FACTORS AFFECTING CONVERGENCE RATE IN MARKOV CHAINS MONTE CARLO

Nour El Orabi
Lobna Aboudoma

Friday, May 21st, 2021

—

DSCI 341101

Fundamentals of Simulation

—

Dr. Ali Hadi

What are Markov Chains?

A discrete-time Markov Chain is a stochastic process that follows the Markov property where the next outcome is only dependent on the previous state. The chain moves in discrete time steps, and it disregards all previous states that led to the current state.

Convergence of Markov chains

A Markov chain converges as the number of steps approaches infinity (P^n as $n \rightarrow \infty$). If the chain converges then there exists a stationary distribution and the matrix must be aperiodic to ensure convergence of the chain. The stationary distribution represents the probability of being in a specific state at any point in the simulation. For each i in the stationary distribution π, i is the proportion of time that the Markov Chain is in the corresponding state. There are numerous factors that affect the convergence rate of Markov Chains which are dependent on the structure and characteristics of the transition matrix. [i]

METHOD OF GENERATING RANDOM MATRICES

For each feature of Markov Chain States, we had a different approach for generating the random matrix, that would ensure the characteristic of interest would be present in a proportion of the generated matrices. The code below is one of our generation methods where we add random numbers generated from a uniform distribution to a 2D array of zeros. Then we round the values to increase the probability of getting zeros in the matrix.

```
def gen_matrix2(n):
    # Start off with a matrix of zeros
    result = np.zeros((n,n))
    # generate random numbers from a uniform distribution and add them to the matrix
    result = result + np.random.uniform(low=0, high=0.25, size=(n, n))
    # round the values in the matrix to increase the probability of getting zeros
    result = np.around(result, decimals=1)
    # Normalize the matrix to have rows with sum = 1 for it to be a stochastic matrix
    result = sklearn.preprocessing.normalize(result, norm="l1")
    return result
```

Factors Affecting Convergence Rate

We will be exploring the factors that affect the convergence rate of Markov Chains in Monte Carlo simulations. The number of states in the Markov Chain will be another factor explored within each of the following factors.

MULTIPLICITY OF EIGENVALUES

The roots of the characteristic equation $|A-\lambda I|=0$, λ, are also known as the eigenvalues. The latter may be either repeated, where they are also known as the multiplicity of eigenvalues, or distinct. In Markov chains, the multiplicity of eigenvalues has an impact on the point of convergence.

PERIODICITY

The periodicity of a state in the Markov Chain is the minimum period it takes to return to the same state. An aperiodic state is a state with a period of 1 which means it returns to itself. A periodic state is when a state can only be returned to in multiples of a period greater than 1.

RECURRENCE

Recurrence in matrices explores the ability to return to any state at any point in the chain. In other words, given that we go from state i to state j, there exists at least one path that takes us back to state i. In non-recurrent matrices, there exists a state, that is not accessible once another state is visited. An example of a recurrent matrix would be the absorbing matrix where each state when entered results in an endless self-loop within the same state, hence the probability of returning to this state is 1.

TRANSIENCE

A state is said to be transient if the probability of leaving a state and not being able to return to the same state is greater than zero. The state would be classified as transient if there exists a state i and j different from i that is accessible from i but, i is not accessible from j. In other words, the probability of returning to this specific state (i) is less than one.

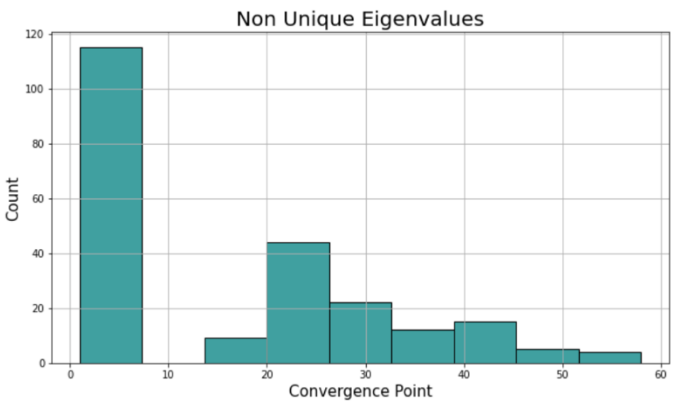
IRREDUCIBILITY

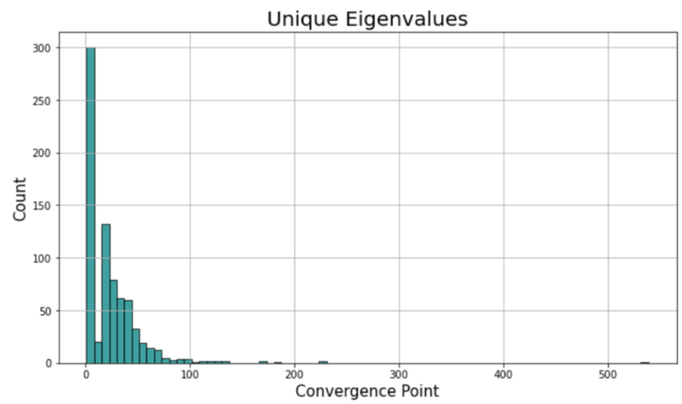
Irreducibility is a characteristic of a Markov Chain where every state can be reached from every other state, hence all states belong to one communicating class. Likewise, a Markov Chain is reducible if there exists more than one communication class

MULTIPLICITY OF EIGENVALUES

A matrix with multiplicity of eigenvalues of 1 has unique eigenvalues and similarly a matrix with repeated eigenvalues has eigenvalues with multiplicity n > 1. Matrices with unique eigenvalues tend to converge at a faster rate.

Number of States / Matrix Size = 3



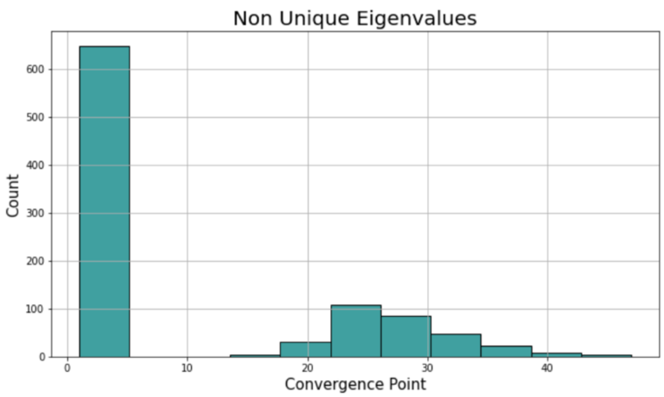


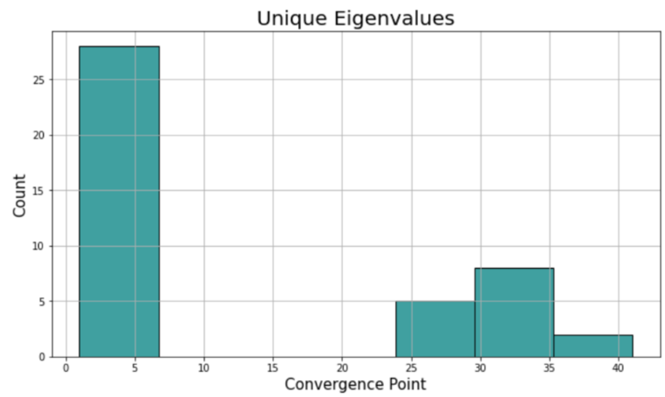
	Unique	Non-Unique
Count	773.000000	227.000000
Mean	50.371281	15.484581
Stdev	226.278791	16.215187
Minimum	1.000000	1.000000
25%	1.000000	1.000000
50%	20.000000	4.000000
75%	35.000000	27.000000
Max	2000.000000	66.000000
Outliers	11.000000	1.000000

Interpretation

As shown in the graphs above, matrices with unique and nonunique eigenvalues have different distributions of data. The mean of the convergence rate of the matrices with unique eigenvalues of multiplicity one is higher than that of the nonunique eigenvalues, which proves that when multiplicity is equal greater 1, the Markov Chain is more likely to converge faster than when the multiplicity is equal to 1. As shown in the table on the left, many matrices with unique eigenvalues didn't converge, due to the fact that the maximum convergent point was 2000(benchmark for nonconvergent points). Furthermore, there exists 11 outliers that most likely didn't converge as well.

Number of States / Matrix Size = 6



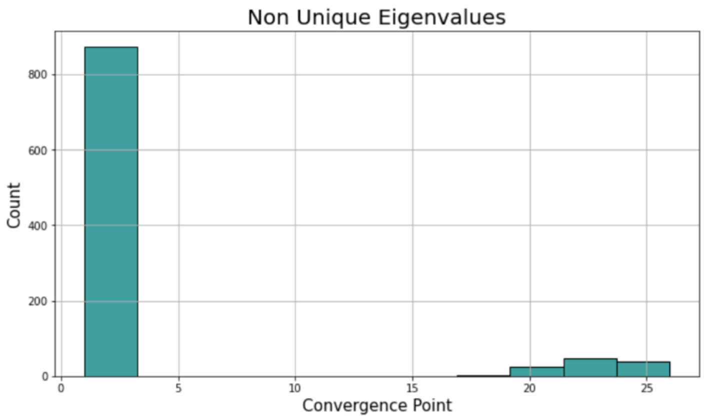


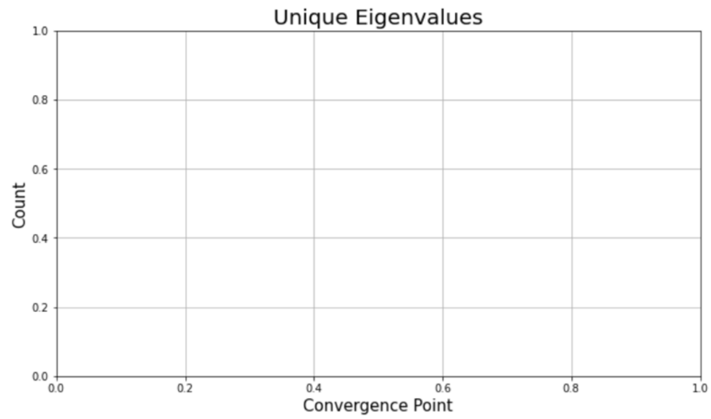
	Unique	Non-Unique
Count	43.000000	957.000000
Mean	11.395349	9.703239
Stdev	14.568622	13.011088
Minimum	1.000000	1.000000
25%	1.000000	1.000000
50%	1.000000	1.000000
75%	27.500000	24.000000
Max	41.000000	65.000000
Outliers	0.000000	2.000000

Interpretation

In the graphs above we see that the distribution of the convergence points has changed. The convergence rate of both unique and non-unique eigenvalues is getting faster as the number of states increases. Matrices with non-unique eigenvalues still tend to converge faster than matrices with unique eigenvalues, yet the difference is not as drastic as the 3x3 matrix.

Number of States / Matrix Size = 10





	Non-Unique
Count	1000.000000
Mean	3.849000
Stdev	7.486404
Minimum	1.000000
25%	1.000000
50%	1.000000
75%	1.000000
Max	30.000000
Outliers	14.000000

Interpretation

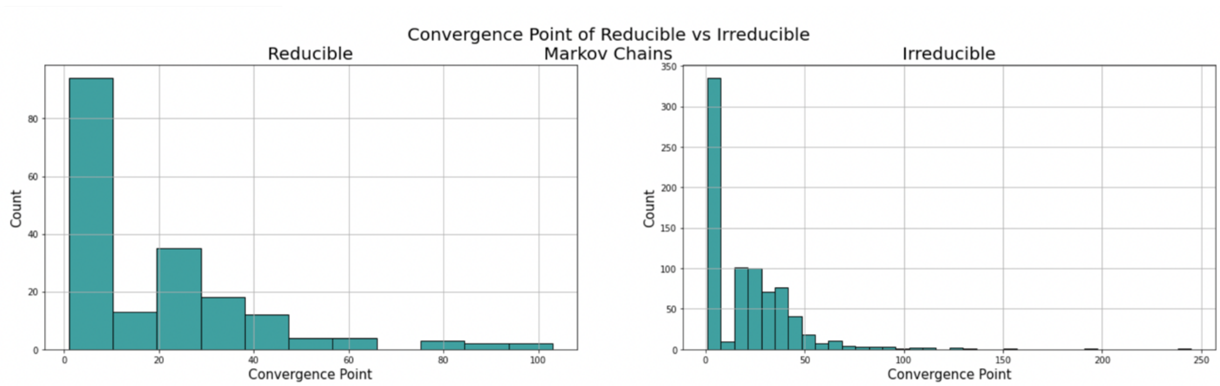
In the graphs above we see that the distribution of the convergence points has changed. The convergence rate of both unique and non-unique eigenvalues is getting faster as the number of states increases. Matrices with non-unique eigenvalues still tend to converge faster than matrices with unique eigenvalues, yet the difference is not as drastic as the 3x3 matrix.

IRREDUCIBILITY

A Markov Chain is said to be irreducible if the transition matrix raised to the power of 4 is a regular matrix. A regular matrix does not include negative values or zeros in the matrix. Hence, we checked for the irreducibility of the Chains by raising the transition matrix to the fourth power and counting the number of zeros in the matrix. If there were no zeros in the matrix, then we concluded that it is a regular matrix, hence, irreducible.

```
def check_irreducible(mat):
    mp = matrix_power(gen_matrix2(3), 4)
    return np.count_nonzero(mp == 0)
```

Number of States / Matrix Size = 3

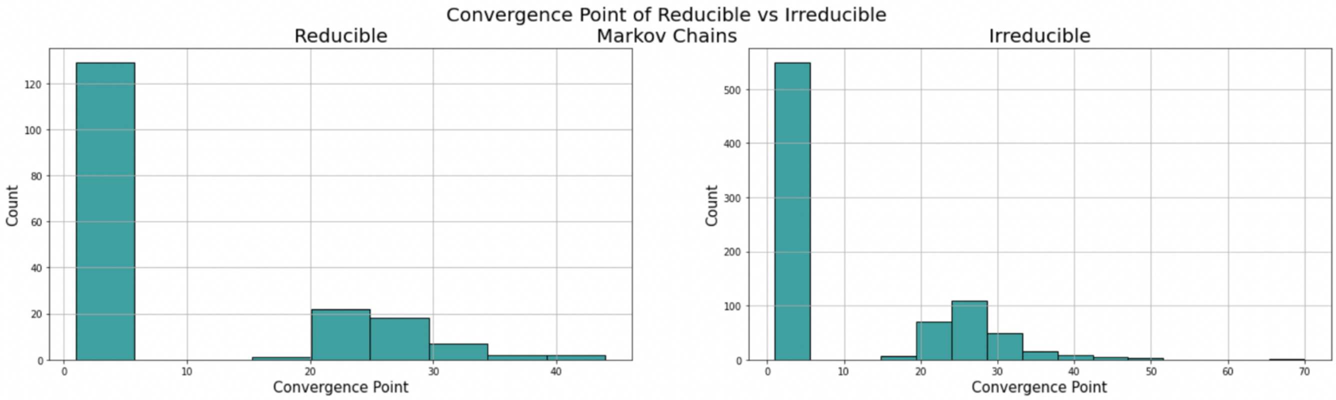


	Reducible	Irreducible
Count	191.000000	809.000000
Mean	20.439791	54.278121
Stdev	27.729052	245.129634
Minimum	1.000000	1.000000
25%	1.000000	1.000000
50%	11.000000	18.000000
75%	29.000000	34.000000
Max	183.000000	2000.000000
Outliers	4.000000	16.000000

Interpretation

Irreducible and reducible Markov Chains have extremely similar distributions of convergence points however irreducible Markov Chains tend to have a higher convergence point with the average being 54 while on the other hand reducible has a lower convergence point with a mean of 20. The data is also more varied in irreducible Markov Chains and there are chains that don't converge with 16 outliers. This could be due to the movement within the chain, not being constrained by the different communication classes like the in the reducible Markov Chains.

Number of States / Matrix Size = 6



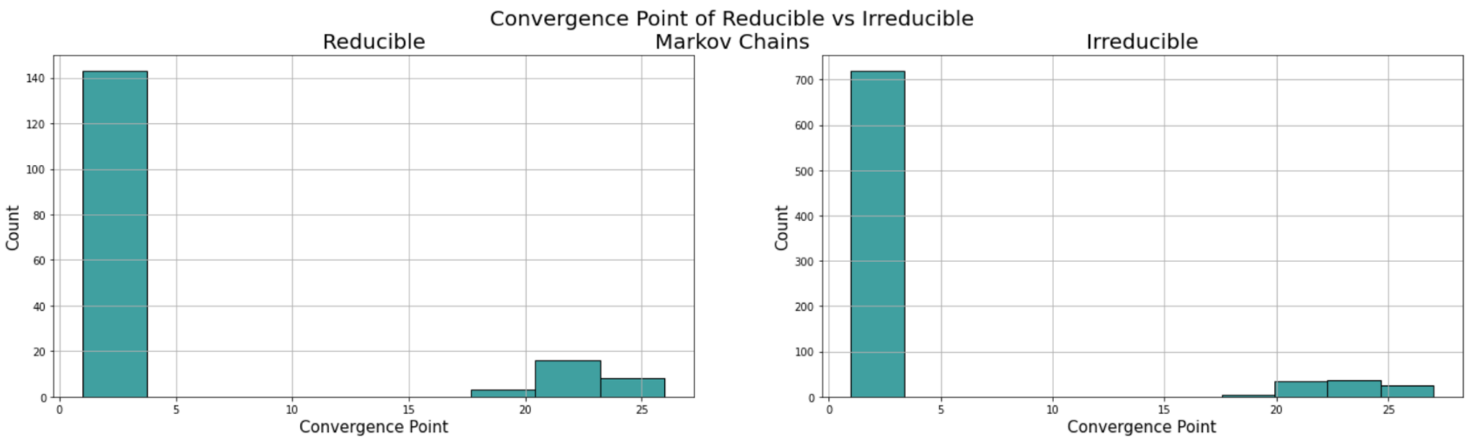

```
irreducible(6)
```

	Reducible	Irreducible
Count	207.000000	793.000000
Mean	10.618357	10.044136
Stdev	13.350548	13.879780
Minimum	1.000000	1.000000
25%	1.000000	1.000000
50%	1.000000	1.000000
75%	24.000000	23.000000
Max	49.000000	151.000000
Outliers	0.000000	3.000000

Interpretation

This distributions of the two graphs are becoming more similar as well as the statistics for the irreducible and reducible Markov Chains. The mean for both distributions is approaching 10, irreducible chains still have a slightly faster convergence rate.

Number of States / Matrix Size = 10



	Reducible	Irreducible
Count	170.000000	830.000000
Mean	4.464706	4.045783
Stdev	8.025215	7.835789
Minimum	1.000000	1.000000
25%	1.000000	1.000000
50%	1.000000	1.000000
75%	1.000000	1.000000
Max	26.000000	40.000000
Outliers	0.000000	10.000000

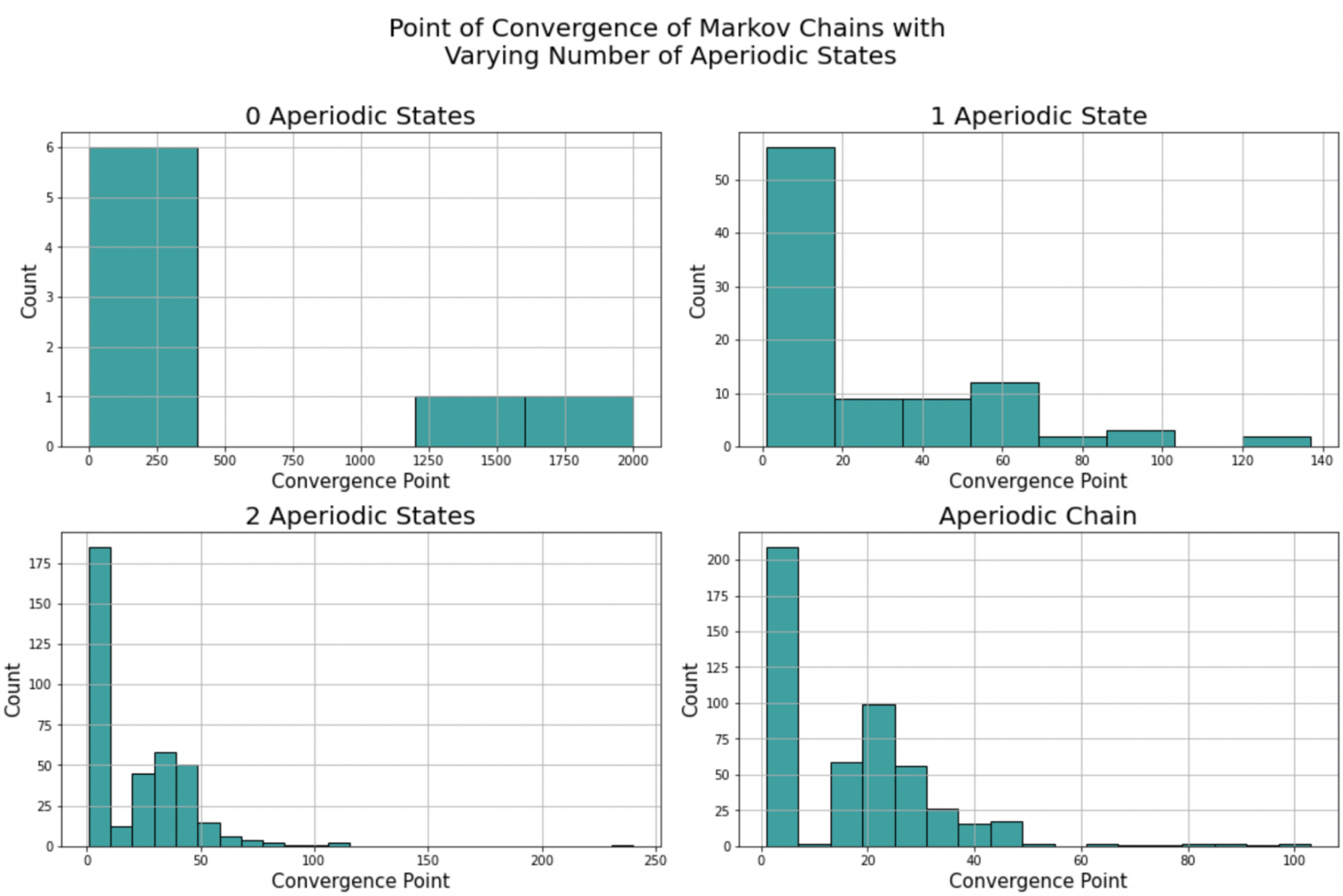
Interpretation

As evident in the graph, as we increase the number of states, there no longer exists unique eigenvalues. Moreover, more non-unique eigenvalues converge at 1, since there aren't any unique eigenvalues.

PERIODICITY

For Markov Chains with 3 States

The period of a state in a Markov Chain is defined as the gcd of the number steps it takes to return to the same state. A state is said to be aperiodic if it has a period of one, in other words, if this state can return to itself. The Markov Chain as a whole is said to be aperiodic if all states in the chain are aperiodic and similarly, periodic if the number of steps is greater than one. In the graphs below, the number of aperiodic states are compared to examine the effect the

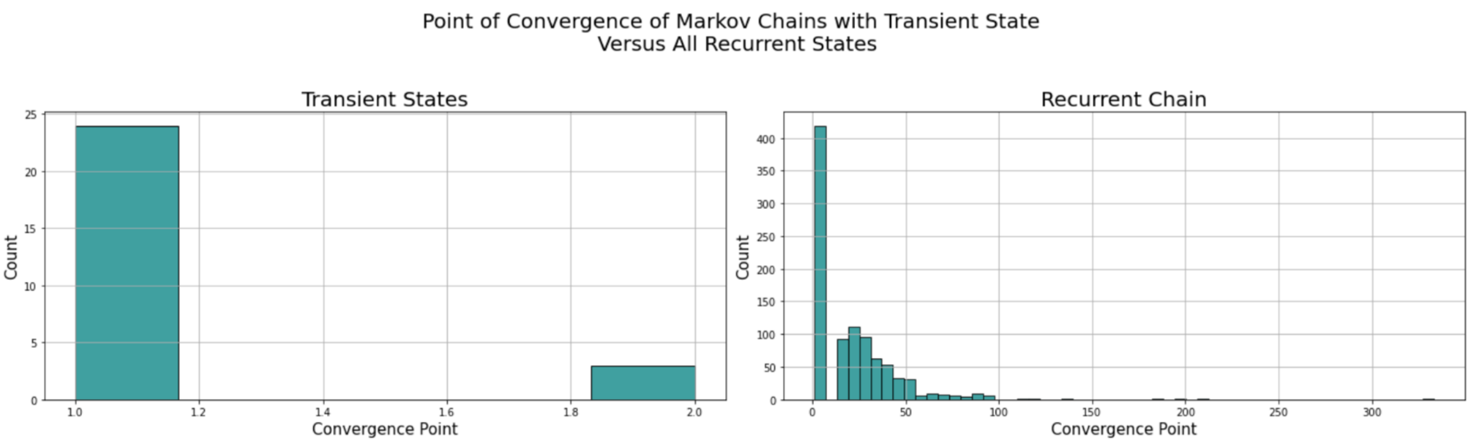


We can deduct from the graphs that Aperiodic Markov Chains converge at the fastest rate. This occurs when all states of the chain are aperiodic, hence the chain as a whole is classified as an aperiodic chain. From the 1,000 matrices randomly generated, the most where aperiodic chains, while completely periodic chains were the least generated and had the highest convergence rate, with an average of 443 iterations of the convergence functions, implying the multiplication of the probability distribution with the transition matrix until convergence. We can generalize that as the number of aperiodic states increases the chain converges faster.

	0 States	1 State	2 States	Aperiodic
Count	8.000000	98.000000	389.000000	505.000000
Mean	443.375000	108.632653	49.632391	18.899010
Stdev	782.115252	382.347777	223.801933	28.066314
Minimum	1.000000	1.000000	1.000000	1.000000
25%	2.000000	1.000000	1.000000	1.000000
50%	46.000000	2.000000	19.000000	17.000000
75%	409.000000	49.250000	38.000000	26.000000
Max	2000.000000	2000.000000	2000.000000	304.000000
Outliers	0.000000	5.000000	7.000000	8.000000

TRANSIENCE & RECURRENCE

In Markov Chains, states are classified as either recurrent or transient, where both describe the likelihood of a state returning to itself through a particular path. When the likelihood (nonzero probability) is less than 1, such that the state never returns to itself, the state is transient. If there exists a multitude of transient states, this suggests that the Markov Chain is absorbing. However, when the likelihood is equal to 1, the state is considered to be recurrent, meaning that a particular state does return to itself. A Markov chain is said to be ergodic when all states of the chain are recurrent. These two types of states are a description of the Markov Chain's overall structure.



	Transient	Recurrent
Count	27.000000	973.000000
Mean	1.111111	60.233299
Stdev	0.320256	273.747582
Minimum	1.000000	1.000000
25%	1.000000	1.000000
50%	1.000000	18.000000
75%	1.000000	32.000000
Max	2.000000	2000.000000
Outliers	0.000000	20.000000

Interpretation

There is an apparent distinction between the points of convergence of transient and recurrent states. This is evident through both the graph and the mean of both types of states. The count of transient states is immensely less than the count of recurrent states, which emphasizes that the probability of having a transient state in a Markov Chain is less than the recurrent chains. As shown in the graph below, the point of convergence occurs after 1 or 2 iterations, this is due to the fact that transient states terminate the circulation of path and given that the matrices' size or the number of states is 3, the transient states are bound to convergence quickly. Conversely, recurrent chains take a longer period to converge, since all states are able to reach each other.

Conclusion

Which Factors had the Greatest Effect on Convergence Rate?

There are multiple factors that impact the convergence rate of Markov Chains, most dependent on the characteristics of the states within the chain resulting in a single communication class or multiple communication classes within the chain. The number of states was also explored within a few of the features of the states such as in the multiplicity of eigenvalues and the irreducibility of the chain. This allows us to explore the effect of the matrix features along with the convergence rate of large Markov Chains.

From the graphs we observed that as the number of states increased, the point of convergence decreased which means that it reaches the stationary distribution in fewer iterations. In small matrices non-unique eigenvalues had a significantly faster convergence rate than matrices with unique eigenvalues of multiplicity one. On the other hand, the difference between average convergence points of irreducible versus reducible chains was not extremely significant yet the distribution of the convergence points varied between the two.

As the matrix size increases, the significance of the effect of the Markov Chain features decreases, since the average convergence rate of chains that do possess the feature and those that don't, eventually end up similar, hence deeming the effect of these features insignificant in large matrices.

Periodicity of the Markov Chain showed notable effects on the convergence rate of Chains with 3 states. With the increase in the number of aperiodic states within the chain, the chain tends to converge faster.

Markov Chains with transient states tend to converge at the first or second iteration unlike chains with all recurrent states that had an average of 50 as a convergence point. Recurrent chains though tend to not converge at times as observed from the high number of outliers.

Hence, there exists numerous factors that affect the convergence rate of Markov Chains, some more significantly than others, but the behavior of the Chain is dependent on these features, hence the convergence rate as well.

Sources

https://www.sciencedirect.com/topics/mathematics/irreducible-markov-chain

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/video-lectures/lecture-8-markov-eigenvalues-and-eigenvectors/MIT6_262S11_lec08.pdf

http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-MCII.pdf

https://www.math.ucdavis.edu/~gravner/MAT135B/materials/ch13.pdf

https://jindal.utdallas.edu/~metin/Probability/Folios/markovChains.pdf

https://www.probabilitycourse.com/chapter11/11_2_4_classification_of_states.php

https://people.engr.tamu.edu/andreas-klappenecker/csce658-s18/markov_chains.pdf

http://courses.washington.edu/inde411/MarkovChains(part3).pdf

http://math.bme.hu/~nandori/Virtual_lab/stat/markov/Periodicity.pdf

http://www.statslab.cam.ac.uk/~yms/M7_2.pdf

https://www.sciencedirect.com/topics/mathematics/irreducible-markov-chain

https://uu.diva-portal.org/smash/get/diva2:536076/FULLTEXT01.pdf