Fintech545 Week3 Project

Elinor Cheng, yc494

February 2023

1 Problem 1

Choose $\lambda = 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 0.97, 0.99$ and calculate the exponentially weighted covariance matrixes.

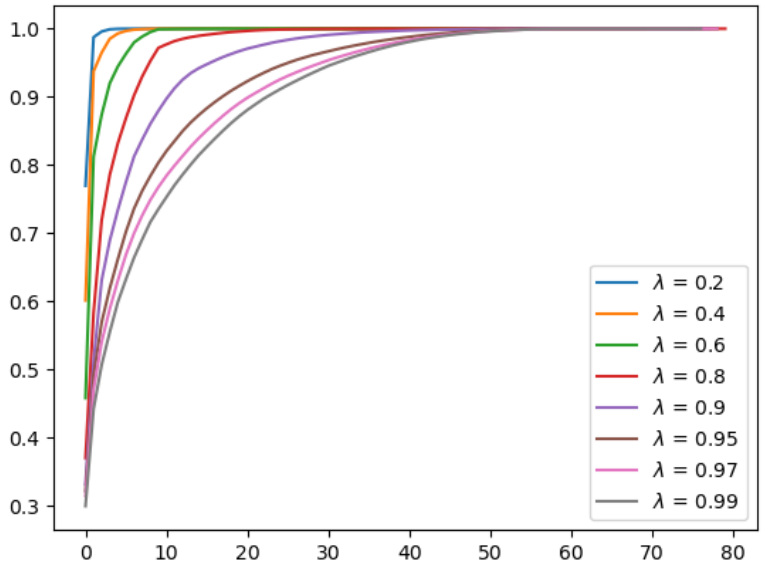

Figure 1: % of variance explained by PCA with different chosen of λ

It is easy to see from the figure above that when λ getting larger, the growth of the cumulative PCA explanations would get slower. When λ getting larger, we will rely more on the previous data, thus the covariance matrix may includes "more information". To be specific, I think if % of variance explained by PCA grows very fast, from the intuition of PCA, it shows that most information of the matrix can be explained by few eigenvalues, and rest of the eigenvalues may not include much information, so the matrix can be easily reduced the dimension without much loss of information. In other word, it may not be "complex" as its dimension shows.

2 Problem 2

	Method	Run Time	Frobenius Norm
$n = 5$	Nearest PSD	0.0009s	1.465e-10
	Higham	0.0028s	1.022e-10
$n = 50$	Nearest PSD	0.0015s	0.03448
	Higham	0.0787s	0.006189
$n = 100$	Nearest PSD	0.0068s	0.07442
	Higham	0.2131s	0.007164
$n = 500$	Nearest PSD	0.1679s	0.3938
	Higham	13.9158s	0.008037
$n = 1000$	Nearest PSD	0.4820s	0.793
	Higham	26.9144s	0.008152

Table 1: Rum time and Frobenius Norm results of using Nearest PSD and Higham Method with different dimensions

From the table above we can see that when the dimension of data is relatively small, the run time and Frobenius norm for both methods are closed. When dimension of data is getting larger:

- the Frobenius norm result of the nearest PSD grows much faster than the result of Higham
- the run time of nearest PSD keeps short but the run time of Higham is growing fast

From my perspective, if the dimension of the data is large(thousands), Higham is a better choice though it will take longer time to execute. Since the run time of Higham is still less than 1 min when $n = 1000$, considering that its outcome is much more closed to the origin matrix, it's worth waiting for a more accurate result.

3 Problem 3

The results of the simulations with different covariance matrix are shown in the following table:

	Results Category	Direct Simulation	PCA with 100% explained	PCA with 75% explained	PCA with 50% explained
Pearson Covariance	Run Time	0.1466s	0.0900s	0.0343s	0.0229s
	Frobenius Norm	8.873e-08	4.461e-08	2.704e-06	1.118e-05
Exponentially Weighted (EW) Covariance	Run Time	0.1745s	0.1362s	0.0171s	0.0112s
	Frobenius Norm	7.02e-08	4.286e-08	2.453e-06	1.193e-05
EW Correlation + Variance	Run Time	0.1059s	0.0889s	0.0418s	0.0106s
	Frobenius Norm	4.989e-08	6.419e-08	2.774e-06	1.255e-05
Pearson Covariance +EW variance	Run Time	0.1082s	0.0723s	0.0122s	0.0058s
	Frobenius Norm	2.076e-14	2.727e-14	1.129e-12	8.639e-12

Table 2: Rum time and Frobenius Norm results of using Nearest PSD and Higham Method with different dimensions

From the table, it is easy to see that, while the growing f norm shows the loss of information when the percent of variance explained by PCA decreases, we can use less time to simulate the same amount of data with using direct simulation. From my perspective, the loss of information is somewhat acceptable since there are no remarkable differencies in the F-norms using different simulation methods (with all 4 covariance matrixes).

Therefore we may draw the conclusion that PCA simulation can be an efficient way to simulate data following a certain known distribution.