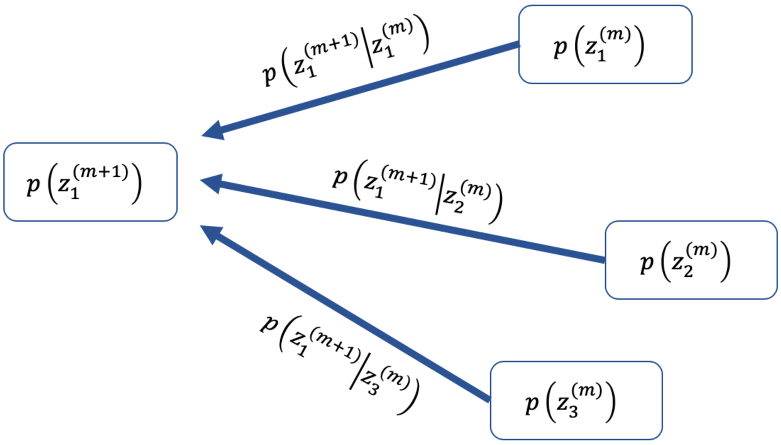
마르코프 체인과 페이지랭크 알고리즘 수식 분석

들어가며

책 또는 논문을 보면 종종 수식이 나온다. 수식이 어려운 것도 있지만 가장 큰 문제는 같은 방정식이라도 저자마다 매개변수를 다르게 사용하고 때로는 무언가를 생략해서 수식을 표현한다. 독자가 책을 건너 뛰어서 수식을 중간부터 봤기 때문에 이해를 못할 수도 있고, 저자가 너무나 당연히 생각해서 생략하는 경우도 있을 것이다. 따라서 본 문서에서는 수식의 기호를 먼저 정의하고 그 수식을 통해 개념들을 설명하겠다. 이 문서의 기본 가정은 어느정도 마르코프 체인와 페이지랭크에 대한 개념은 안다고 가정할 것이다.

1. 마르코프 체인의 수식



[그림 1] 마르코프 체인 예시

[그림 1]은 마르코프 체인의 예시이다. 마르코프 체인는 현재 상태의 확률 $p\left(z_k^{(m)}\right)$와 그 조건부 확률 $p\left(z_k^{(m+1)} \middle| z_k^{(m)}\right)$을 통해서 다음 상태의 확률 $p\left(z_1^{(m+1)}\right)$를 구하는 것이다. 정확한 수식은 다음과 같다.

$$P\left(z_1^{(m+1)}\right) = \sum_k p\left(z_k^{(m+1)} \middle| z_k^{(m)}\right) p\left(z_k^{(m)}\right) \qquad (식\ 1)$$

(식 1)에서 (m)은 현재 상태를 나타내고 $(m+1)$은 다음 상태를 나타내는 변수이다. 여기서 각각의 확률은 벡터 값이 아니라 스칼라 값이다. 일반적으로 마르코프 체인에서 조건부 확률을 $T_{I,j} \equiv p\left(z_i^{(m+1)} \middle| z_j^{(m)}\right)$과 같이 전이 확률(transition probability)로 표현하여 다음 상태의 확률을 표현하면 다음과 같다.

$$P\left(z_1^{(m+1)}\right) = \sum_k T_{1,k} p\left(z_k^{(m)}\right) \qquad (식\ 2)$$

(식 2)의 식의 형태는 스칼라로 표현한 것이다. 이 식을 벡터와 행렬로 표현하면 다음과 같다.

$$\mathrm{T} = \begin{bmatrix} \mathrm{T}_{1,1} & \mathrm{T}_{1,2} & \cdots & \mathrm{T}_{1,j} & \cdots & \mathrm{T}_{1,S} \\ \mathrm{T}_{2,1} & \mathrm{T}_{2,2} & \cdots & \mathrm{T}_{2,j} & \cdots & \mathrm{T}_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{T}_{I,1} & \mathrm{T}_{I,2} & \cdots & \mathrm{T}_{I,j} & \cdots & \mathrm{T}_{I,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{T}_{S,1} & \mathrm{T}_{S,2} & \cdots & \mathrm{T}_{S,j} & \cdots & \mathrm{T}_{S,S} \end{bmatrix} \qquad \text{(식 3)}$$

$$\mathrm{p}\left(\mathrm{z}^{(\mathrm{m}+1)}\right) = \begin{bmatrix} \mathrm{z}_1^{(\mathrm{m}+1)} \\ \mathrm{z}_2^{(\mathrm{m}+1)} \\ \vdots \\ \mathrm{z}_S^{(\mathrm{m}+1)} \end{bmatrix} \qquad \text{(식 4)}$$

$$\mathrm{p}\left(\mathrm{z}^{(\mathrm{m})}\right) = \begin{bmatrix} \mathrm{z}_1^{(\mathrm{m})} \\ \mathrm{z}_2^{(\mathrm{m})} \\ \vdots \\ \mathrm{z}_S^{(\mathrm{m})} \end{bmatrix} \qquad \text{(식 5)}$$

$$\mathrm{p}\left(\mathrm{z}^{(\mathrm{m}+1)}\right) = \mathrm{Tp}\left(\mathrm{z}^{(\mathrm{m})}\right) \qquad \text{(식 6)}$$

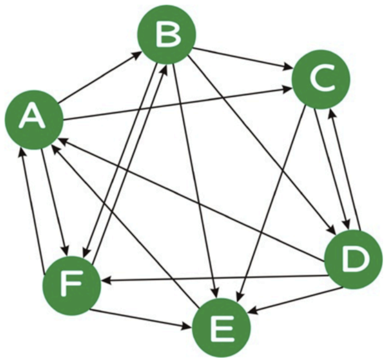
(식 3)은 전이행렬(transition matrix)로 표현하며 (식 4)와 (식 5)는 각각 다음 상태의 벡터와 현재 상태의 벡터이다. 최종적으로 (식 6)은 (식2)를 벡터로 표현한 식이다.

개인적으로 (식 2)와 (식 6)과 같이 스칼라 형태와 벡터 형태의 변환이 자유롭게 되어야 한다고 생각한다. 어떤 개념을 설명하기 위해서는 때로는 스칼라의 형태로 설명하는 것이 쉬울 때가 있고, 때로는 벡터 형태로 설명하는 것이 쉬울 때가 있기 때문이다. 때로는 이러한 변환이 어떤 수식의 본질적인 의미를 볼 수 있게 해줄 수가 있다. 예를 들면 구글의 페이지랭크 논문(링크)에 나온 수식은 다음과 같다.(위키에서 논문에 언급된 수식에 실수가 있는 것 같다고 해서 다음과 같이 수정했다.)

$$PR(A) = \frac{(1-d)}{S} + d\left(\frac{PR(T1)}{C(T1)} + \cdots + \frac{PR(Tn)}{C(Tn)}\right) \qquad \text{(식 6)}$$

(식 6)이 마르코프 체인를 나타내는 식이라고 바로 알아차리기 어려울 것이다. 하지만 위의 수식을 벡터로 변형해보고, 전이 행렬의 본질을 이해 했다면 위의 식이 마르코프 체인의 응용이라는 것을 쉽게 알아차릴 것이다. 여기서 본질을 이해한다는 것은 수학을 통해 수식의 구조 분석을 하고, 이를 통해 어떤 값이라도 전이행렬을 계속 곱하면 반듯이 수렴한다는 것을 알아낼 것이다. 이때 만족하는 조건을 찾아 내는 것이 전이 행렬의 본질을 이해한 것이라고 생각한다.

2. 예시를 통한 마르코프 체인



$$T = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 1 & 1/3 \\ 1/2 & 0 & 0 & 0 & 0 & 1/3 \\ 0 & 1/4 & 0 & 1/3 & 0 & 0 \\ 0 & 1/4 & 1/2 & 0 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/3 & 0 & 1/3 \\ 1/2 & 1/4 & 0 & 0 & 0 & 0 \end{bmatrix}$$

1노드 → A
2노드 → B
3노드 → C
4노드 → D
5노드 → E
6노드 → F

$$T_{i,j} = \frac{1}{j\text{노드에서 나가는 모든 링크의 수}}$$

$T_{i,j}$ 는 j 노드에서 i 노드로 전이될 확률

[그림2] 페이지 랭크 전이행렬 예시

[그림 2]는 페이지랭크 알고리즘에서 정의한 전이행렬이다. 이 전이행렬에 서로 다른 초기값을 가진 값을 곱하면 어떻게 될까? 그 결과는 다음과 같다.

$$p(z^{(0)}) = \begin{bmatrix} 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix} \Rightarrow p(z^{(1)}) = Tp(z^{(0)}) = \begin{bmatrix} 0.2778 \\ 0.1389 \\ 0.0972 \\ 0.1250 \\ 0.2361 \\ 0.1250 \end{bmatrix} \Rightarrow p(z^{(100)}) = T^{100}p(z^{(0)}) = \begin{bmatrix} 0.2709 \\ 0.1970 \\ 0.0788 \\ 0.0887 \\ 0.1798 \\ 0.1847 \end{bmatrix}$$

$$p(y^{(0)}) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow p(y^{(1)}) = Tp(y^{(0)}) = \begin{bmatrix} 0 \\ 0.5 \\ 0 \\ 0 \\ 0 \\ 0.5 \end{bmatrix} \Rightarrow p(y^{(100)}) = T^{100}p(y^{(0)}) = \begin{bmatrix} 0.2709 \\ 0.1970 \\ 0.0788 \\ 0.0887 \\ 0.1798 \\ 0.1847 \end{bmatrix}$$

[그림 3] 서로 다른 초기값의 전이행렬 곱

[그림 3]을 보면 서로 다른 초기값을 가지고 전이행렬을 곱해도 결국 하나의 값으로 수렴한다는 것을 알 수 있다. 그럼 여기서 궁금한게 있다. 이러한 성질은 항상 성립하는 걸까?

3. 마르코프 체인 성질 증명 1 (수렴이 되는 이유)

A는 $S \times S$ matrix

$$A\vec{u}_1 = \lambda_1\vec{u}_1$$
$$A\vec{u}_2 = \lambda_2\vec{u}_2$$
$$\vdots$$
$$A\vec{u}_S = \lambda_S\vec{u}_S$$

$\vec{u}_i^T\vec{u}_i = 1$(단위 벡터)

$\vec{u}_i^T\vec{u}_j = ?$(직교 안할 수 있음)

$\vec{u}_i \neq c\vec{u}_j$, c는 상수(서로 독립)

$A(c\vec{u}_S) = \lambda_S(c\vec{u}_S)$, c는 상수

$$AU = U\Lambda$$
$$U = [\vec{u}_1\ \vec{u}_2 \cdots \vec{u}_S]\ ,\ \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_S \end{bmatrix}$$
$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_S|$$

$$A = U\Lambda U^{-1}$$
$$A^{-1} = U\Lambda^{-1}U^{-1}$$
$$A^n = U\Lambda^n U^{-1}$$
$$\Lambda = U^{-1}AU$$

[그림 4] 고유값과 고유벡터의 간단한 정리

[그림 4]는 고유값과 고유벡터의 일반적인 성질과 수학기호 표시 방법을 정리한 것이다. 마크코프 체인의 경우 다음과 같은 차분 방정식(Difference Equation)으로 표현할 수 있다.

$$X_{n+1} = AX_n,\ \ X_0 = p(z_0) \qquad \text{(식 7)}$$

식 7의 일반해는 다음과 같다.(차분방정식 해법 링크)

$$X_{n+1} = \mathrm{A}X_n, \qquad X_0 = p(z_0)$$
$$X_0 = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_S\vec{u}_S$$
$$X_0 = U\vec{c}, \qquad \vec{c} = [c_1\ c_2 \cdots c_S]^T$$

$$X_n = A^n X_0$$
$$X_n = U\Lambda^n U^{-1}X_0$$
$$X_n = U\Lambda^n\vec{c}$$
$$X_n = \sum_{i=1}^{S} c_i\lambda_i^n\vec{u}_i$$

[그림 5] 차분 방정식 일반해

차분 방정식을 풀기위해서는 고유값과 고유벡터를 구해야 된다는 것을 알 수 있다. 또한 선형종속과 선형독립의 개념을 이해해야 최종적인 일반해 형태를 이해할 수 있다. [그림 5]의 식에서 X_n이 0 벡터가 아닌 값에 수렴하기 위한 조건은 다음과 같다.

- 적어도 1개의 고유값이 1이 되어야 한다.
- 고유값의 크기는 1보다 작아야 한다.

위의 조건이 반드시 성립하는 것을 보이기 위해서 전이행렬의 다음의 특성을 이용할 것이다.

$$\sum_{k=1}^{S} A_{k,j} = 1 \qquad \text{(식 8)}$$

$$A_{i,j} \geq 0$$

식 8은 전이행렬을 만들기 위한 기본 조건이다. 이 조건을 이용하면 고유값의 크기는 1보다 작거나 같다는 것을 증명할 수 있다.

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,j} & \cdots & A_{1,S} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,j} & \cdots & A_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{i,1} & A_{i,2} & \cdots & A_{i,j} & \cdots & A_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{S,1} & A_{S,2} & \cdots & A_{S,j} & \cdots & A_{S,S} \end{bmatrix} \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,j} & \cdots & U_{1,S} \\ U_{2,1} & U_{2,2} & \cdots & U_{2,j} & \cdots & U_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{i,1} & U_{i,2} & \cdots & U_{i,j} & \cdots & U_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{S,1} & U_{S,2} & \cdots & U_{S,j} & \cdots & U_{S,S} \end{bmatrix} = \begin{bmatrix} \lambda_1 U_{1,1} & \lambda_2 U_{1,2} & \cdots & \lambda_j U_{1,j} & \cdots & \lambda_S U_{1,S} \\ \lambda_1 U_{2,1} & \lambda_2 U_{2,2} & \cdots & \lambda_j U_{2,j} & \cdots & \lambda_S U_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_1 U_{i,1} & \lambda_2 U_{i,2} & \cdots & \lambda_j U_{i,j} & \cdots & \lambda_S U_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_1 U_{S,1} & \lambda_2 U_{S,2} & \cdots & \lambda_j U_{S,j} & \cdots & \lambda_S U_{S,S} \end{bmatrix}$$

$$AU = U\Lambda$$

[그림 6] 행렬 기호 정의

[그림 6]은 고유값의 크기는 1보다 작거나 같다는 것을 증명하기 위해 사용된 행렬 기호들이다.

$$\sum_{k=1}^{S} A_{y,k} U_{k,x} = \lambda_x U_{y,x} \qquad \text{(식 9)}$$

$$\sum_{l=1}^{S} \left(\sum_{k=1}^{S} A_{l,k} U_{k,x} \right) = \sum_{l=1}^{S} \lambda_x U_{l,x} \qquad \text{(식 10)}$$

식 9는 행렬의 곱에 대한 결과이고 식 10은 행렬의 곱셈을 한 후 열의 합을 구한식이다. 식 8~10을 이용하면 [그림 7]과 같이 유도할 수 있다.(참고)

$$\sum_{l=1}^{S}|\lambda_x U_{l,x}| = \sum_{l=1}^{S}|\lambda_x||U_{l,x}| = \sum_{l=1}^{S}\left|\sum_{k=1}^{S} A_{l,k}U_{k,x}\right| \leq \sum_{l=1}^{S}\sum_{k=1}^{S} A_{l,k}|U_{k,x}| = \sum_{k=1}^{S}|U_{k,x}|\sum_{l=1}^{S} A_{l,k}$$

$$\sum_{l=1}^{S}|\lambda_x||U_{l,x}| \leq \sum_{k=1}^{S}|U_{k,x}|$$

$$|\lambda_x|\sum_{l=1}^{S}|U_{l,x}| \leq \sum_{k=1}^{S}|U_{k,x}|$$

$$|\lambda_x|(|U_{1,x}| + |U_{2,x}| + \cdots + |U_{S,x}|) \leq |U_{1,x}| + |U_{2,x}| + \cdots + |U_{S,x}|$$

$$|\lambda_x| \leq 1$$

[그림 7] 고유값이 1보다 작거나 같다는 것의 증명

[그림 7]의 증명을 통해서 고유값의 크기는 1보다 작거나 같다는 것을 알았다. 이를 통해 [그림 5]의 식 X_n은 발산하지 않는다는 것을 알 수 있다. 또한 식 8의 특성에 의해서 고유값을 1을 반드시 가지게 된다.

$$(A - I)^T\vec{1} = 0 \qquad \text{(식 11)}$$

식 11은 항상 참인 값이다. 동시에 적어도 하나의 행은 선형종속 관계라는 것을 의미한다. 이 수식을 통해서 적어도 하나의 행은 0으로 이루어지고 이것은 determinant 값이 0이라는 것을 의미하고,(참고) 이는 1이 행렬 A의 고유값이라는 것을 의미한다. 또한 P-F theorem에 의하면 오직 단 하나의 고유값만 1이고 나머지는 1보다 작다고 한다.

$$X_\infty = c_1\vec{u}_1 \qquad \text{(식 12)}$$

따라서 [그림 5]의 차분방정식의 일반해는 수렴하게 되고 그 결과는 식 12와 같게 된다.

4. 마르코프 체인 성질 증명 2 (초기값의 영향이 없는 이유)

다음은 초기값에 상관없이 항상 같은 값으로 수렴되는 이유를 증명하겠다. 증명방법은 [그림 8]과 같다.

$$\mathrm{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,j} & \cdots & A_{1,S} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,j} & \cdots & A_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{i,1} & A_{i,2} & \cdots & A_{i,j} & \cdots & A_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{S,1} & A_{S,2} & \cdots & A_{S,j} & \cdots & A_{S,S} \end{bmatrix} \qquad \sum_{i=1}^{S} A_{i,j} = 1 \qquad A_{i,j} \geq 0$$

$$X_{n+1} = \mathrm{A}X_n \qquad \vec{X}_0 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_S \end{bmatrix} \qquad \sum_{k=1}^{S} x_k = 1$$

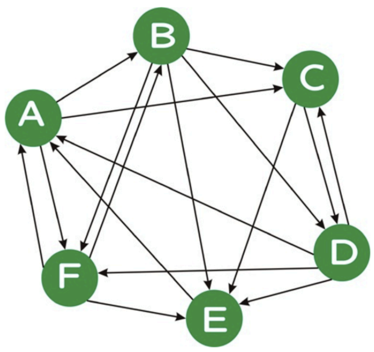
$$X_1 = \mathrm{A}X_0$$

$$\sum_{m=1}^{S} \sum_{l=1}^{S} A_{m,l}\, x_l = \sum_{l=1}^{S} x_l \sum_{m=1}^{S} A_{m,l} = \sum_{l=1}^{S} x_l = 1$$

[그림 8] X_n의 모든 원소의 합이 항상 1인 이유 증명

먼저 AX_0의 모든 원소값의 합이 항상 1인것 을 증명한다. 이는 즉 X_1의 모든 원소값의 합이 항상 1인 것을 의미한다. 그리고 $X_{n+1} = AX_n$의 식에 의해서 모든 X_n의 모든 원소값의 합이 항상 1이 된다는 것을 알 수 있다. 이 성질을 이용하면 식 12의 모든 원소의 합도 1이 된다고 판단해야하며, 이를 통해서 c_1의 값을 초기값에 상관없이 구할 수 있다.

5. 페이지랭크와 마르코프 체인 간의 관계



<Page Rank example>

$$PR(A) = \frac{(1-d)}{S} + d\left(\frac{PR(B)}{C(B)} + \cdots + \frac{PR(F)}{C(F)}\right)$$

$$X_n = [PR(A) \quad PR(B) \quad PR(C) \quad PR(D) \quad PR(E) \quad PR(F)]^T$$

$$J_S = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix},\ S \times S \text{ matrix}$$

T는 페이지랭크 알고리즘의 전이확률

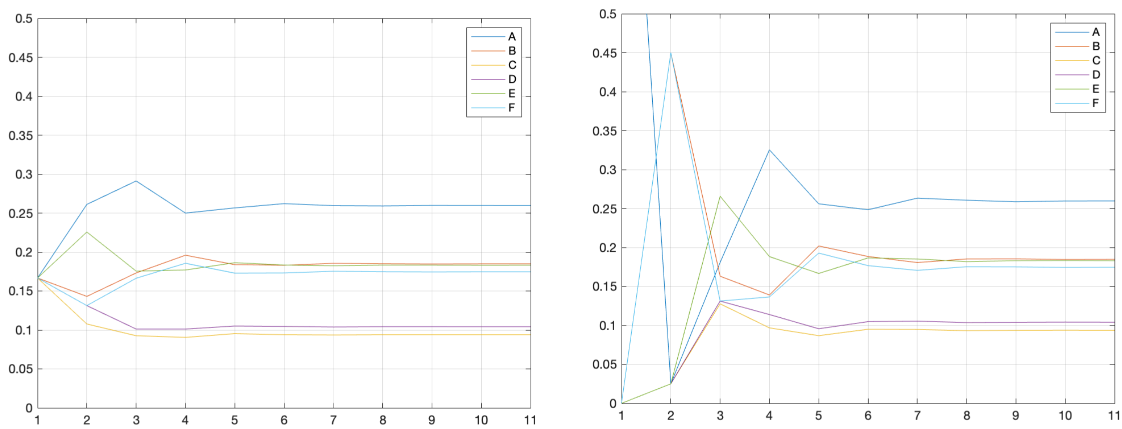
$$X_{n+1} = \frac{(1-d)}{S}\vec{1} + dTX_n,\quad J_S X_n = \vec{1}$$

$$X_{n+1} = \frac{(1-d)}{S} J_S X_n + dTX_n$$

$$X_{n+1} = \left(\frac{(1-d)}{S} J_S + dT\right) X_n$$

[그림 9] 페이지랭크 수식에서 마르코프 체인과의 관계

[그림 9]는 페이지랭크 수식이 마르코프 체인과 관계가 있다는 것을 유도한 수식이다. 이를 유도하기 위해서는 수식의 제약조건들을 활용하면 된다.



$X_n = [1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6]^T$ $\qquad$ $X_n = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^T$

[그림 10] 서로 다른 초값의 페이지랭크 값의 변화

[그림 10]은 식 7에서 서로 다른 초기값의 페이지랭크 값의 변화를 그래프로 그린

것이다. [그림 10]의 왼쪽 그래프는 초기값이 모두 같을 경우이다. [그림 10]의 오른쪽 그래프는 A노드의 초기값이 1이고 나머지는 모두 0인 경우이다. 이것으로 초기값에 상관 없이 하나의 값으로 수렴한다는 것을 알 수 있다.

마치며

시작은 마르코프 체인을 이해하고 실제 적용된 알고리즘을 통하여 수학이 왜 필요한지 결론을 낼려고 했다. 하지만 이해를 하면할 수록 명확한 결론을 얻지 못했다. 오히려 생각이 더 많아졌다. 하나의 예시를 통해서 수학이 왜 필요한지 설명하는 것은 지금은 못할 것 같다. 다음은 몬테카를로 마코프 체인 알고리즘에서 random work 현상을 피하기 위해서 헤밀토니안 방정식을 어떻게 적용하는지 정리 하겠다.