7 The log-normal stock price model

7.1 Continuous random variables A *density curve* is a graph of a line that

1. lies on or above the horizontal axis;
2. has total area 1 beneath it;
3. does not 'double back' (i.e. each vertical slice contains just one point).

In other words, it is the graph of a nonnegative function $f(x)$ with $\int_{-\infty}^{\infty} f(x)dx = 1$. Such a function is called a *probability density function* or pdf for short.

Given a probability density function f, we say a random quantity X has a *continuous* probability distribution with probability density function f (or with the graph of f as its density curve) if for any two numbers a and b the probability that X lies between a and b equals the area below the curve between points a and b on the horizontal axis, i.e.

$$P[a < X < b] = \int_a^b f(x)dx$$

Note also that for a continuous random quantity, $P[X = c] = 0$ for any fixed number c.

For a large number of repeated samples of a random quantity X with a given density curve, we'd expect that the histogram (with narrow bandwidths and with vertical scale chosen so the total area is one) to be approximately the same as the density curve.

If X is a continuous random quantity with pdf f, and $g(x)$ is a real-valued function defined for all real numbers x, then the expected value of the random quantity $g(X)$ is a number given by

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

In particular, taking $g(x) = x$ we have

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

For example, let $a < b$ and suppose X has pdf

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

Then we say X has a *uniform* distribution on the range of values from a to b. In this case

$$\begin{aligned} P[X \le (a+b)/2] &= 1/2 \\ E[X] &= \frac{b+a}{2} \\ E[X^2] &= \frac{b^2+ab+a^2}{3} \end{aligned}$$

The *variance* of a random quantity X (discrete or continuous), with $E[X] = \mu$, is a nonnegative number, denoted $\mathrm{Var}[X]$, given by

$$\mathrm{Var}[X] = E[(X-\mu)^2]$$

or by the alternative formula (which turns out to be equivalent)

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

The *standard deviation* of a random quantity X is a number given by

$$\mathrm{SD}[X] = \sqrt{\mathrm{Var}(X)}$$

Variance and standard deviation measure the 'spread' of the distribution of X. In the example above, (uniform distribution on the range from a to b) we have

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$
$$\mathrm{SD}(X) = \frac{b-a}{\sqrt{12}}$$

Standard deviation has the advantage of being in the same units as the original X. That is, if X is a RV and a is constant then

$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X) \qquad \mathrm{SD}(aX) = |a|\mathrm{SD}(X);$$

On the other hand, adding a constant does not change the variance; that is,

$$\mathrm{SD}(a+X) = \mathrm{SD}(X); \qquad \mathrm{Var}(a+X) = \mathrm{Var}(X)$$

7.2 The standard normal distribution.

A *normal* density curve is a density curve with a special kind of 'bell shape'. Each normal density curve has two *parameters* μ and σ . These are numbers which specify the location and spread of the curve (σ must be positive, but μ could be positive, negative or zero.) We first consider the case with $\mu = 0$ and $\sigma = 1$, called the *standard normal.*

Definition. A random quantity, Z, has a **standard normal distribution** if the probability density function of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty.$$

Then for any number z, we write $\Phi(z)$ for $P[Z \le z]$. We have

$$\Phi(z) = P[Z \le z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Note that by symmetry,

$$\Phi(-z) = 1 - \Phi(z)$$

Some probabilities for the standard normal distribution

There is no simple formula for $\Phi(z)$ but it is available from tables. Here are some values:

$$P[-1 \le Z \le 1] = 0.683$$
$$P[-2 \le Z \le 2] = 0.955$$
$$P[-3 \le Z \le 3] = 0.997$$

(So note that

$$\Phi(1) - \Phi(-1) = 0.683,$$

etc.)

If Z is standard normal, and t is any number, then

$$E[Z] = 0,$$
$$\text{Var}[Z] = 1,$$
$$E[e^{tZ}] = e^{t^2/2}$$

7.3 The normal distribution.

Definition Suppose that the random quantity Z has a standard normal distribution. For any real values $\sigma > 0, \mu$, the random quantity

$$X = \sigma Z + \mu$$

is said to have a **normal distribution**, parameters μ, σ, written $X \sim N(\mu, \sigma^2)$.

(So a standard normal distribution is N(0,1).)

If $X \sim N(\mu, \sigma^2)$, then

$$E[X] = \mu, \quad \text{Var}[X] = \sigma^2.$$

Note that if $X \sim N(\mu, \sigma^2)$, then for any number c,

$$\begin{aligned} P(X \leq c) &= P(\sigma Z + \mu \leq c) \\ &= P(Z \leq \frac{c-\mu}{\sigma}) \\ &= \Phi(\frac{c-\mu}{\sigma}) \end{aligned}$$

Example. Suppose $X \sim N(1,4)$. Then

$$P[X \leq -2] = \Phi\left(\frac{-2-1}{2}\right) = \Phi(-1.5) = 1 - 0.9332 = 0.0668$$

Theorem If $X \sim N(\mu, \sigma^2)$, then the pdf of X is

$$f(x) = \frac{1}{(\sqrt{2\pi})\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty.$$

The graph of this function f is called the normal curve with parameters μ and σ. It is *symmetric* about μ with maximum height $\frac{0.40}{\sigma}$. It is also *unimodal* (i.e. it has just one peak). There is a formula for the height of the normal curve at each point. For example at $\mu \pm \sigma$ the height is 0.61 times the maximum height, at $\mu \pm 2\sigma$ the height is about 0.14 times the maximum height.

If σ is big the curve is quite flat, if σ is small the curve is quite peaked. Whatever σ, the total area under the curve is always 1.

7.4 Sums of independent random variables

Suppose X and Y are **independent** random quantities with

$$X \sim N(\mu, \sigma^2)$$
$$Y \sim N(\nu, \tau^2)$$

Then

$$X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2).$$

In other words, the sum of two independent normal random quantities is itself normal, with mean given by the sum of their means and variance given by the sum of their variances. Note that the variances are additive but the standard deviations are not:

$$\begin{aligned} \mathrm{SD}[X + Y] &= \sqrt{\sigma^2 + \tau^2} \\ &= \sqrt{(\mathrm{SD}[X])^2 + (\mathrm{SD}[Y])^2} \end{aligned}$$

The above fact extends to more than two random quantities: if $X_1, \ldots, X_n$ are independent normal random quantities with
$X_i \sim N(\mu_i, \sigma_i^2)$ then

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

The reason for the importance of the normal distribution in probability theory is the **central limit theorem** which says that given **any** random quantity X with finite variance, the sum of a large number of independent copies of X is approximately normally distributed. (by 'independent copies' we mean independent random variables all with the same distribution as X).

A special case of the central limit theorem is the **normal approximation to the binomial distribution**, which says that if $X \sim Bi(n, p)$, and n is large and p is not very near 0 or 1. Then approximately the distribution of X is normal N(μ, σ^2), where

$$\mu = np, \sigma^2 = np(1 - p).$$

So, for any c, approximately

$$P(X \leq c) = \Phi\left(\frac{c - np}{\sqrt{np(1 - p)}}\right).$$