# 1

# Preliminaries from Probability Theory

This chapter reviews some important results from probability theory and fixes notation. First we introduce discrete and continuous random variables and their distributions. Then we discuss functionals of random variables such as moments. Furthermore, we introduce certain classes of distributions and also multivariate distributions together with copulas.

## 1.1 Discrete Random Variables and Distributions

In financial markets one can observe the prices of assets such as stocks, commodities, currencies, futures, bonds etc. It is a challenge to model these random quantities in a satisfactory manner.

### Log-Returns

Let us assume that we observe an asset price at times $t_i = i\Delta$ for $i \in \{0, 1, \ldots\}$ with time step size $\Delta > 0$. The time $\Delta$ between two successive observations is typically the length of one day. If $X_{t_i}$ denotes the asset price at time $t_i$, then the *log-return* $R_{t_i}$ at this time is defined as

$$R_{t_i} = \ln(X_{t_{i+1}}) - \ln(X_{t_i}) = \ln\left(\frac{X_{t_{i+1}}}{X_{t_i}}\right) \tag{1.1.1}$$

for $i \in \{0, 1, \ldots\}$.

We define the daily log-return of an asset price as the daily increment of the natural logarithm of this price because, as we shall see later on, this reflects well the growth nature of economies and financial markets. Typically log-returns exhibit considerable variability.

We focus in this book on the modeling of log-returns while we introduce the basic concepts of probability, statistics, stochastic processes, stochastic calculus and stochastic differential equations. It will turn out that stochastic

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

differential equations provide an ideal mathematical framework for the modeling of financial quantities. In this context log-returns will also allow us to apply the powerful tools of stochastic calculus. This is not so conveniently achieved when using, so-called, *returns* that are of the form

$$\tilde{R}_{t_i} = \frac{X_{t_{i+1}} - X_{t_i}}{X_{t_i}}$$

and closely approximate log-returns when these are small. As we shall see, log-returns are more tractable in continuous time.

### Relative Frequencies and Probabilities

Let us interpret an asset's log-return $R_{t_i}$ as the *outcome* of an experiment based on observations of the data. Suppose, for simplicity, that we classify the log-returns as strictly negative, zero or positive. We denote these *elementary outcomes* or *states* by $\omega_1$, $\omega_2$, $\omega_3$, indicating that we observe a negative, zero or strictly positive log-return, respectively. We call the set of outcomes or states $\Omega = \{\omega_1, \omega_2, \omega_3\}$ the *sample space* for our experiment.

If we repeat our experiment $N$ times, that is, we observe for a stock daily log-returns on $N$ different days, and count the number $N(\omega_i)$ of times, that the outcome $\omega_i$ occurs, we can form the *relative frequency*

$$f_i(N) = \frac{N(\omega_i)}{N}.$$

For smaller $N$ this number usually varies considerably. As $N$ becomes larger, our experience would indicate that the relative frequency should approach a limit $p_i$, written as

$$\lim_{N\to\infty} f_i(N) = p_i,$$

which we call the *probability* of outcome $\omega_i$.

To illustrate the above example let us look at the daily IBM share price in US dollars over the period from 1977 until 1997, which is shown in Fig. 1.1.1. The corresponding log-returns are plotted in Fig. 1.1.2. In Fig. 1.1.3 we then display the relative frequencies $f_1(t_i)$, $f_2(t_i)$ and $f_3(t_i)$, $i \in \{0, 1, \ldots\}$, of negative, zero and strictly positive log-returns, respectively, during the time period. Note that after some wild fluctuations for small time $t$, at the beginning of the period, the relative frequency for negative log-returns stabilizes around a value close to $p_1 = 0.465$. Similarly, we obtain at the end of the period a value $p_3 = 0.463$ for the relative frequency of strictly positive log-returns. The value $p_2 = 0.072$ is then obtained for the rather small probability of zero log-returns. Clearly, we have $0 \le p_i \le 1$ for each $i \in \{1, 2, 3\}$ and $\sum_{i=1}^{3} p_i = 1$, that is, the probabilities $p_1$, $p_2$ and $p_3$ add up to one.

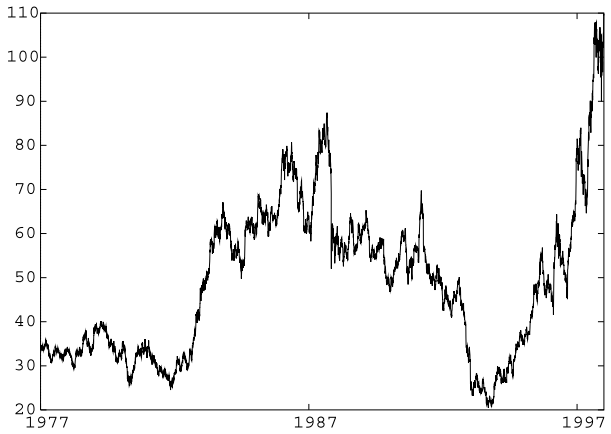


**Fig. 1.1.1.** IBM share price from 1977 until 1997

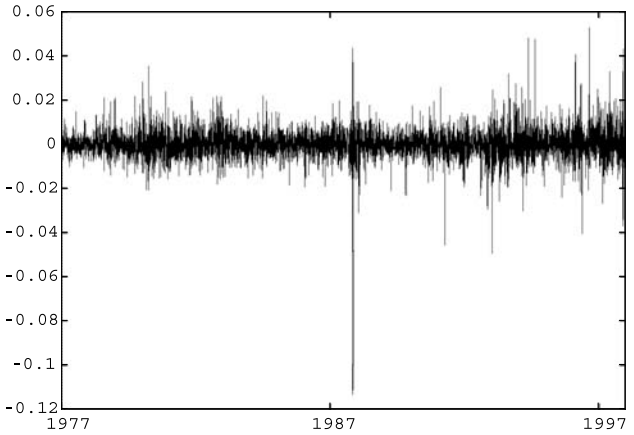


**Fig. 1.1.2.** Log-returns of IBM stock

### Probability Space

To analyze a model one is often interested in combinations of outcomes. We call such a combination an *event* if we can identify it either by its occurrence or its non-occurrence. Obviously, if a subset $A$ of the set of outcomes $\Omega$ is an event, then its *complement* $A^c = \{\omega_i \in \Omega : \omega_i \notin A\}$, which denotes the set of all $\omega_i$ from the sample space $\Omega$ that do not belong to the set $A$, must also be an event. In the case of the above example we might consider the event $A = \{\omega_1, \omega_2\}$ that corresponds to the occurrence of either a negative or zero log-return. The complement of this event is then $A^c = \{\omega_i \in \Omega : \omega_i \notin \{\omega_1, \omega_2\}\} = \{\omega_3\}$. This is the event $\{\omega_3\}$ of a strictly positive log-return.

In particular, the whole sample space $\Omega$ is an event, which is called the *sure event* since one of its outcomes must always occur. The complement of $\Omega$

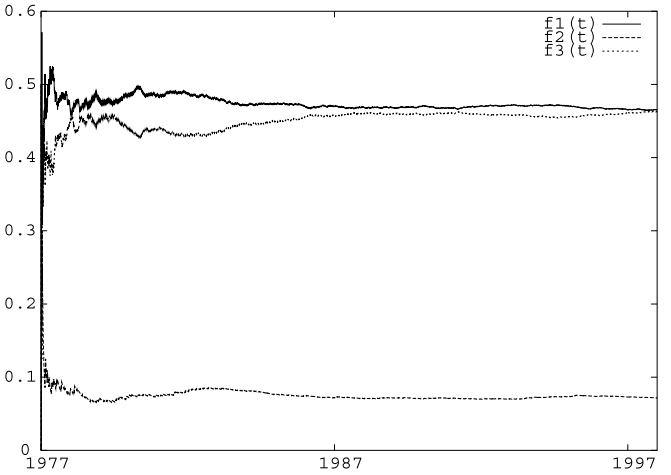


**Fig. 1.1.3.** Relative frequency over time

is the *empty set* $\emptyset$, which is also defined as an event but never occurs. If $A$ and $B$ are events, then the event $A \cup B$ occurs if either $A$ or $B$ occurs, whereas the event $A \cap B$ occurs if both $A$ and $B$ occur. With $A = \{\omega_1, \omega_2\}$, as in our example, and the event $B = \{\omega_2\}$ indicating a zero log-return we note that $A \cup B = \{\omega_1, \omega_2\} \cup \{\omega_2\} = \{\omega_1, \omega_2\}$ stands for an event consisting of either negative or zero log-returns and $A \cap B = \{\omega_1, \omega_2\} \cap \{\omega_2\} = \{\omega_2\}$ is the event which indicates only a zero log-return.

In the above discussion we have only mentioned experiments with a finite number of outcomes. However, the introduction of probabilities based on an infinite set of outcomes and the use of relative frequencies to define probabilities can lead to conceptual subtleties and other mathematical problems. To resolve these difficulties, Kolmogorov developed in the late 1920s an axiomatic approach to probability theory. In this approach the probabilities represent numbers assigned to corresponding events. In what follows we shall employ this axiomatic framework.

Let us denote by $P(A)$ the *probability* of the occurrence of an event $A$ that is taken from the *collection of events* $\mathcal{A}$ that corresponds to the sample space $\Omega$. Then from corresponding properties of relative frequencies we would expect these probabilities to satisfy the following relationships

$$0 \leq P(A) \leq 1, \tag{1.1.2}$$

$$P(A^c) = 1 - P(A), \tag{1.1.3}$$

$$P(\emptyset) = 0, \qquad P(\Omega) = 1, \tag{1.1.4}$$

and

$$P(A \cup B) = P(A) + P(B) \tag{1.1.5}$$

if $A$ and $B$ are exclusive, that is $A \cap B = \emptyset$ for events $A$ and $B$ taken from $\mathcal{A}$.

The above relationships allow us, for a given finite sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, consistently to allocate probabilities to each event. One can deduce that

$$\bigcup_{i=1}^{n} A_i \quad \text{and} \quad \bigcap_{i=1}^{n} A_i$$

are events if $A_1, A_2, \ldots, A_n$ are events, and that

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

if $A_1, A_2, \ldots, A_n$ are *mutually exclusive*, that is if $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$.

For the above example suppose we assign the probabilities $p_i = P(\{\omega_i\})$ for each outcome $\omega_i$, $i \in \{1, 2, 3\}$, as obtained from frequency records. Then the event $A = \{\omega_1, \omega_2\}$ of non-strictly positive outcomes has, according to (1.1.5), the probability

$$P(A) = P(\{\omega_1, \omega_2\}) = P(\{\omega_1\} \cup \{\omega_2\}) = P(\{\omega_1\}) + P(\{\omega_2\}) = p_1 + p_2.$$

The essential probabilistic information that characterizes an experiment can be succinctly summarized in the corresponding triplet $(\Omega, \mathcal{A}, P)$ consisting of the sample space $\Omega$, the collection of events $\mathcal{A}$ and the probability measure $P$, where these have to satisfy certain relationships. In the above analysis we have considered finite collections of events. To cover the case of infinite collections we must specify these properties to avoid contradictions. We assume that the collection of events $\mathcal{A}$ is a *sigma-algebra*, which means that

$$\Omega \in \mathcal{A}, \tag{1.1.6}$$

$$\text{if } A \in \mathcal{A} \text{ then } A^c \in \mathcal{A}, \tag{1.1.7}$$

$$\text{if } A \in \mathcal{A} \text{ and } B \in \mathcal{A} \text{ then } A \cup B \in \mathcal{A} \tag{1.1.8}$$

$$\text{and if } A_i \in \mathcal{A} \text{ for any } i \in \mathcal{N} = \{1, 2, \ldots\} \text{ then } \left(\bigcup_{i=1}^{\infty} A_i\right) \in \mathcal{A}. \tag{1.1.9}$$

In the case of infinite collections, equation (1.1.5) is replaced by what is called *countably additive probabilities*. This means,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \tag{1.1.10}$$

for any sequence $(A_i)_{i \in \mathcal{N}}$ of mutually exclusive events.

It can be shown by DeMorgan's law that a sigma-algebra is closed under finite and countable intersections of events. In addition, if a set function defined on a sigma-algebra satisfies (1.1.2) and (1.1.10) with $P(\Omega) = 1$, then

(1.1.3)–(1.1.5) also hold and hence this set function would be a probability measure.

A triplet $(\Omega, \mathcal{A}, P)$ is then called a *probability space* and the relations (1.1.2)–(1.1.5) can be shown to form a consistent set of rules for modeling probabilities in this space. This kind of structure will be used for all of our modeling work described in this book. Since the models that we can construct will always remain abstract objects, they can reflect reality only to a limited extent. It will be our aim to introduce more and more flexible mathematical structures that provide the potential to model successfully complex stochastic phenomena in finance. However, the reader should never believe that there is anything like a perfect model. Even if some model were to become very successful, the market would regularly demand further modifications and extensions to the model.

The relations (1.1.2)–(1.1.5) allow us to prove in a straightforward manner that if $A, B \in \mathcal{A}$ and $A \subseteq B$ then

$$P(A) \leq P(B). \tag{1.1.11}$$

Furthermore, if $A, B \in \mathcal{A}$ then

$$P(A \cap B^c) = P(A) - P(A \cap B). \tag{1.1.12}$$

There may be some events $A$ with $P(A) = 0$. These are then called *null events*. On the other hand, there may be some event $B$ for which $P(B) = 1$. In this case we say $B$ has occurred *almost surely* (a.s.) or *with probability one*.

### Probabilities

The probability $P(A)$ of an event $A$ can be interpreted as a measure of the likelihood that $A$ occurs. If we have some additional information, such as that another event has occurred, then our estimate of this likelihood may change. For instance, if we know in the above example that the event $A = \{\omega_1, \omega_2\}$ of having no strictly positive log-return has occurred, then conditioned on this information, the conditional probabilities of observing negative or zero log-returns will add up to one. We denote by $P(\{\omega_1\} \,|\, A)$ the *conditional probability* that a negative log-return, the outcome $\omega_1$, will be observed, given that the event $A = \{\omega_1, \omega_2\}$ has occurred. Note that this conditional probability can be expressed by the ratio

$$P(\{\omega_1\} \,|\, A) = \frac{P(\{\omega_1\} \cap A)}{P(A)} = \frac{P(\{\omega_1\})}{P(\{\omega_1, \omega_2\})},$$

where $P(A) > 0$. This relation is readily suggested from the ratio of relative frequencies

$$\frac{f_1(N)}{f_1(N) + f_2(N)} = \frac{\frac{N(\omega_1)}{N}}{\frac{N(\omega_1)}{N} + \frac{N(\omega_2)}{N}} = \frac{N(\omega_1)}{N(\omega_1) + N(\omega_2)},$$

where $N(\omega_1)$ and $N(\omega_2)$ denote the number of outcomes $\omega_1$ and $\omega_2$, respectively, that have occurred out of $N$ repetitions of the experiment.

In general, the *conditional probability* $P(A \,|\, B)$ for the event $A$ given that the event $B$ has occurred is defined by the formula

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)} \tag{1.1.13}$$

provided $P(B) > 0$. This formula is also called the *Bayes formula*. As a consequence of (1.1.13) one obtains

$$P(A)\, P(B \,|\, A) = P(B)\, P(A \,|\, B), \tag{1.1.14}$$

which is sometimes called *Bayes' Theorem*.

Conditional probabilities have similar properties to ordinary probabilities, for instance, they sum to one, when conditioned on the same $B$.

The likelihood for the occurrence of an event could be unaffected by whether or not another event $B$ has occurred. In such a case the conditional probability $P(A \,|\, B)$ should equal $P(A)$, which implies together with (1.1.13) that

$$P(A \cap B) = P(A)\, P(B). \tag{1.1.15}$$

We say that the events $A$ and $B$ are *independent* if and only if (1.1.15) holds. By assuming $P(B) > 0$ and rearranging formula (1.1.15) we see that events $A$ and $B$ are independent if

$$P(A) = \frac{P(A \cap B)}{P(B)}. \tag{1.1.16}$$

For instance, if we extend slightly our example and consider the log-returns from two different days to be independent, then the event characterizing the log-return from the first day does not affect the event that describes the log-return for the second day. In this example the second log-return is assumed to be not influenced by the outcome of the first log-return and vice versa.

More generally we say that $m$ events $A_1, A_2, \ldots, A_m$ are *independent* if

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})\, P(A_{i_2}) \,\ldots\, P(A_{i_k}) \tag{1.1.17}$$

for all $k \in \mathcal{N}$ and non-empty subsets $\{i_1, i_2, \ldots, i_k\}$ of the set of indices $\{1, 2, \ldots, m\}$.

One can show that if $A_1, A_2, B \in \mathcal{A}$ and $P(B) > 0$, then

$$P\left(A_1 \cap A_2^c \,|\, B\right) = P\left(A_1 \,|\, B\right) - P\left(A_1 \cap A_2 \,|\, B\right). \tag{1.1.18}$$

A sequence of events $(A_i)_{i \in \mathcal{N}}$ with $A_i \in \mathcal{A}$ for all $i \in \mathcal{N}$ is called a *partition* of $\Omega$ if

$$\bigcup_{i=1}^{\infty} A_i = \Omega, \tag{1.1.19}$$

and $A_\ell \cap A_m = \emptyset$ for all $\ell \neq m$. This allows us to formulate the following statement on the *total probability*. If $(A_i)_{i\in\mathcal{N}}$ is a partition of $\Omega$ with $P(A_i) > 0$ for all $i \in \mathcal{N}$, then for any event $B \in \mathcal{A}$ one obtains the representation

$$P(B) = \sum_{i=1}^{\infty} P\left(B \,\middle|\, A_i\right)\, P(A_i). \tag{1.1.20}$$

This formula can be very helpful for calculating the probabilities of certain events.

### Random Variables and Distributions

We are often interested in assigning some numerical quantity to the outcomes of a probabilistic experiment. For instance, in our stock log-return example, the quantity $X(\omega)$ might take the value 1 for a strictly positive log-return, 0 for a zero log-return and $-1$ for a negative log-return.

These assigned quantities correspond to the values taken by a function $X : \Omega \to \Re$, where $\Re = (-\infty, \infty)$ is the *set of real numbers*. In our example we have

$$X(\omega) = \begin{cases} 1 & \text{for} \quad \omega = \omega_1 \\ 0 & \text{for} \quad \omega = \omega_2 \\ -1 & \text{for} \quad \omega = \omega_3. \end{cases} \tag{1.1.21}$$

More generally, given a probability space $(\Omega, \mathcal{A}, P)$ we say, that a function $X : \Omega \to \Re$ is an $\mathcal{A}$-*measurable function* or a *random variable* if the set $\{\omega \in \Omega : a < X(\omega) \leq b\}$ is an event for each $a, b \in \Re$ with $a < b$. This means that this set is an element of $\mathcal{A}$. Using this definition it can be shown that if $X$ is a random variable, then it holds for any Borel subset of the real line

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A},$$

see Shiryaev (1984). We say that two random variables $X$ and $Y$ are *independent* if the corresponding events $\{\omega \in \Omega : X(\omega) \leq a\}$ and $\{\omega \in \Omega : Y(\omega) \leq b\}$ are independent for all $a, b \in \Re$.

Now it is appropriate to introduce for a random variable $X$ its *distribution function* $F_X : \Re \to [0, 1]$ that is defined for each real valued $x \in \Re$ by the relation

$$\begin{aligned} F_X(x) &= P(\{\omega \in \Omega : X(\omega) \leq x\}) \\ &= P(X \leq x). \end{aligned} \tag{1.1.22}$$

Here we have used in the last term an abbreviated notation for the probability of an event, which will also be used in other parts of the book. In Fig. 1.1.4 we show the three probabilities, $p_1 = 0.465$, $p_2 = 0.072$ and $p_3 = 0.463$ for the stock log-return example with possible outcomes $-1, 0, 1$, respectively, that is

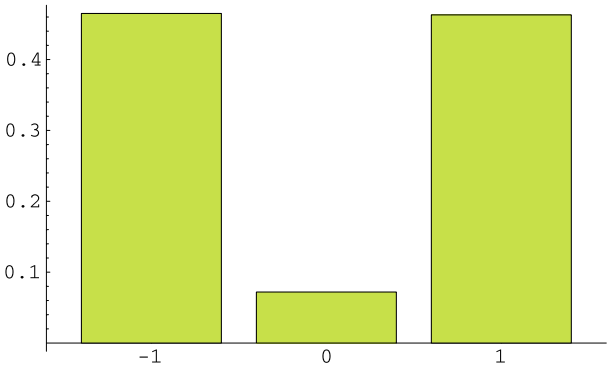


**Fig. 1.1.4.** Probabilities for the stock log-return example

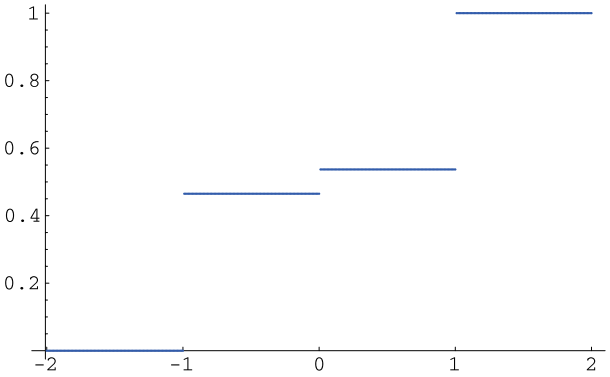


**Fig. 1.1.5.** Distribution for the stock log-return example

$$P(X=x)=\begin{cases} p_1 & \text{for} \quad x=-1 \\ p_2 & \text{for} \quad x=0 \\ p_3 & \text{for} \quad x=1. \end{cases} \tag{1.1.23}$$

The distribution function is then according to (1.1.22) given by

$$F_X(x)=\begin{cases} 0 & \text{for} \quad x<-1 \\ p_1 & \text{for} \quad -1\le x<0 \\ p_1+p_2 & \text{for} \quad 0\le x<1 \\ 1 & \text{for} \quad 1\le x \end{cases} \tag{1.1.24}$$

for $x \in \Re$, which we plot in Fig. 1.1.5.

### Two-Point Distribution

A simple random variable is the *indicator function* $\mathbf{1}_A : \Omega \to \{0,1\}$ of an event $A \in \mathcal{A}$, where

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{for} \quad \omega \in A \\ 0 & \text{for} \quad \omega \notin A. \end{cases} \tag{1.1.25}$$

Here the corresponding distribution function is of the form

$$F_{\mathbf{1}_A}(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - P(A) & \text{for} \quad 0 \le x < 1 \\ 1 & \text{for} \quad 1 \le x, \end{cases} \tag{1.1.26}$$

where $P(A)$ denotes the probability of the event $A$. This is an example of a *two-point random variable* which takes two distinct real values $x_1$ and $x_2$ with probabilities $p_1$ and $p_2 = 1 - p_1$, respectively, where $x_1 < x_2$.

It can be shown that for any random variable $X$ the limit of the value of the distribution function $F_X(x)$ for $x$ tending to minus infinity, $x \to -\infty$, is zero. That is

$$\lim_{x \to -\infty} F_X(x) = 0. \tag{1.1.27}$$

Similarly, it can be verified that

$$\lim_{x \to \infty} F_X(x) = 1 \tag{1.1.28}$$

and $F_X(x)$ is non-decreasing in $x \in \Re$.

The above examples indicate that a distribution function does not have to be continuous. However, one can show that it is always *right-continuous*, that is

$$\lim_{h \to 0+} F_X(x + h) = F_X(x) \tag{1.1.29}$$

for all $x \in \Re$.

### Poisson Distribution

An important discrete random variable is the *Poisson random variable* $X$ characterized by its *mean* $\lambda > 0$. It can be used to model, for instance, the number of trades per day that occur for a given stock or the number of bankruptcies that occur during a year. A Poisson random variable $X$ takes values $0, 1, \ldots$ without any upper bound. The corresponding probabilities $p_n = P(X = n)$ are the *Poisson probabilities* that are given by

$$p_n = \frac{\lambda^n}{n!} \exp\{-\lambda\} \tag{1.1.30}$$

for $n \in \{0, 1, \ldots\}$, where $\lambda > 0$, $n! = 1 \cdot 2 \cdot \ldots \cdot n$ for $n \in \mathcal{N}$ and $0! = 1$. These probabilities are displayed in Fig. 1.1.6 for the *intensity parameter* $\lambda = 2$. We write $X \sim P(\lambda)$ to indicate that $X$ has a Poisson distribution with intensity $\lambda$.

Let $\Omega = \mathcal{N} = \{1, 2, \ldots\}$ denote the set of natural numbers. A *discrete real valued random variable* $X$ is a measurable function from $\Omega$ into a finite or

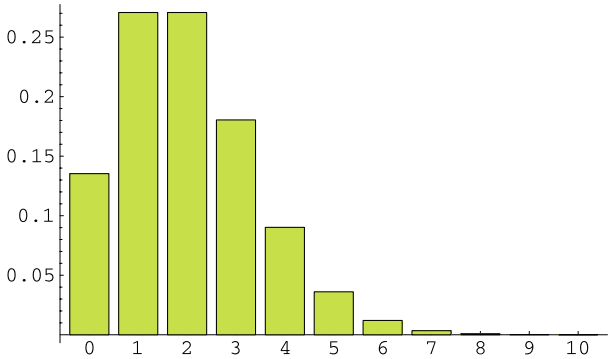


**Fig. 1.1.6.** Poisson probabilities for $\lambda = 2$

possibly infinite set of distinct real values $x_1 < x_2 < \cdots < x_n < \cdots$ with probabilities $p_n = P(X = x_n)$ for $n \in \mathcal{N}$. Its distribution function $F_X$ has the representation

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < x_1 \\ \sum_{i=1}^{n} p_i & \text{for} \quad x_n \le x < x_{n+1}, \end{cases} \tag{1.1.31}$$

for $n \in \mathcal{N}$. $F_X$ is a right-continuous step-function with steps of height $p_n$ at $x = x_n$. For this random variable the set $\{x_1, x_2, \ldots\}$ could be used as the sample space $\Omega$, with all of its subsets being events.

## 1.2 Continuous Random Variables and Distributions

The modeling of events in a financial context often requires random variables that take any value in $\Re = (-\infty, \infty)$ or subintervals of $\Re$. We call a random variable $X$ a *continuous random variable* if the probability $P(X = x)$ is zero for all $x \in \Re$. If $X$ is a continuous random variable, then the corresponding distribution function $F_X$ will also be continuous.

In cases where the distribution function $F_X$ is differentiable, there exists a nonnegative function $f_X$, called the *density function*, such that

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{1.2.1}$$

for all $x \in \Re$. However, $F_X$ could be differentiable *Lebesgue almost everywhere*, that is except possibly on a set of Lebesgue measure zero. It can be shown that if $F_X$ is absolutely continuous, then it can be expressed as integral of the form

$$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds \tag{1.2.2}$$

for all $x \in \Re$, where $f_X$ is the corresponding density function.

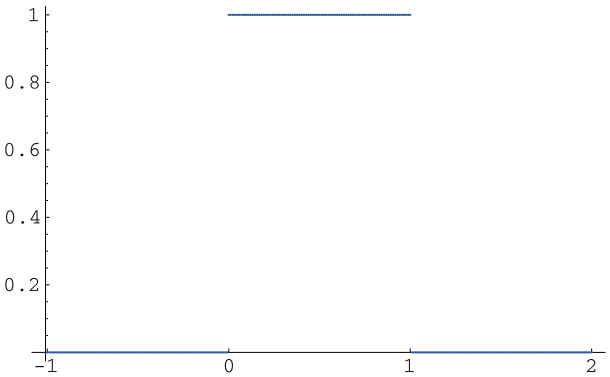


**Fig. 1.2.1.** The uniform density with $a = 0$ and $b = 1$

We shall now describe some commonly occurring examples of continuous random variables.

### Uniform Distribution

Consider a random variable $X$ which takes values only in a finite interval $[a, b)$, such that the probability of its being in a given subinterval is proportional to the length of the subinterval. Then the distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{x-a}{b-a} & \text{for} \quad a \le x < b \\ 1 & \text{for} \quad b \le x, \end{cases}$$

which is differentiable everywhere except at $x = a$ and $x = b$. The corresponding density function is then of the form

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x \notin [a, b) \\ \dfrac{1}{b-a} & \text{for} \quad x \in [a, b). \end{cases} \tag{1.2.3}$$

We say that the random variable $X$ is in this case *uniformly distributed* on $[a, b)$ and use the abbreviation $X \sim U(a, b)$ to denote this fact. For example, log-returns of a stock could be modeled by a $U(-a, a)$ distributed random variable with a parameter $a > 0$ that describes the largest possible absolute log-return. The density for a $U(0, 1)$ distributed random variable is shown in Fig. 1.2.1.

### Exponential Distribution

The waiting time between two events when there is no memory kept on the time when the first event occurred, for instance, bankruptcies, catastrophes

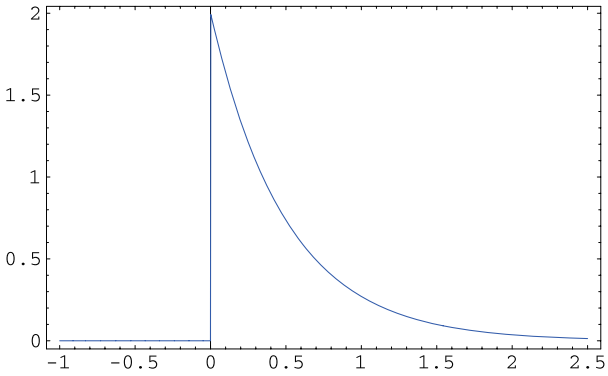


**Fig. 1.2.2.** The exponential density for intensity $\lambda = 2$

or changes in credit ratings, can be often modeled by a random variable $X$ with an *exponential distribution* given by the distribution function

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - \exp\{-\lambda\, x\} & \text{for} \quad x \geq 0 \end{cases} \tag{1.2.4}$$

for some intensity parameter $\lambda > 0$. $F_X$ is differentiable everywhere except when $x = 0$ and has as corresponding density function

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \lambda \exp\{-\lambda\, x\} & \text{for} \quad x \geq 0. \end{cases} \tag{1.2.5}$$

We write $X \sim Exp(\lambda)$ to indicate that $X$ is an exponentially distributed random variable. A larger intensity parameter $\lambda$ means that it is more likely that the waiting time between two events is shorter. In Fig. 1.2.2 we plot the density of the exponential distribution for the intensity $\lambda = 2$.

### Gaussian Distribution

The *Gaussian density function* given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \tag{1.2.6}$$

for $x \in \Re$ has a bell-shaped graph which is symmetric about $x = \mu$. In Fig. 1.2.3 we show the density of an $N(0,1)$ distributed random variable which is also called a *standard Gaussian random variable*. The corresponding standard Gaussian distribution function $F_X(x)$ is everywhere differentiable and has a sigmoidal-shaped graph, see Fig. 1.2.4. A random variable $X$ with the density function (1.2.6) is called a *Gaussian random variable* and we summarize this fact by writing $X \sim N(\mu, \sigma^2)$.

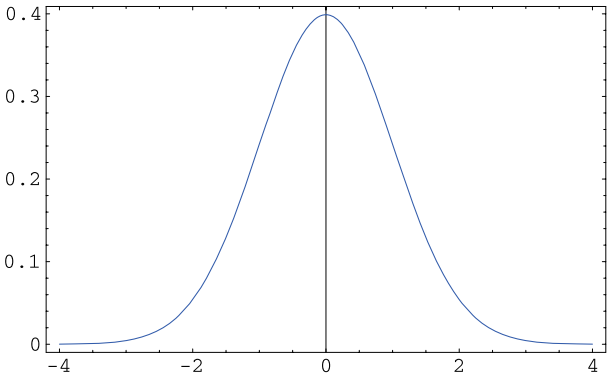


**Fig. 1.2.3.** The standard Gaussian density

Gaussian random variables occur so commonly in many applications, including financial ones, that they are often said to be *normally distributed*. The log-returns of stocks have been widely modeled as normally distributed random variables resulting in the well-known *lognormal asset price model* which we shall discuss later in detail. For this standard market model the increments of the logarithm of the stock price, the log-returns, are assumed to be normally distributed.

Unfortunately, the Gaussian distribution has no explicit analytic representation. Since it is often used in finance, for instance, in option pricing and Value at Risk calculations, it is useful to have an accurate approximation for the standard Gaussian distribution function $N : \Re \to (0,1)$. This function can be approximated, for instance, by the expression

$$\begin{aligned} N(x) = \int_{-\infty}^{x} N'(z)\,dz = 1 - 0.5\,(1 + 0.0498673470\,x + 0.0211410061\,x^2 \\ + 0.0032776263\,x^3 + 0.0000380036\,x^4 \\ + 0.0000488906\,x^5 + 0.0000053830\,x^6)^{-16} \\ + \varepsilon(x), \end{aligned} \tag{1.2.7}$$

for $x \geq 0$, where we have an error term $\varepsilon(x)$ with $|\varepsilon(x)| < 0.00000015$, as established in Abramowitz & Stegun (1972). To obtain values for $N(x)$ for $x < 0$ we can use the relation $N(x) = 1 - N(-x)$. Here $N'(\cdot)$ denotes the standard Gaussian density function

$$N'(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\,x^2\right\} \tag{1.2.8}$$

for $x \in \Re$. In Fig. 1.2.4 we graph the standard Gaussian distribution function. For statistical and other studies it is helpful to know that, for $X \sim N(\mu, \sigma^2)$,

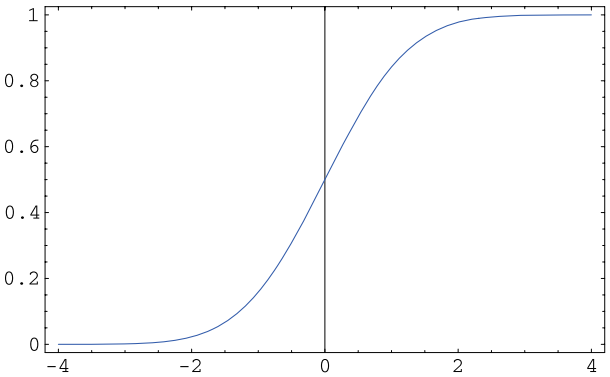


**Fig. 1.2.4.** The standard Gaussian distribution

we have, so-called, *k-sigma rules*, where $|X - \mu| < k\,\sigma$ approximately with probability 0.95 for $k = 2$, 0.9973 for $k = 3$ and 0.99994 for $k = 4$.

### Gamma Distribution

A *gamma distributed* random variable $X$ takes only positive real values and has a density function

$$f_X(x) = \frac{\alpha^p}{\Gamma(p)} \exp\{-\alpha\, x\}\, x^{p-1} \tag{1.2.9}$$

for $0 < x < \infty$ and parameters $\alpha > 0$ and $p > 0$. Here $\Gamma$ denotes the *gamma function* given by

$$\Gamma(p) = \int_0^\infty t^{p-1}\, e^{-t}\, dt \tag{1.2.10}$$

for $p > 0$. We use the abbreviation $X \sim G(p, \alpha)$ to indicate that a random variable $X$ is gamma distributed with the density function (1.2.9) for given parameters $\alpha$ and $p$. We plot in Fig. 1.2.5 the density of the gamma distribution for $\alpha = 0.5$ and $p = 2$.

In the special case $\alpha = 0.5$ the gamma distribution is equivalent to the *chi-square distribution* with $n = 2p$ degrees of freedom. For $n \in \mathcal{N}$ this distribution is obtained as that of a random variable $X$, that is the sum of the squares of $n = 2p$ independent standard Gaussian random variables. We abbreviate this by writing $X \sim \chi^2(n)$. Thus, Fig. 1.2.5 also shows a chi-square density with $n = 4$ degrees of freedom.

Let $X$ denote a chi-square distributed random variable with $n$ degrees of freedom. Its distribution function has the form

$$F_X(x) = \chi^2(x; n) = \int_0^x \frac{\exp\left\{-\frac{u}{2}\right\} \left(\frac{u}{2}\right)^{\frac{n}{2}-1}}{2\,\Gamma\left(\frac{n}{2}\right)}\, du = 1 - \frac{\Gamma\left(\frac{x}{2}; \frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \tag{1.2.11}$$

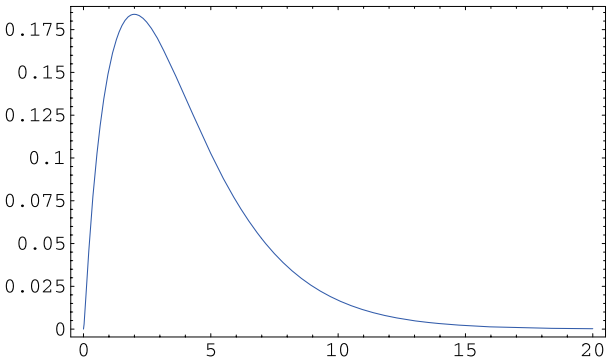


**Fig. 1.2.5.** The gamma density for $\alpha = 0.5$ and $p = 2$

for $x \geq 0$, where

$$\Gamma(u; a) = \int_u^\infty t^{a-1} \exp\{-t\}\, dt \tag{1.2.12}$$

is the *incomplete gamma function* for $u \geq 0$, $a > -1$, see Abramowitz & Stegun (1972) and Johnson, Kotz & Balakrishnan (1995).

### Non-Central Chi-Square Distribution

For a *non-central chi-square distributed* random variable $X \sim \chi^2(n, \ell)$ with $n \geq 0$ degrees of freedom and non-centrality parameter $\ell > 0$ its distribution function has the form

$$F_X(x) = \chi^2(x; n, \ell) = \sum_{k=0}^{\infty} \frac{\exp\left\{-\frac{\ell}{2}\right\} \left(\frac{\ell}{2}\right)^k}{k\,!} \left(1 - \frac{\Gamma\left(\frac{x}{2}; \frac{n+2k}{2}\right)}{\Gamma\left(\frac{n+2k}{2}\right)}\right) \tag{1.2.13}$$

for $x \geq 0$. In some sense, the non-central chi-square distribution is a weighted sum of central chi-square distributions with Poisson probabilities as weights. The corresponding density function is given as

$$f_X(x) = \frac{1}{2} \left(\frac{x}{\ell}\right)^{\frac{n}{4} - \frac{1}{2}} \exp\left\{-\frac{\ell + x}{2}\right\} I_{\frac{n}{2}-1}\left(\sqrt{\ell\, x}\right), \tag{1.2.14}$$

for $x > 0$. Here $I_\nu(\cdot)$ is the *modified Bessel function of the first kind* with index $\nu$, which is of the form

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^j}{j\,!\,\Gamma(j + \nu + 1)}. \tag{1.2.15}$$

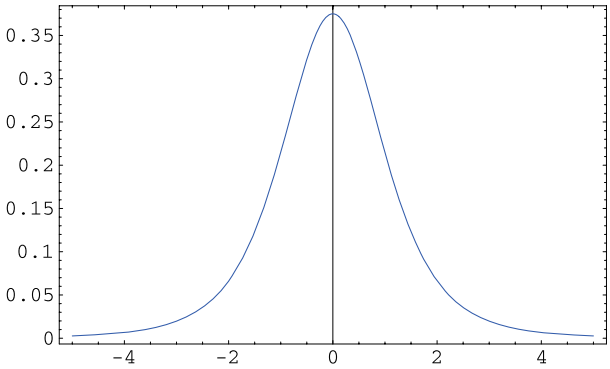


**Fig. 1.2.6.** Student $t$ density for $n = 4$ degrees of freedom

### Central Student $t$ Distribution

Let $Y \sim N(0,1)$ be a standard Gaussian distributed random variable and $Z \sim \chi^2(n)$ be an independent chi-square distributed random variable with $n > 0$ degrees of freedom. Then the random variable

$$X = \frac{Y}{\sqrt{\frac{Z}{n}}} \tag{1.2.16}$$

turns out to be a *central Student t*, or in short a *Student t*, distributed with $n$ degrees of freedom. Its density function is given by

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi\, n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{1.2.17}$$

for $x \in \Re$. We write $X \sim t(n)$ if the random variable $X$ has a Student $t$ distribution with $n$ degrees of freedom. In Fig. 1.2.6 we plot the density of the Student $t$ distribution for $n = 4$ degrees of freedom. As will be shown later, this distribution seems to model log-returns of indices extremely well.

It is interesting to express the Student $t$ distribution function $F_{t(n)}(x)$ in terms of rational and trigonometric functions for small integers $n$, see Shaw (2005). For $n = 1$ one obtains in this way the standard *Cauchy distribution*

$$F_{t(1)}(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x), \tag{1.2.18}$$

where $\tan^{-1}(\cdot)$ expresses the inverse function of $\tan(\cdot)$. Further Student $t$ distribution functions are given by

$$F_{t(2)}(x) = \frac{1}{2} + \frac{x}{2\sqrt{x^2+2}}, \tag{1.2.19}$$

$$F_{t(3)}(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{\sqrt{3}}\right) + \frac{\sqrt{3}\,x}{\pi\,(x^2+3)}, \tag{1.2.20}$$

$$F_{t(4)}(x) = \frac{1}{2} + \frac{x\,(x^2+6)}{2\,(x^2+4)^{\frac{3}{2}}}, \tag{1.2.21}$$

$$F_{t(5)}(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{\sqrt{5}}\right) + \frac{\sqrt{5}\,x\,(3x^2+25)}{3\,\pi\,(x^2+5)^2}, \tag{1.2.22}$$

$$F_{t(6)}(x) = \frac{1}{2} + \frac{x\,(2x^4+30x^2+135)}{4\,(x^2+6)^{\frac{5}{2}}}. \tag{1.2.23}$$

**Symmetric Generalized Hyperbolic Distribution (*)**

Various authors have proposed asset price models with log-returns that relate to the rich class of *symmetric generalized hyperbolic* (SGH) distributions. This class of distributions was extensively examined by Barndorff-Nielsen (1977), see Hurst & Platen (1997) for a study on log-returns. We shall use this class later on to identify the distribution that fits best observed log-returns.

The SGH density function for a random variable $X$ has the form

$$f_X(x) = \frac{1}{\delta\,K_\lambda(\alpha\,\delta)}\sqrt{\frac{\alpha\,\delta}{2\,\pi}}\left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\frac{1}{2}\left(\lambda-\frac{1}{2}\right)} K_{\lambda-\frac{1}{2}}\left(\alpha\,\delta\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) \tag{1.2.24}$$

for $x \in \Re$, where $\lambda \in \Re$ and $\alpha, \delta \geq 0$. We set $\alpha \neq 0$ if $\lambda \geq 0$ and $\delta \neq 0$ if $\lambda \leq 0$. Here $K_\lambda(\cdot)$ is the *modified Bessel function of the third kind* with index $\lambda$, see Abramowitz & Stegun (1972). It can be defined by the integral representation

$$K_\lambda(z) = \frac{1}{2}\int_0^\infty u^{\lambda-1}\exp\left\{-\frac{1}{2}\,z\left(u+\frac{1}{u}\right)\right\}du \tag{1.2.25}$$

for $z \in (0,\infty)$. For $\lambda = \eta + \frac{1}{2}$, where $\eta$ is a nonnegative integer, one has the explicit expression

$$K_{\eta+\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2\,z}}\,\exp\{-z\}\sum_{\ell=0}^{\eta}\frac{(\eta+\ell)!}{(\eta-\ell)!\,\ell!}\,(2\,z)^{-\ell}. \tag{1.2.26}$$

The SGH density is a four parameter density. The parameter $\mu$ is a *location parameter*. The two *shape parameters* for its tails are $\lambda$ and $\bar{\alpha} = \alpha\,\delta$, defined so that they are invariant under scale transformations. The other parameters contribute to the scaling of the density. We define the parameter $c$ as the *unique scale parameter* such that

$$c^2 = \begin{cases} \frac{2\,\lambda}{\alpha^2} & \text{if } \delta = 0 \text{ for } \lambda > 0,\ \bar{\alpha} = 0, \\ \frac{\delta^2\,K_{\lambda+1}(\bar{\alpha})}{\bar{\alpha}\,K_\lambda(\bar{\alpha})} & \text{otherwise.} \end{cases} \tag{1.2.27}$$

It can be shown that as $\lambda \to \pm\infty$ and/or $\bar{\alpha} \to \infty$ the SGH density asymptotically approaches the Gaussian density.

To illustrate certain typical SGH densities we shall describe four special cases of the SGH density in the sequel. These coincide with log-return densities of important asset price models suggested in the literature.

### Student $t$ Density (*)

Praetz (1972) and Blattberg & Gonedes (1974) proposed for log-returns a Student $t$ density with degrees of freedom $n > 0$. This is also the log-return density that arises from observations over long periods of time generated by the *minimal market model* (MMM), which will be derived in Chap. 13, see also Platen (2001). This density is obtained from the above SGH density for the shape parameters $\lambda = -\frac{1}{2}n < 0$ and $\bar{\alpha} = 0$, where $\alpha = 0$ and $\delta = \varepsilon\sqrt{n}$. Using these parameter values the Student $t$ density function for $X$ has then the form

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\varepsilon\sqrt{\pi\, n}\,\Gamma(\frac{n}{2})}\left(1 + \frac{(x-\mu)^2}{\varepsilon^2\, n}\right)^{-\frac{n+1}{2}} \tag{1.2.28}$$

for $x \in \Re$, where $\Gamma(\cdot)$ is again the gamma function, see (1.2.10). Equation (1.2.28) expresses a generalization of the probability density (1.2.17) of a central Student $t$ distributed random variable with $n$ degrees of freedom. The Student $t$ density is a three parameter density. The degree of freedom $n = -2\lambda$ is the shape parameter, with smaller $n$ implying larger tail heaviness for the density. This means that there is a larger probability of extreme values. Furthermore, when the degrees of freedom increase, that is $n \to \infty$, then the Student $t$ density asymptotically approaches the Gaussian density. We plot in Fig. 1.2.7 the central Student $t$ density in logarithmic scale in dependence on the degrees of freedom $n$.

### Normal-Inverse Gaussian Density (*)

Barndorff-Nielsen (1995) proposed log-returns to follow a normal-inverse Gaussian mixture distribution. The corresponding density arises from the SGH density when the shape parameter $\lambda = -\frac{1}{2}$ is chosen. For this parameter value it follows by (1.2.24) that the probability density function of $X$ is then

$$f_X(x) = \frac{\sqrt{\bar{\alpha}}\,\exp\{\bar{\alpha}\}}{c\,\pi}\left(1 + \frac{(x-\mu)^2}{\bar{\alpha}\,c^2}\right)^{-\frac{1}{2}} K_1\left(\bar{\alpha}\sqrt{1 + \frac{(x-\mu)^2}{\bar{\alpha}\,c^2}}\right) \tag{1.2.29}$$

for $x \in \Re$, where $c^2 = \frac{\delta^2}{\bar{\alpha}}$. The normal-inverse Gaussian density is a three parameter density. The parameter $\bar{\alpha}$ is the shape parameter for the tails with smaller $\bar{\alpha}$ implying larger tail heaviness. Furthermore, when $\bar{\alpha} \to \infty$ the normal-inverse Gaussian density asymptotically approaches the Gaussian density. Figure 1.2.8 shows the normal-inverse Gaussian density in logarithmic scale in dependence on the shape parameter $\bar{\alpha}$.

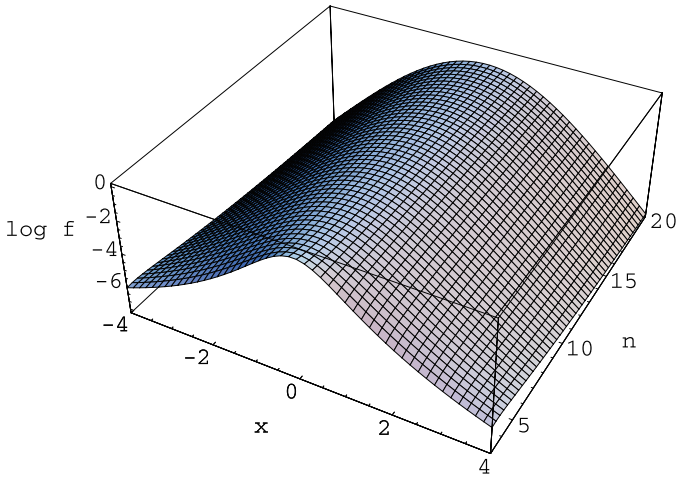


**Fig. 1.2.7.** Student $t$ density under log scale

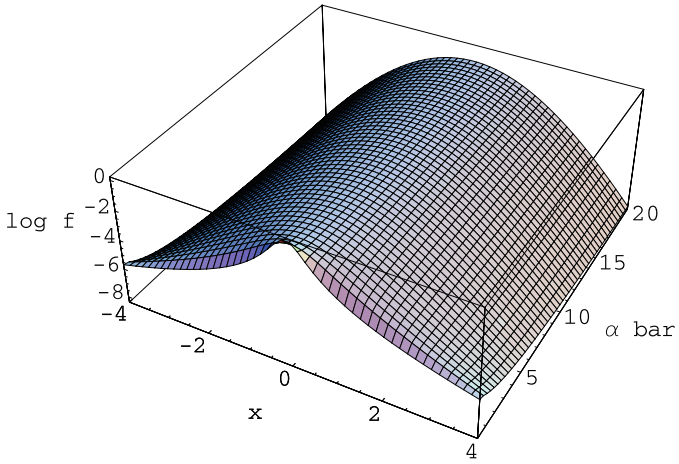


**Fig. 1.2.8.** Normal-inverse Gaussian density under log scale

### Hyperbolic Density (*)

Eberlein & Keller (1995) and Küchler, Neumann, Sørensen & Streller (1999) proposed models, where log-returns appear to be hyperbolicly distributed. This occurs for the choice of the shape parameter $\lambda = 1$ in the SGH density. Using this parameter value the probability density function of $X$ is

$$f_X(x) = \frac{1}{2\,\delta\,K_1(\bar{\alpha})}\,\exp\left\{-\bar{\alpha}\,\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right\} \tag{1.2.30}$$

for $x \in \Re$, where

$$\delta^2 = \frac{c^2\,\bar{\alpha}\,K_1(\bar{\alpha})}{K_2(\bar{\alpha})}.$$

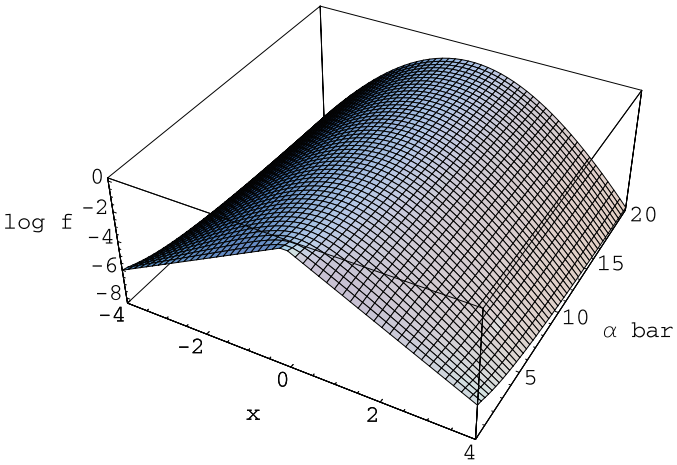


**Fig. 1.2.9.** Hyperbolic density under log scale

The hyperbolic density is a three parameter density. The parameter $\bar{\alpha}$ is the shape parameter with smaller $\bar{\alpha}$ implying larger tail heaviness. Furthermore, when $\bar{\alpha} \to \infty$ the hyperbolic density asymptotically approaches the Gaussian density. In Fig. 1.2.9 we graph the hyperbolic density in a logarithmic scale.

### Variance Gamma Density (*)

Madan & Seneta (1990) proposed that log-returns are distributed with a normal-variance gamma mixture distribution. This case is obtained when the shape parameters are such that $\lambda > 0$ and $\bar{\alpha} = 0$, that is, $\delta = 0$ and $\alpha = \frac{\sqrt{2\,\lambda}}{c}$. With these parameter values the probability density function of $X$ is

$$f_X(x) = \frac{\sqrt{\lambda}}{c\sqrt{\pi}\,\Gamma(\lambda)\,2^{\lambda-1}} \left(\sqrt{2\,\lambda}\,\frac{|x-\mu|}{c}\right)^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\sqrt{2\,\lambda}\,\frac{|x-\mu|}{c}\right) \quad (1.2.31)$$

for $x \in \Re$. The variance gamma density is a three parameter density. The parameter $\lambda$ is the shape parameter with smaller $\lambda$ implying larger tail heaviness. Furthermore, when $\lambda \to \infty$ the variance gamma density asymptotically approaches the Gaussian density. Figure 1.2.10 plots the logarithm of the variance gamma density.

The densities of the Student $t$, normal inverse Gaussian, hyperbolic and variance gamma distribution look very similar when plotted directly. However, their tail densities highlight significant differences. One can see, for instance, that for large $\bar{\alpha}$ and/or large $|\lambda|$ the densities are all close to the Gaussian density. Therefore, we have plotted the corresponding densities in logarithmic scale. In general, it is a challenging problem to identify for log-returns the type of distributions that fits best observed data, as will be discussed later on.

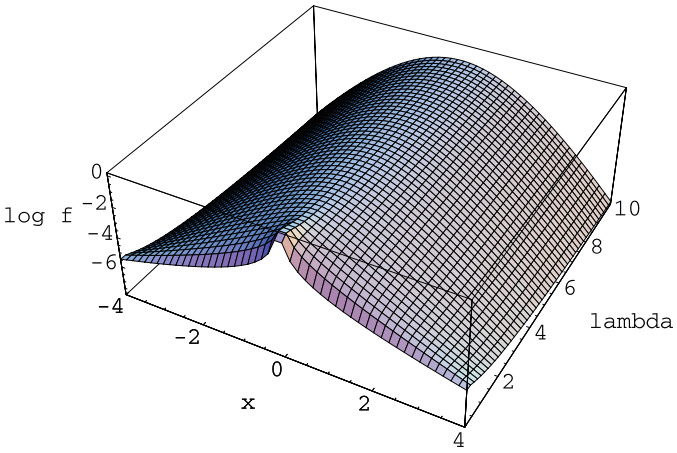


**Fig. 1.2.10.** Variance gamma density under log scale

## 1.3 Moments of Random Variables

Figure 1.1.2 clearly shows that stock log-returns can vary considerably. Therefore, it is important to provide measures for the variability of random variables. Moments, which we shall introduce in the following, provide the most common variability measures.

### Mean

The first of these moments is the arithmetic average that is weighted by the likelihood of occurrence. It is usually called the *mean*, *expectation* or simply *first moment* of the given random variable $X$ and is denoted by $E(X)$. For a *discrete random variable* $X$ the *mean* is defined as

$$\mu_X = E(X) = \sum_{i=0}^{\infty} x_i \, p_i, \tag{1.3.1}$$

where the summation is over all indices of the possible values taken by the random variable. This definition of the mean is readily suggested by the relative frequency interpretation of the probabilities that we discussed in Sect. 1.1.

For example, in the case of a two-point distributed random variable $X$, which takes the value $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$, we have the mean

$$\mu_X = x_1 \, p_1 + x_2 \, (1 - p_1) = x_2 + (x_1 - x_2) \, p_1. \tag{1.3.2}$$

Another example is obtained by computing the mean for the Poisson distribution with the probabilities (1.1.30). Here we have for $X \sim P(\lambda)$ the mean

$$\mu_X = \sum_{i=0}^{\infty} i \, \frac{\lambda^i}{i!} \, \exp\{-\lambda\} = \lambda. \tag{1.3.3}$$

When a *continuous random variable* has a probability density $f_X$, then the corresponding expression for its *mean* is

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \, f_X(x) \, dx. \tag{1.3.4}$$

One may say that the product $f_X(x)\,dx$ approximates the probability that $X$ takes its value in the interval $(x, x + dx)$. Note the similarity between (1.3.4) and (1.3.1).

Since $X$ is a random variable defined on $\Omega$, then (1.3.1) and (1.3.4) can both be equivalently expressed as an integral with respect to the measure $P$. That is, we can write

$$E(X) = \int_{\Omega} X(\omega) \, dP(\omega). \tag{1.3.5}$$

Of course, the above definitions for the mean assume that the summation over the possibly infinite series (1.3.1) and the integral (1.3.4) actually exist, that is, they are finite and well defined for each subset of $\Omega$. This is not always the case, as can be seen from Exercise 1.12 at the end of this chapter. To ensure that the corresponding means are well defined and exist, a necessary and sufficient condition is that $X$ is *integrable*, that is,

$$E(|X|) = \int_{\Omega} |X(\omega)| \, dP(\omega) < \infty. \tag{1.3.6}$$

If $E(|X|) = \infty$, then we say $X$ is not integrable and $E(X)$ does not exist. However, there is no problem in formally defining the mean, even if $E(X) < \infty$ or $E(|X|) = \infty$.

Furthermore, for $p \geq 1$ we say that $X$ is *p-integrable*, if

$$E(|X|^p) = \int_{\Omega} |X(\omega)|^p \, dP(\omega) < \infty. \tag{1.3.7}$$

In particular, if (1.3.7) holds for the case $p = 2$ we call the random variable $X$ *square integrable*.

Let us now compute the means of certain continuous random variables introduced in Sect. 1.2:

The mean of a $U(a, b)$ uniformly distributed random variable $X$ is according to (1.3.4) and (1.2.3) of the form

$$\mu_X = \int_{-\infty}^{\infty} x \, f_X(x) \, dx = \int_a^b \frac{x}{b-a} \, dx = \frac{(a+b)}{2}. \tag{1.3.8}$$

For a random variable $X \sim Exp(\lambda)$ with the exponential distribution (1.2.4) one obtains

$$\mu_X = \int_0^\infty x\,\lambda\,\exp\{-\lambda\,x\}\,dx = \frac{1}{\lambda}. \tag{1.3.9}$$

For a Gaussian distributed random variable $X \sim N(\mu, \sigma^2)$ with density (1.2.6) its mean is given by

$$\mu_X = \int_{-\infty}^\infty \frac{x}{\sqrt{2\,\pi}\,\sigma}\,\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \mu. \tag{1.3.10}$$

One can show that a gamma distributed random variable $X \sim G(p, \alpha)$ with density (1.2.9) has mean

$$\mu_X = \int_0^\infty x\,\frac{\alpha^p}{\Gamma(p)}\,\exp\{-\alpha\,x\}\,x^{p-1}\,dx = \frac{p}{\alpha}. \tag{1.3.11}$$

Finally, we mention that a central Student $t$ distributed random variable $X \sim t(n)$ with $n > 1$ degrees of freedom has mean zero, that is

$$\mu_X = \int_{-\infty}^\infty x\,\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\,\sqrt{\pi\,n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 0. \tag{1.3.12}$$

We could add further examples but it should now be clear how to obtain the mean of a random variable with a given density.

### Variance

A measure for the spread around the mean $\mu_X$ of the values taken by a random variable $X$ is given by its *variance* $\sigma_X^2$, denoted also by $\mathrm{Var}(X)$, which is defined as

$$\sigma_X^2 = \mathrm{Var}(X) = E\left((X - E(X))^2\right) = E((X-\mu_X)^2), \tag{1.3.13}$$

provided that the expression (1.3.13) is finite. Consequently, the variance, also called the *second central moment*, is always nonnegative. The square root of the variance, $\sigma_X = \sqrt{\sigma_X^2}$, is called the *standard deviation* of $X$. Note that if $\mathrm{Var}(X) = 0$, then

$$P(X = E(X)) = 1. \tag{1.3.14}$$

For a two-point distributed random variable $X$, taking values $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$, its variance is given by

$$\sigma_X^2 = p_1\,(1-p_1)\,(x_2-x_1)^2. \tag{1.3.15}$$

For a Poisson distributed random variable $X \sim P(\lambda)$ with intensity $\lambda$ we obtain from (1.1.30) and (1.3.3) the variance

$$\sigma_X^2 = \sum_{i=0}^{\infty} (i-\lambda)^2 \, \frac{\lambda^i}{i\,!} \, \exp\{-\lambda\} = \lambda, \tag{1.3.16}$$

which equals its mean as given by (1.3.3).

It is easy to check that a $U(a,b)$ uniformly distributed random variable $X$ with density (1.2.3) and mean (1.3.8) has variance

$$\sigma_X^2 = \int_a^b \left( x - \frac{a+b}{2} \right)^2 \frac{1}{b-a} \, dx = \frac{(b-a)^2}{12}. \tag{1.3.17}$$

The variance of an exponentially distributed random variable $X \sim Exp(\lambda)$ is, according to (1.2.4) and (1.3.9), given by

$$\sigma_X^2 = \int_0^{\infty} \left( x - \frac{1}{\lambda} \right)^2 \lambda \, \exp\{-\lambda\, x\} \, dx = \lambda^{-2}. \tag{1.3.18}$$

An $N(\mu, \sigma^2)$ distributed Gaussian random variable $X$ with density (1.2.6) can be shown to have a variance that equals $\sigma^2$, that is

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \, \frac{1}{\sqrt{2\pi}\,\sigma} \, \exp\left\{ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right\} dx = \sigma^2. \tag{1.3.19}$$

The variance of a gamma distributed random variable $X \sim G(p, \alpha)$ with density (1.2.9) is of the form

$$\sigma_X^2 = \int_0^{\infty} \left( x - \frac{p}{\alpha} \right)^2 \frac{\alpha^p}{\Gamma(p)} \, \exp\{-\alpha\, x\} \, x^{p-1} \, dx = \frac{p}{\alpha^2}. \tag{1.3.20}$$

Finally, for a central Student $t$ distributed random variable $X \sim t(n)$ we obtain from (1.2.17) and (1.3.12) the variance

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 \, \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\,\sqrt{\pi\, n}} \left( 1 + \frac{x^2}{n} \right)^{-\frac{n+1}{2}} dx = \frac{n}{n-2}, \tag{1.3.21}$$

as long as we have degrees of freedom $n > 2$. A central Student $t$ distribution with $n \leq 2$ degrees of freedom has no finite variance.

### Skewness

Some random variables have probability densities with non-symmetric shapes. One way to measure their asymmetry is to compute the skewness $\beta_X$ of the corresponding density. The *skewness* of a random variable $X$ is measured using the centralized and normalized third moment, that is

$$\beta_X = E\left( \left( \frac{X - \mu_X}{\sigma_X} \right)^3 \right). \tag{1.3.22}$$

For a random variable $X$ we say that its density is called *positively skewed* if $\beta_X > 0$, *negatively skewed* if $\beta_X < 0$ and *symmetric* if $\beta_X = 0$.

For a two-point distributed random variable, taking values $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$, we obtain, using (1.3.2) and (1.3.15), the expression

$$\beta_X = \sqrt{p_1\,(1-p_1)}\,(2p_1 - 1). \tag{1.3.23}$$

Consequently, there is no skewness for $p_1 = 0.5$ in the two-point distribution.

For a Poisson distributed random variable $X$ with probabilities given in (1.1.30) its skewness, using (1.3.3) and (1.3.16), has the form

$$\beta_X = \sum_{i=0}^{\infty} \left( \frac{i-\lambda}{\sqrt{\lambda}} \right)^3 \frac{\lambda^i}{i\,!}\, \exp\{-\lambda\} = \frac{1}{\sqrt{\lambda}}, \tag{1.3.24}$$

which means that the corresponding Poisson distribution is positively skewed.

Furthermore, we note from (1.2.3), (1.3.8) and (1.3.17) that a $U(a,b)$ uniformly distributed random variable $X$ has zero skewness since

$$\beta_X = \int_a^b \left( \frac{x - \frac{a+b}{2}}{\frac{(b-a)}{\sqrt{12}}} \right)^3 \frac{1}{b-a}\, dx = 0. \tag{1.3.25}$$

This confirms the view that the shape of the uniform density in Fig. 1.2.1 is symmetric around its mean. On the other hand, an exponentially distributed random variable $X$ with density (1.2.5) can be shown to have fixed skewness with value

$$\beta_X = \int_0^{\infty} \left( \frac{x - \frac{1}{\lambda}}{\frac{1}{\lambda}} \right)^3 \lambda\, \exp\{-\lambda\, x\}\, dx = 2, \tag{1.3.26}$$

see also Fig. 1.2.2.

One can show for an $N(\mu, \sigma^2)$ distributed Gaussian random variable $X$, using (1.3.10) and (1.3.19), that its density (1.2.6) is symmetric and thus has no skewness. That is, we have

$$\beta_X = \int_{-\infty}^{\infty} \left( \frac{x-\mu}{\sigma} \right)^3 \frac{1}{\sqrt{2\pi}\,\sigma}\, \exp\left\{ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right\} dx = 0. \tag{1.3.27}$$

The Gaussian distribution is obviously not a reasonable distribution if one has to model a strongly skewed random variable.

By (1.3.11) and (1.3.20) a gamma distributed random variable $X \sim G(p, \alpha)$ has positive skewness

$$\beta_X = \int_0^{\infty} \left( \frac{x - \frac{p}{\alpha}}{\frac{\sqrt{p}}{\alpha}} \right)^3 \frac{\alpha^p}{\Gamma(p)}\, \exp\{-\alpha\, x\}\, x^{p-1}\, dx = \frac{2}{\sqrt{p}} \tag{1.3.28}$$

for $p > 0$. This is also indicated by inspection of its density, as displayed in Fig. 1.2.5.

Finally, we mention that the density of a central Student $t$ distributed random variable $X \sim t(n)$ with $n > 3$ degrees of freedom is symmetrically skewed, that is,

$$\beta_X = \int_{-\infty}^{\infty} \left( \frac{x}{\sqrt{\frac{n}{n-2}}} \right)^3 \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi\, n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 0. \tag{1.3.29}$$

This fact is also apparent from the shape of the density shown in Fig. 1.2.6.

**Kurtosis**

Extreme values of returns are very important in a range of financial applications. A large negative log-return of a stock index, which may arise in a major market correction, can considerably change the overall short term performance of a portfolio. The likelihood of such extreme values can be reflected by the *kurtosis* $\kappa_X$, which is the centralized and normalized fourth moment, that is

$$\kappa_X = E\left( \left( \frac{X - \mu_X}{\sigma_X} \right)^4 \right). \tag{1.3.30}$$

For a two-point distributed random variable $X$ taking values $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$ we obtain, using (1.3.2) and (1.3.15),

$$\kappa_X = \frac{\left(\frac{1}{3} - p_1 + p_1^2\right)}{3p_1(1 - p_1)}. \tag{1.3.31}$$

A Poisson distributed random variable $X$ with intensity $\lambda$ yields according to (1.3.3) and (1.3.16) a kurtosis of the form

$$\kappa_X = \sum_{i=0}^{\infty} \left( \frac{i - \lambda}{2} \right)^4 \frac{\lambda^i}{i!} \exp\{-\lambda\} = 3 + \frac{1}{\lambda}. \tag{1.3.32}$$

The kurtosis of a $U(a, b)$ uniformly distributed random variable $X$ by (1.3.8) and (1.3.17) is given by the constant

$$\kappa_X = \int_b^a \left( \frac{x - \frac{a+b}{2}}{\frac{b-a}{\sqrt{12}}} \right)^4 dx = 1.8. \tag{1.3.33}$$

For an exponentially distributed random variable $X$ it can be shown, using (1.3.9) and (1.3.18), that it has a constant kurtosis with

$$\kappa_X = \int_0^{\infty} \left( \frac{x - \frac{1}{\lambda}}{\frac{1}{\lambda}} \right)^4 \lambda \exp\{-\lambda\, x\}\, dx = 9. \tag{1.3.34}$$

An $N(\mu, \sigma^2)$ distributed Gaussian random variable $X$ has by (1.3.10) and (1.3.19) the constant kurtosis

$$\kappa_X = \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^4 \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = 3. \tag{1.3.35}$$

If the kurtosis $\kappa_X$ of a random variable $X$ is greater than 3, then this random variable, its density and also its distribution are called *leptokurtic*.

The kurtosis $\kappa_X$ of a gamma distributed random variable $X \sim G(\alpha, \beta)$ is by (1.3.11) and (1.3.20) of the value

$$\kappa_X = \int_0^{\infty} \left(\frac{x - \frac{p}{\alpha}}{\frac{\sqrt{p}}{\alpha}}\right)^4 \frac{\alpha^p}{\Gamma(p)} \exp\{-\alpha\, x\}\, x^{p-1}\, dx = \frac{3(p+2)}{p}, \tag{1.3.36}$$

which is larger for smaller $p > 0$. Thus a gamma distributed random variable is leptokurtic.

Finally, by (1.3.12) and (1.3.21) we have for a Student $t$ distributed random variable $X \sim t(n)$ the kurtosis

$$\kappa_X = \int_0^{\infty} \left(\frac{x}{\sqrt{\frac{n}{n-2}}}\right)^4 \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi\, n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 3\left(\frac{n-2}{n-4}\right). \tag{1.3.37}$$

This is finite only for $n > 4$ degrees of freedom. This type of random variable is also leptokurtic. The Student $t$ density approaches asymptotically a Gaussian density as $n \to \infty$. This is also reflected in its limiting kurtosis of three as $n \to \infty$.

In Fig. 1.3.1 we plot the kurtosis

$$\kappa_X = \frac{3\, K_\lambda(\bar{\alpha})\, K_{\lambda+2}(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})^2} \tag{1.3.38}$$

for $(\bar{\alpha}, \lambda) \in [0, \infty) \times \Re$ of a symmetric generalized hyperbolic distributed random variable, with density given in (1.2.24), in dependence on the two shape parameters $\lambda$ and $\bar{\alpha}$. Note that the kurtosis is not finite for a Student $t$ distribution with degrees of freedom not greater than four. The hyperbolic distribution yields only a kurtosis of six, which limits its applicability as a log-return distribution because a much higher kurtosis is typically observed for log-returns.

It is an empirical stylized fact, which we shall document later on, that the probability densities of log-returns of stock indices, stock prices and exchange rates have much thicker tails than that of a Gaussian density, which means they are leptokurtic. In some cases the kurtosis of a fitted model may not even be finite. For convenience Table 1.3.1 summarizes the moments for several distributions discussed previously.

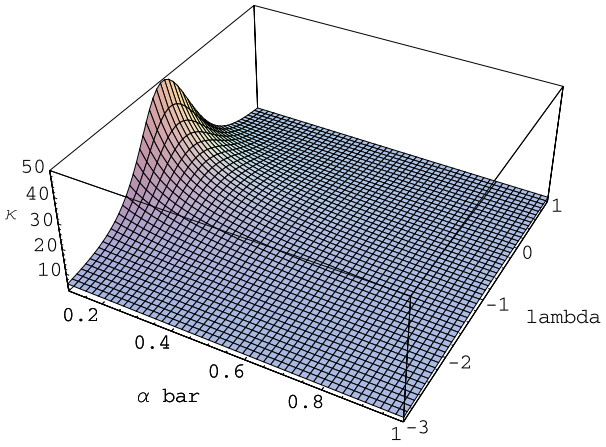


**Fig. 1.3.1.** Kurtosis of SGH random variable in dependence on shape parameters $\lambda$ and $\bar{\alpha}$

**Table 1.3.1.** Moments of some distributions

| $X$ distributed as | | $\mu_X$ | $\sigma_X^2$ | $\beta_X$ | $\kappa_X$ |
|---|---|---|---|---|---|
| Poisson | $P(\lambda)$ | $\lambda$ | $\lambda$ | $\lambda^{\frac{1}{2}}$ | $3+\lambda^{-1}$ |
| Uniform | $U(a,b)$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | 0 | 1.8 |
| Exponential | $Exp(\lambda)$ | $\lambda^{-1}$ | $\lambda^{-2}$ | 2 | 9 |
| Normal | $N(\mu,\sigma^2)$ | $\mu$ | $\sigma^2$ | 0 | 3 |
| Gamma | $G(p,\alpha)$ | $\frac{p}{\alpha}$ | $\frac{p}{\alpha^2}$ | $2\,p^{-\frac{1}{2}}$ | $3\left(\frac{p+2}{p}\right)$ |
| Chi-square | $\chi^2(n)$ | $n$ | $2\,n$ | $\frac{2\sqrt{2}}{n}$ | $3\,\frac{(n+4)}{n}$ |
| Central Student $t$ | $t(n)$ | 0 | $\frac{n}{n-2}$ | 0 | $3\left(\frac{n-2}{n-4}\right)$ |

Finally, let us mention that sometimes the notion of *excess kurtosis* $\underline{\kappa}_X$ of a random variable $X$ is used. This is simply the difference between the kurtosis $\kappa_X$ and the value 3 for the Gaussian kurtosis, that is

$$\underline{\kappa}_X = \kappa_X - 3. \tag{1.3.39}$$

### Higher Order Moments

In general, a new random variable is obtained when we transform or combine random variables by functions or arithmetic operations. For a general transformation of a random variable, however, we need to observe some restrictions on the transforming function $g$. These restrictions follow from measurability constraints to ensure that the resulting variable is still a random variable as defined in Sect. 1.1. More precisely, the function $g$ should be Borel measurable. This is the case when $g$ is, for instance, continuous or piecewise continuous. For more details on these issues the reader is referred to Shiryaev (1984).

When $Y = g(X)$ is a random variable, its expected value, or mean, is

$$E(g(X)) = \sum_{i \in \mathcal{N}} g(x_i)\, p_i \tag{1.3.40}$$

when $X$ is discrete, or

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \tag{1.3.41}$$

when $X$ is continuous with a density function $f_X$. It should be noted that these expectations may be undefined or infinite.

Typical functions of a random variable $X$ are the polynomials $g(x) = x^p$ or $g(x) = (x - \mu_X)^p$ for integers $p \geq 1$. The resulting expected value of $Y = g(X)$ is then called the $p$th *moment*

$$\alpha_p = E(X^p) \tag{1.3.42}$$

or the $p$th *central moment*

$$m_p = E((X - \mu_X)^p), \tag{1.3.43}$$

respectively. For instance, the *variance*

$$\sigma_X^2 = m_2 = \mathrm{Var}(X) = E\left((X - \mu_X)^2\right) \tag{1.3.44}$$

is the *second central moment* of $X$. We have the following important relationships between moments and central moments:

$$\begin{gathered} m_1 = 0, \quad m_2 = \alpha_2 - \alpha_1^2, \quad m_3 = \alpha_3 - 3\,\alpha_1\,\alpha_2 + 2\,\alpha_1^3 \\ m_4 = \alpha_4 - 4\,\alpha_1\,\alpha_3 + 6\,\alpha_1^2\,\alpha_2 - 3\,\alpha_1^4. \end{gathered} \tag{1.3.45}$$

If we use the transform function

$$g(x) = \left(\frac{x - \mu_X}{\sigma_X}\right)^p, \tag{1.3.46}$$

then we obtain the $p$th *normalized central moment*, $p \geq 1$. As previously mentioned, the *skewness* $\beta_X$ is the third normalized central moment and the *kurtosis* $\kappa_X$ is the fourth normalized central moment. Obviously, the first normalized central moment is zero and the second normalized central moment equals one.

Moments provide important information about the given random variable. Note that the higher order moments need not always provide additional information. For example, the Gaussian distribution is completely characterized by its first two moments, its mean $\mu$ and variance $\sigma^2$.

For an $N(\mu, \sigma^2)$ Gaussian distributed random variable $X$ one can show that its $p$th normalized central moment has the form

$$E\left(\left(\frac{X - \mu}{\sigma}\right)^p\right) = \begin{cases} 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2j-1) & \text{for} \quad p = 2j \\ 0 & \text{for} \quad p = 2j - 1, \end{cases} \tag{1.3.47}$$

where $j \in \mathcal{N}$.

The Poisson distribution is already fully characterized by its mean $\lambda$. For a gamma distributed random variable $X \sim \Gamma(p;\alpha)$ the $r$th moment has the form

$$E(X^r) = \frac{\Gamma(p+r)}{\alpha^r\,\Gamma(p)} \tag{1.3.48}$$

for $\alpha > 0,\ p > 0$ and $r > -p$. With (1.3.46) and (1.3.20) we then obtain in this case the $r$th *normalized moment*

$$E\left(\left(\frac{X}{\sigma_X}\right)^r\right) = p^{-\frac{r}{2}}\,\frac{\Gamma(p+r)}{\Gamma(p)} \tag{1.3.49}$$

for $r > -p$, which does not depend on $\alpha$.

### Properties of Moments

General properties of moments can be used to gain an understanding of, and insight into, many of the problems that arise in quantitative finance. Using basic properties of integrals, or equivalently those of infinite series in the discrete case, the first moment, see (1.3.5), inherits the *additivity property*. That is

$$E(a\,X_1 + b\,X_2) = a\,E(X_1) + b\,E(X_2) \tag{1.3.50}$$

for any two random variables $X_1$, $X_2$ and any two real numbers $a$, $b$, provided the expectations are finite.

When $P(X_1 \le X_2) = 1$, then we have for the first moment the *monotonicity property*

$$E(X_1) \le E(X_2). \tag{1.3.51}$$

Moreover, *Jensen's inequality*

$$g(E(X)) \le E(g(X)) \tag{1.3.52}$$

holds for any convex function $g : \Re \to \Re$, which is a function satisfying the relation

$$g(\lambda\,x + (1-\lambda)\,y) \le \lambda\,g(x) + (1-\lambda)\,g(y)$$

for all $x,\ y \in \Re$ and $\lambda \in [0,1]$. In particular, for $g(x) = |x|,\ g(x) = x^2$ and $g(x) = \exp\{x\}$ this implies

$$|E(X)| \le E\,(|X|) \tag{1.3.53}$$

$$|E(X)| \le \sqrt{E(X^2)}. \tag{1.3.54}$$

and

$$\exp\{E(X)\} \le E(\exp\{X\}). \tag{1.3.55}$$

If $E(|X|^s)$ is finite for some $s > 0$, then for all $r \in (0,s]$ and $a \in \Re$ we have the *Lyapunov inequality*

$$(E(|X-a|^r))^{\frac{1}{r}} \le (E(|X-a|^s))^{\frac{1}{s}}\,. \tag{1.3.56}$$

The Lyapunov inequality shows that, if the $s$th moment of a random variable is finite, then any $r$th moment for $r \in (0,s]$ is also finite. For any random variable $X$ we have the *Markov inequality*

$$P(X \ge a) \le \frac{1}{a}\,E(|X|) \tag{1.3.57}$$

for all $a > 0$. From this we can deduce the widely used *Chebyshev inequality*

$$P\left(|X - E(X)| \ge a\right) \le \frac{1}{a^2}\,\mathrm{Var}(X) \tag{1.3.58}$$

for all $a > 0$. For two random variables $X_1$ and $X_2$ the *Cauchy-Schwartz inequality* provides the estimate

$$|E(|X_1\,X_2|)| \le \sqrt{E\left((X_1)^2\right)\,E\left((X_2)^2\right)}. \tag{1.3.59}$$

Further properties of moments can be found at the end of the following section.

### Conditional Expectations

The notion of conditional expectation is central to many of the ideas that arise in probability theory and stochastic calculus. The mean value or expectation $E(X)$ is the coarsest estimate that we have for an integrable random variable $X$, that is, for which $E(|X|) < \infty$, see (1.3.6). If we know that some event $A$ has occurred we may be able to improve on this estimate. For instance, suppose that the event $A = \{\omega \in \Omega :\, X(\omega) \in [a,b]\}$ has occurred. Then in evaluating our estimate of the value of $X$ we need only to consider corresponding values of $X$ and weight them according to their likelihood of occurrence, which is now the conditional probability, see (1.1.13), given this event. The resulting estimate is called the *conditional expectation* of $X$ given event $A$ and is denoted by $E(X\,|\,A)$.

For a discrete random variable $X$ with possible values in a set of real numbers $\mathcal{X} = \{\ldots, x_{-1}, x_0, x_1, \ldots\}$ the conditional probability for the outcome $x_i$ given the event $A = \{\omega \in \Omega :\, X(\omega) \in [a,b]\}$ satisfies

$$P(X = x_i\,|\,A) = \begin{cases} 0 & \text{for} \quad x_i \notin [a,b] \\ \dfrac{p_i}{\sum_{a \le x_j \le b}\, p_j} & \text{for} \quad x_i \in [a,b] \end{cases} \tag{1.3.60}$$

and so the conditional expectation is given by

$$E(X\,|\,A) = \sum_{x_i \in \mathcal{X}} x_i\,P(X = x_i\,|\,A) = \frac{\sum_{a \le x_i \le b}\, x_i\,p_i}{\sum_{a \le x_j \le b}\, p_j}. \tag{1.3.61}$$

More generally, for an integrable random variable $X$ and an event $A \in \mathcal{A}$ the *conditional expectation* $E(X\,|\,A)$ is given by

$$E(X \mid A) = \frac{\int_A X(\omega)\, dP(\omega)}{P(A)}. \tag{1.3.62}$$

For a continuous random variable $X$ with a density function $f_X$ the corresponding *conditional density* is

$$f_X(x \mid A) = \begin{cases} 0 & \text{for} \quad x < a \ \text{ or } \ b < x \\ \frac{f_X(x)}{\int_a^b f_X(s)\, ds} & \text{for} \quad x \in [a, b] \end{cases}$$

with the *conditional expectation*

$$E(X \mid A) = \int_{-\infty}^{\infty} x\, f_X(x \mid A)\, dx = \frac{\int_a^b x\, f_X(x)\, dx}{\int_a^b f_X(x)\, dx}, \tag{1.3.63}$$

which is conditioned on the event $A$ and is thus a number.

More generally let $(\Omega, \mathcal{A}, P)$ be a given probability space with an integrable, see (1.3.6), random variable $X$. We denote by $\mathcal{S}$ a sub-sigma-algebra of $\mathcal{A}$, thus representing a coarser type of information than is given by $\mathcal{A}$. We then define the *conditional expectation* of $X$ with respect to the sub-sigma-algebra $\mathcal{S}$, which we denote by $E(X \mid \mathcal{S})$, as an $\mathcal{S}$-measurable function satisfying

$$\int_S E(X \mid \mathcal{S})(\omega)\, dP(\omega) = \int_S X(\omega)\, dP(\omega), \tag{1.3.64}$$

see Sect. 1.1, for all $S \in \mathcal{S}$. The Radon-Nikodym theorem, see Shiryaev (1984), guarantees the existence and uniqueness of the random variable $E\,(X \mid \mathcal{S})$ a.s. Note that $E\,(X \mid \mathcal{S})$ is a random variable defined on the coarser probability space $(\Omega, \mathcal{S}, P)$ and thus on $(\Omega, \mathcal{A}, P)$. However, $X$ is usually not a random variable on $(\Omega, \mathcal{S}, P)$, but when it is we have

$$E\left(X \,\middle|\, \mathcal{S}\right) = X, \tag{1.3.65}$$

which is the case when $X$ is $\mathcal{S}$-measurable.

Let us consider an example with a random variable $X(\omega) = \omega$ for $\omega \in [0, 1]$ with probability density $f_X(x) = 2x$ for $x \in [0, 1]$. We define the sigma-algebra $\mathcal{S}$ generated by the event $A = \{\omega \in [0, 0.5]\}$. It is then an easy calculation by using (1.3.63) to obtain the conditional expectation

$$E\,(X \mid \mathcal{S})\,(\omega) = \begin{cases} E(X \mid A) = \frac{1}{3} & \text{for} \quad \omega \in [0, 0.5] \\ E(X \mid A^c) = \frac{7}{9} & \text{for} \quad \omega \notin [0, 0.5], \end{cases}$$

where $P(A) = \frac{1}{4}$, $P(A^c) = \frac{3}{4}$ and $E(X) = \frac{2}{3}$.

For nested sigma-algebras $\mathcal{S} \subset \mathcal{T} \subset \mathcal{A}$ and an integrable random variable $X$ we have the *law of iterated conditional expectations*

$$E\left(E\left(X \,\middle|\, \mathcal{T}\right) \,\middle|\, \mathcal{S}\right) = E\left(X \,\middle|\, \mathcal{S}\right) \tag{1.3.66}$$

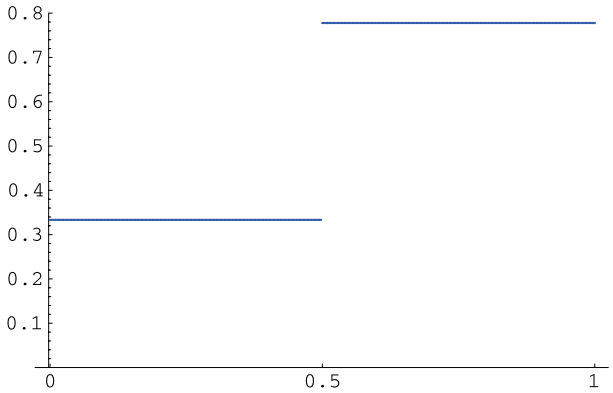


**Fig. 1.3.2.** Conditional expectation

a.s. and when $X$ is independent of the events in $\mathcal{S}$, see (1.1.16), we have

$$E\left(X \,|\, \mathcal{S}\right) = E(X), \tag{1.3.67}$$

a.s. Setting $\mathcal{S} = \{\emptyset, \Omega\}$ it can be seen that

$$E(E\left(X \,|\, \mathcal{S}\right)) = E(X). \tag{1.3.68}$$

This property is easy to check for the above example, where

$$E(X) = \frac{1}{4}\,\frac{1}{3} + \frac{3}{4}\,\frac{7}{9} = \frac{2}{3}.$$

Conditional expectations have similar properties to those of ordinary integrals such as linearity

$$E\left(\alpha\, X + \beta\, Y \,|\, \mathcal{S}\right) = \alpha\, E\left(X \,|\, \mathcal{S}\right) + \beta\, E\left(Y \,|\, \mathcal{S}\right), \tag{1.3.69}$$

where $X$ and $Y$ are integrable random variables and $\alpha, \beta \in \Re$ are deterministic constants. In addition, if $X$ is $\mathcal{S}$-measurable, then

$$E(X\, Y \,|\, \mathcal{S}) = X\, E(Y \,|\, \mathcal{S}). \tag{1.3.70}$$

Furthermore, we have the order preserving property

$$E\left(X \,|\, \mathcal{S}\right) \leq E\left(Y \,|\, \mathcal{S}\right) \tag{1.3.71}$$

if $X \leq Y$ a.s.

The conditional expectation $E(X \,|\, \mathcal{S})$ is in some sense obtained by smoothing $X$ over the events in $\mathcal{S}$. Thus the finer the information set $\mathcal{S}$, the more $E(X \,|\, \mathcal{S})$ resembles the random variable $X$.

**Least-Squares Estimate**

Let $\mathcal{S} \subset \mathcal{A}$ be a given sigma-algebra and $X$ a square integrable random variable on $(\Omega, \mathcal{A}, P)$. We shall show below that

$$E\left(\left(X - E\left(X \,|\, \mathcal{S}\right)\right)^2\right) \leq E\left((X-Y)^2\right) \tag{1.3.72}$$

for all $\mathcal{S}$-measurable, square integrable random variables $Y$. Consequently, $E(X \,|\, \mathcal{S})$ is the *least-squares estimate* or best forecast for $X$ amongst the random variables $Y$ which are $\mathcal{S}$-measurable.

The conditional expectation $E(X \,|\, \mathcal{S})$ can therefore be interpreted as the best estimate, in a least-squares sense, for $X$ under the information given by the events in $\mathcal{S}$. In the case where $\mathcal{S}$ is the sigma-algebra of events generated by a random variable $Y$ we may also write $E(X \,|\, Y)$ for the conditional expectation $E(X \,|\, \mathcal{S})$. This notion of a least-squares estimate, or best forecast, is central to many ideas that arise in stochastic calculus and quantitative finance.

Since the inequality (1.3.72) has fundamental importance we derive it in the following few lines:

Let $Y$ be any square integrable $\mathcal{S}$-measurable random variable and $X$ be a square integrable random variable. Then with $Z = E(X \,|\, \mathcal{S})$ we obtain

$$\begin{aligned} E((X-Y)^2) &= E((X - Z + Z - Y)^2) \\ &= E((X-Z)^2) + 2\,E((X-Z)\,(Z-Y)) + E((Z-Y)^2). \end{aligned} \tag{1.3.73}$$

Using the above described properties of conditional expectations it follows that

$$\begin{aligned} E((X-Z)\,(Z-Y)) &= E\left(E\left((X-Z)\,(Z-Y) \,|\, \mathcal{S}\right)\right) \\ &= E\left(E(X-Z \,|\, \mathcal{S})\,(Z-Y)\right) = E((Z-Z)\,(Z-Y)) = 0. \end{aligned}$$

Consequently, (1.3.73) is minimized by choosing $Y = Z = E(X \,|\, \mathcal{S})$, which proves (1.3.72).

**Moment Generating Functions (*)**

The *cumulants* $k_1, k_2, \ldots$ of a random variable $X$ appear as coefficients of the power series expansion of its *Laplace transform* $\psi_X$, which is also called the *moment generating function*, and has the form

$$\psi_X(\lambda) = E(\exp\{\lambda\, X\}) = 1 + k_1\,\lambda + k_2\,\frac{\lambda^2}{2} + k_3\,\frac{\lambda^3}{3!} + k_4\,\frac{\lambda^4}{4!} + \ldots \tag{1.3.74}$$

for $\lambda \in \Re$ if $\psi_X(\lambda)$ is finite. Note that $\psi_X(\lambda)$ is always finite for $\lambda = 0$ but may be infinite for other values of $\lambda$. The derivatives of the Laplace transform

with respect to $\lambda$ can be used to find the moments. The first four cumulants are related to the first moment and the central moments up to order four, see (1.3.45), by the equations

$$k_1 = \alpha_1 = \mu_X, \quad k_2 = m_2, \quad k_3 = m_3, \quad k_4 = m_4 - 3m_2^2. \tag{1.3.75}$$

The Laplace transform of an $N(\mu, \sigma^2)$ Gaussian distributed random variable $X$ is given by

$$\psi_X(\lambda) = E\left(\exp\{\lambda X\}\right) = \exp\left\{\lambda\,\mu + \frac{\lambda^2\,\sigma^2}{2}\right\} \tag{1.3.76}$$

for $\lambda \in \Re$. This Laplace transform can be used to obtain expectations for asset prices under the standard market model, which is the *lognormal* or *Black-Scholes model*. Under this model returns are normalized increments of exponentials of Gaussian random variables or, equivalently, the log-returns are Gaussian.

### Characteristic Functions (*)

Another important functional of a random variable $X$ is its *characteristic function* $\phi_X$, which is defined as the expectation

$$\phi_X(\theta) = E(\exp\{\imath\,\theta\,X\}), \tag{1.3.77}$$

for all $\theta \in \Re$, where $\imath$ denotes the *imaginary unit*, that is $\imath = \sqrt{-1}$. This function always exists and its absolute value is less than or equal to one, that is

$$|\phi_X(\theta)| \leq 1. \tag{1.3.78}$$

It can be used to identify uniquely the distribution of a given random variable. In this sense the characteristic function encapsulates all of the information content of the distribution of a random variable. For instance, the $p$th moment of $X$, if it exists, can be obtained by the formula

$$\alpha_p = E(X^p) = (-\imath)^p\,\frac{d^p}{(d\theta)^p}\,\phi_X(0). \tag{1.3.79}$$

The mean, variance, skewness and kurtosis can then be derived from these moments according to (1.3.45). For example, the characteristic function of the Poisson distribution with intensity $\lambda$ is from (1.1.31) given by

$$\begin{aligned}\phi_X(\theta) &= \sum_{n=0}^{\infty} \exp\{\imath\,\theta\,n\}\,\frac{\lambda^n}{n!}\,\exp\{-\lambda\} \\ &= \exp\{-\lambda\,(1 - \exp\{\imath\,\theta\})\}\end{aligned} \tag{1.3.80}$$

for $\theta \in \Re$. By using (1.2.6) the characteristic function of an $N(\mu, \sigma^2)$ Gaussian distributed random variable $X$ takes the form

$$\phi_X(\theta) = \int_{-\infty}^{\infty} \exp\{\imath\,\theta\,x\}\,\frac{1}{\sqrt{2\,\pi}\,\sigma}\,\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$$
$$= \exp\left\{\imath\,\theta\,\mu - \theta^2\,\frac{\sigma^2}{2}\right\} \tag{1.3.81}$$

for $\theta \in \Re$. Note the similarity with the Laplace transform (1.3.76). For a $G(p, \alpha)$ gamma distributed random variable, see (1.2.9), we obtain the expression

$$\phi_X(\theta) = \int_0^{\infty} \exp\{\imath\,\theta\,x\}\,\frac{\alpha^p}{\Gamma(p)}\,\exp\{-\alpha\,x\}\,x^{p-1}\,dx$$
$$= \left(\frac{\alpha}{\alpha - \imath\,\theta}\right)^p \tag{1.3.82}$$

for $\theta \in \Re$. For $p = 1$ and $\alpha = \lambda$ this is the characteristic function of an exponential distributed random variable $X \sim Exp(\lambda)$, see (1.2.5). Characteristic functions are often used to analyze and characterize properties of random variables. They are closely related to Fourier transforms of the corresponding density function. A characteristic function $\phi_X(\theta)$ uniquely determines the density function $f_X(x)$ of a continuous random variable $X$. Indeed, the corresponding density function can be found by the *inverse Fourier transform*

$$f_X(x) = \frac{1}{2\,\pi}\int_{-\infty}^{\infty} \exp\{-\imath\,x\,\theta\}\,\phi_X(\theta)\,d\theta, \tag{1.3.83}$$

see Lukacs (1960).

We mentioned at the end of Sect. 1.2 that the SGH distribution covers log-return distributions for a number of important asset price models. Using the notation and parametrization given there one obtains for the SGH distribution the characteristic function

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\,\frac{K_\lambda(\sqrt{(\alpha\,\delta)^2 + \delta^2\,\Delta\,\theta^2})\,(\alpha\,\delta)^\lambda}{K_\lambda(\alpha\,\delta)\,((\alpha\,\delta)^2 + \delta^2\,\Delta\,\theta^2)^{\frac{1}{2}\lambda}} \tag{1.3.84}$$

for $\theta \in \Re$. Recall that $K_\lambda$ is the modified Bessel function of the third kind with index $\lambda$.

If one searches in probability or statistics textbooks and encyclopedias, then the characteristic function of the Student $t$ distribution is notably absent or erroneous. However, a simple closed form solution has been found in Hurst (1997) that is given by the formula

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\,\frac{K_{\frac{1}{2}\,n}(\varepsilon\,\sqrt{n\,\Delta}\,|\theta|)\,(\varepsilon\,\sqrt{n\,\Delta}\,|\theta|)^{\frac{1}{2}\,n}}{\Gamma(\frac{1}{2}\,n)\,2^{\frac{1}{2}\,n-1}} \tag{1.3.85}$$

for all degrees of freedom $n > 0$ and $\theta \in \Re$. For the normal-inverse Gaussian distribution the characteristic function is

$$\phi_X(\theta) = \exp\left\{\imath\,\mu\,\Delta\,\theta + \bar{\alpha}\left(1 - \sqrt{1 + \frac{c^2\,\Delta\,\theta^2}{\bar{\alpha}}}\right)\right\} \tag{1.3.86}$$

for $\theta \in \Re$, where the parametrization is again as in (1.2.29). Furthermore, the hyperbolic distribution has the characteristic function

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\,\frac{\bar{\alpha}\,K_1\left(\sqrt{\bar{\alpha}^2 + \delta^2\,\Delta\,\theta^2}\right)}{K_1(\bar{\alpha})\,\sqrt{\bar{\alpha}^2 + \delta^2\,\Delta\,\theta^2}} \tag{1.3.87}$$

for $\theta \in \Re$. Finally, a variance gamma distributed random variable $X$ has the characteristic function

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\left(1 + \frac{c^2\,\Delta\,\theta^2}{2\,\lambda}\right)^{-\lambda} \tag{1.3.88}$$

for $\theta \in \Re$. A convenient proof for the above results can be obtained by interpreting the above distributions as normal mixture distributions. This means that the random variable is assumed to be conditionally Gaussian distributed with independent random variance. For instance, a Student $t$ distribution with $n$ degrees of freedom is obtained when the inverse of the variance is chi-square distributed with $n$ degrees of freedom. If instead the variance is chi-square distributed, then a variance gamma distribution arises.

### Gaussian Shift (*)

In the context of option pricing, see Buchen & Konstandatos (2005), and other applications it can be useful to apply the following basic relation for *shifted Gaussian random variables*. Let $X \sim N(0,1)$ denote a standard Gaussian random variable, $\theta \in \Re$ a real valued constant and $H(\cdot)$ a real valued function of $x \in \Re$ with $|E(H(X+\theta))| < \infty$. Then it can be shown by exploiting the structure of the Gaussian density that the expectation of a shifted standard Gaussian random variable is of the form

$$E(H(X+\theta)) = E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\,X\right\}H(X)\right). \tag{1.3.89}$$

Interestingly, this allows one also to include the case of more general Gaussian random variables $Y = a + bX$ for $a, b \in \Re$ with mean $E(Y) = a$ and variance $\mathrm{Var}(Y) = b^2$, where we derive the following relation from (1.3.89) for a real valued function $G(y) = G(a + b\,x)$

$$E(G(Y+\theta)) = E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\,X\right\}G(Y)\right). \tag{1.3.90}$$

This is an important relation because the function $G$ can be freely chosen. We shall see later on that the Gaussian shift forms, in principle, the basis for the probability measure transformation that is used in standard derivative pricing.

## 1.4 Joint Distributions and Random Vectors

For many practical applications we need to consider several random variables $X_1$, $X_2$, ..., $X_n$. For instance, these may represent the daily log-returns of all stocks in a market. This leads us to the introduction of joint distributions. The random variables may sometimes be interpreted as components of a vector-valued random variable, which is then called a random vector.

### Joint Distributions

As in the case of a single random variable, we can similarly form a distribution function for $n$ random variables $X_1$, $X_2$, ..., $X_n$, which are defined on the same probability space. The distribution function $F_{X_1,X_2,\cdots,X_n} : \Re^n \to [0,1]$ is called the *joint distribution function* and is defined by the relation

$$F_{X_1,X_2,\cdots,X_n}(x_1,x_2,\ldots,x_n) = P\left(X_i \le x_i,\ i \in \{1,2,\ldots,n\}\right). \tag{1.4.1}$$

Its properties can be illustrated by considering the case of two random variables $X_1$ and $X_2$. Then $F_{X_1,X_2}(x_1,x_2)$ satisfies the limit condition

$$\lim_{x_i \to -\infty} F_{X_1,X_2}(x_1,x_2) = 0 \tag{1.4.2}$$

for $i = 1$ and fixed $x_2 \in \Re$ or $i = 2$ and fixed $x_1 \in \Re$, and also the limit condition

$$\lim_{x_1,x_2 \to \infty} F_{X_1,X_2}(x_1,x_2) = 1. \tag{1.4.3}$$

Furthermore, $F_{X_1,X_2}(x_1,x_2)$ is nondecreasing and continuous from the right in $x_1$ and $x_2$. Additionally, it can be seen that

$$F_{X_1,X_2}(x_1,x_2) = F_{X_2,X_1}(x_2,x_1) \tag{1.4.4}$$

for $(x_1,x_2) \in \Re^2$. The *marginal distribution* $F_{X_1}$ satisfies

$$F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_{X_1,X_2}(x_1,x_2). \tag{1.4.5}$$

for $x_1 \in \Re$.

For continuous random variables the joint distribution function is often differentiable, except possibly at some isolated or boundary points. For a wide class of continuous random variables there is a density function $f_{X_1,X_2} : \Re^2 \to \Re^+ = [0,\infty)$ given by

$$f_{X_1,X_2}(x_1,x_2) = \frac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1\, \partial x_2}, \tag{1.4.6}$$

satisfying

$$F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1,X_2}(s_1,s_2)\, ds_1\, ds_2. \tag{1.4.7}$$

Here $\frac{\partial}{\partial x_1}$ and $\frac{\partial^2}{\partial x_1\, \partial x_2}$ denote first and second order partial derivatives.

**Correlated Random Variables**

Let us consider two random variables $X_1$ and $X_2$ with means $\mu_{X_1}$ and $\mu_{X_2}$ and variances $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$, respectively. Their *covariance* is then defined as

$$\mathrm{Cov}(X_1, X_2) = E((X_1 - \mu_{X_1})\,(X_2 - \mu_{X_2})). \tag{1.4.8}$$

Obviously, we have for two random variables $X_1$ and $X_2$

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1) \tag{1.4.9}$$

and for any constant $a_i \in \Re$, $i \in \{1, 2, 3, 4\}$,

$$\mathrm{Cov}(a_1\, X_1 + a_2, a_3\, X_2 + a_4) = a_1\, a_3\, \mathrm{Cov}(X_1, X_2). \tag{1.4.10}$$

If $X_1$ and $X_2$ are independent, then

$$\mathrm{Cov}(X_1, X_2) = 0. \tag{1.4.11}$$

If $X_1 = X_2$, then

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Var}(X_1). \tag{1.4.12}$$

We define the *correlation* $\varrho_{X_1,X_2}$ of $X_1$ and $X_2$ in the form

$$\varrho_{X_1,X_2} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}. \tag{1.4.13}$$

By the Cauchy-Schwartz inequality (1.3.59) it follows that

$$-1 \le \varrho_{X_1,X_2} \le 1. \tag{1.4.14}$$

If $X_2 = a_1\, X_1 + a_2$ and $a_1 > 0$, then by (1.4.13) and (1.4.8) we have the correlation

$$\varrho_{X_1,X_2} = 1. \tag{1.4.15}$$

The correlation $\varrho_{X_1,X_2}$ provides a measure of the degrees of linear dependence between $X_1$ and $X_2$ using second moments. If $\varrho_{X_1,X_2} \neq 0$, then we call $X_1$ and $X_2$ *correlated*. Two independent random variables are always uncorrelated. For Gaussian random variables also the converse is true, that is, two uncorrelated Gaussian random variables are independent. Note however, in general, two uncorrelated random variables can be still dependent. This is important for log-returns. These can be highly dependent even if they are uncorrelated. This point is often missed in practice. A simple example is given when $X_1$ is $N(0, 1)$ Gaussian distributed and $X_2 = \frac{1}{\sqrt{2}}\,((X_1)^2 - 1)$. Obviously, by (1.3.47) the correlation is zero. However, both random variables $X_1$ and $X_2$ are strongly dependent.

**Bivariate Gaussian Density**

Let $\boldsymbol{A}^\top$ denote the *transpose* of the vector or matrix $\boldsymbol{A}$. A matrix $\boldsymbol{A}$ is *regular* if it is invertible. This is the case if its *determinant* $\det(\boldsymbol{A})$ is not equal to zero.

An important example of a two-dimensional density function is the *bivariate Gaussian density* given by

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{\det(\boldsymbol{D})}} \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{2} C^{i,j}(x_i-\mu_i)(x_j-\mu_j)\right\} \tag{1.4.16}$$

for $(x_1,x_2)^\top \in \Re^2$, with *mean vector* $\boldsymbol{\mu} = (\mu_1,\,\mu_2)^\top \in \Re^2$, *covariance matrix* $\boldsymbol{D} = [D^{i,j}]_{i,j=1}^2$, with components $D^{i,j} = E((X_i-\mu_i)(X_j-\mu_j))$, $i,j \in \{1,2\}$, which is here a $2\times 2$ regular matrix, and the *inverse* of the *matrix* $\boldsymbol{C} = [C^{i,j}]_{i,j=1}^2$. We say that two random variables $X_1$ and $X_2$ having the density (1.4.16) are *jointly Gaussian* distributed with *mean vector* $\boldsymbol{\mu}$ and *covariance matrix* $\boldsymbol{D}$.

If the random vector $\boldsymbol{Z} = (Z_1, Z_2)^\top$ has independent standard Gaussian components $Z_1$ and $Z_2$, then there exists an upper triangular, invertible $2\times 2$ matrix $\boldsymbol{S}$ such that $\boldsymbol{D} = \boldsymbol{S}^\top\boldsymbol{S}$ and the vector

$$\boldsymbol{X} = (X_1, X_2)^\top = \boldsymbol{S}^\top\boldsymbol{Z} + \boldsymbol{\mu} \tag{1.4.17}$$

is jointly Gaussian with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{D}$. $\boldsymbol{S}$ is sometimes called the *Cholesky decomposition* of the covariance matrix $\boldsymbol{D}$.

As an example let us construct pairs of correlated Gaussian random variables $X_1$, $X_2$ with means $\mu_1 = E(X_1) = 0$, $\mu_2 = E(X_2) = 0$ and variances $E(X_1^2) = 1$, $E(X_2^2) = \frac{1}{3}$ and covariance $E(X_1X_2) = \frac{1}{2}$ out of independent standard Gaussian distributed random variables $Z_1$ and $Z_2 \sim N(0,1)$. Some *Value at Risk* (VaR) evaluations are based on constructions of this type.

We note that for

$$X_1 = S^{1,1}\,Z_1 + S^{2,1}\,Z_2 \quad \text{and} \quad X_2 = S^{1,2}\,Z_1 + S^{2,2}\,Z_2 \tag{1.4.18}$$

with

$$\boldsymbol{S} = \begin{pmatrix} S^{1,1} & S^{1,2} \\ S^{2,1} & S^{2,2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{\sqrt{12}} \end{pmatrix}$$

we have

$$\boldsymbol{D} = \boldsymbol{S}^\top\boldsymbol{S} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \tag{1.4.19}$$

and

$$\boldsymbol{C} = \boldsymbol{D}^{-1} = \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix},$$

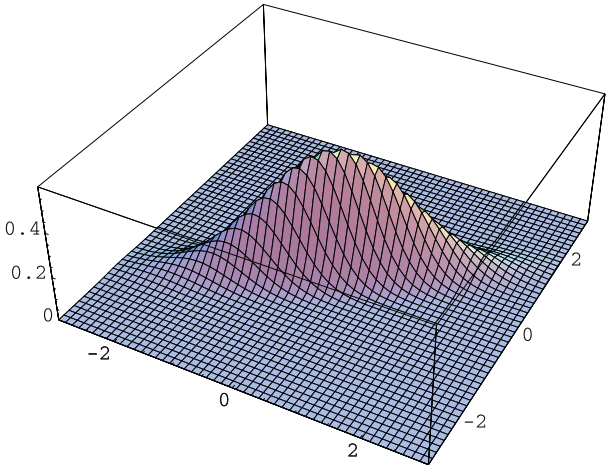


**Fig. 1.4.1.** Bivariate Gaussian density

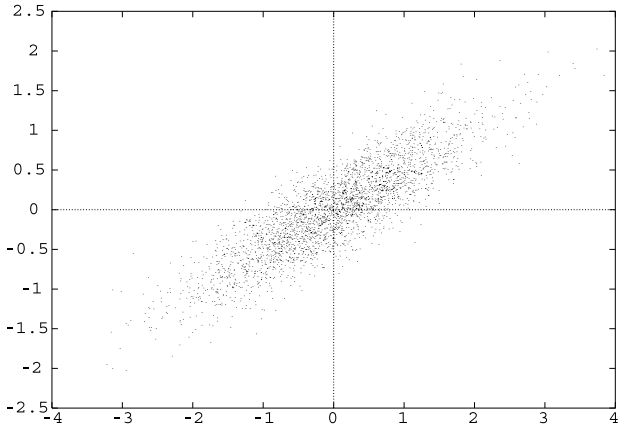


**Fig. 1.4.2.** Points with correlated Gaussian coordinates

where

$$\det(\boldsymbol{C}) = C^{1,1}\,C^{2,2} - C^{1,2}\,C^{2,1} = 12.$$

In Fig. 1.4.1 we show the two-dimensional Gaussian joint density for $X_1$ and $X_2$ with mean vector zero and covariance matrix $\boldsymbol{D}$ given in (1.4.19). We remark that the lines of the Gaussian density that have the same level have an elliptic shape.

Figure 1.4.2 shows 3000 simulated realizations of pairs $(X_1, X_2)$ of these Gaussian random variables using $X_1$ as the $x$-coordinate and $X_2$ as the $y$-coordinate. Note that the points are concentrated mostly in the area where the density given in Fig. 1.4.1 is largest.

For the bivariate Gaussian density with covariance matrix (1.4.19) the correlation coefficient is according to (1.4.18) and (1.4.13) given by

$$\varrho_{X_1,X_2} = \frac{S^{1,1}\, S^{1,2}}{\sigma_{X_1}\, \sigma_{X_2}} = \frac{1}{2}\sqrt{3} \approx 0.866.$$

This means that Fig. 1.4.2 displays a set of 3000 outcomes of correlated Gaussian random variables with the above correlation coefficient.

### Conditional Expectation for the Bivariate Gaussian Case

For given random variables $X_1$ and $X_2$ with bivariate Gaussian distribution one can prove that if $\mathrm{Cov}(X_1, X_2) = 0$, then $X_1$ and $X_2$ are independent. Furthermore, if $\mathrm{Var}(X_2) > 0$, then

$$E(X_1 \,|\, X_2) = E(X_1) + \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}\,(X_2 - E(X_2)) \tag{1.4.20}$$

and

$$E\left((X_1 - E(X_1 \,|\, X_2))^2\right) = \mathrm{Var}(X_1) - \frac{(\mathrm{Cov}(X_1, X_2))^2}{\mathrm{Var}(X_2)}. \tag{1.4.21}$$

Here $E(X_1 \,|\, X_2)$ denotes the conditional expectation of $X_1$ given the information generated by $X_2$.

We emphasize that the above constructions use jointly Gaussian distributed random variables. Now consider two independent $N(0,1)$ standard Gaussian random variables $Y_1$ and $Y_2$. From these we construct $X_1 = |Y_2|\,\mathrm{sgn}(Y_1)$ and $X_2 = Y_2$. Using these definitions it can be shown that $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$ with

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= E(X_1\, Y_2) - E(X_1)\, E(Y_2) = E(Y_2\, |Y_2|\,\mathrm{sgn}(Y_1)) \\ &= E(Y_2\, |Y_2|)\, E(\mathrm{sgn}(Y_1)) = 0,\end{aligned}$$

but $X_1$ and $X_2$ are dependent random variables. As a consequence, $X_1$ and $X_2$ are not jointly Gaussian distributed and

$$\mathrm{Cov}(|X_1|, |X_2|) = E(|Y_2|^2) - (E(|Y_2|))^2 > 0.$$

Note that these types of effects need to be taken into account if one is modeling log-returns of securities.

### Properties of Independent Random Variables

Recall the definition of independent random variables in Sect. 1.1. It can be shown that two random variables $X_1$ and $X_2$ are independent if their joint and marginal distribution functions satisfy the relation

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)\, F_{X_2}(x_2) \tag{1.4.22}$$

for all $x_1, x_2 \in \Re$. This is equivalent to saying that

$$E(g_1(X_1)\, g_2(X_2)) = E(g_1(X_1))\, E(g_2(X_2)) \tag{1.4.23}$$

for all measurable functions $g_1$, $g_2$ for which the above expectations exist. If both $F_{X_1}$ and $F_{X_2}$ have density functions $f_{X_1}$ and $f_{X_2}$, respectively, and if $X_1$ and $X_2$ are independent, then their joint distribution function $F_{X_1,X_2}$ has a density function $f_{X_1,X_2}$ which satisfies the equation

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2). \tag{1.4.24}$$

Moreover, choosing $g_1$ and $g_2$ to equal the identity function in (1.4.23) it can be seen that for two independent random variables $X_1$ and $X_2$ the product $X_1X_2$ has an expectation given by

$$E(X_1X_2) = E(X_1)\, E(X_2), \tag{1.4.25}$$

and the sum $X_1 + X_2$ has a variance satisfying the *additivity property*

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2). \tag{1.4.26}$$

The Gaussian random variables $X_1$ and $X_2$ obtained from (1.4.18) in the corresponding example are by (1.4.19) not independent since $E(X_1X_2) = \frac{1}{2}$ but $E(X_1) = E(X_2) = 0$. They are correlated, as will be shown in the next subsection.

### First and Second Moments of Random Vectors

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ denote a *random vector*. Then the expectation is taken componentwise and we obtain

$$E(\boldsymbol{X}) = (E(X_1), E(X_2), \ldots, E(X_n))^\top. \tag{1.4.27}$$

In the case when $\boldsymbol{B} = [B^{i,j}]_{i,j=1}^{n,m}$ is an $n \times m$ *random matrix*, where $B^{i,j}$ is some random variable we obtain its expectation as the $n \times m$ matrix

$$E(\boldsymbol{B}) = [E(B^{i,j})]_{i,j=1}^{n,m}. \tag{1.4.28}$$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)^\top$ with $n, m \in \mathcal{N}$ denote two random vectors. Their *covariance matrix* $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})$ is defined as

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) &= E\left((\boldsymbol{X} - E(\boldsymbol{X}))\,(\boldsymbol{Y} - E(\boldsymbol{Y}))^\top\right) \\ &= \Big[E\big((X_i - E(X_i))\,(Y_j - E(Y_j))\big)\Big]_{i,j=1}^{n,m} \end{aligned} \tag{1.4.29}$$

The matrix $\mathrm{Cov}(\boldsymbol{X}) = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X})$ is called the *autocovariance matrix* of the vector $\boldsymbol{X}$.

If $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ is an $n$-dimensional vector, $\boldsymbol{A} = [A^{i,j}]_{i,j=1}^{n,m}$ a deterministic $n \times m$ matrix and $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)^\top$ a deterministic $m$-dimensional vector, then for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ it is straightforward to show that

$$E(\boldsymbol{Y}) = E(\boldsymbol{A}\,\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\,E(\boldsymbol{X}) + \boldsymbol{b} \tag{1.4.30}$$

and

$$\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y}) = \boldsymbol{A}\,\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X})\,\boldsymbol{A}^\top. \tag{1.4.31}$$

For example, if $\boldsymbol{X}$ is a vector of $n$ independent random variables with variance $\mathrm{Var}(X_i) = 1,\ i \in \{1, 2, \ldots, n\}$, then

$$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}) = \boldsymbol{I}, \tag{1.4.32}$$

where $\boldsymbol{I}$ is the identity matrix or unit matrix and we have for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ the autocovariance matrix

$$\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y}) = \boldsymbol{A}\,\boldsymbol{A}^\top. \tag{1.4.33}$$

To construct from such a vector $\boldsymbol{X}$ an $n$-dimensional vector $\boldsymbol{Y}$ with given autocovariance matrix $\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y})$ it is sufficient to find an upper triangular $n \times n$-matrix $\boldsymbol{A}$ that satisfies (1.4.33). This matrix is then the Cholesky decomposition of $\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y})$, see (1.4.17).

For any $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ one has the equality

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j), \tag{1.4.34}$$

and if $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i). \tag{1.4.35}$$

### Multivariate Joint Distributions

The properties (1.4.2)–(1.4.5) of joint distribution functions generalize to any number $n \geq 2$ of random variables $X_1$, $X_2$, …, $X_n$. With the notation introduced in (1.4.1) the joint distributions $F_{X_1, X_2, \cdots, X_n}$ satisfy

$$\lim_{x_i \to -\infty} F_{X_1, X_2, \cdots, X_n}(x_1, x_2, \ldots, x_n) = 0 \tag{1.4.36}$$

for $i \in \{1, 2, \ldots, n\}$ and fixed $x_j,\ j \in \{1, 2, \ldots, i-1, i+1, \ldots, n\}$. We also have the limit condition

$$\lim_{x_1, \ldots, x_n \to +\infty} F_{X_1, X_2, \cdots, X_n}(x_1, x_2, \ldots, x_n) = 1. \tag{1.4.37}$$

In addition, $F_{X_1,X_2,\cdots,X_n}$ is non-decreasing and continuous from the right in $x_i$ for $i \in \{1,2,\ldots,n\}$. For any permutation $\{i_1,i_2,\ldots,i_n\}$ of the set $\{1,2,\ldots,n\}$ we have

$$F_{X_{i_1},X_{i_2},\cdots,X_{i_n}}(x_{i_1},x_{i_2},\ldots,x_{i_n}) = F_{X_1,X_2,\cdots,X_n}(x_1,x_2,\ldots,x_n). \quad (1.4.38)$$

Furthermore, if $\{i_1,i_2,\ldots,i_k\}$ is any subset of the set $\{1,2,\ldots,n\}$, then the marginal distribution $F_{X_{i_1},X_{i_2},\cdots,X_{i_k}}$ for $k \in \{1,2,\ldots,n\}$ satisfies

$$F_{X_{i_1},X_{i_2},\cdots,X_{i_k}}(x_{i_1},x_{i_2},\ldots,x_{i_k}) = \lim_{x_i \to +\infty} F_{X_1,X_2,\cdots,X_n}(x_1,x_2,\ldots,x_n), \quad (1.4.39)$$

where this limit has to be taken for all $i \notin \{i_1,i_2,\ldots,i_k\}$.

The properties (1.4.22)–(1.4.26) can also be generalized to $n$ random variables. Thus, the random variables $X_1$, $X_2$, ..., $X_n$ are independent if their joint distribution satisfies the equation

$$F_{X_1,X_2,\cdots,X_k}(x_1,x_2,\ldots,x_k) = F_{X_1}(x_1)\,F_{X_2}(x_2)\cdots F_{X_k}(x_k) \quad (1.4.40)$$

for all $k \in \{1,2,\ldots,n\}$. If in this case each $F_{X_i}$ has a density function $f_{X_i}$, then $F_{X_1,X_2,\cdots,X_n}$ has a joint density function $f_{X_1,X_2,\ldots,X_n}$ that takes the form

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)\,f_{X_2}(x_2)\cdots f_{X_n}(x_n). \quad (1.4.41)$$

In addition, for $n$ independent random variables $X_1,X_2,\cdots,X_n$ the product $g_1(X_1)\ g_2(X_2)\ \cdots g_n(X_n)$ involving measurable functions $g_1,g_2,\ldots,g_n$ has expectation

$$E(g_1(X_1)g_2(X_2)\cdots g_n(X_n)) = E(g_1(X_1))\,E(g_2(X_2))\cdots E(g_n(X_n)), \quad (1.4.42)$$

whereas their sum has variance

$$\mathrm{Var}\left(\sum_{i=1}^{n} g_i(X_i)\right) = \sum_{i=1}^{n} \mathrm{Var}(g_i(X_i)). \quad (1.4.43)$$

**Multivariate Gaussian Density**

Consider a *random vector* $\boldsymbol{X} = (X_1,X_2,\ldots,X_n)^\top$ with *mean vector*

$$\boldsymbol{\mu} = (\mu_1,\mu_2,\ldots,\mu_n)^\top = (E(X_1),E(X_2),\ldots,E(X_n))^\top \quad (1.4.44)$$

and an $n \times n$ *autocovariance matrix* $\boldsymbol{D} = \mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}) = \left[D^{\ell,m}\right]_{\ell,m=1}^{n}$, where

$$D^{\ell,m} = E\left((X_\ell - \mu_\ell)\,(X_m - \mu_m)\right) = E(X_\ell\,X_m) - E(X_\ell)\,E(X_m). \quad (1.4.45)$$

If $\boldsymbol{D}$ is regular, that is $\det(\boldsymbol{D}) \neq 0$, and its density is for $\boldsymbol{x} = (x_1,x_2,\ldots,x_n)^\top \in \Re^n$ given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$
$$= \frac{\exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^\top\,\boldsymbol{D}^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\right\}}{\sqrt{(2\,\pi)^n\,\det(\boldsymbol{D})}}, \tag{1.4.46}$$

then $\boldsymbol{X}$ has an $n$-dimensional Gaussian density. The components of a Gaussian distributed random vector are *independent* if and only if they are pairwise *uncorrelated*. Furthermore, if $\boldsymbol{X}$ is an $n$-dimensional Gaussian random vector, $\boldsymbol{A}$ a deterministic matrix with $m$ rows and $n$ columns and $\boldsymbol{b}$ a deterministic $m$-dimensional vector, then $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ is an $m$-dimensional Gaussian random vector with mean $\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}$ and covariance matrix $\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^\top$.

### Conditional Expectation for Multivariate Gaussian Case (*)

We can generalize the relationships (1.4.20) and (1.4.21) on conditional expectations for bivariate Gaussian random variables to the case where $X_1$ is a scalar random variable and $\boldsymbol{X_2} = (X_2^1, X_2^2, \ldots, X_2^n)^\top$ is an $n$-dimensional random vector such that $X_1$ and the components of $\boldsymbol{X_2}$ are jointly Gaussian distributed. One can prove that if $\mathrm{Cov}(X_1, X_2^i) = 0$ for all $i \in \{1, 2, \ldots, n\}$, then the random variable $X_1$ and the components of the random vector $\boldsymbol{X_2}$ are independent. In the case when the autocovariance matrix of $\boldsymbol{X_2}$ is invertible, that is $\mathrm{Cov}(\boldsymbol{X_2}, \boldsymbol{X_2})^{-1}$ exists, then one has the following conditional expectations

$$E(X_1 \,|\, \boldsymbol{X_2}) = E(X_1) + \mathrm{Cov}(X_1, \boldsymbol{X_2})\,(\mathrm{Cov}(\boldsymbol{X_2}, \boldsymbol{X_2}))^{-1}\,(\boldsymbol{X_2} - E(\boldsymbol{X_2})) \tag{1.4.47}$$

and

$$E\left((X_1 - E(X_1 \,|\, \boldsymbol{X_2}))^2\right) = \mathrm{Var}(X_1) - \mathrm{Cov}(X_1, \boldsymbol{X_2})\,(\mathrm{Cov}(\boldsymbol{X_2}, \boldsymbol{X_2}))^{-1}$$
$$\times\, \mathrm{Cov}(X_1, \boldsymbol{X_2})^\top. \tag{1.4.48}$$

These relationships are quite helpful in statistical analysis and for the pricing of derivatives for multiple securities.

### Multivariate Gaussian Shift (*)

The following relationships can be used for Value at Risk calculations and also in multi-asset option pricing. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \in \Re$ denote an $n$-dimensional vector with correlated $N(0,1)$ distributed components. The correlation matrix equals the covariance matrix $\boldsymbol{D}$ with components

$$D^{\ell,m} = \varrho_{X_\ell, X_m},$$

see (1.4.13). We denote according to (1.4.46) the corresponding joint density by

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = N'_{n,\boldsymbol{D}}(\boldsymbol{x}) = \frac{\exp\{-\frac{1}{2}\,\boldsymbol{x}^\top \boldsymbol{D}^{-1}\,\boldsymbol{x}\}}{\sqrt{(2\pi)^n\,\det(\boldsymbol{D})}} \tag{1.4.49}$$

for $\boldsymbol{x} \in \Re^n$. The associated Gaussian distribution function for $X$ is given by

$$\begin{aligned} F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) &= N_{n,\boldsymbol{D}}(\boldsymbol{x}) \\ &= P(X_i < x_i,\, i \in \{1,2,\ldots,n\}) \\ &= E\left(\prod_{i=1}^{n} \mathbf{1}_{\{X_i<x_i\}}\right) \\ &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} N'_{n,\boldsymbol{D}}(\boldsymbol{y})\, dy_1 \cdots dy_n \end{aligned} \tag{1.4.50}$$

for $\boldsymbol{x} = (x_1,x_2,\ldots,x_n)^\top \in \Re^n$. We say that the $n$-dimensional vector $\boldsymbol{x} \sim N_n(\boldsymbol{0},\boldsymbol{D})$ is Gaussian distributed with mean vector $\boldsymbol{\mu} = (0,\ldots,0)^\top$ and covariance matrix $\boldsymbol{D}$.

Let $\boldsymbol{X} \sim N_n(\boldsymbol{0},\boldsymbol{D})$ and $\boldsymbol{b} = (b^1,b^2,\ldots,b^n)^\top \in \Re^n$ be an $n$-dimensional deterministic vector, then the scalar random variable

$$Z = \boldsymbol{b}^\top \boldsymbol{X}$$

is Gaussian with

$$Z \sim N(0, \boldsymbol{b}^\top \boldsymbol{D}\,\boldsymbol{b}). \tag{1.4.51}$$

More generally, let $\boldsymbol{B} = [B^{i,j}]_{i,j=1}^{m,n}$ be a deterministic $m \times n$ matrix, then we obtain

$$\boldsymbol{Y} = \boldsymbol{B}\,\boldsymbol{X} \sim N_m(\boldsymbol{0}, \boldsymbol{B}\,\boldsymbol{D}\,\boldsymbol{B}^\top), \tag{1.4.52}$$

where the mean vector is a vector of zeros and the covariance matrix $\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^\top$ is an $m \times m$ matrix. Additionally, let us normalize the vector $\boldsymbol{Y}$ by using the diagonal matrix $\boldsymbol{A} = [A^{i,j}]_{i,j=1}^{m}$, where $A^{i,i} = \sqrt{(\boldsymbol{B}\,\boldsymbol{D}\,\boldsymbol{B}^\top)^{i,i}}$ and $A^{i,j} = 0$ for $i \neq j$. We set

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^{-1}\,\boldsymbol{Y} = \boldsymbol{A}^{-1}\,\boldsymbol{B}\,\boldsymbol{X},$$

where $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_m)^\top \sim N_m(\boldsymbol{0}, \boldsymbol{A}^{-1}\,\boldsymbol{B}\,\boldsymbol{D}\,\boldsymbol{B}^\top(\boldsymbol{A}^{-1})^\top)$ turns out to be an $m$-dimensional Gaussian vector with zero mean vector and standard variances for its components. Therefore, it follows for $\tilde{\boldsymbol{y}} = (\tilde{y}_1,\tilde{y}_2,\ldots,\tilde{y}_m)^\top = \boldsymbol{A}^{-1}\boldsymbol{y}$ that

$$\begin{aligned} P\left(\tilde{Y}_i < \tilde{y}_i,\, i \in \{1,2,\ldots,m\}\right) &= P(Y_i < y_i,\, i \in \{1,2,\ldots,m\}) \\ &= N_{m,\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^\top(\boldsymbol{A}^{-1})^\top}(\boldsymbol{A}^{-1}\,\boldsymbol{y}), \end{aligned} \tag{1.4.53}$$

where the multivariate Gaussian distribution function is given in (1.4.50).

From the properties of the probability density $N'_{n,\boldsymbol{D}}(\boldsymbol{x})$ of an $n$-dimensional vector $\boldsymbol{X}$ of standard Gaussian random variables with covariance matrix $\boldsymbol{D}$, see (1.4.49), we have the relation

$$N'_{n,\boldsymbol{D}}(\boldsymbol{x}) = \exp\left\{\boldsymbol{\theta}^\top \boldsymbol{x} - \frac{1}{2}\,\boldsymbol{\theta}^\top \boldsymbol{D}\,\boldsymbol{\theta}\right\} N'_{n,\boldsymbol{D}}(\boldsymbol{x} - \boldsymbol{D}\,\boldsymbol{\theta}) \tag{1.4.54}$$

for any vectors $\boldsymbol{\theta}, \boldsymbol{x} \in \Re^n$. This yields the *multivariate Gaussian shift* property for $\boldsymbol{X} \sim N_n(\boldsymbol{0}, \boldsymbol{D})$, a deterministic vector $\boldsymbol{\theta} = (\theta^1, \theta^2, \ldots, \theta^n)^\top$ and a scalar function $H(\boldsymbol{x})$ of an $n$-dimensional vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ in the form

$$E(H(\boldsymbol{X} + \boldsymbol{\theta})) = E\left(\exp\left\{-\frac{1}{2}\,\boldsymbol{\theta}^\top \boldsymbol{D}\,\boldsymbol{\theta} + \boldsymbol{\theta}^\top \boldsymbol{X}\right\} H(\boldsymbol{X})\right). \tag{1.4.55}$$

This result can be employed in the pricing of derivatives involving several securities, see Buchen (2004) and Buchen & Konstandatos (2005).

### Multivariate Characteristic Functions (*)

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\top$ be a random vector. The *characteristic function* $\phi_{\boldsymbol{X}}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$ is defined for all values of $\boldsymbol{\theta} \in \Re^p$ by

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = E(\exp\{\imath\, \boldsymbol{\theta}^\top \boldsymbol{X}\}), \tag{1.4.56}$$

where $\imath$ is the imaginary unit. Note that

$$|\phi_{\boldsymbol{X}}(\boldsymbol{\theta})| \leq 1 \tag{1.4.57}$$

for all $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top \in \Re^p$. This characteristic function uniquely identifies the distribution of the corresponding random vector. For a continuous $n$-dimensional random vector we have

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{\imath \sum_{k=1}^{p} \theta_k\, x_k\right\} f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1, \ldots, dx_p. \tag{1.4.58}$$

The characteristic function $\phi_{\boldsymbol{X}}(\boldsymbol{\theta})$ of a $p$-dimensional jointly Gaussian distributed random vector $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{D}$ is of the form

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = \exp\left\{\imath\, \boldsymbol{\mu}^\top \boldsymbol{\theta} - \frac{1}{2}\,\boldsymbol{\theta}^\top \boldsymbol{D}\,\boldsymbol{\theta}\right\}. \tag{1.4.59}$$

for all $\boldsymbol{\theta} \in \Re^p$.

Let us give another example using a $p$-dimensional Student $t$ distributed random variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\top$ with $n > 0$ degrees of freedom, zero mean vector $\boldsymbol{\mu} = (0, \ldots, 0)^\top$ and regular covariance matrix $\boldsymbol{D}$. This random variable can be obtained from a multivariate Gaussian vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)^\top$, with mean vector $\boldsymbol{\mu}_{\boldsymbol{Y}} = (0, \ldots, 0)^\top$ and covariance matrix $\boldsymbol{D}$, scaled by the inverse of the square root of an independent scalar $\chi^2(n)$ distributed random variable $Z \in (0, \infty)$ such that

$$\boldsymbol{X} = \frac{\boldsymbol{Y}}{\sqrt{\frac{Z}{n}}}, \tag{1.4.60}$$

see (1.2.16). If $\boldsymbol{Y}$ is a vector of independent standard Gaussian random variables, then $\boldsymbol{X}$ has the characteristic function

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = E\left(e^{\imath\,\boldsymbol{\theta}^\top \boldsymbol{X}}\right) = \frac{K_{\frac{n}{2}}\left(\{n\,\boldsymbol{\theta}^\top\boldsymbol{\theta}\}^{\frac{1}{2}}\right)\,(n\,\boldsymbol{\theta}^\top\boldsymbol{\theta})^{\frac{n}{4}}}{\Gamma(\frac{n}{2})\,2^{\frac{n}{2}-1}} \tag{1.4.61}$$

for $n > 0$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top \in \Re^p$, where $K_\lambda(\cdot)$ is again the modified Bessel function of the third kind with index $\lambda$. Its probability density function is then of the form

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\Gamma(\frac{1}{2}(n+p))}{(\pi\,n)^{\frac{p}{2}}\,\Gamma(\frac{n}{2})}\left(1 + \frac{\boldsymbol{x}^\top\boldsymbol{x}}{n}\right)^{-\frac{1}{2}(n+p)} \tag{1.4.62}$$

for $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^\top \in \Re^p$.

### Further Properties of Moments (*)

When we are considering $n$ different random variables $X_1$, $X_2$, ..., $X_n$, then it is often convenient to use vector notation. For vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ in $\Re^n$ recall that the *inner product* $(\boldsymbol{x}, \boldsymbol{y})$ and the *Euclidean norm* $|\boldsymbol{x}|$ are defined by

$$(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top\,\boldsymbol{y} = \sum_{i=1}^{n} x_i\,y_i \quad \text{and} \quad |\boldsymbol{x}| = \sqrt{\boldsymbol{x}^\top\,\boldsymbol{x}} = \sqrt{\sum_{i=1}^{n}(x_i)^2}, \tag{1.4.63}$$

respectively. Note that for $n = 1$ the Euclidean norm coincides with the absolute value operator.

The following moment inequalities are often useful and follow from more general inequalities for integrals, see Shiryaev (1984). Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ be random vectors, then

$$E(|\boldsymbol{X} + \boldsymbol{Y}|^r) \le c_r\,(E(|\boldsymbol{X}|^r) + E(|\boldsymbol{Y}|^r)) \tag{1.4.64}$$

with $c_r = 1$ for $r \le 1$ and $c_r = 2^{r-1}$ for $r \ge 1$. Furthermore,

$$(E(|\boldsymbol{X} + \boldsymbol{Y}|^r))^{\frac{1}{r}} \le (E(|\boldsymbol{X}|^r))^{\frac{1}{r}} + (E(|\boldsymbol{Y}|^r))^{\frac{1}{r}} \tag{1.4.65}$$

for $r \ge 1$, and

$$E(|(\boldsymbol{X}, \boldsymbol{Y})|) \le (E(|\boldsymbol{X}|^p))^{\frac{1}{p}}\;(E(|\boldsymbol{Y}|^q))^{\frac{1}{q}} \tag{1.4.66}$$

for $p$, $q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$.

## 1.5 Copulas (*)

*Copulas* play an important role in the analysis and modeling of the dependence structures of financial random variables. They are used, for instance, in *Value at Risk* (VaR) and *credit risk* modeling applications. Since they are widely used in different areas in quantitative finance we summarize below a few basic facts on copulas.

### Copula Function (*)

A copula function can be considered as a basic building block for constructing multivariate densities and distributions, see Nelsen (1999). A *copula function* $C : [0,1]^n \to [0,1]$ in $\Re^n$, $n \in \{2,3,\ldots\}$, is a multivariate distribution function with the property that its marginal distributions are standard uniform distributions.

By this definition a copula has the $U(0,1)$ uniform density as the density for all of its marginal distributions, see (1.4.5). The following theorem by Sklar (1959) makes clear that copulas are universal tools for analyzing multivariate distributions.

**Theorem 1.5.1.** (Sklar) *Let $F_{X_1,X_2,\ldots,X_n} : \Re^n \to [0,1]$ be a multivariate $n$-dimensional distribution function with marginal distributions $F_{X_i} : \Re \to [0,1]$, $i \in \{1,2,\ldots,n\}$, then there exists a copula $C : [0,1]^n \to [0,1]$ such that*

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = C\left(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\right) \quad (1.5.1)$$

*for $(x_1,x_2,\ldots,x_n)^\top \in \Re^n$. Moreover, if the marginal distributions have a density, then the copula is unique.*

The proof of this important result exploits the essential fact that one has for $(u_1,u_2,\ldots,u_n)^\top \in [0,1]^n$ the relation

$$C(u_1,u_2,\ldots,u_n) = F_{X_1,X_2,\ldots,X_n}\left(F_{X_1}^{-1}(u_1), F_{X_2}^{-1}(u_2), \ldots, F_{X_n}^{-1}(u_n)\right). \quad (1.5.2)$$

**Corollary 1.5.2.** *For any copula $C : [0,1]^n \to [0,1]$ in $\Re^n$, $n \in \{2,3,\ldots\}$, and distribution functions $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$ the function*

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = C\left(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\right) \quad (1.5.3)$$

*for $(x_1,x_2,\ldots,x_n)^\top \in \Re^n$ defines a multivariate distribution function with marginal distributions $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$.*

This means that every multivariate distribution with continuous marginal distribution function admits a unique copula representation. Furthermore, the above result shows that copulas and marginal distribution functions are the building blocks for general multivariate distributions.

### Gaussian Copula (*)

One of the most common copulas that arise in finance is the *Gaussian copula* $C_{N,\boldsymbol{D}}$, which is defined as

$$C_{N,\boldsymbol{D}}(u_1,u_2,\ldots,u_n) = N_{n,\boldsymbol{D}}\left(N_{X_1}^{-1}(u_1), N_{X_2}^{-1}(u_2), \ldots, N_{X_n}^{-1}(u_n)\right) \quad (1.5.4)$$

for $(u_1,u_2,\ldots,u_n)^\top \in [0,1]^n$. Here $\boldsymbol{D}$ is the regular $n \times n$ covariance matrix of the multivariate Gaussian random variable $\boldsymbol{X}$, see (1.4.45). It is common

in standard VaR calculations to use the Gaussian copula if one has to deduce from the log-returns of the constituents of a portfolio the VaR number of the portfolio.

As an example, let us consider points $(X_1, X_2)^\top$ with Gaussian marginals that have the bivariate Gaussian copula. Here we set

$$X_i = \varrho\, Z_0 + \sqrt{1-\varrho^2}\, Z_i \tag{1.5.5}$$

for $i \in \{1, 2\}$, where $Z_0$, $Z_1$, $Z_2$ are independent standard Gaussian random variables. The parameter $\varrho \in [-1, 1]$ measures the correlation between $X_i$ and $Z_0$ for $i \in \{1, 2\}$. In Fig. 1.4.2 we have plotted 3000 of such points that relate to a bivariate Gaussian distribution with correlation $\varrho \approx 0.866$. The corresponding bivariate Gaussian copula is then

$$C_{N,\boldsymbol{D}}(u_1, u_2) = N_{2,\boldsymbol{D}}\left(N_{X_1}^{-1}(u_1), N_{X_2}^{-1}(u_2)\right) \tag{1.5.6}$$

for $(u_1, u_2)^\top \in [0, 1]^2$. Although the Gaussian copula is widely used in VaR calculations, it usually provides a poor fit to multivariate log-return data.

### Student $t$ Copula (*)

It has been reported in Breymann, Dias & Embrechts (2003) that a good fit for multivariate log-returns of currencies is obtained by the *Student t copula* $C_{t,\boldsymbol{D},\delta}$ with approximately $\delta \approx 4$ degrees of freedom. This copula is defined by the function

$$C_{t,\boldsymbol{D},\delta}(u_1, u_2, \ldots, u_n) = t_{\delta,\boldsymbol{D}}\left(t_{X_1}^{-1}(u_1), t_{X_2}^{-1}(u_2), \ldots, t_{X_n}^{-1}(u_n)\right) \tag{1.5.7}$$

for $(u_1, u_2, \ldots, u_n)^\top \in [0, 1]^n$. Here $t_{\delta,\boldsymbol{D}}$ is the Student $t$ distribution with $\delta > 2$ degrees of freedom and $\boldsymbol{D}$ as the covariance matrix of the components $(X_1, X_2, \ldots, X_n)^\top$, see (1.4.60) and (1.4.62). For currency log-returns Breymann et al. (2003) identified a Student $t$ copula with approximately four degrees of freedom.

The isolines of the bivariate $t$ density have an elliptical shape as is the case for the Gaussian density. This is not surprising due to the representation (1.4.60) of multivariate Student $t$ distributed random variables as multivariate Gaussian random variables with independent inverse chi-square distributed variance.

According to (1.5.7) the bivariate $t$ copula with covariance matrix $\boldsymbol{D}$ and $\delta$ degrees of freedom is obtained from the expression

$$C_{t,\boldsymbol{D},\delta}(u_1, u_2) = t_{\delta,\boldsymbol{D}}\left(t_{X_1}^{-1}(u_1), t_{X_2}^{-1}(u_2)\right) \tag{1.5.8}$$

for $(u_1, u_2)^\top \in [0, 1]^2$.

## 1.6 Exercises for Chapter 1

**1.1.** Show that $\mathrm{Var}(X) = E(X^2) - (E(X))^2$.

**1.2.** Calculate the first and second moments and the variance for a Poisson random variable with intensity $\lambda > 0$.

**1.3.** Calculate the first and second moments and the variance for a $U(a, b)$ uniformly distributed random variable.

**1.4.** Determine for an exponentially distributed random variable with intensity parameter $\lambda > 0$ the first and second moments and the variance.

**1.5.** Calculate the first and second moments and the variance for an $N(0, 1)$ standard Gaussian distributed random variable.

**1.6.** Determine the even moments for a standard Gaussian distributed random variable.

**1.7.** If a random variable $Y$ is $N(\mu, \sigma^2)$ Gaussian distributed show that $X = \frac{Y-\mu}{\sigma}$ is $N(0, 1)$ distributed.

**1.8.** If a random variable $Y$ is $N(0, 1)$ Gaussian distributed what is the distribution of $Y^2$?

**1.9.** Compute the expectation of the exponential $Y = \exp\{X\}$ of a Gaussian $N(\mu, \sigma^2)$ distributed random variable.

**1.10.** (*) Show for a standard Gaussian random variable $X \sim N(0, 1)$, a deterministic constant $\theta \in \Re$ and a real valued function $H(x)$ for $x \in \Re$ with $|E(H(X + \theta))| < \infty$ that

$$E(H(X + \theta)) = E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\, X\right\} H(X)\right).$$

**1.11.** (*) Prove that for a correlated pair of Gaussian random variables the corresponding joint density is, in general, not the product of their marginal densities. When are these random variables independent?

**1.12.** (*) Compute the mean for the Cauchy distribution with density $p(x) = [\pi\,(1 + x^2)]^{-1}$. Is this mean finite?

**1.13.** (*) Compute the conditional expectation $E(X|A)$ for a random variable $X(\omega) = \omega \in [0, 1]$ with density $f_X(x) = x$ with respect to the event $A = \{\omega \in [0, 0.5]\}$.