Chapter 6

Geometric Brownian Motions

1 Normal Distributions

We begin by recalling the normal distribution briefly. Let Z be a random variable distributed as standard normal, i.e., $Z \sim \mathcal{N}(0,1)$. The probability density function of Z is given by

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty. \tag{1}$$

This is the famous bell-curve of normal distributions. Since probability adds up to 1, we have $\int_{-\infty}^{\infty} p_X(x)dx = 1$. Clearly $\mathbb{E}[Z]$, the mean or expected value of Z, is equal to 0, and $\mathrm{Var}(Z)$, the variance of Z, is equal to 1.

Next we want to derive the nth moment $\mathbb{E}[Z^n]$ of Z. For this we need the moment generating function $\mathbb{E}[e^{tZ}]$. If it exists for some $t > 0$, then one can justify that

$$\mathbb{E}[e^{tZ}] = \sum_{k=0}^{\infty} \frac{\mathbb{E}[Z^k]}{k!} t^k. \tag{2}$$

Take note that the right hand side of (2) is the power series expansion in t of the function $\mathbb{E}[e^{tZ}]$. Its coefficients are precisely the moments of the random variable Z. Observe that the kth order derivative of the function $\mathbb{E}[e^{tZ}]$ at $t = 0$ is given by

$$\frac{d^k}{dt^k} \mathbb{E}[e^{tZ}]_{|t=0} = \mathbb{E}[Z^k], \tag{3}$$

which is precisely the kth moment of Z.

By definition,

$$\begin{aligned} \mathbb{E}[e^{tZ}] &\equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz - z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(z-t)^2 - t^2]} dz \\ &= e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} dz = e^{\frac{t^2}{2}}, \end{aligned} \tag{4}$$

where the last equality follows from $\int_{-\infty}^{\infty} p_Z(z)dz = 1$. From that and (3), we get

$$\mathbb{E}[Z^n] = \begin{cases} 0, & \text{when } n \text{ is odd,} \\ \dfrac{(2k)!}{2^k k!}, & \text{when } n = 2k. \end{cases} \tag{5}$$

In particular, $\mathbb{E}[Z^2] = 1$ and $\mathbb{E}[Z^4] = 3$.

Let X be a normally distributed random variable with mean μ and standard deviation σ, i.e., $X \sim \mathcal{N}(\mu, \sigma^2)$. Its probability density function is given by

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty. \tag{6}$$

Note that

$$Z \equiv \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1). \tag{7}$$

Hence by (5),

$$\mathbb{E}[(X-\mu)^n] = \begin{cases} 0, & \text{when } n \text{ is odd,} \\ \sigma^n \dfrac{(2k)!}{2^k k!}, & \text{when } n = 2k. \end{cases}$$

From that, one can compute the moments of X by recursion.

Recall that for $0 < \alpha < 1/2$, the 100α percentage point z_α of the standard normal distribution is defined as the number such that

$$\alpha = \text{Prob}\{Z > z_\alpha\} = \frac{1}{\sqrt{2\pi}} \int_{z_\alpha}^{\infty} e^{-u^2/2} du.$$

Referring to a statistical table, we have for examples, $z_{0.050} = 1.6449$ and $z_{0.025} = 1.960$. More precisely, if $Z \sim \mathcal{N}(0,1)$, then within 95% confidence,

$$-1.96 \leq Z \leq 1.96.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then by (7), we have

$$\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma. \tag{8}$$

If X signifies the daily return of a stock, then we say the stock will have a 2.5% *value at risk* (VAR) of \$$(\mu - 1.96\sigma)$. That implies that on average the stock holder expects a daily return of no more than \$$(\mu - 1.96\sigma)$ in 1 day out of 40 trading days.

2 A Simple Model for Asset Prices

The asset price and stock price were previously denoted by $S(t)$. Here we are going to take the "*volatility*" into account. The standard, generic symbol for such a "randomness" is ω. Thus we denote the asset price at time t by $S(t, \omega)$ or simply by $S_t(\omega)$. It is a stochastic process, i.e. for each fixed t, $S_t(\omega)$ is a random variable depending on ω. For those who are not familiar with stochastic processes, you may consider each ω as a realization of the process $S(t)$. More precisely, if we are allowed to move back in time (say a year ago) and start the process $S(t)$ for a year, then because of the randomness in $S(t)$, we should not be getting the same set of $S(t)$ as we already had. If we can repeat this experiment, i.e. move back in time a year and start $S(t)$ again, we should be getting another new set of $S(t)$. The parameter ω can be considered as the index of these experiments that you have taken, e.g. $S_t(1)$ is the first experiment, and $S_t(2)$ as the second experiment. However, ω need not be an integer. In the following, most of the time, we will simply write $S_t(\omega)$ as $S(t)$ or S_t, with the understanding that it is a stochastic process depending on ω.

We are going to build a model for a *non-dividend-paying stock* $S(t)$. But in general, what can we say about the stock price other than the fact that it is stochastic? Let us look at the price charts of HSBC and Yahoo from February 01 to February 06 in Figure 1. What do you see? We see fluctuation of the prices, that is for sure. But we also see one important thing. Both stock prices have an upward trend. In the short term, the price may rise and fall, but in the long term, the price is going to be higher. It is like putting money in the bank, our return will be positive and increasing according to the interest rate. For stocks, the long term returns seem to the positive although the return rate may be different for different stocks. From the charts, the return rate of Yahoo surely is much higher than that of HSBC.

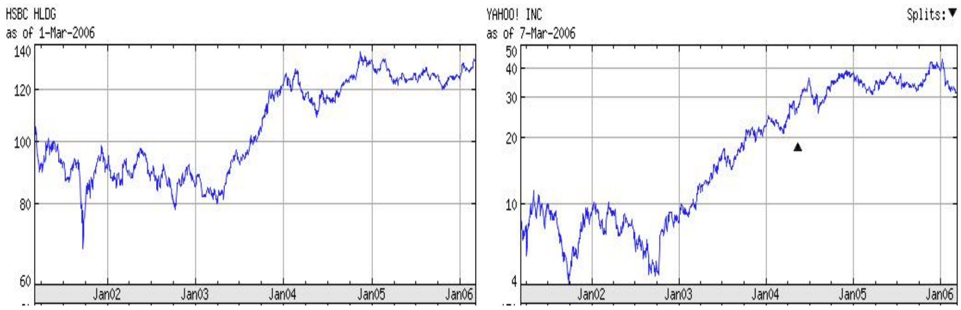

Figure 1. HSBC and Yahoo prices for the last 5 years.

We will build our model of the stock prices based on the above observation. We first recall the *return* is defined to be the change in the price divided by the original value, i.e.

$$\text{return} = \frac{\text{change in price}}{\text{original value}}.$$

Consider a small subsequent time interval $(t, t + \Delta t)$, during which $S(t)$ becomes $S(t + \Delta t) = S(t) + \Delta S(t)$, where $\Delta S(t) = S(t + \Delta t) - S(t)$. The return now can be expressed as

$$\frac{S(t + \Delta t) - S(t)}{S(t)} = \frac{\Delta S(t)}{S(t)}. \tag{9}$$

Let us adopt the differential notation used in calculus. Namely, we use the notation dt for the small change in any quantity over this time interval when we intend to consider it as an *infinitesimal* change. Then (9) can be written as dS/S. How might we model this return dS/S on the asset $S(t)$? According to the observation in Figure 1, the commonest model decomposes this return into two parts.

1. One is a predictable, deterministic and anticipated return akin to the return on money invested in a risk-free bank. It gives a contribution μdt to the return dS/S, where μ is a measure of the average rate of growth of the asset price. Such a parameter μ is also known as the *drift*. In simple models, μ is taken to be a constant. In more complicated models, μ can be a function of S and t.
2. The second contribution to dS/S models the random change in the asset price in response to external effects, such as unexpected news. It is represented by

a random sample $dX(t,\omega)$ drawn from a normal distribution with mean 0 and adds a term $\sigma dX(t,\omega)$ to dS/S. Here σ is a number called the *volatility*, which measures the standard deviation of the returns, and ω emphasizes that $dX(t,\omega)$ is a random process. For simplicity, we will sometimes simplify the writing of $dX(t,\omega)$ by $dX_t(\omega)$, $dX(t)$ or dX_t.

Putting these two contributions together, we obtain the *stochastic differential equation*

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dX(t), \tag{10}$$

which is the mathematical representation of our simple recipe for the price of a non-dividend-paying stock. Equivalently, (10) can also be expressed as

$$dS(t) = \mu S(t)dt + \sigma S(t)dX(t), \tag{11}$$

and

$$S_t(\omega) = S_{t_0}(\omega) + \mu \int_{t_0}^{t} S_u(\omega)du + \sigma \int_{t_0}^{t} S_u(\omega)dX_u(\omega). \tag{12}$$

Later in Definition 5, we will give more precise definitions of the drift parameter μ and the volatility σ. The only symbol in (10) whose role is not yet clear is $dX(t)$. Suppose we were to cross out this term by taking $\sigma = 0$. We then would be left with the ordinary differential equation

$$\frac{dS}{S} = \mu dt \quad \text{or} \quad \frac{dS}{dt} = \mu S.$$

When μ is a constant, this can be solved exactly to give exponential growth in the value of the asset, i.e.,

$$S(t) = S(t_0)e^{\mu(t-t_0)},$$

where $S(t_0)$ is the value of the asset at time t_0. Thus, if $\sigma = 0$, the asset price is totally deterministic, and we can predict the future price of the asset with certainty. In fact, it increases exponentially (or *geometrically* with common factor e^{μ}). It's like putting money in the bank with interest rate μ.

The model (10)–(12) for stock prices seems to have been first mentioned by Paul Samuelson in 1965 in his paper "Rational Theory of Warrant Prices," *Industrial Management Review*, **6** (1965), pp. 13–31. For reasons that will be clear in Chapter 7, the model is sometimes also referred as *geometric Brownian motion*.

Those who are not familiar with stochastic differential equations may be tempted to integrate (10) directly and get

$$S(t) = S(t_0)e^{\mu t + \sigma X(t)}.$$

We will see in the next chapter that this is *not* the correct solution to (10), and in fact $dS/S \neq d\ln S$. Thus there are difference between stochastic differential equations and ordinary differential equations. The difference can fortunately be handled easily by Itô's Lemma discussed in the next chapter.

3 Wiener Process and its Generalizations

The term $dX(t)$ in (10) which gives the randomness to $S(t)$ is certainly the main feature of the geometric Brownian motion. In fact, $X(t)$ is a *Wiener process* and is known to follow *Brownian motions*. Historically speaking, such a random process was observed by Robert Brown, an English botanist, in the summer of 1827, that "pollen grains suspended in water performed a continual swarming motion." Hence it was named after Robert Brown, called Brownian motion.

Brownian motion is a process of tremendous practical and theoretical significance. It was used as an model to explain the ceaseless irregular motions of tiny particles suspended in a fluid. It had also been used as a model of the stock market in Louis Bachelier's (1900) work. His paper was at first largely ignored by academics for many decades, but now his work stands as the innovative first step in a mathematical theory of stock markets that has greatly altered the financial world today.

In 1905, Albert Einstein gave a satisfactory explanation and asserted that the Brownian motion originates in the continued bombardment of the pollen grains by the molecules of the surrounding water, with successive molecular impacts coming from different directions and contributing different impulses to the particles. (Incidentally, 1905 is the same year in which Einstein set forth his theory of relativity and his quantum explanation for the photoelectric effect.) But, Brownian motion is complicated and it is not surprising that it took more than another decade to get a clear picture of the Brownian motion stochastic process.

A rigorous mathematical foundation upon which Brownian motion could be built had to wait until 1920's. In 1923, Norbert Wiener (1894–1964) laid a rigorous mathematical foundation and gave a proof of its existence. Hence, it explains why it is now also called a Wiener process. In the sequel, we will use both Brownian motion and Wiener process interchangeably.

Definition 1. We say a stochastic process $\{X_t(\omega), t \geq 0\}$ is a *standard Wiener process* if it satisfies the following conditions:

(a) $X_0(\omega) = 0$ for all ω;
(b) for all ω, the map $t \mapsto X_t(\omega)$ is a continuous function for $t \geq 0$;
(c) for every t and $h \geq 0$, the change $[X_{t+h}(\omega) - X_t(\omega)] \sim \mathcal{N}(0, h)$; and
(d) $X_u(\omega) - X_v(\omega)$ and $X_t(\omega) - X_s(\omega)$ are independent for all $0 \leq v \leq u \leq s \leq t$.

Condition (a) says that the starting point of a standard Wiener process is the origin. We in fact frequently speak of $\{\xi + X_t(\omega) : t \in \mathbb{R}^+\}$ as a Wiener process started at ξ. Note that this starting point ξ can be a fixed, real number, or a random variable independent of $X_t(\omega)$.

The conditions (b)–(c) are the really essential ones. For each fixed ω, the function $t \mapsto X_t(\omega)$ is called *sample path* (realization, trajectory) of the Wiener process associated with ω. By (b), Wiener paths are continuous. Again for simplicity, we will simply write $X_t(\omega)$ as X_t or $X(t)$. By (c), the increments are normally distributed with mean 0 and variance equal to the time difference. Thus by (6), for any event A,

$$\text{Prob}\{X_{t+h} - X_t \in A\} = \frac{1}{\sqrt{2\pi h}} \int_A e^{-x^2/2h} dx.$$

Example 1. Say, a Wiener process $Y(t)$ is initially equal to 25 and the time t is measured in years. Let us compute the probability distribution of $Y(t)$ at the end of one year. Let $\{X(t), t \geq 0\}$ be a standard Wiener process starting at 0. Then, $Y(t)$ is given by $Y(t) = X(t) + 25$. By Condition (c), $X(1) = X(1) - X(0) \sim \mathcal{N}(0, 1)$. Therefore, the probability distribution of $Y(1)$ is normal with mean 25 and variance 1. $\square$

From (d) and (a), one can deduce that, for $0 \leq t_0 < t_1 < \ldots < t_n$, the random variables

$$X_{t_0}, (X_{t_1} - X_{t_0}), \ldots, (X_{t_n} - X_{t_{n-1}})$$

are independent, i.e., X_t has independent, normally distributed, increments. In particular, we have $X_{t+h} - X_t$ is independent on X_t. Hence a Wiener process has what the so-called *Markov property*, which means that only the present value of the process is relevant for predicting the future, while the past history of the process and the way that the present has emerged from the past are irrelevant. A stochastic process which satisfies the Markov property is called a *Markov process.* Generally speaking, the following expression says that a Wiener process is Markov: for $t \geq 0$ and $h > 0$,

$$\text{Prob}\{X_{t+h} \in A \mid X_s, 0 \leq s \leq t\} = \text{Prob}\{X_{t+h} \in A \mid X_t\}. \tag{13}$$

Take note that stock prices are usually assumed to follow a Markov process. Suppose that the price of XYZ stock is \$50 now. If the stock price follows a Markov process, our prediction for the future should be unaffected by the price one week ago, one month ago, or one year ago. The only relevant piece of information is the price now which is \$50. Obviously, predictions for the future are uncertain and must be expressed in terms of probability distributions. The Markov property implies that the probability distribution of the price at any particular future time is not dependent on the particular path followed by the price in the past.

Definition 2. Let $\{X(t), t \geq 0\}$ be a standard Wiener process. The process

$$Y(t) = \sigma X(t) + \mu t + \zeta, \tag{14}$$

where μ is any real parameter and $\sigma > 0$, is called a *generalized Wiener process*, starting at ζ, with a *drift* parameter μ and a *variance rate* σ^2.

Using the properties of X_t, we see that

$$Y(t + h) - Y(t) \sim \mathcal{N}(\mu h, \sigma^2 h).$$

In differential notations, (14) becomes

$$\begin{cases} Y(0) &= \zeta, \\ dY(t) &= \mu dt + \sigma dX(t), \quad \text{for } t \geq 0. \end{cases}$$

Example 2. Consider the situation where the cash position $Y(t)$ of a company, measured in thousands of dollars, follows a generalized Wiener process with a drift parameter $\mu = 20$ per year and a variance rate of 900 per year. Initially, the cash position is 50. Let us compute the cash position of the company at the end of 5 years. Clearly, $Y(t) = \sigma X(t) + \mu t + \zeta$, where $\zeta = 50$, $\mu = 20$ and $\sigma = 30$. At the end of 5 years, the probability distribution of $Y(5)$ is normal with mean

$$\zeta + \mu t = 50 + 20 \times 5 = 150,$$

and variance

$$\sigma^2 t = 900 \times 5 = 4500.$$

In other words, $Y(5) \sim \mathcal{N}(150, 4500)$. $\square$

Definition 3. Let $\{X(t), t \geq 0\}$ be a standard Wiener process. The process

$$dY(t) = a(Y,t)dt + b(Y,t)dX(t),$$

where both the drift $a(Y,t)$ and the variance rate $b(Y,t)$ are functions of the underlying process $Y(t)$ and time t is called an *Itô process.*

The geometric Brownian motion in (11) is an Itô process.

4 Geometric Brownian Motions

Let us consider the geometric Brownian motion $S(t)$ in §2 again. The process $X(t)$ in (10)–(12) is a standard Wiener process. Note that $dX(t)$ $(= dX_t)$ can be considered as the limit of $X(t+dt) - X(t)$ as $dt \to 0$. Hence by Condition (c) in Definition 1, we have, as $dt \to 0$:

(i) $dX(t)$ is a random variable, drawn from a normal distribution;
(ii) the mean of $dX(t)$ is zero, i.e., $\mathbb{E}[dX(t)] = 0$;
(iii) the variance of $dX(t)$ is dt, i.e., $\mathrm{Var}(dX(t)) = dt$.

In short, $dX_t \sim \mathcal{N}(0, dt)$ and we can write

$$dX_t = \varepsilon\sqrt{dt}, \tag{15}$$

where $\varepsilon \sim \mathcal{N}(0,1)$ is a standard normal random variable. How about the random variable $(dX_t)^2$? First it mean is $\mathbb{E}[(dX_t)^2] = \mathrm{Var}(dX_t) = dt$. By (5), we also get

$$\mathrm{Var}((dX_t)^2) = \mathrm{Var}(\varepsilon^2(dt)) = \mathbb{E}[(\varepsilon^2 - \mathbb{E}[\varepsilon^2])^2](dt)^2 = 2(dt)^2.$$

Using this we can estimate the mean and variance of the return on $S(t)$.

Proposition 4. *The return dS/S on the asset $S(t)$ satisfies*

$$\frac{dS}{S} \sim \mathcal{N}(\mu dt, \sigma^2 dt). \tag{16}$$

Proof. Since $dS/S = \mu dt + \sigma dX_t$ where dX_t is normally distributed, dS/S is also normally distributed. By (10) and (ii) above,

$$\mathbb{E}\left[\frac{dS}{S}\right] = \mathbb{E}[\mu dt + \sigma dX_t] = \mu dt. \tag{17}$$

By (17), (10) and (iii), the variance is given by

$$\mathrm{Var}\left(\frac{dS}{S}\right) = \mathbb{E}\left[\left(\frac{dS}{S} - \mathbb{E}\left[\frac{dS}{S}\right]\right)^2\right] = \mathbb{E}\left[\left(\frac{dS}{S} - \mu dt\right)^2\right] = \mathbb{E}[(\sigma dX_t)^2] = \sigma^2 dt. \tag{18}$$

The standard deviation of dS/S is thus equal to $\sigma\sqrt{dt}$. $\square$

Equation (17) says that on average, the return on the asset $S(t)$ is increasing as a rate of μdt. Thus the next value for S is "higher" than the old one by an amount $\mu S dt$. If we recall the definition of interest rate in (3.4), this indicates that asset "on average"

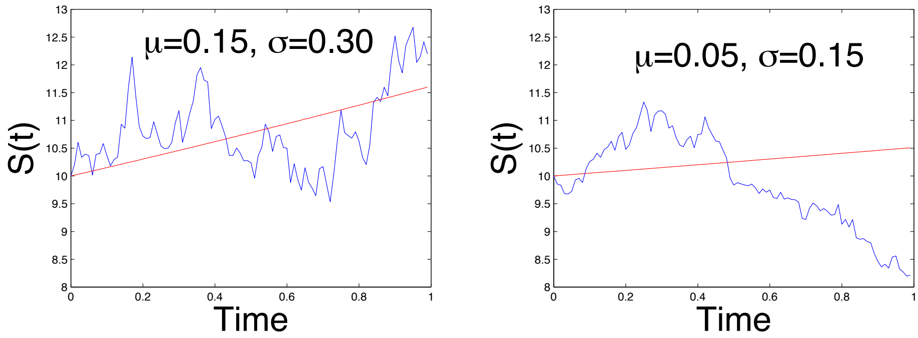

Figure 2. Asset prices for two different stocks with different μ and σ.

is earning an interest rate μ. From (17) and (18), we also see that if we compare two different asset-price processes as described by (10), the one with the larger value of μ usually rises more steeply and the one with the larger value of σ appears more jagged, see Figure 2 for two charts simulated according to the geometric Brownian motion (10). Note that the simulated price charts do have some resemblance to the real price charts in Figure 1.

Recall in (13) that $X(t)$ is Markov: it does not depend on past history. Moreover (11) does not refer to the past history of the asset price. Therefore the next asset price $S(t+dt)$ depends solely on today's price $S(t)$ and not on the past asset price. Thus the geometric Brownian motion $S(t)$ also has the *Markov property*.

The discrete version of (16) is

$$\frac{\Delta S}{S} \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t),$$

where the variable ΔS is the change in the stock price S in a small interval of time Δt. By (7), we then have

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \varepsilon \sqrt{\Delta t}, \tag{19}$$

where $\varepsilon \sim \mathcal{N}(0,1)$ is a random number drawing from a standard normal distribution. The left-hand side of (19) is the return provided by the stock in a short period of time, Δt. The term $\mu \Delta t$ is the expected value of this return, and the term $\sigma \varepsilon \sqrt{\Delta t}$ is the stochastic component of the return. The variance of the stochastic component (and, therefore, of the whole return) is $\sigma^2 \Delta t$. If we let $\Delta t = 1$ unit time, we have the following definition.

Definition 5. The *drift* μ of a stock is the expected return of the stock per unit time, and the *volatility* σ of the stock is the standard deviation of the return of the stock per unit time.

Example 3. Consider a stock that pays no dividends, has a volatility of 30% per annum, and provides an expected return of 15% per annum with continuous compounding. That is, it follows the model (10) with $\sigma = 0.30$ and $\mu = 0.15$. To be precise,

$$\frac{dS(t)}{S(t)} = 0.15 dt + 0.30 dX_t.$$

If S is the stock price at a particular time and ΔS is the increase in the stock price in the next small interval of time, then

$$\frac{\Delta S}{S} = 0.15\Delta t + 0.30\varepsilon\sqrt{\Delta t},$$

where $\varepsilon \sim \mathcal{N}(0,1)$. Let us consider a time interval of one week or 0.0192 year and suppose that the initial stock price is \$100. Then, $\Delta t = 0.0192$, $S = 100$, and hence

$$\Delta S = 100(0.00288 + 0.0416\varepsilon) = 0.288 + 4.16\varepsilon,$$

showing that the price increase is a random variable following the normal distribution with mean \$0.288 and standard deviation \$4.16. □

A *Monte Carlo simulation* of a stochastic process is a procedure for sampling random outcomes for the process. As an example, we can use (19) to simulate the stock price $S(t)$ that follows a geometric Brownian motion. We just need to draw random numbers ε in $\mathcal{N}(0,1)$.

Example 4. Suppose that the expected return from a stock is 14% per annum and that the standard deviation of the return (i.e., the volatility) is 20% per annum. This means that $\mu = 0.14$ and $\sigma = 0.20$. Suppose that $\Delta t = 0.01$ so that we are considering changes in the stock price in time intervals of length 0.01 year (or 3.65 days). From (19),

$$\Delta S = 0.14 \times 0.01S + 0.2\sqrt{0.01}S\varepsilon = 0.0014S + 0.02S\varepsilon. \tag{20}$$

A path for the stock price can be simulated by sampling repeatedly for ε from a standard normal distribution $\mathcal{N}(0,1)$ and substituting it into (20). For instance, the following table is one particular set of results from doing this.

Stock Price at Start of Period	Random Sample for ε	Change in Stock-price During Period
20.000	0.52	0.236
20.236	1.44	0.611
20.847	−0.86	−0.329
20.518	1.46	0.628
21.146	−0.69	−0.262
20.883	−0.74	−0.280
20.603	0.21	0.115
20.719	−1.10	−0.427
20.292	0.73	0.325
20.617	1.16	0.507
21.124	2.56	1.111

Here the initial stock price is assumed to be \$20. For the first period, ε is sampled as 0.52. From (20), the change during the first time period is

$$\Delta S = 0.0014 \times 20 + 0.02 \times 20 \times 0.52 = 0.236.$$

At the beginning of the second time period, the stock price is, therefore, \$20.236. The value of ε sampled for the next period is 1.44. From (20), the change during the second time period is

$$\Delta S = 0.0014 \times 20.236 + 0.02 \times 20.236 \times 1.44 = 0.611.$$

At the beginning of the next period the stock price is, therefore, \$20.847; and so on. $\square$

Note that, because the process we are simulating is Markov, the samples ε that we generate should be independent of each other. Also the above table only shows one possible pattern of stock price movements. Different random samples would lead to different price movements. In fact, if we start all over again, we should be getting a completely new path for $S(t)$.

5 An Important Fact: $(dX_t)^2 = dt$

In (15), we have seen that $dX_t = \varepsilon\sqrt{dt}$, where $\varepsilon \sim \mathcal{N}(0,1)$. Hence $(dX_t)^2 = \varepsilon^2 dt$ where $\mathbb{E}[\varepsilon^2] = \mathrm{Var}[\varepsilon] = 1$. In fact, we can show that the smaller dt becomes, the more certainly $(dX_t)^2$ is equal to dt. To prove this, we need to establish a famous inequality first.

Proposition 6. (Chebyshev's inequality) *For any continuous random variable R and any positive δ, we have*

$$\mathrm{Prob}\{|R| \geq \delta\} \leq \frac{1}{\delta^2}\mathbb{E}[R^2]. \tag{21}$$

Proof. Let $p(r)$ be the probability density function of R. We have

$$\mathbb{E}[R^2] \equiv \int_{-\infty}^{\infty} r^2 p(r)dr \geq \int_{|r|\geq\delta} r^2 p(r)dr \geq \int_{|r|\geq\delta} \delta^2 p(r)dr = \delta^2 \mathrm{Prob}\{|R| \geq \delta\}. \quad \square$$

Theorem 7. *Let $\{X(t), t \geq 0\}$ be a standard Wiener process starting at 0. Then*

$$\mathrm{Prob}\left\{\left|\int_0^t (dX_s)^2 - t\right| \geq \delta\right\} = 0 \tag{22}$$

for any $\delta > 0$. In other words, as $dt \to 0$, $(dX_t)^2 \to dt$ with probability 1.

Proof. Let us take the partition $\{0, t/n, 2t/n, \ldots, t\}$ of the interval $[0,t]$, and denote $\zeta_i = i \cdot t/n$. Note that by the definition of integration

$$\int_0^t (dX_s)^2 = \lim_{n\to\infty} \sum_{i=1}^n \left(X(\zeta_i) - X(\zeta_{i-1})\right)^2,$$

where we will assume here and in the following that all limits exist. By Definition 1(c), if we define

$$Z_{n,i} \equiv \frac{X(\zeta_i) - X(\zeta_{i-1})}{\sqrt{t/n}},$$

then for each n, the sequence $Z_{n,1}, Z_{n,2}, \ldots$ is a set of independent, identically distributed $\mathcal{N}(0,1)$ random variables. Moreover, we have

$$\int_0^t (dX_s)^2 - t = \lim_{n\to\infty} \frac{t}{n}\sum_{i=1}^n Z_{n,i}^2 - t = \lim_{n\to\infty} \frac{t}{n}\sum_{i=1}^n (Z_{n,i}^2 - 1) = \lim_{n\to\infty} R_n, \tag{23}$$

where

$$R_n \equiv \frac{t}{n}\sum_{i=1}^n (Z_{n,i}^2 - 1).$$

To prove (22), we will apply the Chebyshev inequality (21) on R_n. Thus we need to compute $\mathbb{E}[R_n^2]$. Note that

$$\begin{aligned}
R_n^2 &= \frac{t^2}{n^2}\left(\sum_{i=1}^{n}(Z_{n,i}^2-1)\right)\left(\sum_{j=1}^{n}(Z_{n,j}^2-1)\right)\\
&= \frac{t^2}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Z_{n,i}^2 Z_{n,j}^2 - 2n\sum_{i=1}^{n} Z_{n,i}^2 + n^2\right).
\end{aligned}$$

Note that for all n and i, $Z_{n,i} \sim \mathcal{N}(0,1)$, hence

$$\begin{aligned}
\mathbb{E}[R_n^2] &= \frac{t^2}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[Z_{n,i}^2 Z_{n,j}^2] - 2n\sum_{i=1}^{n} \mathbb{E}[Z_{n,i}^2] + n^2\right)\\
&= \frac{t^2}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[Z_{n,i}^2 Z_{n,j}^2] - n^2\right).
\end{aligned}$$

Recall that $Z_{n,i}$ and $Z_{n,j}$ are independent of each other if $i \neq j$, hence

$$\begin{aligned}
\mathbb{E}[R_n^2] &= \frac{t^2}{n^2}\left(\sum_{i=1}^{n} \mathbb{E}[Z_{n,i}^4] + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \mathbb{E}[Z_{n,i}^2]\mathbb{E}[Z_{n,j}^2] - n^2\right)\\
&= \frac{t^2}{n^2}\left(\sum_{i=1}^{n} \mathbb{E}[Z_{n,i}^4] + n(n-1) - n^2\right).
\end{aligned}$$

By (5), we then have,

$$\mathbb{E}[R_n^2] = \frac{t^2}{n^2}\left(3n + n(n-1) - n^2\right) = \frac{2t^2}{n}.$$

Putting this into the Chebyshev inequality (21), we then have

$$0 \leq \mathrm{Prob}\{|R_n| \geq \delta\} \leq \frac{2t^2}{\delta^2 n}, \quad \text{for all } \delta > 0.$$

Taking limit as $n \to \infty$, we have

$$0 \leq \lim_{n\to\infty} \mathrm{Prob}\{|R_n| \geq \delta\} \leq \lim_{n\to\infty} \frac{2t^2}{\delta^2 n} = 0.$$

Recall that probability is just an integration and limit and integration can interchange order under sufficiently nice conditions, which we assume are satisfied here. Hence

$$\mathrm{Prob}\{|\lim_{n\to\infty} R_n| \geq \delta\} = \mathrm{Prob}\{\lim_{n\to\infty} |R_n| \geq \delta\} = \lim_{n\to\infty} \mathrm{Prob}\{|R_n| \geq \delta\} = 0.$$

Hence by (23), we have (22). Thus with probability 1,

$$\int_0^t (dX_s)^2 = t, \tag{24}$$

or in differential form $(dX_t)^2 = dt$. $\square$

Let us illustrate (24) by a numerical example. We set the time interval $t_i - t_{i-1} = 0.001$ for all $i = 1, \cdots, 1000$, and simulate $X(t_i) - X(t_{i-1}) \sim \mathcal{N}(0, 0.001)$. Then we compute the sum $\sum_i^n (X(\frac{i}{1000}) - X(\frac{i-1}{1000}))^2$ for all $n = 1, \cdots, 1000$. The sum for each n is shown as the red curve in Figure 3. We see that the red curve is indeed very close to the function $y(t) = t$.

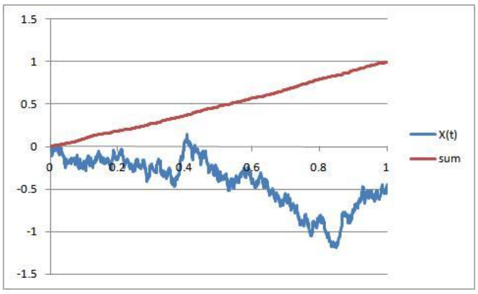

Figure 3. A numerical result to show that $\int_0^t (dX_s)^2 = t$

From Theorem 7, we can derive the following estimates:

Proposition 8. *As $dt \to 0$, we have*

(i) $dX_t = O(\sqrt{dt})$; *i.e.* $\lim_{dt \to 0}\{dX_t/\sqrt{dt}\} = \text{constant}$.
(ii) $dX_t dt = o(dt)$; *i.e.* $\lim_{dt\to 0}\{(dX_t dt)/dt\} = 0$.
(iii) $(dX_t)^2 = dt$;

Example 5. Squaring both sides of (11), we get

$$(dS)^2 = \mu^2 S^2 (dt)^2 + 2\mu\sigma S^2 dt dX_t + \sigma^2 S^2 (dX_t)^2.$$

By Proposition 8, it becomes

$$(dS)^2 = \sigma^2 S^2 dt + o(dt).$$

In the limit $dt \to 0$, we have

$$(dS)^2 = \sigma^2 S^2 dt. \quad \square$$