

Log-Normal From Scratch

Assumptions

Prior to exploring our predictions, it is necessary and helpful to introduce relevant mathematical context. We need to establish key assumptions in order for our predictions to possess relevant results.

1. Efficient Market Hypothesis - Daily Percent Change in Stock Prices Are Independent. Let X_i and X_j be the following

$$X_i \sim \frac{S_{i+1} - S_i}{S_i} \quad X_j \sim \frac{S_{j+1} - S_j}{S_j} \quad (1)$$

where S_i is the stock price on day i .

Everyone trades on day i with all information up to that point, therefore the only information affecting the percent change in price from day i to day $i + 1$ is the information contained by day $i + 1$. Similarly for some $j \neq i$, we have that everyone trades on day j with all information up to that point, therefore the only information affecting the percent change in price from day j to day $j + 1$ is the information contained by day j . Thus since $i + 1 \neq j + 1$, we have that the information affecting the percent change in price for different days (i.e. X_i, X_j) is always independent.

2. Assume that the percent changes in stock prices for different non-overlapping time periods with the same interval of time are identically distributed (i.e. they have the same pdf's).
3. Stock price is continuous

Log-Normality

Now we can get into some prediction formulas. Based on our previous assumptions, the transitive property allows us to say that all X_i s are independent from one another and identically distributed.

Definitions

1. X_i is defined as under assumptions above.
2. $\forall i : E[X_i] = \mu$, since X_i are I.I.D
3. $\forall i : \text{Var}(X_i) = \sigma^2$, since X_i are I.I.D
4. $T, t \in \mathcal{R}, t < T, S_t =$ price at closing on day t (usually $t=0$, so its just the starting price of the stock),
 $S_T =$ price at closing on day T

S_T can be written in terms of μ , σ , T , t , and $Z \sim \mathcal{N}(0, 1)$

$$S_T = S_t \cdot \frac{S_{t+1}}{S_t} \cdot \frac{S_{t+2}}{S_{t+1}} \cdot \frac{S_{t+3}}{S_{t+2}} \dots \frac{S_T}{S_{T-1}}$$

by Taylor expansion around 1, $\ln x \approx \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} \cdot (x-1)^i \approx (x-1)$

$$\begin{aligned} \frac{S_T}{S_t} &= \frac{S_{t+1}}{S_t} \cdot \frac{S_{t+2}}{S_{t+1}} \cdot \frac{S_{t+3}}{S_{t+2}} \dots \frac{S_T}{S_{T-1}} \\ \ln \left(\frac{S_T}{S_t} \right) &= \ln \left(\frac{S_{t+1}}{S_t} \cdot \frac{S_{t+2}}{S_{t+1}} \cdot \frac{S_{t+3}}{S_{t+2}} \dots \frac{S_T}{S_{T-1}} \right) \\ &= \ln \left(\frac{S_{t+1}}{S_t} \right) + \ln \left(\frac{S_{t+2}}{S_{t+1}} \right) + \ln \left(\frac{S_{t+3}}{S_{t+2}} \right) + \dots + \ln \left(\frac{S_T}{S_{T-1}} \right) \\ &\approx \left(\frac{S_{t+1}}{S_t} - 1 \right) + \left(\frac{S_{t+2}}{S_{t+1}} - 1 \right) + \left(\frac{S_{t+3}}{S_{t+2}} - 1 \right) + \dots + \left(\frac{S_T}{S_{T-1}} - 1 \right) \\ &= \frac{S_{t+1} - S_t}{S_t} + \frac{S_{t+2} - S_{t+1}}{S_{t+1}} + \frac{S_{t+3} - S_{t+2}}{S_{t+2}} + \dots + \frac{S_T - S_{T-1}}{S_{T-1}} \\ &= X_t + X_{t+1} + X_{t+2} + \dots + X_T \end{aligned}$$

We know from our assumptions that each of the terms in the sum above are independent and identically distributed. Therefore, by the central limit theorem, if we assume T is sufficiently larger than t , we can say that the sum of these I.I.D random variables follows a normal distribution with mean $E[X_i] \cdot n = n\mu$ and variance $n \cdot \text{Var}(X_i) = n \cdot \sigma^2$, where $n = T - t$. So we write

$$\ln \left(\frac{S_T}{S_t} \right) \approx \mathcal{N}(n\mu, n\sigma^2)$$

This formulation is why the system is referred to as a **log-normal** model. From here, we can continue with the calculations.

$$\begin{aligned} \ln \left(\frac{S_T}{S_t} \right) &\approx \mathcal{N}(n\mu, n\sigma^2) \\ &= n \cdot \mu + \sqrt{n} \cdot \sigma \cdot Z, \text{ where } Z \sim \mathcal{N}(0, 1) \\ \ln(S_T) - \ln(S_t) &\approx n \cdot \mu + \sqrt{n} \cdot \sigma \cdot Z \\ \ln(S_T) &\approx \ln(S_t) + n \cdot \mu + \sqrt{n} \cdot \sigma \cdot Z \\ S_T &\approx e^{\ln(S_t) + n \cdot \mu + \sqrt{n} \cdot \sigma \cdot Z} \\ &= S_t \cdot e^{n \cdot \mu + \sqrt{n} \cdot \sigma \cdot Z} \\ &= S_t \cdot e^{(T-t) \cdot \mu + \sqrt{T-t} \cdot \sigma \cdot Z} \end{aligned}$$

In conclusion,

$$S_T \approx S_t \cdot e^{(T-t) \cdot \mu + \sqrt{T-t} \cdot \sigma \cdot Z}$$

In reality, we will have a "real rate of return", call it r , so

$$S_T = S_t \cdot e^{r \cdot (T-t)}$$

Thus we would expect the following

$$\begin{aligned}
S_t * e^{r \cdot (T-t)} &= E \left[S_t \cdot e^{(T-t) \cdot \mu + \sqrt{T-t} \cdot \sigma * Z} \right] \\
&= S_t \cdot e^{(T-t) \cdot \mu} \cdot E \left[e^{\sqrt{T-t} \cdot \sigma * Z} \right] \\
&= S_t \cdot e^{(T-t) \cdot \mu} \cdot M_Z \left(\sigma \cdot \sqrt{T-t} \right)
\end{aligned}$$

Where $M_Z(t)$ is the moment generating function of the standard normal random variable.

$$\begin{aligned}
&= S_t \cdot e^{(T-t) \cdot \mu} \cdot e^{\frac{(\sigma \cdot \sqrt{T-t})^2}{2}} \\
&= S_t \cdot e^{(T-t) \cdot \mu} \cdot e^{\frac{\sigma^2 \cdot (T-t)}{2}} \\
&= S_t \cdot e^{(T-t) \cdot \left(\mu + \frac{\sigma^2}{2} \right)} \\
S_t * e^{r \cdot (T-t)} &= S_t \cdot e^{(T-t) \cdot \left(\mu + \frac{\sigma^2}{2} \right)} \\
\implies r &= \mu + \frac{\sigma^2}{2} \\
\implies \mu &= r - \frac{\sigma^2}{2}
\end{aligned}$$

Thus, we say

$$S_T \approx S_t \cdot e^{(T-t) \cdot \left(r - \frac{\sigma^2}{2} \right) + \sqrt{T-t} \cdot \sigma * Z}$$

Where r and σ are the real rate of return and standard deviation of daily returns. However, it is impossible to know what these truly are and so the best we can do is estimate through historical data. We can also run multiple iterations of estimating future stock prices to get a better mean estimate. This is the motivation behind using Monte-Carlo simulations and random walks.

Brownian Motion

It should also be noted that we can frame our result in terms of Brownian motion. Recall $\mathcal{N}(a, b)$ means a normal distribution with mean a and standard deviation b .

$$\begin{aligned}
B_t &\sim \mathcal{N} \left(0, \sigma \sqrt{T-t} \right) \\
S_T &\approx S_t \cdot e^{(T-t) \cdot \left(r - \frac{\sigma^2}{2} \right) + B_t}
\end{aligned}$$

For daily returns $T - t = 1$, so we get

$$\begin{aligned}
B_t &\sim \mathcal{N} (0, \sigma) \\
S_t &\approx S_{t-1} \cdot e^{\left(r - \frac{\sigma^2}{2} \right) + B_t}
\end{aligned}$$

Monte Carlo Simulations

We can observe from the assumptions that daily returns, in percentages, are I.I.D. Therefore, by the central limit theorem if we take enough days n , then the sum of these daily returns will follow a normal distribution. In particular this normal distribution will have a mean of $n \cdot E[X_i]$ and a variance of $n \cdot \text{Var}(X_i)$. More explicitly we have

$$X_t + X_{t+1} + X_{t+2} + \dots + X_T \sim \mathcal{N} (n \cdot E[X_i], n \cdot \text{Var}(X_i)).$$

Thus, where $n = T - t$,

$$\sum_{i=0}^n X_i \sim \mathcal{N}(n \cdot E[X_i], n \cdot \text{Var}(X_i)).$$

Since, by definition the mean daily return would be

$$\frac{\sum_{i=0}^n X_i}{n}.$$

Therefore our mean daily return in expectation follows a distribution like the one below

$$\frac{\sum_{i=0}^n X_i}{n} \sim \mathcal{N}\left(E[X_i], \frac{\text{Var}(X_i)}{n}\right).$$

Now the trick is to identify $E[X_i]$ and $\text{Var}(X_i)$. For $E[X_i]$ we will simply take the historical sample mean percent change in daily price (mean daily returns). This is motivated by the fact that the sample mean is the Minimum-Variance Unbiased Estimator for μ . For $\text{Var}(X_i)$ we will similarly take the sample variance of daily percent changes in stock price (variance of daily returns). To recap, we now have the distribution Q we would expect our daily returns to follow to be a normal distribution with mean:

$$\frac{\text{historical mean annual returns}}{252}$$

and standard deviation:

$$\frac{\text{historical standard deviation of returns}}{\sqrt{252}}$$

We can use these facts to conduct a random walk, that is starting from our initial stock price, call it S_0 , we can iterate T times and increase S_0 by our expected returns each time. In other words, $S_T = S_0 * Q_0 * Q_1 * Q_2 * \dots * Q_{T-1}$, where Q_i is a sampling of our distribution of daily returns as described above. If we record each S_i up to S_T and plot them we will get a random walk. However, one random walk is not particularly insightful. What we want to do is generate many random walks in a Monte Carlo simulation. That is we generate random walks as described above N times, then we do things like take the mean final price of all these random walks and expect this value to be more insightful than the final price of one random walk.