DSL 베이지안 통계 <기초반>

Monte Carlo Estimation

& Markov Chains

발표자 : 5기 박채은

몬테카를로 추정

몬테카를로 방법을 이용하여 알고 싶은 값을 추정하는 것

- 몬테카를로 방법

무작위 추출된 난수(random number)를 이용하여 함수의 값을 계산하는 통계학적 방법

```
theta = rgamma(n=m, shape=a, rate=b)
mean(theta) # sample mean
```

- R을 이용한다면?

1. rgamma, rbeta 등을 이용하여 random하게 sample을 뽑는다.
2. 뽑힌 sample을 사용해서 연산을 하여 구하고자 하는 값을 추정한다.

- Important quantities 예시)

평균, 분산, 어떤 사건의 확률, 백분위수(quantiles)

몬테카를로 추정이 필요한 이유는?

Integration을 이용한 계산이 힘들어서

예시) θ ~ Gamma(a, b)

평균: $$\mathrm{E}(\theta) = \int_0^\infty \theta f(\theta) d\theta = \int_0^\infty \theta \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta \, .$$

분산: $$\mathrm{Var}(\theta) = \int_0^\infty (\theta - E(\theta))^2 f(\theta) d\theta \, .$$

계산 복잡!

=> 이렇게 하지 말고 분포에서 sample을 뽑아서 원하는 값을 추정하자

1. Monte Carlo estimation

=> 분포에서 sample m개을 뽑아서 표본 평균과 표본 분산을 계산 가능 -> 추정

θ ~ Gamma(a, b)

θ_i^* i = 1, 2, ⋯, m

표본 평균: $\bar{\theta}^* = \frac{1}{m}\sum_{i=1}^{m}\theta_i^*$ ⟶

표본 분산: $\frac{1}{m}\sum_{i=1}^{m}(\theta_i^* - \bar{\theta}^*)^2$

CLT

$$\bar{X} \sim N(E(X), \frac{Var(X)}{m})$$

몬테카를로 추정을 통해 다양한 integrals 계산 가능

예) 어떤 사건의 확률

$$\boxed{\int_0^\infty I_{[\theta<5]}(\theta)p(\theta)d\theta} = \int_0^5 1 \cdot p(\theta)d\theta + \int_5^\infty 0 \cdot p(\theta)d\theta = \boxed{P(\theta < 5)\,.}$$

평균 구하기 : $E(I_{[\theta<5]}\theta)$

θ ~ Gamma(a, b)

θ_i^* i = 1, 2, ⋯, m $\longrightarrow$ $\frac{1}{m}\sum_{i=1}^{m} I_{\theta^*<5}(\theta_i^*)$

몬테카를로 추정을 통해 다양한 integrals 계산 가능

예) 백분위수 구하기

$$P(\theta < z) = 0.9$$

z 찾고 싶다!

θ ~ Gamma(a, b)

θ_i^* i = 1, 2, ⋯, m

$\theta_{[1]}^*, \theta_{[2]}^*, \theta_{[3]}^*, \ldots\ldots\ldots\ldots\ldots\ldots, \theta_{[m]}^*$

Monte Carlo error(=Monte Carlo standard deviation)

Monte Carlo 방법을 이용해서 구한 approximation이 실제 값과 얼마나 차이가 있나

예) E(θ)를 알고 싶을 때 <= 표본 평균으로 추정할 수 있다

$$\overline{\theta^*} \sim N(E(\theta), \frac{Var(\theta)}{m}) \quad \text{by CLT}$$

Monte Carlo standard deviation(se): $\boxed{\frac{sd(\theta^*)}{\sqrt{m}}}$

만약 m이 크다면, 실제 값이
추정치 ± 2*se 안에 있다고 생각 가능

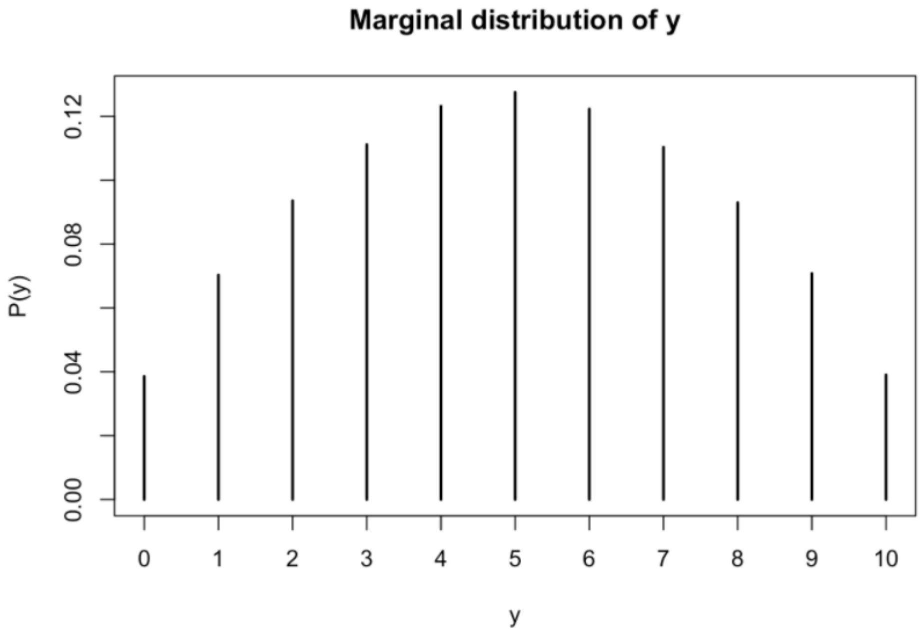
Marginalization

$$\{(y_1^*, \phi_1^*), \cdots, (y_m^*, \phi_m^*)\}$$

몬테카를로 추정의 장점:
결합된 데이터가 생성되어 있다면 marginalization 쉽게 할 수 있다
=> phi 버리고 y만 marginal dist'n의 sample로 보기

$$p(y) = \int_0^1 p(y, \phi) d\phi$$

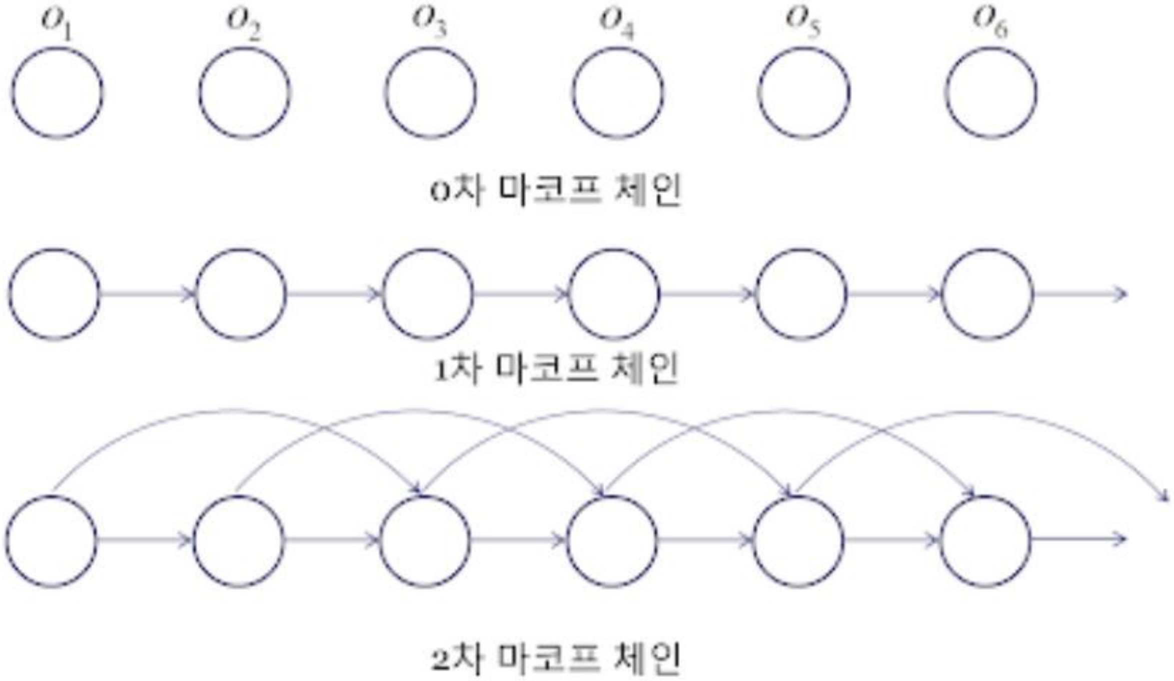
Integral로 계산하는 것은 힘듦.



Markov Chain

마르코프 성질: 각 상태는 이전의 상태에 영향을 받는다.
각 상태는 바로 이전의 상태에만 영향을 받는다.(1차 마르코프 체인)

0차 마르코프 체인
1차 마르코프 체인
2차 마르코프 체인



Markov Chain

Markov assumption: X_{t+1} 에서의 확률분포는 X_t에 의해서만 결정되어야 한다.

$p(X_{t+1}|X_t, X_{t-1}, \cdots, X_2, X_1) = p(X_{t+1}|X_t)$ for all $t=2,\cdots,n$

$$p(X_1, X_2, \cdots, X_n) = p(X_1)\cdot p(X_2|X_1)\cdot p(X_3|X_2, X_1)\cdot \cdots \cdot p(X_n|X_{n-1}, X_{n-2}, \cdots, X_2, X_1)$$

$$p(X_1, X_2, \cdots X_n) = p(X_1)\cdot p(X_2|X_1)\cdot p(X_3|X_2)\cdot p(X_4|X_3)\cdot \cdots \cdot p(X_n|X_{n-1})$$

가정 하나 더) 전이 확률은 시간에 따라 바뀌지 않는다.

Markov Chain 예시

Discrete Markov chain 예시

초기 숫자 정한 후(1, 2, 3, 4, 5만 가능) 동전 던져서 앞면 나오면 이전 숫자 +1, 뒷면 나오면 이전 숫자 -1로 만드는 시행을 n번 반복
상태: 가능한 숫자 1, 2, 3, 4, 5

Continuous Markov chain 예시

Random walk: $X_t = 0$, p(Xt+1|Xt=xt)=N(xt,1)
즉 다음 상태의 확률 분포가 평균이 현재 상태, 분산이 1인 정규분포 따른다.

Transition matrix

각각의 상태에 따른 확률: 전이 확률
전이 확률로 만든 행렬: 전이 행렬
상태에서 상태로 변하는 것을 행렬의 행, 열 성분을 통해 표현함.

예) discrete Markov chain 예시 이어서
전이 확률: 25개 존재 가능
p(Xt+1=5|Xt=4)

$$Q = \begin{pmatrix} 0 & .5 & 0 & 0 & .5 \\ .5 & 0 & .5 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & .5 & 0 & \boxed{.5} \\ .5 & 0 & 0 & .5 & 0 \end{pmatrix}$$

Transition matrix

각각의 상태에 따른 확률: 전이 확률
전이 확률로 만든 행렬: 전이 행렬
상태에서 상태로 변하는 것을 행렬의 행, 열 성분에 표현함.

$$Q = \begin{pmatrix} 0 & .5 & 0 & 0 & .5 \\ .5 & 0 & .5 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ .5 & 0 & 0 & .5 & 0 \end{pmatrix}$$

예) discrete Markov chain 예시 이어서

Chain에서 여러 step 이후의 확률 구하고 싶을 때 전이 행렬 유용
P(Xt+2=3|Xt=1)

$$Q^2: \begin{array}{cccccc} & [,1] & [,2] & [,3] & [,4] & [,5] \\ [1,] & 0.50 & 0.00 & \boxed{0.25} & 0.25 & 0.00 \\ [2,] & 0.00 & 0.50 & 0.00 & 0.25 & 0.25 \\ [3,] & 0.25 & 0.00 & 0.50 & 0.00 & 0.25 \\ [4,] & 0.25 & 0.25 & 0.00 & 0.50 & 0.00 \\ [5,] & 0.00 & 0.25 & 0.25 & 0.00 & 0.50 \end{array}$$

Stationary distribution

Stationary: 움직이지 않는, 정지한

평형 상태에 도달한 state의 확률분포

먼 미래에는 현재 상태의 중요도가 떨어짐.

초기값에 연연하지 않음.

모든 Markov chain에 존재하는 것은 아님. 수렴하는 조건이 있음.

```
c(0.2, 0.2, 0.2, 0.2, 0.2) %*% Q
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.2  0.2  0.2  0.2  0.2
```

Simulating long Markov chain

```
table(x) / n
```

```
## x
##      1      2      3      4      5
## 0.1996 0.2020 0.1980 0.1994 0.2010
```

$\approx$

Stationary distribution

```
c(0.2, 0.2, 0.2, 0.2, 0.2) %*% Q
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.2  0.2  0.2  0.2  0.2
```


THANK YOU :)

Marginalization

Hierarchical model에서도 Monte Carlo sample을 얻을 수 있다

예) Likelihood: $y|\phi \sim iid \ \mathrm{Bin}(10, \phi)$
Prior: $\phi \sim \mathrm{Beta}(2,2)$

Joint distribution:
$p(y, \phi) = p(y|\phi)p(\phi)$

θ_i^* from beta(2, 2) ➡ y_i^* from Bin(10, θ_i^*) => $\{(y_1^*, \phi_1^*), \cdots, (y_m^*, \phi_m^*)\}$

m번 반복