Chapter 1

Probability Concepts

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 the 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 $t-1$ (today) and plan to sell it at the end of month t. The return on this investment, $R_t = (P_t - P_{t-1})/P_t$, 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 1.1describes 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. 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 probability of failure. A mathematical model

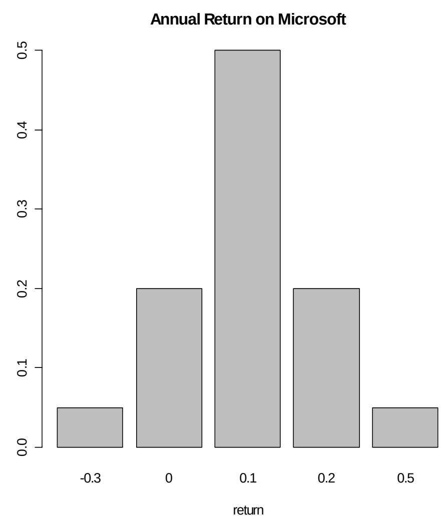

Figure 1.1: Discrete distribution for Microsoft stock.

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

To be completed.

1.1.2 Continuous Random Variables

Definition 8 *A continuous random variable X is one that can take on any real value. That is, $S_X = \{x : x \in \mathbb{R}\}$.*

Definition 9 *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*

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

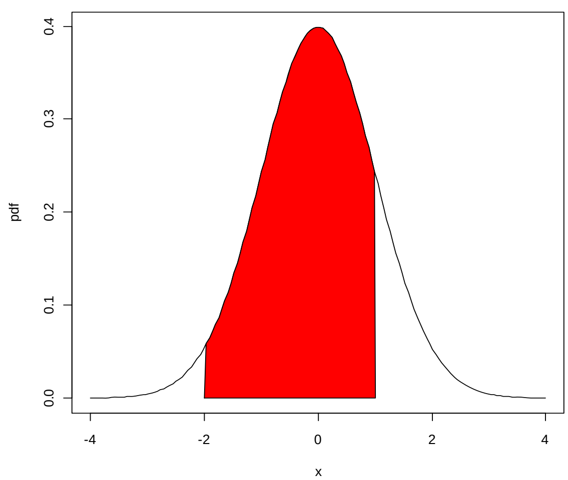

Figure 1.2: $\Pr(-2 \leq X \leq 1)$ is represented by the area under the probability curve.

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

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

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 the 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 xxx. Notice that the area under the curve over the interval $[a, b]$ (area of rectangle) integrates to one:

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

[Insert figure here]

Example 10 *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 $\delta \rightarrow 0$, $\Pr(0 \leq X \leq \delta) \rightarrow \Pr(X = 0)$. Using the above result we see that

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

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

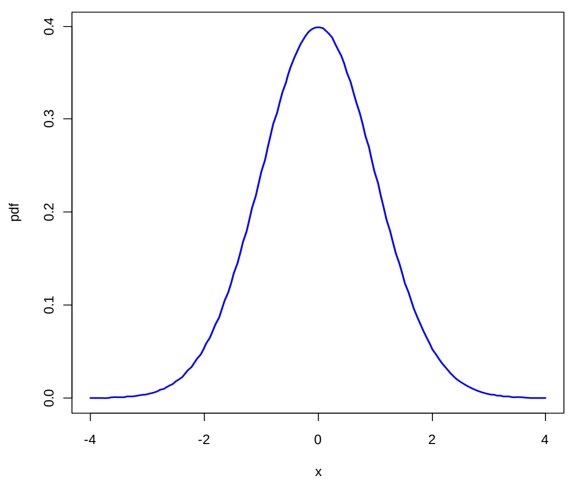

Figure 1.3: Standard normal density.

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. The pdf of a normal random variable is given by

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

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 1.3. 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

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

1.1.3 The Cumulative Distribution Function

Definition 11 *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 12 $F_X(x)$ *for a discrete random variable*

The cdf for the discrete distribution of Microsoft 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 graphed Figure xxx. Notice that the cdf in this case is a discontinuous step function with jumps at the four return values.

Insert figure here

■

Example 13 *$F(x)$ 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 14 *$F_X(x)$ 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:

$$F_X(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz. \tag{1.2}$$

The cdf $\Phi(x)$, however, does not have an analytic representation like the cdf of the uniform distribution and must be approximated using numerical techniques. A graphical representation of $\Phi(x)$ is given in Figure 1.4.

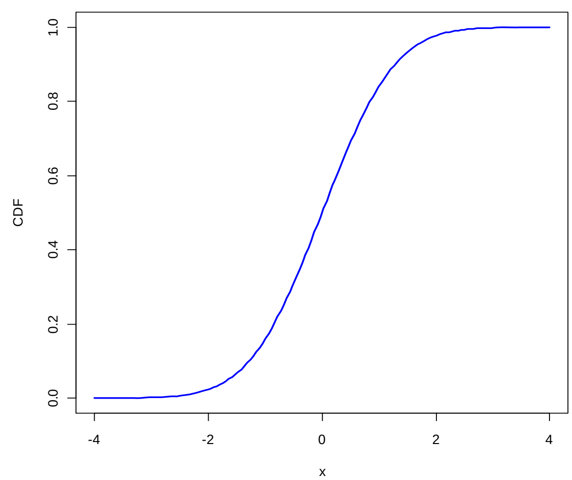

Figure 1.4: Standard normal cdf $\Phi(x)$.

1.1.4 Quantiles of the Distribution of a Random Variable

Consider a random variable X with cdf $F_X(x) = \Pr(X \leq x)$. 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[2] then q_α may be determined analytically as

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

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

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

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

Example 15 *Quantiles from a uniform distribution*

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

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

Given $\alpha \in [0, 1]$ such that $F_X(x) = \alpha$, solving for x gives the inverse cdf:

$$x = F_X^{-1}(\alpha) = \ \alpha(b-a) + a. \tag{1.3}$$

Using (1.3), 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 16 *Quantiles from a standard normal distribution*

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

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

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

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

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 (), the cdf (), and the inverse cdf (), respectively, with the default values `mean=1` and `sd = 0`. The function `rnorm()` returns a specified number of simulated values from the normal distribution.

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 17 *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

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 R

```
> 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 18 *Plotting the standard normal curve*

The graphs of the standard normal pdf and cdf in Figures 1.3 and 1.4 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 19 *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)
# create plot layout but do not plot anything
> plot(x.vals, d.vals, type="n", xlab="x", ylab="pdf")
> i = x.vals >= lb & x.vals <= ub
```

```
# add normal curve
> lines(x.vals, d.vals)
# 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

$$\mu_X = E[X] = \sum_{x \in S_X} x \cdot \Pr(X = x). \tag{1.4}$$

Eq. (1.4) shows that $E[X]$ is a probability weighted average of the possible values of X.

Example 20 *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 21 *Expected value of Bernoulli random variable*

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. ■

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

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x\cdot f(x)dx. \tag{1.5}$$

Example 22 *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 23 *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

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

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

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

Variance and Standard Deviation

The variance of a random variable X, denoted $\mathrm{var}(X)$ 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

$$\sigma_X^2 = \mathrm{var}(X) = E[(X - \mu_X)^2] = \sum_{x \in S_X} (x - \mu_X)^2 \cdot \Pr(X = x),$$

for X discrete, and

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx$$

for X continuous.

Because σ_X^2 represents an average squared deviation, it is not in the same units as X. The standard deviation of X, denoted $\mathrm{SD}(X)$ 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.

Example 24 *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

$$\begin{aligned}
\mathrm{var}(X) &= (-0.3-0.1)^2 \cdot (0.05) + (0.0-0.1)^2 \cdot (0.20) + (0.1-0.1)^2 \cdot (0.5) \\
&\quad +(0.2-0.1)^2 \cdot (0.2) + (0.5-0.1)^2 \cdot (0.05) \\
&= 0.020 \\
\mathrm{SD}(X) &= \sigma_X = \sqrt{0.020} = 0.141.
\end{aligned}$$

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 25 *Variance and standard deviation of a Bernoulli random variable*

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

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

■

Example 26 *Variance and standard deviation of a uniform random variable*

To be completed

Example 27 *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 $\mathrm{SD}(X) = 1$.■

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

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

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 28 *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 1.5 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
```

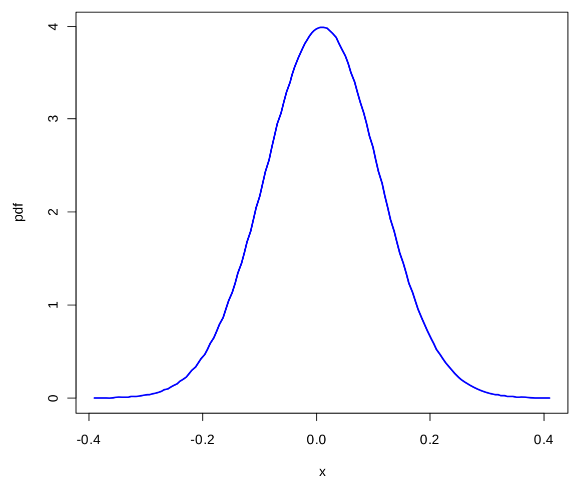

Figure 1.5: Normal distribution for the monthly returns on Microsoft: $R \sim N(0.01, (0.10)^2)$.

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 1% and 5% chances of gaining more than 17.5% and 24.3%, respectively.

■

Example 29 *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, 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 30 *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. For example, suppose $r_t = -2$. This implies a simple return of $R_t = e^{-2} - 1 = -0.865$. 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

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

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

$$\begin{aligned} \mu_Y &= E[Y] = e^{\mu_X + \sigma_X^2/2}, \\ \sigma_Y^2 &= \mathrm{var}(Y) = e^{2\mu_X + \sigma_X^2}(e^{\sigma_X^2} - 1). \end{aligned} \tag{1.6}$$

Example 31 *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

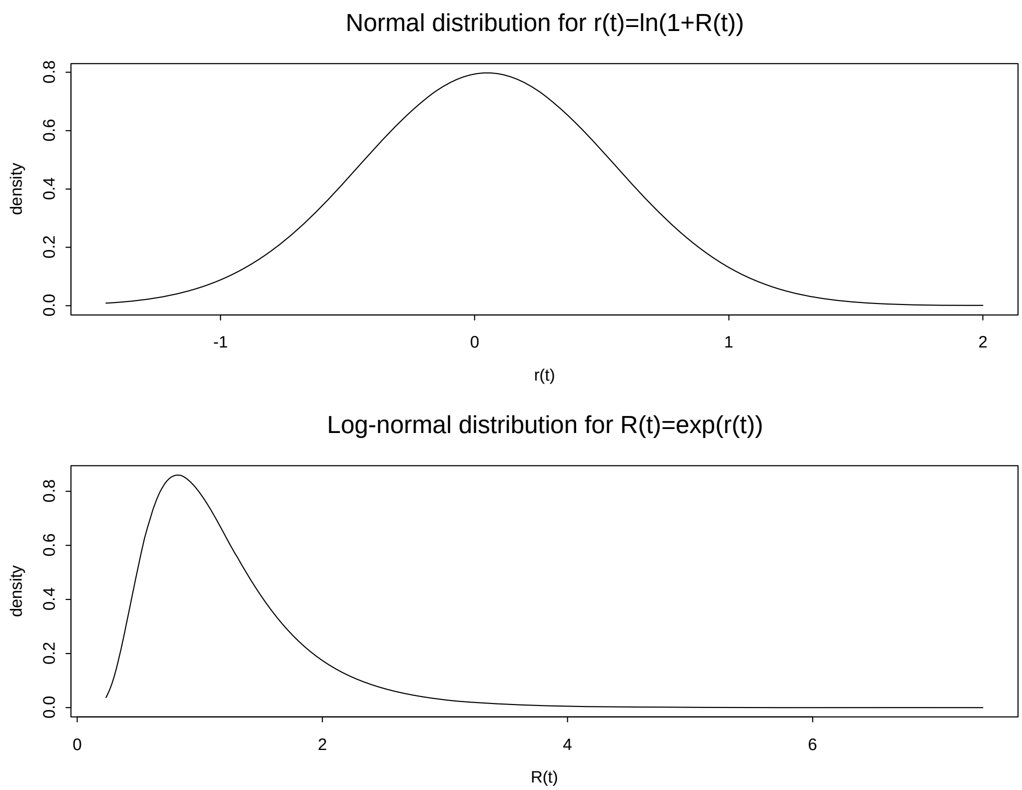

Figure 1.6: Normal distribution for r_t and log-normal distribution for $R_t = e^{r_t}$.

appropriate since the smallest value that R_t can take on is -1. Using (1.6), the mean and variance for $1 + R_t$ are given by

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

The pdfs for r_t and R_t are shown in figure 1.6.

Using standard deviation as a measure of risk

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

denote the monthly return on Amazon and R_B denote the monthly return on Boeing. These returns are to be treated as random variables since the returns will not be realized until the end of the month. We assume that $R_A \sim N(\mu_A, \sigma_A^2)$ and $R_B \sim N(\mu_B, \sigma_B^2)$. Hence, μ_i gives the expected return, $E[R_i]$, on asset $i = A, B$ and σ_i gives the typical size of the deviation of the return on asset i from its expected value. Figure xxx shows the pdfs for the two returns. Notice that $\mu_A > \mu_B$ but also that $\sigma_A > \sigma_B$. The return we expect on asset A is bigger than the return we expect on asset B but the variability of the return on asset A is also greater than the variability on asset B. The high return variability of asset A reflects the risk associated with investing in asset A. In contrast, if we invest in asset B 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 return you must take on extra risk.

insert figure here

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. For a discrete random variable X with sample space S_X

$$\text{skew}(X) = \frac{E[(X - \mu_X)^3]}{\sigma_X^3} = \frac{\sum_{x \in S_X} (x - \mu_X)^3 \cdot \Pr(X = x)}{\sigma_X^3}.$$

For a continuous random variable X with pdf $p(x)$

$$\text{skew}(X) = \frac{E[(X - \mu_X)^3]}{\sigma_X^3} = \frac{\int_{-\infty}^{\infty} (x - \mu_X)^3 \cdot p(x)dx}{\sigma_X^3}.$$

If X has a symmetric distribution then skew$(X) = 0$ since positive and negative values in the formula for skewness 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 6.

insert figure 6 here

Example 32 *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 33 *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 34 *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.

Kurtosis

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

$$\text{kurt}(X) = \frac{E[(X-\mu_X)^4]}{\sigma_X^4} = \frac{\sum_{x \in S_X}(x-\mu_X)^4 \cdot \Pr(X=x)}{\sigma_X^4}$$

where σ_X^4 is just SD(X) raised to the fourth power. For a continuous random variable X with pdf $p(x)$

$$\text{kurt}(X) = \frac{E[(X-\mu_X)^4]}{\sigma_X^4} = \frac{\int_{-\infty}^{\infty}(x-\mu_X)^4 \cdot p(x)dx}{\sigma_X^4}.$$

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.

Example 35 *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 36 *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

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

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

$$\text{excess kurt}(X) = \text{kurt}(X) - 3$$

If excess 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-t Distribution To be completed

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:

- $E[Y] = aE[X] + b$ or $\mu_Y = a\mu_X + b$.
- $\text{var}(Y) = a^2\text{var}(X)$ or $\sigma_Y^2 = a^2\sigma_X^2$.

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

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

In the second result notice 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 useful to go through the derivations, at least for the case that X is a discrete random variable.

Proof. Consider the first result. By the definition of $E[g(X)]$ with $g(X) = b + aX$ we have

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

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

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

Notice that our proof 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 37 *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.

Standardizing a Random Variable

Let X be a random variable with $E[X] = \mu_X$ and $\mathrm{var}(X) = \sigma_X^2$. 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, using the results of the previous section,

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

Hence, standardization creates a new random variable with mean zero and variance 1. In addition, if X is normally distributed then so is Z.
Let $X \sim N(2,4)$ and suppose we want to find $\Pr(X > 5)$. Since X is not standard normal we can't use the standard normal tables to evaluate $\Pr(X > 5)$ directly. We solve the problem by standardizing X as follows:

$$\begin{aligned}\Pr(X > 5) &= \Pr\left(\frac{X-2}{\sqrt{4}} > \frac{5-2}{\sqrt{4}}\right) \\ &= \Pr\left(Z > \frac{3}{2}\right)\end{aligned}$$

where $Z \sim N(0,1)$ is the standardized value of X. $\Pr\left(Z > \frac{3}{2}\right)$ can be found directly from the standard normal tables.

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 in the linear regression model is constructed by the above 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

$$X = \mu_X + \sigma_X Z.$$

This result is useful for modeling purposes. For example, in Chapter 3 we will consider the Constant Expected Return (CER) model of asset returns. Let R denote the monthly continuously compounded return on an asset and let $\mu = E[R]$ and $\sigma^2 = \text{var}(R)$. A simplified version of the CER model is

$$R = \mu + \sigma \cdot \varepsilon$$

where ε is a random variable with mean zero and variance 1. The random variable ε is often interpreted as representing the random news arriving in a given month that makes the observed return differ from the expected value μ. The fact that ε has mean zero means that new, on average, is neutral. The value of σ represents the typical size of a news shock.

1.1.8 Value at Risk: An Introduction

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, $E[R] = \mu = 0.05$ and $\text{var}(R) = \sigma^2 = (0.10)^2$. 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:

- What is the probability distribution of end of month wealth, W_1?
- 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?
- What is the monthly VaR on the $\$10,000$ investment in Microsoft stock with 5% probability? That is, what is the loss that would occur if the return on Microsoft stock is equal to its 5% quantile, $q_{.05}$?

To answer the first question, 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

$$\begin{aligned} W_1 &= W_0(1+R) = W_0 + W_0 R \\ &= \$10,000 + \$10,000 \cdot R. \end{aligned}$$

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

$$\begin{aligned} E[W_1] &= W_0 + W_0 E[R] \\ &= \$10,000 + \$10,000(0.05) = \$10,500 \end{aligned}$$

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 we have

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

To answer the second question, 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

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

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

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

The third question can be answered in two equivalent ways. First, use $R \sim N(0.05, (0.10)^2)$ and solve for the the 5% quantile of Microsoft Stock:

$$\Pr(R < q_{.05}^R) = 0.05 \Rightarrow q_{.05}^R = -0.114.$$

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:

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

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 may be computed using

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

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 xxx, the normal distribution may not be appropriate for characterizing the distribution of simple returns and is more appropriate for characterizing 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$ may be computed as follows:

- Compute the $\alpha \cdot 100\%$ quantile, q_α^r, from the Normal distribution for the continuously compounded return r

  $$q_\alpha = \mu_r + \sigma_r z_\alpha$$

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

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

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

- Compute VaR using the simple return quantile

  $$\text{VaR}_\alpha = |W_0 \cdot q_{.05}^R|.$$

1.1.9 Log-Normal Distribution and Jensen's Inequality

(discuss Jensen's inequality: $E[g(X)] < g(E[X])$ for a convex function. Use this to illustrate the difference between $E[W_0 \exp(R)]$ and $W_0 \exp(E[R])$ where R is a continuously compounded return.) Note, this is where the log-normal distribution will come in handy.