CS 5060 **Name: Braxton Francom**
Spring 2024
Midterm Exam
3/4/2024
Time Limit: 3/17/2024 By End of Day______________________

This exam contains 14 pages (including this cover page) and 5 questions.
The total of points is 80.

This test is open notes but not open-neighbor. Please read each question carefully before deciding on an answer. A clear and concise statement or paragraphs are much better than long answers with no focus on short written responses to these questions.

I have taken the liberty of organizing this as I would an assignment rather than a classical exam. I hope that you take the time to learn and understand from this, if you do so, I imagine that this will be a moderately pleasant venture.

Mario Mercy Option: As long as you get 3 of these questions with 75%, the overall exam score will be at least 75%.

Grade Table

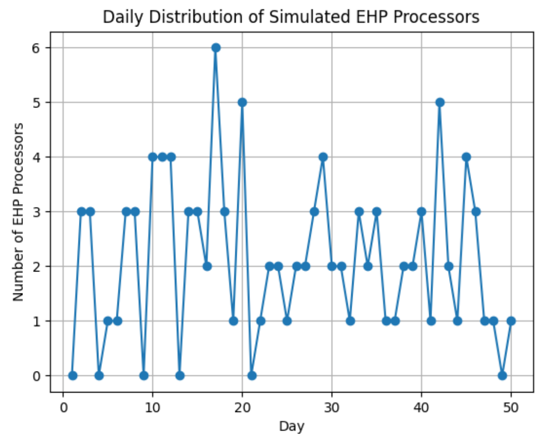
Question	Points	Score
1	20	
2	20	
3	20	
4	20	
5	0	
Total:	80	

1. (20 points) Processor Binning:

   You are tasked with fulfilling a procurement request for a mission-critical component, the processor that your company makes is typically binned into three categories: Class 7, Class 5, and Class 3. (Side note: it is interesting that both team AMD and team Intel uses the same numerical designations...)

   Processors that passes the highest levels are given the 7 designation, with those that pass mid levels are given 5, and those that are very unstable have cores turned off and given the 3 designation. Our technique is very good, and we can have a success rate of 58% viable processors. Of these processors, 67% are Class 3, 24% are Class 5, 8.8% are Class 7, and 0.2% are Extreme High Performance (EHP).

   PART 1 - Set up the simulation with a random float drawn between 1-1000 to emulate the value of the processor. (Note that we must deliver one processor within one week to our client, but they can only accept our EHP products.) Run a simulation assuming we can fabricate and bin-test 1000 units per day. Plot the daily distribution of simulated EHP processors of at least 50 days.



PART 2 - Compute the simple early stopping rule for this problem. Show some plots of the performance.

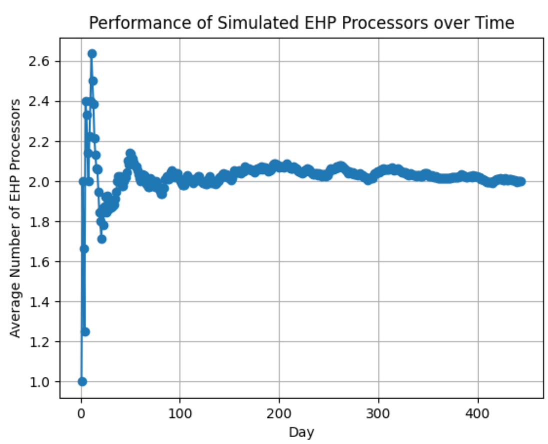


Figure 1: Performance of Simulated EHP Processors over Time

PART 3 - Compute the stopping rule where we get a reward based on the following function:

$$0 \text{ if } val < 998$$
$$5000 * (val - 997) * *2 \text{ if } val >= 998$$

Again, show some plots of performance.

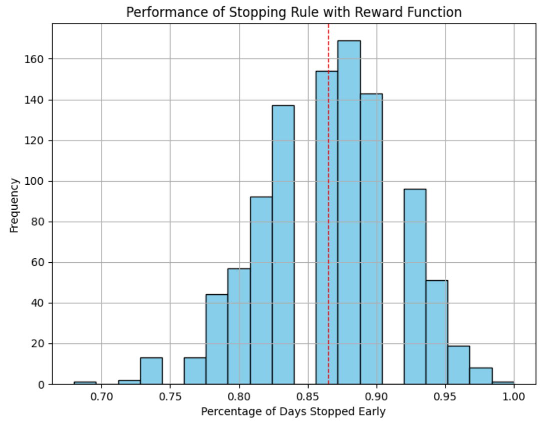


Figure 2: Performance of Stopping Rule with Reward Function

PART 4 - We invest in a slower but better lithography method, increasing our base success rate to 69%. In addition 48% are now Class 3, 33% are Class 5, and 14% are class 5. The remaining 5% are EHPs. Calculate the new results for the reward function in PART 3.

How does this compare with the first simple early-stopping problem?

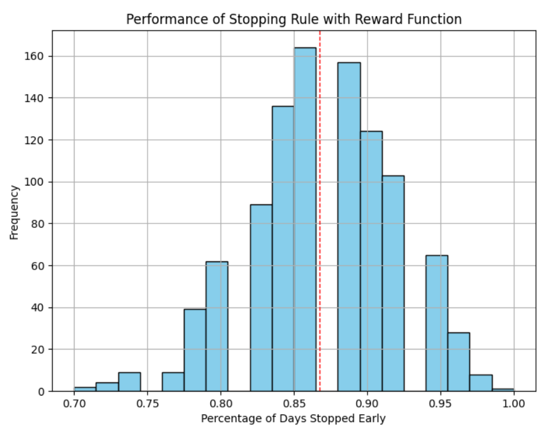


Figure 3: Performance of Stopping Rule with Reward Function

In comparison to the first simple early-stopping problem, it appears the updated reward function gave a more stable performance in terms of the percentage of days stopped early. The histogram of the percentage of days stopped early showed a narrower distribution and a slightly higher mean percentage compared to the first problem. This tells us that the updated method led to more consistent outcomes, with processors reaching the desired performance threshold more reliably across iterations.

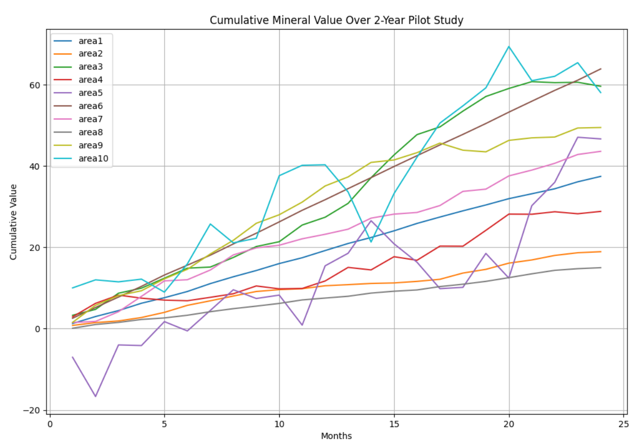
2. (20 points) Mineral Samples and Geologic Surveys:

   You are part of the algorithms team behind a large mining group, based on several survey team results, you have identified 10 potentially valuable locations to start mineral excavation. This is a very expensive process, and developed infrastructure capital is often not movable once installed. You are interested in conducting some pilot mining using inefficient, but cheaper (lots of small trucks) methods for a few months.

   The distribution of mineral value in each area looks as follows:

   area1 = beta(2, 2)+1
   area2 = beta(3, 7)*3
   area3 = normal(2.4, 1.8)
   area4 = uniform(-1, 4)
   area5 = normal(0, 9)
   area6 = beta(7, 3)+2
   area7 = uniform(0, 4)
   area8 = beta(3, 7)*2
   area9 = normal(2, 1.4)
   area10 = normal(1.3, 7)

   PART 1 - You have enough teams of trucks to collect data from each location simultaneously. Plot the results of conducting an equal test with one data point collected each month after your 2-year pilot study is completed.



   PART 2 - You can remove resources to dedicate them to another operational site to get more data points per month. Each data point you get will be different as each team will start excavation in a different sub-area. You have funding for a 2-year pilot study, compute an epsilon greedy approach where each of your 10 teams can move. Tell me what your logic is to move these teams and when

For each month of the pilot study, my code iterates through the following steps:

- **Decision Making**: It decides whether to explore or exploit based on a random number drawn from a uniform distribution. If the random number is less than epsilon, I choose to explore; otherwise, I choose to exploit.
- **Exploration**: During exploration, a random team is selected, and a random location is chosen for it to move to. This allows for the discovery of potential mineral deposits in different areas.
- **Exploitation**: During exploitation, the code identifies the location with the highest estimated mineral value based on the data collected so far, and then moves a random team to that location. This strategy aims to focus resources on areas with higher expected mineral value but does not specifically select the team based on its estimation.

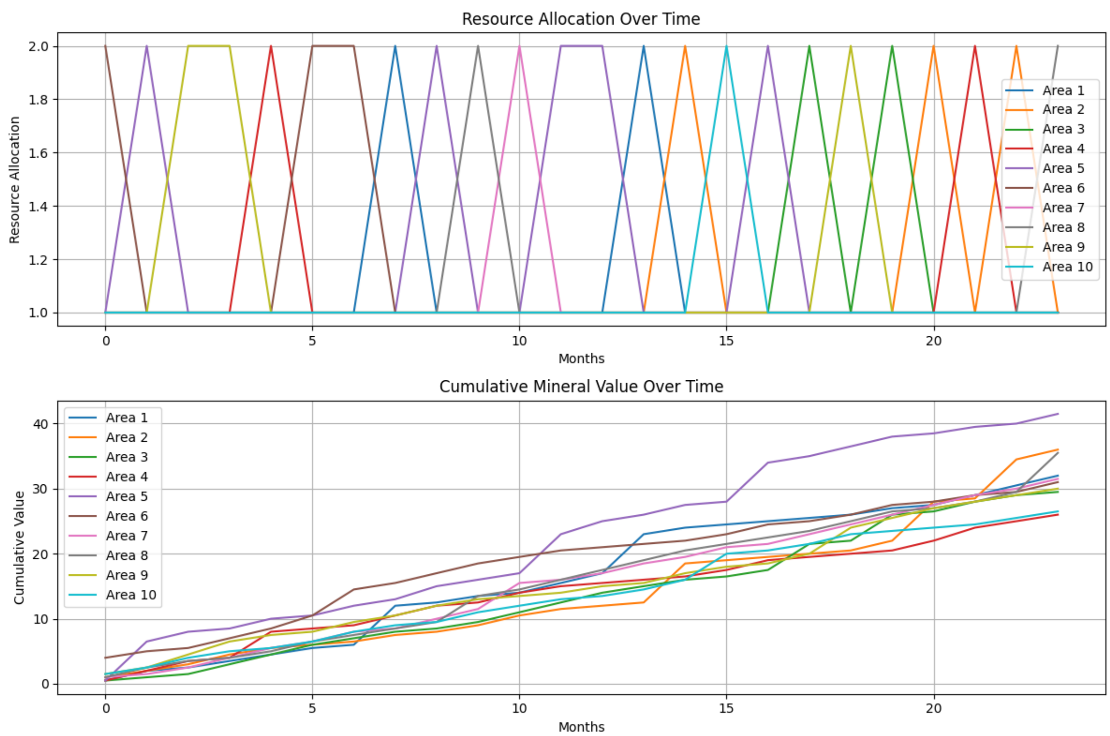


Figure 4: Epsilon Greedy Approach

PART 3 - Lock the teams to be unable to move for 6 months. Allow each team to move in an epsilon-greedy fashion. How does your algorithm perform? show plots and comment on your changes.

Initially, for the first 6 months, all teams are locked, resulting in uniform resource allocation across all areas. After unlocking, the epsilon-greedy approach allows teams to

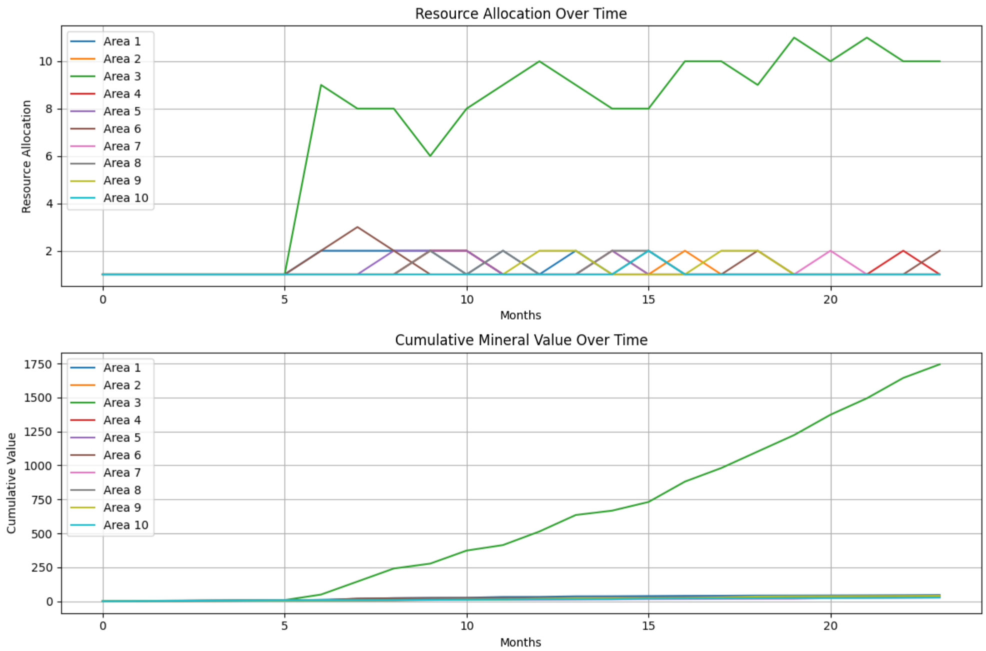


Figure 5: Simulation where teams are locked for first 6 months

explore and exploit different areas based on their cumulative mineral value. In my algorithm we can sees some fluctuation in resource allocation as teams adapt their movement strategies to maximize mineral collection. Overall, my algorithm performs well by exploiting an area that shows promise, but it sometimes misses out on better opportunities by doing so.

PART 4 - By the end of the 2-year pilot, operational funds are allocated to open an operational mine in two locations. Which did you pick, and how quickly were you able to determine what locations were best?

By the end of the 2-year pilot study, after analyzing the cumulative mineral values over time, it was clear that area3 and area6 consistently yielded the best results. These two locations demonstrated higher cumulative mineral values compared to others throughout the study period. The selection process was relatively quick, as the performance of each area became apparent within the first few months of the pilot study. Because of these reasons, I would pick to allocate operational funds to open mines in area3 and area6.

3. (20 points) Tech Sector before the Covid changes:

Suppose that you are a bank that services many firms tied closely to the technology sector. We will be looking at specifically AAPL (Apple) and TSLA (Tesla) shares provided in the Canvas midterm module.

PART 1 - Plot the value of the assets over time, use the Close price column. Also, plot a distribution of the difference between the high and low prices each day. Plot the distribution of the change in prices of the close each day. Comment on the similarities and differences in these stocks. Explain the distributions which govern the change in Close price of these two stocks.

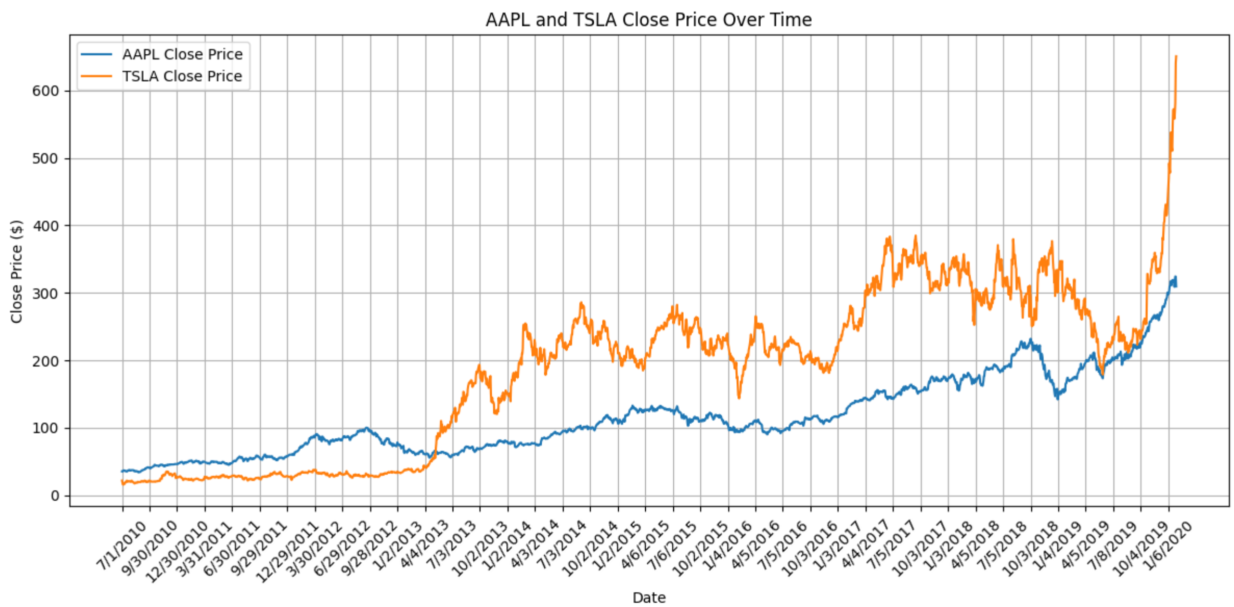


Figure 6: AAPL and TSLA Close Price Over Time

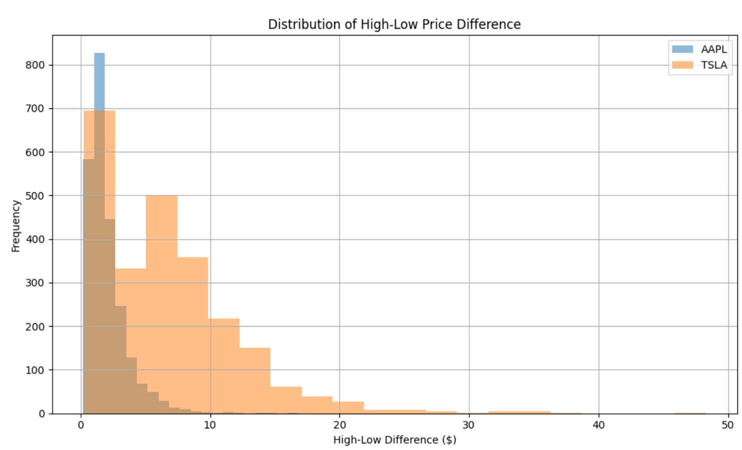


Figure 7: Distribution of High-Low Price Difference

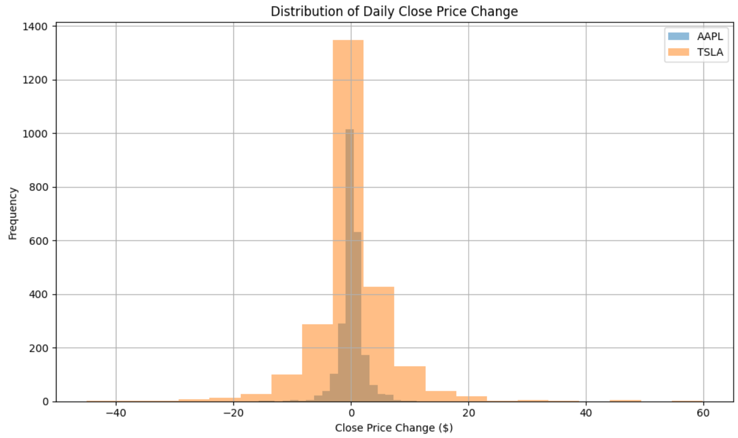


Figure 8: Distribution of Daily Close Price Change

The upward trend observed in both AAPL and TSLA stocks signifies their overall positive performance during the given period. However, a particularly intriguing aspect lies in the distribution of the high-low price difference. AAPL exhibits a relatively tight distribution, with values predominantly clustering below 10. This indicates a relatively stable price range within which the stock tends to fluctuate. Conversely, the distribution for TSLA presents a striking contrast, showcasing a much broader spread. This difference shows TSLA's characteristic high volatility, suggesting significant fluctuations in its trading prices. The extent of TSLA's volatility becomes even more apparent when examining the 'Distribution of Daily Close Price Change'. In this distribution we see a wide-ranging spread, highlighting the considerable variability in TSLA's daily price movements compared to AAPL.

PART 2 - Compute a 1-year European Call option moving forward from the final date in the data provided (Jan 31, 2020). The strike price is the same as the close price on that day. Assume volatility at 8.2%. Calculate the drift based on a regression or analysis in PART 1. Use the distribution of the stocks calculated. Plot the simulated paths for 1000 runs and the computed answer.

Computed Answers:

```
Simulated European Call option price for AAPL: 10.2259836840406
Simulated European Call option price for TSLA: 20.2338120417510
```

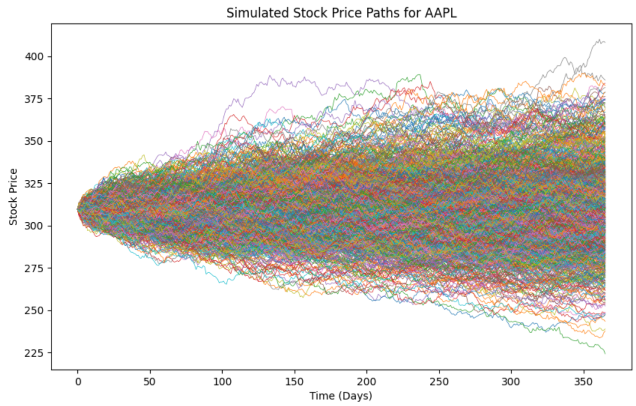


Figure 9: Simulated Stock Price Paths for AAPL

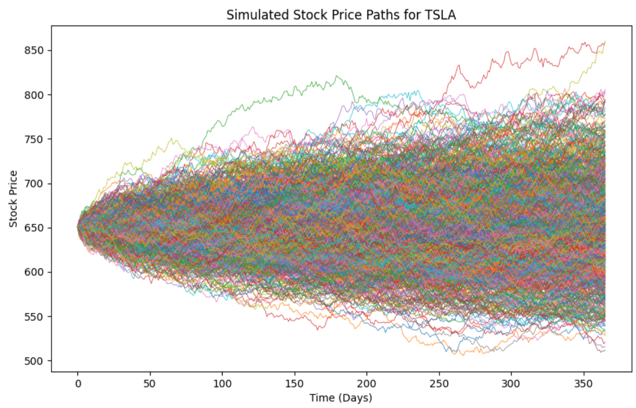


Figure 10: Simulated Stock Price Paths for TSLA

PART 3 - Find the option price for a new financial asset called the AAPL+TESLA security. This security is the value of the average between the two options.

Option price for AAPL+TSLA before COVID slide: 15.23539

Now add a simulated slide down in March of 2020 of 35% for AAPL and 41% for TSLA. What Happens to the price of the option if you could predict this slide in price from COVID? (Assume the same 1-year window of expiry)

Option price for AAPL+TSLA after COVID slide: 4.66815

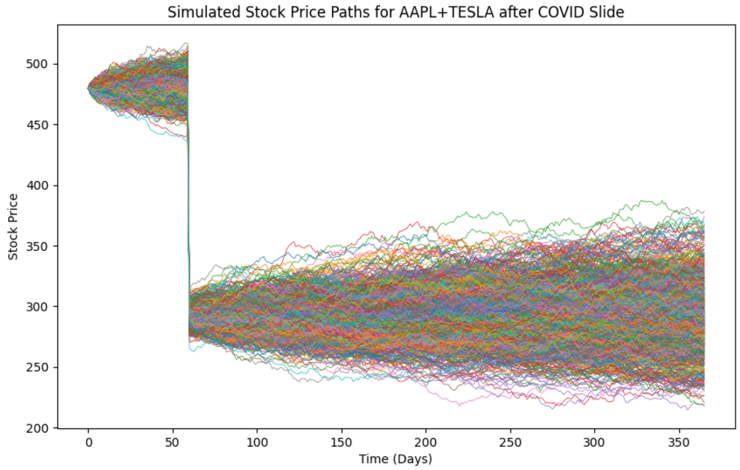


Figure 11: Simulated Slide

PART 4 - BONUS QUESTION: Do PART 3 again but Calculate the joint probabilities through a correlation matrix and simulate random draws that reflect the correlation in these tech-heavy industries. Add the expected mid-COVID growth to the stock prices starting in April of 2020, where firms grew 3x faster than anticipated after a 30% slump.

```
Option price for AAPL+TESLA before mid-COVID growth: 164.199209
Option price for AAPL+TESLA after mid-COVID growth: 165.144907
```

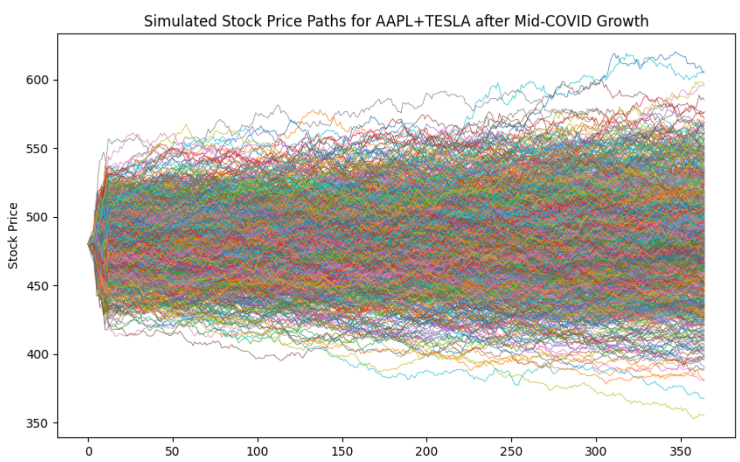


Figure 12: Part 4

4. (20 points) Insurance: Value of Subdivision

We are going to look at a simple insurance dataset (in the Canvas midterm module, a data file is provided named insurance.csv) where we only consider a few factors for 10,000 people:

- age (INT age of claimant)
- sex (BINARY As identified by the claimant)
- bmi (FLOAT bmi of claimant, NA if not claimant that year)
- children (INT children count per claimant, NA if not claimant that year)
- smoker (BINARY smoking status of claimant, NA if not claimant that year)
- region (STRING claimant regional ID, NA if not claimant that year)

PART 1 - Calculate the basic insurance rate for this population if we are looking at zero deductible policies and we need an 11 % profit margin. What is the standard deviation and volatility of this portfolio?

Basic Insurance Rate: 1970.8965739742491
Standard Deviation of Charges: 6326.251703908578
Volatility of Portfolio: 3.5629162301391624

This approach calculates volatility as the standard deviation of charges normalized by the mean of charges. This represents the relative variability of charges compared to their average value.

PART 2 - Calculate the tiered insurance rate (with the same 11% profit margin) for this population by breaking up the insurance product by age categories of [18-22, 23-30, 31-48, and 49+]. What is the standard deviation and volatility of this portfolio?

Age Range: (18, 22)
Total Charges: 1859253.200169
Basic Insurance Rate for Age Group: 1250.7703346591456
Standard Deviation of Charges: 5085.018810453363
Volatility of Portfolio: 4.512715662657135

Age Range: (23, 30)
Total Charges: 2313259.91069
Basic Insurance Rate for Age Group: 1543.1000606165264
Standard Deviation of Charges: 5501.748381892145
Volatility of Portfolio: 3.9575792003146764

Age Range: (31, 48)
Total Charges: 6335341.15741
Basic Insurance Rate for Age Group: 1963.7611518361075
Standard Deviation of Charges: 6271.636972767756
Volatility of Portfolio: 3.544991728379606

Age Range: (49+)
Total Charges: 7247970.72249
Basic Insurance Rate for Age Group: 2591.061997411884
Standard Deviation of Charges: 7279.252771539429
Volatility of Portfolio: 3.118401097495757

PART 3 - Calculate the same in PART 2, but add sex to the calculation. Compare just using the sex parameter as compared to both age and sex.

ALL MALES:
Total Charges: 9434763.79614
Basic Insurance Rate: 2075.012445753002
Std Dev Charges: 6716.154381692353
Volatility: 3.5927164576466866

ALL FEMALES:
Total Charges: 8321061.194619
Basic Insurance Rate: 1864.8047498540461
Std Dev Charges: 5901.664154539797
Volatility: 3.5128863823689285

Gender and Age Range: male - (18, 22)
Total Charges: 1055081.46705
Basic Insurance Rate: 1431.712015190098
Std Dev Charges: 5692.3624418387135
Volatility: 4.41326345200925

Gender and Age Range: female - (18, 22)
Total Charges: 804171.733119
Basic Insurance Rate: 1072.8733458678967
Std Dev Charges: 4404.697439519455
Volatility: 4.557121468901331

Gender and Age Range: male - (23, 30)
Total Charges: 1271847.90377
Basic Insurance Rate: 1635.8646270969873
Std Dev Charges: 5984.667601295805
Volatility: 4.060837875825341

Gender and Age Range: female - (23, 30)
Total Charges: 1041412.00692

Basic Insurance Rate: 1443.1552155820225
Std Dev Charges: 4930.975133187574
Volatility: 3.792649840253531

Gender and Age Range: male - (31, 48)
Total Charges: 3372321.8704999997
Basic Insurance Rate: 2085.391240253482
Std Dev Charges: 6592.3816719199385
Volatility: 3.5089548256382224

Gender and Age Range: female - (31, 48)
Total Charges: 2963019.2869100003
Basic Insurance Rate: 1841.5181458399218
Std Dev Charges: 5931.642850473379
Volatility: 3.575378053645197

Gender and Age Range: male - (49, 130)
Total Charges: 3735512.5548199997
Basic Insurance Rate: 2639.3500546468495
Std Dev Charges: 7633.287563188588
Volatility: 3.210240786447341

Gender and Age Range: female - (49, 130)
Total Charges: 3512458.16767
Basic Insurance Rate: 2541.6092347546937
Std Dev Charges: 6900.08522273683
Volatility: 3.013482360901599

PART 4 - Comment on the change in risk for the portfolio and the anticipated benefits from subdividing the insurance product. Calculate what the prices would be if you could break the product into known claimants (ie: you know which people will be claimants and who will not be this year). What does the cost rise to for the annual rate on the claimants?

By subdividing the insurance product based on age groups and sex, we are likely to have a better understanding of the risk profile of each subgroup. This subdivision allows us to tailor insurance rates more accurately to each subgroup's risk, potentially reducing situations where we lose money. The more factors we consider (such as age and sex), the better we can predict and manage potential claims, leading to a more stable and predictable portfolio. In addition to this, subdividing the insurance product allows for more precise pricing, which can attract customers who are more likely to be profitable for the insurance company(Non-claimants).

```
Annual Rate on the Claimants: 13270.42
```

Bonus:

5. (3 points) I could really use help in understanding how to improve the course, particularly as I missed a few days. Thank you for being here and working hard on these topics.

   What is making sense, and what doesn't? How can I do this better? (I know I can certainly post lectures faster... and clarify assignments better...)

   Overall this class is great! I have thoroughly enjoyed it. As far as how to improve...This could just be a me problem, but I have felt quite confused on some of the past assignments. The lecture material you cover in class is very interesting, but when it comes time to implement the algorithms I never know where to start. I think it would be very helpful to supplement lecturing with a little bit of implementing algorithms together. I might be the only one that thinks that though so take it with a grain of salt. ;)