INDIAN INSTITUTE OF TECHNOLOGY KANPUR

STOCK PRICE PREDICTION
USING
GEOMETRIC BROWNIAN MOTION

Course Name: IME625A
Course Instructor: Dr. Avijit Khanra
TA: Abhinav, Saurabh

Group Members:
Shivam Goel (180714)
Akhil Shukla (180057)
Rohan Gupta (170585)

Links:

GitHub Repository: IME625A Project Github Repo Link

Google Colab: IME625A Project Colaboratory Link

ABSTRACT

On the stock exchange, the stock is the most traded instrument. They are a source of income to the companies and investors, but they are intertwined with risk. Thus, predicting stock prices and that too with accuracy becomes very important. It can help to anticipate losses and thus provide optimal benefit to the investors.

In this project report, we start with the aim of predicting Stock prices using Geometric Brownian Motion (GBM). First, we establish the concept of random walk and, using this, define Brownian Motion. Then, we see Itô's Process and Itô's Lemma, which are used to find the differential of time-dependent Stochastic Process. After we have laid these groundworks, we obtain the expression for GBM, using Itô's Lemma, and state the correspondence with Brownian Motion with Drift. But the expression for GBM contains parameters, so using the historical data, we calculate these parameters. Then we see how these parameters can be estimated. We write the code of our GBM model and parameter estimation in Python, predict the stock prices and compare our result with the actual stock prices.

INTRODUCTION

To make critical financial decisions, simulated price paths of financial assets are often used to make price predictions. There are several models that are used for predicting stock prices. These models rely on two assumptions: stock prices may follow a past pattern, or they are completely random (like random walk).

Geometric Brownian Motion (or GBM) incorporates both theories while predicting stock prices through its two components:

- *certain*, which represents the return that the stock will earn over a short period, also known as the *drift* of the stock
- *uncertain*, which incorporates the idea of random walks and includes the stocks volatility and an element of random volatility

We first understand Brownian Motion, and from there, we read about GBM and how it is used to predict stock prices.

BROWNIAN MOTION (1D)

Let us consider a symmetric random walk, which in each unit time is equally likely to take a unit step either to the left or the right, i.e., it is a Markov Chain with $P_{i,i+1} = \frac{1}{2} = P_{i,i-1}$. Now suppose we speed up the process by taking smaller steps (Δx) in smaller time intervals (Δt). If we take its limiting behavior, we obtain Brownian Motion.

Formally, let us denote the position at time t by $X(t)$. Then

$$X(t) = \Delta x \left(X_1 + \ \dots\dots\dots + X_{\frac{t}{\Delta t}}\right),$$

where $X_i = \pm 1$ with equal probability

Assuming X_i to be independent, we obtain:

$$E[X_t] = 0,$$

$$Var\big(X(t)\big) = (\Delta x)^2 \left[\frac{t}{\Delta t}\right]$$

Now let $\Delta x \ = \ \sigma\sqrt{\Delta t}$, where σ is a positive constant and as $\Delta t \to 0$,

$$E[X_t] = 0,$$

$$Var\big(X(t)\big) \to \sigma^2 t$$

Thus, we get

$$X(t) = \sigma\sqrt{\Delta t} * (X_1 + \ \dots\dots\dots + X_{\frac{t}{\Delta t}})$$

$$X(t) = \sigma\sqrt{t} * \left(\frac{X_1 + \ \dots\dots\dots + X_{\frac{t}{\Delta t}}}{\sqrt{\frac{t}{\Delta t}}}\right)$$

As $\Delta t \to 0$, using the Central Limit Theorem, the quantity between brackets in the above equation tends to the standard normal distribution with mean zero and variance one. Thus, we can write:

$$X(t) = \sigma\sqrt{\Delta t} * \varepsilon ,$$

where ε is a random number drawn from normal distribution.

Using the above results and central limit theorem, we define $X(t)$ as *Brownian motion* process on $[0, T]$ if:

(i) $X(0) = 0$

(ii) The mapping $t \to X(t)$, with probability 1, is a continuous function on $[0, T]$

(iii) $X(t)$ has stationary and independent increment properties

(iv) $X(t) \sim N(0, \sigma^2 t)$ for any $0 < \text{t} \leq \text{T}$

Suppose $X(t)$ is normal with mean 0 and variance t, its density function is given by

$$f_t(\text{x}) = \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}}$$

BROWNIAN MOTION WITH DRIFT

We call a process $X(t)$ a Brownian motion with drift μ and diffusion coefficient σ^2 if

$$W_t = \frac{X(t) - \mu t}{\sigma}$$

is a standard Brownian motion. Brownian Motion with drift is represented as $X \sim BM(\mu, \sigma^2)$.
For time-dependent drift and diffusion coefficient, Brownian Motion may be defined by the following equation:

$$dX(t) = \mu_t * dt + \sigma_t * dW_t$$

The process X(t) has continuous sample paths and independent increments, and each increment $X(t) - X(s)$ is normally distributed with:

Mean: $E[X(t) - X(s)] = \int_s^t \mu(k) * dk$, and

Variance: $Var[X(t) - X(s)] = \int_s^t \sigma^2(k) * dk$

ITÔ'S PROCESSES

An Itô process is defined to be a stochastic process that can be expressed as the sum of an integral with respect to Brownian Motion (W_t) and an integral with respect to time,

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

Itô's Lemma [1]– It is an identity used in calculus to find the differential of a time-dependent function of a Stochastic Process. Using Taylor Series expansion, for a function f which is a function of x and t, we can write:

$$df_{x,t} = \frac{\partial f_{x,t}}{\partial t} dt + \frac{\partial f_{x,t}}{\partial x} dx + \frac{1}{2}\left(\frac{\partial^2 f_{x,t}}{\partial x^2} dx^2 + \cdots\right) \ldots$$

Let us assume X_t is an Itô's process. Thus, we substitute x by X_t and dx by $\mu_{x,t} * dt + \sigma_{x,t} * dW_t$.
Note: In this case, μ and σ are a function of both X_t and t (that is why two variables in the subscripts).

$$df_{x,t} = \frac{\partial f_{x,t}}{\partial t} dt + \frac{\partial f_{x,t}}{\partial x}\left(\mu_{x,t} dt + \sigma_{x,t} dW_t\right) + \frac{1}{2}\left(\frac{\partial^2 f_{x,t}}{\partial x^2}\left(\mu_{x,t}^2 dt^2 + 2\mu_{x,t}\sigma_{x,t} dt * dW_t + \sigma_{x,t}^2 dW_t^2\right)\right) + \cdots$$

Ignoring higher-order terms in the limit $dt \to 0$ and substituting dt for dW_t due to the quadratic variance of Brownian Motion[2], we obtain

$$df_{x,t} = \left(\frac{\partial f_{x,t}}{\partial t} + \mu_{x,t} \frac{\partial f_{x,t}}{\partial x} + \frac{\sigma_{x,t}^2}{2} \frac{\partial^2 f_{x,t}}{\partial x^2}\right) dt + \sigma_{x,t} \frac{\partial f_{x,t}}{\partial x} dW_t$$

[1] Link for rigorous proof in Additional Reading Section
[2] Link for more theoretical details in Additional Reading Section

GEOMETRIC BROWNIAN MOTION (GBM)

A stochastic process S_t is a geometric Brownian motion if $log(S_t)$ is a Brownian Motion with initial value $log(S_0)$. We write the GBM Stochastic Differential Equation (SDE) as:

$$dS_t = \mu S_t dt + \sigma S_t dW$$

In Geometric Brownian Motion, $\frac{S(t_2)- S(t_1)}{S(t_1)}, \frac{S(t_3)- S(t_2)}{S(t_2)}, \dots \dots \dots , \frac{S(t_n)- S(t_{n-1})}{S(t_{n-1})}$ are independent, for $t_1 < t_2 < t_3 \dots \dots \dots < t_n$, rather than $S(t_{i+1}) - S(t_i)$.

One of the advantages of GBM is that it is always positive. This property makes it desirable in stock price modelling.

Now, in Itô's Lemma, we put $f(S_t)) = \log(S_t))$. Here, f is only a function of x and not a function of t, $\frac{\partial f_x}{\partial t} = 0$ and $\frac{\partial f_x}{\partial x} = \frac{1}{x}$. Also, in Geometric Brownian Motion, $\mu_{x,t} = S * \mu_t$ and $\sigma_{x,t} = S * \sigma_t$. Putting these values in the SDE, we get

$$d(log\ (S_t)) = \left(0 + S_t \mu_t * \frac{1}{S_t} + S_t^2 \frac{\sigma_t^2}{2} * \left(-\frac{1}{S_t^2}\right)\right) dt + S_t \sigma_t * \frac{1}{S_t} dW_t$$

$$d(log(S_t)) = (\mu_t - \frac{\sigma_t^2}{2})dt + \sigma_t dW_t$$

From the above equation, we can say $d(\log(S_t))$ has a Brownian motion with drift. And thus, it has a normal distribution with mean, $\left(\mu_t - \frac{\sigma_t^2}{2}\right) t$ and variance, $(\sigma_t^2 t)$. Also, we can substitute $dW_t = \varepsilon\sqrt{dt}$

$$\log(S_t) - \log(S_{t-1}) = \left(\mu_t - \frac{\sigma_t^2}{2}\right) dt + \sigma_t \varepsilon \sqrt{dt}$$

$$\log\left(\frac{S_t}{S_{t-1}}\right) = \left(\mu_t - \frac{\sigma_t^2}{2}\right) dt + \sigma_t \varepsilon \sqrt{dt}$$

$$\frac{S_t}{S_{t-1}} = e^{\left(\mu_t - \frac{\sigma_t^2}{2}\right) dt + \sigma_t \varepsilon \sqrt{dt}}$$

$$S_t = S_{t-1} e^{\left(\mu_t - \frac{\sigma_t^2}{2}\right) dt + \sigma_t \varepsilon \sqrt{dt}}$$

Extending the above GBM model to any period of time t:

$$S_t = S_0 e^{\left(\mu_t - \frac{\sigma_t^2}{2}\right) t + \sigma_t \varepsilon \sqrt{t}}$$

PARAMETER ESTIMATION

To apply GBM, we need two parameters: drift and volatility of the stock. If they are not already known, they need to be estimated from the historical data.

Let S_i denote the stock closing price at the end of an i^{th} trading period, and τ be the length of the time interval between two consecutive trading periods expressed in years, $\tau = t_i - t_{i-1}$.

If u_i is the logarithm of the daily return on the stock return over time τ, i.e.,

$$u_i = ln\left(\frac{S_i}{S_{i-1}}\right)$$

Then the unbiased estimator $\bar{u}$ of the logarithm of the returns is given by

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$$

From the formula of GBM, we concluded that $d(\log(S_t))$ has a normal distribution with mean, $(\mu_t - \frac{\sigma_t^2}{2})t$ and variance, $(\sigma_t^2 t)$. Using the above result, let's obtain the value of these parameters and see what they represent:

1. VOLATILITY: The meaning of volatility is variation or spread of the distribution. Thus, the value of volatility is always positive or zero because it is related to the standard deviation of the distribution. It also gives an idea about the stability of stock price. The most usual method of measuring the stock volatility is the standard deviation of the price returns. To do this, we observe the historical data at fixed intervals of time, such as the daily closing time, and use the results we obtained above.

   The estimator of the SD of the u_i's is given by

$$\nu = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2}$$

   Now σ can be simply estimated by:

$$\sigma = \frac{\nu}{\sqrt{\tau}}$$

2. DRIFT: The meaning of drift parameter is trend or growth rate. If the drift is positive, the trend is going up over time. If the drift is negative, the trend is going down. Equating the mean $\bar{u}$ and $\left(\mu - \frac{\sigma^2}{2}\right)\tau$:

$$\bar{u} = \left(\mu - \frac{1}{2}\sigma^2\right)\tau$$

$$\Longrightarrow \quad \mu = \frac{\bar{u}}{\tau} + \frac{1}{2}\sigma^2$$

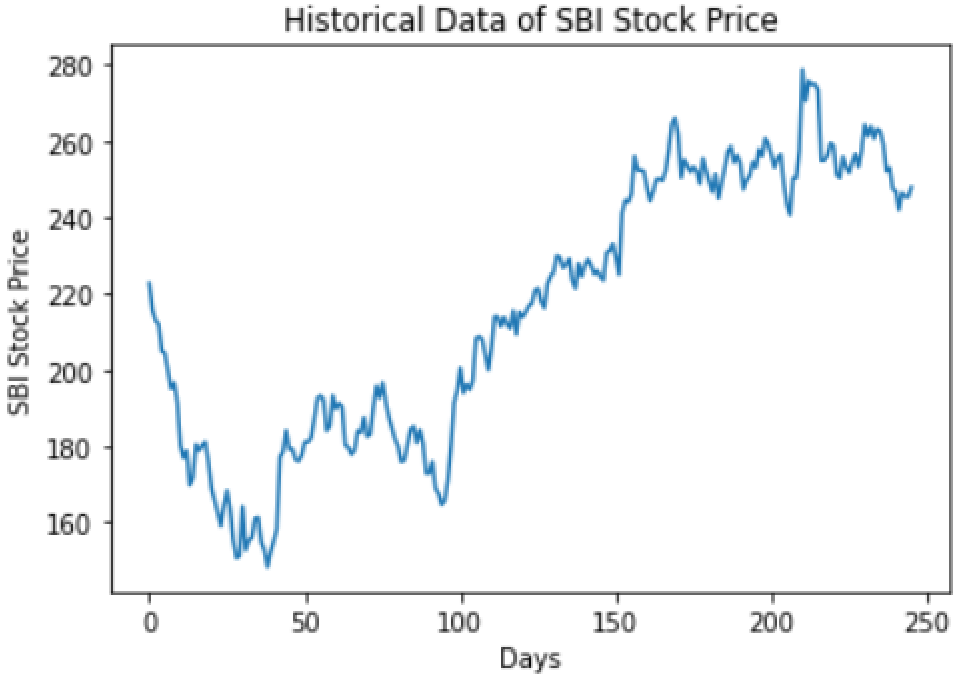
GBM MODEL IMPLEMENTATION

For testing, we took the data of stock prices of SBI from the year 2016 and using that as our historical data, we try to predict the stock prices for the month of January 2017. Then we compared the predicted stock prices with the actual stock prices. The code implementing GBM is as follows:

```
"""##**PARAMETER ESTIMATION**"""

prices = np.array(data)
U_list = np.log(prices[1:]/prices[:-1])
n = len(prices) - 1
Eu = sum(U_list) / float(n)
sigma2 = np.sum((U_list - Eu)**2) / ((n-1))
mu = Eu + sigma2 / 2.0
print(f'\u03bc = {mu}\n\u03c3^2 = {sigma2}\n')

"""# **SIMULATION**"""

def stock_price(data, days, num_samples):

   samples = []
   for num in range(num_samples):
      S_prev = prices[-1]
      for i in range(days):
         S = S_prev * np.exp((mu - sigma2/2) + np.sqrt(sigma2)*np.random.normal(0,1))
         S_prev = S
      samples.append(S)

   samples = np.array(samples)
   expected_price = np.mean(samples)
   return expected_price

days = np.arange(1, len(actual_prices)+1,1)
num_samples = 50
np.random.seed(0)
prices = np.array(data)
expected_prices = []
for i in range(0,len(days)):
   expected_prices.append(stock_price(data, days[i], num_samples))
expected_prices = np.array(expected_prices)
expected_prices = np.insert(expected_prices,0,prices[-1])
actual_prices = np.insert(actual_prices,0,prices[-1])

"""##**DATA COMPARISON**##"""

plt.plot(actual_prices, label = 'Actual')
plt.plot(expected_prices, label = 'GBM', ls='--')

plt.ylabel('SBI Stock Price')
plt.xlabel('Prediction Day')
plt.title('Geometric Brownian Motion - SBI')

plt.legend(loc = 'upper left')
```

RESULTS

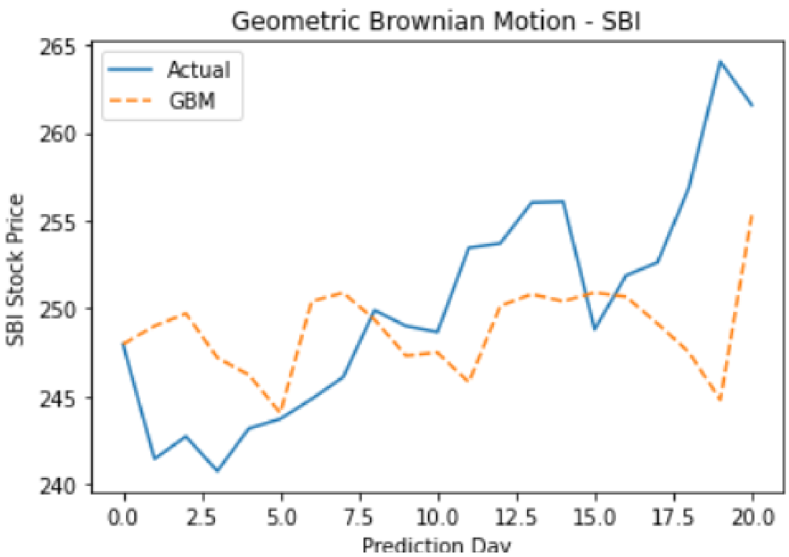
The distribution of historical data over 1 year is as follows:



From the historical data shown above, we estimate our required parameters and get the following values:

$$\mu = 0.0007253773919 52974$$
$$\sigma^2 = 0.0005785778513477582$$

To predict the price for a particular day, we take 50 samples for that day and assign the average as the stock price. We do this for all trading days of January and then plot them along with the actual data. The resulting graph is as follows:



We can get a gist of the stock prices in the future using the GBM model.

REFERENCES

1. Introduction to Probability Models(10th edition) by Sheldon Ross
2. Reddy, Krishna and Clinton, Vaughan, Simulating Stock Prices Using Geometric Brownian Motion: Evidence from Australian Companies, Australasian Accounting, Business and Finance Journal, 10(3), 2016, 23-47 [Link]
3. AIP Conference Proceedings 2242, 030016 (2020) [Link]
4. Blog on Random Walk: Introduction, GBM, Simulation [Link]
5. Wikipedia [Link]
6. Ito's Lemma [Link]

FOR ADDITIONAL READING:

1. Rigorous Proof of Itô's Lemma [Link]
2. Quadratic Variation of Brownian Motion [Link]