베이즈데이터분석 / 이재용 교수

10강

메트로폴리스 - 헤이스팅스 알고리듬

한국방송통신대학교

목차

- 메트로폴리스-헤이스팅스 알고리듬
- 정규분포의 예
- 코시 모형의 예

목차

- 메트로폴리스-헤이스팅스 알고리듬
- 정규분포의 예
- 코시 모형의 예

메트로폴리스-헤이스팅스 알고리듬의 유도

목표

주어진 목표 분포 $\pi(x)$를 정상분포(stationary distribution)으로 갖는 마르코프체인의 커널 $K(x, dy)$를 구하고자 한다.

상세평형조건(Detailed balance condition)

앞에서 상세평형조건

$$\pi(dx)K(x, dy) \ = \ \pi(dy)K(y, dx), \forall x, y \ \in \ S$$

을 만족하면 π가 커널 $K(x, dy)$의 정상분포가 된다.

문제

$q(x,y)$가 주어진 임의의 커널이라고 하자.

$q(x,y)$로부터 상세평형조건을 만족하는 커널 $K(x,y)$를 구할 수 있나?

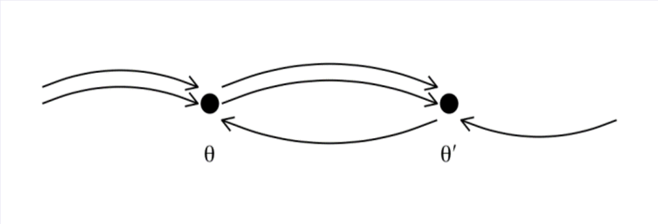
메트로폴리스-헤이스팅스 알고리듬의 유도

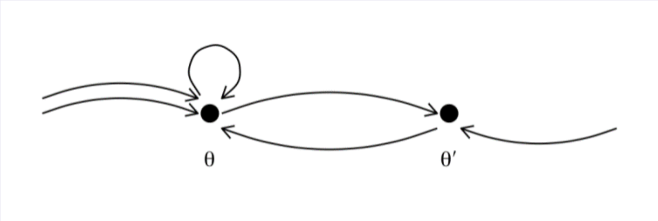
- $\pi(x)q(x,y) > \pi(y)q(y,x)$이라고 가정하자. $\alpha(x,y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}$라고 하면,

$$\pi(x)q(x,y)\alpha(x,y) = \pi(y)q(y,x)$$

  가 만족한다.

- 위의 식을 이용해서 커널을 구할 수 있다. 상세평형조건이 성립하지 않게 하는 확률만큼 자기 자신으로 되돌린다.





메트로폴리스-헤이스팅스 알고리듬

단계 1 (초기화)$x^{(0)}$를 정한다.

단계 2 (메트로폴리스 – 헤이스팅스 반복)

$t = 1, 2, \ldots, m$에 대해서 다음을 수행한다.

(i) (후보값과 균등확률변수 추출)서로 독립이 되도록 x, u를 추출한다.

$$x \sim q\left(x^{(t-1)}, \cdot\right),$$

$$u \sim U(0, 1).$$

(ii) (합격확률의 계산)

$$\alpha(x^{(t-1)}, x) = \min\{1, \frac{\pi(x)q(x, x^{(t-1)})}{\pi(x^{(t-1)})q(x^{(t-1)}, x)}\}$$

를 계산한다.

(iii) $x^{(t)}$값의 결정

$$x^{(t)} = \begin{cases} x, & \text{if } u \le \alpha(x^{(t-1)}, x) \\ x^{(t-1)}, & \text{if } u > \alpha(x^{(t-1)}, x). \end{cases}$$

임의보행 메트로폴리스-헤이스팅스(random-walk MH)

0에 대해 대칭인 분포 g에 대해,

제안 커널이 $q(x^{(t-1)}, \cdot) = g(\cdot - x^{(t-1)})$의 형태를 가지면,

이를 통해 생성되는 MH 커널을 임의보행메트로폴리스 커널이라고 한다.

예

많이 쓰는 예는 $q(x^{(t-1)}, x) = U(x^{(t-1)} - d, x^{(t-1)} + d)$와

$N(x^{(t-1)}, d^2)$가 있다. 여기서 $d > 0$이다.

독립 메트로폴리스-헤이스팅스(independent MH)

어떤 분포 g에 대해, 제안 커널$q(x^{(t-1)},\cdot) = g(\cdot)$와 같아서,

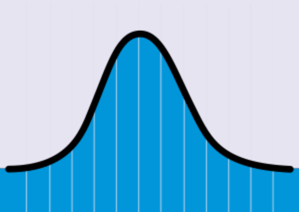
체인의 이전 값 $x^{(t-1)}$에 의존하지 않을 때,

이로부터 생성되는 MH 커널을 독립 MH커널이라고 한다.

이 경우, $g(\cdot) \approx \pi(\cdot)$인 것이 좋다.

목차

- 메트로폴리스-헤이스팅스 알고리듬
- 정규분포의 예
- 코시 모형의 예



예. 정규분포. 임의보행 메트로폴리스

목표 밀도함수

$$\pi(\theta) \propto e^{-\frac{1}{2}\theta^2}, \theta \in \mathbb{R}$$

제안 전이핵: 임의보행 메트로폴리스

$$Q(\theta^{(t-1)}, \theta^*) = N(\theta^* | \theta^{(t-1)}, d^2), d > 0$$

$$q(\theta^{(t-1)}, \theta^*) = e^{-\frac{1}{2d^2}(\theta^* - \theta^{(t-1)})^2}$$

예. 정규분포. 임의보행 메트로폴리스

합격 확률

$$\alpha\left(\theta^{(t-1)}, \theta^*\right) = \min\left\{1, \frac{\pi(\theta^*)q\left(\theta^*, \theta^{(t-1)}\right)}{\pi(\theta^{(t-1)})q(\theta^{(t-1)}, \theta^*)}\right\}$$

$$= \min\{1, e^{\frac{1}{2}(\theta^{(t-1)})^2 - (\theta^*)^2}\}.$$

알고리듬

단계 1 (초기화) $\theta^{(0)}$와 $d > 0$를 정한다.

단계 2 (메트로폴리스 – 헤이스팅스 반복)

$t = 1, 2, \ldots, m$에 대해서 다음을 수행한다.

예. 정규분포. 임의보행 메트로폴리스

(i) (후보값과 균등확률변수 추출) 서로 독립이 되도록 θ^*와 u를 추출한다.

$$\theta^* \sim N(\theta^{(t-1)}, d^2),$$
$$u \sim U(0,1).$$

(ii) (합격확률의 계산)

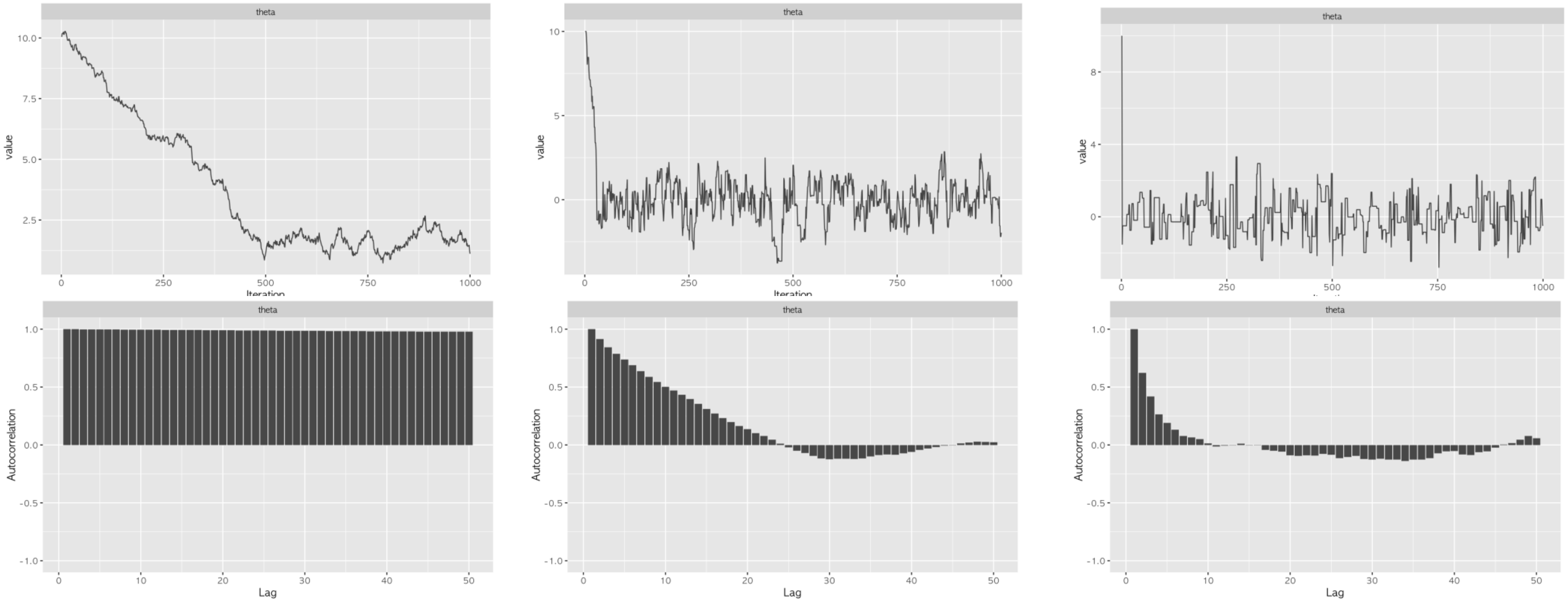
$$\alpha(\theta^{(t-1)}, \theta^*) = \min\{1, e^{\frac{1}{2}(\theta^{(t-1)})^2 - (\theta^*)^2}\}$$

를 계산한다.

(iii) $\theta^{(t)}$값의 결정

$$\theta^{(t)} = \begin{cases} \theta^*, & \text{if } u \le \alpha\left(\theta^{(t-1)}, \theta^*\right) \\ \theta^{(t-1)}, & \text{if } u > \alpha(\theta^{(t-1)}, \theta^*). \end{cases}$$

임의보행 제안커널의 표준편차와 수렴성



왼쪽부터 제안커널의 표준편차는 $d = 0.1, 1, 4$이고

합격 확률은 $0.88, 0.699, 0.286$이다.

예. 정규분포. 독립 메트로폴리스

목표 밀도함수

$$\pi(\theta) = e^{-\frac{1}{2}\theta^2}, \theta \in \mathbb{R}$$

제안 전이핵: 독립 메트로폴리스

$$Q(\theta^{(t-1)}, \theta^*) = N(\theta^* | \mu, d^2), \mu \in \mathbb{R}, d > 0$$

$$q(\theta^{(t-1)}, \theta^*) = e^{-\frac{1}{2d^2}(\theta^* - \mu)^2}$$

예. 정규분포. 독립 메트로폴리스

합격 확률

$$\alpha\left(\theta^{(t-1)}, \theta^*\right) = \min\left\{1, \frac{\pi(\theta^*)q\left(\theta^*, \theta^{(t-1)}\right)}{\pi(\theta^{(t-1)})q(\theta^{(t-1)}, \theta^*)}\right\}$$

$$= \min\left\{1, \frac{e^{-\frac{1}{2}(\theta^*)^2} e^{-\frac{1}{2d^2}(\theta^{(t-1)}-\mu)^2}}{e^{-\frac{1}{2}(\theta^{(t-1)})^2} e^{-\frac{1}{2d^2}(\theta^*-\mu)^2}}\right\}.$$

$$= \min\left\{1, exp\left[-\frac{1}{2}\left((\theta^*)^2 - \left(\theta^{(t-1)}\right)^2\right) - \frac{1}{2d^2}(\theta^{(t-1)} - \theta^*)(\theta^{(t-1)} + \theta^* - 2\mu)\right]\right\}$$

목차

- 메트로폴리스-헤이스팅스 알고리듬
- 정규분포의 예
- 코시 모형의 예

코시 모형

모형

$$x_1, \ldots, x_n | \mu, \sigma \sim Ca(\mu, \sigma), \mu \in \mathbb{R}, \sigma > 0$$

사전분포

$$\pi(\mu, \sigma) d\mu d\sigma = \frac{1}{\sigma} d\mu d\sigma$$

위 모형의 사후분포를 임의보행 메트로폴리스 알고리듬으로 근사해보자.

코쉬 모형: 사후분포의 유도

- 코쉬분포의 밀도함수는

$$f\,(x|\mu,\sigma)\ =\frac{1}{\sigma}\frac{1}{1\ +\left(\frac{x-\mu}{\sigma}\right)^{2}},x\ \in\ \mathbb{R}$$

이다.

- 사전분포는 부적절 사전분포이므로,
  사후분포가 적절 분포임을 보여야 한다.
  여기서는 적절 분포라는 것을 가정하고 시작한다.

- (사후분포)

$$\pi(\mu,\sigma|x) = \pi_n(\mu,\sigma)$$

$$\propto \frac{1}{\sigma}\prod_{i=1}^{n}\frac{1}{\sigma}\frac{1}{1+\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\propto \frac{1}{\sigma^{n+1}}\prod_{i=1}^{n}\frac{1}{1+\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

코쉬 모형: 사후분포의 유도

- (변수의 변환) $\xi = \log\sigma$로 변수를 변환한다.

$$d\xi = \frac{1}{\sigma}d\sigma, \sigma = e^{\xi}, d\sigma = \sigma d\xi = e^{\xi}d\xi$$

임을 이용한다. 사후분포를 π_n으로 표시하자.

$$\pi_n(\mu,\sigma)d\mu d\sigma$$

$$= \frac{1}{\sigma^{n+1}}\prod_{i=1}^{n}\frac{1}{1 + \left(\frac{x_i - \mu}{\sigma}\right)^2}d\mu d\sigma$$

$\sigma = e^{\xi}, d\sigma = e^{\xi}d\xi$로 대치한다.

$$= e^{-(n+1)\xi}\prod_{i=1}^{n}\frac{1}{1 + e^{-2\xi}(x_i - \mu)^2}d\mu e^{\xi}d\xi$$

코쉬 모형: 사후분포의 유도

$$= e^{-n\xi} \prod_{i=1}^{n} \frac{1}{1 + e^{-2\xi}(x_i - \mu)^2} d\mu d\xi$$

이다. 따라서,

$$\pi_n(\mu, \xi) = e^{-n\xi} \prod_{i=1}^{n} \frac{1}{1 + e^{-2\xi}(x_i - \mu)^2}$$

이다. 이를 이용해 사후표본을 추출한다.

코쉬 모형: 임의보행 알고리듬

- (μ의 추출을 위한 제안커널과 합격확률)

$\mu^{(t-1)}$이 주어져 있을 때, μ의 제안커널로 $N(\mu^{(t-1)}, b^2), b > 0$을 쓰기로 한다.

μ의 합격확률을 계산해보자.

$$\alpha(\mu^{(t-1)}, \mu) = \min\left\{1, \frac{\prod_{i=1}^{n} \frac{1}{1+e^{-2\xi}(x_i-\mu)^2} \frac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2b^2}(\mu^{(t-1)}-\mu)^2}}{\prod_{i=1}^{n} \frac{1}{1+e^{-2\xi}(x_i-\mu^{(t-1)})^2} \frac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2b^2}(\mu-\mu^{(t-1)})^2}}\right\}$$

$$= \min\left\{1, \prod_{i=1}^{n} \frac{1+e^{-2\xi}(x_i-\mu^{(t-1)})^2}{1+e^{-2\xi}(x_i-\mu)^2}\right\}$$

- (ξ의 추출을 위한 제안커널과 합격확률)

$\xi^{(t-1)}$이 주어져 있을 때, ξ의 제안커널로 $N\left(\xi^{(t-1)}, d^2\right), d > 0$을 쓰기로 한다.

ξ의 합격확률을 계산해보자.

$$\alpha\left(\xi^{(t-1)}, \xi\right) = \min\left\{1, \frac{e^{-n\xi} \prod_{i=1}^{n} \frac{1}{1+e^{-2\xi}(x_i-\mu)^2}}{e^{-n\xi^{(t-1)}} \prod_{i=1}^{n} \frac{1}{1+e^{-2\xi^{(t-1)}}(x_i-\mu)^2}}\right\}$$

$$= \min\left\{1, e^{n\left(\xi^{(t-1)}-\xi\right)} \prod_{i=1}^{n} \frac{1 + e^{-2\xi^{(t-1)}}(x_i-\mu)^2}{1 + e^{-2\xi}(x_i-\mu)^2}\right\}$$

코쉬 모형: 임의보행 알고리듬

단계 1. (초기화) $\mu^{(0)} = \bar{x}, \xi^{(0)} = \log s$로 정한다.

단계 2. (MH 단계) $t = 1, 2, \ldots, m$에 대해 다음을 수행한다.

(i) ($\mu^{(t)}$의 추출)

1. 아래의 분포에서 μ, U를 추출한다.

$$\mu \sim N(\mu^{(t-1)}, b^2)$$
$$U \sim U(0, 1)$$

2. $\alpha(\mu^{(t-1)}, \mu)$를 계산한다.

3. $U \le \alpha(\mu^{(t-1)}, \mu)$이면, $\mu^{(t)} = \mu;$

그렇지 않으면, $\mu^{(t)} = \mu^{(t-1)}$로 놓는다.

코쉬 모형: 임의보행 알고리듬

(ii) ($\xi^{(t)}$의 추출)

1. 아래의 분포에서 ξ, U를 추출한다.

$$\xi \sim N(\xi^{(t-1)}, d^2)$$
$$U \sim U(0,1)$$

2. $\alpha(\xi^{(t-1)}, \xi)$를 계산한다.

3. $U \leq \alpha(\xi^{(t-1)}, \xi)$이면, $\xi^{(t)} = \xi$;
그렇지 않으면, $\xi^{(t)} = \xi^{(t-1)}$로 놓는다.

코쉬 모형: 코드

```
# 자료
x = rcauchy(10, location=1, scale = 2)
n=length(x)

# MH 샘플러의 초기화.
m = 5000
mu.jump = 2
xi.jump = 2

po.mu = NULL
po.xi = NULL
mu = median(x)
sig = mad(x)
xi = log(sig)
```

코쉬 모형: 코드

```
# 사후표본 추출
for(j in 1:m) {
   muc = rnorm(1, mu, mu.jump)
   u = runif(1, 0,1)
   log.accept.mu = sum(log((1+exp(-2*xi)*(x-mu)^2)/(1+exp(-2*xi)*(x-muc)^2)))
   if(u < exp(log.accept.mu)) mu = muc

   xic = rnorm(1, xi, xi.jump)
   u = runif(1, 0, 1)
   log.accept.xi = n*(xi-xic)+sum(log((1+exp(-2*xi)*(x-mu)^2)/(1+exp(-2*xic)*(x-mu)^2)))
   if(u < exp(log.accept.xi)) xi = xic

   po.mu = c(po.mu, mu)
   po.xi = c(po.xi, xi)
}
po.sig = exp(po.xi)
post.df = data.frame(mu = po.mu, sig=po.sig, xi=po.xi)
```

코쉬 모형: 코드

```
# 사후표본의 요약
library(dplyr)
library(coda)
library(ggmcmc)
post.df %>% as.mcmc %>% summary
post.df %>% as.mcmc %>% ggs %>% ggs_density
post.df %>% as.mcmc %>% ggs %>% ggs_traceplot
post.df %>% as.mcmc %>% ggs %>% ggs_autocorrelation
```

다음시간

11강

해밀턴몬테카를로