Probability, Monte Carlo Simulation & Application in Finance – Day 4 (Risk)

Warm Up Problem: Suppose there is an amoeba. Every minute, this amoeba can either die, do nothing, split into 2, or split into 3 amoebas; all these scenarios being equally likely to happen. All further amoebas behave the same way:

1. The expected value of a random variable X with possible outcomes (values) x_1, x_2,……x_k with respective probabilities p_1, p_2,……p_k is the probability weighted sum of each possible value. It is also referred by few other ways, such as the mean value (mu), E[X], X bar, the first moment.

   $\mu \equiv \mathrm{E}[X] = \overline{X} = \langle X \rangle$ = $x_1 * p_1 + x_2 * p_2 + \ldots\ldots\ldots + x_k * p_k$

   What is the expected number of amoebas after 1 minute & 2 minutes?

2. The *moments* of a random variable X are the expected value of the powers of the random variable itself.

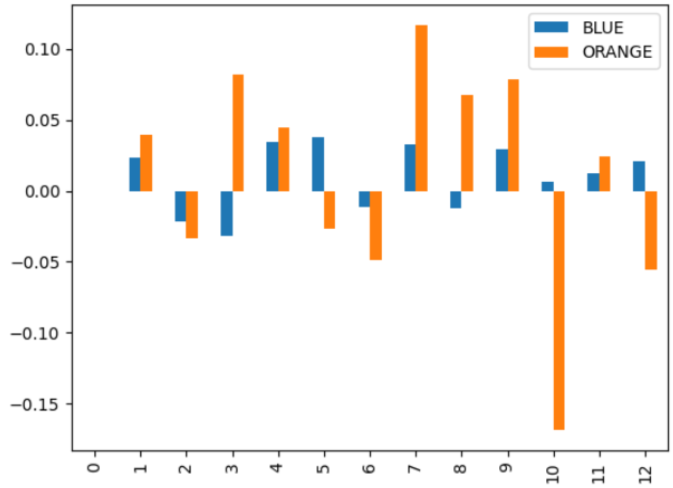
   $\mu_\ell \equiv \mathrm{E}\left[X^\ell\right] \equiv \langle X^\ell \rangle$ = $x_1^\ell * p_1 + x_2^\ell * p_2 \ldots\ldots\ldots\ldots\ldots\ldots + x_k^\ell * p_k$

   The 2nd moment of random variable X will be equal to: $x_1^{\mathbf{2}} * p_1 + x_2^{\mathbf{2}} * p_2 \ldots\ldots\ldots\ldots\ldots\ldots + x_k^{\mathbf{2}} * p_k$

   What is the second moment of the number of amoebas after 1 minute & 2 minutes?

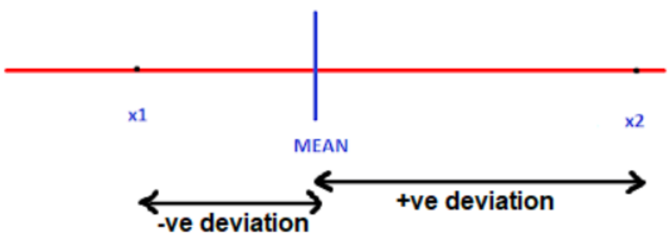
3. What is the probability that the entire population of amoeba will eventually die off?

Measure of Risk: Standard Deviation



Look at the bar chart (from last lesson) of the two series of returns above: one blue and one orange. Both have the same average return, but the blue series is much less volatile than the orange series. This means the orange series deviates from the average more often and more significantly than the blue series.

Variance of a population is a way to measure how spread-out individual values in the population are from its average value.



The variance of X is the expectation of X minus its mean, the whole quantity squared. We look at X minus mean to measure how much X deviates from its mean value. We take the square so that we don't have positive and negative fluctuations canceling each other out. Looking more closely, we can see variance is the second moment of the deviation from mean.

$$\sigma^2 = \text{Var}(X) \equiv \text{E}\left[(X-\mu)^2\right] = \sum_k (x_k - \mu)^2\, p(x_k)$$

Problem (a): Let X be a variable that always takes a value of m. Find its variance.

Problem (b): Let X be a random variable that either takes the value X=1 with probability 1/3 or the value X=10 with probability 2/3. What is the variance of X?

Note that if all possible values of variable X are equally probable, then the variance will be just the average of all values of the deviation from mean. To make the variance comparable to the original unit of the population, we take the positive square root of the variance to find 'Standard Deviation", also a key measure of volatility/risk in finance.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

When a wide-range of outcomes are possible, the likelihood of negative outcomes increases, which is the reason for expressing financial risk in terms of volatility, such as standard deviation. Higher volatility means more deviations from the mean. If returns numbers are pretty close to each other, the standard deviation will be small. But if return numbers are "volatile", i.e., few are very large or small compared to others, then the standard deviation will be high. Assets with high "volatility" (high standard deviation of returns) will have high risk while assets with low volatility (low standard deviation of returns) will have low risk ("Safe Haven" assets).

Problem (c): Let X be a random variable and let f(X) and g(X) be functions of X. Recall that expected value of f(X) i.e. $E[f(X)] = f(X_1)*p_1 + f(X_2)*p_2 + \dots\dots\dots + f(X_k)* p_k$, where p_k is the probability of $f(X_k)$.

Based on definition above for expected value, show that:

1. The expectation of any constant, c, times any function f, is just the constant times the expectation.
   $\mathrm{E}[cf(X)] = c\,\mathrm{E}[f(X)]$

2. If we have the sum of any two functions, f and g, the expectation of the sum is the sum of the expectations.
   $\mathrm{E}[f(X) + g(X)] = \mathrm{E}[f(X)] + \mathrm{E}[g(X)]$

3. A function that takes other functions as inputs, such as E[f(X)], is often called an "operator". Note that the expectation operator E here preserves addition and multiplication by constants. Such an operator is called linear. Use this linearity property to derive variance as the expected value of X squared minus the square of the expected value.

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

Problem (d): Find the volatility of the gold returns in below table (it's the same gold prices as in earlier lesson on returns)

Year	Average Gold Price (USD per ounce)	Return (%)	Return – μ	(Return – μ)^2
2024	$2,662.28	30.2		
2023	$2,045.00	13.6		
2022	$1,800.00	1.7		
2021	$1,770.00	0		
2020	$1,770.00	27.2		
2019	$1,392.00	9.8		
2018	$1,268.00	0.9		
2017	$1,257.00	0.6		
2016	$1,250.00	7.8		
		μ= ?		∑ = ?

σ =?

Problem (e): What is the unit of the volatility calculated above?

Problem (f): If the volatility of Oil in these same 10-year period is 34%, which one out of oil and gold is a more volatile asset? Which one is riskier asset? Which one is a safe haven for investment?

Problem (g): Below is a python example of finding standard deviation. Find the standard deviation in python for the gold returns above.

```
import numpy as np
data = [10, 15, 20, 25, 30]
# Compute Standard Deviation
data_std_dev = np.std(data)
print("Standard Deviation of Data:", data_std_dev)
Standard Deviation of Data: 7.0710678118654755
```

Problem (h): A teacher can use the standard deviation of marks of her students in an exam as a metric to assess the overall level of understanding of the subject. Fill out the last column in below table on how a teacher can interpret the class performance based on the mean and standard deviation pattern:

Mean Score of the Class	Standard Deviation of the Class	Class Performance
Low	Low	
Low	High	
High	Low	
High	High	

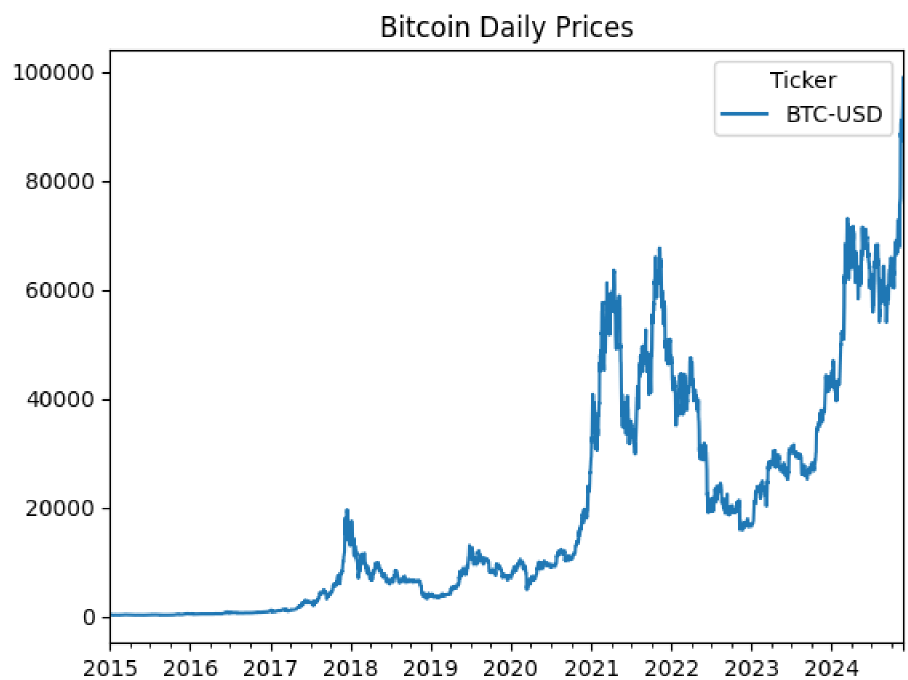
Problem (i): Some countries do not have a central bank. For example, Panama and Ecuador use the US dollar as their official currency and do not have a central bank that issues their own currency. Can you guess potential reasons or benefits for doing so?

Problem (j) Bitcoin Volatility: Imagine you have Minecraft coins or reward points which you can use within the game to buy skins, textures, and other cool stuff. What if you can convert them into actual money and use outside the Minecraft world, say to buy things, send money to friends, or even invest in them - Cryptocurrency is one such type of digital or virtual money that allows this and uses cryptography for security. Unlike traditional money issued by governments, cryptocurrencies operate on technology called blockchain, which is a decentralized network of computers that ensures transactions are secure and transparent. Bitcoin is the first and most well-known cryptocurrency. It was created in 2009 by an unknown person (or group) using the name Satoshi Nakamoto. Bitcoin is often referred to as "digital gold" because it's limited in supply (only 21 million bitcoins will ever exist).

Below code shows how to pull bitcoin prices using Yahoo Finance and plots its daily prices. Find the daily return, create a line chart of daily returns. Find volatility of daily returns for the whole period.

```
import pandas as pd
import yfinance as yf

df = yf.download('BTC-USD',
                      start='2015-01-01',
                      progress=False)
df.Close.plot(title='Bitcoin Daily Prices')
```



Problem (k): Below code shows the volatility numbers for Gold price returns from earlier table. Compare the volatility numbers with that of Bitcoin calculated above. Does the result make sense? Which one is more volatile? Which one will you prefer to invest in?

```
# Find volatility of Gold returns
import numpy as np
data = [30.2,13.6,1.7,0,27.2,9.8,0.9,0.6,7.8]
# Compute Standard Deviation
data_std_dev = np.std(data)
print("Standard Deviation of Data:", data_std_dev)
Standard Deviation of Data: 10.85
```

Annualizing Volatility

Problem (l): Imagine you roll two fair six-sided dice. Let X be the outcome of the first die and Y be the outcome of the second die.

1. Calculate the variance of X

2. Calculate the variance of Y

3. Calculate the variance of X+Y

4. Since these die rolls are independent of each other (i.e., the occurrence in first die roll doesn't impact the occurrences in second die roll), the variance of their sum is the sum of their individual variances. Select True/False

When we compare the volatility of different assets, like gold and Bitcoin, we need to make sure we're looking at the same time period. For example, if we have daily returns for Bitcoin and annual returns for gold as in our case, we need to adjust the gold volatility to an annual basis to compare them.

Assuming the variance of a sum of independent random variables (here daily returns) equals the sum of the individual variances:

$$\sigma_A^2 = Var[R_A] = Var\left[\sum_{t=1}^{T} R_{d,t}\right] = \sum_{t=1}^{T} Var[R_{d,t}] = T \cdot Var[R_d] = T \cdot \sigma_d^2$$

New Standard Deviation = Old Standard Deviation * √(New Period Length / Old Period Length)

Problem (l): We annualize volatility by scaling (multiplying) it by the square root of the number of periods per observation. 'Annualize' the Bitcoin volatility calculated based on daily returns:

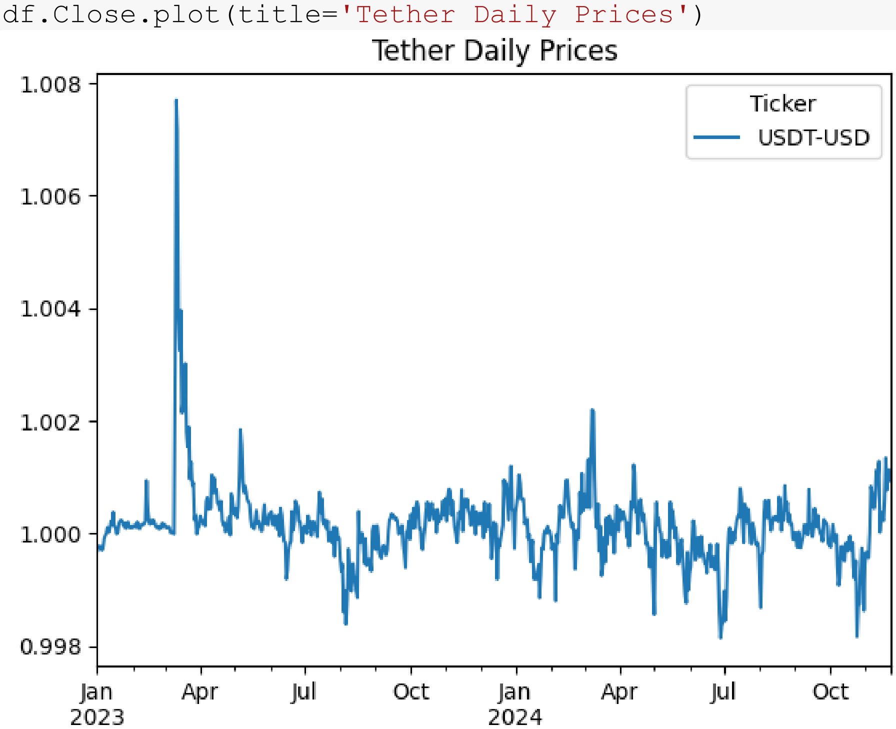
1. What is the old (daily) period length (in number of days)? ___________
2. What will be the new period (annual) length (in number of days)? ____________
3. Find the new (annual) standard deviation

4. Compare the annual volatility numbers of Gold and Bitcoin. Does the result make sense now? Which one is more volatile? Which one will you prefer to invest in?

Problem (m): Stablecoins are a type of cryptocurrency designed to have a stable value, usually pegged to a currency like the US dollar. They are supposed to provide stability in the otherwise volatile cryptocurrency market. Tether is one of the most popular stablecoins. It's pegged to the US dollar, meaning its value is always around one dollar. Can you guess what will be its volatility? Validate your answer using data from Yahoo Finance.

```
import pandas as pd
import yfinance as yf

df = yf.download('USDT-USD',
                      start='2023-01-01',
                      progress=False)
df.Close.plot(title='Tether Daily Prices')
```




```
(df.tail(252).Close.pct_change()*100).std()*np.sqrt(252) #Annualized
volatility of Tether
```

Risk Adjusted Returns

Problem (n) Compare returns that have different risks: Take a look at below graph of returns of 2 different assets – US Small Caps and US Large Caps. Clearly US Small caps are far more volatile but they also give a higher return. How do we compare these two and which one will you invest in? (Let's call it your Try 1 of Investment Decision Making)

Problem (o): One way to think about is to see how much return you get for each unit of risk we take for each asset. Calculate this "Return on Risk" ratio for each of these assets. Compare them and decide which asset you will invest in. (Let's call it your Try 2 of Investment Decision Making)

Problem (p) Risk Free Rate of Return:

Government needs money for a lot of things, like building and maintaining infrastructure (roads, bridges, schools), funding public services (healthcare, education, defense), paying off debt, etc. Taxes are a major source of income for the government, but they aren't always enough to cover all the expenses. To fill this gap, they also borrow money from public by issuing 'debt or fixed income' securities:

1. **Treasury Bills (T-Bills)**: These are short-term debts that expires or 'matures' in a year or less. They do not pay an interest but are sold at a discount (lower value than the amount repaid at the end of period).
2. **Treasury Notes (T-Notes)**: These are medium-term debts that expires or 'matures' in 2 to 10 years. They pay interest every six months.
3. **Treasury Bonds (T-Bonds)**: These are long-term debts that expires or 'matures' in 20 to 30 years. They also pay interest every six months.

There is always a pre-defined amount of money that the issuer will pay back to the investor when the asset matures or expires – this is known as 'face value' or 'par value' or 'principal'. The interest rate paid by notes or bonds is also called as 'coupon'. Bonds which don't pay any interest (as we see in case of US Treasury bills) are called as 'zero-coupon bonds'.

Fixed income securities like US treasuries in general have a face value of $1,000 and prices are quoted in 100. For example, if the price of $1,000 13-week treasury bill is 98.896250. So, anyone investing/buying this bill will have to pay (98.896250/100) *1000=988.9625$. At the end of 13-week, the investor will receive $1,000 back. Find:

1. The rate of return for this treasury bill. What will be your initial value/investment amount?

2. Note that the above return is for 13-week period. Find the annual rate of return.

This annual rate of return (also called 'yield') on 3-month (13 weeks) US Treasury bill is often used as 'risk-free' rate because it carries minimal risk as the market considers there to be virtually no chance of the U.S. government defaulting on its obligations to repay the principal. The risk-free rate represents the interest an investor would expect from an absolutely risk-free investment over a specified period of time. Investors will not accept additional risk unless the potential rate of return is greater than the risk-free rate.

Problem (q): The risk-free rate is the return on an investment that carries (select one option from below):

a) No Risk
b) High Risk
c) Low Risk

Problem (r) Excess Return & Sharpe Ratio: By comparing the expected return of an investment to the risk-free rate, investors can assess whether the potential return justifies the level of risk taken. So, investors invariably look at how much of 'excess return' (Expected Return – Risk Free Return) you get for each unit of risk we take for each asset – This is called Sharpe Ratio.

Excess Return = Expected Return – Risk Free Return
Sharpe Ratio = Excess Return / Volatility

Calculate the Sharpe ratio for each of the high cap and small cap asset discussed earlier. The risk-free rate of return will be as calculated in earlier problem above. Compare the Sharpe ratios and decide which asset you will invest in. (Let's call it your Try 3 of Investment Decision Making)

Problem (s): Is your result (preferred asset for investment) same in each of the tries 1,2,3 of your investment decision making? Yes/No. If no, why?

Problem (t): While making investment decisions (e.g., Asset Selection - Which asset to select out of potential assets for investment, Asset Allocation – How much to invest in selected list of assets), will you maximize or minimize the Sharpe ratio?

Problem (u): If the risk-free rate goes up and the return and volatility of an asset are unchanged, the Sharpe Ratio ____:

a) Goes up
b) Goes down
c) Could go up or down

Risk Measures beyond volatility: VaR & Conditional VaR

Problem (v): Asset A loses 1% a month for 12 months and Asset B gains 1% per month for 12 months. Which is the more volatile asset?

a) Asset A
b) Asset B
c) Neither
d) It depends

Problem (w): In the above problem, since both assets produce the same return each month, their deviations from their mean are zero each month, and so the volatility of both of these assets is 0. Will it deter/stop you from investing in asset B over asset A?

As you can see from above, volatility is not necessarily a bad thing or risk as it's just the deviation from the mean and there is no harm if the deviation from the mean is towards the upside. Uncertainty on the downside is what investors are most concerned about. Large losses are particularly important to know about. There are few risk measures which tries to capture the downside risk.

Problem (x): Consider the following sequence of monthly returns on an asset: -4%, +5%, +2%, -7%, +1%, +0.5%, -2%, -1%, -2%, +5%. Find the worst return after excluding the 10% of worst return.

The absolute value of the number you arrived in above problem is the maximum loss which can occur by investing in the asset with 100-10%=90% confidence level. This is called 'Value at Risk'. VaR provides an estimate of potential loss at a given confidence level.

Problem (y): Consider the following sequence of monthly returns on an asset: -4%, +5%, +2%, -7%, +1%, +0.5%, -2%, -1%, -2%, +5%. Find the average of worst returns after excluding the 20% of worst return.

The absolute value of the number you arrived in above problem is the average loss beyond VaR which can occur by investing in the asset with 100-20%=80% confidence level. This is called 'Conditional VaR'.

Homework (Putting it all together): Below is the python code to get Gold and Bitcoin daily prices since 2010. Find their:

a) Annualized return for the period. (Ans: 7.8%, 69.6%)
b) Annualized volatility for the period. (Ans: 14.75%, 68.92%)
c) Sharpe Ratio for each assuming Risk Free rate is 4.5%. (Ans 0.2 & 0.9)
d) Value at Risk at 95% confidence level, i.e., potential % loss at 95% confidence level (Ans: 0.015%,0.062%. Note that since we use daily prices and returns to come up this number, this is 1-Day VaR, i.e., potential loss over a single day)
e) Conditional VaR at 95% confidence level, i.e., average % loss beyond VaR with 95% confidence. (0.02%,0.098%. Note that since we use daily prices and returns to come up this number, this is 1-Day CVaR, i.e., potential average loss beyond VaR over a single day)

```python
# Gold and Bitcoin Prices
import pandas as pd
import yfinance as yf

df_gold = yf.download('GC=F',start='2010-01-01',progress=False)
df_bitcoin = yf.download('BTC-USD',start='2010-01-01',progress=False)
df=pd.concat([df_gold['Close'],df_bitcoin['Close']],axis=1).dropna()
df.columns=['Gold','Bitcoin']
df.tail()
```