LOGNORMAL MODEL FOR STOCK PRICES

MICHAEL J. SHARPE
MATHEMATICS DEPARTMENT, UCSD

1. Introduction

What follows is a simple but important model that will be the basis for a later study of stock prices as a geometric Brownian motion. Let S_0 denote the price of some stock at time $t = 0$. We then follow the stock price at regular time intervals $t = 1$, $t = 2, \dots, t = n$. Let S_t denote the stock price at time t. For example, we might start time running at the close of trading Monday, March 29, 2004, and let the unit of time be a trading day, so that $t = 1$ corresponds to the closing price Tuesday, March 30, and $t = 5$ corresponds to the price at the closing price Monday, April 5. The model we shall use for the (random) evolution of the the price process $S_0, S_1, \dots, S_n$ is that for $1 \le k \le n$, $S_k = S_{k-1}X_k$, where the X_k are strictly positive and IID—i.e., independent, identically distributed. We shall return to this model after the next section, where we set down some reminders about normal and related distributions.

2. Properties of the Normal and Lognormal Distributions

First of all, a random variable Z is called standard normal (or $N(0,1)$, for short), if its density function $f_Z(z)$ is given by the standard normal density function $\phi(z) :== \frac{e^{-z^2/2}}{\sqrt{2\pi}}$. The function $\Phi(z) := \int_{-\infty}^{z} \phi(u)\,du$ denotes the distribution function of a standard normal variable, so an equivalent condition is that the distribution function (also called the cdf) of Z satisfies $F_Z(z) = P(Z \le z) = \Phi(z)$. You should recall that

$$\int_{-\infty}^{\infty} \phi(z)\,dz = 1 \qquad \text{i.e., } \phi \text{ is a probability density} \tag{2.1}$$

$$\int_{-\infty}^{\infty} z\phi(z)\,dz = 0 \qquad \text{with mean 0} \tag{2.2}$$

$$\int_{-\infty}^{\infty} z^2\phi(z)\,dz = 1 \qquad \text{and second moment 1.} \tag{2.3}$$

In particular, if Z is $N(0,1)$, then the mean of Z, $E(Z) = 0$ and the second moment of Z, $E(Z^2) = 1$. In particular, the variance $V(Z) = E(Z^2) - (E(Z))^2 = 1$. Recall that standard deviation is the square root of variance, so Z has standard deviation 1. More generally, a random variable V has a normal distribution with mean μ and standard deviation $\sigma > 0$ provided $Z := (V - \mu)/\sigma$ is standard normal. We write for short $V \sim N(\mu, \sigma^2)$. It's easy to check that in this case, $E(V) = \mu$ and $\mathrm{Var}(V) = \sigma^2$. There are three essential facts you should remember when working with normal variates.

Theorem 2.4. *Let $V_1, \dots, V_k$ be independent, with each $V_j \sim N(\mu_j, \sigma_j^2)$. Then*

$$V_1 + \cdots + V_k \sim N(\mu_1 + \cdots + \mu_k, \sigma_1^2 + \cdots + \sigma_k^2).$$

Theorem 2.5. *(Central Limit Theorem:) If a random variable V may be expressed a sum of independent variables, each of small variance, then the distribution of V is approximately normal.*

This statement of the CLT is very loose, but a mathematically correct version involves more than you are assumed to know for this course. The final point to remember is a few special cases, assuming $V \sim N(\mu, \sigma^2)$.

$$P(|V - \mu| \le \sigma) \approx 0.68; \qquad P(|V - \mu| \le 2\sigma) \approx 0.95; \qquad P(|V - \mu| \le 3\sigma) \approx 399/400. \tag{2.6}$$

We'll say that a random variable $X = \exp(\sigma Z + \mu)$, where $Z \sim N(0, 1)$, is lognormal(μ,σ^2). Note that the parameters μ and σ are the mean and standard deviation respectively of $\log X$. Of course, $\sigma Z + \mu \sim N(\mu, \sigma^2)$, by definition. The parameter μ affects the scale by the factor $\exp(\mu)$, and we'll see below that the parameter σ affects the shape of the density in an essential way.

Proposition 2.7. *Let X be lognormal(μ,σ^2). Then the distribution function F_X and the density function f_X of X are given by*
(2.8)
$$F_X(x) = P(X \le x) = P(\log X \le \log x) = \Phi(\sigma Z + \mu \le \log x) = P\Big(Z \le \frac{\log x - \mu}{\sigma}\Big) = \Phi\Big(\frac{\log x - \mu}{\sigma}\Big), \quad x > 0.$$

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{\phi\Big(\frac{\log x - \mu}{\sigma}\Big)}{\sigma x}, \qquad x > 0. \tag{2.9}$$

These permit us to work out a formulas for the moments of X. First of all, for any positive integer k,

$$E(X^k) = \int_0^\infty x^k f_X(x)\, dx = \int_0^\infty \frac{x^k \phi\Big(\frac{\log x - \mu}{\sigma}\Big)}{\sigma x}\, dx$$

hence after making the substitution $x = \exp(\sigma z + mu)$, so that $dx = \sigma \exp(\sigma z + \mu)$, we find

$$E(X^k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{z^2}{2} + k\sigma z + k\mu}\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(z - k\sigma)^2 + \frac{k\sigma^2}{2} + k\mu}\, dz = e^{\frac{k^2\sigma^2}{2} + k\mu}. \tag{2.10}$$

(We completed the square in the exponent, then used the fact that by a trivial substitution, $\int_{-\infty}^\infty \phi(z - a)\, dz = 1$.) In particular, setting $k = 1$ and $k = 2$ give

$$E(X) = e^{\frac{\sigma^2}{2} + \mu}; \qquad E(X^2) = e^{2\sigma^2 + 2\mu}; \qquad V(X) = E(X^2) - (E(X))^2 = e^{\sigma^2 + 2\mu}\big(e^{\sigma^2} - 1\big). \tag{2.11}$$

The median of X (which continues to be assumed lognormal(μ,σ^2)) is that x such that $F_X(x) = 1/2$. By (2.8), this is the same as requiring $\Phi(\frac{\log x - \mu}{\sigma}) = 1/2$, hence that $\frac{\log x - \mu}{\sigma} = 0$, and so $\log x = \mu$, or $x = e^\mu$. That is,

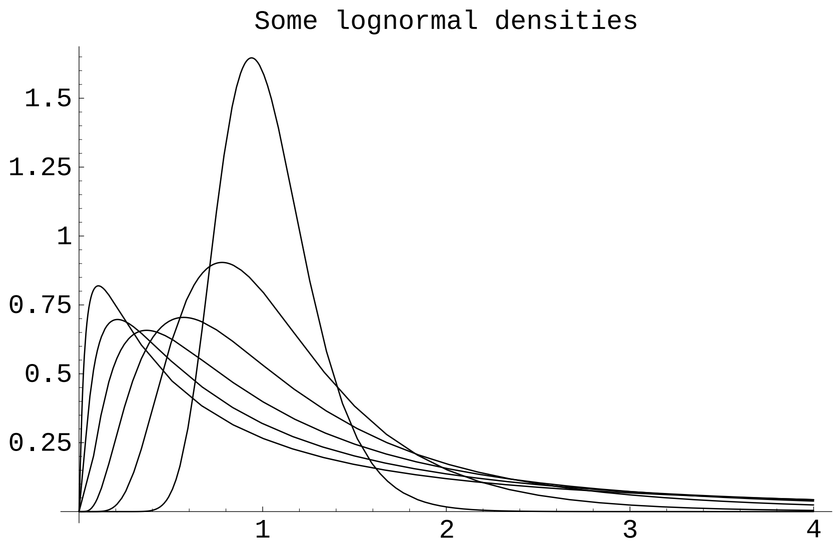
$$X \text{ has median } e^\mu. \tag{2.12}$$

The two theorems above for normal variates have obvious counterparts for lognormal variates. We'll state them somewhat informally as:

Theorem 2.13. *A product of independent lognormal variates is also lognormal with respective parameters $\mu = \sum \mu_j$ and $\sigma^2 = \sum \sigma_j^2$.*

Theorem 2.14. *A random variable which is a product of a large number of independent factors, each close to 1, is approximately lognormal.*

Here is a sampling of lognormal densities with $\mu = 0$ and σ varying over $\{.25, .5, .75, 1.00, 1.25, 1.50\}$.



The smaller σ values correspond to the rightmost peaks, and one sees that for smaller σ, the density is close to the normal shape. If you think about modeling men's heights, the first thing one thinks about is modeling with a normal distribution. One might also consider modeling with a lognormal, and if we take the unit of measurement to be 70 inches (the average height of men), then the standard deviation will be quite small, in those units, and we'll find little difference between those particular normal and lognormal densities.

3. Lognormal Price Model

We continue now with the model described in the introduction: $S_k = S_{k-1}X_k$. The first natural question here is which specific distributions should be allowed for the X_k. Let's suppose we follow stock prices not just at the close of trading, but at all possible $t \geq 0$, where the unit of t is trading days, so that, for example, $t = 1.3$ corresponds to .3 of the way through the trading hours of Wednesday, March 31. Note that $\frac{S_1}{S_0} = \frac{S_1}{S_{.5}}\frac{S_{.5}}{S_0}$, and under the time homogeneity postulated above, one should suppose that $\frac{S_1}{S_{.5}}$ and $\frac{S_{.5}}{S_0}$ are IID. Continuing in this way, we see that for any positive integer m, setting $h = 1/m$,

$$\frac{S_1}{S_0} = \frac{S_{mh}}{S_{(m-1)h}}\frac{S_{(m-1)h}}{S_{(m-2)h}}\cdots\frac{S_h}{S_0}$$

where the factors $\frac{S_{kh}}{S_{(k-1)h}}$ are IID. Consequently, taking logarithms, we find

$$\log(X_1) = \log\left(\frac{S_1}{S_0}\right) = \sum_{k==1}^{m}\log\left(\frac{S_{kh}}{S_{(k-1)h}}\right)$$

so that for arbitrarily large m, $\log X_1$ may be represented as the sum of m IID random variables. In view of the Central Limit Theorem, under mild additional conditions—for example, if $\log X_1$ has finite variance, then $\log X_1$ must have a normal distribution. Therefore, it is reasonable to hypothesize that the X_k are lognormal, and we may write $X_k = \exp(\sigma Z_k + \mu)$, where the Z_k are IID standard normal.

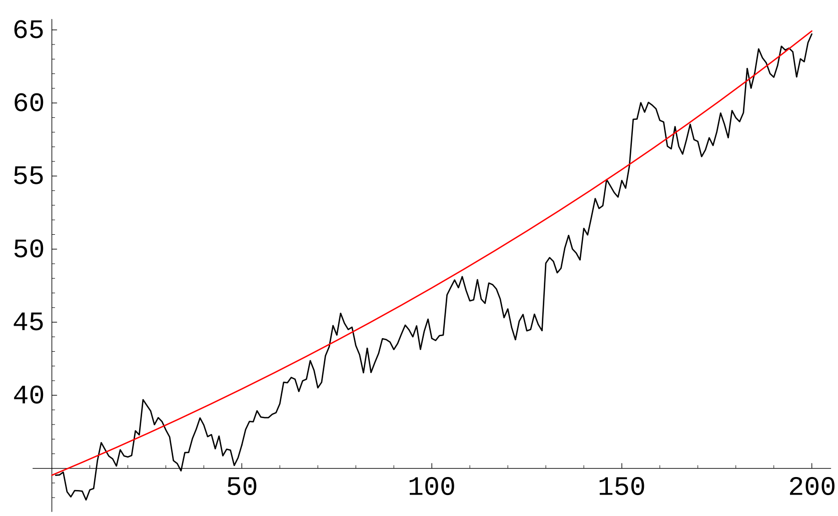
The first issue is the estimation of the parameters μ and σ from data. The thing you need to recall is that if you have a sample of n IID normal variates $Y_1, \ldots, Y_n$ with unknown mean μ and unknown standard deviation σ, then the sample mean $\bar{Y} := \frac{Y_1+\cdots+Y_n}{n}$ is an unbiased estimator of μ and $\frac{\sum(Y_k-\bar{Y})^2}{n-1}$ is an unbiased estimator of σ^2. If we denote by $\bar{Y^2}$ the mean value of the Y_k^2, it is elementary algebra to verify that

$$\frac{\sum(Y_k-\bar{Y})^2}{n-1} = \frac{n}{n-1}\left(\bar{Y^2} - \bar{Y}^2\right). \tag{3.1}$$

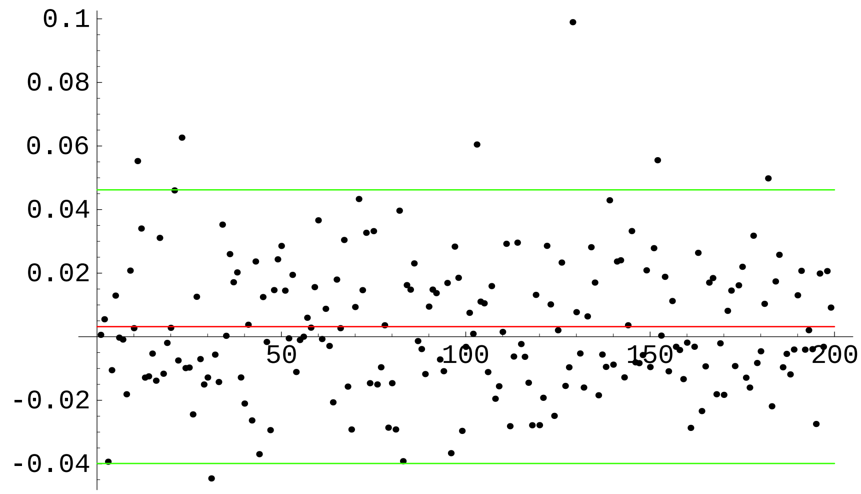
When n is large, the factor $n/(n-1)$ is close to 1, and may be ignored. Be aware that when Excel computes the variance (VAR) of a list of numbers y_1 through y_n, it uses this formula.

So, if we have a sample of stock prices S_0 through S_n, we compute the n ratios $X_1 := \frac{S_1}{S_0}$ through $X_n := \frac{S_n}{S_{n-1}}$ and then set $Y_k := \log X_k$. (In the financial literature, $R_k := \frac{S_k - S_{k-1}}{S_{k-1}} = X_k - 1$ is called the **return** for the k^{th} day. In practice, X_k is quite close to 1 most of the time, and so Y_k is mostly close to 0. For this reason, since $\log(1+z)$ is close to z when z is small, Y_k is mostly very close to the return R_k.) Apply the estimators described above to estimate μ by $\bar{Y}$ and σ^2 by formula (3.1). This kind of calculation can be conveniently handled by an Excel spreadsheet, or a computer algebra system such as *Mathematica*^{TM}. Stock price data is available online, for example at `http://biz.yahoo.com`. Spreadsheet files of stock price histories may be downloaded from that site in CSV (comma separated value) format, which may be imported from Excel or *Mathematica*^{TM}.

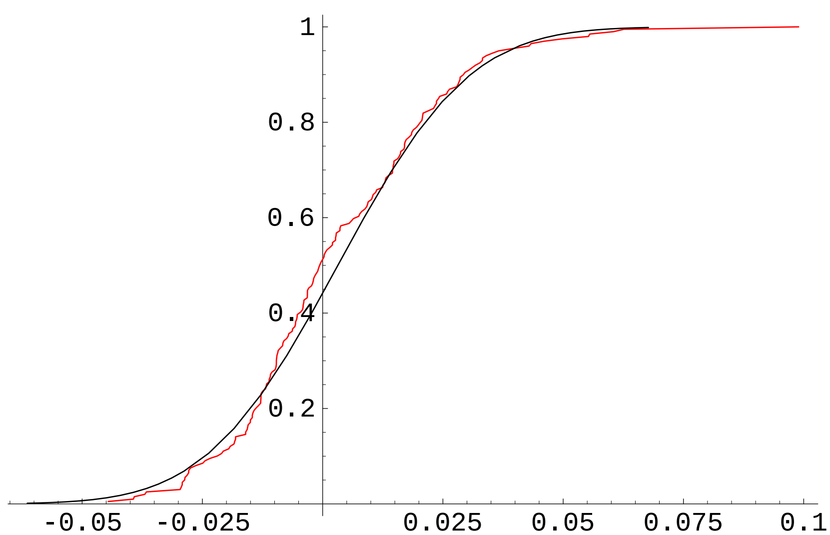
The parameters μ and σ arising from this stock price model are called the *drift* and *volatility* respectively. The idea is that stocks price movement is governed by a deterministic exponential growth rate μ, though subject to random fluctation whose magnitude is governed by σ. The following picture of Qualcomm stock (QCOM) over roughly the last nine months is shown in the following picture, along with the deterministic growth rate $S_0 e^{k\mu}$. You might at this point check out the last page of this handout, where I've graphed the result of 10 simulations starting at the same initial price, but using independent lognormal multipliers with the same drift and volatility as this data.



The graph below shows a plot of the values $\log X_j$ versus time j, along with a horizontal red line at their mean μ and horizontal green lines at levels $\mu \pm 2\sigma$. Note that of the 199 points in the plot, only 7 are outside these levels. This is not far from the roughly 5% of outliers you would expect, based on the normal frequencies.



The empirical cdf of the $\log X_k$ is pictured next (in red) compared with a normal cdf having the estimated μ and σ.



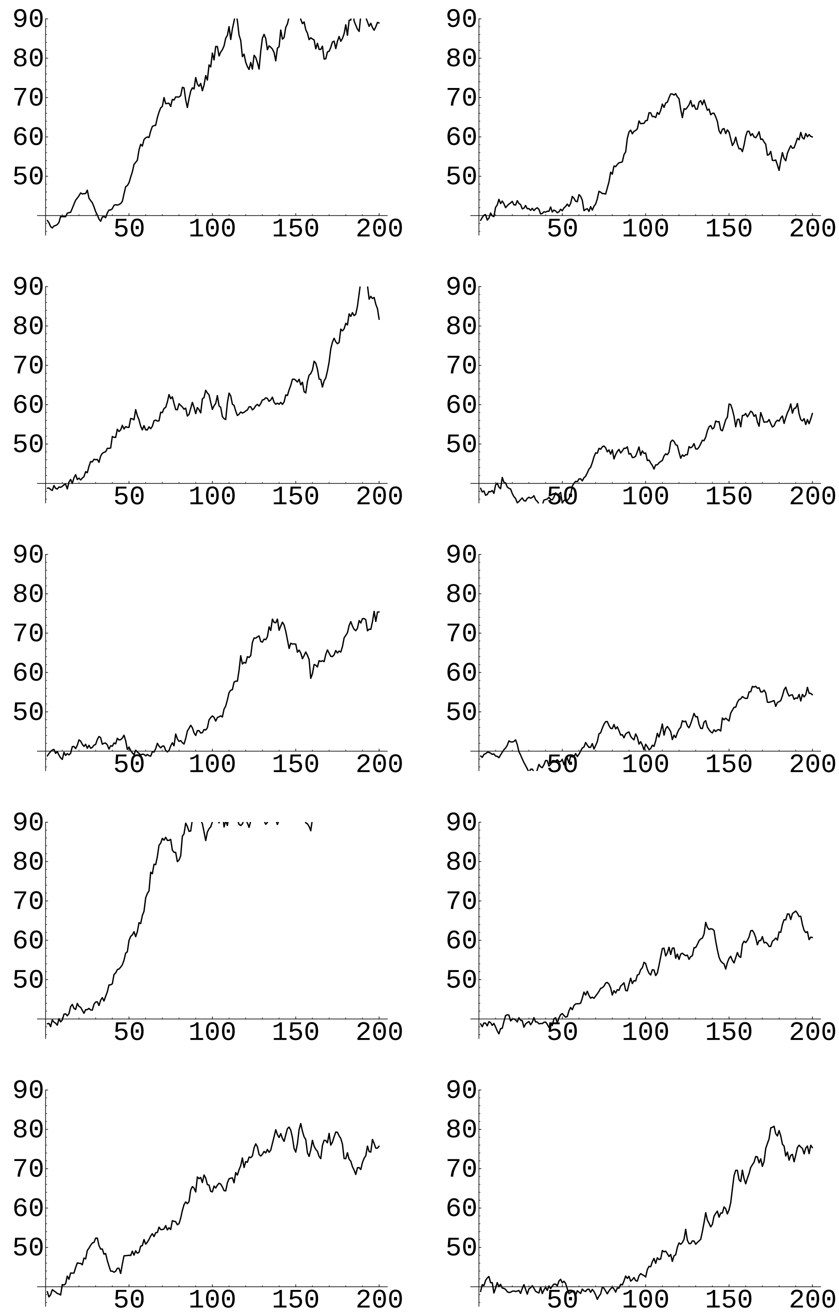
The following is only for those who already know about such matters. To test whether the $\log X_k$ are normal, one computes the maximal difference D between the empirical cdf and the normal cdf with the estimated μ and σ using the $n = 199$ data. Then $\sqrt{n}D$ has a known approximate distribution under the null hypothesis, and approximately,

$$P(\sqrt{n}D > 1.22) = .1, \quad P(\sqrt{n}D > 1.36) = .05, \quad P(\sqrt{n}D > 1.63) = .01. \tag{3.2}$$

In our case, the observed value of $\sqrt{n}D$ is about 1.06, which is not sufficient to reject the null hypothesis at any reasonable level.

We emphasize that the justifications given here are quite crude. In particular, the hypothesized independence of the day to day returns is difficult to reconcile with the well known herd mentality of stock investors. There is an extensive literature on models for stock prices that are much more sophisticated, though of course less easy to work with. From the point of view of the mathematical modeler, a mathematically simple model that yields approximately correct insights is certainly worthwhile, though one should always keep its limitations in mind.

Finally, here is the simulated stock price based on the same initial price as the earlier graph of QCOM, but using independent lognormal multipliers with the same drift and volatility as the QCOM data.



4. A continuous time version of the lognormal model

This time, we take the postulates from the beginning of the last section, and enhance them a bit so that the time parameter t may take any positive real value. It is in this application supposed to be a clock that ticks only during stock trading hours. We suppose S_t represents the dollar value of a particular stock at time t. Here are the spruced up hypotheses.

Definition 4.1. *We shall say that the stock price process $S_t(>0)$ follows a lognormal model provided the following conditions hold.*

(4.2) *For all $s, t \geq 0$, the random variable S_{t+s}/S_t has a distribution depending only on s, not on t. (Loosely speaking, a stock should have the same chance of going up 10% in the next hour, no matter what time we start at.)*

(4.3) *For any n, if we consider the process S_t at the times $0 < t_1 < \cdots < t_n$, the ratios S_{t_1}/S_0, S_{t_2}/S_{t_1} through $S_{t_n}/S_{t_{n-1}}$ are mutually independent. (Loosely speaking, a prediction of the stock price percentage increase from time t_{n-1} to time t_n should not be influenced by knowledge of the actual percentage increases during any preceding periods.)*

Let's examine the consequences of this definition. First of all, arguing just as in the preceding section, the distribution of S_t/S_0 is necessarily lognormal with some parameters (μ_t, σ_t^2). Let us define $\mu := \mu_1$ and $\sigma^2 := \sigma_1^2$, so that S_1/S_0 is lognormal (μ, σ^2). Now, for any integer $m > 0$

$$\frac{S_1}{S_0} = \frac{S_{1/m}}{S_0}\frac{S_{2/m}}{S_{1/m}}\cdots\frac{S_{m/m}}{S_{(m-1)/m}}$$

and by the first hypothesis in the definition, all the random variables on the right side have the same distribution, namely lognormal $(\mu_{1/m}, \sigma_{1/m}^2$. Moreover, by the second condition in the definition, they are mutually independent. As a product of independent lognormals is also lognormal, and the parameters add, we have

$$\mu = m\mu_{1/m}; \qquad \sigma^2 = m\sigma_{1/m}^2.$$

Consequently,

$$\mu_{1/m} = \frac{1}{m}\mu; \qquad \sigma_{1/m}^2 = \frac{1}{m}\sigma^2.$$

Similarly, we may write

$$\frac{S_{k/m}}{S_0} = \frac{S_{1/m}}{S_0}\frac{S_{2/m}}{S_{1/m}}\cdots\frac{S_{k/m}}{S_{(k-1)/m}}$$

and deduce by the same reasoning that

$$\mu_{k/m} = \frac{k}{m}\mu; \qquad \sigma_{k/m}^2 = \frac{k}{m}\sigma^2.$$

Writing this another way, we have proved that

$$\mu_t = t\mu; \qquad \sigma_t^2 = t\sigma^2 \quad \text{for } t \text{ of the form } k/m.$$

As the integers $k, m > 0$ are completely arbitrary, it follows (with some mild but unspecified assumption) that in fact

$$\mu_t = t\mu; \qquad \sigma_t^2 = t\sigma^2 \quad \text{for } t > 0. \tag{4.4}$$

As a consequence, we have shown:

Proposition 4.5. *If S_t satisfies Definition 4.1, then (a) S_1/S_0 lognormal with some parameters (μ, σ^2); (b) S_{t+s}/S_t is then lognormal $(s\mu, s\sigma^2)$.*

We further analyze this process by studying its natural logarithm $V_t := \log S_t/S_0$. The new process V_t clearly has the following properties:

(4.6) $V_0 = \log(S_0/S_0) = 0$;

(4.7) $V_t \sim N(t\mu, t\sigma^2)$;

(4.8) $V_{t+s} - V_t = \log\left(S_{t+s}/S_t\right) \sim N(s\mu, s\sigma^2)$, and so has the same distribution as V_s;

(4.9) If $0 < t_1 < \cdots < t_n$, then V_{t_1}, $V_{t_2} - V_{t_1}$ through $V_{t_n} - V_{t_{n-1}}$ are independent.

We make one further algebraic simplification, setting $B_t := \left(\frac{V_t - t\mu}{\sigma}\right)$, so that the process B_t has the following properties:

(4.10) $B_0 = 0$;
(4.11) $B_{t+s} - B_t \sim N(0, s)$, and so has the same distribution as B_s;
(4.12) If $0 < t_1 < \cdots < t_n$, then B_{t_1}, $B_{t_2} - B_{t_1}$ through $B_{t_n} - B_{t_{n-1}}$ are independent.

A process B_t, defined for $t \geq 0$, satisfing these conditions is called a **standard Brownian motion**. It may be proved (quite tricky proof, though) that the process may also be assumed to have continuous sample paths. That is, we may assume that for every ω, $t \to B_t(\omega)$ is continuous.

There is a substantial literature available based on this mathematical model of Brownian motion, and the methods developed to study it are fundamental to the study of mathematical finance at a more advanced level.

Now let's go back and write

$$V_t = t\mu + \sigma B_t.$$

That is, the process V_t has a uniform drift component $t\mu$ and a scaled Brownian component σB_t. Finally, we express S_t in terms of B_t by

$$\frac{S_t}{S_0} = \exp\bigl(V_t\bigr) = \exp\bigl(t\mu + \sigma B_t\bigr).$$

It is a consequence of this representation that for any $t, s \geq 0$, we have

$$\frac{S_{t+s}}{S_t} = \exp\bigl(s\mu + \sigma(B_{t+s} - B_t)\bigr). \tag{4.13}$$

5. A Black–Scholes formula

We now have all the tools available to perform a simple calculation that will prove to solve one of the option pricing problems to be studied later. We assume that the stock price process S_t satisfies (4.13), with given drift μ and volatility σ. Fix $x > 0$, $T > 0$ and let $s = S_0$ denote the price at time 0. We shall prove that

$$E(S_T - x)^+ = se^{T(\mu+\sigma^2/2)}\Bigl(\Phi\bigl(\frac{\log(s/x) + T\mu}{\sqrt{T}\sigma} + \sqrt{T}\sigma\bigr)\Bigr) - x\Bigl(\Phi\bigl(\frac{\log(s/x) + T\mu}{\sqrt{T}\sigma}\bigr)\Bigr). \tag{5.1}$$

We have $E(S_T - x)^+ = Eh(S_T/S_0)$, where $h(y) := (sy - x)^+$, which vanishes for $y < x/s$ and takes the value $(sy - x)$ for $y \geq x/s$. In view of (4.13), writing $B_T = \sqrt{T}Z$ where $Z \sim N(0, 1)$, we have $S_T/S_0 = \exp(T\mu + \sigma\sqrt{T}Z)$, so that

$$E(S_T - x)^+ = Eh(S_T/S_0) = E\bigl((\exp(\tau Z + \nu) - x)^+\bigr); \qquad \tau := \sqrt{T}\sigma; \quad \nu := T\mu + \log s.$$

From this we may calculate, since $e^{\tau Z+\nu} - x \geq 0$ if and only if $Z \geq \frac{\log x - \nu}{\tau}$,

$$E(S_T - x)^+ = \int_w^\infty \bigl(e^{\tau z+\nu} - x\bigr)\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz; \quad w := \frac{\log x - \nu}{\tau}.$$

The latter integral expands to

$$\int_w^\infty e^{\tau z+\nu}\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz - x\int_w^\infty \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz.$$

The second term reduces at once to $-x(1 - \Phi(w)) = -x\Phi(-w)$, and the first terms permits a completion of the square in the exponent to give

$$\int_w^\infty e^{\tau z+\nu}\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = e^{\nu+\tau^2/2}\int_w^\infty \frac{e^{-(z-\tau)^2/2}}{\sqrt{2\pi}}\,dz.$$

Changing the variable $u := z - \tau$ reduces this to

$$e^{\nu+\tau^2/2}\int_{w-\tau}^\infty \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,du = e^{\nu+\tau^2/2}\bigl(1 - \Phi(w - \tau)\bigr) = e^{\nu+\tau^2/2}\Phi(\tau - w).$$

Taking these components together yields finally

$$E(S_T - x)^+ = e^{\nu+\tau^2/2}\Phi(\tau - w) - x\Phi(-w), \tag{5.2}$$

and after substituting back for τ, ν, and noting that $w = -\frac{\log(s/x)+T\mu}{\sigma\sqrt{T}}$, this amounts precisely what was claimed in (5.1).

We shall show later, based on a "no arbitrage" argument, that if the bank interest rate is r, then

$$\mu + \frac{\sigma^2}{2} = r. \tag{5.3}$$

If we substitute $\mu = r - \sigma^2/2$ into (5.1), we find

$$E(S_T - x)^+ = se^{Tr}\Phi\Big(\frac{\log(s/x) + T(r-\sigma^2/2)}{\sqrt{T}\sigma} + \sqrt{T}\sigma\Big) - x\Phi\Big(\frac{\log(s/x) + T(r-\sigma^2/2)}{\sqrt{T}\sigma}\Big). \tag{5.4}$$

Even without the argument based on "no arbitrage", a simple argument may be given for (5.3). First of all, let's switch to stock prices based on present value, which is to say that we take the interest rate r to be the rate of inflation, fix a present time $t = t_0$, and let $U_t := e^{-r(t-t_0)}S_t$ denote the price of stock "measured in dollars as of time t_0." If, as in the preceding discussion, S_t/S_{t_0} is a lognormal process with drift μ and volatility σ, say (with B denoting a standard Brownian motion)

$$\frac{S_t}{S_{t_0}} = e^{\sigma B_{t-t_0} + \mu(t-t_0)},$$

then

$$\frac{U_t}{U_{t_0}} = e^{\sigma B_{t-t_0} + (\mu - r)(t-t_0)},$$

so that in fact, U_t/U_{t_0} is again a lognormal with the same volatility, but with drift $\mu - r$. However, in the financial world in which U_t represents a stock price, the interest rate is effectively 0. By our formula for the lognormal mean, we have

$$E\frac{U_t}{U_{t_0}} = e^{(t-t_0)(\mu - r + \sigma^2/2)}.$$

Fix $t > t_0$. If $\mu - r + \sigma^2/2 > 0$, we would have $E\frac{U_t}{U_{t_0}} > 0$, so a stock investment would be guaranteed to lead to greater wealth over the long run (by the law of large numbers). This is not consistent with the reality of a market. Similarly, if $\mu - r + \sigma^2/2 < 0$, a long term loss would be guaranteed, and no rational person would invest. For these reasons, one should assume that (5.3) is valid. (A "no arbitrage" argument depends on knowing more about trading possibilities, which we have not yet studied.)