

Chapter 2

Fundamentals of Probability

The preceding chapter laid out the foundations for understanding how bond prices move in relation to yields. Next we have to characterize movements in bond yields or, more generally, any relevant risk factor in financial markets.

This is done with the tools of probability, a mathematical abstraction that describes the distribution of risk factors. Each risk factor is viewed as a random variable whose properties are described by a probability distribution function. These distributions can be processed with the price-yield relationship to create a distribution of the profit-and-loss profile for the trading portfolio.

This chapter reviews the fundamental tools of probability theory for risk managers. Section 2.1 lays out the foundations, characterizing random variables by their probability density and distribution functions. These functions can be described by their principal moments, mean, variance, skewness, and kurtosis. Distributions with multiple variables are described in Section 2.2. Section 2.3 then turns to functions of random variables. Section 2.4 presents some examples of important distribution functions for risk management, including the uniform, normal, lognormal, Student's, and binomial. Finally, Section 2.5 discusses limit distributions, which can be used to characterize the average and tails of independent random variables.

2.1 Characterizing Random Variables

The classical approach to probability is based on the concept of the **random variable** (rv). This can be viewed as the outcome from throwing a die, for example. Each realization is generated from a fixed process. If the die is perfectly symmetric, we could say that the probability of observing a face with a six in one throw is $p = 1/6$.

Although the event itself is random, we can still make a number of useful statements from a fixed data-generating process.

The same approach can be taken to financial markets, where stock prices, exchange rates, yields, and commodity prices can be viewed as random variables. The assumption of a fixed data-generating process for these variables, however, is more tenuous than for the preceding experiment.

2.1.1 Univariate Distribution Functions

A random variable X is characterized by a **distribution function**,

$$F(x) = P(X \leq x) \quad (2.1)$$

which is the probability that the realization of the random variable X ends up less than or equal to the given number x . This is also called a **cumulative distribution function**.

When the variable X takes discrete values, this distribution is obtained by summing the step values less than or equal to x . That is,

$$F(x) = \sum_{x_j \leq x} f(x_j) \quad (2.2)$$

where the function $f(x)$ is called the **frequency function** or the **probability density function** (p.d.f.). This is the probability of observing x .

When the variable is continuous, the distribution is given by

$$F(x) = \int_{-\infty}^x f(u) du \quad (2.3)$$

The density can be obtained from the distribution using

$$f(x) = \frac{dF(x)}{dx} \quad (2.4)$$

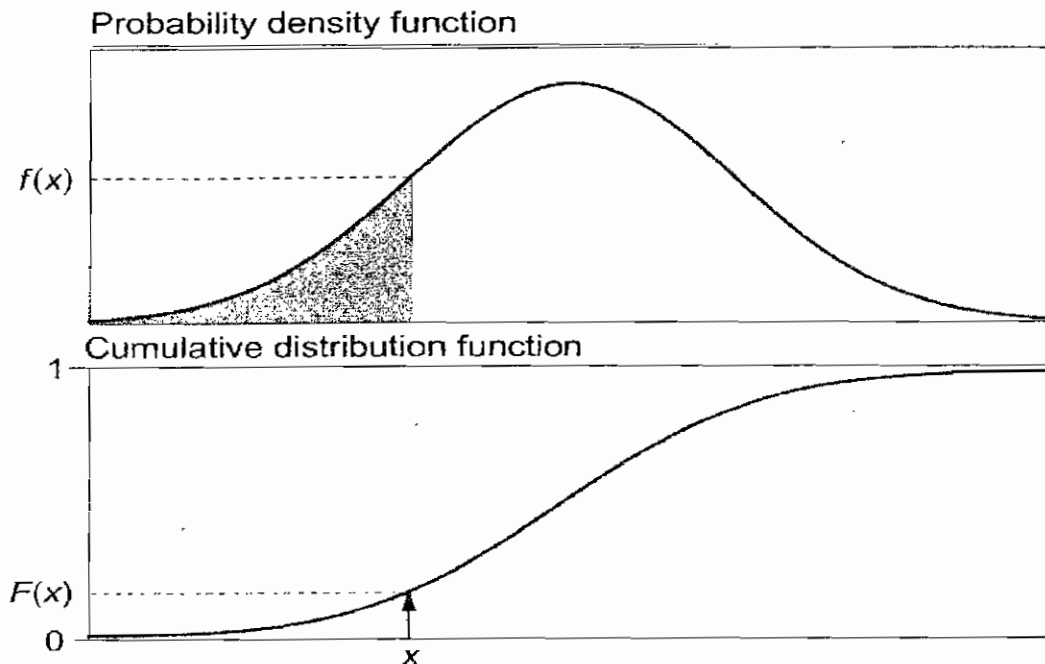
Often, the random variable will be described interchangeably by its distribution or its density.

These functions have notable properties. The density $f(u)$ must be positive for all u . As x tends to infinity, the distribution tends to unity because it represents the total probability of any draw for x :

$$\int_{-\infty}^{\infty} f(u) du = 1 \quad (2.5)$$

Figure 2-1 gives an example of a density function $f(x)$ on the top panel, and of a cumulative distribution function $F(x)$ on the bottom panel. $F(x)$ measures the area under the $f(x)$ curve to the left of x , which is represented by the shaded area. Here, this area is 0.24. For small values of x , $F(x)$ is close to zero. Conversely, for large values of x , $F(x)$ is close to unity.

FIGURE 2-1 Density and Distribution Functions



Example: Density functions

A gambler wants to characterize the probability density function of the outcomes from a pair of dice. Because each die has six faces, there are $6^2 = 36$ possible throw combinations. Out of these, there is one occurrence of an outcome of two (each die showing a one), so the frequency of an outcome of two is 1. We can have two occurrences of a three (a one and a two and vice versa), and so on.

The gambler builds the frequency table (Table 2-1) for each value, from 2 to 12. From this, the gambler can compute the probability of each outcome. For instance, the probability of observing three is equal to 2, the frequency $n(x)$, divided by the total number of outcomes, or 36, which gives 0.0556. We can verify that all the probabilities add up to 1, since all occurrences must be accounted for. From the table, we see that the probability of an outcome of three or less is 8.33%.

TABLE 2-1 Probability Density Function

Outcome x_i	Frequency $n(x)$	Probability $f(x)$	Cumulative Probability $F(x)$
2	1	1/36 0.0278	0.0278
3	2	2/36 0.0556	0.0833
4	3	3/36 0.0833	0.1667
5	4	4/36 0.1111	0.2778
6	5	5/36 0.1389	0.4167
7	6	6/36 0.1667	0.5833
8	5	5/36 0.1389	0.7222
9	4	4/36 0.1111	0.8333
10	3	3/36 0.0833	0.9167
11	2	2/36 0.0556	0.9722
12	1	1/36 0.0278	1.0000
Sum	36	1 1.0000	

2.1.2 Moments

A random variable is characterized by its distribution function. Instead of having to report the whole function, it is convenient to summarize it by a few parameters, or **moments**.

For instance, the expected value for x , or **mean**, is given by the integral

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad (2.6)$$

which measures the *central tendency* or *center of gravity* of the population.

The distribution can also be described by its **quantile**, which is the cutoff point x with an associated probability c :

$$F(x) = \int_{-\infty}^x f(u)du = c \quad (2.7)$$

So there is a probability of c that the random variable will fall *below* x . Because the total probability adds up to 1, there is a probability of $p = 1 - c$ that the random variable will fall *above* x . Define this quantile as $Q(X, c)$. The 50% quantile is known as the **median**.

In fact, value at risk (VAR) can be interpreted as the cutoff point such that a loss will not happen with probability greater than $p = 95\%$, say. If $f(u)$ is the distribution

of profits and losses on the portfolio, VAR is defined from

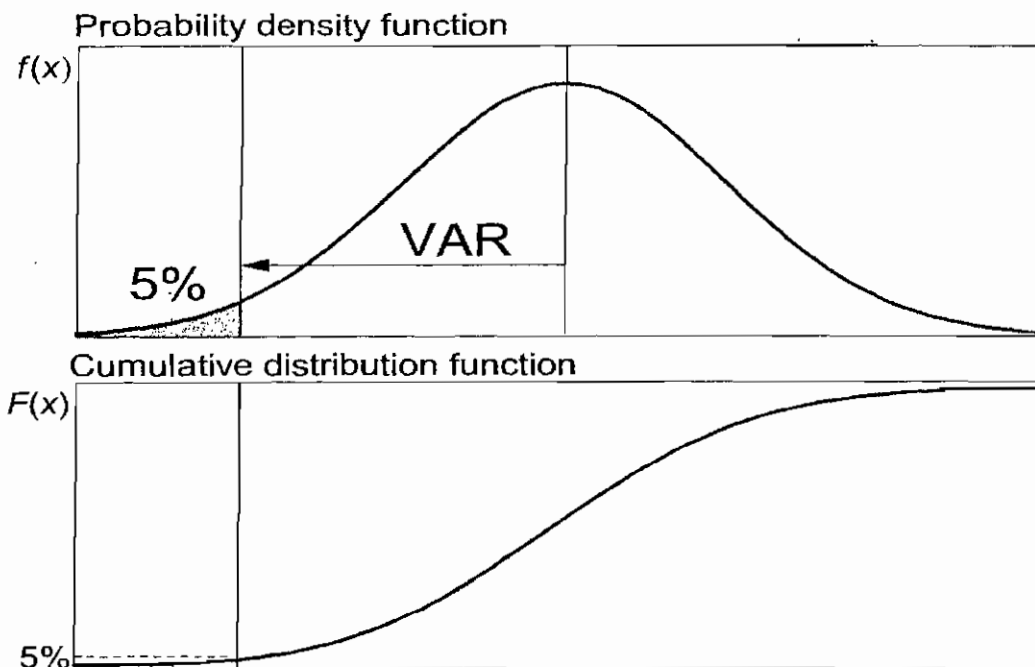
$$F(x) = \int_{-\infty}^x f(u)du = (1 - p) \quad (2.8)$$

where p is the right-tail probability, and c the usual left-tail probability. VAR can then be defined as the deviation between the expected value and the quantile,

$$\text{VAR}(c) = E(X) - Q(X, c) \quad (2.9)$$

Figure 2-2 shows an example with $c = 5\%$.

FIGURE 2-2 VAR as a Quantile



Another useful moment is the squared dispersion around the mean, or **variance**, which is

$$\sigma^2 = V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx \quad (2.10)$$

The **standard deviation** is more convenient to use since it has the same units as the original variable X :

$$\text{SD}(X) = \sigma = \sqrt{V(X)} \quad (2.11)$$

Next, the scaled third moment is the **skewness**, which describes departures from symmetry. It is defined as

$$\gamma = \left(\int_{-\infty}^{+\infty} [x - E(X)]^3 f(x)dx \right) / \sigma^3 \quad (2.12)$$

Negative skewness indicates that the distribution has a long left tail, which indicates a high probability of observing large negative values. If this represents the distribution of profits and losses for a portfolio, this is a dangerous situation. Figure 2-3 displays distributions with various signs for the skewness.

FIGURE 2-3 Effect of Skewness

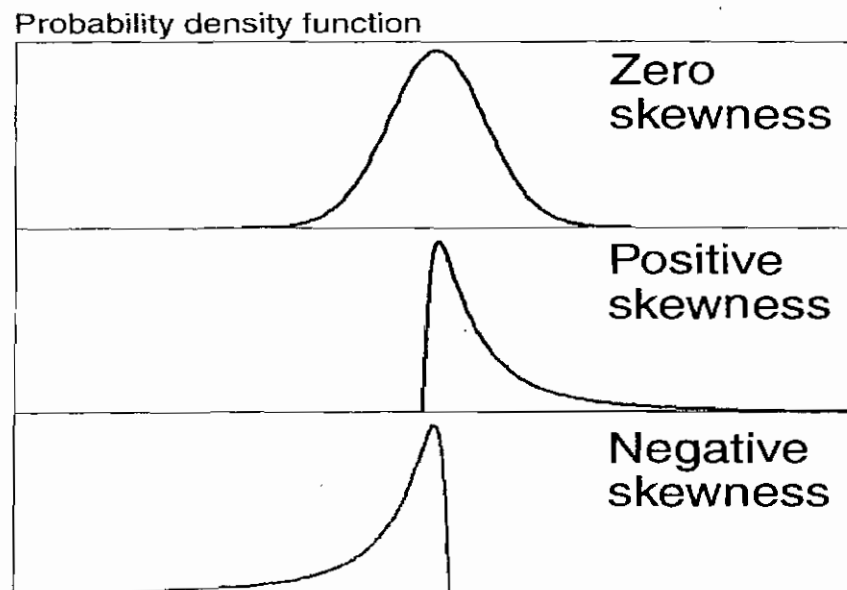
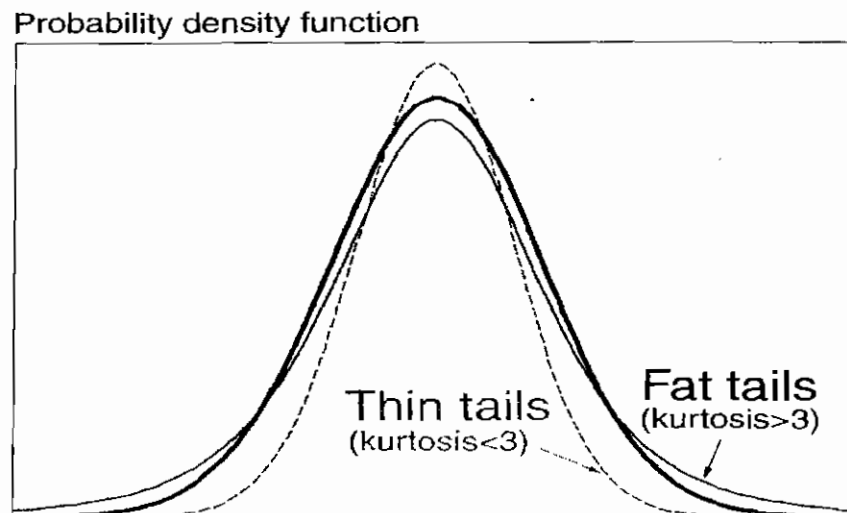


FIGURE 2-4 Effect of Kurtosis



The scaled fourth moment is the **kurtosis**, which describes the degree of “flatness” of a distribution, or the width of its tails. It is defined as

$$\delta = \left(\int_{-\infty}^{+\infty} [x - E(X)]^4 f(x) dx \right) / \sigma^4 \quad (2.13)$$

Because of the fourth power, large observations in the tail will have a large weight and hence create high kurtosis. Such a distribution is called **leptokurtic**, or **fat-tailed**. This parameter is very important for risk measurement. A kurtosis of 3 is considered average. High kurtosis indicates a higher probability of extreme movements. Figure 2-4 displays distributions with various values for the kurtosis.

Example: Computing moments

Our gambler wants to know the expected value of the outcome of throwing two dice. He computes the product of each outcome and the associated probability, as shown in Table 2-2. For instance, the first entry is $xf(x) = 2 \times 0.0278 = 0.0556$. Summing across all events, the mean is $\mu = 7.000$. This is also the median, since the distribution is perfectly symmetric.

TABLE 2-2 Computing Moments of a Distribution

Outcome x_i	Prob. $f(x)$	Mean $xf(x)$	Variance $(x - \mu)^2 f(x)$	Skewness $(x - \mu)^3 f(x)$	Kurtosis $(x - \mu)^4 f(x)$
2	0.0278	0.0556	0.6944	-3.4722	17.3611
3	0.0556	0.1667	0.8889	-3.5556	14.2222
4	0.0833	0.3333	0.7500	-2.2500	6.7500
5	0.1111	0.5556	0.4444	-0.8889	1.7778
6	0.1389	0.8333	0.1389	-0.1389	0.1389
7	0.1667	1.1667	0.0000	0.0000	0.0000
8	0.1389	1.1111	0.1389	0.1389	0.1389
9	0.1111	1.0000	0.4444	0.8889	1.7778
10	0.0833	0.8333	0.7500	2.2500	6.7500
11	0.0556	0.6111	0.8889	3.5556	14.2222
12	0.0278	0.3333	0.6944	3.4722	17.3611
Sum	1.0000	7.0000	$\sigma^2 = 5.8333$	0.0000	80.5000
Denominator				$\sigma^3 = 14.0888$	$\sigma^4 = 34.0278$
		Mean $\mu = 7.00$	StdDev $\sigma = 2.4152$	Skewness $\gamma = 0.0000$	Kurtosis $\delta = 2.3657$

Next, we can use Equation (2.10) to compute the variance. The first term is $(x - \mu)^2 f(x) = (2 - 7)^2 0.0278 = 0.6944$. These terms add up to 5.8333 or, taking the square root, $\sigma = 2.4152$. The skewness terms sum to zero, because for each entry with a positive deviation $(x - \mu)^3$, there is an identical one with a negative sign

and with the same probability. Finally, the kurtosis terms $(x - \mu)^4 f(x)$ sum to 80.5. Dividing by $\sigma^4 = 34.0278$ gives a kurtosis of $\delta = 2.3657$.

2.2 Multivariate Distribution Functions

In practice, portfolio payoffs depend on numerous random variables. To simplify, start with two random variables. This could represent two currencies, or two interest rate factors, or default and credit exposure, to give just a few examples.

We can extend Equation (2.1) to

$$F_{12}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) \quad (2.14)$$

which defines a joint bivariate distribution function. In the continuous case, this is also

$$F_{12}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{12}(u_1, u_2) du_1 du_2 \quad (2.15)$$

where $f(u_1, u_2)$ is now the **joint density**. In general, adding random variables considerably complicates the characterization of the density or distribution functions.

The analysis simplifies considerably if the variables are **independent**. In this case, the joint density separates out into the product of the densities:

$$f_{12}(u_1, u_2) = f_1(u_1) \times f_2(u_2) \quad (2.16)$$

and the integral reduces to

$$F_{12}(x_1, x_2) = F_1(x_1) \times F_2(x_2) \quad (2.17)$$

In other words, the joint probability reduces to the product of the probabilities.

This is very convenient because we only need to know the individual densities to reconstruct the joint density. For example, a credit loss can be viewed as a combination of (1) default, which is a random variable with a value of 1 for default and 0 otherwise, and (2) the exposure, which is a random variable representing the amount at risk—for instance, the positive market value of a swap. If the two variables are independent, we can construct the distribution of the credit loss easily. In the case of the two dice, the events are indeed independent. As a result, the probability of a joint event is simply the product of probabilities. For instance, the probability of throwing two ones is equal to $1/6 \times 1/6 = 1/36$.

It is also useful to characterize the distribution of x_1 abstracting from x_2 . By integrating over all values of x_2 , we obtain the **marginal density**

$$f_1(x_1) = \int_{-\infty}^{\infty} f_{12}(x_1, u_2) du_2 \quad (2.18)$$

and similarly for x_2 . We can then define the **conditional density** as

$$f_{1.2}(x_1 | x_2) = \frac{f_{12}(x_1, x_2)}{f_2(x_2)} \quad (2.19)$$

Here we keep x_2 fixed and divide the joint density by the marginal probability of observing x_2 . This normalization is necessary to ensure that the conditional density is a proper density function that integrates to 1. This relationship is also known as **Bayes' rule**.

When dealing with two random variables, the co-movement can be described by the **covariance**

$$\text{Cov}(X_1, X_2) = \sigma_{12} = \int_1 \int_2 [x_1 - E(X_1)][x_2 - E(X_2)] f_{12}(x_1, x_2) dx_1 dx_2 \quad (2.20)$$

It is often useful to scale the covariance into a unitless number, called the **correlation coefficient**, obtained as

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} \quad (2.21)$$

The correlation coefficient is a measure of linear dependence. One can show that the correlation coefficient always lies in the interval $[-1, +1]$. A correlation of 1 means that the two variables always move in the same direction. A correlation of -1 means that the two variables always move in opposite directions.

If the variables are independent, the joint density separates out and this becomes

$$\text{Cov}(X_1, X_2) = \left\{ \int_1 [x_1 - E(X_1)] f_1(x_1) dx_1 \right\} \left\{ \int_2 [x_2 - E(X_2)] f_2(x_2) dx_2 \right\} = 0$$

by Equation (2.6), since the average deviation from the mean is zero. In this case, the two variables are said to be **uncorrelated**. Hence independence implies zero correlation (the reverse is not true, however).

Example: Multivariate functions

Consider two variables, such as the Canadian dollar and the euro. Table 2-3a describes the joint density function $f_{12}(x_1, x_2)$, assuming two payoffs only for each variable. Note first that the density indeed sums to $0.30 + 0.20 + 0.15 + 0.35 = 1.00$.

TABLE 2-3a Joint Density Function

x_2	x_1	
	-5	+5
-10	0.30	0.15
+10	0.20	0.35

From this, we can compute the marginal density for each variable, along with its mean and standard deviation. For instance, the marginal probability of $x_1 = -5$ is given by $f_1(x_1) = f_{12}(x_1, x_2 = -10) + f_{12}(x_1, x_2 = +10) = 0.30 + 0.20 = 0.50$. The marginal probability of $x_1 = +5$ must be 0.50 as well. Table 2-3b shows that the means and standard deviations are, respectively, $\bar{x}_1 = 0.0$, $\sigma_1 = 5.0$ and $\bar{x}_2 = 1.0$, $\sigma_2 = 9.95$.

TABLE 2-3b Marginal Density Functions

Variable 1				Variable 2			
x_1	Prob. $f_1(x_1)$	Mean $x_1 f_1(x_1)$	Variance $(x_1 - \bar{x}_1)^2 f_1(x_1)$	x_2	Prob. $f_2(x_2)$	Mean $x_2 f_2(x_2)$	Variance $(x_2 - \bar{x}_2)^2 f_2(x_2)$
-5	0.50	-2.5	12.5	-10	0.45	-4.5	54.45
+5	0.50	+2.5	12.5	+10	0.55	+5.5	44.55
Sum	1.00	0.0	25.0	Sum	1.00	1.0	99.0
		$\bar{x}_1 = 0.0$	$\sigma_1 = 5.0$			$\bar{x}_2 = 1.0$	$\sigma_2 = 9.95$

Finally, Table 2-3c details the computation of the covariance, which gives $\text{Cov} = 15.00$. Dividing by the product of the standard deviations, we get $\rho = \text{Cov}/(\sigma_1 \sigma_2) = 15.00/(5.00 \times 9.95) = 0.30$. The positive correlation indicates that when one variable goes up, the other is more likely to go up than down.

TABLE 2-3c Covariance and Correlation

	$(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)f_{12}(x_1, x_2)$	
	$x_1 = -5$	$x_1 = +5$
$x_2 = -10$	$(-5 - 0)(-10 - 1)0.30 = 16.50$	$(+5 - 0)(-10 - 1)0.15 = -8.25$
$x_2 = +10$	$(-5 - 0)(+10 - 1)0.20 = -9.00$	$(+5 - 0)(+10 - 1)0.35 = 15.75$
Sum	Cov = 15.00	

Example 2-1: FRM Exam 1999—Question 21

The covariance between variable A and variable B is 5. The correlation between A and B is 0.5. If the variance of A is 12, what is the variance of B ?

- a) 10.00
- b) 2.89
- c) 8.33
- d) 14.40

Example 2-2: FRM Exam 2000—Question 81

Which one of the following statements about the correlation coefficient is *false*?

- a) It always ranges from -1 to $+1$.
- b) A correlation coefficient of zero means that two random variables are independent.
- c) It is a measure of linear relationship between two random variables.
- d) It can be calculated by scaling the covariance between two random variables.

2.3 Functions of Random Variables

Risk management is about uncovering the distribution of portfolio values. Consider a security that depends on a unique source of risk, such as a bond. The risk manager could model the change in the bond price as a random variable directly. The problem with this choice is that the distribution of the bond price is not stationary, because the price converges to the face value at expiration.

Instead, the practice is to model the change in yields as a random variable because its distribution is better behaved. The next step is to use the relationship between the bond price and the yield to uncover the distribution of the bond price.

This illustrates a general principle of risk management, which is to model the risk factor first, then to derive the distribution of the instrument from information about the function that links the instrument value to the risk factor. This may not be easy to do, unfortunately, if the relationship is highly nonlinear. In what follows, we first focus on the mean and variance of simple transformations of random variables.

2.3.1 Linear Transformation of Random Variables

Consider a transformation that multiplies the original random variable by a constant and adds a fixed amount, $Y = a + bX$. The expectation of Y is

$$E(a + bX) = a + bE(X) \quad (2.22)$$

and its variance is

$$V(a + bX) = b^2V(X) \quad (2.23)$$

Note that adding a constant never affects the variance since the computation involves the *difference* between the variable and its mean. The standard deviation is

$$SD(a + bX) = bSD(X) \quad (2.24)$$

Example: Currency position plus cash

A dollar-based investor has a portfolio consisting of \$1 million in cash plus a position in 1,000 million Japanese yen. The distribution of the dollar/yen exchange rate X has a mean of $E(X) = 1/100 = 0.01$ and volatility of $SD(X) = 0.10/100 = 0.001$.

The portfolio value can be written as $Y = a + bX$, with fixed parameters (in millions) $a = \$1$ and $b = ¥1,000$. Therefore, the portfolio expected value is $E(Y) = \$1 + ¥1,000 \times 1/100 = \11 million, and the standard deviation is $SD(Y) = ¥1,000 \times 0.001 = \1 million.

2.3.2 Sum of Random Variables

Another useful transformation is the summation of two random variables. A portfolio, for instance, could contain one share of Intel plus one share of Microsoft. Each stock price behaves as a random variable.

The expectation of the sum $Y = X_1 + X_2$ can be written as

$$E(X_1 + X_2) = E(X_1) + E(X_2) \quad (2.25)$$

and its variance is

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2) \quad (2.26)$$

When the variables are uncorrelated, the variance of the sum reduces to the sum of variances. Otherwise, we have to account for the cross-product term.

Key concept:

The expectation of a sum is the sum of expectations. The variance of a sum, however, is the sum of variances only if the variables are uncorrelated.

2.3.3 Portfolios of Random Variables

More generally, consider a linear combination of a number of random variables. This could be a portfolio with fixed weights, for which the rate of return is

$$Y = \sum_{i=1}^N w_i X_i \quad (2.27)$$

where N is the number of assets, X_i is the rate of return on asset i , and w_i is its weight.

This can be written in matrix notation, replacing a string of numbers by a single vector:

$$Y = w_1 X_1 + w_2 X_2 + \cdots + w_N X_N = [w_1 w_2 \cdots w_N] \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = w'X \quad (2.28)$$

where w' represents the transposed vector (i.e., horizontal) of weights and X is the vertical vector containing individual asset returns. The appendix for this chapter provides a brief review of matrix multiplication.

The portfolio expected return is now

$$E(Y) = \mu_p = \sum_{i=1}^N w_i \mu_i \quad (2.29)$$

which is a weighted average of the expected returns $\mu_i = E(X_i)$. The variance is

$$V(Y) = \sigma_p^2 = \sum_{i=1}^N w_i^2 \sigma_i^2 + \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_i w_j \sigma_{ij} = \sum_{i=1}^N w_i^2 \sigma_i^2 + 2 \sum_{i=1}^N \sum_{j < i}^N w_i w_j \sigma_{ij} \quad (2.30)$$

Using matrix notation, the variance can be written as

$$\sigma_p^2 = [w_1 \cdots w_N] \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1N} \\ \vdots & & & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \cdots & \sigma_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}$$

Defining Σ as the covariance matrix, the variance of the portfolio rate of return can be written more compactly as

$$\sigma_p^2 = w' \Sigma w \quad (2.31)$$

This is a useful expression for describing the risk of the total portfolio.

Example: Computing the risk of a portfolio

Consider a portfolio invested in Canadian dollars and euros. The joint density function is given in Table 2-3a. Here x_1 describes the payoff on the Canadian dollar, with $\mu_1 = 0.00$, $\sigma_1 = 5.00$, and $\sigma_1^2 = 25$. For the euro, $\mu_2 = 1.00$, $\sigma_2 = 9.95$, and $\sigma_2^2 = 99$. The covariance was computed as $\sigma_{12} = 15.00$, with the correlation $\rho = 0.30$. If we have 60% invested in Canadian dollar and 40% in euros, what is the portfolio volatility?

Following Equation (2.31), we write

$$\sigma_p^2 = [0.60 \ 0.40] \begin{bmatrix} 25 & 15 \\ 15 & 99 \end{bmatrix} \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix} = [0.60 \ 0.40] \begin{bmatrix} 25 \times 0.60 + 15 \times 0.40 \\ 15 \times 0.60 + 99 \times 0.40 \end{bmatrix}$$

$$\sigma_p^2 = [0.60 \ 0.40] \begin{bmatrix} 21.00 \\ 48.60 \end{bmatrix} = 0.60 \times 21.00 + 0.40 \times 48.60 = 32.04$$

Therefore, the portfolio volatility is $\sigma_p = \sqrt{32.04} = 5.66$. Note that this is barely higher than the volatility of the Canadian dollar alone, even though the risk of the euro is much higher. The portfolio risk has been kept low due to a diversification effect, or low correlation between the two assets.

2.3.4 Product of Random Variables

Some risks result from the product of two random variables. A credit loss, for instance, arises from the product of the occurrence of default and the loss given default.

Using Equation (2.20), the expectation of the product $Y = X_1 X_2$ can be written as

$$E(X_1 X_2) = E(X_1)E(X_2) + \text{Cov}(X_1, X_2) \quad (2.32)$$

When the variables are independent, this reduces to the product of the means.

The variance is more complex to evaluate. With independence, it reduces to

$$V(X_1 X_2) = E(X_1)^2 V(X_2) + V(X_1) E(X_2)^2 + V(X_1) V(X_2) \quad (2.33)$$

2.3.5 Distributions of Transformations of Random Variables

The preceding results focus on the mean and variance of simple transformations only. They do not fully describe the distribution of the transformed variable $Y = g(X)$.

This, unfortunately, is usually complicated for all but the simplest transformations $g(\cdot)$ and densities $f(X)$.

Even if there is no closed-form solution for the density, we can describe the cumulative distribution function of Y when $g(X)$ is a one-to-one transformation from X to Y . This implies that the function can be inverted, or that for a given y we can find x such that $x = g^{-1}(y)$. We can then write

$$P[Y \leq y] = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y)) \quad (2.34)$$

where $F(\cdot)$ is the cumulative distribution function of X . Here we have assumed the relationship is positive. Otherwise, the right-hand term is changed to $1 - F_X(g^{-1}(y))$.

This allows us to derive the quantile of, say, the bond price from information about the probability distribution of the yield. Suppose we consider a zero-coupon bond, for which the market value V is

$$V = \frac{100}{(1+r)^T} \quad (2.35)$$

where r is the yield. This equation describes V as a function of r , or $Y = g(X)$. Using $r = 6\%$ and $T = 30$ years, the current price is $V = \$17.41$. The inverse function $X = g^{-1}(Y)$ is

$$r = (100/V)^{1/T} - 1 \quad (2.36)$$

We wish to estimate the probability that the bond price could fall below a cutoff price $V = \$15$. We invert the price-yield function and compute the associated yield level, $g^{-1}(y) = (100/\$15)^{1/30} - 1 = 6.528\%$. Lower prices are associated with higher yield levels. Using Equation (2.34), the probability is given by

$$P[V \leq \$15] = P[r \geq 6.528\%]$$

Assuming the yield change is normal with volatility 0.8%, this gives a probability of 25.5%.¹ Even though we do not know the density of the bond price, this method allows us to trace out its cumulative distribution by changing the cutoff price of \$15. Taking the derivative, we can recover the density function of the bond price. Figure 2-5 shows that this p.d.f. is skewed to the right.

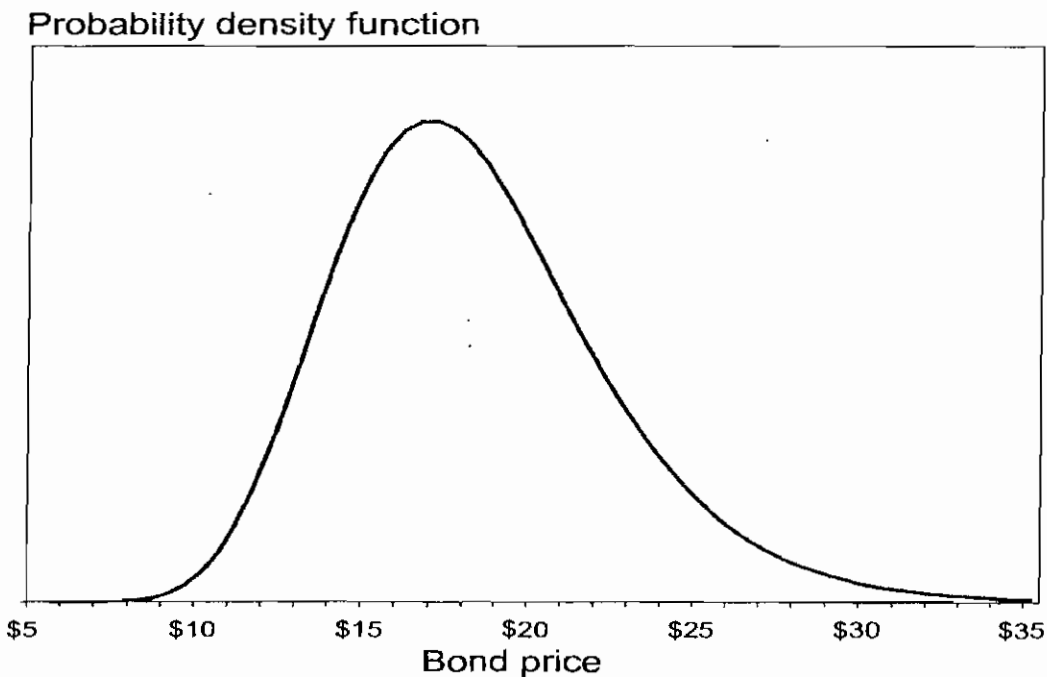
On the extreme right, if the yield falls to zero, the bond price will go to \$100. On the extreme left, if the yield goes to infinity, the bond price will fall to, but not go

¹We shall see later that this is obtained from the standard normal variable $z = (6.528 - 6.000)/0.80 = 0.660$. Using standard normal tables, or the =NORMSDIST(-0.660) Excel function, this gives 25.5%.

below, zero. Relative to the current value of \$17.41, there is a greater likelihood of large movements up than down.

This method, unfortunately, cannot be easily extended. For general density functions and transformations, risk managers turn to numerical methods, especially when the number of random variables is large. This is why credit risk models, for instance, all describe the distribution of credit losses through simulations.

FIGURE 2-5 Density Function for the Bond Price



Example 2-3: FRM Exam 2002—Question 70

Given that x and y are random variables, and a , b , c , and d are constant, which one of the following definitions is *wrong*?

- a) $E(ax + by + c) = aE(x) + bE(y) + c$, if x and y are correlated
- b) $V(ax + by + c) = V(ax + by) + c$, if x and y are correlated
- c) $\text{Cov}(ax + by, cx + dy) = acV(x) + bdV(y) + (ad + bc)\text{Cov}(x, y)$, if x and y are correlated
- d) $V(x - y) = V(x + y) = V(x) + V(y)$, if x and y are uncorrelated

2.4 Important Distribution Functions

2.4.1 Uniform Distribution

The simplest continuous distribution function is the **uniform distribution**. This is defined over a range of values for x , $a \leq x \leq b$. The density function is

$$f(x) = \frac{1}{(b-a)}, \quad a \leq x \leq b \quad (2.37)$$

which is constant and indeed integrates to unity. This distribution puts the same weight on each observation within the allowable range, as shown in Figure 2-6. We denote this distribution as $U(a, b)$.

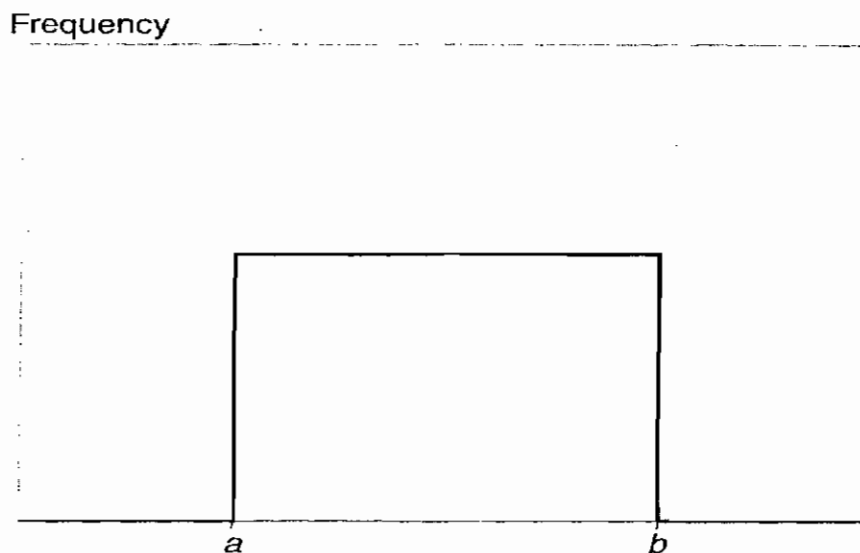
Its mean and variance are given by

$$E(X) = \frac{a+b}{2} \quad (2.38)$$

$$V(X) = \frac{(b-a)^2}{12} \quad (2.39)$$

The uniform distribution $U(0, 1)$ is widely used as a starting distribution for generating random variables in simulations. We assume that the p.d.f. $f(Y)$ and cumulative distribution $F(Y)$ are known. As any cumulative distribution function ranges from zero to unity, we can draw X from $U(0, 1)$ and then compute $y = F^{-1}(x)$. The random variable Y will then have the desired distribution $f(Y)$.

FIGURE 2-6 Uniform Density Function



Realization of the uniform random variable

Financial Risk Manager Handbook, Third Edition

Example 2-4: FRM Exam 2002—Question 119

The random variable X with density function $f(x) = 1/(b - a)$ for $a < x < b$, and 0 otherwise, is said to have a uniform distribution over (a, b) . Calculate its mean.

- a) $(a + b)/2$
- b) $a - b/2$
- c) $a + b/4$
- d) $a - b/4$

2.4.2 Normal Distribution

Perhaps the most important continuous distribution is the **normal distribution**, which adequately represents many random processes. This has a bell-like shape, with more weight in the center and tails tapering off to zero. The daily rate of return in a stock price, for instance, has a distribution similar to the normal p.d.f.

The normal distribution can be characterized by its first two moments only, the mean μ and variance σ^2 . The first parameter represents the location and the second the dispersion. The normal density function has the expression

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right] \quad (2.40)$$

Its mean is $E[X] = \mu$ and its variance $V[X] = \sigma^2$. We denote this distribution as $N(\mu, \sigma^2)$. Because the function can be fully specified by these two parameters, it is called a **parametric function**.

Instead of having to deal with different parameters, it is often more convenient to use a **standard normal variable** such as ϵ , which has been standardized, or normalized, so that $E(\epsilon) = 0$, $V(\epsilon) = \sigma(\epsilon) = 1$. Figure 2-7 plots the **standard normal density**.

First, note that the function is symmetrical around the mean. Its mean of zero is the same as its **mode** (which is the most likely, or highest, point on this curve) and **median** (the value such that the area to the left is a 50% probability). The skewness of a normal distribution is 0, which indicates that it is symmetric around the mean. The kurtosis of a normal distribution is 3. Distributions with fatter tails have a greater kurtosis coefficient.

About 95% of the distribution is contained between the values of $\epsilon_1 = -2$ and $\epsilon_2 = +2$, and 68% of the distribution falls between the values of $\epsilon_1 = -1$ and

$\epsilon_2 = +1$. Table 2-4 gives the values corresponding to right-tail probabilities such that

$$\int_{-a}^{\infty} f(\epsilon)d\epsilon = c \quad (2.41)$$

For instance, the value -1.645 is the quantile that corresponds to a 95% probability.²

FIGURE 2-7 Normal Density Function

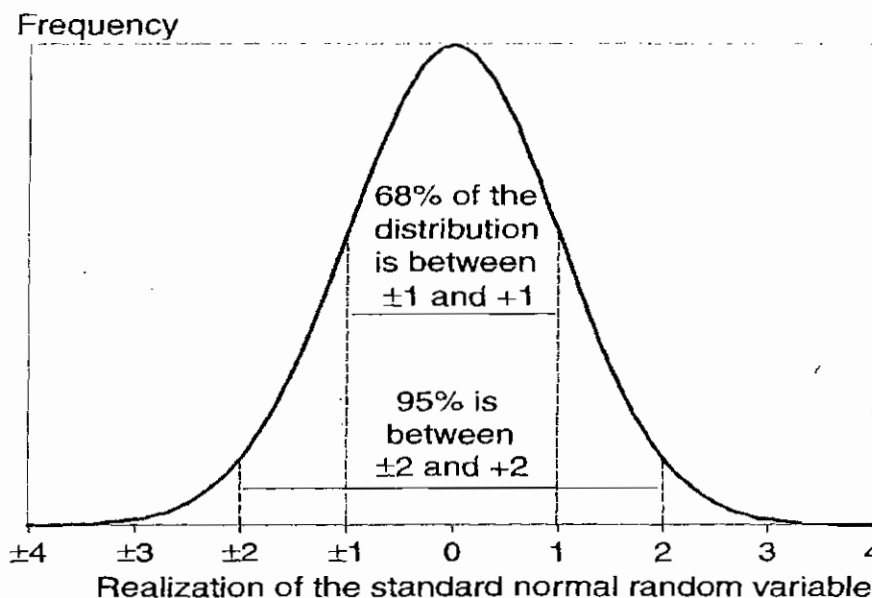


TABLE 2-4 Lower Quantiles of the Standardized Normal Distribution

c	Confidence Level (%)								
	99.99	99.9	99	97.72	97.5	95	90	84.13	50
$-a$	-3.715	-3.090	-2.326	-2.000	-1.960	-1.645	-1.282	-1.000	-0.000

The distribution of any normal variable can be recovered from the standard normal by defining

$$X = \mu + \epsilon\sigma \quad (2.42)$$

Using Equations (2.22) and (2.23), we can show that X indeed has the desired moments, with $E(X) = \mu + E(\epsilon)\sigma = \mu$ and $V(X) = V(\epsilon)\sigma^2 = \sigma^2$.

²More generally, the cumulative distribution can be found from the Excel function =NORMDIST. For example, we can verify that =NORMSDIST(-1.645) yields 0.04999, or a 5% left-tail probability.

Define, for instance, the random variable as the change in the dollar value of a portfolio. The expected value is $E(X) = \mu$. To find the quantile of X at the specified confidence level c , we replace ϵ by $-\alpha$ in Equation (2.42). This gives $Q(X, c) = \mu - \alpha\sigma$. Using Equation (2.9), we can compute VAR as

$$\text{VAR} = E(X) - Q(X, c) = \mu - (\mu - \alpha\sigma) = \alpha\sigma \quad (2.43)$$

For example, a portfolio with a standard deviation of \$10 million would have a VAR, or potential downside loss, of \$16.45 million at the 95% confidence level.

Key concept:

With normal distributions, the VAR of a portfolio is obtained from the product of the portfolio standard deviation and a standard normal deviate factor that reflects the confidence level, for instance, 1.645 at the 95% level.

An important property of the normal distribution is that it is one of the few distributions that is *stable* under addition. In other words, a linear combination of jointly normally distributed random variables has a normal distribution.³ This is extremely useful because we need to know only the mean and variance of the portfolio to reconstruct its whole distribution.

Key concept:

A linear combination of jointly normal variables has a normal distribution.

Example 2-5: FRM Exam 2000—Question 108

The distribution of one-year returns for a portfolio of securities is normally distributed with an expected value of €45 million and a standard deviation of €16 million. What is the probability that the value of the portfolio one year hence will be between €39 million and €43 million?

- a) 8.6%
- b) 9.6%
- c) 10.6%
- d) 11.6%

³Strictly speaking, this is true only under either of the following conditions: (1) The univariate variables are independently distributed, or (2) the variables are multivariate normally distributed (this invariance property also holds for jointly elliptically distributed variables).

Example 2-6: FRM Exam 1999—Question 12

For a standard normal distribution, what is the approximate area under the cumulative distribution function between the values -1 and 1 ?

- a) 50%
- b) 68%
- c) 75%
- d) 95%

Example 2-7: FRM Exam 1999—Question 13

What is the kurtosis of a normal distribution?

- a) 0
- b) Cannot be determined, because it depends on the variance of the particular normal distribution considered
- c) 2
- d) 3

Example 2-8: FRM Exam 1999—Question 16

If a distribution with the same variance as a normal distribution has kurtosis greater than 3, which of the following is *true*?

- a) It has fatter tails than the normal distribution.
- b) It has thinner tails than the normal distribution.
- c) It has the same tail fatness as the normal distribution since the variances are the same.
- d) Cannot be determined from the information provided.

2.4.3 Lognormal Distribution

The normal distribution is a good approximation for many financial variables, such as the rate of return on a stock, $r = (P_1 - P_0)/P_0$, where P_0 and P_1 are the stock prices at time 0 and 1.

Strictly speaking, this is inconsistent with reality since a normal variable has infinite tails on both sides. In theory, r could end up below -1 , which implies $P_1 < 0$. In reality, due to the limited liability of corporations, stock prices cannot turn negative. In many situations, however, this is an excellent approximation. For instance, with short horizons or small price moves, the probability of having a negative price is so small as to be negligible. If this is not the case, we need to resort to distributions that prevent prices from going negative. One such distribution is the lognormal.

A random variable X is said to have a **lognormal distribution** if its logarithm $Y = \ln(X)$ is normally distributed. Define here $X = (P_1/P_0)$. Because the argument

X in the logarithm function must be positive, the price P_1 can never go below zero.

The lognormal density function has the expression

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right], \quad x > 0 \quad (2.44)$$

Note that this is more complex than simply plugging $\ln(x)$ in Equation (2.40), because x also appears in the denominator. Its mean is

$$E[X] = \exp\left[\mu + \frac{1}{2}\sigma^2\right] \quad (2.45)$$

and variance $V[X] = \exp[2\mu + 2\sigma^2] - \exp[2\mu + \sigma^2]$. The parameters were chosen to correspond to those of the normal variable, $E[Y] = E[\ln(X)] = \mu$ and $V[Y] = V[\ln(X)] = \sigma^2$.

Conversely, if we set $E[X] = \exp[r]$, the mean of the associated normal variable is $E[Y] = E[\ln(X)] = (r - \sigma^2/2)$. We will see later that this adjustment is also used in the Black-Scholes option valuation model, where the formula involves a trend in $(r - \sigma^2/2)$ for the log-price ratio.

FIGURE 2-8 Lognormal Density Function

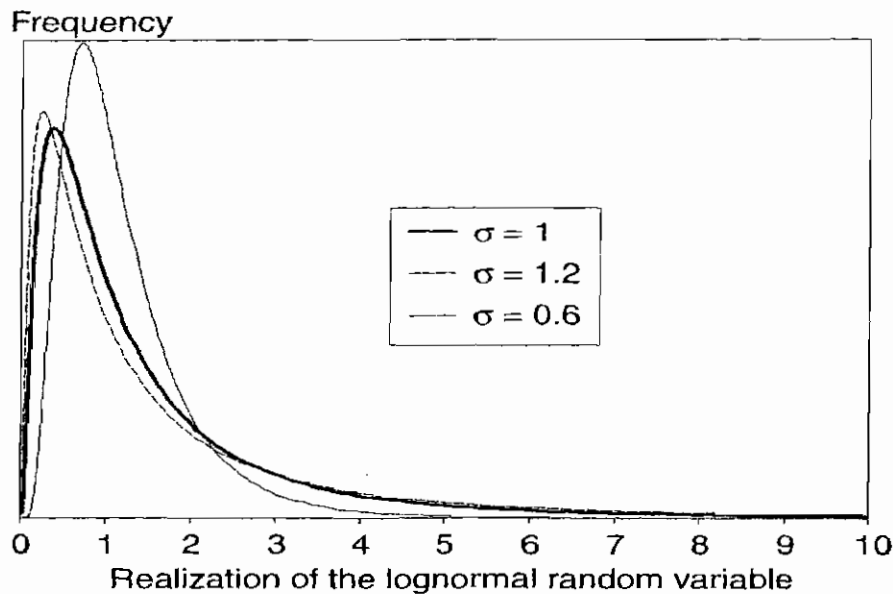


Figure 2-8 depicts the lognormal density function with $\mu = 0$ and $\sigma = 1.0, 1.2, 0.6$. Note that the distribution is skewed to the right. The tail increases for greater values of σ . This explains why as the variance increases, the mean is pulled up in Equation (2.45).

We also note that the distribution of the bond price in our previous example, Equation (2.35), resembles a lognormal distribution. Using continuous compounding instead of annual compounding, the price function is

$$V = 100 \exp(-rT) \quad (2.46)$$

which implies $\ln(V/100) = -rT$. Thus if r is normally distributed, V has a lognormal distribution.

Example 2-9: FRM Exam 2001—Question 72

The lognormal distribution is

- a) Positively skewed
- b) Negatively skewed
- c) Not skewed; that is, its skew equals 2
- d) Not skewed; that is, its skew equals 0

Example 2-10: FRM Exam 1999—Question 5

Which of the following statements best characterizes the relationship between the normal and lognormal distributions?

- a) The lognormal distribution is the logarithm of the normal distribution.
- b) If the natural log of the random variable X is lognormally distributed, then X is normally distributed.
- c) If X is lognormally distributed, then the natural log of X is normally distributed.
- d) The two distributions have nothing to do with one another.

Example 2-11: FRM Exam 2002—Question 125

Consider a stock with an initial price of \$100. Its price one year from now is given by $S = 100 \times \exp(r)$, where the rate of return r is normally distributed with a mean of 0.1 and a standard deviation of 0.2. With 95% confidence, after rounding, S will be between

- a) \$67.57 and \$147.99
- b) \$70.80 and \$149.20
- c) \$74.68 and \$163.56
- d) \$102.18 and \$119.53

Example 2-12: FRM Exam 1998—Question 16

Which of the following statements are *true*?

- I. The sum of two random normal variables is also a random normal variable.
 - II. The product of two random normal variables is also a random normal variable.
 - III. The sum of two random lognormal variables is also a random lognormal variable.
 - IV. The product of two random lognormal variables is also a random lognormal variable.
- a) I and II only
 - b) II and III only
 - c) III and IV only
 - d) I and IV only

Example 2-13: FRM Exam 2000—Question 128

For a lognormal variable X , we know that $\ln(X)$ has a normal distribution with a mean of zero and a standard deviation of 0.5. What are the expected value and the variance of X ?

- a) 1.025 and 0.187
- b) 1.126 and 0.217
- c) 1.133 and 0.365
- d) 1.203 and 0.399

2.4.4 Student's t Distribution

Another important distribution is the **Student's t distribution**. This arises in hypothesis testing, because it describes the distribution of the ratio of the estimated coefficient to its standard error.

This distribution is characterized by a parameter k known as the **degrees of freedom**. Its density is

$$f(x) = \frac{\Gamma\{(k+1)/2\}}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+x^2/k)^{(k+1)/2}} \quad (2.47)$$

where Γ is the gamma function, defined as $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$. As k increases, this function converges to the normal p.d.f.

The distribution is symmetrical with mean zero and variance

$$V[X] = \frac{k}{k-2} \quad (2.48)$$

provided $k > 2$. Its kurtosis is

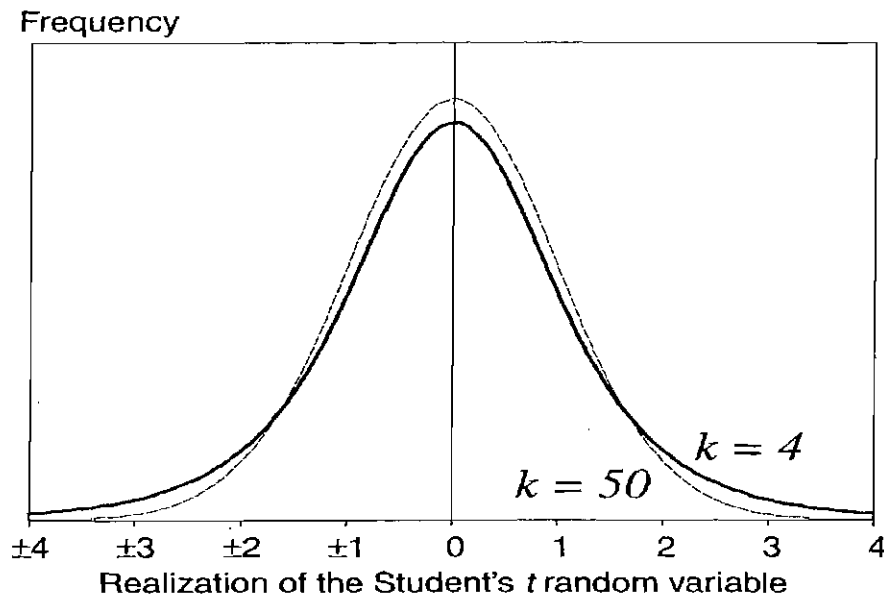
$$\delta = 3 + \frac{6}{k-4} \quad (2.49)$$

provided $k > 4$. It has fatter tails than the normal, which often allows a better representation of typical financial variables. Typical estimated values of k are around 4 to 6. Figure 2-9 displays the density for $k = 4$ and $k = 50$. The latter is close to the normal. With $k = 4$, however, the p.d.f. has fatter tails. As was done for the normal density, we can use the Student's t to compute VAR as a function of the volatility,

$$\text{VAR} = \alpha_k \sigma \quad (2.50)$$

where the multiplier now depends on the degrees of freedom k .

FIGURE 2-9 Student's t Density Function



Another distribution derived from the normal is the **chi-square distribution**, which can be viewed as the sum of independent squared standard normal variables:

$$x = \sum_{j=1}^k z_j^2 \quad (2.51)$$

where k again represents the degrees of freedom. Its mean is $E[X] = k$ and variance $V[X] = 2k$. For k sufficiently large, $\chi^2(k)$ converges to a normal distribution $N(k, 2k)$. This distribution describes the sample variance.

Another associated distribution is the **F distribution**, which can be viewed as the ratio of independent chi-square variables divided by their degrees of freedom:

$$F(a, b) = \frac{\chi^2(a)/a}{\chi^2(b)/b} \quad (2.52)$$

This distribution appears in joint tests of regression coefficients.

Example 2-14: FRM Exam 1999—Question 3

It is often said that distributions of returns from financial instruments are leptokurtic. For such distributions, which of the following comparisons with a normal distribution of the same mean and variance *must* hold?

- a) The skew of the leptokurtic distribution is greater.
- b) The kurtosis of the leptokurtic distribution is greater.
- c) The skew of the leptokurtic distribution is smaller.
- d) The kurtosis of the leptokurtic distribution is smaller.

2.4.5 Binomial Distribution

Consider now a random variable that can take discrete values between zero and n . This could be, for instance, the number of times VAR is exceeded over the last year, also called the number of **exceptions**. Thus, the binomial distribution plays an important role in the backtesting of VAR models.

A binomial variable can be viewed as the result of n independent **Bernoulli trials**, where each trial results in an outcome of $y = 0$ or $y = 1$. This applies, for example, to credit risk. In case of default, we have $y = 1$; otherwise $y = 0$. Each Bernoulli variable has expected value $E[Y] = p$ and variance $V[Y] = p(1 - p)$.

A random variable is defined to have a **binomial distribution** if the discrete density function is given by

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n \quad (2.53)$$

where $\binom{n}{x}$ is the number of combinations of n things taken x at a time, or

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad (2.54)$$

and the parameter p is between 0 and 1. This distribution also represents the total number of successes in n repeated experiments where each success has a probability of p .

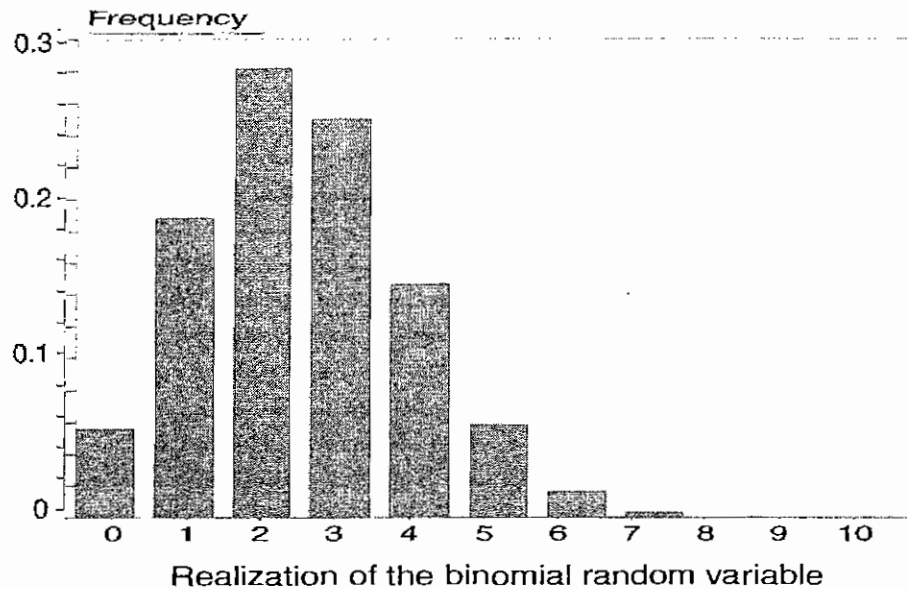
The binomial variable has mean and variance

$$E[X] = pn \quad (2.55)$$

$$V[X] = p(1 - p)n \quad (2.56)$$

It is presented in Figure 2-10 for the case where $p = 0.25$ and $n = 10$. The probability of observing $X = 0, 1, 2, \dots$ is 5.6%, 18.8%, 28.1%, and so on.

FIGURE 2-10 Binomial Density Function with $p = 0.25, n = 10$



For instance, we want to know the probability of observing $x = 0$ exceptions out of a sample of $n = 250$ observations when the true probability is 1%. We should expect to observe 2.5 exceptions on average across many such samples. There will be, however, some samples with no exceptions at all. This probability is

$$f(X = 0) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \frac{250!}{1 \times 250!} 0.01^0 0.99^{250} = 0.081$$

So we would expect to observe 8.1% of samples with zero exceptions, under the null hypothesis. We can repeat this calculation with different values for x . For example, the probability of observing eight exceptions is $f(X = 8) = 0.02\%$ only. We can use this information to test the null hypothesis. Because this probability is so low, observing eight exceptions would make us question whether the true probability is 1%.

2.5 Limit Distributions

2.5.1 Distribution of Averages

The normal distribution is extremely important because of the **central limit theorem** (CLT), which states that the mean of n independent and identically distributed variables converges to a normal distribution as the number of observations n increases. This very powerful result is valid for any underlying distribution, as long as the realizations are independent. For instance, the distribution of total credit losses converges to a normal distribution as the number of loans increases to a large value, assuming defaults are always independent of each other.

Define \bar{X} as the mean $\frac{1}{n} \sum_{i=1}^n X_i$, where each variable has mean μ and standard deviation σ . We have

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \quad (2.57)$$

Standardizing the variable, we can write

$$\frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} \rightarrow N(0, 1) \quad (2.58)$$

If \bar{X} is the sum instead of the mean, this result still holds, but without the \sqrt{n} term in the denominator. Thus, the normal distribution is the limiting distribution of the average, which explains why it has such a prominent place in statistics.

As an example, consider the binomial variable, which is the sum of independent Bernoulli trials. When n is large, we can use the CLT and approximate the binomial distribution by the normal distribution. Using Equation (2.58) for the sum, we have

$$z = \frac{x - pn}{\sqrt{p(1-p)n}} \rightarrow N(0, 1) \quad (2.59)$$

which is much easier to evaluate than the binomial distribution.

Consider, for example, the issue of whether the number of exceptions x we observe is compatible with a 99% VAR. For our example, the mean and variance of x are $E[X] = 0.01 \times 250 = 2.5$ and $V[X] = 0.01(1 - 0.01) \times 250 = 2.475$. We observe $x = 8$, which gives $z = (8 - 2.5)/\sqrt{2.475} = 3.50$. We can now compare this number to the standard normal distribution. Say we decide to reject the hypothesis that VAR is correct if the statistic falls outside a 95% two-tailed confidence band. This interval is $(-1.96, +1.96)$ for the standardized normal distribution. Here the value of

3.50 is much higher than the cutoff point of +1.96. As a result, we would reject the null hypothesis that the true probability of observing an exception is 1% only. In other words, there are simply too many exceptions to be explained by bad luck. It is more likely that the VAR model underestimates risk.

2.5.2 Distribution of Tails

The CLT deals with the mean, or center of the distribution. For risk management purposes, it is also useful to examine the tails of the distribution.

Another powerful theorem is given by **extreme value theory** (EVT). The EVT theorem says that the limit distribution for values x beyond a cutoff point u belongs to the family

$$\begin{aligned} F(y) &= 1 - (1 + \xi y)^{-1/\xi}, & \xi &\neq 0 \\ F(y) &= 1 - \exp(-y), & \xi &= 0 \end{aligned} \quad (2.60)$$

where $y = (x - u)/\beta$. To simplify, we define the loss x as a positive number so that y is also positive. The distribution is characterized by $\beta > 0$, a *scale* parameter, and by ξ , a *shape* parameter that determines the speed at which the tail disappears.

This distribution is called the **generalized Pareto distribution**, because it subsumes other distributions as special cases. For instance, the normal distribution corresponds to $\xi = 0$, in which case the tails disappear at an exponential speed. Typical financial data have $\xi > 0$, which implies *fat tails*, as in the case of the Student's t distribution. This distribution is often used to smooth out the tails of empirical data.

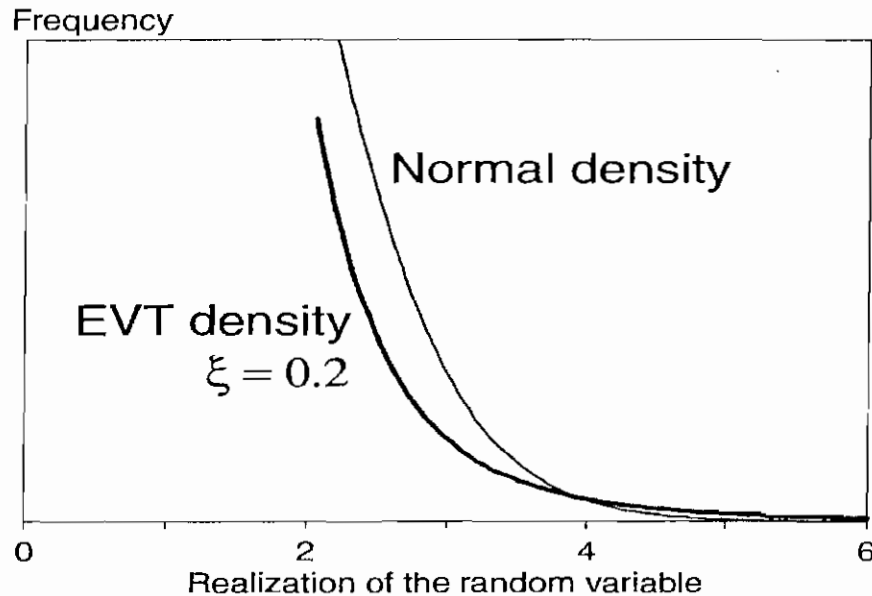
Figure 2-11 illustrates the shape of the density function for U.S. stock market data. The normal density falls off fairly quickly. The EVT density is defined only when the loss x exceeds an arbitrary cutoff point, which is taken as 2 in this case. With $\xi = 0.2$, the EVT density has a fatter tail than the normal density, implying a higher probability of experiencing large losses. This is an important observation for risk management purposes.

Example 2-15: FRM Exam 2001—Question 68

EVT, extreme value theory, helps quantify two key measures of risk:

- The magnitude of an X -year return in the loss in excess of VAR
- The magnitude of VAR and the level of risk obtained from scenario analysis
- The magnitude of market risk and the magnitude of operational risk
- The magnitude of market risk and the magnitude of credit risk

FIGURE 2-11 EVT and Normal Densities



2.6 Answers to Chapter Examples

Example 2-1: FRM Exam 1999—Question 21

c) From Equation (2.21), we have $\sigma_B = \text{Cov}(A, B) / (\rho\sigma_A) = 5 / (0.5\sqrt{12}) = 2.89$, for a variance of $\sigma_B^2 = 8.33$.

Example 2-2: FRM Exam 2000—Question 81

b) Correlation is a measure of linear association. Independence implies zero correlation, but the reverse is not always true.

Example 2-3: FRM Exam 2002—Question 70

b) Statement a) is correct, as it is a linear operation. Statement c) is correct, as shown by Equation (2.30). Statement d) is correct, as the covariance term is zero if the variables are uncorrelated. Statement b) is false, as adding a constant c to a variable cannot change the variance. The constant drops out because it is also in the expectation.

Example 2-4: FRM Exam 2002—Question 119

a) The mean is the center of the distribution, which is the average of a and b .

Example 2-5: FRM Exam 2000—Question 108

b) First, we compute the standard variate for each cutoff point: $\epsilon_1 = (43 - 45)/16 =$

-0.125 and $\epsilon_2 = (39 - 45)/16 = -0.375$. Next, we compute the cumulative distribution function for each: $F(\epsilon_1) = 0.450$ and $F(\epsilon_2) = 0.354$. Hence, the difference is a probability of $0.450 - 0.354 = 0.096$.

Example 2-6: FRM Exam 1999—Question 12

b) See Figure 2-7.

Example 2-7: FRM Exam 1999—Question 13

d) Note that b) is not correct because the kurtosis involves σ^4 in the denominator and is hence scale-free.

Example 2-8: FRM Exam 1999—Question 16

a) As in Equation (2.13), the kurtosis adjusts for σ . Greater kurtosis than for the normal implies fatter tails.

Example 2-9: FRM Exam 2001—Question 72

a) The lognormal distribution has a long left tail, as in Figure 2-8. So it is positively skewed.

Example 2-10: FRM Exam 1999—Question 5

c) X is said to be lognormally distributed if its logarithm $Y = \ln(X)$ is normally distributed.

Example 2-11: FRM Exam 2002—Question 125

c) Note that this is a two-tailed confidence band, so $\alpha = 1.96$. We find the extreme values from $\$100 \exp(\mu \pm \alpha\sigma)$. The lower limit is then $V_1 = \$100 \exp(0.10 - 1.96 \times 0.2) = \$100 \exp(-0.292) = \$74.68$. The upper limit is $V_2 = \$100 \exp(0.10 + 1.96 \times 0.2) = \$100 \exp(0.492) = \$163.56$.

Example 2-12: FRM Exam 1998—Question 16

d) Normal variables are stable under addition, so I is true. For lognormal variables X_1 and X_2 , we know that their logs, $Y_1 = \ln(X_1)$ and $Y_2 = \ln(X_2)$, are normally distributed. Hence, the sum of their logs, or $\ln(X_1) + \ln(X_2) = \ln(X_1 X_2)$, must also be normally distributed. The product is itself lognormal, so IV is true.

Example 2-13: FRM Exam 2000—Question 128

c) Using Equation (2.45), we have $E[X] = \exp[\mu + 0.5\sigma^2] = \exp[0 + 0.5 \cdot 0.5^2] = 1.1331$. Assuming there is no error in the answers listed for the variance, it is sufficient to find the correct answer for the expected value.

Example 2-14: FRM Exam 1999—Question 3

b) Leptokurtic refers to a distribution with fatter tails than the normal, which implies greater kurtosis.

Example 2-15: FRM Exam 2001—Question 68

a) EVT allows risk managers to approximate distributions in the tails beyond the usual VAR confidence levels. Answers c) and d) are too general. Answer b) is also incorrect, as EVT is based on historical data rather than scenario analyses.

Appendix: Review of Matrix Multiplication

This appendix briefly reviews the mathematics of matrix multiplication. Suppose we have two matrices, A and B , that we wish to multiply to obtain the new matrix $C = AB$. The respective dimensions are $(n \times m)$ for A , that is, n rows and m columns, and $(m \times p)$ for B . The number of columns for A must exactly match (or conform to) the number of rows for B . If so, this will result in a matrix C of dimensions $(n \times p)$.

We can write the matrix A in terms of its individual components a_{ij} , where i denotes the row and j denotes the column:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

As an illustration, take a simple example where the matrices are of dimension (2×3) and (3×2) .

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$C = AB = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

To multiply the matrices, each row of A is multiplied element by element by each column of B . For instance, c_{12} is obtained by taking

$$c_{12} = [a_{11} \quad a_{12} \quad a_{13}] \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

The matrix C is then

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

Matrix multiplication can be easily implemented in Excel using the function =MMULT. First, we highlight the cells representing the output matrix C , say f1:g2. Then we enter the function, for instance =MMULT(a1:c2; d1:e3), where the first range represents the first matrix, A , here 2 by 3, and the second range represents matrix B , here 3 by 2. The final step is to hit the three keys Control-Shift-Return simultaneously.