Chapter 1

Review of Random Variables

Updated: January 16, 2015

This chapter reviews basic probability concepts that are necessary for the modeling and statistical analysis of financial data.

1.1 Random Variables

We start with the basic definition of a random variable:

Definition 1 *A Random variable X is a variable that can take on a given set of values, called the sample space and denoted S_X, where the likelihood of the values in S_X is determined by X's probability distribution function (pdf).*

Example 2 *Future price of Microsoft stock*

Consider the price of Microsoft stock next month. Since the price of Microsoft stock next month is not known with certainty today, we can consider it a random variable. The price next month must be positive and realistically it can't get too large. Therefore the sample space is the set of positive real numbers bounded above by some large number: $S_P = \{P : P \in [0, M], M > 0\}$. It is an open question as to what is the best characterization of the probability distribution of stock prices. The log-normal distribution is one possibility.[1] ■

[1] If P is a positive random variable such that $\ln P$ is normally distributed then P has a log-normal distribution.

Example 3 *Return on Microsoft stock*

Consider a one-month investment in Microsoft stock. That is, we buy one share of Microsoft stock at the end of month $t-1$ (e.g., end of February) and plan to sell it at the end of month t (e.g., end of March). The return over month t on this investment, $R_t = (P_t - P_{t-1})/P_{t-1}$, is a random variable because we do not know what the price will be at the end of the month. In contrast to prices, returns can be positive or negative and are bounded from below by -100%. We can express the sample space as $S_{R_t} = \{R_t : R_t \in [-1, M],\ M > 0\}$. The normal distribution is often a good approximation to the distribution of simple monthly returns, and is a better approximation to the distribution of continuously compounded monthly returns. ■

Example 4 *Up-down indicator variable*

As a final example, consider a variable X defined to be equal to one if the monthly price change on Microsoft stock, $P_t - P_{t-1}$, is positive, and is equal to zero if the price change is zero or negative. Here, the sample space is the set $S_X = \{0, 1\}$. If it is equally likely that the monthly price change is positive or negative (including zero) then the probability that $X = 1$ or $X = 0$ is 0.5. This is an example of a bernoulli random variable. ■

The next sub-sections define discrete and continuous random variables.

1.1.1 Discrete Random Variables

Consider a random variable generically denoted X and its set of possible values or sample space denoted S_X.

Definition 5 *A discrete random variable X is one that can take on a finite number of n different values $S_X = \{x_1, x_2, \ldots, x_n\}$ or, at most, a countably infinite number of different values $S_X = \{x_1, x_2, \ldots.\}$.*

Definition 6 *The pdf of a discrete random variable, denoted $p(x)$, is a function such that $p(x) = \Pr(X = x)$. The pdf must satisfy (i) $p(x) \geq 0$ for all $x \in S_X$; (ii) $p(x) = 0$ for all $x \notin S_X$; and (iii) $\sum_{x \in S_X} p(x) = 1$.*

Example 7 *Annual return on Microsoft stock*

State of Economy	S_X = Sample Space	$p(x) = \Pr(X = x)$
Depression	-0.30	0.05
Recession	0.0	0.20
Normal	0.10	0.50
Mild Boom	0.20	0.20
Major Boom	0.50	0.05

Table 1.1: Probability distribution for the annual return on Microsoft

Let X denote the annual return on Microsoft stock over the next year. We might hypothesize that the annual return will be influenced by the general state of the economy. Consider five possible states of the economy: depression, recession, normal, mild boom and major boom. A stock analyst might forecast different values of the return for each possible state. Hence, X is a discrete random variable that can take on five different values. Table 1.1 describes such a probability distribution of the return and a graphical representation of the probability distribution is presented in Figure 1.1.

■

The Bernoulli Distribution

Let $X = 1$ if the price next month of Microsoft stock goes up and $X = 0$ if the price goes down (assuming it cannot stay the same). Then X is clearly a discrete random variable with sample space $S_X = \{0, 1\}$. If the probability of the stock price going up or down is the same then $p(0) = p(1) = 1/2$ and $p(0) + p(1) = 1$.

The probability distribution described above can be given an exact mathematical representation known as the *Bernoulli* distribution. Consider two mutually exclusive events generically called "success" and "failure". For example, a success could be a stock price going up or a coin landing heads and a failure could be a stock price going down or a coin landing tails. The process creating the success or failure is called a *Bernoulli trial*. In general, let $X = 1$ if success occurs and let $X = 0$ if failure occurs. Let $\Pr(X = 1) = \pi$, where $0 < \pi < 1$, denote the probability of success. Then $\Pr(X = 0) = 1 - \pi$ is the

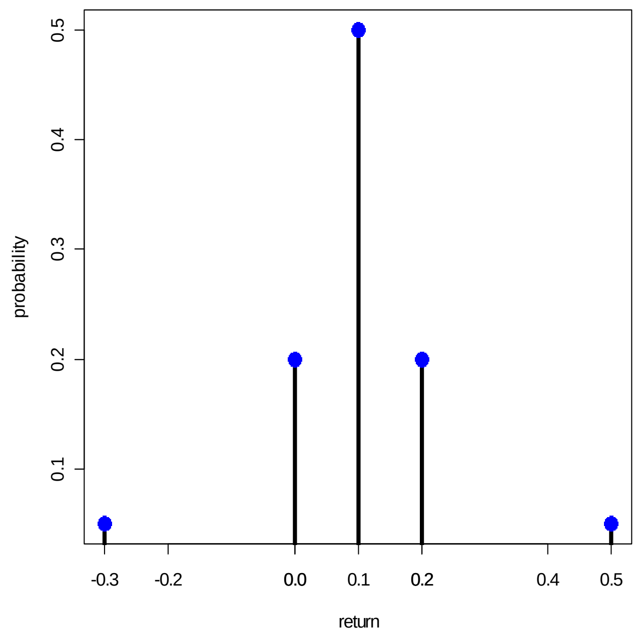

Figure 1.1: Discrete distribution for Microsoft stock.

probability of failure. A mathematical model describing this distribution is

$$p(x) = \Pr(X = x) = \pi^x(1-\pi)^{1-x}, \; x = 0, 1. \tag{1.1}$$

When $x = 0$, $p(0) = \pi^0(1-\pi)^{1-0} = 1-\pi$ and when $x = 1, p(1) = \pi^1(1-\pi)^{1-1} = \pi$.

The Binomial Distribution

Consider a sequence of independent Bernoulli trials with success probability π generating a sequence of $0-1$ variables indicating failures and successes. A *binomial* random variable X counts the number of successes in n Bernoulli trials, and is denoted $X \sim B(n, \pi)$. The sample space is $S_X = \{0, 1, \ldots, n\}$ and

$$\Pr(X = k) = \binom{n}{k}\pi^k(1-\pi)^{n-k}.$$

The term $\binom{n}{k}$ is the binomial coefficient, and counts the number of ways k objects can be chosen from n distinct objects. It is defined by

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

where $n!$ is the factorial of n, or $n(n-1)\cdots 2\cdot 1$.

Example 8 *Binomial tree model for stock prices*

To be completed.

1.1.2 Continuous Random Variables

Definition 9 *A continuous random variable X is one that can take on any real value. That is, $S_X = \{x : x \in \mathbb{R}\}$.*

Definition 10 *The probability density function (pdf) of a continuous random variable X is a nonnegative function f, defined on the real line, such that for any interval A*

$$\Pr(X \in A) = \int_A f(x)dx.$$

That is, $\Pr(X \in A)$ *is the "area under the probability curve over the interval A". The pdf $f(x)$ must satisfy (i) $f(x) \geq 0$; and (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$.*

A typical "bell-shaped" pdf is displayed in Figure 1.2 and the area under the curve between -2 and 1 represents $\Pr(-2 \leq X < 1)$. For a continuous random variable, $f(x) \neq \Pr(X = x)$ but rather gives the height of the probability curve at x. In fact, $\Pr(X = x) = 0$ for all values of x. That is, probabilities are not defined over single points. They are only defined over intervals. As a result, for a continuous random variable X we have

$$\Pr(a \leq X \leq b) = \Pr(a < X \leq b) = \Pr(a < X < b) = \Pr(a \leq X < b).$$

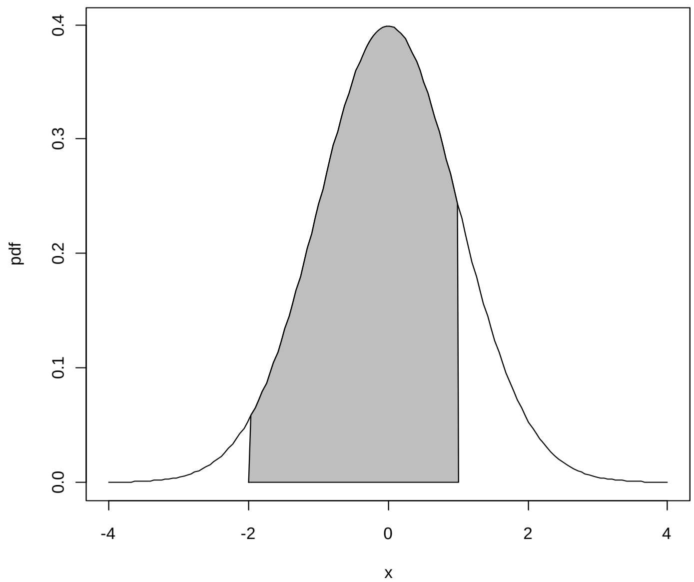

Figure 1.2: $\Pr(-2 \leq X \leq 1)$ is represented by the area under the probability curve.

The Uniform Distribution on an Interval

Let X denote the annual return on Microsoft stock and let a and b be two real numbers such that $a < b$. Suppose that the annual return on Microsoft stock can take on any value between a and b. That is, the sample space is restricted to the interval $S_X = \{x \in \mathcal{R} : a \leq x \leq b\}$. Further suppose that the probability that X will belong to any subinterval of S_X is proportional to the length of the interval. In this case, we say that X is *uniformly distributed* on the interval $[a, b]$. The pdf of X has a very simple mathematical form:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise,} \end{cases}$$

and is presented graphically in Figure 1.3. Notice that the area under the curve over the interval $[a, b]$ (area of rectangle) integrates to one:

$$\int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b dx = \frac{1}{b-a} [x]_a^b = \frac{1}{b-a} [b-a] = 1.$$

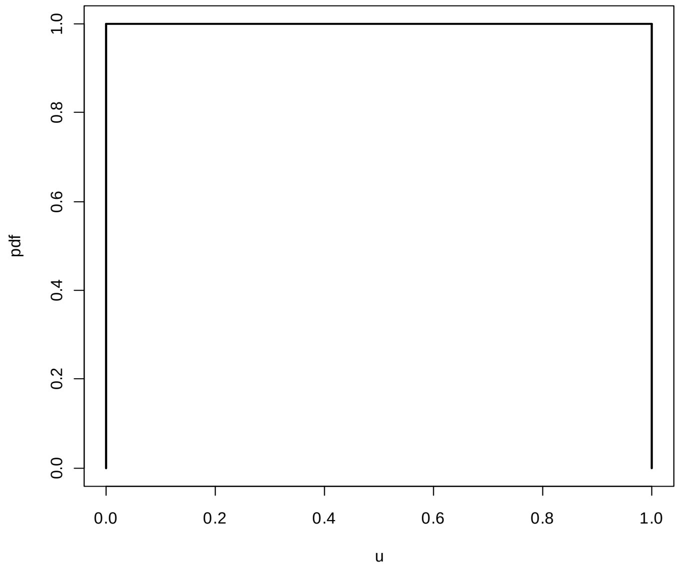

Figure 1.3: Uniform distribution over the interval [0,1]. That is, $X \sim U(0,1)$.

Example 11 *Uniform distribution on* $[-1,1]$

Let $a = -1$ and $b = 1$, so that $b - a = 2$. Consider computing the probability that the return will be between -50% and 50%.We solve

$$\Pr(-50\% < X < 50\%) = \int_{-0.5}^{0.5} \frac{1}{2} dx = \frac{1}{2} [x]_{-0.5}^{0.5} = \frac{1}{2} [0.5 - (-0.5)] = \frac{1}{2}.$$

Next, consider computing the probability that the return will fall in the interval $[0, \delta]$ where δ is some small number less than $b = 1$:

$$\Pr(0 \leq X \leq \delta) = \frac{1}{2} \int_0^{\delta} dx = \frac{1}{2} [x]_0^{\delta} = \frac{1}{2} \delta.$$

As $\delta \to 0$, $\Pr(0 \leq X \leq \delta) \to \Pr(X = 0)$. Using the above result we see that

$$\lim_{\delta \to 0} \Pr(0 \leq X \leq \delta) = \Pr(X = 0) = \lim_{\delta \to 0} \frac{1}{2} \delta = 0.$$

Hence, probabilities are defined on intervals but not at distinct points. ■

The Standard Normal Distribution

The normal or Gaussian distribution is perhaps the most famous and most useful continuous distribution in all of statistics. The shape of the normal distribution is the familiar "bell curve". As we shall see, it can be used to describe the probabilistic behavior of stock returns although other distributions may be more appropriate.

If a random variable X follows a *standard normal distribution* then we often write $X \sim N(0,1)$ as short-hand notation. This distribution is centered at zero and has inflection points at ± 1.[2] The pdf of a normal random variable is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \quad -\infty \leq x \leq \infty. \tag{1.2}$$

It can be shown via the change of variables formula in calculus that the area under the standard normal curve is one:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx = 1.$$

The standard normal distribution is illustrated in Figure 1.4. Notice that the distribution is *symmetric* about zero; i.e., the distribution has exactly the same form to the left and right of zero.

The normal distribution has the annoying feature that the area under the normal curve cannot be evaluated analytically. That is

$$\Pr(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx,$$

does not have a closed form solution. The above integral must be computed by numerical approximation. Areas under the normal curve, in one form or another, are given in tables in almost every introductory statistics book and standard statistical software can be used to find these areas. Some useful approximate results are:

$$\begin{aligned}
\Pr(-1 \leq X \leq 1) &\approx 0.67, \\
\Pr(-2 \leq X \leq 2) &\approx 0.95, \\
\Pr(-3 \leq X \leq 3) &\approx 0.99.
\end{aligned}$$

[2] An inflection point is a point where a curve goes from concave up to concave down, or vice versa.

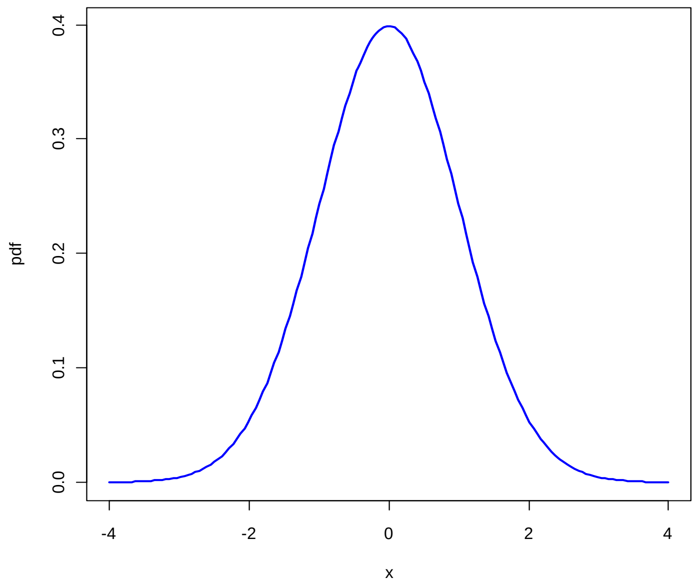

Figure 1.4: Standard normal density.

1.1.3 The Cumulative Distribution Function

Definition 12 *The cumulative distribution function (cdf) of a random variable X (discrete or continuous), denoted F_X, is the probability that $X \leq x$:*

$$F_X(x) = \Pr(X \leq x), \quad -\infty \leq x \leq \infty.$$

■

The cdf has the following properties:

(i) If $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$

(ii) $F_X(-\infty) = 0$ and $F_X(\infty) = 1$

(iii) $\Pr(X > x) = 1 - F_X(x)$

(iv) $\Pr(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

(v) $F'_X(x) = \frac{d}{dx}F_X(x) = f(x)$ if X is a continuous random variable and $F_X(x)$ is continuous and differentiable.

Example 13 *$F_X(x)$ for a discrete random variable*

The cdf for the discrete distribution of Microsoft from Table 1.1 is given by

$$F_X(x) = \begin{cases} 0, & x < -0.3 \\ 0.05, & -0.3 \leq x < 0 \\ 0.25, & 0 \leq x < 0.1 \\ 0.75 & 0.1 \leq x < 0.2 \\ 0.95 & 0.2 \leq x < 0.5 \\ 1 & x > 0.5 \end{cases}$$

and is illustrated in Figure 1.5. ■

Example 14 *$F_X(x)$ for a uniform random variable*

The cdf for the uniform distribution over $[a, b]$ can be determined analytically:

$$\begin{aligned} F_X(x) &= \Pr(X < x) = \int_{-\infty}^{x} f(t)dt \\ &= \frac{1}{b-a}\int_{a}^{x} dt = \frac{1}{b-a}[t]_a^x = \frac{x-a}{b-a}. \end{aligned}$$

We can determine the pdf of X directly from the cdf via

$$f(x) = F'_X(x) = \frac{d}{dx}F_X(x) = \frac{1}{b-a}.$$

■

Example 15 *$F_X(x)$ for a standard normal random variable*

The cdf of standard normal random variable X is used so often in statistics that it is given its own special symbol:

$$\Phi(x) = F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz. \tag{1.3}$$

The cdf $\Phi(x)$, however, does not have an analytic representation like the cdf of the uniform distribution and so the integral in (1.3) must be approximated using numerical techniques. A graphical representation of $\Phi(x)$ is given in Figure 1.6. ■

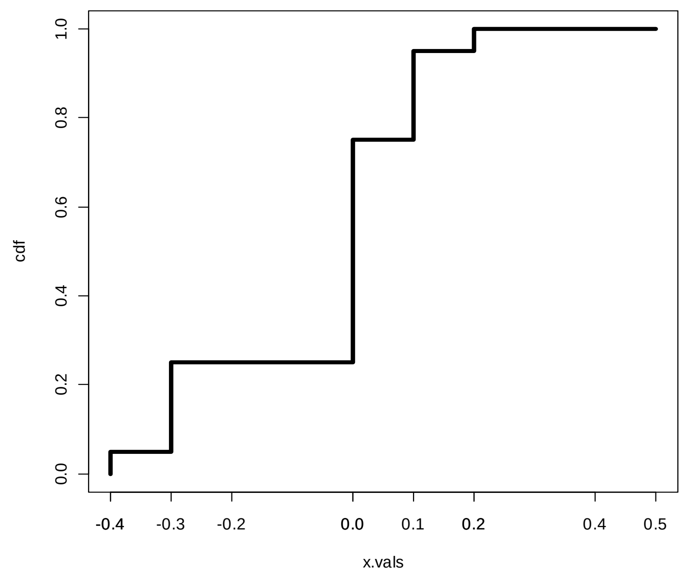

Figure 1.5: CDF of Discrete Distribution for Microsoft Stock Return.

1.1.4 Quantiles of the Distribution of a Random Variable

Consider a random variable X with continuous cdf $F_X(x)$. For $0 \leq \alpha \leq 1$, the $100 \cdot \alpha\%$ *quantile* of the distribution for X is the value q_α that satisfies

$$F_X(q_\alpha) = \Pr(X \leq q_\alpha) = \alpha.$$

For example, the 5% quantile of X, $q_{0.05}$, satisfies

$$F_X(q_{0.05}) = \Pr(X \leq q_{0.05}) = 0.05.$$

The *median* of the distribution is 50% quantile. That is, the median, $q_{0.5}$, satisfies

$$F_X(q_{0.5}) = \Pr(X \leq q_{0.5}) = 0.5.$$

If F_X is invertible[3] then q_α may be determined analytically as

$$q_\alpha = F_X^{-1}(\alpha) \tag{1.4}$$

[3] The inverse of $F(x)$ will exist if F is strictly increasing and is continuous.

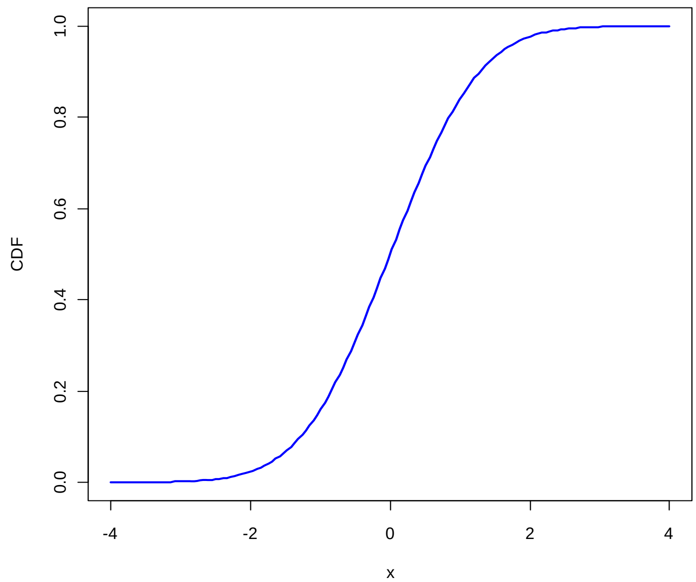

Figure 1.6: Standard normal cdf $\Phi(x)$.

where F_X^{-1} denotes the inverse function of F_X. Hence, the 5% quantile and the median may be determined from

$$q_{0.05} = F_X^{-1}(.05), q_{0.5} = F_X^{-1}(.5).$$

In inverse cdf F_X^{-1} is sometimes called the *quantile function*.

Example 16 *Quantiles from a uniform distribution*

Let $X \sim U[a, b]$ where $b > a$. Recall, the cdf of X is given by

$$F_X(x) = \frac{x - a}{b - a}, \; a \leq x \leq b,$$

which is continuous and strictly increasing. Given $\alpha \in [0, 1]$ such that $F_X(x) = \alpha$, solving for x gives the inverse cdf:

$$x = F_X^{-1}(\alpha) = \; \alpha(b - a) + a. \tag{1.5}$$

Using (1.5), the 5% quantile and median, for example, are given by

$$\begin{aligned} q_{0.05} &= F_X^{-1}(.05) = .05(b-a)+a = .05b+.95a, \\ q_{0.5} &= F_X^{-1}(.5) = .5(b-a)+a = .5(a+b). \end{aligned}$$

If $a=0$ and $b=1$, then $q_{0.05}=0.05$ and $q_{0.5}=0.5$. ■

Example 17 *Quantiles from a standard normal distribution*

Let $X \sim N(0,1)$. The quantiles of the standard normal distribution are determined by solving

$$q_\alpha = \Phi^{-1}(\alpha), \tag{1.6}$$

where Φ^{-1} denotes the inverse of the cdf Φ. This inverse function must be approximated numerically and is available in most spreadsheets and statistical software. Using the numerical approximation to the inverse function, the 1%, 2.5%, 5%, 10% quantiles and median are given by

$$\begin{aligned} q_{0.01} &= \Phi^{-1}(.01) = -2.33,\ q_{0.025} = \Phi^{-1}(.025) = -1.96, \\ q_{0.05} &= \Phi^{-1}(.05) = -1.645,\ q_{0.10} = \Phi^{-1}(.10) = -1.28, \\ q_{0.5} &= \Phi^{-1}(.5) = 0. \end{aligned}$$

Often, the standard normal quantile is denoted z_α. ■

1.1.5 R Functions for Discrete and Continuous Distributions

R has built-in functions for a number of discrete and continuous distributions. These are summarized in Table 1.2. For each distribution, there are four functions starting with `d`, `p`, `q` and `r` that compute density (pdf) values, cumulative probabilities (cdf), quantiles (inverse cdf) and random draws, respectively. Consider, for example, the functions associated with the normal distribution. The functions `dnorm()`, `pnorm()` and `qnorm()` evaluate the standard normal density (1.2), the cdf (1.3), and the inverse cdf or quantile function (1.6), respectively, with the default values `mean=0` and `sd = 1`. The function `rnorm()` returns a specified number of simulated values from the normal distribution.

Distribution	**Function (root)**	Parameters	Defaults
beta	beta	shape1, shape2	_, _
binomial	binom	size, prob	_, _
Cauchy	cauchy	location, scale	0, 1
chi-squared	chisq	df, ncp	_, 1
F	f	df1, df2	_, _
gamma	gamma	shape, rate, scale	_, 1, 1/rate
geometric	geom	prob	_
hyper-geometric	hyper	m, n, k	_, _, _
log-normal	lnorm	meanlog, sdlog	0, 1
logistic	logis	location, scale	0, 1
negative binomial	nbinom	size, prob, mu	_, _, _
normal	norm	mean, sd	0, 1
Poisson	pois	Lambda	1
Student's t	t	df, ncp	_, 1
uniform	unif	min, max	0, 1
Weibull	weibull	shape, scale	_, 1
Wilcoxon	wilcoxon	m, n	_, _

Table 1.2: Probability distributions in base R.

Finding Areas Under the Normal Curve Most spreadsheet and statistical software packages have functions for finding areas under the normal curve. Let X denote a standard normal random variable. Some tables and functions give $\Pr(0 \leq X \leq z)$ for various values of $z > 0$, some give $\Pr(X \geq z)$ and some give $\Pr(X \leq z)$. Given that the total area under the normal curve is one and the distribution is symmetric about zero the following results hold:

- $\Pr(X \leq z) = 1 - \Pr(X \geq z)$ and $\Pr(X \geq z) = 1 - \Pr(X \leq z)$
- $\Pr(X \geq z) = \Pr(X \leq -z)$
- $\Pr(X \geq 0) = \Pr(X \leq 0) = 0.5$

The following examples show how to compute various probabilities.

Example 18 *Finding areas under the normal curve using R*

First, consider finding $\Pr(X \geq 2)$. By the symmetry of the normal distribution, $\Pr(X \geq 2) = \Pr(X \leq -2) = \Phi(-2)$. In R use

```
> pnorm(-2)
[1] 0.0228
```

Next, consider finding $\Pr(-1 \leq X \leq 2)$. Using the cdf, we compute $\Pr(-1 \leq X \leq 2) = \Pr(X \leq 2) - \Pr(X \leq -1) = \Phi(2) - \Phi(-1)$. In R use

```
> pnorm(2) - pnorm(-1)
[1] 0.8186
```

Finally, using R the exact values for $\Pr(-1 \leq X \leq 1)$, $\Pr(-2 \leq X \leq 2)$ and $\Pr(-3 \leq X \leq 3)$ are

```
> pnorm(1) - pnorm(-1)
[1] 0.6827
> pnorm(2) - pnorm(-2)
[1] 0.9545
> pnorm(3) - pnorm(-3)
[1] 0.9973
```

■

Plotting Distributions

When working with a probability distribution, it is a good idea to make plots of the pdf or cdf to reveal important characteristics. The following examples illustrate plotting distributions using R.

Example 19 *Plotting the standard normal curve*

The graphs of the standard normal pdf and cdf in Figures 1.4 and 1.6 were created using the following R code:

```
# plot pdf
> x.vals = seq(-4, 4, length=150)
> plot(x.vals, dnorm(x.vals), type="l", lwd=2, col="blue",
+      xlab="x", ylab="pdf")
# plot cdf
> plot(x.vals, pnorm(x.vals), type="l", lwd=2, col="blue",
+      xlab="x", ylab="CDF")
```

■

Example 20 *Shading a region under the standard normal curve*

Figure 1.2 showing $\Pr(-2 \leq X \leq 1)$ as a red shaded area is created with the following code

```
> lb = -2
> ub = 1
> x.vals = seq(-4, 4, length=150)
> d.vals = dnorm(x.vals)
# plot normal density
> plot(x.vals, d.vals, type="l", xlab="x", ylab="pdf")
> i = x.vals >= lb & x.vals <= ub
# add shaded region between -2 and 1
> polygon(c(lb, x.vals[i], ub), c(0, d.vals[i], 0), col="red")
```

■

1.1.6 Shape Characteristics of Probability Distributions

Very often we would like to know certain shape characteristics of a probability distribution. We might want to know where the distribution is centered, and how spread out the distribution is about the central value. We might want to know if the distribution is symmetric about the center or if the distribution has a long left or right tail. For stock returns we might want to know about the likelihood of observing extreme values for returns representing market crashes. This means that we would like to know about the amount of probability in the extreme tails of the distribution. In this section we discuss four important shape characteristics of a probability distribution:

1. expected value (mean): measures the center of mass of a distribution
2. variance and standard deviation: measures the spread about the mean
3. skewness: measures symmetry about the mean
4. kurtosis: measures "tail thickness"

Expected Value

The expected value of a random variable X, denoted $E[X]$ or μ_X, measures the center of mass of the pdf. For a discrete random variable X with sample space S_X, the expected value is defined as

$$\mu_X = E[X] = \sum_{x \in S_X} x \cdot \Pr(X = x). \tag{1.7}$$

Eq. (1.7) shows that $E[X]$ is a probability weighted average of the possible values of X.

Example 21 *Expected value of discrete random variable*

Using the discrete distribution for the return on Microsoft stock in Table 1.1, the expected return is computed as:

$$\begin{aligned} E[X] &= (-0.3) \cdot (0.05) + (0.0) \cdot (0.20) + (0.1) \cdot (0.5) + (0.2) \cdot (0.2) + (0.5) \cdot (0.05) \\ &= 0.10. \end{aligned}$$

■

Example 22 *Expected value of Bernoulli and binomial random variables*

Let X be a Bernoulli random variable with success probability π. Then

$$E[X] = 0 \cdot (1-\pi) + 1 \cdot \pi = \pi$$

That is, the expected value of a Bernoulli random variable is its probability of success. Now, let $Y \sim B(n, \pi)$. It can be shown that

$$E[Y] = \sum_{k=0}^{n} k \binom{n}{k} \pi^k (1-\pi)^{n-k} = n\pi. \tag{1.8}$$

■

For a continuous random variable X with pdf $f(x)$, the expected value is defined as

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx. \tag{1.9}$$

Example 23 *Expected value of a uniform random variable*

Suppose X has a uniform distribution over the interval $[a, b]$. Then

$$\begin{aligned} E[X] &= \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{1}{2} x^2 \right]_a^b \\ &= \frac{1}{2(b-a)} \left[b^2 - a^2 \right] \\ &= \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}. \end{aligned}$$

If $b = -1$ and $a = 1$, then $E[X] = 0$. ■

Example 24 *Expected value of a standard normal random variable*

Let $X \sim N(0, 1)$. Then it can be shown that

$$E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0.$$

Hence, the standard normal distribution is centered at zero. ■

Expectation of a Function of a Random Variable

The other shape characteristics of the distribution of a random variable X are based on expectations of certain functions of X. Let $g(X)$ denote some function of the random variable X. If X is a discrete random variable with sample space S_X then

$$E[g(X)] = \sum_{x \in S_X} g(x) \cdot \Pr(X = x), \tag{1.10}$$

and if X is a continuous random variable with pdf f then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)dx. \tag{1.11}$$

Variance and Standard Deviation

The variance of a random variable X, denoted $\text{var}(X)$ or σ_X^2, measures the spread of the distribution about the mean using the function $g(X) = (X - \mu_X)^2$. If most values of X are close to μ_X then on average $(X - \mu_X)^2$ will be small. In contrast, if many values of X are far below and/or far above μ_X then on average $(X - \mu_X)^2$ will be large. Squaring the deviations about μ_X guarantees a positive value. The variance of X is defined as

$$\sigma_X^2 = \text{var}(X) = E[(X - \mu_X)^2]. \tag{1.12}$$

Because σ_X^2 represents an average squared deviation, it is not in the same units as X. The standard deviation of X, denoted $\text{sd}(X)$ or σ_X, is the square root of the variance and is in the same units as X. For "bell-shaped" distributions, σ_X measures the typical size of a deviation from the mean value.

The computation of (1.12) can often be simplified by using the result

$$\text{var}(X) = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2 \tag{1.13}$$

Example 25 *Variance and standard deviation for a discrete random variable*

Using the discrete distribution for the return on Microsoft stock in Table 1.1 and the result that $\mu_X = 0.1$, we have

$$\begin{aligned} \text{var}(X) &= (-0.3 - 0.1)^2 \cdot (0.05) + (0.0 - 0.1)^2 \cdot (0.20) + (0.1 - 0.1)^2 \cdot (0.5) \\ &\quad + (0.2 - 0.1)^2 \cdot (0.2) + (0.5 - 0.1)^2 \cdot (0.05) \\ &= 0.020. \end{aligned}$$

Alternatively, we can compute var(X) using (1.13)

$$\begin{aligned}E[X^2] - \mu_X^2 &= (-0.3)^2 \cdot (0.05) + (0.0)^2 \cdot (0.20) + (0.1)^2 \cdot (0.5) \\ &+ (0.2)^2 \cdot (0.2) + (0.5)^2 \cdot (0.05) - (0.1)^2 \\ &= 0.020.\end{aligned}$$

The standard deviation is sd$(X) = \sigma_X = \sqrt{0.020} = 0.141$.Given that the distribution is fairly bell-shaped we can say that typical values deviate from the mean value of 10% by about 14.1%. ■

Example 26 *Variance and standard deviation of a Bernoulli and binomial random variables*

Let X be a Bernoulli random variable with success probability π. Given that $\mu_X = \pi$ it follows that

$$\begin{aligned}\text{var}(X) &= (0-\pi)^2 \cdot (1-\pi) + (1-\pi)^2 \cdot \pi \\ &= \pi^2(1-\pi) + (1-\pi)^2\pi \\ &= \pi(1-\pi)\left[\pi + (1-\pi)\right] \\ &= \pi(1-\pi), \\ \text{sd}(X) &= \sqrt{\pi(1-\pi)}.\end{aligned}$$

Now, let $Y \sim B(n, \pi)$. It can be shown that

$$\text{var}(Y) = n\pi(1-\pi)$$

and so sd$(Y) = \sqrt{n\pi(1-\pi)}$. ■

Example 27 *Variance and standard deviation of a uniform random variable*

Let $X \sim U[a, b]$. Using (1.13) and $\mu_X = \frac{a+b}{2}$, after some algebra, it can be shown that

$$\text{var}(X) = E[X^2] - \mu_X^2 = \frac{1}{b-a}\int_a^b x^2 dx - \left(\frac{a+b}{2}\right)^2 = \frac{1}{12}(b-a)^2,$$

and sd$(X) = (b-a)/\sqrt{12}$.

Example 28 *Variance and standard deviation of a standard normal random variable*

Let $X \sim N(0, 1)$.Here, $\mu_X = 0$ and it can be shown that

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1.$$

It follows that sd$(X) = 1$. ■

The General Normal Distribution

Recall, if X has a standard normal distribution then $E[X] = 0$, $\text{var}(X) = 1$. A general normal random variable X has $E[X] = \mu_X$ and $\text{var}(X) = \sigma_X^2$ and is denoted $X \sim N(\mu_X, \sigma_X^2)$. Its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\right\}, \quad -\infty \leq x \leq \infty. \tag{1.14}$$

Showing that $E[X] = \mu_X$ and $\text{var}(X) = \sigma_X^2$ is a bit of work and is good calculus practice. As with the standard normal distribution, areas under the general normal curve cannot be computed analytically. Using numerical approximations, it can be shown that

$$\begin{aligned}
\Pr(\mu_X - \sigma_X &< X < \mu_X + \sigma_X) \approx 0.67, \\
\Pr(\mu_X - 2\sigma_X &< X < \mu_X + 2\sigma_X) \approx 0.95, \\
\Pr(\mu_X - 3\sigma_X &< X < \mu_X + 3\sigma_X) \approx 0.99.
\end{aligned}$$

Hence, for a general normal random variable about 95% of the time we expect to see values within $\pm$ 2 standard deviations from its mean. Observations more than three standard deviations from the mean are very unlikely.

Example 29 *Normal distribution for monthly returns*

Let R denote the monthly return on an investment in Microsoft stock, and assume that it is normally distributed with mean $\mu_R = 0.01$ and standard deviation $\sigma_R = 0.10$. That is, $R \sim N(0.01, (0.10)^2)$. Notice that $\sigma_R^2 = 0.01$ and is not in units of return per month. Figure 1.7 illustrates the distribution. Notice that essentially all of the probability lies between -0.4 and 0.4. Using the R function `pnorm()`, we can easily compute the probabilities $\Pr(R < -0.5)$, $\Pr(R < 0)$, $\Pr(R > 0.5)$ and $\Pr(R > 1)$:

```
> pnorm(-0.5, mean=0.01, sd=0.1)
[1] 1.698e-07
> pnorm(0, mean=0.01, sd=0.1)
[1] 0.4602
> 1 - pnorm(0.5, mean=0.01, sd=0.1)
[1] 4.792e-07
> 1 - pnorm(1, mean=0.01, sd=0.1)
[1] 0
```

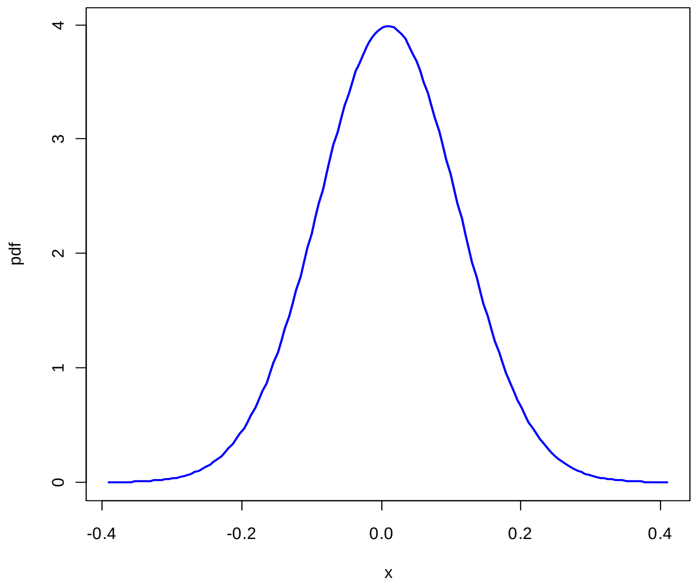

Figure 1.7: Normal distribution for the monthly returns on Microsoft: $R \sim N(0.01, (0.10)^2)$.

Using the R function **qnorm()**, we can find the quantiles $q_{0.01}$, $q_{0.05}$, $q_{0.95}$ and $q_{0.99}$:

```
> a.vals = c(0.01, 0.05, 0.95, 0.99)
> qnorm(a.vals, mean=0.01, sd=0.10)
[1] -0.2226 -0.1545  0.1745  0.2426
```

Hence, over the next month, there are 1% and 5% chances of losing more than 22.2% and 15.5%, respectively. In addition, there are 5% and 1% chances of gaining more than 17.5% and 24.3%, respectively.

■

Example 30 *Risk-return tradeoff*

Consider the following investment problem. We can invest in two non-dividend paying stocks, Amazon and Boeing, over the next month. Let R_A

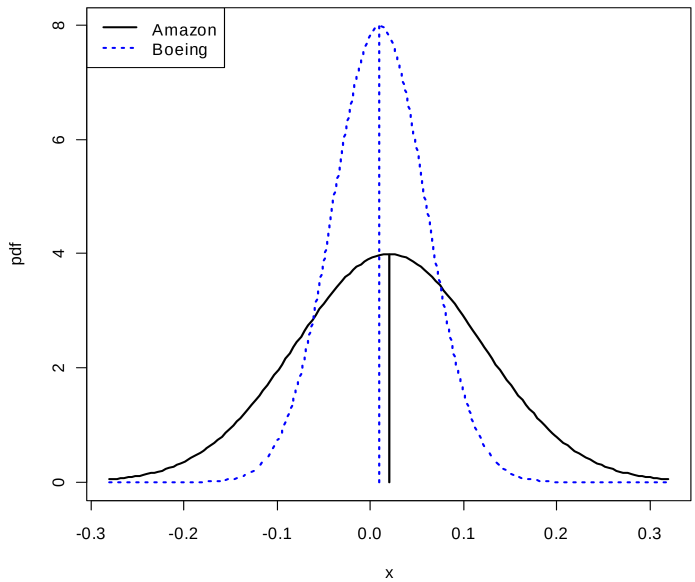

Figure 1.8: Risk-return tradeoff between one-month investments in Amazon and Boeing stock.

denote the monthly return on Amazon and R_B denote the monthly return on Boeing. Assume that $R_A \sim N(0.02, (0.10)^2)$ and $R_B \sim N(0.01, (0.05)^2)$. Figure 1.8 shows the pdfs for the two returns. Notice that $\mu_A = 0.02 > \mu_B = 0.01$ but also that $\sigma_A = 0.10 > \sigma_B = 0.05$. The return we expect on Amazon is bigger than the return we expect on Boeing but the variability Amazon's return is also greater than the variability of Boeing's return. The high return variability (volatility) of Amazon reflects the higher risk associated with investing in Amazon compared to investing in Boeing. If we invest in Boeing we get a lower expected return, but we also get less return variability or risk. This example illustrates the fundamental "no free lunch" principle of economics and finance: you can't get something for nothing. In general, to get a higher expected return you must be prepared to take on higher risk.

Example 31 *Why the normal distribution may not be appropriate for simple returns*

Let R_t denote the simple annual return on an asset, and suppose that $R_t \sim N(0.05, (0.50)^2)$. Because asset prices must be non-negative, R_t must always be larger than -1. However, the normal distribution is defined for $-\infty \leq R_t \leq \infty$ and based on the assumed normal distribution $\Pr(R_t < -1) = 0.018$. That is, there is a 1.8% chance that R_t is smaller than -1. This implies that there is a 1.8% chance that the asset price at the end of the year will be negative! This is why the normal distribution may not appropriate for simple returns. ■

Example 32 *The normal distribution is more appropriate for continuously compounded returns*

Let $r_t = \ln(1+R_t)$ denote the continuously compounded annual return on an asset, and suppose that $r_t \sim N(0.05, (0.50)^2)$. Unlike the simple return, the continuously compounded return can take on values less than -1. In fact, r_t is defined for $-\infty \leq r_t \leq \infty$. For example, suppose $r_t = -2$. This implies a simple return of $R_t = e^{-2} - 1 = -0.865$[4]. Then $\Pr(r_t \leq -2) = \Pr(R_t \leq -0.865) = 0.00002$. Although the normal distribution allows for values of r_t smaller than -1, the implied simple return R_t will always be greater than -1. ■

The Log-Normal distribution

Let $X \sim N(\mu_X, \sigma_X^2)$, which is defined for $-\infty < X < \infty$. The log-normally distributed random variable Y is determined from the normally distributed random variable X using the transformation $Y = e^X$. In this case, we say that Y is log-normally distributed and write

$$Y \sim \ln N(\mu_X, \sigma_X^2),\ 0 < Y < \infty. \tag{1.15}$$

The pdf of the log-normal distribution for Y can be derived from the normal distribution for X using the change-of-variables formula from calculus and is given by

$$f(y) = \frac{1}{y\sigma_X\sqrt{2\pi}} e^{-\frac{(\ln y - \mu_X)^2}{2\sigma_X^2}} \tag{1.16}$$

[4] If $r_t = -\infty$ then $R_t = e^{r_t} - 1 = e^{-\infty} - 1 = 0 - 1 = -1$.

Due to the exponential transformation, Y is only defined for non-negative values. It can be shown that

$$\begin{aligned}\mu_Y &= E[Y] = e^{\mu_X+\sigma_X^2/2}, \\ \sigma_Y^2 &= \mathrm{var}(Y) = e^{2\mu_X+\sigma_X^2}(e^{\sigma_X^2}-1).\end{aligned} \quad (1.17)$$

Example 33 *Log-normal distribution for simple returns*

Let $r_t = \ln(P_t/P_{t-1})$ denote the continuously compounded monthly return on an asset and assume that $r_t \sim N(0.05, (0.50)^2)$. That is, $\mu_r = 0.05$ and $\sigma_r = 0.50$. Let $R_t = \frac{P_t - P_{t-1}}{P_t}$ denote the simple monthly return. The relationship between r_t and R_t is given by $r_t = \ln(1 + R_t)$ and $1 + R_t = e^{r_t}$. Since r_t is normally distributed $1 + R_t$ is log-normally distributed. Notice that the distribution of $1 + R_t$ is only defined for positive values of $1 + R_t$. This is appropriate since the smallest value that R_t can take on is -1. Using (1.17), the mean and variance for $1 + R_t$ are given by

$$\begin{aligned}\mu_{1+R} &= e^{0.05+(0.5)^2/2} = 1.191 \\ \sigma_{1+R}^2 &= e^{2(0.05)+(0.5)^2}(e^{(0.5)^2}-1) = 0.563\end{aligned}$$

The pdfs for r_t and $1 + R_t$ are shown in Figure 1.9. ■

Skewness

The skewness of a random variable X, denoted skew(X), measures the symmetry of a distribution about its mean value using the function $g(X) = (X - \mu_X)^3/\sigma_X^3$, where σ_X^3 is just sd(X) raised to the third power:

$$\mathrm{skew}(X) = \frac{E[(X-\mu_X)^3]}{\sigma_X^3} \quad (1.18)$$

When X is far below its mean $(X - \mu_X)^3$ is a big negative number, and when X is far above its mean $(X - \mu_X)^3$ is a big positive number. Hence, if there are more big values of X below μ_X then skew$(X) < 0$. Conversely, if there are more big values of X above μ_X then skew$(X) > 0$. If X has a symmetric distribution then skew$(X) = 0$ since positive and negative values in (1.18) cancel out. If skew$(X) > 0$ then the distribution of X has a "long right tail" and if skew$(X) < 0$ the distribution of X has a "long left tail". These cases are illustrated in Figure xxx.

insert figure xxx here

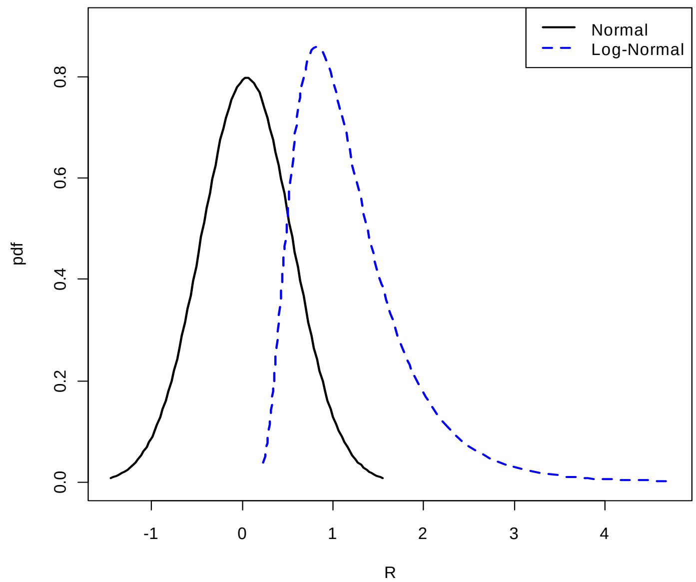

Figure 1.9: Normal distribution for r_t and log-normal distribution for $1+R_t = e^{r_t}$.

Example 34 *Skewness for a discrete random variable*

Using the discrete distribution for the return on Microsoft stock in Table 1, the results that $\mu_X = 0.1$ and $\sigma_X = 0.141$, we have

$$\begin{aligned} \text{skew}(X) &= [(-0.3-0.1)^3 \cdot (0.05) + (0.0-0.1)^3 \cdot (0.20) + (0.1-0.1)^3 \cdot (0.5) \\ &\quad +(0.2-0.1)^3 \cdot (0.2) + (0.5-0.1)^3 \cdot (0.05)]/(0.141)^3 \\ &= 0.0 \end{aligned}$$

■

Example 35 *Skewness for a normal random variable*

Suppose X has a general normal distribution with mean μ_X and variance σ_X^2. Then it can be shown that

$$\text{skew}(X) = \int_{-\infty}^{\infty} \frac{(x-\mu_X)^3}{\sigma_X^3} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2} dx = 0.$$

This result is expected since the normal distribution is symmetric about it's mean value μ_X. ■

Example 36 *Skewness for a log-Normal random variable*

Let $Y = e^X$, where $X \sim N(\mu_X, \sigma_X^2)$, be a log-normally distributed random variable with parameters μ_X and σ_X^2. Then it can be shown that

$$\text{skew}(Y) = \left(e^{\sigma_X^2} + 2\right)\sqrt{e^{\sigma_X^2} - 1} > 0.$$

Notice that skew(Y) is always positive, indicating that the distribution of Y has a long right tail, and that it is an increasing function of σ_X^2. This positive skewness is illustrated in Figure 1.9. ■

Kurtosis

The kurtosis of a random variable X, denoted kurt(X), measures the thickness in the tails of a distribution and is based on $g(X) = (X - \mu_X)^4/\sigma_X^4$:

$$\text{kurt}(X) = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}, \tag{1.19}$$

where σ_X^4 is just sd(X) raised to the fourth power. Since kurtosis is based on deviations from the mean raised to the fourth power, large deviations get lots of weight. Hence, distributions with large kurtosis values are ones where there is the possibility of extreme values. In contrast, if the kurtosis is small then most of the observations are tightly clustered around the mean and there is very little probability of observing extreme values. Figure xxx illustrates distributions with large and small kurtosis values.

Insert Figure Here

Example 37 *Kurtosis for a discrete random variable*

Using the discrete distribution for the return on Microsoft stock in Table 1, the results that $\mu_X = 0.1$ and $\sigma_X = 0.141$, we have

$$\begin{aligned}\text{kurt}(X) &= [(-0.3 - 0.1)^4 \cdot (0.05) + (0.0 - 0.1)^4 \cdot (0.20) + (0.1 - 0.1)^4 \cdot (0.5) \\ &\quad +(0.2 - 0.1)^4 \cdot (0.2) + (0.5 - 0.1)^4 \cdot (0.05)]/(0.141)^4 \\ &= 6.5\end{aligned}$$

■

Example 38 *Kurtosis for a normal random variable*

Suppose X has a general normal distribution mean μ_X and variance σ_X^2. Then it can be shown that

$$\text{kurt}(X) = \int_{-\infty}^{\infty} \frac{(x-\mu_X)^4}{\sigma_X^4} \cdot \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}(\frac{x-\mu_X}{\sigma_X})^2} dx = 3.$$

Hence a kurtosis of 3 is a benchmark value for tail thickness of bell-shaped distributions. If a distribution has a kurtosis greater than 3 then the distribution has thicker tails than the normal distribution and if a distribution has kurtosis less than 3 then the distribution has thinner tails than the normal. ■

Sometimes the kurtosis of a random variable is described relative to the kurtosis of a normal random variable. This relative value of kurtosis is referred to as *excess kurtosis* and is defined as

$$\text{ekurt}(X) = \text{kurt}(X) - 3 \tag{1.20}$$

If the excess kurtosis of a random variable is equal to zero then the random variable has the same kurtosis as a normal random variable. If excess kurtosis is greater than zero, then kurtosis is larger than that for a normal; if excess kurtosis is less than zero, then kurtosis is less than that for a normal.

The Student's-t Distribution The kurtosis of a random variable gives information on the tail thickness of its distribution. The normal distribution, with kurtosis equal to three, gives a benchmark for the tail thickness of symmetric distributions. A distribution similar to the standard normal distribution but with fatter tails, and hence larger kurtosis, is the Student's t distribution. If X has a Student's t distribution with degrees of freedom parameter v, denoted $X \sim t_v$, then its pdf has the form

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}, \quad -\infty < x < \infty,\ v > 0. \tag{1.21}$$

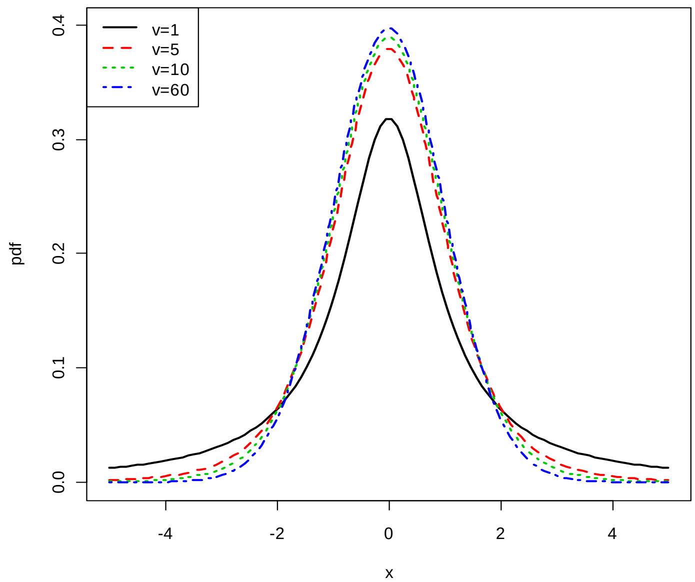

Figure 1.10: Student's t density with $v = 1, 5, 10$ and 60.

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ denotes the gamma function. It can be shown that

$$
\begin{aligned}
E[X] &= 0, \;\; v > 1 \\
\mathrm{var}(X) &= \frac{v}{v-2}, \;\; v > 2, \\
\mathrm{skew}(X) &= 0, \;\; v > 3, \\
\mathrm{kurt}(X) &= \frac{6}{v-4} + 3, \;\; v > 4.
\end{aligned}
$$

The parameter v controls the scale and tail thickness of distribution. If v is close to four, then the kurtosis is large and the tails are thick. If $v < 4$, then $\mathrm{kurt}(X) = \infty$. As $v \to \infty$ the Student's t pdf approaches that of a standard normal random variable and $\mathrm{kurt}(X) = 3$. Figure 1.10 shows plots of the Student's t density for various values of v as well as the standard normal density.

Example 39 *Computing tail probabilities and quantiles from the Student's* *t distribution*

The R functions `pt()` and `qt()` can be used to compute the cdf and quantiles of a Student's t random variable. For $v = 1, 2, 5, 10, 60, 100$ and ∞ the 1% quantiles can be computed using

```
> v = c(1, 2, 5, 10, 60, 100, Inf)
> qt(0.01, df=v)
[1] -31.821  -6.965  -3.365  -2.764  -2.390  -2.364  -2.326
```

For $v = 1, 2, 5, 10, 60, 100$ and ∞ the values of $\Pr(X < -3)$ are

```
> pt(-3, df=v)
[1] 0.102416 0.047733 0.015050 0.006672 0.001964 0.001704 0.001350
```

■

1.1.7 Linear Functions of a Random Variable

Let X be a random variable either discrete or continuous with $E[X] = \mu_X$, $\text{var}(X) = \sigma_X^2$, and let a and b be known constants. Define a new random variable Y via the linear function of X

$$Y = g(X) = aX + b.$$

Then the following results hold:

$$\begin{aligned}
\mu_Y &= E[Y] = a \cdot E[X] + b = a \cdot \mu_X + b, \\
\sigma_Y^2 &= \text{var}(Y) = a^2 \text{var}(X) = a^2 \sigma_X^2, \\
\sigma_Y &= \text{sd}(Y) = a \cdot \text{sd}(X) = a \cdot \sigma_X.
\end{aligned}$$

The first result shows that expectation is a linear operation. That is,

$$E[aX + b] = aE[X] + b.$$

The second result shows that adding a constant to X does not affect its variance, and that the effect of multiplying X by the constant a increases the variance of X by the square of a. These results will be used often enough that it instructive to go through the derivations.

Consider the first result. Let X be a discrete random variable. By the definition of $E[g(X)]$, with $g(X) = b + aX$, we have

$$\begin{aligned} E[Y] &= \sum_{x \in S_X} (ax+b) \cdot \Pr(X=x) \\ &= a \sum_{x \in S_X} x \cdot \Pr(X=x) + b \sum_{x \in S_X} \Pr(X=x) \\ &= a \cdot E[X] + b \cdot 1 \\ &= a \cdot \mu_X + b. \end{aligned}$$

If X is a continuous random variable then by the linearity of integration

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} (ax+b) f(x)dx = a \int_{-\infty}^{\infty} x f(x)dx + b \int_{-\infty}^{\infty} f(x)dx \\ &= aE[X] + b. \end{aligned}$$

Next consider the second result. Since $\mu_Y = a\mu_X + b$ we have

$$\begin{aligned} \text{var}(Y) &= E[(Y - \mu_y)^2] \\ &= E[(aX + b - (a\mu_X + b))^2] \\ &= E[(a(X - \mu_X) + (b - b))^2] \\ &= E[a^2(X - \mu_X)^2] \\ &= a^2 E[(X - \mu_X)^2] \quad \text{(by the linearity of } E[\cdot]) \\ &= a^2 \text{var}(X) \end{aligned}$$

Notice that the derivation of the second result works for discrete and continuous random variables.

A normal random variable has the special property that a linear function of it is also a normal random variable. The following proposition establishes the result.

Proposition 40 *Let $X \sim N(\mu_X, \sigma_X^2)$ and let a and b be constants. Let $Y = aX + b$. Then $Y \sim N(a\mu_X + b, a^2\sigma_X^2)$.*

The above property is special to the normal distribution and may or may not hold for a random variable with a distribution that is not normal.

Example 41 *Standardizing a Random Variable*

Let X be a random variable with $E[X] = \mu_X$ and $\text{var}(X) = \sigma_X^2$. Define a new random variable Z as

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{1}{\sigma_X}X - \frac{\mu_X}{\sigma_X},$$

which is a linear function $aX + b$ where $a = \frac{1}{\sigma_X}$ and $b = -\frac{\mu_X}{\sigma_X}$. This transformation is called *standardizing* the random variable X since

$$\begin{aligned} E[Z] &= \frac{1}{\sigma_X}E[X] - \frac{\mu_X}{\sigma_X} = \frac{1}{\sigma_X}\mu_X - \frac{\mu_X}{\sigma_X} = 0, \\ \text{var}(Z) &= \left(\frac{1}{\sigma_X}\right)^2 \text{var}(X) = \frac{\sigma_X^2}{\sigma_X^2} = 1. \end{aligned}$$

Hence, standardization creates a new random variable with mean zero and variance 1. In addition, if X is normally distributed then $Z \sim N(0,1)$. ■

Example 42 *Computing probabilities using standardized random variables*

Let $X \sim N(2,4)$ and suppose we want to find $\Pr(X > 5)$ but we only know probabilities associated with a standard normal random variable $Z \sim N(0,1)$. We solve the problem by standardizing X as follows:

$$\Pr(X > 5) = \Pr\left(\frac{X-2}{\sqrt{4}} > \frac{5-2}{\sqrt{4}}\right) = \Pr\left(Z > \frac{3}{2}\right) = 0.06681.$$

■

Standardizing a random variable is often done in the construction of test statistics. For example, the so-called *t-statistic* or *t-ratio* used for testing simple hypotheses on coefficients is constructed by the standardization process.

A non-standard random variable X with mean μ_X and variance σ_X^2 can be created from a standard random variable via the linear transformation:

$$X = \mu_X + \sigma_X \cdot Z. \tag{1.22}$$

This result is useful for modeling purposes as illustrated in the next example.

Example 43 *Quantile of general normal random variable*

Let $X \sim N(\mu_X, \sigma_X^2)$. Quantiles of X can be conveniently computed using

$$q_\alpha = \mu_X + \sigma_X z_\alpha \tag{1.23}$$

where $\alpha \in (0,1)$ and z_α is the $\alpha \times 100\%$ quantile of a standard normal random variable. This formula is derived as follows. By the definition z_α and (1.22)

$$\alpha = \Pr(Z \leq z_\alpha) = \Pr\left(\mu_X + \sigma_X \cdot Z \leq \mu_X + \sigma_X \cdot z_\alpha\right) = \Pr(X \leq \mu_X + \sigma_X \cdot z_\alpha),$$

which implies (1.23).

Example 44 *Constant expected return model for asset returns*

Let r denote the monthly continuously compounded return on an asset, and assume that $r \sim N(\mu_r, \sigma_r^2)$. Then r can be expressed as

$$r = \mu_r + \sigma_r \cdot \varepsilon,\ \varepsilon \sim N(0,1).$$

The random variable ε can be interpreted as representing the random news arriving in a given month that makes the observed return differ from its expected value μ. The fact that ε has mean zero means that news, on average, is neutral. The value of σ_r represents the typical size of a news shock. The bigger is σ_r, the larger is the impact of a news shock and vice-versa. ■

1.1.8 Value at Risk: An Introduction

As an example of working with linear functions of a normal random variable, and to illustrate the concept of *Value-at-Risk* (VaR), consider an investment of \$10,000 in Microsoft stock over the next month. Let R denote the monthly simple return on Microsoft stock and assume that $R \sim N(0.05, (0.10)^2)$. That is, $E[R] = \mu_R = 0.05$ and $\mathrm{var}(R) = \sigma_R^2 = (0.10)^2$. Let W_0 denote the investment value at the beginning of the month and W_1 denote the investment value at the end of the month. In this example, $W_0 = \$10,000$. Consider the following questions:

(i) What is the probability distribution of end of month wealth, W_1?

(ii) What is the probability that end of month wealth is less than $\$9,000$, and what must the return on Microsoft be for this to happen?

(iii) What is the loss in dollars that would occur if the return on Microsoft stock is equal to its 5% quantile, $q_{.05}$? That is, what is the monthly 5% VaR on the \$10,000 investment in Microsoft?

To answer (i), note that end of month wealth, W_1, is related to initial wealth W_0 and the return on Microsoft stock R via the linear function

$$W_1 = W_0(1+R) = W_0 + W_0 R = \$10,000 + \$10,000 \cdot R.$$

Using the properties of linear functions of a random variable we have

$$E[W_1] = W_0 + W_0 E[R] = \$10,000 + \$10,000(0.05) = \$10,500,$$

and

$$\begin{aligned} \text{var}(W_1) &= (W_0)^2 \text{var}(R) = (\$10,000)^2(0.10)^2, \\ \text{sd}(W_1) &= (\$10,000)(0.10) = \$1,000. \end{aligned}$$

Further, since R is assumed to be normally distributed it follows that W_1 is normally distributed:

$$W_1 \sim N(\$10,500, (\$1,000)^2).$$

To answer (ii), we use the above normal distribution for W_1 to get

$$\Pr(W_1 < \$9,000) = 0.067.$$

To find the return that produces end of month wealth of \$9,000, or a loss of \$10,000 − \$9,000 = \$1,000, we solve

$$R = \frac{\$9,000 - \$10,000}{\$10,000} = -0.10.$$

If the monthly return on Microsoft is -10% or less, then end of month wealth will be \$9,000 or less. Notice that $R = -0.10$ is the 6.7% quantile of the distribution of R:

$$\Pr(R < -0.10) = 0.067.$$

Question (iii) can be answered in two equivalent ways. First, we use $R \sim N(0.05, (0.10)^2)$ and solve for the 5% quantile of Microsoft Stock using (1.23)[5]:

$$\begin{aligned} &\Pr(R < q_{.05}^R) = 0.05 \\ &\Rightarrow q_{.05}^R = \mu_R + \sigma_R \cdot z_{.05} = 0.05 + 0.10 \cdot (-1.645) = -0.114. \end{aligned}$$

[5] Using R, $z_{.05} = qnorm(0.05) = -1.645$.

That is, with 5% probability the return on Microsoft stock is -11.4% or less. Now, if the return on Microsoft stock is -11.4% the loss in investment value is $\$10,000 \cdot (0.114) = \$1,144$. Hence, $\$1,144$ is the 5% VaR over the next month on the $\$10,000$ investment in Microsoft stock. For the second method, use $W_1 \tilde{} N(\$10,500, (\$1,000)^2)$ and solve for the 5% quantile of end of month wealth directly:

$$\Pr(W_1 < q_{.05}^{W_1}) = 0.05 \Rightarrow q_{.05}^{W_1} = \$8,856.$$

This corresponds to a loss of investment value of $\$10,000 - \$8,856 = \$1,144$. Hence, if W_0 represents the initial wealth and $q_{.05}^{W_1}$ is the 5% quantile of the distribution of W_1 then the 5% VaR is

$$5\% \text{ VaR} = W_0 - q_{.05}^{W_1}.$$

In general if W_0 represents the initial wealth in dollars and q_α^R is the $\alpha \times 100\%$ quantile of distribution of the simple return R, then the $\alpha \times 100\%$ VaR is defined as

$$\text{VaR}_\alpha \ = |W_0 \cdot q_\alpha^R|. \tag{1.24}$$

In words, VaR_α represents the dollar loss that could occur with probability α. By convention, it is reported as a positive number (hence the use of the absolute value function).

Value-at-Risk Calculations for Continuously Compounded Returns

The above calculations illustrate how to calculate value-at-risk using the normal distribution for simple returns. However, as argued in Example 31, the normal distribution may not be appropriate for characterizing the distribution of simple returns and may be more appropriate for characterizing the distribution of continuously compounded returns. Let R denote the simple monthly return, let $r = \ln(1 + R)$ denote the continuously compounded return and assume that $r \sim N(\mu_r, \sigma_r^2)$. The $\alpha \times 100\%$ monthly VaR on an investment of $\$W_0$ is computed using the following steps:

1. Compute the $\alpha \cdot 100\%$ quantile, q_α^r, from the normal distribution for the continuously compounded return r :

$$q_\alpha^r = \mu_r + \sigma_r z_\alpha,$$

where z_α is the $\alpha \cdot 100\%$ quantile of the standard normal distribution.

2. Convert the continuously compounded return quantile, q_α^r, to a simple return quantile using the transformation

$$q_\alpha^R = e^{q_\alpha^r} - 1$$

3. Compute VaR using the simple return quantile (1.24).

Example 45 *Computing VaR from simple and continuously compounded returns using R*

Let R denote the simple monthly return on Microsoft stock and assume that $R \sim N(0.05, (0.10)^2)$. Consider an initial investment of $W_0 = \$10,000$. To compute the 1% and 5% VaR values over the next month use

```
> mu.R = 0.05
> sd.R = 0.10
> w0 = 10000
> q.01.R = mu.R + sd.R*qnorm(0.01)
> q.05.R = mu.R + sd.R*qnorm(0.05)
> VaR.01 = abs(q.01.R*w0)
> VaR.05 = abs(q.05.R*w0)
> VaR.01
[1] 1826
> VaR.05
[1] 1145
```

Hence with 1% and 5% probability the loss over the next month is at least \$1,826 and \$1,145, respectively.

Let r denote the continuously compounded return on Microsoft stock and assume that $r \sim N(0.05, (0.10)^2)$. To compute the 1% and 5% VaR values over the next month use

```
> mu.r = 0.05
> sd.r = 0.10
> q.01.R = exp(mu.r + sd.r*qnorm(0.01)) - 1
> q.05.R = exp(mu.r + sd.r*qnorm(0.05)) - 1
> VaR.01 = abs(q.01.R*w0)
> VaR.05 = abs(q.05.R*w0)
> VaR.01
```

```
[1] 1669
> VaR.05
[1] 1082
```

Notice that when $1+R = e^r$ has a log-normal distribution, the 1% and 5% VaR values (losses) are slightly smaller than when R is normally distributed. This is due to the positive skewness of the log-normal distribution. ■

1.2 Bivariate Distributions

So far we have only considered probability distributions for a single random variable. In many situations we want to be able to characterize the probabilistic behavior of two or more random variables simultaneously. In this section, we discuss bivariate distributions.

1.2.1 Discrete Random Variables

Let X and Y be discrete random variables with sample spaces S_X and S_Y, respectively. The likelihood that X and Y takes values in the joint sample space $S_{XY} = S_X \times S_Y$ is determined by the joint probability distribution $p(x,y) = \Pr(X = x, Y = y)$. The function $p(x,y)$ satisfies

(i) $p(x,y) > 0$ for $x, y \in S_{XY}$;

(ii) $p(x,y) = 0$ for $x, y \notin S_{XY}$;

(iii) $\sum_{x,y \in S_{XY}} p(x,y) = \sum_{x \in S_X} \sum_{y \in S_Y} p(x,y) = 1$.

Example 46 *Bivariate discrete distribution for stock returns*

Let X denote the monthly return (in percent) on Microsoft stock and let Y denote the monthly return on Apple stock. For simplicity suppose that the sample spaces for X and Y are $S_X = \{0,1,2,3\}$ and $S_Y = \{0,1\}$ so that the random variables X and Y are discrete. The joint sample space is the two dimensional grid $S_{XY} = \{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1), (3,0), (3,1)\}$. Table 1.3 illustrates the joint distribution for X and Y. From the table, $p(0,0) = \Pr(X = 0, Y = 0) = 1/8$. Notice that sum of all the entries in the table sum to unity. The bivariate distribution is illustrated graphically in Figure 1.11 as a 3-dimensional barchart. ■

		Y		
	%	0	1	Pr(X)
	0	1/8	0	1/8
X	1	2/8	1/8	3/8
	2	1/8	2/8	3/8
	3	0	1/8	1/8
	Pr(Y)	4/8	4/8	1

Table 1.3: Discrete bivariate distribution for Microsoft and Apple stock prices.

Marginal Distributions

The joint probability distribution tells the probability of X and Y occuring together. What if we only want to know about the probability of X occurring, or the probability of Y occuring?

Example 47 *Find* $\Pr(X = 0)$ *and* $\Pr(Y = 1)$ *from joint distribution*

Consider the joint distribution in Table 1.3. What is $\Pr(X = 0)$ regardless of the value of Y? Now X can occur if $Y = 0$ or if $Y = 1$ and since these two events are mutually exclusive we have that $\Pr(X = 0) = \Pr(X = 0, Y = 0) + \Pr(X = 0, Y = 1) = 0 + 1/8 = 1/8$. Notice that this probability is equal to the horizontal (row) sum of the probabilities in the table at $X = 0$. We can find $\Pr(Y = 1)$ in a similar fashion: $\Pr(Y = 1) = \Pr(X = 0, Y = 1) + \Pr(X = 1, Y = 1) + \Pr(X = 2, Y = 1) + \Pr(X = 3, Y = 1) = 0 + 1/8 + 2/8 + 1/8 = 4/8$. This probability is the vertical (column) sum of the probabilities in the table at $Y = 1$. ■

The probability $\Pr(X = x)$ is called the *marginal probability* of X and is given by

$$\Pr(X = x) = \sum_{y \in S_Y} \Pr(X = x, Y = y). \tag{1.25}$$

Similarly, the marginal probability of $Y = y$ is given by

$$\Pr(Y = y) = \sum_{x \in S_X} \Pr(X = x, Y = y). \tag{1.26}$$

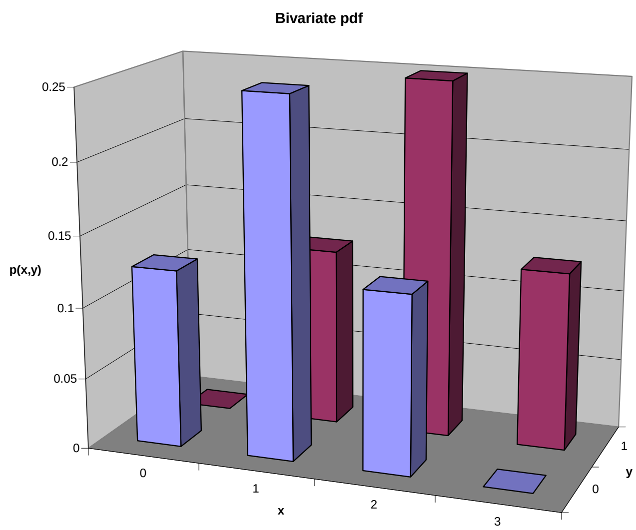

Figure 1.11: Discrete bivariate distribution.

Example 48 *Marginal probabilities of discrete bivariate distribution*

The marginal probabilities of $X = x$ are given in the last column of Table 1.3, and the marginal probabilities of $Y = y$ are given in the last row of Table 1.3. Notice that these probabilities sum to 1. For future reference we note that $E[X] = 3/2, \mathrm{var}(X) = 3/4, E[Y] = 1/2$, and $\mathrm{var}(Y) = 1/4$. ■

Conditional Distributions

For random variables in Table 1.3, suppose we know that the random variable Y takes on the value $Y = 0$. How does this knowledge affect the likelihood that X takes on the values $0, 1, 2$ or 3? For example, what is the probability that $X = 0$ *given that* we know $Y = 0$? To find this probability, we use Bayes' law and compute the *conditional probability*

$$\Pr(X = 0|Y = 0) = \frac{\Pr(X = 0, Y = 0)}{\Pr(Y = 0)} = \frac{1/8}{4/8} = 1/4.$$

The notation $\Pr(X = 0|Y = 0)$ is read as "the probability that $X = 0$ given that $Y = 0$". Notice that $\Pr(X = 0|Y = 0) = 1/4 > \Pr(X = 0) = 1/8$.

Hence, knowledge that $Y = 0$ increases the likelihood that $X = 0$. Clearly, X *depends* on Y.

Now suppose that we know that $X = 0$. How does this knowledge affect the probability that $Y = 0$? To find out we compute

$$\Pr(Y = 0|X = 0) = \frac{\Pr(X = 0, Y = 0)}{\Pr(X = 0)} = \frac{1/8}{1/8} = 1.$$

Notice that $\Pr(Y = 0|X = 0) = 1 > \Pr(Y = 0) = 1/2$. That is, knowledge that $X = 0$ makes it certain that $Y = 0$.

In general, the conditional probability that $X = x$ given that $Y = y$ (provided $\Pr(Y = y) \neq 0$) is

$$\Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}, \tag{1.27}$$

and the conditional probability that $Y = y$ given that $X = x$ (provided $\Pr(X = x) \neq 0$) is

$$\Pr(Y = y|X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}. \tag{1.28}$$

Example 49 *Conditional distributions*

For the bivariate distribution in Table 1.3, the conditional probabilities along with marginal probabilities are summarized in Tables 1.4 and 1.5. Notice that the marginal distribution of X is centered at $x = 3/2$ whereas the conditional distribution of $X|Y = 0$ is centered at $x = 1$ and the conditional distribution of $X|Y = 1$ is centered at $x = 2$. ■

Conditional Expectation and Conditional Variance

Just as we defined shape characteristics of the marginal distributions of X and Y, we can also define shape characteristics of the conditional distributions of $X|Y = y$ and $Y|X = x$. The most important shape characteristics are the *conditional expectation* (*conditional mean*) and the *conditional variance*. The conditional mean of $X|Y = y$ is denoted by $\mu_{X|Y=y} = E[X|Y = y]$, and the

x	$\Pr(X = x)$	$\Pr(X\|Y = 0)$	$\Pr(X\|Y = 1)$
0	1/8	2/8	0
1	3/8	4/8	2/8
2	3/8	2/8	4/8
3	1/8	0	2/8

Table 1.4: Conditional probability distribution of X from bivariate discrete distribution.

y	$\Pr(Y = y)$	$\Pr(Y\|X = 0)$	$\Pr(Y\|X = 1)$	$\Pr(Y\|X = 2)$	$\Pr(Y\|X = 3)$
0	1/2	1	2/3	1/3	0
1	1/2	0	1/3	2/3	1

Table 1.5: Conditional distribution of Y from bivariate discrete distribution.

conditional mean of $Y|X = x$ is denoted by $\mu_{Y|X=x} = E[Y|X = x]$. These means are computed as

$$\mu_{X|Y=y} = E[X|Y = y] = \sum_{x \in S_X} x \cdot \Pr(X = x|Y = y), \tag{1.29}$$

$$\mu_{Y|X=x} = E[Y|X = x] = \sum_{y \in S_Y} y \cdot \Pr(Y = y|X = x). \tag{1.30}$$

Similarly, the conditional variance of $X|Y = y$ is denoted by $\sigma^2_{X|Y=y} = \mathrm{var}(X|Y = y)$ and the conditional variance of $Y|X = x$ is denoted by $\sigma^2_{Y|X=x} = \mathrm{var}(Y|X = x)$. These variances are computed as

$$\sigma^2_{X|Y=y} = \mathrm{var}(X|Y = y) = \sum_{x \in S_X} (x - \mu_{X|Y=y})^2 \cdot \Pr(X = x|Y = y), \tag{1.31}$$

$$\sigma^2_{Y|X=x} = \mathrm{var}(Y|X = x) = \sum_{y \in S_Y} (y - \mu_{Y|X=x})^2 \cdot \Pr(Y = y|X = x). \tag{1.32}$$

Example 50 *Compute conditional expectation and conditional variance*

For the random variables in Table 1.3, we have the following conditional moments for X:

$$\begin{aligned}
E[X|Y &= 0] = 0 \cdot 1/4 + 1 \cdot 1/2 + 2 \cdot 1/4 + 3 \cdot 0 = 1, \\
E[X|Y &= 1] = 0 \cdot 0 + 1 \cdot 1/4 + 2 \cdot 1/2 + 3 \cdot 1/4 = 2, \\
\mathrm{var}(X|Y &= 0) = (0-1)^2 \cdot 1/4 + (1-1)^2 \cdot 1/2 + (2-1)^2 \cdot 1/2 + (3-1)^2 \cdot 0 = 1/2, \\
\mathrm{var}(X|Y &= 1) = (0-2)^2 \cdot 0 + (1-2)^2 \cdot 1/4 + (2-2)^2 \cdot 1/2 + (3-2)^2 \cdot 1/4 = 1/2.
\end{aligned}$$

Compare these values to $E[X] = 3/2$ and $\mathrm{var}(X) = 3/4$. Notice that as y increases, $E[X|Y = y]$ increases.

For Y, similar calculations gives

$$\begin{aligned}
E[Y|X &= 0] = 0, E[Y|X = 1] = 1/3, E[Y|X = 2] = 2/3, E[Y|X = 3] = 1, \\
\mathrm{var}(Y|X &= 0) = 0, \mathrm{var}(Y|X = 1) = 0.2222, \mathrm{var}(Y|X = 2) = 0.2222, \mathrm{var}(Y|X = 3) = 0.
\end{aligned}$$

Compare these values to $E[Y] = 1/2$ and $\mathrm{var}(Y) = 1/4$. Notice that as x increases $E[Y|X = x]$ increases. ■

Conditional Expectation and the Regression Function

Consider the problem of predicting the value Y given that we know $X = x$. A natural predictor to use is the conditional expectation $E[Y|X = x]$. In this prediction context, the conditional expectation $E[Y|X = x]$ is called the *regression function*. The graph with $E[Y|X = x]$ on the vertical axis and x on the horizontal axis gives the *regression line*. The relationship between Y and the regression function may expressed using the trivial identity

$$\begin{aligned}
Y &= E[Y|X = x] + Y - E[Y|X = x] \\
&= E[Y|X = x] + \varepsilon, \qquad (1.33)
\end{aligned}$$

where $\varepsilon = Y - E[Y|X]$ is called the *regression error*.

Example 51 *Regression line for bivariate discrete distribution*

For the random variables in Table 1.3, the regression line is plotted in Figure 1.12. Notice that there is a linear relationship between $E[Y|X = x]$ and x. When such a linear relationship exists we call the regression function a *linear regression*. Linearity of the regression function, however, is not guaranteed. It may be the case that there is a non-linear (e.g., quadratic) relationship between $E[Y|X = x]$ and x. In this case, we call the regression function a *non-linear regression*.

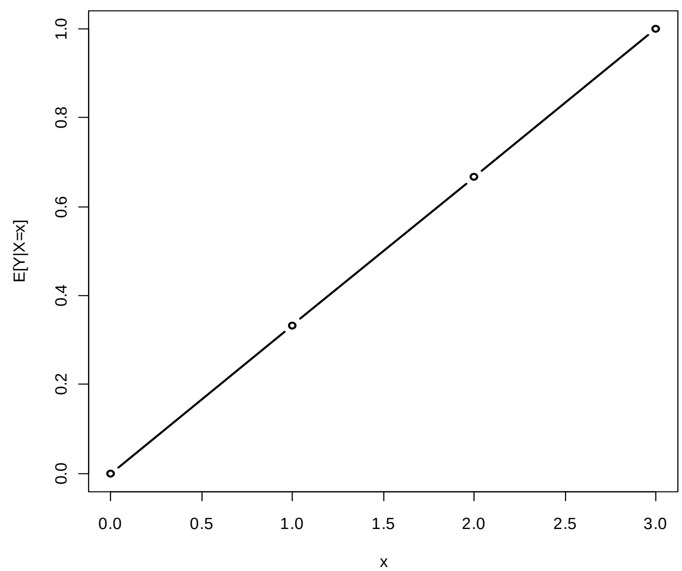

Figure 1.12: Regression function $E[Y|X = x]$ from discrete bivariate distribution.

Law of Total Expectations

For the random variables in Table 1.3, notice that

$$\begin{aligned} E[X] &= E[X|Y=0] \cdot \Pr(Y=0) + E[X|Y=1] \cdot \Pr(Y=1) \\ &= 1 \cdot 1/2 + 2 \cdot 1/2 = 3/2, \end{aligned}$$

and

$$\begin{aligned} E[Y] &= E[Y|X=0] \cdot \Pr(X=0) + E[Y|X=1] \cdot \Pr(X=1) \\ +E[Y|X &= 2] \cdot \Pr(X=2) + E[Y|X=3] \cdot \Pr(X=3) = 1/2. \end{aligned}$$

This result is known as the *law of total expectations*. In general, for two random variables X and Y (discrete or continuous) we have

$$\begin{aligned} E[X] &= E[E[X|Y]], \qquad (1.34) \\ E[Y] &= E[E[Y|X]], \end{aligned}$$

where the first expectation is taken with respect to Y and the second expectation is taken with respect to X.

1.2.2 Bivariate Distributions for Continuous Random Variables

Let X and Y be continuous random variables defined over the real line. We characterize the joint probability distribution of X and Y using the joint probability function (pdf) $f(x,y)$ such that $f(x,y) \geq 0$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$$

The three-dimensional plot of the joint probability distribution gives a probability surface whose total volume is unity. To compute joint probabilities of $x_1 \leq X \leq x_2$ and $y_1 \leq Y \leq y_2$, we need to find the volume under the probability surface over the grid where the intervals $[x_1, x_2]$ and $[y_1, y_2]$ overlap. Finding this volume requries solving the double integral

$$\Pr(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x,y)dxdy.$$

Example 52 *Bivariate standard normal distribution*

A standard bivariate normal pdf for X and Y has the form

$$f(x,y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}, -\infty \leq x, y \leq \infty \tag{1.35}$$

and has the shape of a symmetric bell (think Liberty Bell) centered at $x = 0$ and $y = 0$. To find $\Pr(-1 < X < 1, -1 < Y < 1)$ we must solve

$$\int_{-1}^{1}\int_{-1}^{1} \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}dxdy$$

which, unfortunately, does not have an analytical solution. Numerical approximation methods are required to evaluate the above integral. The function `pmvnorm()` in the R package mvtnorm can be used to evaluate areas under the bivariate standard normal surface. To compute $\Pr(-1 < X < 1, -1 < Y < 1)$ use

```
> library(mvtnorm)
> pmvnorm(lower=c(-1, -1), upper=c(1, 1))
[1] 0.4661
attr(,"error")
[1] 1e-15
attr(,"msg")
[1] "Normal Completion"
```

Here, $\Pr(-1 < X < 1, -1 < Y < 1) = 0.4661$. The attribute `error` gives the estimated absolute error of the approximation, and the attribute `message` tells the status of the algorithm used for the approximation. See the online help for `pmvnorm` for more details. ■

Marginal and Conditional Distributions

The marginal pdf of X is found by integrating y out of the joint pdf $f(x, y)$ and the marginal pdf of Y is found by integrating x out of the joint pdf:

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy, \tag{1.36}$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)dx. \tag{1.37}$$

The conditional pdf of X given that $Y = y$, denoted $f(x|y)$, is computed as

$$f(x|y) = \frac{f(x, y)}{f(y)}, \tag{1.38}$$

and the conditional pdf of Y given that $X = x$ is computed as

$$f(y|x) = \frac{f(x, y)}{f(x)}. \tag{1.39}$$

The conditional means are computed as

$$\mu_{X|Y=y} = E[X|Y = y] = \int x \cdot p(x|y)dx, \tag{1.40}$$

$$\mu_{Y|X=x} = E[Y|X = x] = \int y \cdot p(y|x)dy \tag{1.41}$$

and the conditional variances are computed as

$$\sigma^2_{X|Y=y} = \text{var}(X|Y=y) = \int (x - \mu_{X|Y=y})^2 p(x|y)dx, \tag{1.42}$$

$$\sigma^2_{Y|X=x} = \text{var}(Y|X=x) = \int (y - \mu_{Y|X=x})^2 p(y|x)dy. \tag{1.43}$$

Example 53 *Conditional and marginal distributions from bivariate standard normal*

Suppose X and Y are distributed bivariate standard normal. To find the marginal distribution of X we use (1.36) and solve

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Hence, the marginal distribution of X is standard normal. Similiar calculations show that the marginal distribution of Y is also standard normal. To find the conditional distribution of $X|Y=y$ we use (1.38) and solve

$$\begin{aligned} f(x|y) &= \frac{f(x,y)}{f(x)} = \frac{\frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2+y^2)+\frac{1}{2}y^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \\ &= f(x) \end{aligned}$$

So, for the standard bivariate normal distribution $f(x|y) = f(x)$ which does not depend on y. Similar calculations show that $f(y|x) = f(y)$. ■

1.2.3 Independence

Let X and Y be two discrete random variables. Intuitively, X is independent of Y if knowledge about Y does not influence the likelihood that $X = x$ for all possible values of $x \in S_X$ and $y \in S_Y$. Similarly, Y is independent of X if knowledge about X does not influence the likelihood that $Y = y$ for all values of $y \in S_Y$. We represent this intuition formally for discrete random variables as follows.

Definition 54 *Let X and Y be discrete random variables with sample spaces S_X and S_Y, respectively. X and Y are independent random variables iff*

$$\begin{aligned}\Pr(X &= x|Y=y) = \Pr(X=x),\ \textit{for all } x \in S_X, y \in S_Y \\ \Pr(Y &= y|X=x) = \Pr(Y=y),\ \textit{for all } x \in S_X, y \in S_Y\end{aligned}$$

Example 55 *Check independence of bivariate discrete random variables*

For the data in Table 1.11, we know that $\Pr(X=0|Y=0) = 1/4 \neq \Pr(X=0) = 1/8$ so X and Y are not independent.

Proposition 56 *Let X and Y be discrete random variables with sample spaces S_X and S_Y, respectively. X and Y are independent if and only if (iff)*

$$\Pr(X=x, Y=y) = \Pr(X=x) \cdot \Pr(Y=y),\ \textit{for all } x \in S_X, y \in S_Y$$

Intuition for the above result follows from

$$\begin{aligned}\Pr(X &= x|Y=y) = \frac{\Pr(X=x, Y=y)}{\Pr(Y=y)} = \frac{\Pr(X=x)\cdot\Pr(Y=y)}{\Pr(Y=y)} = \Pr(X=x), \\ \Pr(Y &= y|X=x) = \frac{\Pr(X=x, Y=y)}{\Pr(X=x)} = \frac{\Pr(X=x)\cdot\Pr(Y=y)}{\Pr(X=x)} = \Pr(Y=y)\end{aligned}$$

which shows that X and Y are independent.

For continuous random variables, we have the following definition of independence.

Definition 57 *Let X and Y be continuous random variables. X and Y are independent iff*

$$\begin{aligned}f(x|y) &= f(x),\ \textit{for } -\infty < x, y < \infty, \\ f(y|x) &= f(y),\ \textit{for } -\infty < x, y < \infty.\end{aligned}$$

As with discrete random variables, we have the following result for continuous random variables.

Proposition 58 *Let X and Y be continuous random variables . X and Y are independent iff*

$$f(x,y) = f(x)f(y)$$

The result in the above proposition is extremely useful in practice because it gives us an easy way to compute the joint pdf for two independent random variables: we simple compute the product of the marginal distributions.

Example 59 *Constructing the bivariate standard normal distribution*

Let $X \sim N(0,1)$, $Y \sim N(0,1)$ and let X and Y be independent. Then

$$f(x,y) = f(x)f(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}.$$

This result is a special case of the bivariate normal distribution[6].

A useful property of the independence between two random variables is the following.

Result: If X and Y are independent then $g(X)$ and $h(Y)$ are independent for any functions $g(\cdot)$ and $h(\cdot)$.

For example, if X and Y are independent then X^2 and Y^2 are also independent.

1.2.4 Covariance and Correlation

Let X and Y be two discrete random variables. Figure 1.13 displays several bivariate probability scatterplots (where equal probabilities are given on the dots). In panel (a) we see no linear relationship between X and Y. In panel (b) we see a perfect positive linear relationship between X and Y and in panel (c) we see a perfect negative linear relationship. In panel (d) we see a positive, but not perfect, linear relationship; in panel (e) we see a negative, but not perfect, linear relationship. Finally, in panel (f) we see no systematic linear relationship but we see a strong nonlinear (parabolic) relationship. The *covariance* between X and Y measures the *direction* of linear relationship between the two random variables. The *correlation* between X and Y measures the *direction* and *strength* of linear relationship between the two random variables.

Let X and Y be two random variables with $E[X] = \mu_X$, $\mathrm{var}(X) = \sigma_X^2$, $E[Y] = \mu_Y$ and $\mathrm{var}(Y) = \sigma_Y^2$.

[6] stuff to add: if X and Y are independent then $f(X)$ and $g(Y)$ are independent for any functions $f(\cdot)$ and $g(\cdot)$.

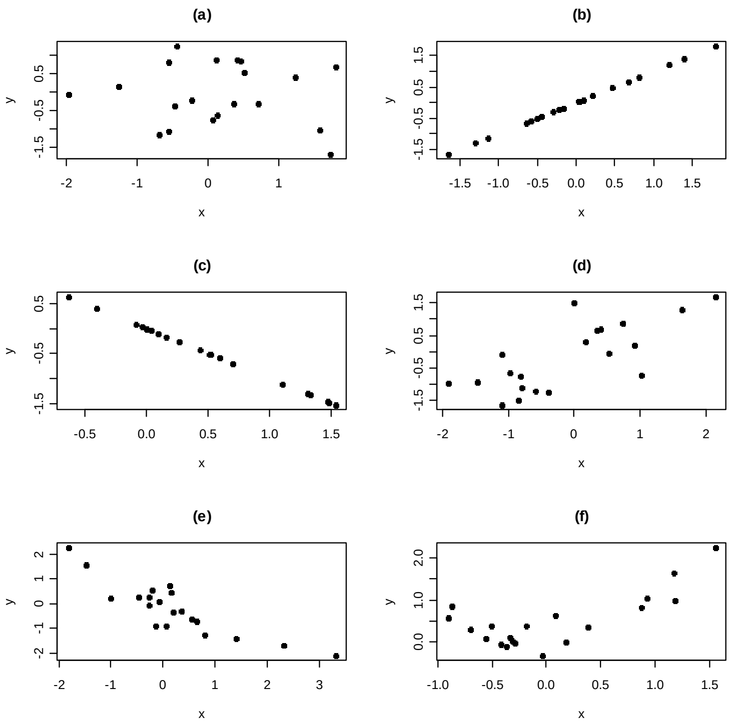

Figure 1.13: Probability scatterplots illustrating dependence between X and Y.

Definition 60 *The covariance between two random variables X and Y is given by*

$$\begin{aligned}\sigma_{XY} &= \mathrm{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \\ &= \sum_{x\in S_X}\sum_{y\in S_Y}(x-\mu_X)(y-\mu_Y)\Pr(X=x,Y=y) \quad \textit{for discrete } X \textit{ and } Y, \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)p(x,y)dxdy \quad \textit{for continuous } X \textit{ and } Y.\end{aligned}$$

Definition 61 *The correlation between two random variables X and Y is given by*

$$\rho_{XY} = \mathrm{cor}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

The correlation coefficient, ρ_{XY}, is a scaled version of the covariance.

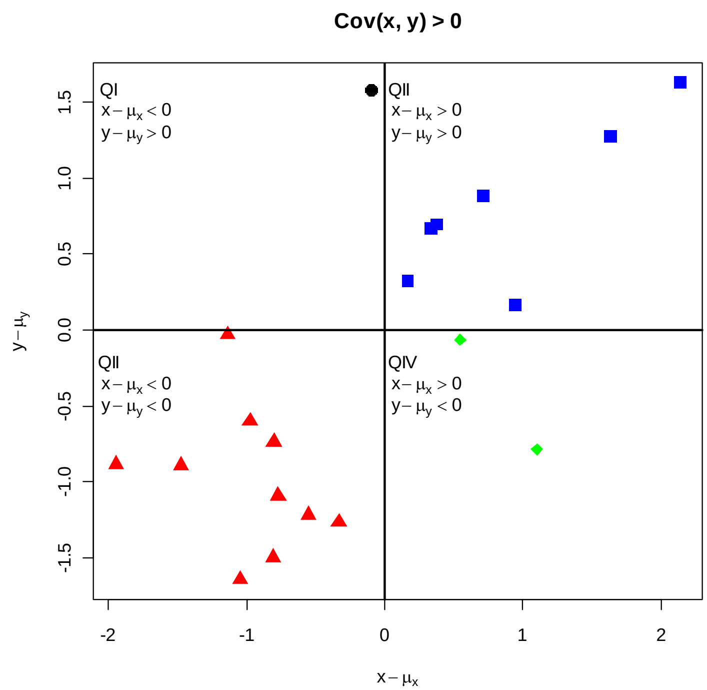

Figure 1.14: Probability scatterplot of discrete distribution with positive covariance. Each pair (X, Y) occurs with equal probability.

To see how covariance measures the direction of linear association, consider the probability scatterplot in Figure 1.14. In the figure, each pair of points occurs with equal probability. The plot is separated into quadrants (right to left, top to bottom). In the first quadrant (black circles), the realized values satisfy $x < \mu_X, y > \mu_Y$ so that the product $(x-\mu_X)(y-\mu_Y) < 0$. In the second quadrant (blue squares), the values satisfy $x > \mu_X$ and $y > \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) > 0$. In the third quadrant (red triangles), the values satisfy $x < \mu_X$ and $y < \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) > 0$. Finally, in the fourth quadrant (green diamonds), $x > \mu_X$ but $y < \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) < 0$. Covariance is then a probability weighted average all of the product terms in the four quadrants. For the values in Figure 1.14, this weighted average is positive because most of the values are in the second and third quadrants.

Example 62 *Calculate covariance and correlation for discrete random variables*

For the data in Table 1.3, we have

$$\begin{aligned}\sigma_{XY} = \mathrm{cov}(X,Y) &= (0-3/2)(0-1/2)\cdot 1/8 + (0-3/2)(1-1/2)\cdot 0 \\ &\quad + \cdots + (3-3/2)(1-1/2)\cdot 1/8 = 1/4 \\ \rho_{XY} = \mathrm{cor}(X,Y) &= \frac{1/4}{\sqrt{(3/4)\cdot(1/2)}} = 0.577\end{aligned}$$

Properties of Covariance and Correlation

Let X and Y be random variables and let a and b be constants. Some important properties of $\mathrm{cov}(X,Y)$ are

1. $\mathrm{cov}(X,X) = \mathrm{var}(X)$
2. $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$
3. $\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y] = E[XY] - \mu_X\mu_Y$
4. $\mathrm{cov}(aX,bY) = a\cdot b\cdot \mathrm{cov}(X,Y)$
5. If X and Y are independent then $\mathrm{cov}(X,Y) = 0$ (no association $\Longrightarrow$ no linear association). However, if $\mathrm{cov}(X,Y) = 0$ then X and Y are not necessarily independent (no linear association $\not\Rightarrow$ no association).
6. If X and Y are jointly normally distributed and $\mathrm{cov}(X,Y) = 0$, then X and Y are independent.

The first two properties are intuitive. The third property results from expanding the definition of covariance:

$$\begin{aligned}\mathrm{cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E\left[XY - X\mu_Y - \mu_X Y + \mu_X\mu_Y\right] \\ &= E[XY] - E[X]\mu_Y - \mu_X E[Y] + \mu_X\mu_Y \\ &= E[XY] - 2\mu_X\mu_Y + \mu_X\mu_Y \\ &= E[XY] - \mu_X\mu_Y\end{aligned}$$

The fourth property follows from the linearity of expectations

$$\begin{aligned}\mathrm{cov}(aX, bY) &= E[(aX - a\mu_X)(bY - b\mu_Y)] \\ &= a \cdot b \cdot E[(X - \mu_X)(Y - \mu_Y)] \\ &= a \cdot b \cdot \mathrm{cov}(X, Y)\end{aligned}$$

The fourth property shows that the value of $\mathrm{cov}(X, Y)$ depends on the scaling of the random variables X and Y. By simply changing the scale of X or Y we can make $\mathrm{cov}(X, Y)$ equal to any value that we want. Consequently, the numerical value of $\mathrm{cov}(X, Y)$ is not informative about the strength of the linear association between X and Y. However, the sign of $\mathrm{cov}(X, Y)$ is informative about the direction of linear association between X and Y. The fifth property should be intuitive. Independence between the random variables X and Y means that there is no relationship, linear or nonlinear, between X and Y. However, the lack of a linear relationship between X and Y does not preclude a nonlinear relationship. The last result illustrates an important property of the normal distribution: lack of covariance implies independence.

Some important properties of $\mathrm{cor}(X, Y)$ are

1. $-1 \leq \rho_{XY} \leq 1$.

2. If $\rho_{XY} = 1$ then X and Y are perfectly positively linearly related. That is, $Y = aX + b$ where $a > 0$.

3. If $\rho_{XY} = -1$ then X and Y are perfectly negatively linearly related. That is, $Y = aX + b$ where $a < 0$.

4. If $\rho_{XY} = 0$ then X and Y are not linearly related but may be nonlinearly related.

5. $\mathrm{cor}(aX, bY) = \mathrm{cor}(X, Y)$ if $a > 0$ and $b > 0$; $\mathrm{cor}(X, Y) = -\mathrm{cor}(X, Y)$ if $a > 0, b < 0$ or $a < 0, b > 0$.

Bivariate normal distribution

Let X and Y be distributed bivariate normal. The joint pdf is given by

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \times \tag{1.44}$$
$$\exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-\frac{2\rho_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$$

where $E[X]=\mu_X$, $E[Y]=\mu_Y$, $\mathrm{sd}(X)=\sigma_X$, $\mathrm{sd}(Y)=\sigma_Y$, and $\rho_{XY}=\mathrm{cor}(X,Y)$. The correlation coefficient ρ_{XY} describes the dependence between X and Y. If $\rho_{XY}=0$ then the pdf collapses to the pdf of the standard bivariate normal distribution.

It can be shown that the marginal distributions of X and Y are normal: $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$. In addition, it can be shown that the conditional distributions $f(x|y)$ and $f(y|x)$ are also normal with means given by

$$\mu_{X|Y=y} = \alpha_X+\beta_X\cdot y, \tag{1.45}$$
$$\mu_{Y|X=x} = \alpha_Y+\beta_Y\cdot x, \tag{1.46}$$

where

$$\begin{aligned}\alpha_X &= \mu_X-\beta_X\mu_Y,\ \beta_X=\sigma_{XY}/\sigma_Y^2,\\ \alpha_Y &= \mu_Y-\beta_Y\mu_X,\ \beta_Y=\sigma_{XY}/\sigma_X^2,\end{aligned}$$

and variances given by

$$\begin{aligned}\sigma_{X|Y=y}^2 &= \sigma_X^2-\sigma_{XY}^2/\sigma_Y^2,\\ \sigma_{X|Y=y}^2 &= \sigma_Y^2-\sigma_{XY}^2/\sigma_X^2.\end{aligned}$$

Notice that the conditional means (regression functions) (1.29) and (1.30) are linear functions of x and y, respectively.

Example 63 *Expressing the bivariate normal distribution using matrix algebra*

The formula for the bivariate normal distribution (1.44) is a bit messy. We can greatly simplify the formula by using matrix algebra. Define the 2×1

vectors $\mathbf{x} = (x, y)'$ and $\boldsymbol{\mu} = (\mu_X, \mu_y)'$, and the 2×2 matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$$

Then the bivariate normal distribution (1.44) may be compactly expressed as

$$f(\mathbf{x}) = \frac{1}{2\pi \det(\mathbf{\Sigma})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

where

$$\det(\mathbf{\Sigma}) = \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 = \sigma_X^2\sigma_Y^2 - \sigma_X^2\sigma_Y^2\rho_{XY}^2 = \sigma_X^2\sigma_Y^2(1 - \rho_{XY}^2).$$

Example 64 *Plotting the bivariate normal distribution*

The R package mvtnorm contains the functions `dmvnorm()`, `pmvnorm()`, and `qmvnorm()` which can be used to compute the bivariate normal pdf, cdf and quantiles, respectively. Plotting the bivariate normal distribution over a specified grid of x and y values in R can be done with the `persp()` function. First, we specify the parameter values for the joint distribution. Here, we choose $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ and $\rho = 0.5$. We will use the `dmvnorm()` function to evaluate the joint pdf at these values. To do so, we must specify a covariance matrix of the form

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

In R this matrix can be created using

```
> sigma = matrix(c(1, 0.5, 0.5, 1), 2, 2)
```

Next we specify a grid of x and y values between -3 and 3:

```
> x = seq(-3, 3, length=100)
> y = seq(-3, 3, length=100)
```

To evaluate the joint pdf over the two-dimensional grid we can use the `outer()` function:

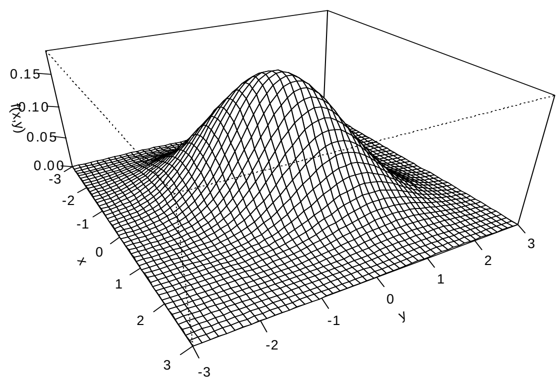

Figure 1.15: Bivariate normal pdf with $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ and $\rho = 0.5$.

```
# function to evaluate bivariate normal pdf on grid of x & y values
> bv.norm <- function(x, y, sigma) {
+   z = cbind(x,y)
+   return(dmvnorm(z, sigma=sigma))
+ }
# use outer function to evaluate pdf on 2D grid of x-y values
> fxy = outer(x, y, bv.norm, sigma)
```

To create the 3D plot of the joint pdf, use the `persp()` function:

```
> persp(x, y, fxy, theta=60, phi=30, expand=0.5, ticktype="detailed",
+       zlab="f(x,y)")
```

The resulting plot is given in Figure 1.15.

Expectation and variance of the sum of two random variables

Let X and Y be two random variables with well defined means, variances and covariance and let a and b be constants. Then the following results hold:

$$\begin{aligned}
E[aX+bY] &= aE[X]+bE[Y] = a\mu_X + b\mu_Y \\
\mathrm{var}(aX+bY) &= a^2\mathrm{var}(X)+b^2\mathrm{var}(Y)+2\cdot a\cdot b\cdot \mathrm{cov}(X,Y) \\
&= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2\cdot a\cdot b\cdot \sigma_{XY}
\end{aligned}$$

The first result states that the expected value of a linear combination of two random variables is equal to a linear combination of the expected values of the random variables. This result indicates that the expectation operator is a linear operator. In other words, expectation is additive. The second result states that variance of a linear combination of random variables is not a linear combination of the variances of the random variables. In particular, notice that covariance comes up as a term when computing the variance of the sum of two (not independent) random variables. Hence, the variance operator is not, in general, a linear operator. That is, variance, in general, is not additive.

It is instructive to go through the derivation of these results. Let X and Y be discrete random variables. Then,

$$\begin{aligned}
E[aX+bY] &= \sum_{x\in S_X}\sum_{y\in S_y}(ax+by)\Pr(X=x,Y=y) \\
&= \sum_{x\in S_X}\sum_{y\in S_y} ax\Pr(X=x,Y=y) + \sum_{x\in S_X}\sum_{y\in S_y} bx\Pr(X=x,Y=y) \\
&= a\sum_{x\in S_X} x\sum_{y\in S_y}\Pr(X=x,Y=y) + b\sum_{y\in S_y} y\sum_{x\in S_X}\Pr(X=x,Y=y) \\
&= a\sum_{x\in S_X} x\Pr(X=x) + b\sum_{y\in S_y} y\Pr(Y=y) \\
&= aE[X]+bE[Y] = a\mu_X + b\mu_Y.
\end{aligned}$$

The result for continuous random variables is similar. Effectively, the summations are replaced by integrals and the joint probabilities are replaced by the joint pdf. Next, let X and Y be discrete or continuous random variables.

Then

$$\begin{aligned}
\mathrm{var}(aX+bY) &= E[(aX+bY-E[aX+bY])^2] \\
&= E[(aX+bY-a\mu_X-b\mu_Y)^2] \\
&= E[(a(X-\mu_X)+b(Y-\mu_Y))^2] \\
&= a^2E[(X-\mu_X)^2]+b^2E[(Y-\mu_Y)^2]+2\cdot a\cdot b\cdot E[(X-\mu_X)(Y-\mu_Y)] \\
&= a^2\mathrm{var}(X)+b^2\mathrm{var}(Y)+2\cdot a\cdot b\cdot \mathrm{cov}(X,Y).
\end{aligned}$$

Linear Combination of two Normal random variables

The following proposition gives an important result concerning a linear combination of normal random variables.

Proposition 65 *Let* $X \sim N(\mu_X, \sigma_X^2)$*,* $Y \sim N(\mu_Y, \sigma_Y^2)$*,* $\sigma_{XY} = \mathrm{cov}(X,Y)$ *and* a *and* b *be constants. Define the new random variable* Z *as*

$$Z = aX + bY.$$

Then

$$Z \sim N(\mu_Z, \sigma_Z^2)$$

where $\mu_Z = a\mu_X + b\mu_Y$ *and* $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$*.*

This important result states that a linear combination of two normally distributed random variables is itself a normally distributed random variable. The proof of the result relies on the change of variables theorem from calculus and is omitted. Not all random variables have the nice property that their distributions are closed under addition.

Example 66 *Portfolio of two assets*

Consider a portfolio of two stocks A (Amazon) and B (Boeing) with investment shares x_A and x_B with $x_A + x_B = 1$. Let R_A and R_B denote the simple monthly returns on these assets, and assume that $R_A \sim N(\mu_A, \sigma_A^2)$ and $R_B \sim N(\mu_B, \sigma_B^2)$. Furthermore, let $\sigma_{AB} = \rho_{AB}\sigma_A\sigma_B = \mathrm{cov}(R_A, R_B)$. The portfolio return is $R_p = x_A R_A + x_B R_B$, which is a linear function of two random variables. Using the properties of linear functions of random variables, we have

$$\begin{aligned}
\mu_p &= E[R_p] = x_A E[R_A] + x_B E[R_B] = x_A\mu_A + x_B\mu_B \\
\sigma_p^2 &= \mathrm{var}(R_p) = x_A^2\mathrm{var}(R_A) + x_B^2\mathrm{var}(R_B) + 2x_Ax_B\mathrm{cov}(R_A,R_B) \\
&= x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + 2x_Ax_B\sigma_{AB}
\end{aligned}$$

■

1.3 Multivariate Distributions

Multivariate distributions are used to characterize the joint distribution of a collection of N random variables $X_1, X_2, \ldots, X_N$ for $N > 1$. The mathematical formulation of this joint distribution can be quite complex and typically makes use of matrix algebra. Here, we summarize some basic properties of multivariate distributions without the use of matrix algebra. In Chapter xxx, we show how matrix algebra greatly simplifies the description of multivariate distribution.

1.3.1 Discrete Random Variables

Let $X_1, X_2, \ldots, X_N$ be N discrete random variables with sample spaces $S_{X_1}, S_{X_2}, \ldots, S_{X_N}$. The likelihood that these random variables take values in the joint sample space $S_{X_1} \times S_{X_2} \times \cdots \times S_{X_N}$ is given by the joint probability function

$$p(x_1, x_2, \ldots, x_N) = \Pr(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N).$$

For $N > 2$ it is not easy to represent the joint probabilities in a table like Table 1.3 or to visualize the distribution as in Figure 1.11.

Marginal distributions for each variable X_i can be derived from the joint distribution as in (1.25) by summing the joint probabilities over the other variables $j \neq i$. For example,

$$p(x_1) = \sum_{x_2 \in S_{X_2}, \ldots, x_N \in S_{X_N}} p(x_1, x_2, \ldots, x_N).$$

With N random variables, there are numerous conditional distributions that can be formed. For example, the distribution of X_1 given $X_2 = x_2, \ldots, X_N = x_N$ is determined using

$$\Pr(X_1 = x_1 | X_2 = x_2, \ldots, X_N = x_N) = \frac{\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N)}{\Pr(X_2 = x_2, \ldots, X_N = x_N)}.$$

Similarly, the joint distribution of X_1 and X_2 given $X_3 = x_3, \ldots, X_N = x_N$ is given by

$$\Pr(X_1 = x_1, X_2 = x_2 | X_3 = x_3, \ldots, X_N = x_N) = \frac{\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N)}{\Pr(X_3 = x_3, \ldots, X_N = x_N)}.$$

1.3.2 Continuous Random Variables

Let $X_1, X_2, \ldots, X_N$ be N continuous random variables each taking values on the real line. The joint pdf is a function $f(x_1, x_2, \ldots, x_N) \geq 0$ such that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_N) dx_1 dx_2 \cdots dx_N = 1.$$

Joint probabilities of $x_{11} \leq X_1 \leq x_{12}$, $x_{21} \leq X_2 \leq x_{22}, \ldots, x_{N1} \leq X_N \leq x_{N2}$ are computed by solving the integral equation

$$\int_{x_{11}}^{x_{12}} \int_{x_{21}}^{x_{22}} \cdots \int_{x_{N1}}^{x_{N2}} f(x_1, x_2, \ldots, x_N) dx_1 dx_2 \cdots dx_N. \tag{1.47}$$

For most multivariate distribution (1.47) cannot be solved analytically and must be approximated numerically.

The marginal pdf for x_i is found by integrating the joint pdf with respect to the other variables. For example, the marginal pdf for x_1 is found by solving

$$f(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_N) dx_2 \cdots dx_N.$$

Conditional pdf for a single random variable or a collection of random variables are defined in the obvious way.

1.3.3 Independence

A collection of N random variables are independent if their joint distribution factors into the product of all of the marginal distributions:

$$\begin{aligned} p(x_1, x_2, \ldots, x_N) &= p(x_1)p(x_2)\cdots p(x_N) \text{ for } X_i \text{ discrete,} \\ f(x_1, x_2, \ldots, x_N) &= f(x_1)f(x_2)\cdots f(x_N) \text{ for } X_i \text{ continuous.} \end{aligned}$$

In addition, if N random variables are independent then any functions of these random variables are also independent.

1.3.4 Dependence Concepts

In general it is difficult to define dependence concepts for collections of more than two random variables. Dependence is typically only defined between pairwise random variables. Hence, covariance and correlation are also useful concepts when dealing with more than two random variables.

For N random variables $X_1, X_2, \ldots, X_N$, with mean values $\mu_i = E[X_i]$ and variances $\sigma_i^2 = \text{var}(X_i)$, the pairwise covariances and correlations are defined as

$$\begin{aligned}\text{cov}(X_i, X_j) &= \sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_i)], \\ \text{cov}(X_i, X_j) &= \rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j},\end{aligned}$$

for $i \neq j$. There are $N(N-1)/2$ pairwise covariances and correlations. Often, these values are summarized using matrix algebra in an $N \times N$ covariance matrix $\mathbf{\Sigma}$ and an $N \times N$ correlation matrix $\mathbf{R}$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{pmatrix}, \tag{1.48}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{12} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & \cdots & 1 \end{pmatrix}. \tag{1.49}$$

1.3.5 Linear Combinations of N Random Variables

Many of the results for manipulating a collection of random variables generalize in a straightforward way to the case of more than two random variables. The details of the generalizations are not important for our purposes. However, the following results will be used repeatedly throughout the book.

Let $X_1, X_2, \ldots, X_N$ denote a collection of N random variables (discrete or continuous) with means μ_i,variances σ_i^2 and covariances σ_{ij}. Define the new random variable Z as a linear combination

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_N X_N$$

where $a_1, a_2, \ldots, a_N$ are constants. Then the following results hold

$$\begin{aligned}\mu_Z &= E[Z] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_N E[X_N] \\ &= \sum_{i=1}^{N} a_i E[X_i] = \sum_{i=1}^{N} a_i \mu_i.\end{aligned} \tag{1.50}$$

$$\begin{aligned}\sigma_Z^2 &= \mathrm{var}(Z) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_N^2\sigma_N^2 \\ &\quad +2a_1a_2\sigma_{12} + 2a_1a_3\sigma_{13} + \cdots + a_1a_N\sigma_{1N} \\ &\quad +2a_2a_3\sigma_{23} + 2a_2a_4\sigma_{24} + \cdots + a_2a_N\sigma_{2N} \\ &\quad +\cdots+ \\ &\quad +2a_{N-1}a_N\sigma_{(N-1)N}.\end{aligned} \tag{1.51}$$

The derivation of these results is very similar to the bivariate case and so is omitted. In addition, if all of the X_i are normally distributed then Z is also normally distributed with mean μ_Z and variance σ_Z^2 as described above.

The variance of a linear combination of N random variables contains N variance terms and $N(N-1)$ covariance terms. For $N = 2, 5, 10$ and 100 the number of covariance terms in $\mathrm{var}(Z)$ is $2, 20, 90$ and 9900, respectively. Notice that when N is large there are many more covariance terms than variance terms in $\mathrm{var}(Z)$.

The expression for $\mathrm{var}(Z)$ is messy. It can be simplified using matrix algebra notation, as explained in detail in Chapter xxx. To preview, define the $N \times 1$ vectors $\mathbf{X} = (X_1, \ldots, X_N)'$ and $\mathbf{a} = (a_1, \ldots, a_N)'$.Then $Z = \mathbf{a}'\mathbf{X}$ and $\mathrm{var}(Z) = \mathrm{var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\mathbf{\Sigma}\mathbf{a}$, where Σ is the $N \times N$ covariance matrix, which is much more compact than (1.51).

Example 67 *Square-root-of-time rule for multi-period continuously compounded returns*

Let r_t denote the continuously compounded monthly return on an asset at times $t = 1, \ldots, 12$. Assume that $r_1, \ldots, r_{12}$ are independent and identically distributed (iid) $N(\mu, \sigma^2)$. Recall, the annual continuously compounded return is equal the sum of twelve monthly continuously compounded returns: $r_A = r(12) = \sum_{t=1}^{12} r_t$.Since each monthly return is normally distributed, the annual return is also normally distributed. The mean of $r(12)$ is

$$\begin{aligned}E[r(12)] &= E\left[\sum_{t=1}^{12} r_t\right] \\ &= \sum_{t=1}^{12} E[R_t] \text{ (by linearity of expectation)} \\ &= \sum_{t=1}^{12} \mu \quad \text{(by identical distributions)} \\ &= 12 \cdot \mu.\end{aligned}$$

Hence, the expected 12-month (annual) return is equal to 12 times the expected monthly return. The variance of $r(12)$ is

$$\begin{aligned}\text{var}(r(12)) &= \text{var}\left(\sum_{t=1}^{12} r_t\right) \\ &= \sum_{t=1}^{12} \text{var}(r_t) \quad \text{(by independence)} \\ &= \sum_{j=0}^{11} \sigma^2 \quad \text{(by identical distributions)} \\ &= 12 \cdot \sigma^2,\end{aligned}$$

so that the annual variance is also equal to 12 times the monthly variance[7]. Hence, the annual standard deviation is $\sqrt{12}$ times the monthly standard deviation: $\text{sd}(r(12)) = \sqrt{12}\sigma$ (this result is known as the square root of time rule). Therefore, $r(12) \sim N(12\mu,\, 12\sigma^2)$.

1.3.6 Covariance Between Linear Combinations of Random Variables

Consider the linear combinations of two random variables

$$\begin{aligned}Y &= X_1 + X_2, \\ Z &= X_3 + X_4\end{aligned}$$

[7] This result often causes some confusion. It is easy to make the mistake and say that the annual variance is $(12)^2 = 144$ time the monthly variance. This result would occur if $r_A = 12r_t$, so that $\text{var}(r_A) = (12)^2\text{var}(r_t) = 144\text{var}(r_t)$.

The covariance between Y and Z is

$$
\begin{aligned}
\mathrm{cov}(Y,Z) &= \mathrm{cov}(X_1+X_2,X_3+X_4) \\
&= E[((X_1+X_2)-(\mu_1+\mu_2))((X_3+X_4)-(\mu_3+\mu_4))] \\
&= E[((X_1-\mu_1)+(X_2-\mu_2))((X_3-\mu_3)+(X_4-\mu_4))] \\
&= E[(X_1-\mu_1)(X_3-\mu_3)]+E[(X_1-\mu_1)(X_4-\mu_4)] \\
&\quad +E[(X_2-\mu_2)(X_3-\mu_3)]+E[(X_2-\mu_2)(X_4-\mu_4)] \\
&= \mathrm{cov}(X_1,X_3)+\mathrm{cov}(X_1,X_4)+\mathrm{cov}(X_2,X_3)+\mathrm{cov}(X_2,X_4).
\end{aligned}
$$

Hence, covariance is additive for linear combinations of random variables.

1.4 Further Reading

Excellent intermediate level treatments of probability theory using calculus are given in Hoel, Port and Stone (1971) and DeGroot (1986). Intermediate treatments with an emphasis towards applications in finance include Ross (1999) and Watsom and Parramore (1998). Intermediate textbooks with an emphasis on econometrics include Amemiya (1994), Goldberger (1991), Ramanathan (1995). Advanced treatments of probability theory applied to finance are given in Neftci (1996). Everything you ever wanted to know about probability distributions is given in Johnson and Kotz (1994, 1995), Kotz, Balakrishnan and Johnson (2000) and .

Bibliography

[1] Amemiya, T. (1994). *Introduction to Statistics and Econometrics.* Harvard University Press, Cambridge, MA.

[2] Goldberger, A.S. (1991). *A Course in Econometrics.* Harvard University Press, Cambridge, MA.

[3] Hoel, P.G., Port, S.C. and Stone, C.J. (1971). *Introduction to Probability Theory.* Houghton Mifflin, Boston, MA.

[4] Johnson, N.L, and Kotz, S. (1994). *Continuous Univariate Distributions, Vol. 1*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

[5] Johnson, N.L, and Kotz, S. (1995). *Continuous Univariate Distributions, Vol. 2*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

[6] Kotz, S., Balakrishnan, N., and Johnson, N.L, and (2000). *Continuous Multivariate Distributions, Vol. 1*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

[7] Neftci, S.N. (1996). *An Introduction to the Mathematics of Financial Derivatives.* Academic Press, San Diego, CA.

[8] Ross, S. (1999). *An Introduction to Mathematical Finance: Options and Other Topics.* Cambridge University Press, Cambridge, UK.

[9] Watsham, T.J., and Parramore, K. (1998). *Quantitative Methods in Finance.* International Thompson Business Press, London, UK.