베이즈데이터분석 / 이재용 교수

09강

깁스추출법

목차

목차

합격불합격 방법으로 1차원 정규 확률변수의 생성

- (목적)

$$y \sim c_1 f(x) = c_1 e^{-y^2/2}, c_1 = \frac{1}{\sqrt{2\pi}}$$

- (제안 밀도함수)

$$x \sim c_2 g(x) = c_1 \frac{1}{1+x^2}, c_2 = \frac{1}{\pi}$$

$$\sup_x \frac{f(x)}{g(x)} \leq M, M = 2/\sqrt{e}.$$

- (합격확률)

$$\mathbb{E}\alpha(x) \approx 0.66$$

합격불합격 방법으로 k - 차원 정규 확률변수의 생성

- (목적)

$$y \sim c_1 f(y) = c_1 e^{-\frac{1}{2}\sum_{i=1}^{k} y_i^2}, c_1 = \frac{1}{(2\pi)^{k/2}}$$

- (제안 밀도함수)

$$x \sim c_2 g(x) = c_2 \prod_{i=1}^{k} \frac{1}{1 + x_i^2}, c_2 = \frac{1}{\pi^k}$$

$$\sup_x \frac{f(x)}{g(x)} \le M = \left(\frac{2}{\sqrt{e}}\right)^k$$

합격불합격 방법으로 k - 차원 정규 확률변수의 생성

- (합격확률)

$$\mathbb{P}\left(u \le \frac{1}{M}\prod_{i=1}^{k}\left((1+x_i^2)e^{-x_i^2/2}\right)\right)$$

$$= \mathbb{E}\mathbb{P}\left(u \le \frac{1}{M}\prod_{i=1}^{k}\left((1+x_i^2)e^{-x_i^2/2}\right)\middle| x\right)$$

$$= \mathbb{E}\left(\frac{1}{M}\prod_{i=1}^{k}\left((1+x_i^2)e^{-x_i^2/2}\right)\right)$$

$$= \left(\frac{\sqrt{e}}{2}\right)^k \prod_{i=1}^{k}\int (1+x_i^2)\cdot e^{-x_i^2/2}\cdot\frac{1}{\pi}\cdot\frac{1}{1+x_i^2}\,dxi$$

$$= \left(\frac{\sqrt{e}}{2\pi}\right)^k \prod_{i=1}^{k}\int e^{-x_i^2/2}\ dx_i$$

$$= \left(\sqrt{\frac{e}{2\pi}}\right)^k$$

$k = 10$일 때, 0.01568; $k = 100$일 때, 9.003×10^{-19}.

차원이 커질수록 확률을 생성하기가 현실적으로 어려워진다.

목차

마르코프 체인(Markov chain)

: t시점의 값이 직전 값에만 의존하는 확률변수들의 수열.

$(\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots)$가 확률변수들의 열일 때,

$$[\theta^{(t)} | \theta^{(t-1)}, \theta^{(t-2)}, \dots, \theta^{(0)}] = [\theta^{(t)} | \theta^{(t-1)}], \forall t \geq 0$$

를 만족하면 $(\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots)$를 마르코프 체인이라 한다.

$[\theta^{(0)}]$는 확률변수 $\theta^{(0)}$의 분포를 의미한다.

필요한 용어들

- **마르코프 체인의 커널(Kernel):** 직전 값이 주어졌을 때 현재 값의 분포. $[\theta^{(t)}|\theta^{(t-1)}]$ 혹은 $K(\theta^{(t-1)}, \theta^{(t)})$로 표현한다.

- **마르코프 체인 몬테 카를로(Markov chain Monte Carlo, MCMC)**
  : 마르코프 체인을 이용한 적분의 근사법, 즉 몬테 카를로 방법.

- **마르코프 체인의 정상분포(stationary distribution)**
  : 마르코프 체인이 오랜 시간 생성될 때 수렴하는 분포. 즉 t가 매우 클 때, $\theta^{(t)}$의 분포.

깁스추출법

- 깁스추출법(Gibbs sampling)은 다차원 확률변수를 추출하는 방법이다.
- $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}$이 사후분포 $\pi(\theta|x)$를 추론하기 위해 깁스추출법으로 추출한 사후표본일 때 다음이 성립한다.
  - 사후표본 $(\theta^{(t)})$은 서로 독립인 랜덤표본이 아니라 마르코프 체인이 된다.
  - 마르코프 체인 $(\theta^{(t)})$의 정상분포(혹은 불변분포)는 사후분포 $\pi(\theta|x)$가 된다.
  - 사후표본의 표본평균은 사후분포의 기대값을 근사하고, 사후표본의 표본분위수는 사후분포의 분위수를 근사한다.

깁스 알고리듬

$(\theta_1, \theta_2, \theta_3)$의 밀도함수가 $\pi(\theta_1, \theta_2, \theta_3)$이고, θ_{-i}는 θ에서 i번째 원소를 제외한 나머지라고 하자.

단계 1. 초기값 $(\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)})$ 을 선정한다.

단계 2. $t = 1, 2, 3, \ldots,$에 대하여, $\theta^{(t)}$의 각 원소를 조건부 분포로 부터 추출한다. 즉,

- $\theta_1^{(t)} \sim \pi_{\theta_1|\theta_{-1}}(\cdot \mid \theta_2^{(t-1)}, \theta_3^{(t-1)})$을 추출한다.
- $\theta_2^{(t)} \sim \pi_{\theta_2|\theta_{-2}}(\cdot \mid \theta_1^{(t)}, \theta_3^{(t-1)})$을 추출한다.
- $\theta_3^{(t)} \sim \pi_{\theta_3|\theta_{-3}}(\cdot \mid \theta_1^{(t)}, \theta_2^{(t)})$을 추출한다.

단계 3. $\theta^{(1)}, \theta^{(2)}, \ldots$는 $\pi(\theta_1, \theta_2, \theta_3)$를 정상분포로 갖는 마르코프 체인이 된다.

깁스추출법의 예: 이변량 정규분포

목표

$$\pi = N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2, & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2, & \sigma_2^2,\end{pmatrix}\right)$$

를 정상분포로 갖는 깁스추출표본을 구하자.

깁스추출법의 예: 이변량 정규분포

마르코프체인의 사용

마르코프체인의 사용

$(\theta_1^{(t)}, \theta_2^{(t)})\ , t \geq 0$ 를 이용해 정규분포의 적률과

분위수를 구한다.

$$\widehat{\int h(\theta_1,\theta_2)\pi(\theta_1,\theta_2)d\theta_1 d\theta_2} = \frac{1}{n}\sum_{t=1}^{n} h(\theta_1^{(t)}, \theta_2^{(t)})$$

$$\text{i.e.},\ \widehat{P(\theta_1 \geq 0, \theta_2 \geq 0)} = \frac{1}{n}\sum_{t=1}^{n} I(\theta_1^{(t)} \geq 0, \theta_2^{(t)} \geq 0)\,.$$

깁스추출법의 예: 이변량 정규분포

알고리듬

단계 1. 초기값 $(\theta_1^{(0)}, \theta_2^{(0)})$을 선정한다.

단계 2. $t = 1, 2, 3, \ldots,$에 대하여, $\theta^{(t)}$의 각 원소를 조건부 분포로 부터 추출한다. 즉,

(i) $\theta_1^{(t)} \sim$

$$N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(\theta_2^{(t-1)} - \mu_2), \sigma_1^2(1 - \rho^2))$$

를 발생한다.

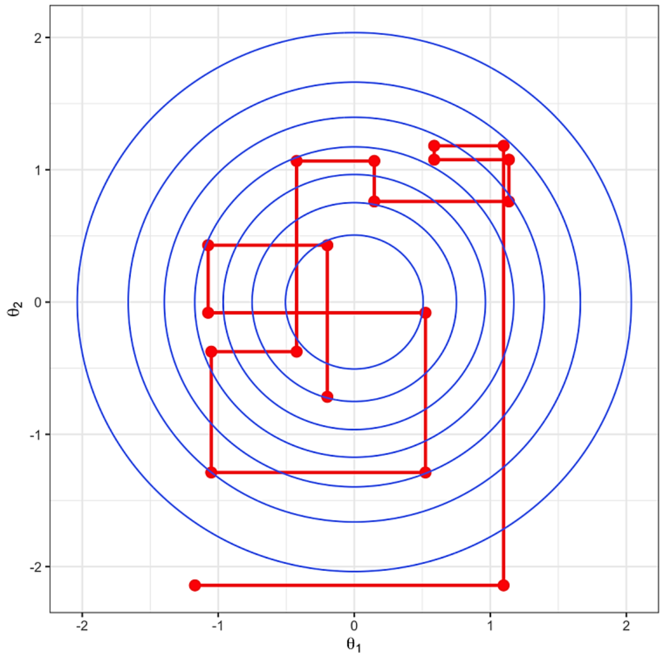
(ii) $\theta_2^{(t)} \sim$

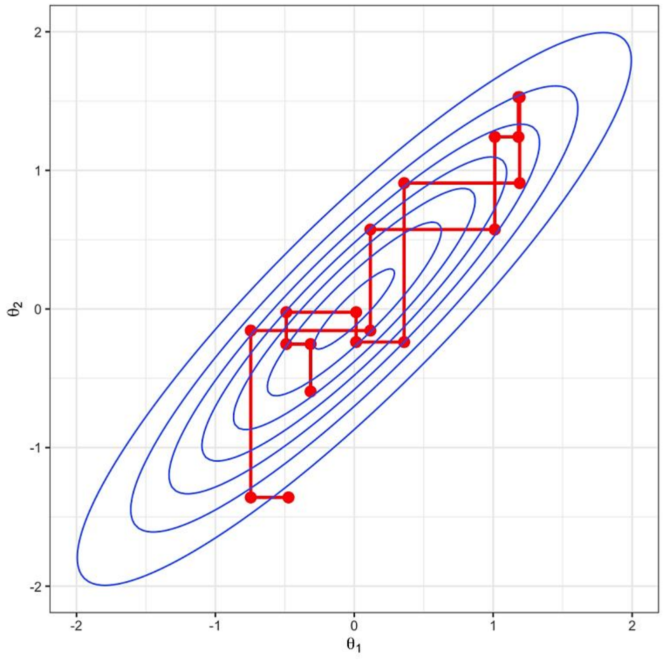
$$N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(\theta_1^{(t)} - \mu_1), \sigma_2^2(1 - \rho^2))$$

를 발생한다.

단계 3. $\theta^{(1)}, \theta^{2}, \ldots.$는 이변량 정규분포를 정상분포로 갖는 마르코프 체인이 된다.

이변량 정규분포 깁스표본의 궤적





ρ = 0일 때와 ρ = 0.99일 때 초기 깁스 표본의 궤적의 그림이다.

목차

MCMC 수렴진단

마르코프 체인 수렴의 문제

- 마르코프 체인은 서로 독립인 표본과 달리 정상분포로 수렴을 하는데 시간이 걸릴 수 있다.
- 마르코프 체인이 수렴했는지 판단하는 것을 수렴 진단이라 한다.

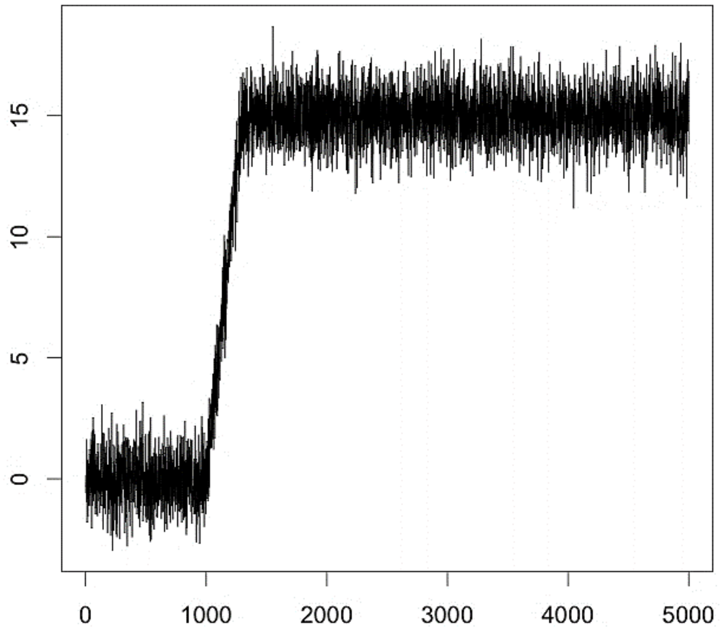
실제적인 방법들

- **번인**(Burn-in): 엠씨엠씨 표본의 앞부분을 버리는 것을 말한다.
- **가늘게하기**(thinning): 엠씨엠씨 표본의 자기상관계수가 높을 때, 모든 표본을 다 사용하지 않고 r번째 표본만 따로 추출해서 사용하는 것을 말한다. 가늘게하기로 컴퓨터의 메모리를 절약할 수 있다.

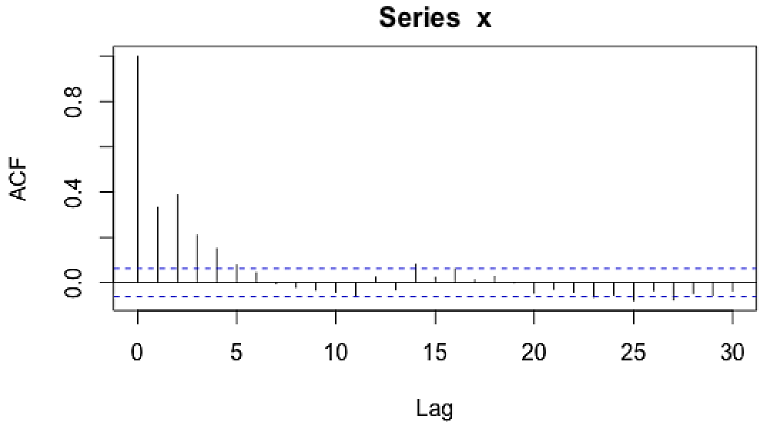
MCMC 수렴진단

그림을 이용한 진단

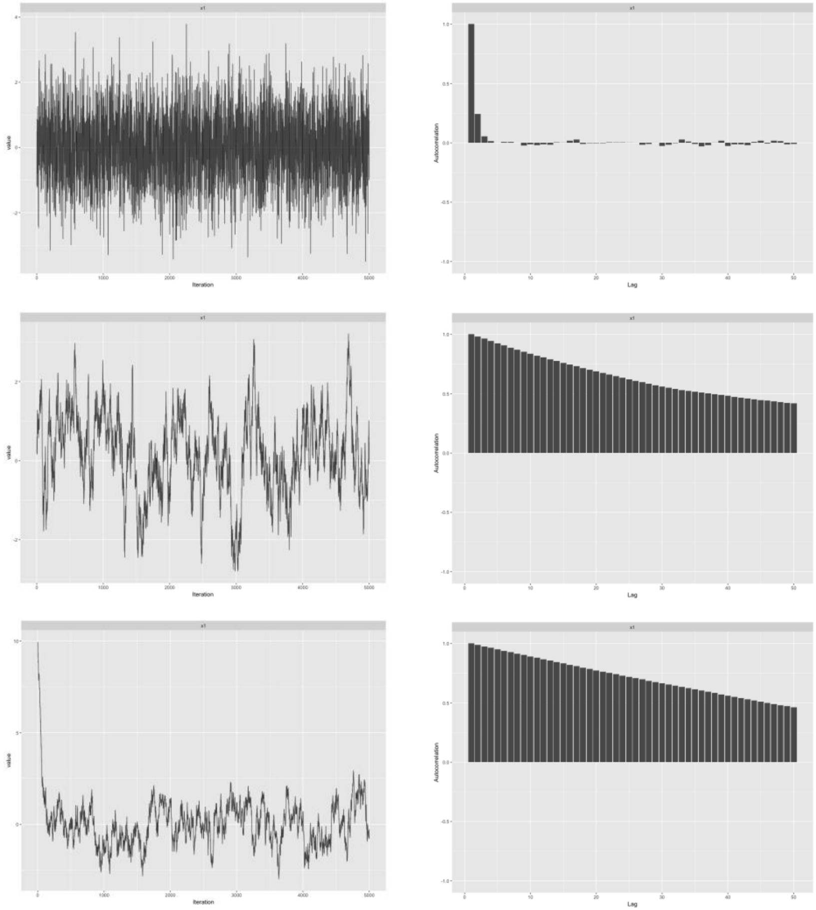
- 시계열 그림



- 자기상관계수 그림



시계열그림과 자기상관계수 그림을 이용한 수렴진단



맨 위 행의 그림은 마르코프체인에 문제가 없어 보인다. 두 번째 행은 자기 상관계수가 너무 크다. 이 때는 가늘게하기를 할 필요가 있다. 세 번째 행은 마르코프체인이 초기값에 의존한다. 번인을 할 필요가 있다.

목차

목표 사후 밀도 함수의 형태

$$\pi(\theta) \propto \prod_{i=1}^{k} f_i(\theta), f_i(\theta) \geq 0 \ (\forall i)$$

잠재변수의 추가

$f_i(\theta)$를 적분의 형태로 써보면

$$\pi(\theta) \propto \prod_{i=1}^{k} f_i\,(\theta), = \prod_{i=1}^{k} \int I\big(0 \ \leq \ w_i \ \leq f_i(\theta)\big) dw_i.$$

$w_1, \dots, w_k$를 잠재 변수로 사후 분포에 추가하면

$$\pi(\theta, w_1, \dots, w_k) \propto \prod_{i=1}^{k} I\big(0 \ \leq \ w_i \ \leq f_i(\theta)\big).$$

$$\pi(\theta) \propto \int \dots \int \pi(\theta, w_1, \dots w_k) dw_1 \dots dw_k$$

이므로, $\pi(\theta)$는 $(\theta, w_1, \dots, w_k) \sim \pi(\theta, w_1, \dots, w_k)$일 때 θ의 주변 분포이다.

깁스 알고리듬

- $w_i, i = 1, 2, \ldots, k$의 완전 조건부 밀도 함수

$$\pi(w_i | w_{-i}, \theta) \propto I\,(0 \leq w_1 \leq f_i(\theta)) \propto U(0, f_i(\theta))$$

- θ의 완전 조건부 밀도함수

$$\pi(\theta | w) \propto \prod_{i=1}^{k} I(f_i(\theta) \geq w_i) \propto U(A)$$

$$A = \{\theta : f_i(\theta) \geq w_i, i = 1, \ldots, k\}$$

예. 절단된 정규분포

$$\pi(x) \propto e^{-\frac{1}{2\sigma^2}(x-\mu)^2} I(x \geq A), A \in \mathbb{R}$$

라 하자. π를 정상분포로 갖는 마르코프 체인 $x^{(t)}$을 생성하고자 한다.

아이디어

$$\pi(x) \propto \int I\left(0 \leq z \leq e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\right) I(x \geq A)dz$$

이라는 것에 착안해서,

$$\pi(x, z) \propto I\left(0 \leq z \leq e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\right) I(x \geq A)$$

이라 놓는다. $\pi(x, z)$를 정상분포로 갖는 마르코프 체인 $\left(x^{(t)}, z^{(t)}\right)$를 생성하면, $x^{(t)}$는 $\pi(x)$를 정상분포로 갖는 마르코프체인이 된다.

그룹화 깁스추출법과 붕괴 깁스추출법

알고리듬 1. 원형의 깁스추출법

(i) $\theta_1 \sim \pi(\theta_1|\theta_2, \theta_3)$

(ii) $\theta_2 \sim \pi(\theta_2|\theta_1, \theta_3)$

(iii) $\theta_3 \sim \pi(\theta_3|\theta_1, \theta_2)$

알고리듬 2. 그룹 깁스 추출법

(i) $(\theta_1, \theta_2) \sim \pi(\theta_1, \theta_2|\theta_3)$

(ii) $\theta_3 \sim \pi(\theta_3|\theta_1, \theta_2)$

몇 개의 변수를 묶어서 한 개의 변수로 본 알고리듬이다.

알고리듬 3. 붕괴 깁스추출법

(i) $(\theta_1, \theta_2) \sim \pi(\theta_1, \theta_2)$

(ii) $\theta_3 \sim \pi(\theta_3|\theta_1, \theta_2)$

그룹 깁스추출법에서 조건부 분포 대신 주변 분포를 사용했다. $\theta(t)$들은 서로 독립이 된다.

알고리듬 4. 수정된 붕괴 깁스추출법

(i) $\theta_1 \sim \pi(\theta_1|\theta_2)$

(ii) $\theta_2 \sim \pi(\theta_2|\theta_1)$

(iii) $\theta_3 \sim \pi(\theta_3|\theta_1, \theta_2)$

그룹 깁스추출법에서 (i) 단계에 깁스추출법을 적용한 것이다.

속도는 붕괴 깁스추출법 > 그룹 깁스추출법 > 본래 깁스추출법의 순으로 빠르다.
수정된 붕괴 깁스 추출법은 붕괴 깁스추출법 보다는 느릴 것이다.

다음시간

10강

메트로폴리스