution*.

Equivalent representations

The probability density function of the normal distribution is defined as follows:

$$f^{\mathrm{N}}_{\mu,\sigma^2}(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1.67}$$

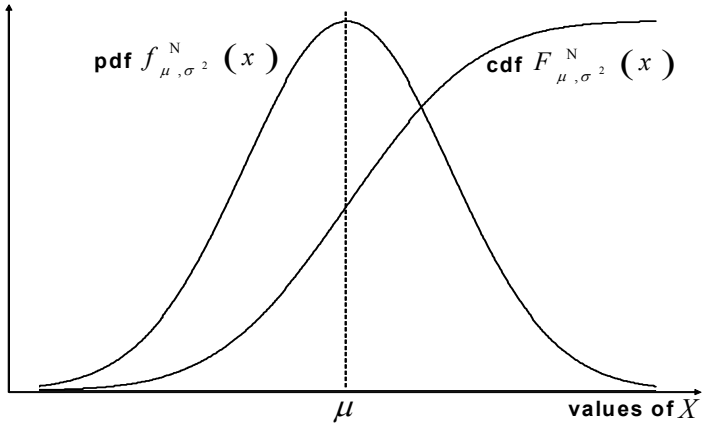

Fig. 1.6. Normal distribution: pdf and cdf

see Figure 1.6.

The cumulative distribution function of the normal distribution can be expressed in terms of the error function ($B.75$) as follows:

$$F^{\mathrm{N}}_{\mu,\sigma^2}(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right], \tag{1.68}$$

see Figure 1.6.

The characteristic function of the normal distribution reads:

$$\phi^{\mathrm{N}}_{\mu,\sigma}(\omega) = e^{i\mu\omega - \frac{\sigma^2}{2}\omega^2}, \tag{1.69}$$

see Abramowitz and Stegun (1974).

Inverting (1.68) we obtain the quantile of the normal distribution:

$$Q^{\mathrm{N}}_{\mu,\sigma^2}(p) = \mu + \sqrt{2\sigma^2}\,\mathrm{erf}^{-1}(2p-1). \tag{1.70}$$

Summary statistics

The standard parameters that summarize the properties of the normal distribution, namely expected value, standard deviation, skewness and kurtosis, can be computed from the characteristic function (1.69) with the technique described in Appendix www.1.6, and read respectively:

$$\mathrm{E}\{X\} = \mu \tag{1.71}$$

$$\mathrm{Sd}\{X\} = \sqrt{\sigma^2} \tag{1.72}$$

$$\mathrm{Sk}\{X\} = 0 \tag{1.73}$$

$$\mathrm{Ku}\{X\} = 3. \tag{1.74}$$

It is possible to compute explicitly also other parameters of location and dispersion. Since the normal distribution is symmetrical, from Appendix www.1.5 we know that the median is equal to the expected value, which in this case is also equal to the mode:

$$\text{Med}\{X\} = \text{Mod}\{X\} = \mu. \tag{1.75}$$

The mean absolute deviation reads:

$$\text{MAD}\{X\} = \sqrt{\frac{2\sigma^2}{\pi}}. \tag{1.76}$$

The interquartile range can be easily derived from the expression of the quantile (1.70) and reads:

$$\text{Ran}\{X\} = \sqrt{2\sigma^2}\left[\text{erf}^{-1}\left(\frac{1}{2}\right) - \text{erf}^{-1}\left(-\frac{1}{2}\right)\right]. \tag{1.77}$$

1.3.3 Cauchy distribution

Like the normal distribution, the *Cauchy distribution* is bell-shaped and depends on two parameters μ and σ^2. The parameter μ is a location parameter that can take on any value and the parameter σ^2 is the square of a dispersion parameter $|\sigma|$.

We use the following notation to indicate that X is Cauchy distributed with the above parameters:

$$X \sim \text{Ca}\left(\mu, \sigma^2\right). \tag{1.78}$$

The case $\mu \equiv 0$ and $\sigma^2 \equiv 1$ is called the *standard Cauchy distribution*.

The Cauchy distribution is used instead of the normal distribution when extreme events are comparatively speaking more likely to occur than in the case of a normal distribution. This phenomenon is also known as *fat tails* behavior.

Equivalent representations

The probability density function of the Cauchy distribution, which we plot in Figure 1.7, is defined as follows:

$$f^{\text{Ca}}_{\mu,\sigma^2}(x) \equiv \frac{1}{\pi\sqrt{\sigma^2}}\left(1 + \frac{(x-\mu)^2}{\sigma^2}\right)^{-1}, \tag{1.79}$$

see Abramowitz and Stegun (1974) and `mathworld.com`.

The cumulative distribution function of the Cauchy distribution, which we plot in Figure 1.7, reads:

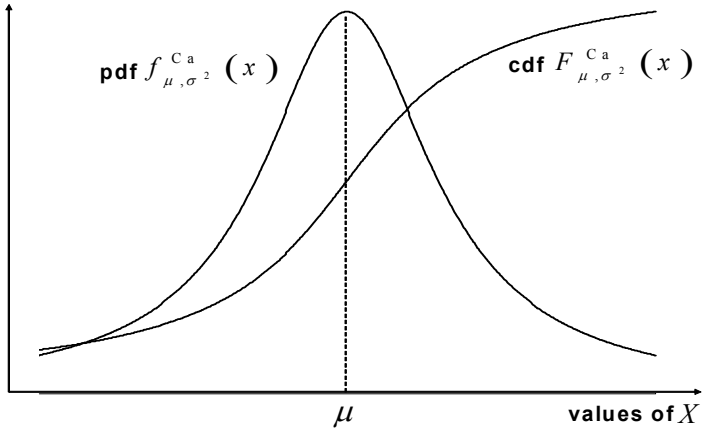

Fig. 1.7. Cauchy distribution: pdf and cdf

$$F_{\mu,\sigma^2}^{\mathrm{Ca}}(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sqrt{\sigma^2}}\right), \tag{1.80}$$

see e.g. `mathworld.com`.

The characteristic function of the Cauchy distribution reads:

$$\phi_{\mu,\sigma^2}^{\mathrm{Ca}}(\omega) = e^{i\mu\omega - \sqrt{\sigma^2}|\omega|}, \tag{1.81}$$

see e.g. Abramowitz and Stegun (1974) and `mathworld.com`.

The quantile of the Cauchy distribution is obtained inverting (1.80) and reads:

$$Q_{\mu,\sigma^2}^{\mathrm{Ca}}(p) = \mu + \sqrt{\sigma^2}\tan\left(\pi p - \frac{\pi}{2}\right). \tag{1.82}$$

Summary statistics

The moments of the Cauchy distribution are not defined. This happens because the probability density function (1.79) decays proportionally to x^{-2} in the tails. Therefore the computation of the generic moment of order k involves integrating a function of the order of x^{k-2} as $|x| \to \infty$, which does not converge for any positive integer k.

The fact that the moments are not defined is reflected also in the expression of the characteristic function (1.81), which is not differentiable in zero. Therefore in particular it cannot be expressed as a Taylor series in terms of the moments as in Appendix www.1.6.

Nevertheless, from the expression of the quantile (1.82) we obtain the median, which is also equal to the mode:

$$\text{Med}\{X\} = \text{Mod}\{X\} = \mu. \tag{1.83}$$

Similarly, from the expression of the quantile (1.82) we obtain the interquartile range:

$$\text{Ran}\{X\} = 2\sqrt{\sigma^2}. \tag{1.84}$$

1.3.4 Student t distribution

Like the normal and the Cauchy distributions, the *Student t distribution*, is bell-shaped. It depends on three parameters (ν, μ, σ^2). The parameter ν, which takes on integer values, is called the *degrees of freedom* of the Student t distribution and determines the thickness of the tails. The parameter μ is a location parameter that can take on any value and σ^2 is the square of a dispersion parameter $|\sigma|$.

We use the following notation to indicate that X is Student t distributed with the above parameters:

$$X \sim \text{St}\left(\nu, \mu, \sigma^2\right). \tag{1.85}$$

The case $\mu = 0$ and $\sigma^2 = 1$ is called the *standard Student t distribution*.

Equivalent representations

On `mathworld.com` we find the standard Student t probability density function. By applying formula $(T.14)$ in Appendix www.1.2 we obtain the probability density function of the general Student t distribution, which reads:

$$f^{\text{St}}_{\nu,\mu,\sigma}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi\sigma^2}} \left(1 + \frac{1}{\nu}\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}}, \tag{1.86}$$

where Γ is the gamma function $(B.80)$. See in Figure 1.8 the bell-shaped profile of this function.

Similarly, we find on `mathworld.com` the standard Student t cumulative distribution function. By applying formula $(T.15)$ in Appendix www.1.2 we obtain the cumulative distribution function of the general Student t distribution. In Figure 1.8 we plot this function, which reads explicitly:

$$F^{\text{St}}_{\nu,\mu,\sigma}(x) = 1 - \frac{1}{2} I\left(\left(1 + \frac{1}{\nu}\frac{(x-\mu)^2}{\sigma^2}\right)^{-1}; \frac{\nu}{2}, \nu\right), \tag{1.87}$$

where I is the regularized beta function $(B.91)$.

The quantile of the Student t distribution cannot be expressed analytically.

On p. 948 of Abramowitz and Stegun (1974) we find the characteristic function of the standard Student t distribution. By applying formula $(T.18)$

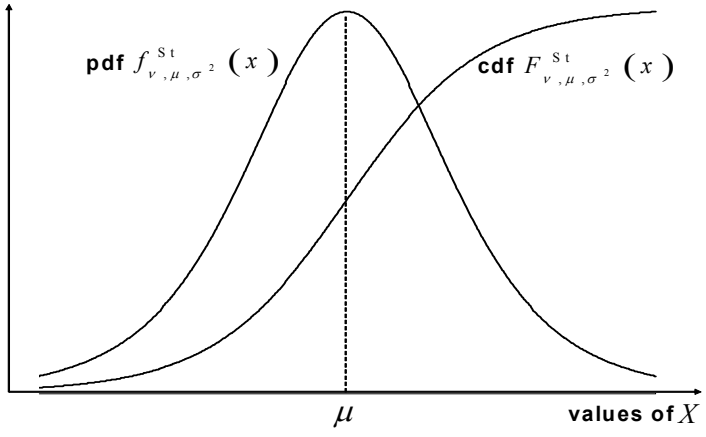

Fig. 1.8. Student t distribution: pdf and cdf

in Appendix www.1.2 we obtain the cumulative distribution function of the general Student t distribution:

$$\phi^{\text{St}}_{\nu,\mu,\sigma^2} = \frac{e^{i\omega\mu}}{\pi\Gamma(\nu/2)}\left(\frac{\sigma^2\omega^2}{4\nu}\right)^{\frac{\nu}{4}} Y_{\frac{\nu}{2}}\left(\sqrt{\frac{\sigma^2\omega^2}{\nu}}\right), \tag{1.88}$$

where Γ denotes the gamma function ($B.80$) and Y_ν is the Bessel function of the second kind ($B.93$).

Summary statistics

The standard parameters that summarize the properties of the Student t distribution, namely expected value, standard deviation, skewness and kurtosis, are computed in Abramowitz and Stegun (1974) and read:

$$\text{E}\{X\} = \mu \tag{1.89}$$

$$\text{Sd}\{X\} = \frac{\nu}{\nu-2}\sqrt{\sigma^2} \tag{1.90}$$

$$\text{Sk}\{X\} = 0 \tag{1.91}$$

$$\text{Ku}\{X\} = 3 + \frac{6}{\nu-4}. \tag{1.92}$$

These parameters are defined for $\nu > 1$, 2, 3 and 4 respectively.

The Student t distribution includes the normal distribution and the Cauchy distribution as special cases. Indeed we show in Appendix www.2.14 in a more general context that the limit $\nu \to \infty$ of the Student t probability

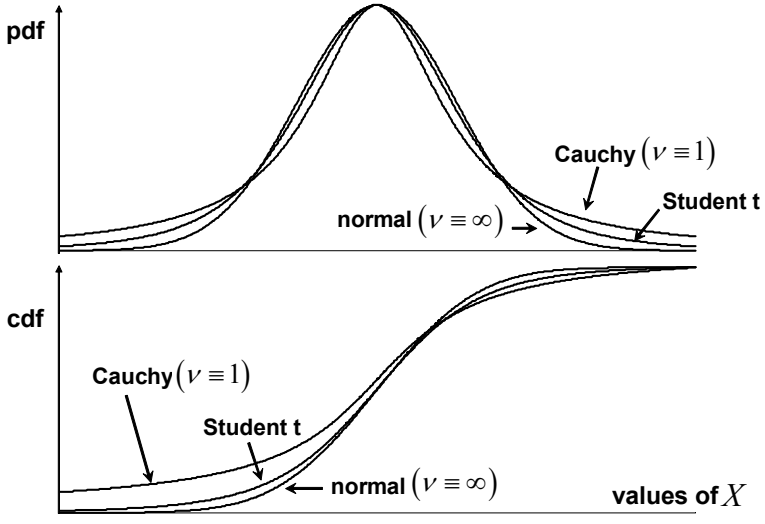

Fig. 1.9. Relations among Cauchy, normal, and Student t distributions

density function (1.86) yields the normal probability density function (1.67). On the other hand, if we set $\nu \equiv 1$ in (1.86) and recall $(B.81)$ and $(B.82)$, we obtain the Cauchy probability density function (1.79).

As we see in Figure 1.9, the lower the degrees of freedom, the "fatter" the tails of the probability density function and the flatter the cumulative distribution function. This is consistent with the above discussion of the Cauchy distribution and with the expression (1.92) of the kurtosis.

1.3.5 Lognormal distribution

The price of a security is a positive random variable. Furthermore, the random changes from the current price are better stated in percentage terms than in absolute terms. In other words, if the price now is, say, 1\$, the chance that the price will double, which corresponds to an absolute change of 1\$ is approximately equal to the chance that the price will become half, which corresponds to an absolute change of 0.5\$.

To model this feature, consider a random variable (the "percentage change") that is normally distributed:

$$Y \sim \mathrm{N}\left(\mu, \sigma^2\right). \tag{1.93}$$

The *lognormal distribution* is defined as the distribution of the variable $X \equiv e^Y$. The rationale behind this name is obviously the fact that by definition X is lognormally distributed if and only if its logarithm is normally distributed.

We use the following notation to indicate that X is lognormally distributed with the above parameters:

$$X \sim \text{LogN}\left(\mu, \sigma^2\right). \tag{1.94}$$

Equivalent representations

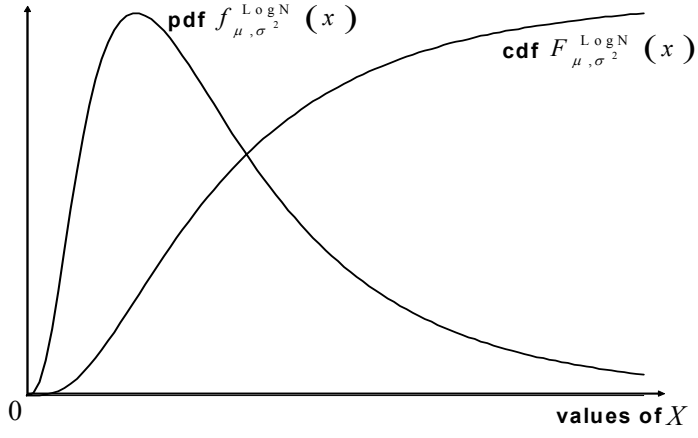

Fig. 1.10. Lognormal distribution: pdf and cdf

The probability density function of the lognormal distribution reads from $(T.21)$ in Appendix www.1.1 as follows:

$$f_{\mu,\sigma^2}^{\text{LogN}}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(\ln(x)-\mu)^2}{\sigma^2}}. \tag{1.95}$$

We notice in Figure 1.10 that the lognormal pdf is not symmetrical.

Applying formula $(T.22)$ in Appendix www.1.3 to the normal cumulative distribution function (1.68), we obtain the cumulative distribution function of the lognormal distribution, which we plot in Figure 1.10:

$$F_{\mu,\sigma^2}^{\text{LogN}}(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\ln(x)-\mu}{\sqrt{2\sigma^2}}\right)\right). \tag{1.96}$$

The characteristic function is not known in analytic form.

Applying formula $(T.23)$ in Appendix www.1.3 to the normal quantile (1.70), we obtain the quantile of the lognormal distribution:

$$Q_{\mu,\sigma^2}^{\text{LogN}}(p) = e^{\mu+\sqrt{2\sigma^2}\,\text{erf}^{-1}(2p-1)}. \tag{1.97}$$

Summary statistics

The standard parameters that summarize the properties of the lognormal distribution, namely expected value, standard deviation, skewness and kurtosis read respectively:

$$\mathrm{E}\{X\} = e^{\mu+\frac{\sigma^2}{2}} \tag{1.98}$$

$$\mathrm{Sd}\{X\} = e^{\mu+\frac{\sigma^2}{2}}\sqrt{e^{\sigma^2}-1} \tag{1.99}$$

$$\mathrm{Sk}\{X\} = \sqrt{e^{\sigma^2}-1}\left(e^{\sigma^2}+2\right) \tag{1.100}$$

$$\mathrm{Ku}\{X\} = e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3. \tag{1.101}$$

The above parameters can be computed with a technique which we discuss in a general multivariate environment in Appendix www.2.16. In particular, we notice that the lognormal distribution is positively skewed, as we see in the profile of the probability density function in Figure 1.10.

It is possible to compute explicitly also other parameters of location and dispersion. The median follows from $(T.9)$ in Appendix www.1.1:

$$\mathrm{Med}\{X\} = e^{\mu}. \tag{1.102}$$

The first-order condition on the density (1.95) yields the mode:

$$\mathrm{Mod}\{X\} = e^{\mu-\sigma^2}. \tag{1.103}$$

Notice that the three location parameters (1.98), (1.102) and (1.102) yield different results.

The expression of the interquartile range follows from the quantile (1.97) and reads:

$$\mathrm{Ran}\{X\} = e^{\mu}\left(e^{\sqrt{2\sigma^2}\,\mathrm{erf}^{-1}\left(\frac{1}{2}\right)}-e^{\sqrt{2\sigma^2}\,\mathrm{erf}^{-1}\left(-\frac{1}{2}\right)}\right). \tag{1.104}$$

1.3.6 Gamma distribution

We introduce here a distribution that is useful in Bayesian analysis, where the parameters of a distribution are considered as random variables. In particular, we will need a distribution to describe the variance, which is always non-negative. The gamma distribution proves particularly suitable in this respect.

Consider a set of ν random variables $(Y_1,\ldots,Y_\nu)$ that are normally identically distributed:

$$Y_t \sim \mathrm{N}\left(\mu,\sigma^2\right), \tag{1.105}$$

for all $t = 1,\ldots,\nu$. Furthermore, assume that these random variables are independent[1].

[1] Refer to Section 2.3 for a formal definition of dependence.

The *non-central gamma distribution* with ν degrees of freedom is defined as the distribution of the following variable:

$$X \equiv Y_1^2 + \cdots + Y_\nu^2. \tag{1.106}$$

As such, the non-central gamma distribution depends on three parameters $\left(\nu, \mu, \sigma^2\right)$. The parameter ν is an integer and is called the *degrees of freedom* of the gamma distribution; the parameter μ can assume any value and is called the *non-centrality parameter*; the parameter σ^2 is a positive scalar and is called the *scale parameter*.

We use the following notation to indicate that X is distributed as a non-central gamma with the above parameters:

$$X \sim \mathrm{Ga}\left(\nu, \mu, \sigma^2\right). \tag{1.107}$$

The special case where the non-centrality parameter is $\mu \equiv 0$ gives rise to the *central gamma distribution with ν degrees of freedom*. We use the following notation to indicate that X is central-gamma distributed with the above parameters:

$$X \sim \mathrm{Ga}\left(\nu, \sigma^2\right). \tag{1.108}$$

The special case where the scale parameter is $\sigma^2 \equiv 1$ gives rise to the *(non-central) chi-square distribution with ν degrees of freedom*.

In particular, when $\mu \equiv 0$ and $\sigma^2 \equiv 1$ we obtain the *chi-square distribution with ν degrees of freedom*, which is denoted as follows:

$$X \sim \chi^2_\nu. \tag{1.109}$$

In view of generalizations to a multivariate setting and applications later on in the book, we focus below on the central gamma distribution, which includes the chi-square distribution as a special case.

Equivalent representations

The results and expressions that follow can be found on `mathworld.com`.

The probability density function of the central gamma distribution reads:

$$f^{\mathrm{Ga}}_{\nu,\sigma^2}(x) = \frac{1}{(2\sigma^2)^{\frac{\nu}{2}}\, \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}\frac{x}{\sigma^2}}, \tag{1.110}$$

where Γ is the gamma function ($B.80$). We plot in Figure 1.11 the profile of this density.

The cumulative distribution function of the central gamma distribution reads:

$$F^{\mathrm{Ga}}_{\nu,\sigma^2}(x) = P\left(\frac{\nu}{2}; \frac{x}{2\sigma^2}\right), \tag{1.111}$$

where P is the lower regularized gamma function ($B.85$), see Figure 1.11 for a plot.

The characteristic function of the central gamma distribution reads:

$$\phi^{\mathrm{Ga}}_{\nu,\sigma^2}(\omega) = \left(1 - 2i\sigma^2\omega\right)^{-\frac{\nu}{2}}. \tag{1.112}$$

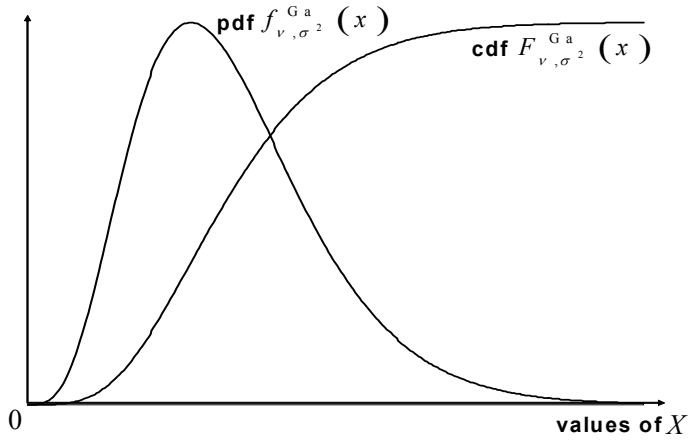

Fig. 1.11. Gamma distribution: pdf and cdf

Summary statistics

The standard parameters that summarize the properties of the gamma distribution, namely expected value, standard deviation, skewness and kurtosis read respectively:

$$\mathrm{E}\{X\} = \nu\sigma^2 \tag{1.113}$$

$$\mathrm{Sd}\{X\} = \sqrt{2\nu}\sigma^2 \tag{1.114}$$

$$\mathrm{Sk}\{X\} = \sqrt{\frac{8}{\nu}} \tag{1.115}$$

$$\mathrm{Ku}\{X\} = 3 + \frac{12}{\nu}. \tag{1.116}$$

The first-order condition on the probability density function yields the mode:

$$\mathrm{Mod}\{X\} = (\nu - 2)\,\sigma^2. \tag{1.117}$$

1.3.7 Empirical distribution

Suppose that our information i_T regarding the random variable X consists of T past measurements of this variable:

$$i_T \equiv \{x_1, \ldots, x_T\}. \tag{1.118}$$

Notice the lower-case notation in (1.118), since the measurements have already taken place, and therefore the outcomes are no longer random variables.

The *empirical distribution* provides a straightforward model for the basic assumption of statistics that we can learn about the future from the past: under the empirical distribution any of the past outcomes is assumed equally likely to occur again in future measurements of X, whereas any other value cannot occur.

We use the following notation to indicate that X is distributed according to an empirical distribution with the above observations:

$$X \sim \mathrm{Em}\,(i_T)\,. \tag{1.119}$$

Equivalent representations

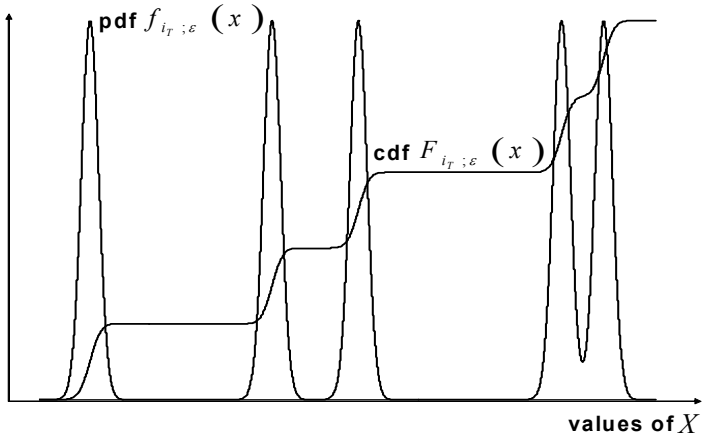

Fig. 1.12. Empirical distribution (regularized): pdf and cdf

The empirical distribution is discrete. Therefore its probability density function is a generalized function. As in $(B.22)$, we can express the empirical pdf as follows:

$$f_{i_T}(x) = \frac{1}{T}\sum_{t=1}^{T} \delta^{(x_t)}(x)\,, \tag{1.120}$$

where δ is the Dirac delta $(B.16)$.

It is impossible to represent graphically this probability density function, unless we regularize it by means of the convolution as in $(B.54)$. The regularized probability density function of the empirical distribution reads in terms of the smooth approximation $(B.18)$ of the Dirac delta as follows:

$$f_{i_T;\epsilon} \equiv f_{i_T} * \delta_\epsilon^{(0)} = \frac{1}{T}\sum_{t=1}^{T} \delta_\epsilon^{(x_t)}, \tag{1.121}$$

where ϵ is a small bandwidth. We plot in Figure 1.12 the regularized version of the empirical probability density function.

From $(B.53)$ the empirical cumulative distribution function reads:

$$F_{i_T}(x) = \frac{1}{T}\sum_{t=1}^{T} H^{(x_t)}(x), \tag{1.122}$$

where H is the Heaviside step function $(B.73)$. In Figure 1.12 we plot the regularized cumulative distribution function ensuing from (1.121).

From the definition of the characteristic function (1.12) in terms of the expectation operator $(B.56)$, and from the property $(B.17)$ of the Dirac delta we obtain:

$$\phi_{i_T}(\omega) = \frac{1}{T}\sum_{t=1}^{T} e^{i\omega x_t}. \tag{1.123}$$

The quantile (1.17) is not defined because the cumulative distribution function (1.122) is not invertible. Nevertheless, using the regularization technique (1.20) and then considering the limit where the bandwidth ϵ tends to zero we can easily obtain the result. Indeed, a comparison of Figure 1.12 with Figure 1.2 shows that the quantile of the empirical distribution reads:

$$Q_{i_T}(p) = x_{[pT]:T}, \tag{1.124}$$

where $[\cdot]$ denotes the integer part and where we denote as follows the ordered set of observations:

$$\begin{aligned} x_{1:T} &\equiv \min\{x_1, \ldots, x_T\} \\ &\vdots \\ x_{T:T} &\equiv \max\{x_1, \ldots, x_T\}. \end{aligned} \tag{1.125}$$

Summary statistics

The standard parameters that summarize the properties of the empirical distribution, namely expected value, standard deviation, skewness and kurtosis, follow from the definition of the expectation operator $(B.56)$, and the property $(B.17)$ of the Dirac delta. We denote these parameters respectively as follows:

$$\widehat{\mathrm{E}}_{i_T} = \frac{1}{T}\sum_{t=1}^{T} x_t \tag{1.126}$$

$$\widehat{\mathrm{Sd}}_{i_T} = \frac{1}{T}\sum_{t=1}^{T} \left(x_t - \widehat{\mathrm{E}}_{i_T}\right)^2 \tag{1.127}$$

$$\widehat{\mathrm{Sk}}_{i_T} = \frac{1}{T}\sum_{t=1}^{T} \left(\frac{x_t - \widehat{\mathrm{E}}_{i_T}}{\widehat{\mathrm{Sd}}_{i_T}}\right)^3 \tag{1.128}$$

$$\widehat{\mathrm{Ku}}_{i_T} = \frac{1}{T}\sum_{t=1}^{T} \left(\frac{x_t - \widehat{\mathrm{E}}_{i_T}}{\widehat{\mathrm{Sd}}_{i_T}}\right)^4. \tag{1.129}$$

These parameters are also called *sample mean*, *sample standard deviation*, *sample skewness* and *sample kurtosis* respectively.

The mode is not defined. From the expression for the quantile (1.124) we obtain the *sample median*:

$$\mathrm{Med}\{X\} = x_{\left[\frac{T}{2}\right]:T}. \tag{1.130}$$

Similarly, from the expression for the quantile we obtain the *sample interquartile range*:

$$\mathrm{Med}\{X\} = x_{\left[\frac{3}{4}T\right]:T} - x_{\left[\frac{1}{4}T\right]:T}. \tag{1.131}$$

2

Multivariate statistics

The financial markets contain many sources of risk. When dealing with several sources of risk at a time we cannot treat them separately: the joint structure of multi-dimensional randomness contains a wealth of information that goes beyond the juxtaposition of the information contained in each single variable.

In this chapter we discuss multivariate statistics. The structure of this chapter reflects that of Chapter 1: to ease the comprehension of the multivariate case refer to the respective section in that chapter. For more on this subject see also references such as Mardia, Kent, and Bibby (1979), Press (1982) and Morrison (2002).

In Section 2.1 we introduce the building blocks of multivariate distributions which are direct generalizations of the one-dimensional case. These include the three equivalent representations of a distribution in terms of the probability density function, the characteristic function and the cumulative distribution function.

In Section 2.2 we discuss the factorization of a distribution into its purely univariate components, namely the marginal distributions, and its purely joint component, namely the copula. To present copulas we use the leading example of vanilla options.

In Section 2.3 we introduce the concept of independence among random variables and the related concept of conditional distribution.

In Section 2.4 we discuss the location summary statistics of a distribution such as its expected value and its mode, and the dispersion summary statistics such as the covariance matrix and the modal dispersion. We detail the geometrical representations of these statistics in terms of the location-dispersion ellipsoid, and their probabilistic interpretations in terms of a multivariate version of Chebyshev's inequality. We conclude introducing more summary statistics such as the multivariate moments, which provide a deeper insight into the shape of a multivariate distribution.

In Section 2.5 we discuss summary statistics for the level of interdependence among the marginal components of a multivariate distribution. We introduce copula-driven measures of dependence such as the Schweizer-Wolff

measure and copula-driven measures of concordance, such as Spearman's rho and Kendall's tau. We also analyze the advantages and potential pitfalls of using the correlation as a measure of interdependence.

In Section 2.6 we present a taxonomy of parametric distributions that represent the multivariate generalization of those introduced in Chapter 1. In particular, in view of their applications to estimation theory, we introduce matrix-variate distributions, such as the Wishart distribution, the matrix-variate normal, Cauchy and Student t distributions. In view of their applications to modeling prices, we introduce generic log-distributions, of which the lognormal is an example, along with a general technique to compute all the moments of these distributions.

In Section 2.7 we discuss a few broad classes of distributions that are very useful in applications, namely elliptical and symmetric stable distributions, which are symmetric and analytically tractable, and infinitely divisible distribution, that allow to model the financial markets at any investment horizon.

2.1 Building blocks

In this section we introduce the multivariate extension of the building blocks of univariate statistics discussed in Section 1.1, namely the concept of multivariate distribution and its equivalent representations in terms of the joint probability density function, the joint cumulative distribution function and the joint characteristic function.

A *random variable* $\mathbf{X}$ of dimension N is a vector that corresponds to a joint measurement of N variables that has yet to take place:

$$\mathbf{X} \equiv (X_1, \ldots, X_N)' . \tag{2.1}$$

A joint measurement corresponds to one point in the space $\mathbb{R}^N$. Therefore the joint measurements of $\mathbf{X}$ can assume a range of values in various regions of $\mathbb{R}^N$, and each of these values has a specific probability to occur.

For example, consider two stocks that trade today on the exchange at the following prices (e.g. in dollars):

$$\widetilde{x}_1 \equiv 100, \quad \widetilde{x}_2 \equiv 50. \tag{2.2}$$

Tomorrow's prices $\mathbf{X} \equiv (X_1, X_2)'$ for these stocks are a bivariate random variable. A joint measurement is a point in the plane $\mathbb{R}^2$ and with each point on the plane is associated a different probability.

The stochastic features of the different possible measurements of a random variable $\mathbf{X}$ can be described in terms of a *multivariate distribution*. A distribution is characterized by a *space of events* $\mathfrak{E}$ and a *probability* $\mathbb{P}$.

The unknown outcome $\mathbf{x}$ of the joint measurement of the entries of $\mathbf{X}$ corresponds to one specific event $\mathfrak{e}$ among many that can take place in a space of events $\mathfrak{E}$. Therefore, a multivariate random variable is a function from the space of events to the range of measurements in $\mathbb{R}^N$: if a specific event $\mathfrak{e}$ takes place, the measurement will take on the value $\mathbf{x} \equiv \mathbf{X}(\mathfrak{e})$. In a different universe a different event $\mathfrak{e}'$ might have taken place and thus the measurement would have assumed a different value $\mathbf{x}' \equiv \mathbf{X}(\mathfrak{e}')$.

The likelihood of different possible events is described by a probability $\mathbb{P}$, which is a measure on the space of events. The following notation stands for the probability of all the events $\mathfrak{e}$ in the space of events $\mathfrak{E}$ that give rise to a joint measurement of $\mathbf{X}$ in the region $\mathcal{R}$ of the space $\mathbb{R}^N$:

$$\mathbb{P}\{\mathbf{X} \in \mathcal{R}\} \equiv \mathbb{P}\left\{\mathfrak{e} \in \mathfrak{E} \text{ such that } \mathbf{X}(\mathfrak{e}) \in \mathcal{R} \subset \mathbb{R}^N\right\}. \tag{2.3}$$

This expression generalizes (1.2).

As in the one-dimensional case, a distribution can be represented in three equivalent ways.

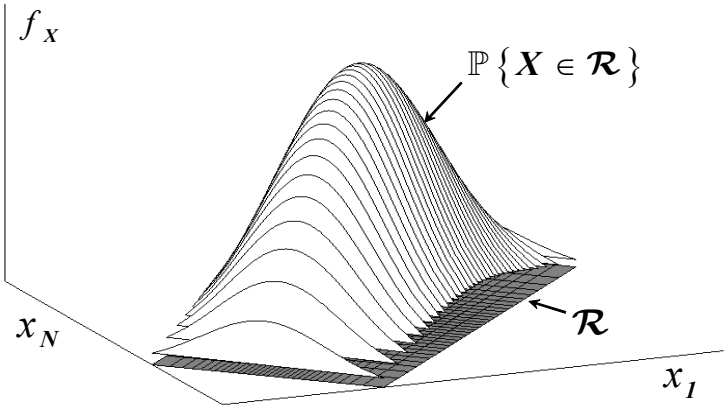

Fig. 2.1. Multivariate probability density function

The most intuitive way to represent the distribution of the random variable $\mathbf{X}$ is through the *probability density function (pdf)* $f_{\mathbf{X}}$. Intuitively, the pdf shows a peak where the outcome of the measurement of $\mathbf{X}$ is more likely to occur. More formally, the probability density function is defined in such a way that the probability that a measurement takes place in a generic region $\mathcal{R}$ is the volume comprised between the region and the density, see Figure 2.1:

$$\mathbb{P}\{\mathbf{X} \in \mathcal{R}\} \equiv \int_{\mathcal{R}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}. \tag{2.4}$$

In particular, since a probability is non-negative, the probability density function is non-negative:

$$f_{\mathbf{X}}(\mathbf{x}) \geq 0. \tag{2.5}$$

Furthermore, since the joint measurement of $\mathbf{X}$ must assume a value in $\mathbb{R}^N$, the following normalization condition must hold:

$$\int_{\mathbb{R}^N} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 1. \tag{2.6}$$

For instance, consider the following function:

$$f_{\mathbf{X}}(x_1, x_2) \equiv \frac{\sqrt{5}}{\sqrt{8}\pi} e^{-\frac{1}{2}u(x_1,x_2)}, \tag{2.7}$$

where u is the following quadratic form:

$$u(x_1, x_2) \equiv \begin{pmatrix} x_1 - \widetilde{x}_1 \\ x_2 - \widetilde{x}_2 \end{pmatrix}' \begin{pmatrix} \frac{10}{3} & -\frac{2}{3}\sqrt{10} \\ -\frac{2}{3}\sqrt{10} & \frac{10}{3} \end{pmatrix} \begin{pmatrix} x_1 - \widetilde{x}_1 \\ x_2 - \widetilde{x}_2 \end{pmatrix}; \tag{2.8}$$

and where $(\widetilde{x}_1, \widetilde{x}_2)$ are the current prices (2.2) of the two stocks in our example. This function has a bell shape which is peaked around the current prices, see Figure 2.1. The function (2.7) satisfies (2.5) and (2.6), as we show in a more general context in Section 2.6.2. Therefore it defines a probability density function, which we can use to model tomorrow's prices $\mathbf{X} \equiv (X_1, X_2)'$ for the two stocks in the example.

The second equivalent way to describe the distribution of a random variable $\mathbf{X}$ is the *cumulative distribution function (cdf)* $F_{\mathbf{X}}$, which is defined as the probability that the joint measurement of the entries of $\mathbf{X}$ be less than a given generic value:

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &\equiv \mathbb{P}\{\mathbf{X} \leq \mathbf{x}\} \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} f_{\mathbf{X}}(u_1, \ldots, u_N)\, du_1 \cdots du_N. \end{aligned} \tag{2.9}$$

The cumulative distribution function is obtained from the probability density function by applying the combined integration operators ($B.27$) as follows:

$$F_{\mathbf{X}} = (\mathcal{I}_1 \circ \cdots \circ \mathcal{I}_N)[f_{\mathbf{X}}]. \tag{2.10}$$

In turn, the probability density function can be recovered from the cumulative distribution function by applying the combined differentiation operators ($B.25$) as follows:

$$f_{\mathbf{X}} = (\mathcal{D}_1 \circ \cdots \circ \mathcal{D}_N)[F_{\mathbf{X}}]. \tag{2.11}$$

Therefore the two representations in terms of pdf and cdf are equivalent. The positivity condition (2.5) and the normalization condition (2.6) on the pdf

transfer to the cdf in a way similar to the one-dimensional case (1.10). Indeed $F_{\mathbf{X}}$ is an increasing function of each coordinate and satisfies the following normalization conditions:

$$F_{\mathbf{X}}(x_1, \ldots, -\infty, \ldots, x_N) = 0, \quad F_{\mathbf{X}}(+\infty, \ldots, +\infty) = 1. \tag{2.12}$$

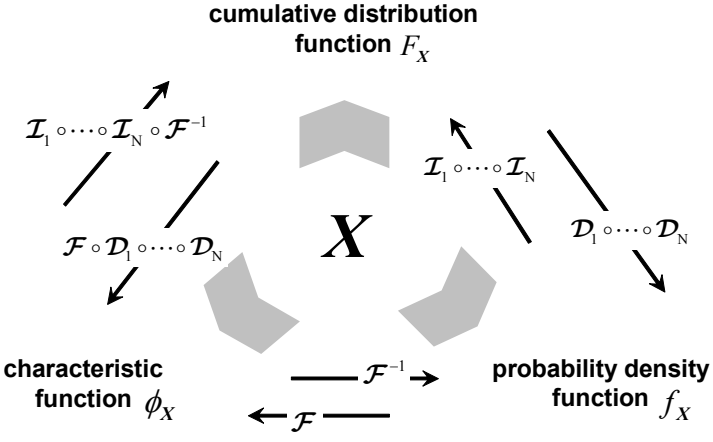

Fig. 2.2. Equivalent representations of a multivariate distribution

The third way to describe the properties of a distribution is by means of the *characteristic function (cf)* $\phi_{\mathbf{X}}$, defined in terms of the expectation operator ($B.56$) as follows:

$$\phi_{\mathbf{X}}(\boldsymbol{\omega}) \equiv \mathrm{E}\left\{e^{i\boldsymbol{\omega}'\mathbf{X}}\right\}, \tag{2.13}$$

where $i \equiv \sqrt{-1}$ is the imaginary unit. The characteristic function assumes values in the complex plane.

A comparison of (2.13) with ($B.34$) and ($B.56$) shows that the characteristic function is the Fourier transform of the probability density function:

$$\phi_{\mathbf{X}} = \mathcal{F}[f_{\mathbf{X}}]. \tag{2.14}$$

Therefore the probability density function can be recovered by means of the inverse Fourier transform ($B.40$) from the characteristic function:

$$f_{\mathbf{X}} = \mathcal{F}^{-1}[\phi_{\mathbf{X}}]. \tag{2.15}$$

At times the characteristic function proves to be the easiest way to represent a distribution.

The characteristic function of the distribution of the example (2.7) reads:

$$\phi_{\mathbf{X}}(\omega_1,\omega_2) = e^{i(\omega_1\widetilde{x}_1+\omega_2\widetilde{x}_2)}e^{-\frac{1}{2}w(\omega_1,\omega_2)}, \tag{2.16}$$

where $(\widetilde{x}_1,\widetilde{x}_2)$ are the current prices (2.2) of the stocks and where w is the following quadratic form:

$$w(\omega_1,\omega_2) = \begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}' \begin{pmatrix}1/2 & 1/\sqrt{10}\\ 1/\sqrt{10} & 1/2\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}. \tag{2.17}$$

This is a specific instance of the more general result (2.157).

We stress that the probability density function $f_{\mathbf{X}}$, the cumulative distribution function $F_{\mathbf{X}}$ and the characteristic function $\phi_{\mathbf{X}}$ are three fully equivalent ways to represent the distribution of the random variable $\mathbf{X}$. We summarize in Figure 2.2 the mutual relationships among these representations.

As in the one-dimensional case discussed in Chapter 1, in the sequel we make the implicit assumption that the probability density function $f_{\mathbf{X}}$ is a smooth and strictly positive function. In general, this is not the case. To make our hypothesis legitimate we regularize whenever necessary the probability density function as discussed in Appendix B.4:

$$f_{\mathbf{X}} \mapsto f_{\mathbf{X};\epsilon} \equiv \frac{1}{(2\pi)^{\frac{N}{2}}\epsilon^N}\int_{\mathbb{R}^N} e^{-\frac{(\mathbf{y}-\mathbf{x})'(\mathbf{y}-\mathbf{x})}{2\epsilon^2}} f(\mathbf{y})\,d\mathbf{y}. \tag{2.18}$$

For the practical as well as "philosophical" motivations behind the regularization, see $(B.54)$ and comments thereafter.

2.2 Factorization of a distribution

The distribution of a multivariate random variable $\mathbf{X}$ can be factored into two separate components. On the one hand the marginal distributions of each entry of the vector $\mathbf{X}$, which represent the purely univariate features of $\mathbf{X}$. On the other hand the copula, a standardized distribution which summarizes the purely "joint" component of the distribution of $\mathbf{X}$. We summarize this schematically as follows:

$$\boxed{\textbf{multivariate = "1-dim" (marginals) + "joint"(copula)}} \tag{2.19}$$

2.2.1 Marginal distribution

Consider an N-dimensional random variable $\mathbf{X}$. We split $\mathbf{X}$ in two sub-sets: the K-dimensional random variable $\mathbf{X}_A$ made of the first K entries and the

$(N-K)$-dimensional random variable $\mathbf{X}_B$ made of the remaining entries:

$$\mathbf{X} \equiv \begin{pmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{pmatrix}. \tag{2.20}$$

The *marginal distribution* of the variable $\mathbf{X}_B$ is the distribution of $\mathbf{X}_B$ obtained disregarding the existence of $\mathbf{X}_A$. In particular, we obtain the marginal distribution of the generic entry X_n by disregarding the remaining $N-1$ entries.

Consider the bivariate example (2.7), which describes the joint stochastic behavior of two stock prices. The marginal distribution of the first stock must be the univariate example (1.6) of Chapter 1, which describes the stochastic behavior of the first stock only. Otherwise, the two models are in contradiction with each other and one of them must be wrong.

We can represent the marginal distribution of $\mathbf{X}_B$ by means of its cumulative distribution function:

$$\begin{aligned} F_{\mathbf{X}_B}(\mathbf{x}_B) &\equiv \mathbb{P}\{\mathbf{X}_B \leq \mathbf{x}_B\} = \mathbb{P}\{\mathbf{X}_A \leq +\infty, \mathbf{X}_B \leq \mathbf{x}_B\} \\ &\equiv F_{\mathbf{X}}(+\infty, \mathbf{x}_B). \end{aligned} \tag{2.21}$$

In words, the marginal cumulative distribution function is the joint cumulative distribution function, where the variables we intend to disregard are set to infinity.

Equivalently, we can represent the marginal distribution of $\mathbf{X}_B$ by means of its probability density function. Applying the differentiation operator to the cumulative distribution function (2.21) as in (2.11) we obtain:

$$f_{\mathbf{X}_B}(\mathbf{x}_B) \equiv \int_{\mathbb{R}^K} f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B)\, d\mathbf{x}_A. \tag{2.22}$$

In words, the marginal pdf averages out of the joint pdf the variables that we intend to disregard.

In our example, the integration of the joint pdf (2.7) yields:

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2)\, dx_2 = \frac{1}{\sqrt{\pi}} e^{-(x-\tilde{x}_1)^2}. \tag{2.23}$$

This computation is a specific instance of the more general result (2.162). Not surprisingly (2.23) is the one-dimensional pdf (1.6) of the first stock price.

Finally, we can represent the marginal distribution of $\mathbf{X}_B$ by means of its characteristic function:

$$\begin{aligned} \phi_{\mathbf{X}_B}(\boldsymbol{\omega}) &\equiv \mathrm{E}\left\{e^{i\boldsymbol{\omega}'\mathbf{X}_B}\right\} = \mathrm{E}\left\{e^{i\boldsymbol{\psi}'\mathbf{X}_A + \boldsymbol{\omega}'\mathbf{X}_B}\right\}\Big|_{\boldsymbol{\psi}=\mathbf{0}} \\ &\equiv \phi_{\mathbf{X}}(\mathbf{0}, \boldsymbol{\omega}). \end{aligned} \tag{2.24}$$

In words, the marginal characteristic function is the joint characteristic function, where the variables we intend to disregard are set to zero.

2.2.2 Copulas

In this section we introduce copulas. For more on this subject consult references such as Nelsen (1999).

Definition

The copula represents the true interdependence structure of a random variable, which in our applications is the market. Intuitively, the copula is a standardized version of the purely joint features of a multivariate distribution, which is obtained by filtering out all the purely one-dimensional features, namely the marginal distribution of each entry X_n.

In order to factor out the marginal components, we simply transform deterministically each entry X_n in a new random variable U_n, whose distribution is the same for each entry. Since the distribution of each U_n is normalized this way, we lose track of the specific marginal distribution of X_n.

In order to map a generic one-dimensional random variable X into a random variable U which has a distribution of our choice, consider the cumulative distribution function F_X defined in (1.7). By means of the function F_X we can define a new random variable, called the *grade* of X:

$$U \equiv F_X(X). \tag{2.25}$$

The grade of X is a deterministic transformation of the random variable X that assumes values in the interval $[0,1]$. We prove in we Appendix www.2.1 that the grade is uniformly distributed on this interval:

$$U \sim \mathrm{U}([0,1]). \tag{2.26}$$

To obtain a random variable Z with a distribution of our choice, we prove in Appendix www.2.1 that it suffices to compute the quantile function Q_Z of that distribution as in (1.17), and then to define Z as the quantile applied to the grade U:

$$Z \equiv Q_Z(U). \tag{2.27}$$

In Figure 2.3 we display the graphical interpretation of the above operations.[1]

In particular, we can standardize each marginal component X_n of the original random variable $\mathbf{X}$ by means of the uniform distribution. Therefore, we consider the vector of the grades:

[1] This technique also allows us to simulate univariate distributions of any kind starting with a uniform random number generator.

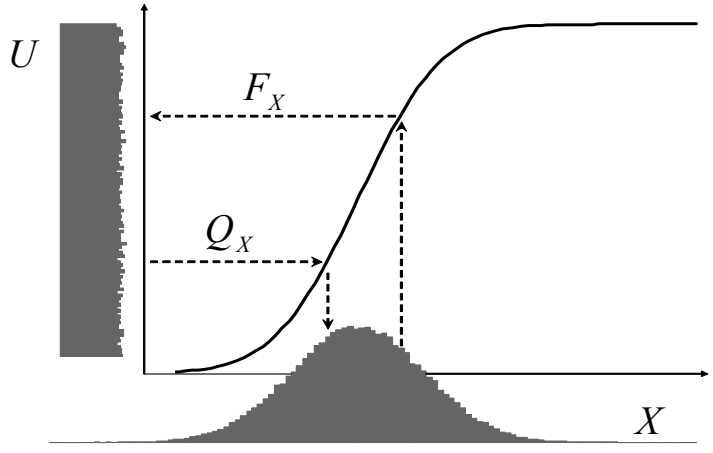

Fig. 2.3. Distribution of the grades: relation with cdf and quantile

$$\mathbf{U} \equiv \begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix} \equiv \begin{pmatrix} F_{X_1}(X_1) \\ \vdots \\ F_{X_N}(X_N) \end{pmatrix}. \tag{2.28}$$

This random variable assumes values on the unit hypercube:

$$[0,1]^N \equiv [0,1] \times \cdots \times [0,1]. \tag{2.29}$$

The *copula* of the multivariate random variable $\mathbf{X}$ is the joint distribution of its grades (2.28).

Representations

Since the copula is a distribution, namely the distribution of the grades $\mathbf{U}$, we can represent it in terms of the probability density function or the cumulative distribution function, or the characteristic function.

In Appendix www.2.3 we prove that the pdf of the copula reads:

$$f_{\mathbf{U}}(u_1,\ldots,u_N) = \frac{f_{\mathbf{X}}(Q_{X_1}(u_1),\ldots,Q_{X_N}(u_N))}{f_{X_1}(Q_{X_1}(u_1))\cdots f_{X_N}(Q_{X_N}(u_N))}, \tag{2.30}$$

where Q_{X_n} is the quantile (1.17) of the generic n-th marginal entry of $\mathbf{X}$.

In Figure 2.4 we plot the probability density function of the copula of the leading example (2.7), which we compute explicitly in a more general setting in (2.176).

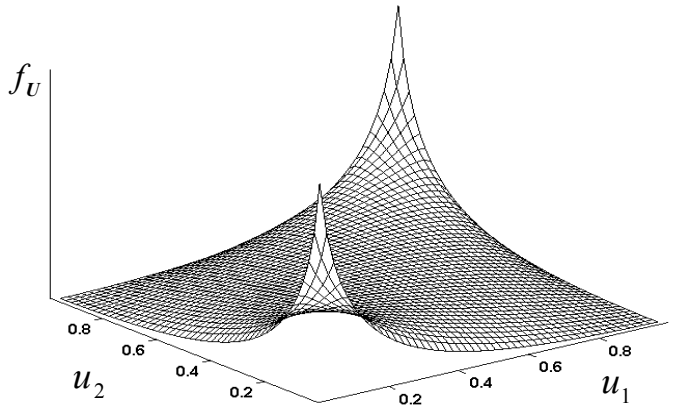

Fig. 2.4. Copula: probability density function

We can also represent the copula of the random variable $\mathbf{X}$ equivalently in terms of its cumulative distribution function. We prove in Appendix www.2.3 that the cdf of the copula of $\mathbf{X}$ reads:

$$F_{\mathbf{U}}(u_1, \ldots, u_N) = F_{\mathbf{X}}(Q_{X_1}(u_1), \ldots, Q_{X_N}(u_N)). \tag{2.31}$$

In particular, since the marginal distribution of the generic n-th entry is uniform, from (2.21) and (1.56) we obtain:

$$F_{\mathbf{U}}(1, \ldots, u_n, \ldots, 1) = u_n, \tag{2.32}$$

see Figure 2.10 for a few examples.

Properties

We can write (2.30) as follows:

$$f_{\mathbf{X}}(x_1, \ldots, x_N) = f_{\mathbf{U}}(F_{X_1}(x_1), \ldots, F_{X_N}(x_N)) \prod_{n=1}^{N} f_{X_n}(x_n). \tag{2.33}$$

This expression formalizes the loose expression (2.19): the joint pdf of a generic variable $\mathbf{X}$ is the product of the pdf of its copula and the pdf of the marginal densities of its entries. In other words, the copula factors out the purely marginal features of a distribution.

The copula contains all the information about the joint features of a distribution in a standardized form. Indeed, given the copula of $\mathbf{X}$, i.e. the distribution of the grades $\mathbf{U}$, from (2.28) we can reconstruct the distribution of $\mathbf{X}$ with a *deterministic* transformation of each grade separately:

$$\mathbf{X} \stackrel{d}{=} \begin{pmatrix} Q_{X_1}(U_1) \\ \vdots \\ Q_{X_N}(U_N) \end{pmatrix}. \tag{2.34}$$

Therefore, the copula is a standardized distribution that summarizes the purely joint features behind a multivariate random variable.

The purely joint features of a distribution characterize the true structure of randomness of a multivariate random variable. In other words, the copula allows to detect the true interdependence structure behind a generic multivariate random variable $\mathbf{X}$. In practical terms, the copula provides an effective tool to monitor and *hedge* the risks in the markets.

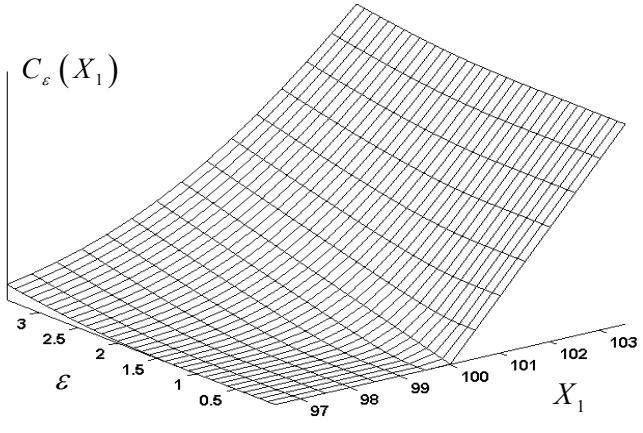

Fig. 2.5. Regularization of call option payoff

To see this, consider two *co-monotonic* random variables $\mathbf{X}$ and $\mathbf{Y}$, namely random variables such that:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} g_1(X_1) \\ \vdots \\ g_N(X_N) \end{pmatrix}, \tag{2.35}$$

where each g_n is an increasing invertible function of its argument.

For instance, in our example (2.7) of two stock prices $\mathbf{X} \equiv (X_1, X_2)'$, consider the payoff of a call option on the first stock with strike K, i.e. the following random variable:

$$C_1 \equiv \max(X_1 - K, 0), \tag{2.36}$$

where the strike price is, say, $K \equiv 100$.

The function C_1 is not strictly increasing in its argument X_1, but it becomes so if we replace it with a regularized version by means of $(B.49)$. In Appendix www.2.7 we show that the regularized call option payoff reads:

$$C_{1;\epsilon} \equiv \frac{(X_1 - K)}{2}\left(1 + \text{erf}\left(\frac{X_1 - K}{\sqrt{2\epsilon^2}}\right)\right) + \frac{\epsilon}{\sqrt{2\pi}} e^{-\frac{(X_1-K)^2}{2\epsilon^2}}. \tag{2.37}$$

This profile is smooth, strictly increasing in X_1, and tends to the exact profile (2.36) as the bandwidth ϵ tends to zero, see Figure 2.5. Therefore the stock price X_1 and the regularized call option payoff C_1 are co-monotonic and so are the pairs (X_1, X_2) and (C_1, X_2).

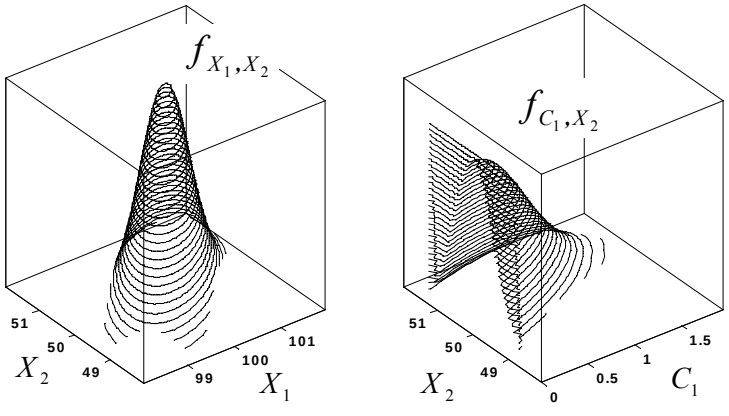

Fig. 2.6. Co-monotonic transformations: effects on the joint distribution

The joint distributions of co-monotonic variables are not equal, see Figure 2.6. Yet, the sources of randomness behind two co-monotonic random variables are the same. The common feature of these variables is their copula, as we show in Appendix www.2.3:

$$(\mathbf{X}, \mathbf{Y}) \text{ co-monotonic } \Leftrightarrow \text{ copula of } \mathbf{X} = \text{copula of } \mathbf{Y}. \tag{2.38}$$

In our example the joint distribution of the first stock price and the second stock price (X_1, X_2) is different than the joint distribution of the call option on the first stock and the second stock price (C_1, X_2). We see this in Figure 2.6, where we plot the two different probability density functions. Nevertheless, the copula of (X_1, X_2) is the same as the copula of (C_1, X_2) and is represented by the probability density function in Figure 2.4.

2.3 Dependence

Loosely speaking, two random variables are independent if any information on either variable does not affect the distribution of the other random variable. To introduce formally the concept of dependence, it is more intuitive to first define conditional distributions.

Consider an N-dimensional random variable $\mathbf{X}$. We split $\mathbf{X}$ in two subsets: the K-dimensional random variable $\mathbf{X}_A$ of the first K entries and the $(N-K)$-dimensional random variable $\mathbf{X}_B$ of the remaining entries:

$$\mathbf{X} \equiv \begin{pmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{pmatrix}. \tag{2.39}$$

The *conditional distribution* of the variable $\mathbf{X}_B$ given $\mathbf{x}_A$ is the distribution of $\mathbf{X}_B$ knowing that the realization of $\mathbf{X}_A$ is the specific value $\mathbf{x}_A$. We denote the conditioned random variable equivalently as $\mathbf{X}_B|\mathbf{x}_A$ or $\mathbf{X}_B|\mathbf{X}_A = \mathbf{x}_A$.

Suppose that in our example (2.7) the two stock prices $\mathbf{X} \equiv (X_1, X_2)'$ appear almost, but not quite, simultaneously on the screen. Before we look at the screen, the probability distribution of second stock price X_2 is represented by its marginal distribution. After we see the price of the first stock we have more information available. The distribution that describes the second stock price X_2, knowing that the price of the first stock is $X_1 \equiv x_1$, is the conditional distribution $X_2|x_1$.

The most intuitive way to represent the conditional distribution is the probability density function:

$$f_{\mathbf{X}_B|\mathbf{x}_A}(\mathbf{x}_B) = \frac{f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B)}{\int f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B)\, d\mathbf{x}_B} = \frac{f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B)}{f_{\mathbf{X}_A}(\mathbf{x}_A)}. \tag{2.40}$$

In words, the conditional pdf of $\mathbf{X}_B$ given knowledge of $\mathbf{X}_A$ is the joint pdf of $\mathbf{X}_A$ and $\mathbf{X}_B$ divided by the marginal pdf of $\mathbf{X}_A$ evaluated at the known point $\mathbf{x}_A$. Geometrically, the conditional pdf of $\mathbf{X}_B$ is a (rescaled) section of the joint pdf, which passes through the known point $\mathbf{x}_A$, see Figure 2.7. Equivalently, we could represent the conditional distribution with the respective cumulative density function or characteristic function, but the representation would be less intuitive.

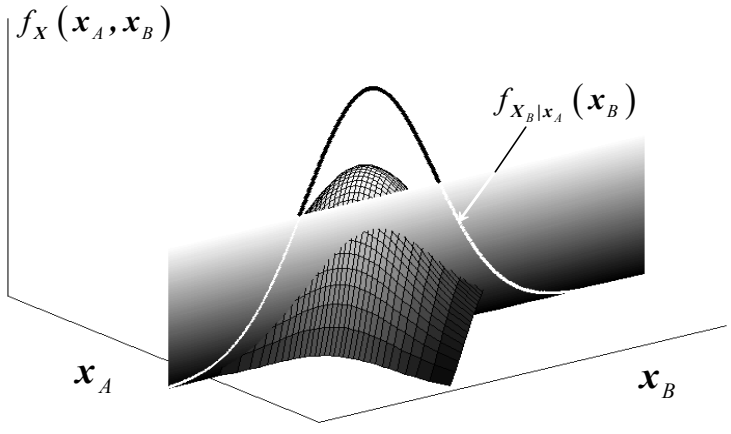

Fig. 2.7. Conditional probability density function

In our example, dividing the joint pdf of the two stock prices (2.7) by the marginal pdf of the first stock price (2.23) and simplifying, we obtain:

$$f_{X_2|x_1}(x_2) = \frac{1}{\sqrt{2\pi\sigma_C^2}} e^{-\frac{1}{2\sigma_C^2}(x_2-\mu_C)^2}, \tag{2.41}$$

where

$$\mu_C \equiv \widetilde{x}_2 + \sqrt{\frac{2}{5}}(x_1 - \widetilde{x}_1), \quad \sigma_C^2 \equiv \frac{3}{10}, \tag{2.42}$$

and where $(\widetilde{x}_1, \widetilde{x}_2)$ are the current prices (2.2). This computation is a specific instance of the more general result (2.173). The conditional pdf of the second stock price depends explicitly on the value x_1 of the first stock price, which is known by assumption.

From (2.40) we derive *Bayes' rule*, which is of the utmost importance in many financial applications:

$$\begin{aligned} f_{\mathbf{X}_A|\mathbf{x}_B}(\mathbf{x}_A) &= \frac{f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B)}{\int f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B)\, d\mathbf{x}_A} \\ &= \frac{f_{\mathbf{X}_B|\mathbf{x}_A}(\mathbf{x}_B)\, f_{\mathbf{X}_A}(\mathbf{x}_A)}{\int f_{\mathbf{X}_B|\mathbf{x}_A}(\mathbf{x}_B)\, f_{\mathbf{X}_A}(\mathbf{x}_A)\, d\mathbf{x}_A}. \end{aligned} \tag{2.43}$$

Bayes' rule expresses the conditional distribution of $\mathbf{X}_A$ given $\mathbf{x}_B$ in terms of the conditional distribution of $\mathbf{X}_B$ given $\mathbf{x}_A$ and the marginal distribution of $\mathbf{X}_A$.

At this point we have the tools to introduce the concept of (in)dependence among random variables. Splitting the multivariate random variable $\mathbf{X}$ into two sub-sets $(\mathbf{X}_A, \mathbf{X}_B)$ as in (2.39), we say that $\mathbf{X}_B$ is *independent* of $\mathbf{X}_A$ if the conditional distribution of $\mathbf{X}_B$ given $\mathbf{x}_A$ does not contain any more information than the marginal distribution of $\mathbf{X}_B$. More precisely, the variable $\mathbf{X}_B$ is independent of the variable $\mathbf{X}_A$ if for arbitrary functions g and h the marginal distribution of $g(\mathbf{X}_B)$ and the conditional distribution of $g(\mathbf{X}_B)$ given $h(\mathbf{x}_A)$ are the same.

The two stock prices in our example are independent if knowing the price (or the return, or any other function) of one stock does not add information regarding the distribution of the other stock and viceversa.

We can check for independence among variables in terms of their probability density function. Indeed, it can be proved that the mutual independence of $\mathbf{X}_A$ and $\mathbf{X}_B$ is equivalent to the joint pdf of $\mathbf{X}_A$ and $\mathbf{X}_B$ being the product of the marginal pdf of $\mathbf{X}_A$ and the marginal pdf of $\mathbf{X}_B$:

$$(\mathbf{X}_A, \mathbf{X}_B) \text{ independent} \Leftrightarrow f_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B) = f_{\mathbf{X}_A}(\mathbf{x}_A) f_{\mathbf{X}_B}(\mathbf{x}_B), \tag{2.44}$$

see Shirayaev (1989).

In particular, (2.40) and (2.44) imply the intuitive result that the marginal distribution of a variable and its conditional distribution given the realization of an independent variable are the same:

$$(\mathbf{X}_A, \mathbf{X}_B) \text{ independent} \Rightarrow f_{\mathbf{X}_B|\mathbf{x}_A}(\mathbf{x}_B) = f_{\mathbf{X}_B}(\mathbf{x}_B). \tag{2.45}$$

In our example the two stock prices are not independent, since the conditional distribution of one stock price (2.41) depends on the other stock price.

Similarly, we can check for independence among variables in terms of their cumulative distribution function. Indeed, substituting (2.44) in the definition of the cdf (2.9) and integrating, the mutual independence of $\mathbf{X}_A$ and $\mathbf{X}_B$ is equivalent to the joint cdf of $\mathbf{X}_A$ and $\mathbf{X}_B$ being the product of the marginal cdf of $\mathbf{X}_A$ and the marginal cdf of $\mathbf{X}_B$:

$$(\mathbf{X}_A, \mathbf{X}_B) \text{ independent} \Leftrightarrow F_{\mathbf{X}}(\mathbf{x}_A, \mathbf{x}_B) = F_{\mathbf{X}_A}(\mathbf{x}_A) F_{\mathbf{X}_B}(\mathbf{x}_B). \tag{2.46}$$

Finally, we can check for independence among variables in terms of their characteristic function. Indeed, from (2.44) for any functions g and h the expectation operator $(B.56)$ can be factored as follows:

$$\begin{aligned}(\mathbf{X}_A, \mathbf{X}_B) \text{ independent} \Rightarrow& \\ \mathrm{E}\{g(\mathbf{X}_A) h(\mathbf{X}_B)\} &= \mathrm{E}\{g(\mathbf{X}_A)\} \mathrm{E}\{h(\mathbf{X}_B)\}.\end{aligned} \tag{2.47}$$

Therefore, from the definition of the characteristic function (2.13) we obtain that the mutual independence of $\mathbf{X}_A$ and $\mathbf{X}_B$ is equivalent to the joint characteristic function of $\mathbf{X}_A$ and $\mathbf{X}_B$ being the product of the marginal cf of $\mathbf{X}_A$ and the marginal cf of $\mathbf{X}_B$:

$$(\mathbf{X}_A, \mathbf{X}_B) \text{ independent} \Leftrightarrow \phi_{\mathbf{X}}(\boldsymbol{\omega}_A, \boldsymbol{\omega}_B) = \phi_{\mathbf{X}_A}(\boldsymbol{\omega}_A)\, \phi_{\mathbf{X}_B}(\boldsymbol{\omega}_B). \quad (2.48)$$

2.4 Shape summary statistics

In this section we discuss multivariate parameters of location and dispersion that summarize the main properties of a multivariate distribution. As in the one-dimensional case, these parameters provide an easy-to-interpret picture of the main properties of a multivariate distribution. After discussing their definition and properties we present a geometrical interpretation that recurs throughout the book. We conclude with a brief introduction to higher-order summary statistics.

2.4.1 Location

Consider an N-dimensional random variable $\mathbf{X}$. Our purpose is to summarize the whole distribution of $\mathbf{X}$ into one *location parameter* $\text{Loc}\{\mathbf{X}\}$, similarly to what we did in Section 1.2.1 for the univariate case.

Theory

As in the one-dimensional case, we require that the location parameter display some intuitive features. For instance, if the distribution is peaked around a specific value, the location parameter should be close to that peak. In particular, a constant $\mathbf{m}$ can be seen as an infinitely peaked random variable, see $(B.22)$ and comments thereafter. Thus the location of a constant should be the constant itself:

$$\text{Loc}\{\mathbf{m}\} = \mathbf{m}. \quad (2.49)$$

This implies that the location parameter must be an N-dimensional vector.

Furthermore, consider a generic *affine transformation*:

$$\mathbf{X} \mapsto \mathbf{Y} \equiv \mathbf{a} + \mathbf{B}\mathbf{X}, \quad (2.50)$$

where $\mathbf{a}$ is a vector and $\mathbf{B}$ is a conformable matrix. A sensible parameter of location should track any invertible affine transformation of the original variable, i.e. a transformation such as (2.50), where $\mathbf{B}$ is an invertible matrix. In other words, if $\mathbf{B}$ is invertible, the location parameters should satisfy the following property:

$$\text{Loc}\{\mathbf{a} + \mathbf{B}\mathbf{X}\} = \mathbf{a} + \mathbf{B}\,\text{Loc}\{\mathbf{X}\}. \quad (2.51)$$

Property (2.51) is called the *affine equivariance* of the location parameter. For the rationale behind this requirement refer to the one dimensional case (1.24).

Examples

An example of location parameter is the multivariate *mode*, defined as the multivariate generalization of (1.30), namely as the highest peak of the joint probability density function:

$$\text{Mod}\{\mathbf{X}\} \equiv \underset{\mathbf{x}\in\mathbb{R}^N}{\text{argmax}}\{f_{\mathbf{X}}(\mathbf{x})\}. \tag{2.52}$$

We prove in Appendix www.2.5 that the mode is affine equivariant, i.e. it satisfies (2.51).

Consider our leading example (2.7) of two stock prices. From the first-order conditions on the joint pdf we obtain:

$$\text{Mod}\{\mathbf{X}\} = (\widetilde{x}_1, \widetilde{x}_2)', \tag{2.53}$$

where $(\widetilde{x}_1, \widetilde{x}_2)$ are the current prices (2.2). This is a specific instance of the more general result (2.158).

Another multivariate location parameter is the multivariate *expected value*, defined as the juxtaposition of the expected value (1.25) of the marginal distribution of each entry:

$$\text{E}\{\mathbf{X}\} \equiv (\text{E}\{X_1\}, \ldots, \text{E}\{X_N\})'. \tag{2.54}$$

Indeed, we prove in Appendix www.2.6 that the expected value is affine equivariant, i.e. it satisfies (2.51).

In our example (2.7) we have:

$$\text{E}\{\mathbf{X}\} = (\widetilde{x}_1, \widetilde{x}_2)', \tag{2.55}$$

where $(\widetilde{x}_1, \widetilde{x}_2)$ are the current prices (2.2). This is a specific instance of the more general result (2.158).

On the other hand, the juxtaposition of the median, or any other quantile, of each entry of a random variable does not satisfy (2.51) and therefore it does not define a suitable location parameter.

Mode and expected value might not be defined: the expectation integral might not converge in the case of the expected value, and the maximum of the probability density function might not be unique in the case of the mode. If they are defined, they both represent suitable location parameters.

Nevertheless, the expected value (2.54) is the benchmark multivariate location parameter.

In the first place, as in the one-dimensional case the expected value is a global parameter that includes information from the whole distribution,

whereas the mode is a local parameter that depends on the value of the probability density function at one single point.

Secondly, the expected value enjoys a purely multivariate feature: the affine equivariance property holds for generic, i.e. not necessarily invertible, affine transformations. In other words the following equality holds for *any* conformable matrix $\widetilde{\mathbf{B}}$ and vector $\widetilde{\mathbf{a}}$:

$$\mathrm{E}\left\{\widetilde{\mathbf{a}}+\widetilde{\mathbf{B}}\mathbf{X}\right\}=\widetilde{\mathbf{a}}+\widetilde{\mathbf{B}}\,\mathrm{E}\{\mathbf{X}\}, \tag{2.56}$$

see Appendix www.2.6. This is not true for other parameters of location.

For example the mode of the sum of two variables in general is not the sum of the modes:

$$\mathrm{Mod}\{X+Y\}\neq\mathrm{Mod}\{X\}+\mathrm{Mod}\{Y\}. \tag{2.57}$$

This implies that the affine equivariance for generic affine transformations (2.56) does not hold for the mode even in the simple case $\mathbf{a}\equiv 0$ and $\mathbf{B}\equiv(1,1)$.

Finally, whenever the characteristic function of $\mathbf{X}$ is known and analytical, i.e. it can be recovered entirely from its Taylor series expansion, computing the expected value is straightforward, as we show in Appendix www.2.10.

2.4.2 Dispersion

Consider an N-dimensional random variable $\mathbf{X}$. Here we extend to a multivariate environment the concept of *dispersion parameter* discussed in Section 1.2.2 for the univariate case.

Theory

As in the univariate case discussed in Chapter 1, we require that the dispersion parameter behaves suitably under invertible affine transformations:

$$\mathbf{X}\mapsto\mathbf{Y}\equiv\mathbf{a}+\mathbf{B}\mathbf{X}, \tag{2.58}$$

where $\mathbf{a}$ is a vector and $\mathbf{B}$ is a conformable invertible matrix.

To determine the nature of the required behavior, we recall the definition (1.35) of the absolute value of the z-score of the variable X in the univariate case:

$$|Z_X|\equiv\sqrt{(X-\mathrm{Loc}\{X\})\,\frac{1}{\mathrm{Dis}\{X\}^2}\,(X-\mathrm{Loc}\{X\})}. \tag{2.59}$$

In that context, the dispersion parameter $\mathrm{Dis}\{X\}$ is properly defined if the absolute value of the z-score is unaffected by affine transformations:

$$|Z_{a+bX}|=|Z_X|, \tag{2.60}$$

see (1.36).

To generalize the absolute value of the z-score to a multivariate environment, we introduce the *Mahalanobis distance* of the point $\mathbf{x}$ from the point $\boldsymbol{\mu}$ through the metric $\boldsymbol{\Sigma}$, denoted and defined as follows:

$$\text{Ma}\left(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \equiv \sqrt{\left(\mathbf{x}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}-\boldsymbol{\mu}\right)}, \tag{2.61}$$

where the metric $\boldsymbol{\Sigma}$ is a symmetric and positive matrix. The points $\mathbf{x}$ which share the same Mahalanobis distance from $\boldsymbol{\mu}$ lie on the surface of an ellipsoid centered in $\boldsymbol{\mu}$, see $(A.73)$. The larger (the eigenvalues of) $\boldsymbol{\Sigma}$, the smaller the Mahalanobis distance of the generic point $\mathbf{x}$ from the center $\boldsymbol{\mu}$. Therefore the matrix $\boldsymbol{\Sigma}$ indeed provides a metric to measure distances.

Comparing (2.59) with (2.61) we see that in a multivariate environment the absolute value of the z-score is replaced by the Mahalanobis distance from the location parameter through the metric provided by the yet to be defined "squared" dispersion parameter:

$$\text{Ma}_{\mathbf{X}} \equiv \text{Ma}\left(\mathbf{X}, \text{Loc}\left\{\mathbf{X}\right\}, \text{DisSq}\left\{\mathbf{X}\right\}\right). \tag{2.62}$$

We remark that considering (2.62) is intuitive, since a natural formulation of the dispersion of the variable $\mathbf{X}$ requires the dispersion parameter to represent a metric, i.e. a distance, between the variable and its location parameter.

In this context the dispersion parameter $\text{DisSq}\left\{\mathbf{X}\right\}$ is properly defined if it satisfies two properties.

In the first place $\text{DisSq}\left\{\mathbf{X}\right\}$ must be a symmetric and positive matrix, in such a way to define a metric in (2.62).

Secondly, $\text{DisSq}\left\{\mathbf{X}\right\}$ must be such that the Mahalanobis distance (2.62) is invariant under invertible affine transformations:

$$\text{Ma}_{\mathbf{a}+\mathbf{B}\mathbf{X}} = \text{Ma}_{\mathbf{X}}. \tag{2.63}$$

Given the affine equivariant properties of the location parameter (2.51), this is true if and only if for all invertible affine transformations (2.58) the dispersion parameter satisfies:

$$\text{DisSq}\left\{\mathbf{a}+\mathbf{B}\mathbf{X}\right\} = \mathbf{B}\,\text{DisSq}\left\{\mathbf{X}\right\}\mathbf{B}'. \tag{2.64}$$

We call this property the *affine equivariance* of a multivariate dispersion parameter.

To summarize, a *dispersion matrix*, or *dispersion parameter*, or *scatter matrix* or *scatter parameter* is a symmetric and positive matrix $\text{DisSq}\left\{\mathbf{X}\right\}$ that is affine equivariant, i.e. it satisfies (2.64).

Examples

An example of scatter matrix is the *modal dispersion*:

$$\text{MDis}\{\mathbf{X}\} \equiv -\left(\left. \frac{\partial^2 \ln f_{\mathbf{X}}}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{\mathbf{x}=\text{Mod}\{\mathbf{X}\}} \right)^{-1}, \tag{2.65}$$

see e.g. O'Hagan (1994). In Appendix www.2.5 we prove that the modal dispersion is indeed a scatter matrix, i.e. it is a symmetric and positive matrix that is affine equivariant. The rationale behind the modal dispersion follows from a second-order Taylor expansion of the pdf $f_{\mathbf{X}}$ around its mode, see (1.39) for the univariate case: the larger in absolute value the (always negative) second derivative in (2.65), the thinner the probability density function of $\mathbf{X}$ around its mode and thus the less disperse the distribution.

Consider our leading example (2.7). From a direct computation of the second derivatives of the log-pdf at the mode (2.53) we obtain:

$$\text{MDis}\{\mathbf{X}\} = \begin{pmatrix} 1/2 & 1/\sqrt{10} \\ 1/\sqrt{10} & 1/2 \end{pmatrix}. \tag{2.66}$$

Another example of scatter parameter is the *covariance* matrix, defined as follows:

$$\text{Cov}\{\mathbf{X}\} \equiv \text{E}\left\{ (\mathbf{X} - \text{E}\{\mathbf{X}\}) (\mathbf{X} - \text{E}\{\mathbf{X}\})' \right\}, \tag{2.67}$$

or component-wise:

$$\begin{aligned} \text{Cov}\{X_m, X_n\} &\equiv [\text{Cov}\{\mathbf{X}\}]_{mn} \\ &\equiv \text{E}\{ (X_m - \text{E}\{X_m\}) (X_n - \text{E}\{X_n\}) \}. \end{aligned} \tag{2.68}$$

In Appendix www.2.6 we prove that the covariance is a scatter matrix, i.e. it is symmetric, positive and affine equivariant.

In our leading example (2.7) we obtain:

$$\text{Cov}\{\mathbf{X}\} = \begin{pmatrix} 1/2 & 1/\sqrt{10} \\ 1/\sqrt{10} & 1/2 \end{pmatrix}. \tag{2.69}$$

This is a specific instance of a more general result, see Section 2.6.2.

Modal dispersion and covariance matrix might not be defined: the expectation integral might not converge in the case of the covariance, and the mode might not be unique in the case of the modal dispersion. When they are defined, they both represent suitable dispersion parameters.

Nevertheless, the covariance is the benchmark multivariate scatter parameter.

In the first place, like the variance in the one-dimensional case, the covariance is a global parameter that includes information from the whole distribution, whereas the modal dispersion is a local parameter that depends on

the shape of the probability density function around one single point, i.e. the mode.

Secondly, from the factorization of the expectation operator in the presence of independent variables (2.47) and the component-wise definition of the covariance matrix (2.68) we obtain that the covariance of independent variables is null:

$$(X_m, X_n) \text{ independent} \Rightarrow \text{Cov}\{X_n, X_n\} = 0. \tag{2.70}$$

This result motivates the name "covariance", as independent variables do not "co-vary".

In the third place, whenever the characteristic function of $\mathbf{X}$ is known and analytical, i.e. it can be recovered entirely from its Taylor series expansion, computing the covariance matrix is straightforward, as we show in Appendix www.2.10.

Finally, the affine equivariance property (2.64) holds in the case of the covariance even for generic, i.e. not necessarily invertible, affine transformations. In other words, the following identity holds for *any* conformable matrix $\widetilde{\mathbf{B}}$ and vector $\widetilde{\mathbf{a}}$:

$$\text{Cov}\left\{\widetilde{\mathbf{a}} + \widetilde{\mathbf{B}}\mathbf{X}\right\} = \widetilde{\mathbf{B}}\,\text{Cov}\{\mathbf{X}\}\,\widetilde{\mathbf{B}}', \tag{2.71}$$

see Appendix www.2.6. This is not true for other dispersion parameters.

For example, since from (2.57) the mode is not affine equivariant for non-invertible transformations, neither can be the modal dispersion.

The generic affine equivariance (2.56) and (2.71) of the expected value and covariance matrix respectively also allows us to build a dispersion parameter with a more intuitive "bottom up" approach.

Indeed, consider a specific type of non-invertible affine transformations, i.e. a linear combinations $\boldsymbol{\alpha}'\mathbf{X}$, where $\boldsymbol{\alpha}$ is an N-dimensional vector of constants. A linear combination of random variables is a univariate random variable. Therefore we can compute the dispersion parameter (1.40) defined in terms of the expectation operator:

$$\text{Dis}\{\boldsymbol{\alpha}'\mathbf{X}\} \equiv \left(\text{E}\left\{|\boldsymbol{\alpha}'(\mathbf{X} - \text{E}\{\mathbf{X}\})|^p\right\}\right)^{\frac{1}{p}}. \tag{2.72}$$

For a general value of p, there exists no result concerning linear combinations that involve equalities. Nevertheless, in the case $p \equiv 2$ the dispersion in (2.72) becomes the standard deviation and a few algebraic manipulations show that there exists a matrix $\mathbf{S}$ such that

$$\text{Sd}\{\boldsymbol{\alpha}'\mathbf{X}\} = \sqrt{\boldsymbol{\alpha}'\mathbf{S}\boldsymbol{\alpha}}. \tag{2.73}$$

From $(B.65)$ and $(B.68)$ the matrix $\mathbf{S}$ coincides with the covariance (2.67). In particular, from (2.73) we obtain that the diagonal elements of the covariance matrix are the variances of the marginal distributions of each entry:

$$\text{Cov}\{X_n, X_n\} = (\text{Sd}\{X_n\})^2 = \text{Var}\{X_n\}. \tag{2.74}$$

2.4.3 Location-dispersion ellipsoid

Consider an N-dimensional random variable $\mathbf{X}$. In this section we propose a graphical interpretation of the parameters of location and dispersion of $\mathbf{X}$. In particular, we will develop our discussion around the benchmark parameters, i.e. the expected value $\mathrm{E}\{\mathbf{X}\}$ defined in (2.54) and the covariance matrix $\mathrm{Cov}\{\mathbf{X}\}$ defined in (2.67), which we denote here as E and Cov respectively to ease the notation.

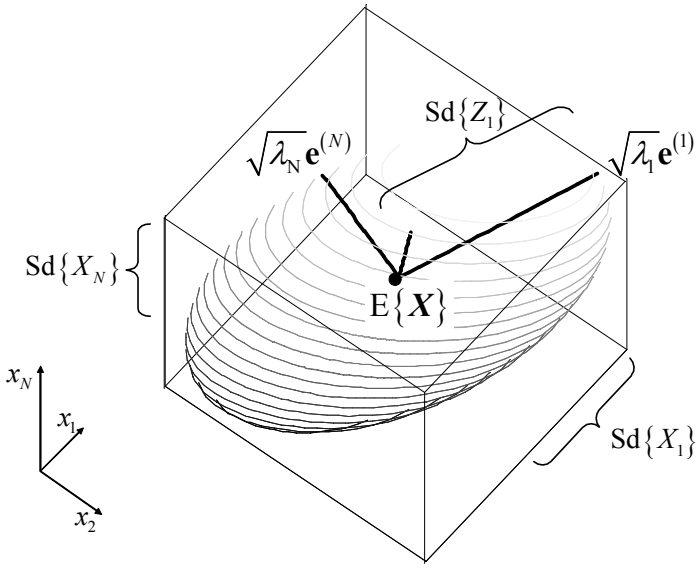

Fig. 2.8. Location-dispersion ellipsoid

A generic representation of expected value and covariance must convey all the information contained in these parameters. On the other hand, a geometrical representation must also provide support to intuition.

We state here and motivate in the sequel that we can effectively represent geometrically E and Cov by means of the *location-dispersion ellipsoid*, defined as follows:

$$\mathcal{E}_{\mathrm{E},\mathrm{Cov}} \equiv \left\{\mathbf{x} \text{ such that } (\mathbf{x}-\mathrm{E})' \,\mathrm{Cov}^{-1}\,(\mathbf{x}-\mathrm{E}) \leq 1\right\}, \tag{2.75}$$

see Figure 2.8.

First of all, we remark that this is indeed the implicit equation of an ellipsoid. The expected value is a vector and the covariance matrix is symmetric and positive definite. Therefore, from $(A.73)$ the locus $\mathcal{E}_{\mathrm{E},\mathrm{Cov}}$ is an ellipsoid centered in the location parameter E.

The fact that the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ is centered in the expected value E shows that on the one hand the ellipsoid conveys all the information about E, and on the other hand the ellipsoid supports intuition regarding the meaning of E, which is the average location of the random variable $\mathbf{X}$.

As far as the dispersion parameter Cov is concerned, we already know from the discussion in Appendix A.5 that the ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ conveys all the information contained in the covariance matrix Cov.

To show that it also support intuition regarding the dispersion properties of the random variable $\mathbf{X}$ we rephrase in this context the analysis of Appendix A.5. Consider the spectral decomposition $(A.70)$ of the covariance matrix:

$$\mathrm{Cov}\{\mathbf{X}\} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}'. \tag{2.76}$$

In this expression $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues of the covariance sorted in decreasing order:

$$\boldsymbol{\Lambda} \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_N); \tag{2.77}$$

and $\mathbf{E}$ is the juxtaposition of the respective eigenvectors:

$$\mathbf{E} \equiv \left(\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(N)}\right), \tag{2.78}$$

which satisfies $\mathbf{E}\mathbf{E}' = \mathbf{I}_N$, the identity matrix.

We know from Appendix A.5 that the principal axes of the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ are parallel to the the eigenvectors $\left\{\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(N)}\right\}$ of the covariance. On the other hand, in this context the eigenvectors define very special directions, namely the directions along which the randomness in $\mathbf{X}$ displays zero covariance. In other words, consider the following random variable:

$$\mathbf{Z} \equiv \mathbf{E}'\mathbf{X} = \begin{pmatrix} \left[\mathbf{e}^{(1)}\right]'\mathbf{X} \\ \vdots \\ \left[\mathbf{e}^{(N)}\right]'\mathbf{X} \end{pmatrix}. \tag{2.79}$$

Each entry of the vector $\mathbf{Z}$ is the projection of the random variable $\mathbf{X}$ on one eigenvector. From $\mathbf{E}\mathbf{E}' = \mathbf{I}_N$, for any $n \neq m$ we have:

$$\mathrm{Cov}\{Z_m, Z_n\} = \left[\mathbf{e}^{(m)}\right]'\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}'\left[\mathbf{e}^{(n)}\right] = [\boldsymbol{\Lambda}]_{mn} = 0. \tag{2.80}$$

Thus the principal axes of the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ define the directions along which the randomness in $\mathbf{X}$ displays zero covariance.

Furthermore from Appendix A.5 the length of the principal axes of the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ are the square root of the eigenvalues of the covariance. On the other hand, in this context the eigenvalues have a very special meaning, namely they represent the variance of $\mathbf{X}$ along the direction of the eigenvectors:

$$\mathrm{Var}\{Z_n\} = \left[\mathbf{e}^{(n)}\right]' \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}' \left[\mathbf{e}^{(n)}\right] = \lambda_n. \tag{2.81}$$

Thus from (2.74) the length of the principal axes of the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ represent the standard deviation of $\mathbf{X}$ along the direction of the principal axes.

In particular, from (A.68) the first eigenvalue corresponds the maximum variance achievable with a projection:

$$\lambda_1 = \max_{\|\mathbf{e}\|=1} \{\mathrm{Var}\{\mathbf{e}'\mathbf{X}\}\}; \tag{2.82}$$

and the first eigenvector $\mathbf{e}^{(1)}$ is the direction of maximal variation, i.e. it satisfies:

$$\mathbf{e}^{(1)} = \operatorname*{argmax}_{\|\mathbf{e}\|=1} \{\mathrm{Var}\{\mathbf{e}'\mathbf{X}\}\}. \tag{2.83}$$

Similarly, from (A.69) the last eigenvalue corresponds the minimum variance achievable with a projection:

$$\lambda_N = \min_{\|\mathbf{e}\|=1} \{\mathrm{Var}\{\mathbf{e}'\mathbf{X}\}\}; \tag{2.84}$$

and the last eigenvector $\mathbf{e}^{(N)}$ is the direction of minimal variation, i.e. it satisfies:

$$\mathbf{e}^{(N)} = \operatorname*{argmin}_{\|\mathbf{e}\|=1} \{\mathrm{Var}\{\mathbf{e}'\mathbf{X}\}\}. \tag{2.85}$$

Moreover, the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ is a suitable generalization of the one-dimensional location-dispersion bar defined in (1.53). Indeed, consider the rectangle with sides parallel to the reference axes of $\mathbb{R}^N$ which enshrouds the ellipsoid, see Figure 2.8. We prove in Appendix www.2.8 that the generic n-th side of this rectangle is centered on the expected value $\mathrm{E}\{X_n\}$ of the n-th marginal component and is long twice the standard deviation $\mathrm{Sd}\{X_n\}$ of the n-th marginal component. In other words, the enshrouding rectangle is defined by the following set of N equations:

$$\mathrm{E}\{X_n\} - \mathrm{Sd}\{X_n\} \leq x_n \leq \mathrm{E}\{X_n\} + \mathrm{Sd}\{X_n\}. \tag{2.86}$$

Each of these equations represents the location-dispersion bar (1.53) of the respective marginal distribution.

Finally, the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ is, among all the ellipsoids of equal volume, the one that contains the highest probability of occurrence of the random variable $\mathbf{X}$ within its boundaries. To make this statement precise, we consider the locus:

$$\mathcal{E}^q_{\mathrm{E,Cov}} \equiv \left\{\mathbf{x} \text{ such that } (\mathbf{x}-\mathrm{E})' \mathrm{Cov}^{-1} (\mathbf{x}-\mathrm{E}) \leq q^2\right\}. \tag{2.87}$$

This locus represents a rescaled version of the location-dispersion ellipsoid (2.75), where all the principal axis are multiplied by a factor q, see Figure 2.9. In Appendix www.2.9 we prove the following results.

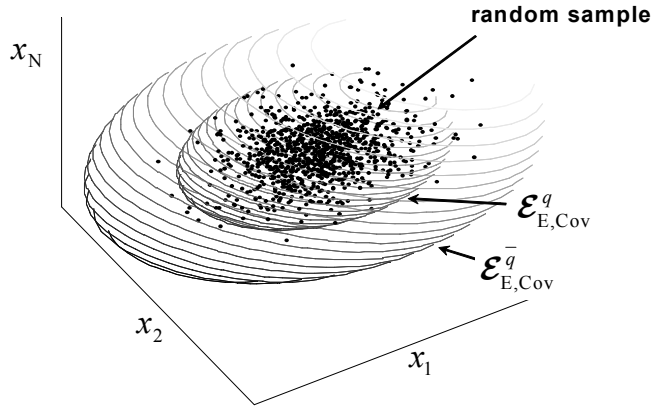

Fig. 2.9. Multivariate Chebyshev inequality

By the *Chebyshev inequality* for *any* vector $\mathbf{v}$ and *any* symmetric and positive matrix $\mathbf{U}$ the probability that observations occur outside the ellipsoid $\mathcal{E}^q_{\mathbf{v},\mathbf{U}}$ with principal axes proportional to q decays as the square of q:

$$\mathbb{P}\left\{\mathbf{X} \notin \mathcal{E}^q_{\mathbf{v},\mathbf{U}}\right\} \leq \frac{a_{\mathbf{v},\mathbf{U}}}{q^2}, \tag{2.88}$$

where the constant a is the expected squared Mahalanobis distance (2.61) of the random variable $\mathbf{X}$ from the point $\mathbf{v}$ through the metric $\mathbf{U}$:

$$a_{\mathbf{v},\mathbf{U}} \equiv \mathrm{E}\left\{\mathrm{Ma}^2\left(\mathbf{X},\mathbf{v},\mathbf{U}\right)\right\}. \tag{2.89}$$

Nevertheless, if we set $\mathbf{v}$ equal to the expected value and $\mathbf{U}$ equal to the covariance matrix in (2.88), the function a reaches a minimum, and is equal to the dimension the random variable $\mathbf{X}$. Therefore the probability of $\mathbf{X}$ not occurring in the ellipsoid is uniformly the minimum possible and reads:

$$\mathbb{P}\left\{\mathbf{X} \notin \mathcal{E}^q_{\mathrm{E,Cov}}\right\} \leq \frac{N}{q^2}. \tag{2.90}$$

In other words, the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ is the one ellipsoid among those of equal volume that enshrouds the most probability.

2.4.4 Higher-order statistics

Similarly to the one-dimensional case discussed in Section 1.2.3, we can gain more insight into the statistical features of a multivariate distribution from

the moments of that distribution of order higher than the expected value and the covariance matrix.

To introduce the higher moments, we recall that the expected value is a vector, namely the vector of expectations of each entry of a multivariate random variable. On the other hand the covariance is a matrix, namely the matrix of (a simple function of) expectations of all the cross products of two entries.

The expectation operator (B.56) applied to the cross products of three, four, etc. entries can be organized in tensors, a straightforward generalization of the concept of vector and matrix, see (A.92) and comments thereafter for a quick review.

The *k-th raw moment* of a multivariate random variable $\mathbf{X}$ is a tensor of order k, defined as follows:

$$\mathrm{RM}^{\mathbf{X}}_{n_1\cdots n_k} \equiv \mathrm{E}\{X_{n_1}\cdots X_{n_k}\}. \tag{2.91}$$

This definition generalizes the one-dimensional raw-moment (1.47). In particular, the expected value (2.54) is the first raw moment.

The *k-th central moment* of a random variable is a location-independent version of the respective raw moment:

$$\mathrm{CM}^{\mathbf{X}}_{n_1\cdots n_k} \equiv \mathrm{E}\{(X_{n_1}-\mathrm{E}\{X_{n_1}\})\cdots(X_{n_k}-\mathrm{E}\{X_{n_k}\})\}. \tag{2.92}$$

This definition generalizes the one-dimensional central-moment (1.48). In particular, the covariance matrix (2.68) is the second central moment.

The central moments of a distribution are tensors that enjoy special transformation properties. For instance, from (2.71) the covariance matrix is equivariant under any, not necessarily invertible, affine transformation. From the linearity of the expectation operator (B.56) and the definition of the central moments (2.92), it follows that all the central moments are affine equivariant, in that for any M-dimensional vector $\mathbf{a}$ and any $M\times N$ matrix $\mathbf{B}$ the following relation holds:

$$\mathrm{CM}^{\mathbf{a}+\mathbf{B}\mathbf{X}}_{m_1\cdots m_k} = \sum_{n_1,\ldots,n_k=1}^{N} B_{m_1,n_1}\cdots B_{m_k,n_k}\,\mathrm{CM}^{\mathbf{X}}_{n_1\cdots n_k}. \tag{2.93}$$

For example, consider $\mathbf{a}\equiv\mathbf{0}$ and $\mathbf{B}'\equiv\mathbf{b}$, an N-dimensional vector. In this case the affine-equivariance property (2.93) yields the expression for the central moments (1.48) of the one-dimensional variable $\mathbf{b}'\mathbf{X}$. For instance, the third central moment reads:

$$\mathrm{CM}^{\mathbf{b}'\mathbf{X}}_{3} = \sum_{l,m,n=1}^{N} b_l b_m b_n\,\mathrm{CM}^{\mathbf{X}}_{lmn}. \tag{2.94}$$

Similarly to the univariate case, it is possible to define normalized version of the higher central moments.

The *co-skewness* is the following three-dimensional tensor:

$$\begin{aligned}\mathrm{Sk}\{X_l, X_m, X_n\} &\equiv [\mathrm{Sk}\{\mathbf{X}\}]_{lmn} \\ &\equiv \frac{\mathrm{CM}^{\mathbf{X}}_{lmn}}{\mathrm{Sd}\{X_l\}\,\mathrm{Sd}\{X_m\}\,\mathrm{Sd}\{X_n\}},\end{aligned} \tag{2.95}$$

which generalizes the univariate skewness (1.49). The co-skewness provides information on the symmetry of the distribution of $\mathbf{X}$. It is also possible to summarize the information provided by the co-skewness in one overall index of symmetry, see Mardia (1970).

The *co-kurtosis* is the following four-dimensional tensor:

$$\begin{aligned}\mathrm{Ku}\{X_l, X_m, X_n, X_p\} &\equiv [\mathrm{Ku}\{\mathbf{X}\}]_{lmnp} \\ &\equiv \frac{\mathrm{CM}^{\mathbf{X}}_{lmnp}}{\mathrm{Sd}\{X_l\}\,\mathrm{Sd}\{X_m\}\,\mathrm{Sd}\{X_n\}\,\mathrm{Sd}\{X_p\}},\end{aligned} \tag{2.96}$$

which generalizes the univariate kurtosis (1.51). The co-kurtosis provides information on the thickness of the tails of the distribution of $\mathbf{X}$. It is also possible to summarize the information provided by the co-kurtosis in one overall index of tail thickness, see Mardia (1970).

Computing the above summary statistics involves in general integrations. Nevertheless, whenever the characteristic function of $\mathbf{X}$ is known and analytical, i.e. it can be recovered entirely from its Taylor series expansion, we can compute these quantities by means of simple differentiation and some algebra, as we show Appendix www.2.10.

Nevertheless, the number of parameters in the higher moments grows as N^k, where N is the dimension of the multivariate distribution of $\mathbf{X}$ and k is the order of the moment. This number becomes intractable for $k > 2$ in any practical application.

2.5 Dependence summary statistics

The N entries of a random variable $\mathbf{X}$ display in general a complex dependence structure that it is important to monitor in view of hedging and managing risk. In this section we describe how to summarize in one number the dependence between two generic entries X_m and X_n. We refer the reader to references such as Nelsen (1999) for more results on this subject.

2.5.1 Measures of dependence

A *measure of dependence* $\mathrm{Dep}\{X_m, X_n\}$ between two random variables X_m and X_n should be a function of the distribution of the variables, normalized in such a way to make it easy to interpret, for instance as follows:

$$0 \leq \text{Dep}\{X_m, X_n\} \leq 1. \tag{2.97}$$

Furthermore, it should display a minimal set of intuitive features, such as the following:

1. Total independence represents one extreme of the spectrum of possible values:
$$(X_m, X_n) \text{ independent} \Leftrightarrow \text{Dep}\{X_m, X_n\} \equiv 0. \tag{2.98}$$
2. Total dependence represents the other extreme of the spectrum of possible values:
$$(X_m, X_n) \text{ co-monotonic} \Leftrightarrow \text{Dep}\{X_m, X_n\} \equiv 1, \tag{2.99}$$
where co-monotonicity is defined in (2.35).
3. The measure of dependence spots the core interdependence structure. In other words, assume that the random variable X_m is a deterministic invertible function of a random variable Y_m, i.e. they are in one-to-one correspondence, and that an analogous relation holds between X_n and another random variable Y_n. The dependence between the first set of variables should be the same as the dependence between the second set of variables:
$$\left.\begin{array}{c}(X_m, Y_m) \text{ one-to-one} \\ (X_n, Y_n) \text{ one-to-one}\end{array}\right\} \Rightarrow \text{Dep}\{X_m, X_n\} = \text{Dep}\{Y_m, Y_n\}. \tag{2.100}$$

In Section 2.2.2 we determined that the core interdependence structure between two generic variables X_m and X_n is driven by their copula. We recall that the copula is the joint distribution of the grades:

$$\begin{pmatrix} U_m \\ U_n \end{pmatrix} \equiv \begin{pmatrix} F_{X_m}(X_m) \\ F_{X_n}(X_n) \end{pmatrix}, \tag{2.101}$$

where F_{X_m} is the cumulative distribution function of X_m, see (2.28). Therefore in order to define a measure of dependence between $(X_m, X_n)'$ it is natural to turn to their copula, which we represent in terms of the cumulative distribution function F_{U_m,U_n}.

As far as the property on independence (2.98) is concerned, since the marginal distribution of each of the grades (2.101) is uniform, from (1.56) and (2.46) we see that X_m and X_n are independent if and only if their copula is uniformly distributed on the unit square, in which case the cumulative distribution function of the copula reads:

$$\Pi(u_m, u_n) \equiv u_m u_n, \tag{2.102}$$

see Figure 2.10.

Intuitively, the measure of dependence between X_m and X_n should be a distance between their copula, as represented by F_{U_m,U_n}, and the copula of two independent variables, as represented by (2.102): the larger the distance,

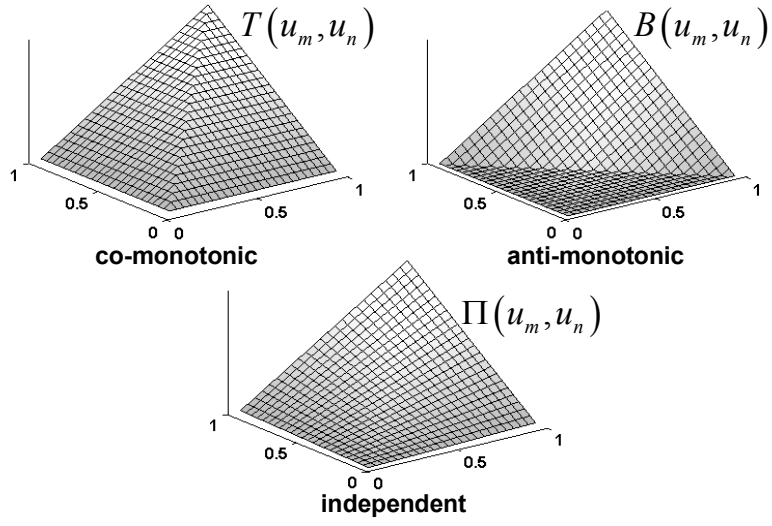

Fig. 2.10. Cumulative distribution function of special bivariate copulas

the higher the level of dependence. We can introduce a distance between these two functions by means of the L_p-norm (B.12), defined in this case on the unit square $\mathbb{Q} \equiv [0,1] \times [0,1]$. This way we obtain the following family of measures of dependence, called the *Schweizer-Wolff measures of dependence*:

$$\begin{aligned} \mathrm{SW}\{X_m, X_n\} &\equiv k_p \left\| F_{U_m,U_n} - \Pi \right\|_p \\ &\equiv k_p \left(\int_{\mathbb{Q}} \left| F_{U_m,U_n}(u_m,u_n) - \Pi(u_m,u_n) \right|^p du_m du_n \right)^{\frac{1}{p}}, \end{aligned} \tag{2.103}$$

where $p \geq 1$ and k_p is a constant yet to be defined. By construction, this measure satisfies (2.98), which is the first property required of a measure of dependence.

To determine the constant in (2.103) we turn to (2.99), the property of a generic measure of dependence which regards total dependence. It can be proved that the *Frechet-Hoeffding bounds* hold on the cumulative distribution function of a generic copula:

$$B(u_m,u_n) \leq F_{U_m,U_n}(u_m,u_n) \leq T(u_m,u_n), \tag{2.104}$$

where the "bottom" bound is defined as follows:

$$B(u_m,u_n) \equiv \max(u_m + u_n - 1, 0); \tag{2.105}$$

and the "top" bound is defined as follows:

$$T(u_m, u_n) \equiv \min(u_m, u_n). \tag{2.106}$$

The lower bound (2.105) is the cumulative distribution function of an "extreme" copula, namely the copula of $(X, -X)$, which does not depend on the distribution of X. On the other hand, the upper bound (2.106) is the cumulative distribution function of the another "extreme" copula, namely the copula of (X, X), which does not depend on the distribution of X. We plot in Figure 2.10 the cumulative distribution functions (2.105) and (2.106): notice how all the copulas in the figure satisfy (2.32).

In order for the Schweizer and Wolff measure of dependence (2.103) to satisfy (2.99) we need to normalize the constant k_p in its definition as follows:

$$k_p \equiv \frac{1}{\|B - \Pi\|_p} = \frac{1}{\|T - \Pi\|_p}. \tag{2.107}$$

To make sure that the Schweizer and Wolff measure of dependence normalized this way is a proper measure of dependence, we now turn to the last property (2.100). The copula is *almost* invariant under one-to-one transformations such as those that appear in (2.100). Indeed, consider two new variables:

$$\begin{pmatrix} Y_m \\ Y_n \end{pmatrix} \equiv \begin{pmatrix} g(X_m) \\ h(X_n) \end{pmatrix}, \tag{2.108}$$

where g and h are *increasing* invertible functions. In other, words (X_m, X_n) and (Y_m, Y_n) are co-monotonic, see (2.35). Now consider the copula of (Y_m, Y_n), which is the distribution of the grades:

$$\begin{pmatrix} V_m \\ V_n \end{pmatrix} \equiv \begin{pmatrix} F_{Y_m}(Y_m) \\ F_{Y_n}(Y_n) \end{pmatrix}. \tag{2.109}$$

Since from (2.38) the copula of (X_m, X_n) is the same as the copula of (Y_m, Y_n), the following relation holds:

$$F_{V_m,V_n} = F_{U_m,U_n}. \tag{2.110}$$

Therefore the Schweizer and Wolff measure of dependence automatically satisfies (2.100) for increasing one-to-one correspondences among the variables. If on the other hand one of the two variables, say Y_n, is an invertible *decreasing* function of X_n then (2.110) must be replaced by the following expression:

$$F_{V_m,V_n} = u_m - F_{U_m,U_n}(u_m, 1 - u_n). \tag{2.111}$$

Nevertheless, the integral in (2.103) is not affected by this change. Therefore the Schweizer and Wolff measure of dependence also satisfies (2.100).

Consider our leading example (2.7) of two stock prices (X_1, X_2). As in (2.36) consider a call option on the first stock price with strike K, i.e. the following random variable:

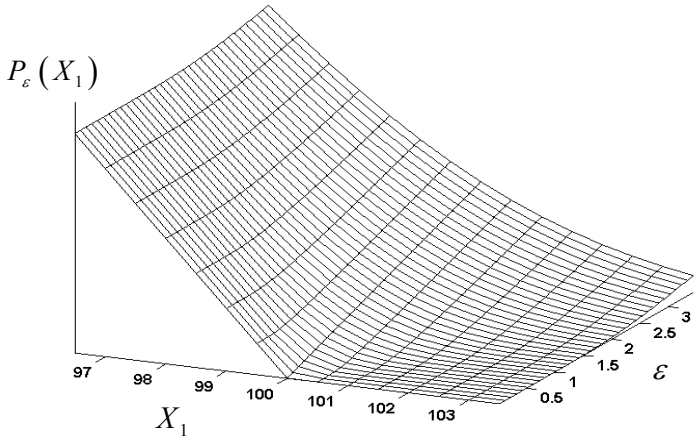

Fig. 2.11. Regularization of put option payoff

$$C_1 \equiv \max(X_1 - K, 0). \tag{2.112}$$

The call option payoff is a strictly increasing function of the underlying X_1, once we replace it by its regularized version as in (2.37), see also Figure 2.5. Consider the payoff of a put option on the first stock with strike K, i.e. the new random variable:

$$P_1 \equiv -\min(X_1 - K, 0). \tag{2.113}$$

The put option payoff is a strictly decreasing function of the underlying X_1, once we replace it by its regularized version by means of $(B.49)$:

$$P_\epsilon \equiv -\frac{(X_1 - K)}{2}\left(1 - \text{erf}\left(\frac{X_1 - K}{\sqrt{2\epsilon^2}}\right)\right) + \frac{\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2\epsilon^2}(X_1 - K)^2}, \tag{2.114}$$

see Figure 2.11 for the plot and Appendix www.2.7 for the proof.

We summarize the Schweizer and Wolff measure of dependence (2.103) between any two of the above securities in the following table:

SW	X_1	C_1	P_1	X_2
X_1	1	1	1	γ_p
C_1		1	1	γ_p
P_1			1	γ_p
X_2				1

(2.115)

The first stock price X_1 and the (regularized) call option payoff C_1 are co-monotonic functions: from (2.99) their measure of dependence is one. The

(regularized) put option payoff P_1 is an invertible function of the first stock price X_1: from (2.100) it is completely equivalent to X_1 and C_1. Therefore, the dependence of any of them with the second stock price X_2 is the same constant γ_p, which depends on the choice of $p \geq 1$ in the definition of the Schweizer and Wolff measure of dependence.

2.5.2 Measures of concordance

Due to (2.100), a measure of dependence does not distinguish between a random variable and any invertible function of that random variable. Nonetheless, in many applications it becomes important to separate increasing invertible functions from decreasing invertible functions.

We recall from (2.35) that two random variables X and Y are *co-monotonic* if

$$Y = g(X), \quad g \text{ invertible, increasing.} \tag{2.116}$$

Similarly, we define two random variables X and Y as *anti-monotonic* if

$$Y = g(X), \quad g \text{ invertible, decreasing.} \tag{2.117}$$

In our example, the (regularized) call option payoff (2.112) and the price of the underlying stock are co-monotonic; the (regularized) put option payoff (2.113) and the price of the underlying stock are anti-monotonic; the (regularized) call option payoff and the (regularized) put option payoff are anti-monotonic. The interests of an investor who owns the call option are very different than the interests of an investor who owns the put option. A dependence parameter such as the Schweizer and Wolff measure does not distinguish between calls and puts, as we see in (2.115).

Therefore, we are led to consider *measures of concordance*, that convey more information than measures of dependence. Ideally, a measure of concordance $\text{Con}\{X_m, X_n\}$ between two random variables X_m and X_n should be a function of their distribution, normalized in such a way to make it easy to interpret, for instance as follows:

$$-1 \leq \text{Con}\{X_m, X_n\} \leq 1. \tag{2.118}$$

Furthermore, it should displays a set of intuitive features:

1'. Independence represents the middle of the spectrum of possible values:

$$(X_m, X_n) \text{ independent} \Leftrightarrow \text{Con}\{X_m, X_n\} = 0. \tag{2.119}$$

2a'. Total concordance represents one extreme of the spectrum of possible values:

$$(X_m, X_n) \text{ co-monotonic} \Leftrightarrow \text{Con}\{X_m, X_n\} = 1. \tag{2.120}$$

2b'. Total discordance represents the other extreme of the spectrum of possible values:

$$(X_m, X_n) \text{ anti-monotonic} \Leftrightarrow \text{Con}\{X_m, X_n\} = -1. \tag{2.121}$$

3. The measure of concordance spots the core interdependence structure:

$$\left.\begin{array}{c}(X_m, Y_m) \text{ co-monotonic} \\ (X_n, Y_n) \text{ co-monotonic}\end{array}\right\} \Rightarrow \text{Con}\{X_m, X_n\} = \text{Con}\{Y_m, Y_n\}. \tag{2.122}$$

4. Concordance and discordance play a symmetric role:

$$\text{Con}\{X_m, -X_n\} = -\text{Con}\{X_m, X_n\}. \tag{2.123}$$

A comparison with the properties of the measures of dependence shows that a measure of concordance satisfying Properties 1'-4 would indeed convey all the information contained in a measure of dependence, and more.

Unfortunately, (2.119) cannot be satisfied together with the other properties. Intuitively, if we want to measure the "direction" of the dependence between two variables with a single number, variables that in some scenarios are concordant and in some other scenarios are discordant display the same amount of concordance as independent variables, although they are not. Therefore, (2.119) must be weakened as follows:

1. Independence implies the middle of the spectrum of possible values:

$$(X_m, X_n) \text{ independent} \Rightarrow \text{Con}\{X_m, X_n\} = 0. \tag{2.124}$$

For similar reasons (2.120) and (2.121) must be weakened as follows:

2a. Total concordance implies one extreme of the spectrum of possible values:

$$(X_m, X_n) \text{ co-monotonic} \Rightarrow \text{Con}\{X_m, X_n\} = 1. \tag{2.125}$$

2b. Total discordance implies the other extreme of the spectrum of possible values:

$$(X_m, X_n) \text{ anti-monotonic} \Rightarrow \text{Con}\{X_m, X_n\} = -1. \tag{2.126}$$

Just like in the case of measures of dependence, to define measures of concordance between X_m and X_n we turn to their copula, i.e. the joint distribution of the grades:

$$\begin{pmatrix} U_m \\ U_n \end{pmatrix} \equiv \begin{pmatrix} F_{X_m}(X_m) \\ F_{X_n}(X_n) \end{pmatrix}, \tag{2.127}$$

which we represent in terms of its cumulative distribution function F_{U_m,U_n}, or its probability density function f_{U_m,U_n}.

A popular measure of concordance is *Kendall's tau*. Kendall's tau is a normalized weighed average of the distance with sign of the cdf of the copula of (X_m, X_n) from the cdf (2.102) of the copula of two independent variables:

$$\tau\{X_m, X_n\} \equiv 4 \int_{\mathbb{Q}} \left(F_{U_m,U_n}(u_m, u_n) - \Pi(u_m, u_n)\right) f_{U_m,U_n}(u_m, u_n)\, du_m du_n, \tag{2.128}$$

where $\mathbb{Q} \equiv [0,1] \times [0,1]$ is the unit square.

This definition reminds us of the Schweizer and Wolff measure of dependence (2.103), but it is different in two respects. One difference is minor: Kendall's tau is a *weighted* average of the difference of the two functions, the weights being provided by the pdf of the copula. The second difference is conceptually more relevant: Kendall's tau evaluates the difference *with sign*, not in absolute value. It can be checked that due to this last feature, Kendall's tau satisfies Properties 1-4 above and thus it defines a suitable measure of concordance.

Consider our leading example (2.7) of two stock prices (X_1, X_2), together with a call option (2.112) and a put option (2.113) on the first stock. We summarize Kendall's τ between any two of the above securities in the following table:

τ	X_1	C_1	P_1	X_2
X_1	1	1	-1	$\frac{2}{\pi}\arcsin\left(\sqrt{\frac{2}{5}}\right)$
C_1		1	-1	$\frac{2}{\pi}\arcsin\left(\sqrt{\frac{2}{5}}\right)$
P_1			1	$-\frac{2}{\pi}\arcsin\left(\sqrt{\frac{2}{5}}\right)$
X_2				1

(2.129)

The first stock price X_1, and the (regularized) call option payoff C_1 are co-monotonic, and therefore due to (2.125) their concordance is 1. The (regularized) put option payoff P_1 and the first stock price X_1 are anti-monotonic, and therefore due to (2.126) their concordance is -1. Similarly, the (regularized) put option payoff P_1 and the (regularized) call option payoff C_1 are anti-monotonic, and therefore their concordance is -1. The value of Kendall's τ between the first stock price X_1 and the second stock price X_2 is a specific instance of the more general result (2.178). Since the (regularized) call option payoff C_1 and the first stock price X_1 are co-monotonic, due to (2.122) the concordance of the second stock price X_2 with C_1 is the same as the concordance of X_2 with X_1. On the other hand, since the (regularized) put option payoff P_1 and the first stock price X_1 are anti-monotonic, due to (2.123) the concordance of the second stock price X_2 with P_1 is the opposite of the concordance of X_2 with X_1.

We mention another popular measure of concordance, *Spearman's rho*, which is the correlation of the grades:

$$\rho\{X_m, X_n\} \equiv \frac{\mathrm{Cov}\{U_m, U_n\}}{\mathrm{Sd}\{U_m\}\,\mathrm{Sd}\{U_n\}}, \tag{2.130}$$

see Section 2.5.3 below for a discussion of the correlation. It is possible to check that Spearman's rho satisfies Properties 1-4 above and therefore it defines a suitable measure of concordance.

Spearman's rho and Kendall's tau evaluation of the concordance between two variables is in general different, although this difference is bounded as follows:

$$\frac{3\tau - 1}{2} \leq \rho \leq \frac{1 + 2\tau - \tau^2}{2}, \tag{2.131}$$

whenever $\tau \geq 0$; and

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho \leq \frac{1 + 3\tau}{2}, \tag{2.132}$$

whenever $\tau \leq 0$.

2.5.3 Correlation

In this section we draw a bridge between the concordance summary statistics of a generic multivariate random variable $\mathbf{X}$ discussed above and the location-dispersion summary statistics of $\mathbf{X}$ introduced in Section 2.4, in particular the expected value (2.54) and the covariance matrix (2.67).

In defining the concordance summary statistics we relied on copulas, because copulas capture the core interdependence among variables: indeed the copula of one random variable with any of a set of co-monotonic variables is the same, although the co-monotonic variables might have very different marginal distributions.

The expected value is a purely "marginal" parameter, since it is the juxtaposition of the expected values of the single marginal entries X_n. Therefore, we cannot find any relation between expected value and parameters of concordance.

On the other hand, the covariance matrix displays both "marginal" and "joint" features.

From (2.74) the diagonal entries of the covariance matrix are the square of the standard deviation of the marginal entries. We can get rid of these purely "marginal" features by normalizing the covariance matrix into what is called the *correlation matrix*:

$$\mathrm{Cor}\{X_m, X_n\} \equiv [\mathrm{Cor}\{\mathbf{X}\}]_{mn} \equiv \frac{\mathrm{Cov}\{X_m, X_n\}}{\mathrm{Sd}\{X_m\}\,\mathrm{Sd}\{X_n\}}. \tag{2.133}$$

The correlation is an extremely popular parameter among finance practitioners.

In our leading example of two stock prices (X_1, X_2) we derive from (2.69) their correlation:

$$\text{Cor}\{X_1, X_2\} = \sqrt{\frac{2}{5}}. \tag{2.134}$$

The correlation displays some features that remind us of the properties of the measures of concordance.

Indeed the correlation is a normalized parameter:

$$-1 \leq \text{Cor}\{X_m, X_n\} \leq 1. \tag{2.135}$$

This follows from the Cauchy-Schwartz inequality (B.69). Furthermore, the following holds.

1. Independence implies the middle of the spectrum of possible correlation values:

$$(X_m, X_n) \text{ independent} \Rightarrow \text{Cor}\{X_m, X_n\} = 0. \tag{2.136}$$

This is true since the covariance of independent variables is zero, see (2.70).

2a. Positive affine concordance represents one extreme of the spectrum of possible correlation values:

$$X_m = a + bX_n \Leftrightarrow \text{Cor}\{X_m, X_n\} = 1, \tag{2.137}$$

where a is a scalar and b is a positive scalar. This follows from (B.70).

2b. Negative affine concordance represents the other extreme of the spectrum of possible correlation values:

$$X_m = a - bX_n \Leftrightarrow \text{Cor}\{X_m, X_n\} = -1, \tag{2.138}$$

where a is a scalar and b is a positive scalar. This follows from (B.71).

3. Correlations are unaffected by positive affine transformations:

$$\left.\begin{array}{l} Y_m \equiv a + bX_m \\ Y_n \equiv c + dX_n \end{array}\right\} \Rightarrow \text{Cor}\{X_m, X_n\} = \text{Cor}\{Y_m, Y_n\}, \tag{2.139}$$

where (a, c) are scalars and (b, d) are positive scalars. This follows from the affine equivariance property (2.71) of the covariance matrix.

4. Correlation and anti-correlation play a symmetric role:

$$\text{Cor}\{X_m, -X_n\} = -\text{Cor}\{X_m, X_n\}. \tag{2.140}$$

This follows from the affine equivariance property (2.71) of the covariance matrix.

A comparison of these properties with the respective properties of the measures of concordance shows that the correlation fails to be a suitable measure of concordance because it only covers affine transformations, whereas the measures of concordance cover more general invertible transformations.

Consider again our leading example (2.7) of two stock prices (X_1, X_2), along with the call option (2.112) and the put option (2.113) on the first stock. We summarize in the following table the correlation between any two of the above securities, which we computed by means of simulations:

Cor	X_1	C_1	P_1	X_2
X_1	1	.86	−.86	.63
C_1		1	−.47	.54
P_1			1	−.54
X_2				1

(2.141)

Although the first stock price X_1, and the (regularized) call option payoff C_1 are co-monotonic, their correlation is not 1. Although the (regularized) put option payoff P_1 and the first stock price X_1 are anti-monotonic, their correlation is not -1. Similarly, although the (regularized) put option payoff P_1 and the (regularized) call option payoff C_1 are anti-monotonic, their correlation is not -1. As far as the second stock is concerned, although the (regularized) call option payoff C_1 and the first stock price X_1 are co-monotonic, the correlation of the second stock price X_2 with C_1 is not the same as the correlation of X_2 with X_1. Similarly, although the (regularized) put option payoff P_1 and the first stock price X_1 are anti-monotonic, the correlation of the second stock price X_2 with P_1 is not the opposite of the correlation of X_2 with X_1.

Furthermore, a measure of concordance is defined in terms of the copula, and as such is not influenced by the marginal distribution of the variables involved. On the other hand, the set of possible values of the correlation does depend on the marginal distributions between the variables involved.

For example, consider two normally distributed random variables:

$$X_1 \sim \mathrm{N}\left(\mu_1, \sigma_1^2\right), \quad X_2 \equiv \mathrm{N}\left(\mu_2, \sigma_2^2\right). \tag{2.142}$$

It is possible to show that the correlation between these variables can take on any value in the interval $[-1, 1]$, see (2.169).

On the other hand, consider two lognormal variables:

$$Y_1 \equiv e^{X_1}, \quad Y_2 \equiv e^{X_2}. \tag{2.143}$$

The correlation between these variables is bounded within an interval smaller than $[-1, 1]$. For instance, the fact that both variables are positive implies that the correlation between the variables (2.143) cannot equal -1.

After the above critiques, one might wonder why correlation is such a popular tool. In the first place, the correlation indeed draws a bridge between the location-dispersion parameters and the dependence-concordance summary statistics. Secondly, for an important class of distribution the correlation completely defines the dependence structure, see Section 2.7.1.

2.6 Taxonomy of distributions

In this section we provide a taxonomy of multivariate distributions, stressing only the features that are needed in the sequel to tackle financial applications. Except for the order statistics, all the distribution introduced are generalizations of the one-dimensional distributions introduced in Section 1.3, to which the reader is referred to support intuition.

2.6.1 Uniform distribution

The simplest multivariate distribution is the uniform distribution. The *uniform distribution* models the situation where the only information available about the N-dimensional random variable $\mathbf{X}$ is that its realization is bound to take place on a given range in $\mathbb{R}^N$, and that all points in that range are equally likely outcomes for the realization.

In particular, consider an ellipsoid $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ centered in $\boldsymbol{\mu}$ with shape defined by the symmetric and positive matrix $\boldsymbol{\Sigma}$ as in $(A.73)$. We use the following notation to indicate that $\mathbf{X}$ is uniformly distributed on the ellipsoid $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$:

$$\mathbf{X} \sim \mathrm{U}\left(\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\right). \tag{2.144}$$

In Appendix www.2.11 we follow Fang, Kotz, and Ng (1990) to prove the results in the sequel.

The probability density function of the uniform distribution on the ellipsoid reads:

$$f^{\mathrm{U}}_{\boldsymbol{\mu},\Sigma}(\mathbf{x}) = \frac{\Gamma\left(\frac{N}{2}+1\right)}{\pi^{\frac{N}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\mathbb{I}_{\mathcal{E}_{\mu,\boldsymbol{\Sigma}}}(\mathbf{x}), \tag{2.145}$$

where Γ is the gamma function $(B.80)$ and $\mathbb{I}$ is the indicator function $(B.72)$.

The characteristic function of the uniform distribution on the ellipsoid reads:

$$\phi^{\mathrm{U}}_{\boldsymbol{\mu},\Sigma}(\boldsymbol{\omega}) = e^{i\boldsymbol{\omega}'\boldsymbol{\mu}}\psi\left(\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}\right), \tag{2.146}$$

where the function ψ is defined in terms of the beta function $(B.88)$ as follows[2]:

$$\psi(\gamma) \equiv \frac{2}{B\left(\frac{1}{2},\frac{N+1}{2}\right)}\int_0^1 \cos\left(\sqrt{\gamma}z\right)\left(1-z^2\right)^{\frac{N-1}{2}}dz. \tag{2.147}$$

[2] There are two minor typos in Fang, Kotz, and Ng (1990)

The mode is not defined, but the standard location parameter, i.e. the expected value, is defined and reads:

$$\mathrm{E}\{\mathbf{X}\} = \boldsymbol{\mu}. \tag{2.148}$$

The modal dispersion is not defined, but the standard scatter parameter, i.e. the covariance matrix, is defined and reads:

$$\mathrm{Cov}\{\mathbf{X}\} = \frac{1}{N+2}\boldsymbol{\Sigma}. \tag{2.149}$$

Now we split $\mathbf{X}$ in two sub-sets: the K-dimensional random variable $\mathbf{X}_A$ made of the first K entries and the $(N-K)$-dimensional random variable $\mathbf{X}_B$ made of the remaining entries. The marginal distribution of $\mathbf{X}_A$ is not uniform. The conditional distribution of $\mathbf{X}_B$ given $\mathbf{X}_A$ is uniform on an ellipsoid of lower dimension.

Bivariate standard uniform distribution

To gain more insight in the properties of the multivariate uniform distribution, we consider more in detail the bivariate uniform distribution on the unit circle.

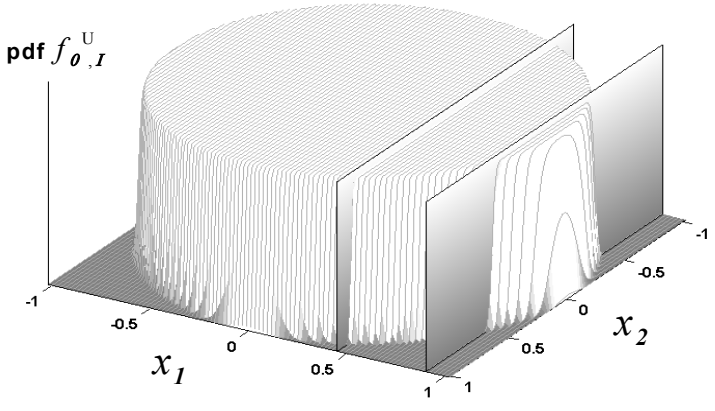

Fig. 2.12. Uniform distribution on the unit circle

The probability density function (2.145) is zero outside the unit circle and constant on the circle:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\pi}\mathbb{I}_{\{x_1^2+x_2^2\leq 1\}}(x_1,x_2), \tag{2.150}$$

see Figure 2.12.

We compute explicitly the marginal density of X_1. From Figure 2.12 we see that if $|x_1| > 1$ the marginal pdf is zero. When $|x_1| \leq 1$ the marginal density in x_1 is proportional to the area of the intersection of the vertical plane through x_1 with the density pie in Figure 2.12. This area is zero in $x_1 \equiv \pm 1$ and it reaches its maximum in $x_1 \equiv 0$. Indeed:

$$f_{X_1}(x_1) \equiv \int_{-\sqrt{1-x_1^2}}^{+\sqrt{1-x_1^2}} \frac{1}{\pi} dx_2 = \frac{2}{\pi}\sqrt{1-x_1^2}. \tag{2.151}$$

This formula shows that the marginal distribution of a uniform distribution is not uniform.

As for the conditional density of X_2 given x_1, we see in Figure 2.12 that the conditional pdf is non-zero only in the following domain:

$$-\sqrt{1-x_1^2} \leq x_2 \leq \sqrt{1-x_1^2}. \tag{2.152}$$

In this region the conditional pdf of X_2 given x_1 reads:

$$f_{X_2|x_1}(x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} = \frac{1}{2\sqrt{1-x_1^2}}. \tag{2.153}$$

Since it does not depend on its argument x_2, this function describes a plateau. A rescaled version of this plateau is represented in Figure 2.12 by the profile of the intersection of the vertical plane through x_1 with the density pie. When suitably rescaled, this plateau becomes taller and thinner as the known variable x_1 approaches the extremes $x_1 \equiv \pm 1$: indeed, if we know that $x_1 \equiv \pm 1$, then X_2 must be zero with certainty and thus the respective conditional probability density function must becomes a Dirac delta centered in zero, see (*B*.22).

From (2.149) the two variables X_1 and X_2 are uncorrelated:

$$\mathrm{Cor}\{X_1, X_2\} = 0. \tag{2.154}$$

Nevertheless, the conditional pdf of X_2 explicitly depends on X_1 and thus X_1 and X_2 are *not* independent.

2.6.2 Normal distribution

The *normal distribution* is the most widely used model to describe the statistical properties of a random variable $\mathbf{X}$ that can take on values in the whole space $\mathbb{R}^N$ in a symmetrical way around a peak.

The normal distribution depends on two parameters: an N-dimensional location vector $\boldsymbol{\mu}$ that determines the peak of the distribution, and an $N \times N$ symmetric and positive scatter matrix $\boldsymbol{\Sigma}$ that determines the shape of the distribution around its peak.

We use the following notation to indicate that $\mathbf{X}$ is normally distributed with the above parameters:

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.155}$$

The *standard normal distribution* corresponds to the specific case $\boldsymbol{\mu} \equiv \mathbf{0}$ and $\boldsymbol{\Sigma} \equiv \mathbf{I}$, the identity matrix.

The following results and more on the normal distribution can be found e.g. in Mardia, Kent, and Bibby (1979), Press (1982) and Morrison (2002).

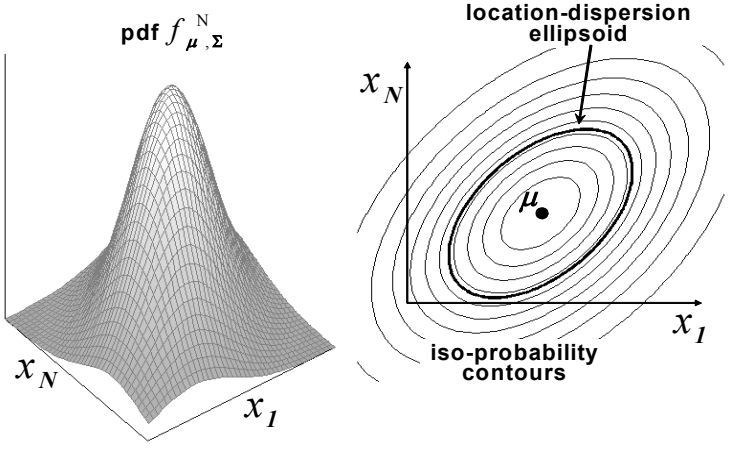

Fig. 2.13. Normal distribution

The multivariate normal probability density function reads:

$$f^{\mathrm{N}}_{\boldsymbol{\mu},\Sigma}(\mathbf{x}) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{2.156}$$

see the left portion of Figure 2.13 for a plot in the bivariate case, and the right portion of that figure for the projection on the plane of the points that share the same values of the pdf.

The characteristic function of the normal distribution reads:

$$\phi^{\mathrm{N}}_{\boldsymbol{\mu},\Sigma}(\boldsymbol{\omega}) = e^{i\boldsymbol{\mu}'\boldsymbol{\omega} - \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}}, \tag{2.157}$$

see also Appendix www.2.12. The expected value and the mode coincide and read:

$$\mathrm{E}\{\mathbf{X}\} = \mathrm{Mod}\{\mathbf{X}\} = \boldsymbol{\mu}. \tag{2.158}$$

The covariance matrix and the modal dispersion are both defined. They coincide and read:

$$\text{Cov}\{\mathbf{X}\} = \text{MDis}\{\mathbf{X}\} = \boldsymbol{\Sigma}. \tag{2.159}$$

In the right portion of Figure 2.13 we plot for the bivariate case the location-dispersion ellipsoid $\mathcal{E}_{\text{E,Cov}}$ defined in (2.75), see the discussion in Section 2.4.3.

Now we split $\mathbf{X}$ in two sub-sets: the K-dimensional random variable $\mathbf{X}_A$ made of the first K entries and the $(N-K)$-dimensional random variable $\mathbf{X}_B$ made of the remaining entries:

$$\mathbf{X} \equiv \begin{pmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{pmatrix}. \tag{2.160}$$

We split accordingly the location and the scatter parameters:

$$\boldsymbol{\mu} \equiv \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \boldsymbol{\Sigma} \equiv \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}. \tag{2.161}$$

The marginal distribution of $\mathbf{X}_A$ is a normal distribution with the following parameters:

$$\mathbf{X}_A \sim \text{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}). \tag{2.162}$$

This is a specific case of a more general result. Indeed, any affine transformation of $\mathbf{X}$ is normally distributed as follows:

$$\mathbf{a} + \mathbf{B}\mathbf{X} \sim \text{N}\left(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\right). \tag{2.163}$$

The conditional distribution of $\mathbf{X}_B$ given $\mathbf{x}_A$ is normal:

$$\mathbf{X}_B|\mathbf{x}_A \sim \text{N}(\boldsymbol{\mu}_B|\mathbf{x}_A, \boldsymbol{\Sigma}_B|\mathbf{x}_A), \tag{2.164}$$

where

$$\boldsymbol{\mu}_B|\mathbf{x}_A \equiv \boldsymbol{\mu}_B + \boldsymbol{\Sigma}_{BA}\boldsymbol{\Sigma}_{AA}^{-1}(\mathbf{x}_A - \boldsymbol{\mu}_A) \tag{2.165}$$

$$\boldsymbol{\Sigma}_B|\mathbf{x}_A \equiv \boldsymbol{\Sigma}_{BB} - \boldsymbol{\Sigma}_{BA}\boldsymbol{\Sigma}_{AA}^{-1}\boldsymbol{\Sigma}_{AB}. \tag{2.166}$$

Notice that the expression of the conditional covariance does not depend on the known variable $\mathbf{x}_A$.

As for independence, two jointly normal random variables are independent if and only if their covariance, or equivalently their correlation, is null:

$$(X_m, X_n) \text{ independent} \Leftrightarrow \text{Cov}\{X_n, X_n\} = 0. \tag{2.167}$$

This is another very special feature of the normal distribution. In general the much weaker relation (2.70) holds.

Bivariate normal distribution

To better understand the properties of the multivariate normal distribution, we consider the bivariate case.

We write the scatter parameter component-wise as follows:

$$\boldsymbol{\Sigma} \equiv \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{2.168}$$

where $|\rho| \leq 1$. This is the most general parametrization of a symmetric and positive 2×2 matrix. Also, it is convenient for notational purposes to assume $\sigma_1, \sigma_2 \geq 0$.

Since from (2.159) the matrix $\boldsymbol{\Sigma}$ is the covariance, it follows immediately that

$$\mathrm{Cor}\{X_1, X_2\} = \rho, \tag{2.169}$$

which shows that the correlation of two jointly normal variables can assume any value in the interval $[-1, 1]$.

In this notation, the expression of the normal probability density function (2.156) reads

$$f^{\mathrm{N}}_{\boldsymbol{\mu},\Sigma}(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} e^{-\frac{1}{2}\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{(1-\rho^2)}} \tag{2.170}$$

where (z_1, z_2) are the z-scores, i.e. the standardized variables:

$$z_i \equiv \frac{x_i - \mu_i}{\sigma_i}, \quad i = 1, 2. \tag{2.171}$$

From (2.162) the marginal distribution of X_1 is normal:

$$X_1 \sim \mathrm{N}\left(\mu_1, \sigma_1^2\right). \tag{2.172}$$

From Figure 2.13, this result is intuitive. Indeed, the marginal density in x_1 is proportional to the area underneath the joint probability density function cut by the vertical plane through x_1: this area decreases at infinity and has a peak at the point $x_1 \equiv \mu_1$.

From (2.164) the conditional distribution of X_2 given x_1 is also normal:

$$X_2|x_1 \sim \mathrm{N}\left(\mu_2|x_1, (\sigma_2|x_1)^2\right). \tag{2.173}$$

The above parameters read explicitly:

$$\mu_2|x_1 \equiv \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1) \tag{2.174}$$

$$\sigma_2|x_1 \equiv \sigma_2\sqrt{1-\rho^2}. \tag{2.175}$$

From Figure 2.13, this result is intuitive. Indeed, the (rescaled) profile of the conditional density of X_2 given x_1 is given by the intersection of the vertical plane through x_1 with the joint probability density function: this intersection has a bell shape peaked in $\mu_2|x_1$.

We now consider independence. If in (2.170) we set $\rho \equiv 0$, the pdf can be factored into the product of the pdf of the two marginal distributions of X_1 and X_2. In other words, from (2.169) we see that the two variables are independent if and only if their correlation is zero, which is stated more generally in (2.167).

To gain further insight in the dependence structure of the bivariate normal distribution, we consider the copula. In Appendix www.2.12 we prove that the probability density function of the copula reads in terms of the inverse of the error function (B.75) as follows:

$$f^{\mathrm{N}}_{U_1,U_2}(u_1,u_2) = \frac{1}{\sqrt{1-\rho^2}} e^{g_\rho\left(\mathrm{erf}^{-1}(2u_1-1),\mathrm{erf}^{-1}(2u_2-1)\right)}, \tag{2.176}$$

where g is defined as follows:

$$g_\rho(v_1,v_2) \equiv -\frac{\rho}{1-\rho^2} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}' \begin{pmatrix} \rho & -1 \\ -1 & \rho \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \tag{2.177}$$

From this expression we see that the copula of two jointly normal variables is completely determined by their correlation.

Therefore it is not surprising that Kendall's tau, the measure of concordance defined in (2.128), reads:

$$\tau\{X_1,X_2\} = \frac{2}{\pi}\arcsin(\rho). \tag{2.178}$$

In other words, the concordance of two jointly normal variables is completely determined by their correlation.

Matrix-variate normal distribution

Consider an $(N \times K)$-matrix-valued random variable:

$$\mathbf{X} \equiv \left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(K)}\right) \equiv \begin{pmatrix} \mathbf{X}_{(1)} \\ \vdots \\ \mathbf{X}_{(N)} \end{pmatrix}, \tag{2.179}$$

where each column $\mathbf{X}^{(k)}$ is an N-dimensional random variable and each row $\mathbf{X}_{(n)}$ is a K-dimensional random variable. The random matrix $\mathbf{X}$ has a *matrix-variate normal distribution* if

$$\mathrm{vec}(\mathbf{X}) \sim \mathrm{N}(\mathrm{vec}(\mathbf{M}), \mathbf{S} \otimes \mathbf{\Sigma}), \tag{2.180}$$

where vec is the operator (A.104) that stacks the columns of a matrix into a vector; $\mathbf{S}$ is a $K \times K$ symmetric and positive definite matrix; $\mathbf{\Sigma}$ is an $N \times N$ symmetric and positive definite matrix; and $\otimes$ denotes the Kronecker product (A.96). We denote a matrix-variate normal distribution with the above parameters as follows:

$$\mathbf{X} \sim \mathrm{N}(\mathbf{M}, \boldsymbol{\Sigma}, \mathbf{S}). \tag{2.181}$$

The following results are proved in Appendix www.2.13.

The probability density function of the matrix-valued random variable (2.181) can be conveniently expressed as follows:

$$f^{\mathrm{N}}_{\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{S}}(\mathbf{X}) \equiv (2\pi)^{-\frac{NK}{2}} |\boldsymbol{\Sigma}|^{-\frac{K}{2}} |\mathbf{S}|^{-\frac{N}{2}} e^{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{S}^{-1}(\mathbf{X}-\mathbf{M})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\right\}}. \tag{2.182}$$

Notice that this density generalizes the vector-variate normal probability density function (2.156). Therefore the multivariate normal distribution (2.155) can be seen as the following special case of the matrix-variate normal distribution:

$$\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 1). \tag{2.183}$$

From the definition (2.180) we see immediately that the matrix $\mathbf{M}$ is the expected value of $\mathbf{X}$:

$$\mathrm{E}\{\mathbf{X}\} = \mathbf{M}. \tag{2.184}$$

The matrix $\boldsymbol{\Sigma}$ defines the overall covariance structure between any two N-dimensional columns $\mathbf{X}^{(j)}, \mathbf{X}^{(k)}$ among the K that constitute the random matrix $\mathbf{X}$:

$$\mathrm{Cov}\left\{\mathbf{X}^{(j)}, \mathbf{X}^{(k)}\right\} = S_{jk}\boldsymbol{\Sigma}. \tag{2.185}$$

Similarly, the matrix $\mathbf{S}$ defines the overall covariance structure between any two K-dimensional rows $\mathbf{X}_{(m)}, \mathbf{X}_{(n)}$ among the N that constitute the random matrix $\mathbf{X}$:

$$\mathrm{Cov}\left\{\mathbf{X}_{(m)}, \mathbf{X}_{(n)}\right\} = \Sigma_{mn}\mathbf{S}. \tag{2.186}$$

2.6.3 Student t distribution

The *Student t distribution* is another model that describes the statistical properties of a random variable $\mathbf{X}$ which can assume values on the whole space $\mathbb{R}^N$ in a symmetrical way around a peak.

Similarly to the normal distribution, the Student t distribution depends on an N-dimensional location parameter $\boldsymbol{\mu}$ that determines the peak of the distribution, and an $N \times N$ symmetric and positive scatter matrix $\boldsymbol{\Sigma}$ that determines the shape of the distribution around its peak. It also depends on an additional parameter ν, the *degrees of freedom* of the Student t distribution, whose integer value determines the relative importance of the peak of the distribution with respect to its tails.

We use the following notation to indicate that $\mathbf{X}$ has a Student t-distribution with the above parameters:

$$\mathbf{X} \sim \mathrm{St}(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.187}$$

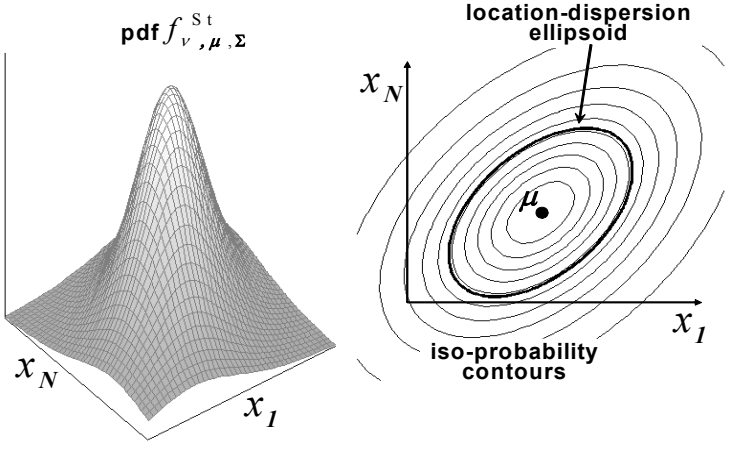

Fig. 2.14. Student t distribution

The *standard Student t distribution* corresponds to the specific case $\boldsymbol{\mu} \equiv \mathbf{0}$ and $\boldsymbol{\Sigma} \equiv \mathbf{I}$, the identity matrix.

The multivariate Student t probability density function reads:

$$f^{\text{St}}_{\nu,\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) = (\nu\pi)^{-\frac{N}{2}} \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left(1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^{-\frac{\nu+N}{2}}, \tag{2.188}$$

where Γ denotes the gamma function (B.80), see Kotz and Nadarajah (2004) and Fang, Kotz, and Ng (1990). In the left portion of Figure 2.14 we plot the bivariate case, and in the right portion we plot the projection on the plane of the points that share the same values of the pdf.

The characteristic function of the Student t distribution is computed in Sutradhar (1986) and Sutradhar (1988). The characteristic function of the Student t distribution assumes a different form depending on whether the degrees of freedom ν are odd or even. We report here the expression for odd degrees of freedom:

$$\phi^{\text{St}}_{\nu,\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) = \frac{\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right)}{2^{\nu-1}\Gamma\left(\frac{\nu}{2}\right)} e^{\left(i\boldsymbol{\mu}'\boldsymbol{\omega} - \sqrt{\nu\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}}\right)} \sum_{r=1}^{\frac{\nu+1}{2}} \binom{\nu-r}{\frac{\nu+1}{2}-r} \frac{\left(2\sqrt{\nu\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}}\right)^{r-1}}{(r-1)!}, \tag{2.189}$$

where Γ is the gamma function (B.80).

The expected value and the mode coincide and read:

$$\mathrm{E}\{\mathbf{X}\} = \mathrm{Mod}\{\mathbf{X}\} = \boldsymbol{\mu}. \tag{2.190}$$

The covariance matrix is defined if $\nu > 0$ and reads:

$$\mathrm{Cov}\{\mathbf{X}\} = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma}. \tag{2.191}$$

In the right portion of Figure 2.14 we plot for the bivariate case the location-dispersion ellipsoid $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ defined in (2.75), see the discussion in Section 2.4.3.

Now we split $\mathbf{X}$ in two sub-sets: the K-dimensional random variable $\mathbf{X}_A$ made of the first K entries and the $(N-K)$-dimensional random variable $\mathbf{X}_B$ made of the remaining entries:

$$\mathbf{X} \equiv \begin{pmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{pmatrix}. \tag{2.192}$$

We split accordingly the location and the scatter parameters:

$$\boldsymbol{\mu} \equiv \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \boldsymbol{\Sigma} \equiv \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}. \tag{2.193}$$

The marginal distribution of $\mathbf{X}_A$ is a Student t distribution with the following parameters:

$$\mathbf{X}_A \sim \mathrm{St}\left(\nu, \boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}\right). \tag{2.194}$$

This is a specific case of a more general result. Indeed, any affine transformation of $\mathbf{X}$ is Student t distributed as follows:

$$\mathbf{a} + \mathbf{B}\mathbf{X} \sim \mathrm{St}\left(\nu, \mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\right). \tag{2.195}$$

On the other hand, unlike in the normal case, the conditional distribution of a Student t distribution is in general not a Student t distribution. Indeed, from the expression of the joint pdf (2.188) and the fact that from (2.194) the marginal pdf is in the form (2.188) it is immediate to compute the pdf of the conditional distribution as the ratio of the joint pdf and the marginal pdf. Nevertheless, the conditional pdf is not of the form (2.188).

As far as independence is concerned, since the generic conditional distribution is not a Student t distribution and the generic marginal distribution is a Student t distribution it follows that marginal and conditional distribution cannot coincide. Therefore random variables that are jointly Student t distributed are not independent.

Just like in the one-dimensional case, the Student t distribution encompasses the normal distribution as a special case. Indeed, as we show more in general in Appendix www.2.14, in the limit $\nu \to \infty$ the Student t probability density function (2.188) yields the normal probability density function (2.156) and thus:

$$\mathrm{St}(\infty, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.196}$$

As the degrees of freedom ν decrease, the tails in the pdf (2.188) of the distribution become thicker and thicker. We can see this by comparing Figure 2.14 with Figure 2.13, see also Section 2.6.4 and refer to Figure 1.9 for the univariate case.

Matrix-variate Student t distribution

The matrix-variate Student t distribution was introduced by Dickey (1967), see Appendix www.2.14 for the relation with the notation in the original paper.

Consider an $(N \times K)$-matrix-valued random variable

$$\mathbf{X} \equiv \left(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)}\right) \equiv \begin{pmatrix} \mathbf{X}_{(1)} \\ \vdots \\ \mathbf{X}_{(N)} \end{pmatrix}, \tag{2.197}$$

where each column $\mathbf{X}^{(k)}$ is an N-dimensional random variable and each row $\mathbf{X}_{(n)}$ is a K-dimensional random variable.

The random matrix $\mathbf{X}$ is distributed according to a *matrix-valued Student t distribution* with the following parameters

$$\mathbf{X} \sim \mathrm{St}(\nu, \mathbf{M}, \boldsymbol{\Sigma}, \mathbf{S}), \tag{2.198}$$

if its probability density function reads:

$$f^{\mathrm{St}}_{\nu,\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{S}}(\mathbf{X}) \equiv \gamma |\boldsymbol{\Sigma}|^{-\frac{K}{2}} |\mathbf{S}|^{-\frac{N}{2}} \left| \mathbf{I}_K + \mathbf{S}^{-1} (\mathbf{X} - \mathbf{M})' \frac{\boldsymbol{\Sigma}^{-1}}{\nu} (\mathbf{X} - \mathbf{M}) \right|^{-\frac{\nu+N}{2}}. \tag{2.199}$$

In this expression $\mathbf{M}$ is an $N \times K$ matrix; $\boldsymbol{\Sigma}$ is an $N \times N$ symmetric and positive definite matrix; $\mathbf{S}$ is a $K \times K$ symmetric and positive definite matrix; ν is a positive integer; and γ is a normalization constant defined in terms of the gamma function (B.80) as follows:

$$\gamma \equiv (\nu\pi)^{-\frac{NK}{2}} \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{\nu+N-1}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)} \cdots \frac{\Gamma\left(\frac{\nu+N-K+1}{2}\right)}{\Gamma\left(\frac{\nu-K+1}{2}\right)}. \tag{2.200}$$

Notice that the density (2.199) generalizes the vector-variate Student t probability density function (2.188). Therefore the multivariate Student t distribution (2.187) can be seen as the following special case of the matrix-variate Student t distribution:

$$\mathrm{St}(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{St}(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}, 1). \tag{2.201}$$

Unlike in the normal case (2.180), by stacking the columns of the matrix $\mathbf{X}$ in (2.198) with the vec operator (A.104) we do not obtain a Student t-distributed variable:

$$\text{vec}(\mathbf{X}) \nsim \text{St}(\nu, \text{vec}(\mathbf{M}), \mathbf{S} \otimes \mathbf{\Sigma}). \tag{2.202}$$

Nevertheless the following results hold:

$$\text{E}\{\mathbf{X}\} = \mathbf{M}, \tag{2.203}$$

which generalizes (2.190); and

$$\text{Cov}\{\text{vec}(\mathbf{X})\} = \frac{\nu}{\nu - 2}\mathbf{S} \otimes \mathbf{\Sigma}, \tag{2.204}$$

which generalizes (2.191). Therefore the matrix $\mathbf{\Sigma}$ defines the overall covariance structure between any two N-dimensional columns $\mathbf{X}^{(j)}, \mathbf{X}^{(k)}$ among the K that constitute the random matrix $\mathbf{X}$:

$$\text{Cov}\left\{\mathbf{X}^{(j)}, \mathbf{X}^{(k)}\right\} = \frac{\nu}{\nu - 2} S_{jk} \mathbf{\Sigma}. \tag{2.205}$$

Similarly, the matrix $\mathbf{S}$ defines the overall covariance structure between any two K-dimensional rows $\mathbf{X}_{(m)}, \mathbf{X}_{(n)}$ among the N that constitute the random matrix $\mathbf{X}$:

$$\text{Cov}\left\{\mathbf{X}_{(m)}, \mathbf{X}_{(n)}\right\} = \frac{\nu}{\nu - 2} \Sigma_{mn} \mathbf{S}. \tag{2.206}$$

These result parallel (2.184)-(2.186) for the normal distribution. Indeed, in the limit $\nu \to \infty$ the matrix-variate Student t distribution (2.198) becomes the matrix-variate normal distribution (2.181):

$$\text{St}(\infty, \mathbf{M}, \mathbf{\Sigma}, \mathbf{S}) = \text{N}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{S}), \tag{2.207}$$

see the proof in Appendix www.2.14.

2.6.4 Cauchy distribution

As in the univariate setting, the special case of the Student t distribution with $\nu \equiv 1$ degrees of freedom is called the *Cauchy distribution*, which we denote as follows:

$$\text{Ca}(\boldsymbol{\mu}, \mathbf{\Sigma}) \equiv \text{St}(1, \boldsymbol{\mu}, \mathbf{\Sigma}). \tag{2.208}$$

The *standard Cauchy distribution* corresponds to the specific case $\boldsymbol{\mu} \equiv \mathbf{0}$ and $\mathbf{\Sigma} \equiv \mathbf{I}$, the identity matrix.

From (2.188), the probability density function of the Cauchy distribution reads:

$$f^{\text{Ca}}_{\boldsymbol{\mu},\mathbf{\Sigma}}(\mathbf{x}) = \frac{\Gamma\left(\frac{1+N}{2}\right)}{\pi^{\frac{N}{2}}\Gamma\left(\frac{1}{2}\right)} |\mathbf{\Sigma}|^{-\frac{1}{2}} \left(1 + (\mathbf{x} - \boldsymbol{\mu})' \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)^{-\frac{N+1}{2}}, \tag{2.209}$$

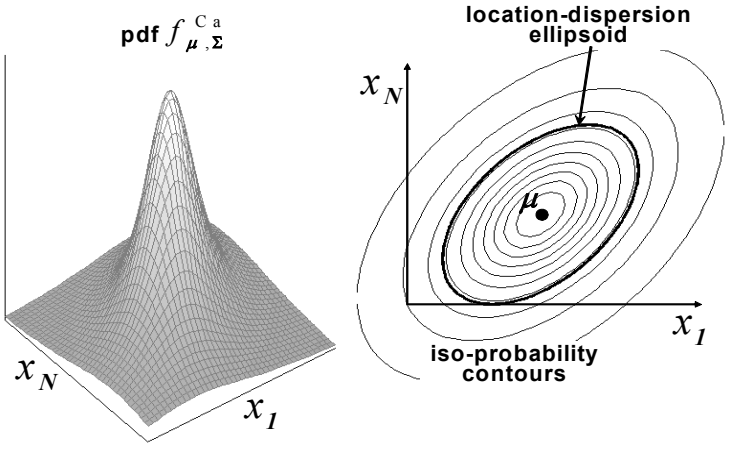

Fig. 2.15. Cauchy distribution

see the left portion of Figure 2.15 for a plot in the bivariate case, and the right portion of that figure for the projection on the plane of the points that share the same values of the pdf.

From (2.189), the characteristic function of the Cauchy distribution reads:

$$\phi^{\text{Ca}}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{\omega}) = e^{i\boldsymbol{\mu}'\boldsymbol{\omega} - \sqrt{\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}}}. \tag{2.210}$$

The tails in the density (2.209) are so thick that the moments are not defined. Nevertheless, the mode is defined and reads:

$$\text{Mod}\{\mathbf{X}\} = \boldsymbol{\mu}. \tag{2.211}$$

Similarly, the modal dispersion (2.65) is defined and reads:

$$\text{MDis}\{\mathbf{X}\} = \frac{1}{N+1}\boldsymbol{\Sigma}, \tag{2.212}$$

see Appendix www.2.15. In the right portion of Figure 2.15 we plot for the bivariate case the location-dispersion ellipsoid $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ defined in (2.75), see the discussion in Section 2.4.3.

2.6.5 Log-distributions

Log-distributions are defined as the exponential of other parametric distributions. As such, they are suitable to model positive quantities such as prices of limited-liability securities.

More precisely, consider a random variable $\mathbf{Y}$, whose distribution is represented by its pdf $f_{\mathbf{Y}}$, or its cdf $F_{\mathbf{Y}}$, or its characteristic function $\phi_{\mathbf{Y}}$. The variable $\mathbf{X} \equiv e^{\mathbf{Y}}$, where the exponential acts component-wise, is *log-*$\mathbf{Y}$ *distributed*, because by definition the logarithm of $\mathbf{X}$ has the same distribution as $\mathbf{Y}$.

The following results are discussed in Appendix www.2.16.

The probability density function of a log-$\mathbf{Y}$ distribution reads:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{f_{\mathbf{Y}}(\ln(\mathbf{x}))}{\prod_{n=1}^{N} x_n}. \tag{2.213}$$

The raw moments of a log-$\mathbf{Y}$ distribution read:

$$\mathrm{E}\{X_{n_1} \cdots X_{n_k}\} = \phi_{\mathbf{Y}}(\boldsymbol{\omega}_{n_1 \cdots n_k}), \tag{2.214}$$

where the vector $\boldsymbol{\omega}$ is defined in terms of the canonical basis $(A.15)$ as follows:

$$\boldsymbol{\omega}_{n_1 \cdots n_k} \equiv \frac{1}{i}\left(\boldsymbol{\delta}^{(n_1)} + \cdots + \boldsymbol{\delta}^{(n_k)}\right). \tag{2.215}$$

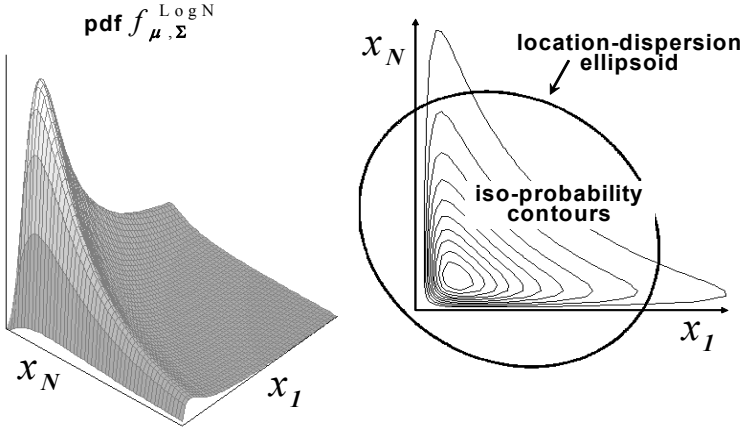

Fig. 2.16. Lognormal distribution

In particular, consider a random variable $\mathbf{Y}$ that is normally distributed with expected value $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.216}$$

We use the following notation to indicate that $\mathbf{X} \equiv e^{\mathbf{Y}}$ has a *lognormal distribution* with the above parameters:

$$\mathbf{X} \sim \mathrm{LogN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.217}$$

The probability density function of the lognormal distribution follows from (2.213) and the pdf (2.156) of the normal distribution:

$$f^{\mathrm{LogN}}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) = \frac{(2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\prod_{n=1}^{N} x_n} e^{-\frac{1}{2}(\ln(\mathbf{x})-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\ln(\mathbf{x})-\boldsymbol{\mu})}, \tag{2.218}$$

see the left portion of Figure 2.16 for a plot in the bivariate case, and the right portion of that figure for the projection on the plane of the points that share the same values of the pdf.

Expected values and covariances of the lognormal distribution follow from (2.214) and the characteristic function (2.157) of the normal distribution:

$$\mathrm{E}\{X_n\} = e^{\mu_n + \frac{\Sigma_{nn}}{2}} \tag{2.219}$$

$$\mathrm{Cov}\{X_m, X_n\} = e^{\mu_m + \mu_n + \frac{\Sigma_{mm}}{2} + \frac{\Sigma_{nn}}{2}} \left(e^{\Sigma_{mn}} - 1\right). \tag{2.220}$$

In the right portion of Figure 2.16 we plot for the bivariate case the location-dispersion ellipsoid $\mathcal{E}_{\mathrm{E,Cov}}$ defined in (2.75), see the discussion in Section 2.4.3.

2.6.6 Wishart distribution

Consider a set of random variables $\{\mathbf{X}_1, \ldots, \mathbf{X}_\nu\}$ that are independent and normally distributed with zero expected value and with the same scatter parameter:

$$\mathbf{X}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad t = 1, \ldots, \nu. \tag{2.221}$$

The *Wishart distribution* with ν degrees of freedom is the distribution of the random matrix $\mathbf{W}$ defined as follows:

$$\mathbf{W} \equiv \mathbf{X}_1 \mathbf{X}_1' + \cdots + \mathbf{X}_\nu \mathbf{X}_\nu'. \tag{2.222}$$

Therefore the Wishart distribution depends on two parameters: the degrees of freedom ν, which takes on integer values, and the scale parameter $\boldsymbol{\Sigma}$, which is a symmetric and positive matrix. We use the following notation to indicate that $\mathbf{W}$ is a Wishart-distributed matrix with the above parameters:

$$\mathbf{W} \sim \mathrm{W}(\nu, \boldsymbol{\Sigma}). \tag{2.223}$$

Notice that by construction $\mathbf{W}$ is a symmetric and positive matrix-valued random variable. This distribution plays a major role in the analysis of the estimation of covariance matrices.

The following results on the Wishart distribution can be found in Anderson (1984) and Mardia, Kent, and Bibby (1979).

The probability density function of the Wishart distribution reads:

$$f^{\mathrm{W}}_{\nu,\mathbf{\Sigma}}(\mathbf{W}) = \frac{1}{\kappa}|\mathbf{\Sigma}|^{-\frac{\nu}{2}}|\mathbf{W}|^{\frac{\nu-N-1}{2}}e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{W}\right)}, \tag{2.224}$$

where κ is a normalization constant defined in terms of the the gamma function $(B.80)$ as follows:

$$\kappa \equiv 2^{\frac{\nu N}{2}}\pi^{\frac{N(N-1)}{4}}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu-1}{2}\right)\cdots\Gamma\left(\frac{\nu-N+1}{2}\right). \tag{2.225}$$

The characteristic function of the Wishart distribution reads:

$$\phi^{\mathrm{W}}_{\nu,\mathbf{\Sigma}}(\mathbf{\Omega}) \equiv \mathrm{E}\left\{e^{i\,\mathrm{tr}(\mathbf{W\Omega})}\right\} = |\mathbf{I}-2i\mathbf{\Sigma\Omega}|^{-\frac{\nu}{2}}. \tag{2.226}$$

The expected value, which is the standard parameter of location, reads component-wise as follows:

$$\mathrm{E}\{W_{mn}\} = \nu\Sigma_{mn}. \tag{2.227}$$

The cross-covariances, which determine the dispersion of $\mathbf{W}$, read:

$$\mathrm{Cov}\{W_{mn}, W_{pq}\} = \nu\left(\Sigma_{mp}\Sigma_{nq} + \Sigma_{mq}\Sigma_{np}\right). \tag{2.228}$$

As in Magnus and Neudecker (1979), we can express this in compact notation as follows:

$$\mathrm{Cov}\{\mathrm{vec}[\mathbf{W}]\} = \nu\left(\mathbf{I}_{N^2} + \mathbf{K}_{NN}\right)(\mathbf{\Sigma}\otimes\mathbf{\Sigma}), \tag{2.229}$$

where vec is the operator $(A.104)$ that stacks the columns of $\mathbf{W}$ into a vector, $\mathbf{I}$ is the identity matrix, $\mathbf{K}$ is the commutation matrix $(A.108)$ and $\otimes$ is the Kronecker product $(A.96)$.

A comparison of (2.224) with (1.110) shows that the Wishart distribution is the multivariate generalization of the gamma distribution (1.108). Furthermore, for a generic vector $\mathbf{a}$ we obtain:

$$\mathbf{W} \sim \mathrm{W}(\nu, \mathbf{\Sigma}) \Rightarrow \mathbf{a}'\mathbf{W}\mathbf{a} \sim \mathrm{Ga}(\nu, \mathbf{a}'\mathbf{\Sigma}\mathbf{a}), \tag{2.230}$$

see Appendix www.2.17.

Since the inverse of a symmetric and positive matrix is a symmetric and positive matrix, the Wishart distribution can be used to model a symmetric and positive matrix also through its inverse. In other words, assume that the inverse of a random matrix $\mathbf{Z}$ is Wishart-distributed:

$$\mathbf{Z}^{-1} \sim \mathrm{W}\left(\nu, \mathbf{\Psi}^{-1}\right). \tag{2.231}$$

Then the distribution of $\mathbf{Z}$ is called *inverse-Wishart*, and is denoted as follows:

$$\mathbf{Z} \sim \mathrm{IW}(\nu, \mathbf{\Psi}). \tag{2.232}$$

We stress that $\mathbf{Z}$ is, like $\mathbf{Z}^{-1}$, a matrix-valued random variable that is symmetric and positive.

In Appendix www.2.17 we prove that the probability density function of the inverse-Wishart distribution reads:

$$f^{\mathrm{IW}}_{\nu,\mathbf{\Psi}}(\mathbf{Z}) = \frac{1}{\kappa}\,|\mathbf{\Psi}|^{\frac{\nu}{2}}\,|\mathbf{Z}|^{-\frac{\nu+N+1}{2}}\,e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Psi}\mathbf{Z}^{-1}\right)}, \tag{2.233}$$

where κ is the normalization constant (2.225).

The 2×2 Wishart distribution

To better understand the Wishart distribution we consider the case of 2×2 matrices:

$$\mathbf{W} \equiv \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \sim \mathrm{W}\left(\nu; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right), \tag{2.234}$$

where $|\rho| \leq 1$.

The symmetry of $\mathbf{W}$ implies $W_{12} \equiv W_{21}$. Therefore this random matrix is completely determined by the three entries (W_{11}, W_{12}, W_{22}).

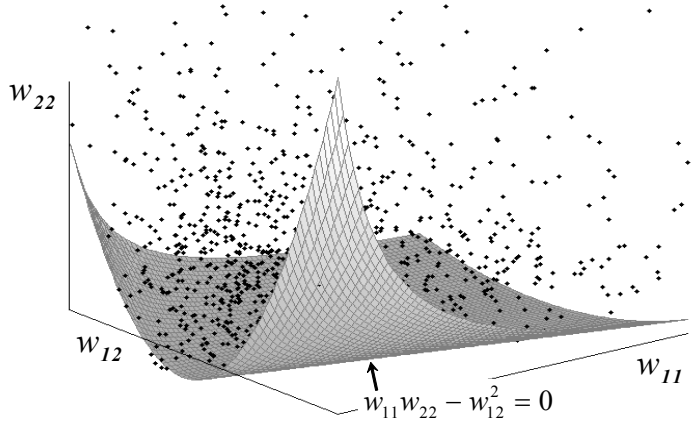

Fig. 2.17. Wishart distribution

Furthermore a symmetric matrix is positive if and only if its eigenvalues are positive. In the 2×2 case, denoting as λ_1 and λ_2 the two eigenvalues, these are positive if and only if the following inequalities are satisfied:

$$\lambda_1\lambda_2 > 0, \qquad \lambda_1 + \lambda_2 > 0. \tag{2.235}$$

On the other hand, the product of the eigenvalues is the determinant of $\mathbf{W}$ and the sum of the eigenvalues is the trace of $\mathbf{W}$, which are both invariants, see Appendix A.4. Therefore the positivity condition is equivalent to the two conditions below:

$$|\mathbf{W}| \equiv W_{11}W_{22} - W_{12}^2 \geq 0 \tag{2.236}$$

$$\mathrm{tr}\,(\mathbf{W}) \equiv W_{11} + W_{22} \geq 0, \tag{2.237}$$

where the first expression follows from $(A.41)$.

In Figure 2.17 we plot a few outcomes of a simulation of (2.234). Notice that all the outcomes lie above the surface $w_{11}w_{22} - w_{12}^2 = 0$: therefore (2.236) is satisfied. Furthermore, all the outcomes satisfy $w_{22} \geq -w_{11}$: therefore (2.237) is also satisfied. In other words, each outcome corresponds to a symmetric and positive 2×2 matrix.

2.6.7 Empirical distribution

The generalization to the multivariate case of the empirical distribution is immediate. Suppose that we can access T past measurements of the N-dimensional random variable $\mathbf{X}$:

$$i_T \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}, \tag{2.238}$$

where we use the lower-case notation because these measurements have already taken place and thus they no longer represent random variables.

The *empirical distribution* models in the most simplistic way the basic assumption of statistics that we can learn from past experience. More precisely, under this distribution any of the past occurrences is an equally likely potential outcome of future measurements of $\mathbf{X}$, whereas different realizations cannot occur.

We use the following notation to indicate that $\mathbf{X}$ is distributed according to an empirical distribution stemming from the above observations:

$$\mathbf{X} \sim \mathrm{Em}\,(i_T)\,. \tag{2.239}$$

The empirical distribution is discrete. Therefore its probability density function is a generalized function. As in $(B.22)$, we can express the empirical pdf as follows:

$$f_{i_T}(\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} \delta^{(\mathbf{x}_t)}(\mathbf{x})\,, \tag{2.240}$$

where δ is the Dirac delta $(B.16)$.

To visualize this probability density function we regularize it by means of the convolution as in $(B.54)$. The regularized probability density function of the empirical distribution reads in terms of the smooth approximation $(B.18)$ of the Dirac delta as follows:

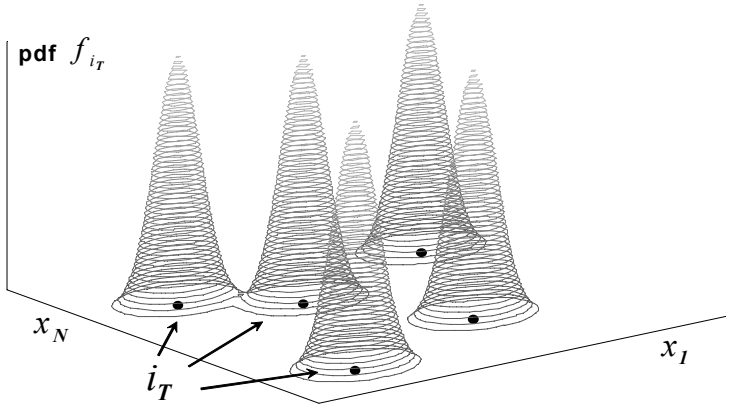

Fig. 2.18. Empirical distribution (regularized)

$$f_{i_T;\epsilon} \equiv f_{i_T} * \delta_\epsilon^{(\mathbf{0})} = \frac{1}{T}\sum_{t=1}^{T} \delta_\epsilon^{(\mathbf{x}_t)}, \tag{2.241}$$

where ϵ is a small bandwidth, see Figure 2.18.

From $(B.53)$ the empirical cumulative distribution function reads:

$$F_{i_T} = \frac{1}{T}\sum_{t=1}^{T} H^{(\mathbf{x}_t)}, \tag{2.242}$$

where H is the Heaviside step function $(B.73)$.

From the definition of the characteristic function (2.13) in terms of the expectation operator $(B.56)$, and from the property $(B.17)$ of the Dirac delta we obtain the characteristic function of the empirical distribution:

$$\phi_{i_T}(\boldsymbol{\omega}) = \frac{1}{T}\sum_{t=1}^{T} e^{i\boldsymbol{\omega}'\mathbf{x}_t}. \tag{2.243}$$

From the same rationale we also obtain the moments of any order of the empirical distribution. In particular, the expected value is called the *sample mean*, which we denote as follows:

$$\widehat{\mathrm{E}}_{i_T} \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t. \tag{2.244}$$

Similarly, the covariance matrix of the empirical distribution is called the *sample covariance*, which we denote as follows:

$$\widehat{\mathrm{Cov}}_{i_T} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_t - \widehat{\mathrm{E}}_{i_T}\right)^2. \tag{2.245}$$

2.6.8 Order statistics

The order statistics are useful in the context of nonparametric estimation. The following results and more can be found in David (1981).

Consider T independent and identically distributed univariate random variables and their respective realizations:

$$\{X_1,\ldots,X_T\}, \quad \{x_1,\ldots,x_T\}, \tag{2.246}$$

where as usual the upper-case notation indicates the random variable, and the lower-case notation indicates the respective realization.

Consider the smallest among the realized variables: this is, say, the realization of the second variable x_2. In a different scenario, the smallest realization might have been the realization of a different random variable, say x_4. In general, the value x_2 in the first scenario is different than the value x_4 in the second scenario. In other words, the minimum among the random variables (2.246) is a random variable.

Similarly, the maximum among the random variables (2.246) is a random variable. More in general, consider the whole set of ordered random variables:

$$\begin{aligned} X_{1:T} &\equiv \min\{X_1,\ldots,X_T\} \\ &\vdots \\ X_{T:T} &\equiv \max\{X_1,\ldots,X_T\}. \end{aligned} \tag{2.247}$$

The generic r-th element $X_{r:T}$, i.e. the r-th smallest random variable, is called the *r-th order statistic*.

The probability density function of the order statistics reads:

$$f_{X_{r:T}}(x) = \frac{T!}{(r-1)!\,(T-r)!} F_X^{r-1}(x)\left(1-F_X(x)\right)^{T-r} f_X(x), \tag{2.248}$$

where f_X and F_X denote respectively the common probability density function and the common cumulative distribution function respectively of all the variables (2.246).

The cumulative distribution function of the order statistics reads:

$$F_{X_{r:T}}(x) = I\left(F_X(x), r, T-r+1\right), \tag{2.249}$$

where I is the regularized beta function ($B.91$).

When defined, the expected value of the generic r-th order statistic can be expressed in terms of the common quantile function Q_X of the variables (2.246) as follows:

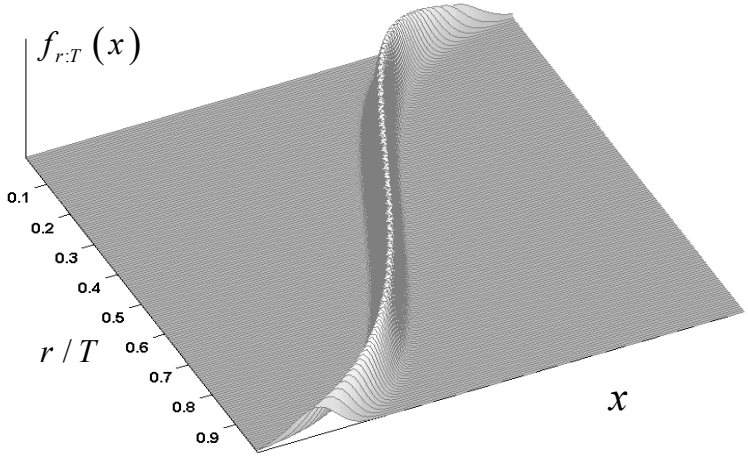

Fig. 2.19. Probability density function of order statistics

$$\mathrm{E}\{X_{r:T}\} = \int_{\mathbb{R}} Q_X(u)\, \widetilde{\delta}_{r,T}(u)\, du, \tag{2.250}$$

where the function δ is defined in terms of the indicator function $(B.72)$ and reads:

$$\widetilde{\delta}_{r,T}(u) \equiv \frac{T!}{(r-1)!\,(T-r)!} u^{r-1} (1-u)^{T-r}\, \mathbb{I}_{[0,1]}(u). \tag{2.251}$$

In the limit of a large sample T this function is a smooth approximation to the Dirac delta $(B.16)$:

$$\widetilde{\delta}_{r,T} \overset{T\to\infty}{\longrightarrow} \delta^{(r/T)}. \tag{2.252}$$

Therefore, when it is defined, the expected value of the r-th order statistic can be approximated by the quantile of any of the variables (2.246) as follows:

$$\mathrm{E}\{X_{r:T}\} \approx Q_X\left(\frac{r}{T}\right), \tag{2.253}$$

see Figure 2.19 and compare with Figure 1.2.

The concentration of the distribution of the order statistics around its expected value and the accuracy of the approximation (2.253) increases with the size T of the sample according to (2.252).

An important case of order statistics are those of the uniform distribution. Consider a set of T random variables that are independent and identically uniformly distributed on the unit interval:

$$U_t \sim \mathrm{U}([0,1]), \quad t = 1, \ldots, T. \tag{2.254}$$

The order statistics of the uniform distribution are important because they represent the grade of any order statistics. In other words, the order statistic from a generic distribution (2.247) has the same distribution as the quantile of the respective order statistics from the uniform distribution:

$$(X_{1:T},\ldots,X_{T:T}) \stackrel{d}{=} (Q_X(U_{1:T}),\ldots,Q_X(U_{T:T})). \tag{2.255}$$

This result is a straightforward consequence of the definition of quantile, see (2.27) and Figure 2.3.

2.7 Special classes of distributions

In this section we put the distributions presented in Section 2.6 in a more general perspective in view of their applications. Refer to Figure 2.20 throughout the discussion.

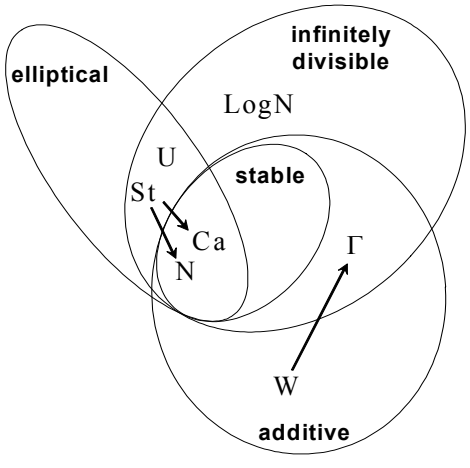

Fig. 2.20. Special classes of distributions

2.7.1 Elliptical distributions

Elliptical distributions are highly symmetrical distributions that are analytically tractable and yet flexible enough to model a wide range of situations. Refer to Fang, Kotz, and Ng (1990) and Fang and Zhang (1990) for more details.

Consider an N-dimensional random variable $\mathbf{X}$, whose distribution we represent by means of its probability density function $f_{\mathbf{X}}$. Consider the iso-probability contours:

$$C_L \equiv \{\mathbf{x} \text{ such that } f_{\mathbf{X}}(\mathbf{x}) = L\}. \tag{2.256}$$

The random variable $\mathbf{X}$ is *elliptically distributed* with location parameter $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ if for all levels $L \in (0, \infty)$ the iso-probability contour is the surface of the following ellipsoid:

$$\mathcal{E}^{q(L)}_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \equiv \left\{\mathbf{x} \text{ such that } (\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \leq q(L)^2\right\}, \tag{2.257}$$

for a suitable function $q(L)$, see $(A.73)$ for the details of the geometrical interpretation of this locus.

Examples of such distributions are the normal, Student t and Cauchy distributions respectively, as we see from the right portion of Figure 2.13, Figure 2.14 and Figure 2.15 respectively.

An equivalent characterization of an elliptical distribution is the following. Consider a random variable $\mathbf{Y}$ whose distribution is *spherically symmetrical*, i.e. such that for any rotation, as represented by the matrix $\boldsymbol{\Gamma}$, the distributions of the original variable and the rotated variable are the same: $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\Gamma}\mathbf{Y}$. The probability density function of a spherically symmetrical random variable must be constant on any sphere centered in zero. Therefore, as we show in Appendix www.2.4, an elliptical random variable with location parameter $\boldsymbol{\mu}$ and scatter parameter $\boldsymbol{\Sigma}$ is an invertible affine transformation of a spherically symmetrical random variable:

$$\mathbf{X} \equiv \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}, \tag{2.258}$$

where $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}$.

To obtain a final, equivalent characterization of elliptical distributions, we notice that in general we can write any non-zero random variable $\mathbf{Y}$ as follows: $\mathbf{Y} = R\mathbf{U}$, where $R \equiv \|\mathbf{Y}\|$ is the norm of $\mathbf{Y}$ and thus it is a univariate random variable, and $\mathbf{U} \equiv \mathbf{Y}/\|\mathbf{Y}\|$.

It can be proved that if $\mathbf{Y}$ is spherically symmetrical, then R and $\mathbf{U}$ are independent and $\mathbf{U}$ is uniformly distributed on the surface of the unit ball $\mathcal{E}_{\mathbf{0},\mathbf{I}}$ in N dimensions. Therefore a final equivalent definition of an elliptical distribution with location parameter $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ is the following:

$$\mathbf{X} \equiv \boldsymbol{\mu} + R\mathbf{A}\mathbf{U}. \tag{2.259}$$

In this expression

$$\mathbf{A}\mathbf{A}' \equiv \boldsymbol{\Sigma}, \quad R \equiv \left\|\mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\|, \quad \mathbf{U} \equiv \frac{\mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})}{\|\mathbf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})\|}, \tag{2.260}$$

and $\mathbf{U}$ is uniformly distributed on the surface of the unit ball and is independent of R.

We show in Appendix www.2.18 that the generic elliptical probability density function must be of the form:

$$f_{\boldsymbol{\mu},\Sigma}(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g_N\left(\mathrm{Ma}^2(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma})\right), \tag{2.261}$$

where g_N is a non-negative univariate function that satisfies

$$\int_0^\infty v^{\frac{N}{2}-1} g_N(v)\, dv < \infty; \tag{2.262}$$

the parameter $\boldsymbol{\mu}$ is the center of the ellipsoid $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$; the parameter $\boldsymbol{\Sigma}$ is a symmetric and positive matrix that determines the shape of the ellipsoid $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$; and Ma is the Mahalanobis distance of the point $\mathbf{x}$ from $\boldsymbol{\mu}$ through the metric $\boldsymbol{\Sigma}$, as defined in (2.61).

For example, for the uniform distribution from (2.145) we obtain:

$$g_N^{\mathrm{U}}\left(\mathrm{Ma}^2\right) \equiv \frac{\Gamma\left(\frac{N}{2}+1\right)}{\pi^{\frac{N}{2}}} \mathbb{I}_{[0,1]}\left(\mathrm{Ma}^2\right). \tag{2.263}$$

For the normal distribution from (2.156) we obtain:

$$g_N^{\mathrm{N}}\left(\mathrm{Ma}^2\right) \equiv \frac{e^{-\frac{\mathrm{Ma}^2}{2}}}{(2\pi)^{\frac{N}{2}}}. \tag{2.264}$$

For the Student t distribution from (2.188) we obtain:

$$g_N^{\mathrm{St}}\left(\mathrm{Ma}^2\right) \equiv \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{N}{2}}} \left(1+\frac{\mathrm{Ma}^2}{\nu}\right)^{-\frac{\nu+N}{2}}, \tag{2.265}$$

which also covers the Cauchy distribution as the special case $\nu \equiv 1$, see (2.209).

Therefore, all the above are elliptical distributions.

Equivalently, elliptical distributions can be represented in terms of their characteristic function. The generic elliptical characteristic function has the following form:

$$\phi_{\boldsymbol{\mu},\Sigma}(\boldsymbol{\omega}) \equiv e^{i\boldsymbol{\omega}'\boldsymbol{\mu}} \psi\left(\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}\right), \tag{2.266}$$

where ψ is a suitable real-valued function.

For example, we see from (2.157) that for the normal distribution we have:

$$\psi^{\mathrm{N}}(\gamma) \equiv e^{-\frac{\gamma}{2}}. \tag{2.267}$$

The expression of ψ for the uniform distribution is given in (2.147). It is immediate to derive the expression of ψ for the Cauchy distribution from (2.210) and for the Student t distribution from (2.189).

Since an elliptical distribution is fully determined by the location parameter $\boldsymbol{\mu}$, the dispersion parameter $\boldsymbol{\Sigma}$ and the generator g of the probability density function (or equivalently the generator ψ of the characteristic function), we use the following notation to denote that a variable $\mathbf{X}$ is elliptically distributed with the above parameters:

$$\mathbf{X} \sim \mathrm{El}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_N\right), \tag{2.268}$$

where we emphasized that the generator g depends on the dimension N of the random variable $\mathbf{X}$.

For example, the normal distribution is elliptical and thus from (2.264) the following notations are equivalent:

$$\mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \equiv \mathrm{El}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \frac{e^{-\frac{1}{2}\cdot}}{(2\pi)^{\frac{N}{2}}}\right). \tag{2.269}$$

Among the most remarkable properties of elliptical distributions we mention their behavior under affine transformations. Indeed, affine transformations of elliptically distributed random variables are elliptical distributed and the new location-dispersion parameters are easily computed in terms of the original ones.

More precisely, if $\mathbf{X}$ is an N-dimensional elliptical variable as in (2.268), then for any K-dimensional vector $\mathbf{a}$ and any $K \times N$ matrix $\mathbf{B}$ the following relation holds:

$$\mathbf{a} + \mathbf{B}\mathbf{X} \sim \mathrm{El}\left(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}', g_K\right). \tag{2.270}$$

Notice nonetheless that the new generator g_K has in general a very different functional form than the original generator g_N.

For example, consider a the bivariate uniform distribution on the unit circle. In the above notation, its distribution reads:

$$(X_1, X_2)' \sim \mathrm{El}\left(\mathbf{0}, \mathbf{I}_2, g_2\right), \tag{2.271}$$

where $\mathbf{I}$ is the identity matrix and from (2.150) the two-dimensional generator is defined in terms of the indicator function $(B.72)$ as follows:

$$g_2\left(r^2\right) \equiv \frac{1}{\pi}\mathbb{I}_{[0,1]}\left(r^2\right). \tag{2.272}$$

Now consider the affine transformation determined by the following choice:

$$\mathbf{a} \equiv 0, \quad \mathbf{B} \equiv (1, 0). \tag{2.273}$$

The outcome of the transformation is the marginal distribution of the first variable X_1. From (2.270) we obtain:

$$X_1 \sim \mathrm{El}(0, 1, g_1), \tag{2.274}$$

where from (2.151) the one-dimensional generator reads:

$$g_1(r^2) \equiv \frac{2}{\pi}\sqrt{1-r^2}. \tag{2.275}$$

Therefore the marginal distribution of a uniform random variable is elliptical, but it is not uniform.

Further remarkable properties of the elliptical distributions regard the moments, when these are defined. As we show in Appendix www.2.18 the following relation holds:

$$\mathrm{E}\{\mathbf{X}\} = \boldsymbol{\mu}, \quad \mathrm{Cov}\{\mathbf{X}\} = \frac{\mathrm{E}\{R^2\}}{N}\boldsymbol{\Sigma}, \tag{2.276}$$

where R is defined in (2.260).

More in general, for the central moments of any order (2.92) we obtain:

$$\mathrm{CM}^{\mathbf{X}}_{m_1\cdots m_k} = \mathrm{E}\{R^k\} \sum_{n_1,\ldots,n_k=1}^{N} A_{m_1 n_1}\cdots A_{m_k n_k}\,\mathrm{E}\{U_{n_1}\cdots U_{n_k}\}. \tag{2.277}$$

In this expression $\mathbf{A}$ and $\mathbf{U}$ are defined in (2.260). The moments of the uniform distribution on the surface of the unit ball are null if any variable appears an odd number of times; otherwise they read:

$$\mathrm{E}\{U_1^{2s_1}\cdots U_N^{2s_N}\} = \frac{\prod_{n=1}^{N}\frac{(2s_n)!}{4^{s_n}s_n!}}{\frac{N}{2}\left(\frac{N}{2}+1\right)\cdots\left(\frac{N}{2}+\left(\sum_{n=1}^{N}s_n\right)-1\right)}. \tag{2.278}$$

Since the copula of a distribution does not depend on purely marginal parameters such as the expected value and the standard deviation, for elliptical random variable the copula is fully determined by the correlations, see for instance the normal case (2.176). As a consequence, since the measures of concordance are defined in terms of the copula of a distribution, the measures of concordance between the entries of an elliptical random variable $\mathbf{X}$ are fully determined by the correlation matrix. For instance, Lindskog, McNeil, and Schmock (2003) prove that Kendall's tau (2.128) is the following function of correlation:

$$\tau\{X_m, X_n\} = \frac{2}{\pi}\arcsin\left(\mathrm{Cor}\{X_m, X_n\}\right), \tag{2.279}$$

which extends the result for the normal case (2.178) to generic elliptical variables.

2.7.2 Stable distributions

In view of our applications, *stable distributions* are analytically tractable distributions that can be projected to specific horizons in the future, see Figure 3.11. For more results on stable distributions see e.g. Embrechts, Klueppelberg, and Mikosch (1997) and references therein.

Consider three independent random variables $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ with the same multivariate distribution. That distribution is stable if for any positive constants α and β there exist constants γ and δ such that the following holds:

$$\alpha\mathbf{X} + \beta\mathbf{Y} \stackrel{d}{=} \gamma + \delta\mathbf{Z}, \tag{2.280}$$

where "$\stackrel{d}{=}$" denotes "equal in distribution". In other words, the distribution is closed under linear combinations.

For example, assume that the three variables are independently normally distributed:

$$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.281}$$

Then

$$\alpha\mathbf{X} + \beta\mathbf{Y} \sim \mathrm{N}((\alpha+\beta)\boldsymbol{\mu}, (\alpha+\beta)\boldsymbol{\Sigma}).$$

Therefore setting $\gamma \equiv 0$ and $\delta \equiv (\alpha + \beta)$ the relation (2.280) is satisfied and thus the normal distribution is stable.

The Cauchy distribution (2.208) is stable. The lognormal distribution (2.217) is not stable, as the sum of lognormal variables is not lognormal. Similarly, for a generic number ν of degrees of freedom the Student t distribution (2.187) is not stable.

In view of our applications we are particularly interested in symmetric stable distributions, such as the normal distribution and the Cauchy distribution. Symmetric stable distributions are best represented in terms of their characteristic function. Indeed, a random variable $\mathbf{X}$ has a *symmetric stable distribution* if and only if its characteristic function has the following form:

$$\begin{aligned}\phi_{\mathbf{X}}(\boldsymbol{\omega}) &\equiv \mathrm{E}\left\{e^{i\boldsymbol{\omega}'\mathbf{X}}\right\} \\ &= e^{i\boldsymbol{\omega}'\boldsymbol{\mu}} \exp\left(-\int_{\mathbb{R}^N} |\boldsymbol{\omega}'\mathbf{s}|^{\alpha} m_{\boldsymbol{\Sigma}}(\mathbf{s})\, d\mathbf{s}\right).\end{aligned} \tag{2.282}$$

In this expression the parameter $\boldsymbol{\mu}$ is a location vector and the parameter α is a scalar that determines such features as the thickness of the tails of the distribution. The (generalized) function $m_{\boldsymbol{\Sigma}}$ defines a symmetric measure that is non-zero on the surface of the ellipsoid $\mathcal{E}_{\mathbf{0},\boldsymbol{\Sigma}}$ with shape parameter $\boldsymbol{\Sigma}$ centered in zero, see $(A.73)$. In formulas:

$$m_{\boldsymbol{\Sigma}}(\mathbf{s}) = m_{\boldsymbol{\Sigma}}(-\mathbf{s}), \text{ for all } \mathbf{s} \in \mathbb{R}^N, \tag{2.283}$$

and

$$m_{\boldsymbol{\Sigma}}(\mathbf{s}) \equiv 0 \text{ for all } \mathbf{s} \text{ such that } \mathbf{s}'\boldsymbol{\Sigma}^{-1}\mathbf{s} \neq 1. \tag{2.284}$$

We use the following notation to indicate that $\mathbf{X}$ has a symmetric stable distribution with the above parameters:

$$\mathbf{X} \sim \mathrm{SS}(\alpha, \boldsymbol{\mu}, m_{\boldsymbol{\Sigma}}). \tag{2.285}$$

Symmetric stable distributions are also called *symmetric-alpha-stable* ($s\alpha s$) *distributions*.

For example, consider a normally distributed random variable:

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.286}$$

Consider the spectral decomposition ($A.70$) of the covariance matrix:

$$\boldsymbol{\Sigma} \equiv \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{E}', \tag{2.287}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues of $\mathbf{S}$:

$$\boldsymbol{\Lambda} \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_N); \tag{2.288}$$

and $\mathbf{E}$ is the juxtaposition of the respective eigenvectors:

$$\mathbf{E} \equiv \left(\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(N)}\right). \tag{2.289}$$

Define N vectors $\left\{\mathbf{v}^{(1)}, \ldots \mathbf{v}^{(N)}\right\}$ as follows:

$$\left(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(N)}\right) \equiv \mathbf{V} \equiv \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}. \tag{2.290}$$

Define the following measure:

$$m_{\boldsymbol{\Sigma}} \equiv \frac{1}{4}\sum_{n=1}^{N}\left(\delta^{(\mathbf{v}_n)} + \delta^{(-\mathbf{v}_n)}\right), \tag{2.291}$$

where $\delta^{(\mathbf{x})}$ is the Dirac delta centered in $\mathbf{x}$ as defined in ($B.16$).

We prove in Appendix www.2.19 the following results. The measure $m_{\boldsymbol{\Sigma}}$ satisfies (2.283) and (2.284). In turn, the characteristic function (2.157) of the normal distribution can be written as follows:

$$\phi^{\mathrm{N}}_{\boldsymbol{\mu},\Sigma}(\boldsymbol{\omega}) = e^{i\boldsymbol{\mu}'\boldsymbol{\omega}} \exp\left(-\int_{\mathbb{R}^N} |\boldsymbol{\omega}'\mathbf{s}|^2 m_{\boldsymbol{\Sigma}}(\mathbf{s})\, d\mathbf{s}\right). \tag{2.292}$$

Therefore the following notations are equivalent:

$$\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv \mathrm{SS}\left(2, \boldsymbol{\mu}, \frac{1}{4}\sum_{n=1}^{N}\left(\delta^{(\mathbf{v}_n)} + \delta^{(-\mathbf{v}_n)}\right)\right). \tag{2.293}$$

We remark that one should not confuse stability with additivity: a distribution is *additive* if the sum of two variables with that distribution belongs to the same class of distributions. Indeed, stable distributions are additive, but the reverse implication is not true.

For example, consider three independent random matrices that are Wishart-distributed with the same scale factor:

$$(\mathbf{W}, \mathbf{S}, \boldsymbol{\Omega}) \sim \mathrm{W}(\nu, \boldsymbol{\Sigma}). \tag{2.294}$$

Then:

$$\mathbf{W} + \mathbf{S} \sim \mathrm{W}(2\nu, \boldsymbol{\Sigma}). \tag{2.295}$$

This follows easily from the definition (2.222) of the Wishart distribution. Therefore the Wishart distribution for a given scale parameter is additive. Nevertheless

$$\mathbf{W} + \mathbf{S} \overset{d}{\neq} \gamma + \delta\boldsymbol{\Omega}. \tag{2.296}$$

Therefore the Wishart distribution for a given scale parameter is not stable.

2.7.3 Infinitely divisible distributions

In view of our applications, *infinitely divisible* distributions can be projected to a generic investment horizon, see Figure 3.11, although the computation might not be straightforward.

More formally, the distribution of a random variable $\mathbf{X}$ is infinitely divisible if, for any integer T, the distribution of $\mathbf{X}$ is the same as the distribution of the sum of T suitably chosen independent and identically distributed random variables:

$$\mathbf{X} \overset{d}{=} \mathbf{Y}_1 + \cdots + \mathbf{Y}_T. \tag{2.297}$$

For example, assume that $\mathbf{X}$ is normally distributed:

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.298}$$

For an arbitrary integer T consider the following set of independent and identically distributed normal random variables:

$$\mathbf{Y}_t \sim \mathrm{N}\left(\frac{\boldsymbol{\mu}}{T}, \frac{\boldsymbol{\Sigma}}{T}\right). \tag{2.299}$$

It is immediate to check that these variables satisfy (2.297). Therefore the normal distribution is stable.

Unlike in the normal case, although for a given T the distribution of all the terms $\mathbf{Y}_\nu$ in (2.297) is the same, this distribution need not be the same for all values of T.

For instance, the lognormal distribution is infinitely divisible, see Thorin (1977). Nevertheless, unlike in the normal case (2.299), the distribution of the divisors depends on T.

Many common distributional models are infinitely divisible. For instance, the elliptical distributions discussed in this book are infinitely divisible.

On the other hand, not all distributions are infinitely divisible. For example, the Wishart distribution is not infinitely divisible, except in the univariate case. Indeed the gamma distribution, which is the one-dimensional Wishart distribution, is infinitely divisible, see Cuppens (1975).