DSL Seminar: MCMC (2)

Kyung-han Kim

Data Science Lab

January, 2023

목차

확률 과정 (Stochastic Process)

- 확률과정 $\{X_t, t \in T\}$은 확률변수를 나열해 놓은 것입니다 (Collection of random variables).
- 첨자 t는 주로 시간을 나타냅니다. 이 경우 확률과정은 시간의 변화에 따른 확률변수 값의 변화를 나타낸다고 볼 수 있습니다.
- X_t는 시점 t에서의 확률과정의 state(상태)를 의미합니다.
  - X_t의 집합은 유한집합일 수도, 무한집합일 수도 있지만 이해를 돕기 위해 설명 과정에서는 유한한 상황을 가정하겠습니다.
- 만약 T가 유한집합 혹은 가산집합 (countable) 이라면, 그 확률과정은 discrete-time stochastic process입니다.
- 만약 T가 무한집합이라면, 그 확률과정은 continuous-time stochastic process입니다.
- 저희는 Discrete-time stochastic process만 다룹니다.

Markov Chain

- 일반적인 확률과정에서, X_n 값은 $X_0, X_1, ..., X_{n-1}$의 영향을 받습니다.
- 마르코프 연쇄 (Markov Chain, MC)는 X_n이 X_{n-1}에만 영향을 받는 특수한 형태의 확률과정입니다.
- 즉, Markov Chain에서는

$$P(X_n = j | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i) = P(X_n = j | X_{n-1} = i)$$

  이 성립합니다.
- 다르게 말하자면 $X_0, X_1, ..., X_{n-2}$는 X_{n-1}이 주어져 있다면 X_n과 독립입니다.
- 한 시점의 값이 직전 시점 값에만 영향을 받는 확률 과정!

Markov Chain의 예시

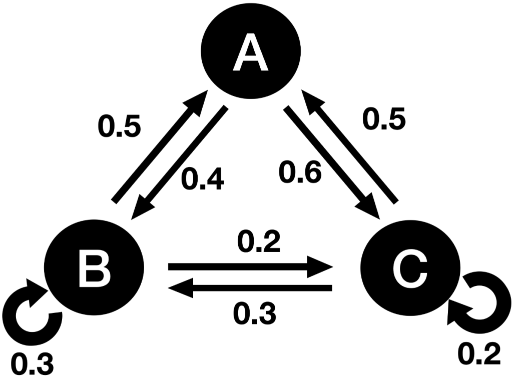

Figure 1: Example of Markov Chain

Transition Probability Matrix $(\mathbf{P})$

- $p_{ij} = P(X_{n+1} = j | X_n = i)$: one-step transition probability
  - $p_{ij} \geq 0$, $\sum_{j=0}^{\infty} p_{ij} = 1$
- $\mathbf{P} = (p_{ij})$: one-step transition probability matrix
  - $\mathbf{P}$의 각 행의 원소들의 합은 항상 1이다.
  - 각 열의 원소들의 합은 1이 아닐 수도 있다.
- n-step transition probability matrix $\mathbf{P}^{(n)} = \mathbf{P} \times \mathbf{P}^{(n-1)}$.
- $\mathbf{P}^{(n)} \times \mathbf{P}^{(m)} = \mathbf{P}^{(n+m)}$. (Chapman-Kolmogorov Equation)
- Markov chain은 아래의 두 가지를 알면 완전히 정의됩니다:
  - 1) 초깃값 (X_0)
  - 2) One-step transition probability matrix $(\mathbf{P})$

1) Irreducibility

- Irreducible Markov chain은 class가 1개인 Markov chain입니다.
- 이는 어떤 두 개의 state를 고르든지 그 두 state가 서로 접근 가능하다 (accessible)는 의미입니다.
- 다시 말해, 만약 어떤 Markov Chain이 irreducible하다면, 모든 state i, j에 대해 i^{th} state에서 시작해서 어떻게든 j^{th} state에 도달할 수 있습니다.
  - 주의] $p_{ij} > 0$인 것과 위 내용은 다릅니다.
- Irreducible Markov chain은 항상 stationary distribution을 가지며, 그 stationary distribution은 유일합니다 (unique).
- 만약 irreducible Markov chain이 자기 자신으로 돌아오는 경로를 하나라도 가지고 있다면 (즉, $p_{ii} > 0$인 i가 존재한다면) 그 Markov chain은 limiting distribution을 가집니다.

2) Limiting Distribution

- Limiting distribution π_j 는 어떤 Markov chain이 시점 $n \to \infty$ 일 때 state j 에 있을 확률을 의미합니다.

$$\pi_j = \lim_{n\to\infty} (\mathbf{P}^{(n)})_{ij} = \lim_{n\to\infty} P(X_n = j | X_0 = i)$$

- Limiting distribution은 초깃값 (X_0) 의 영향을 받지 않습니다.
- Limiting distribution은 존재할 수도 있고, 존재하지 않을 수도 있습니다.
  - Irreducible Markov chain이 자기 자신으로 돌아오는 경로를 하나라도 가지고 있으면 항상 limiting distribution을 갖습니다.

- Stationary distribution π_j 는 어떤 Markov chain이 특정한 state j 를 장기적으로 방문하는 비율(proportion) 내지는 확률(probability)로 정의됩니다.

$$\pi_j = \lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n} P(X_t = j | X_0 = i)$$

- Stationary distribution은 $\pi = \pi \mathbf{P}$ 를 만족시키므로,
이 식을 이용해 stationary distribution을 찾을 수 있습니다.
($\pi = [\pi_0, \pi_1, \cdots, \pi_N]$, N 은 전체 state의 개수)

Limiting Distribution vs. Stationary Distribution

- (1) Limiting distribution:
  - Markov chain이 Irreducible하더라도 존재하지 않을 수 있습니다.
    - Aperiodicity까지 만족시켜야 합니다
  - 단일 시점에서 정의되는 개념입니다.

$$\pi_j = \lim_{n \to \infty} (\mathbf{P}^{(n)})_{ij} = \lim_{n \to \infty} P(X_n = j | X_0 = i)$$

- (2) Stationary distribution:
  - Irreducible Markov chain은 언제나 stationary distribution을 가집니다.
  - 확률과정 전체에서 특정 state를 방문하는 비율로 정의됩니다.

$$\pi_j = \lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n} P(X_t = j | X_0 = i)$$

- 만약 어떤 Markov chain이 irreducible하면서 aperiodic하다면, stationary distribution과 limiting distribution이 같아집니다.

4) Aperiodicity

- Aperiodic Markov chain은 모든 state의 주기 (period) 가 1인 Markov chain입니다.
- i^{th} state의 주기는 $(\mathbf{P}^{(n)})_{ii} > 0$을 만족시키는 모든 자연수 n의 최대공약수로 정의됩니다.
- 만약 $period \geq 2$라면, Markov chain이 특정한 패턴을 가지고 움직이게 됩니다. 즉, 주기성이 생깁니다.
- Irreducible Markov chain이 자기 자신으로 돌아오는 경로를 적어도 하나 가지고 있다면, 그 Markov chain은 반드시 Aperiodic합니다.

5) Ergodicity

- Ergodic $=$ Irreducible $+$ Aperiodic $(+$ Positive Recurrent$)$
- Ergodic Markov chain은 항상 유일한 stationary distribution을 가지며, 이것을 limiting distribution으로도 해석할 수 있습니다.
- 정리하자면,
  - 1) Irreducible: state i 에서 출발해서 state j 에 도달할 수 있다.
  - 2) Aperiodic: Markov chain에 주기성이 없다. 특정한 패턴이 없다.
  - 3) Positive Recurrent: Markov chain이 state i에서 출발하면, 유한 시간 안에 다시 state i로 돌아올 수 있다.

- 원래의 stationary distribution과 reversed version의 stationary distribution이 같아지는 Markov chain입니다.
- 더 정확하게 표현하자면, time-reversible Markov chain은

$$P(X_n = j|X_{n+1} = i) = P(X_{n+1} = j|X_n = i)$$

  을 만족합니다.
- $P(X_n = j|X_{n+1} = i) = \frac{P(X_n=j)P(X_{n+1}=i|X_n=j)}{P(X_{n+1}=i)}$ 이므로, time-reversible MC는 $\frac{\pi_j p_{ji}}{\pi_i} = p_{ij}, \pi_j p_{ji} = \pi_i p_{ij}$ 을 만족합니다.
- 위 조건식을 **detailed balance condition**이라고 합니다.

야바위는 Markov Chain?

- 통계전산에서 사용한 야바위 예시를 같이 살펴 봅시다!
  1. 컵이 3개, 공이 1개일 때의 야바위는 Markov Chain인가?
  2. 만약 그렇다면 Transition probability matrix $\mathbf{P}$를 구하여라.
  3. 이 Markov chain은 Irreducible한가?
  4. 이 Markov chain은 Aperiodic한가?
  5. 이 Markov chain은 Time-reversible MC인가?

Monte Carlo Methods

- Monte Carlo 방법론은 random simulation을 통해 문제를 해결하는 방법론을 의미합니다.
- 통계전산 수업에서는 그 중 Monte Carlo Integration을 주로 사용합니다.
- Monte Carlo Integration은 $\int_0^1 g(x)dx$의 근사값을 대수의 법칙을 이용해 구하는 방법입니다.

$$\int_0^1 g(x)dx = \int_0^1 1 \times g(x)dx = \int_0^1 f(x)g(x)dx = E[g(U)]$$

가 성립하므로 (단, $f(x)$는 U(0,1)의 pdf), $n \to \infty$이면

$$\sum_{i=1}^{n} \frac{g(U_i)}{n} \to E[g(U)] = \theta$$

Example of Monte Carlo Integration: Normal PDF

- 정규분포의 pdf를 0부터 ∞까지 적분하면 0.5가 됨을 Monte Carlo Integration을 통해 확인해 봅시다.

$$\int_0^\infty \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)dx\text{에서 } y = \frac{x}{x+1}\text{로 놓으면}$$

$$x = -\frac{y}{y-1}\text{이므로 } \frac{dx}{dy} = \frac{1}{(y-1)^2}\text{이다.}$$

$$\text{즉, } \int_0^\infty \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)dx = \int_0^1 \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(-\frac{y}{y-1})^2)\frac{1}{(y-1)^2}dy$$

$$= \mathsf{E}[\frac{1}{\sqrt{2\pi}}\frac{1}{(Y-1)^2}\exp(-\frac{Y^2}{2(Y-1)^2})],\ Y \sim \mathsf{U}(0,1)$$

Example of Monte Carlo Integration: Normal PDF (cont'd)

- 실제로 R에서 U(0,1)을 따르는 난수를 생성하고, 그 난수와 앞서 구한 식을 이용하면 0.5에 가까운 값이 계산되는지 확인해 봅시다.

Example of Monte Carlo Integration: π estimation

- 반지름의 길이가 1이면서 원점을 중심으로 하는 원과 $-1 \leq x \leq 1$, $-1 \leq y \leq 1$ 범위의 정사각형을 생각해 봅시다.
- 그러면 정사각형의 넓이는 4, 반지름의 넓이는 π가 됩니다.
- 만약 $X \sim Unif(-1, 1)$, $Y \sim Unif(-1, 1)$이라면 점 (X, Y)가 단위원 안에 들어갈 확률은 $\frac{\pi}{4}$가 될 것입니다.
- 정리하자면,

$$\frac{\sum_{i=1}^{n} I(x_i^2 + y_i^2 \leq 1)}{n} \to E[I(X^2+Y^2 \leq 1)] = P(X^2+Y^2 \leq 1) = \frac{\pi}{4},$$

$$4 \times \frac{\sum_{i=1}^{n} I(x_i^2 + y_i^2 \leq 1)}{n} \to \pi$$

가 성립합니다.

Example of Monte Carlo Integration: π estimation (cont'd)

Algorithm

1. Pick n.
2. Generate $(u_{11}, u_{21}), \cdots, (u_{1n}, u_{2n})$, where u_{1i}, u_{2i} are pseudo r.n. on $(0, 1)$.
3. $\pi = 4 \times \frac{\#[(2u_1-1)^2+(2u_2-1)^2 \leq 1]}{n}$