긴 스 샘 플 링

ToBig's 9기 이잉걸

Gibbs Sampling

다양한 알고리즘과 방법론을 중심으로

Contents

'몬테 카를로 방법'

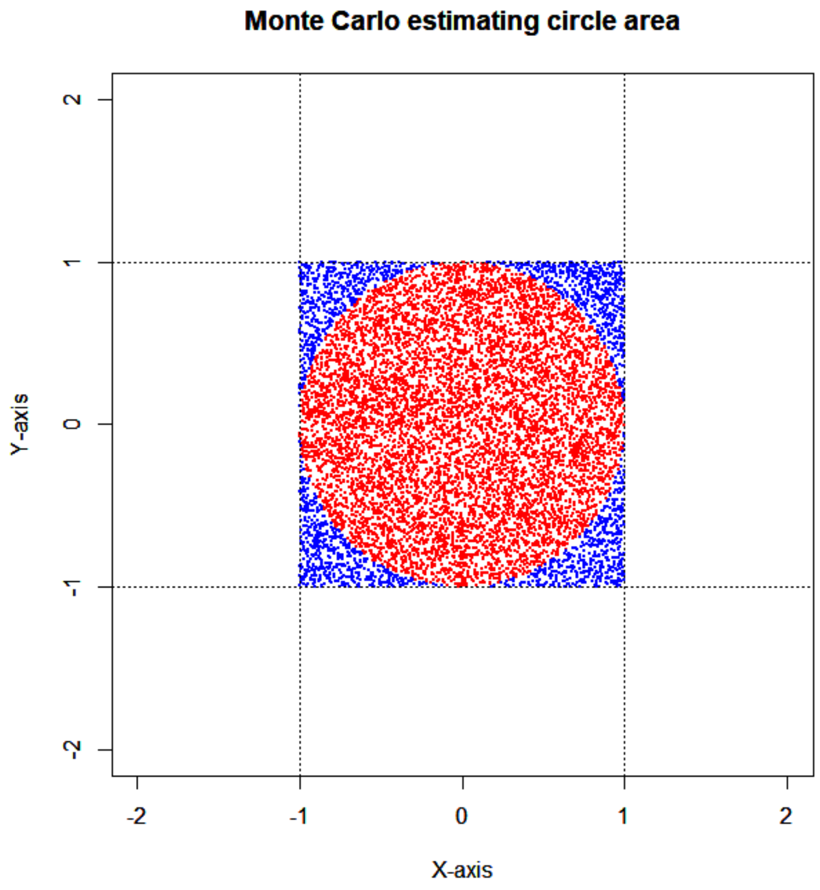
무작위 추출된 난수를 이용해 <u>원하는 함수의 값</u>을 계산하기 위한 시뮬레이션 방법이다.

자유도가 높거나 닫힌 꼴(closed form)의 해가 없는 문제들에 널리 쓰이는 방법이지만,

어느 정도의 오차를 감안해야만 하는 특징이 있다.

무작위로 뽑힌 난수의 개수가 늘어날수록 더 정확한 결과를 얻을 수 있으나, 그만큼 더 많은 시간이 걸리는 것을 감안해야 한다.

또한 시뮬레이션 기반 방법이기 때문에, 해석적인 방법과 달리 항상 어느 정도 오차가 있을 수 있음을 감안해야만 한다.



예시) > 내가 돌을 던진 모든 횟수는 정사각형의 면적에 대응하고,
저 원의 내부에 떨어진 돌의 횟수는 원의 면적에 대응한다.

원 내부에 위치한 (x,y)의 개수 : 모든 점의 개수 = 원의 면적 : 정사각형 면적

$$\text{원의 면적} = \text{정사각형 면적} \times \frac{\text{원 내부에 위치한 (x,y)의 개수}}{\text{모든 점의 개수}}$$

'마르코프 연쇄'

확률론에서, **마르코프 연쇄**(Марков連鎖, 영어: Markov chain)는 Markov Property를 갖는 이산 시간 확률 과정이다.

마르코프 연쇄는 시간에 따른 계의 상태의 변화를 나타낸다.

매 시간마다 계는 상태를 바꾸거나 같은 상태를 유지한다. 상태의 변화를 '전이' 라 한다.

마르코프 성질은 과거와 현재 상태가 주어졌을 때의 미래 상태의 조건부 확률 분포가 과거 상태와는 독립적으로 현재 상태에 의해서만 결정된다는 것을 뜻한다.

– 체인? :=〉 상태값의 시퀀스! Ex) 상태 : (비 or 맑음) –〉 P(Day10=비|Day9=비,…,Day1=맑음)

– Markov Property(memory = 1)

$$Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, \cdots, X_1 = x_1) = Pr(X_n = x_n \mid X_{n-1} = x_{n-1})$$

=> 현재의 상태는 바로 이전의 상태에만 영향을 받고, 그 이전의 상태와는 독립이다.

– Markov Property(memory = m)

$$Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, \cdots, X_1 = x_1) = Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots, X_{n-m} = x_{n-m})$$

- **체인? :=〉 상태값의 시퀀스!** **Ex) 상태 : (비 or 맑음) –〉 P(Day10=비|Day9=비,…,Day1=맑음)**

- **Markov Property(memory = 1)**

$$Pr(X_n = x_n \mid \underline{X_{n-1} = x_{n-1}}, \boxed{\cdots, X_1 = x_1}) = Pr(X_n = x_n \mid \underline{X_{n-1} = x_{n-1}})$$

=> 현재의 상태는 바로 이전의 상태에만 영향을 받고, 그 이전의 상태와는 독립이다.

- **Markov Property(memory = m)**

$$Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, \cdots, X_1 = x_1) = Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots, X_{n-m} = x_{n-m})$$

Unit 01 | Monte Carlo & Markov Chain

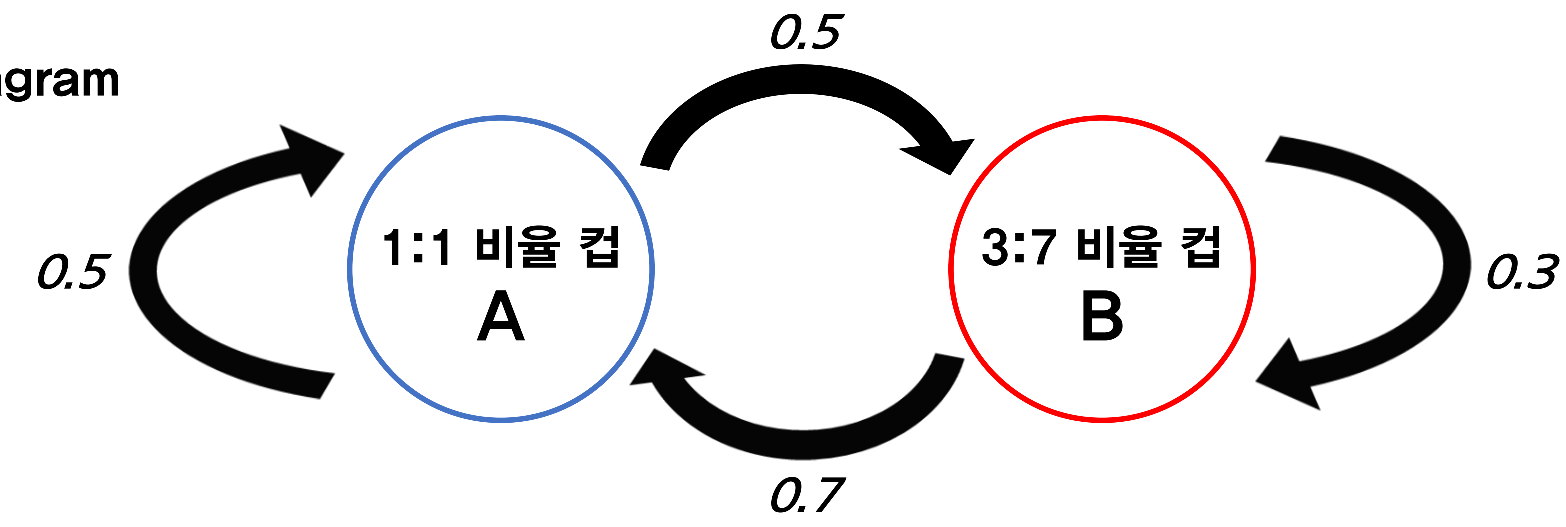
1 : 1 3 : 7

– 예시

(Cup) A B

① 왼쪽 컵에서 흰 공이 나오면 다시 공을 넣고 왼쪽 컵에서 뽑는다.
② 왼쪽 컵에서 검은 공이 나오면 다시 공을 넣고 왼쪽 컵에서 뽑는다.
③ 오른쪽 컵에서 검은 공이 나오면 다시 공을 넣고 오른쪽 컵에서 뽑는다.
④ 오른 쪽컵에서 흰 공이 나오면 다시 공을 넣고 왼쪽 컵에서 뽑는다.

– State Diagram



– 전이 확률

$$A_{ij} = \begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix}$$

a_{ij} : i상태에서 j상태로 전이할 확률

과정이 두 번 반복된다면?

$$\begin{bmatrix} \boxed{0.5 \quad 0.5} \\ 0.7 \quad 0.3 \end{bmatrix} \begin{bmatrix} \boxed{0.5} & 0.5 \\ \boxed{0.7} & 0.3 \end{bmatrix} = \begin{bmatrix} \boxed{0.60} & 0.40 \\ 0.56 & 0.44 \end{bmatrix}$$

a_{ij} : i상태에서 j상태로 두 번의 공 뽑기 후 전이할 확률

과정이 무한히 반복된다면?

$$\begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix} \cdots \begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix} = \begin{array}{cc} W & B \\ \multicolumn{2}{c}{\begin{bmatrix} \frac{7}{12} & \frac{5}{12} \\ \frac{7}{12} & \frac{5}{12} \end{bmatrix}} \end{array}$$

'정상 분포(Stationary distribution)'

‘Markov Chain Monte Carlo Method’ (MCMC)

우리가 샘플을 얻으려고 하는 목표 분포를 정상 분포(Stationary distribution)로 가지는 마코프 체인을 만든다.

체인의 시뮬레이션을 시작하고, 초기값에 영향을 받는 burn-in period를 지나고 나면 목표 분포를 따르는 샘플을 얻을 수 있다.

‘Markov Chain Monte Carlo Method’ (MCMC)

우리가 샘플을 얻으려고 하는 목표 분포를 정상 분포(Stationary distribution)로 가지는 마코프 체인을 만든다.

체인의 시뮬레이션을 시작하고, 초기값에 영향을 받는 burn-in period를 지나고 나면 목표 분포를 따르는 샘플을 얻을 수 있다.

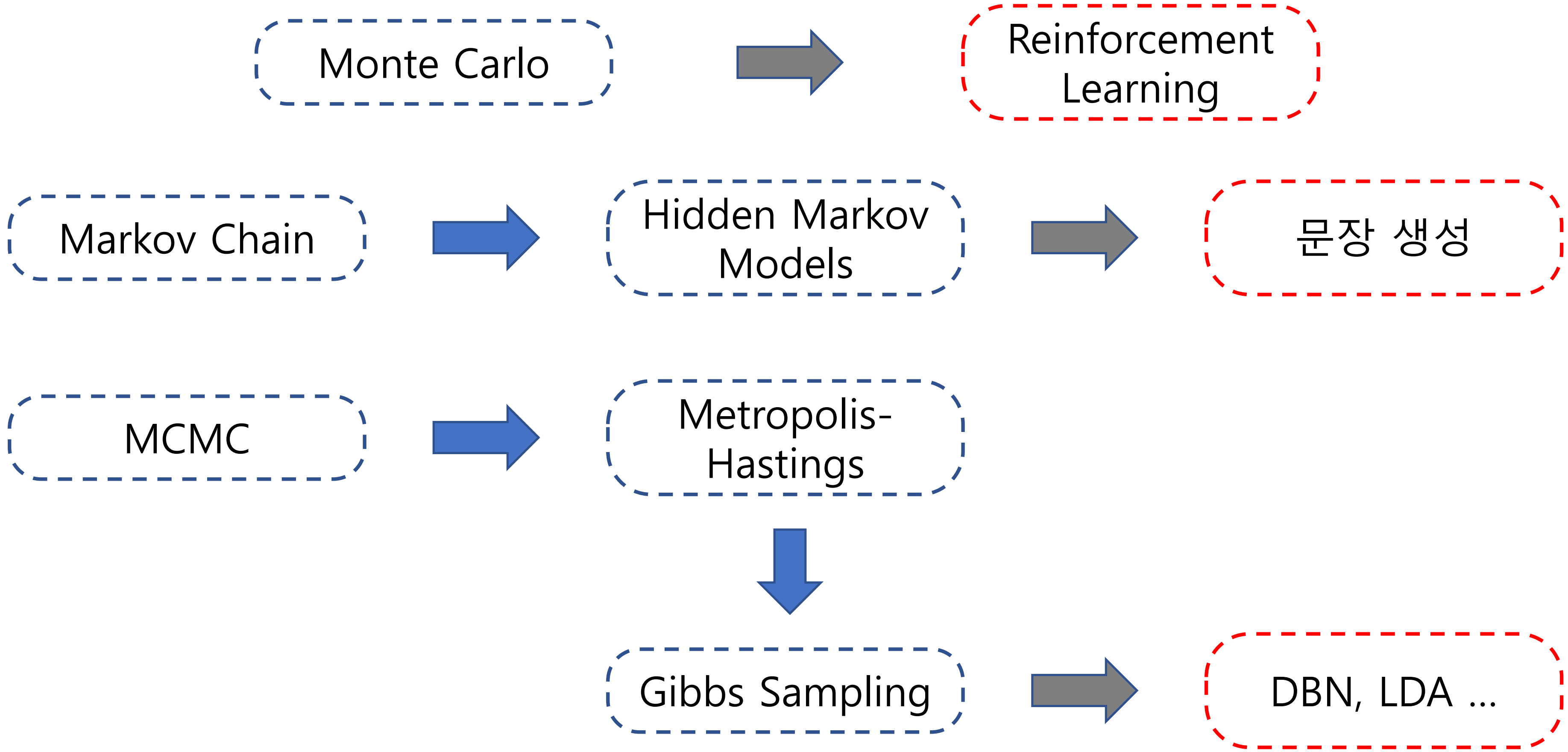
① 마코프 체인을 위한 초기값 X_0 를 지정한다.

② 마코프 체인이 정상 분포에 도달할 때까지 n개의 관측값(마코프 수열), $\{X_k, k = 1, \ldots, n\}$을 생성한다.

③ 적당한 수렴진단 방법을 이용하여 알고리즘의 수렴 여부를 확인한다. 만약 수렴하지 않으면, 더 많은 관측값을 생성한다.

➢ 마코프 체인이 정상 분포에 수렴하는지 여부를 확인하고, 초기 B개의 관측값을 제거한 후(burn-in), $\{X_{B+1}, \ldots, X_n\}$ 을 분석을 위한 표본으로 채택한다.

Unit 01 | Monte Carlo & Markov Chain



M.H. Algorithm

MH 알고리즘의 기본 아이디어는 마코프 체인의 정상 분포가 f(x)가 되도록 마코프 체인 $\{X_k, k = 1, \ldots, n\}$을 생성,

적절한 전이확률을 통해 다음 상태 X_{t+1}를 생성하는 것으로, 이를 위하여 거절법(rejection sampling)을 사용한다.

1) 현재 상태(current state) X_t에 기반한 적절한 제안 분포 $g(\cdot \mid X_t)$로부터 후보(candidate) 난수를 생성한다!

2) 후보난수 Y가 채택된다면, 마코프 체인의 다음 상태 X_{t+1}은 Y가 되고, $X_{t+1} = Y$
기각된다면, 현재 상태 X_t에 머물게 된다($X_{t+1} = X_t$).

– Process

1 우리가 잘 알고 있는, 표본을 쉽게 추출할 수 있는 분포 $g(\cdot \mid X_t)$ (proposal distribution)를 하나 정한다.
(체인의 시작점 X_0 을 정하거나, 분포 $g(\cdot \mid X_t)$ 로부터 추출한다.)

2 체인이 정상 분포로 수렴할 때까지 다음의 과정을 반복한다.

어렵지 않아요!

– Process

1 우리가 잘 알고 있는, 표본을 쉽게 추출할 수 있는 분포 $g(\,\cdot\,|\,X_t)$ (proposal distribution)를 하나 정한다.
(체인의 시작점 X_0 을 정하거나, 분포 $g(\,\cdot\,|\,X_t)$ 로부터 추출한다.)

2 체인이 정상 분포로 수렴할 때까지 다음의 과정을 반복한다.

현재 상태 : $X_t = x_t$

제안 분포 : $g(\,\cdot\,|\,X_t = x_t)$

값 하나 추출 : $Y = y$

– Process

1 우리가 잘 알고 있는, 표본을 쉽게 추출할 수 있는 분포 $g(\,\cdot \mid X_t)$ (proposal distribution)를 하나 정한다.
(체인의 시작점 X_0을 정하거나, 분포 $g(\,\cdot \mid X_t)$ 로부터 추출한다.)

2 체인이 정상 분포로 수렴할 때까지 다음의 과정을 반복한다.

현재 상태 : $X_t = x_t$

제안 분포 : $g(\,\cdot \mid X_t = x_t)$

값 하나 추출 : $Y = y$

→

값 계산->

$$\frac{f(y)g(X_t = x_t \mid y)}{f(x_t)g(y \mid X_t = x_t)}$$

값 하나 추출 :

$U \sim Unif(0,1)$

→

– Process

1 우리가 잘 알고 있는, 표본을 쉽게 추출할 수 있는 분포 $g(\cdot \mid X_t)$ (proposal distribution)를 하나 정한다.
(체인의 시작점 X_0 을 정하거나, 분포 $g(\cdot \mid X_t)$ 로부터 추출한다.)

2 체인이 정상 분포로 수렴할 때까지 다음의 과정을 반복한다.

현재 상태 : $X_t = x_t$

제안 분포 : $g(\cdot \mid X_t = x_t)$

값 하나 추출 : $Y = y$

값 계산->

$$\frac{f(y)g(X_t = x_t \mid y)}{f(x_t)g(y \mid X_t = x_t)}$$

값 하나 추출 :

$U \sim Unif(0,1)$

$$X_{t+1} = \begin{cases} y, & if\ U \le \alpha(x_t, y) \\ x_t, & otherwise \end{cases}$$

$$\alpha(x_t, y) := min\left(1, \frac{f(y)g(X_t = x_t \mid y)}{f(x_t)g(y \mid X_t = x_t)}\right)$$

– Example

▶ **목표 분포(정상 분포) :** $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

– Example

▶ **목표 분포(정상 분포) :** $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

▶ **제안 분포 :** $g(y|X_t) = \begin{cases} \frac{1}{2}, & X_t - 1 \le y \le X_t + 1 \\ 0, & \text{그 외.} \end{cases}$

– Example

▶ **목표 분포(정상 분포) :** $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

▶ **제안 분포 :** $g(y|X_t) = \begin{cases} \frac{1}{2}, & X_t - 1 \le y \le X_t + 1 \\ 0, & \text{그 외.} \end{cases}$

▶ Acceptance probability : $\alpha(x_t, y) = \left(\frac{e^{-y^2/2}}{\sqrt{2\pi}}\right) \cdot \frac{1}{2} \Big/ \left(\frac{e^{-x_t^2/2}}{\sqrt{2\pi}}\right) \cdot \frac{1}{2} = \boxed{\exp\left(-\frac{y^2 - x_t^2}{2}\right)}$

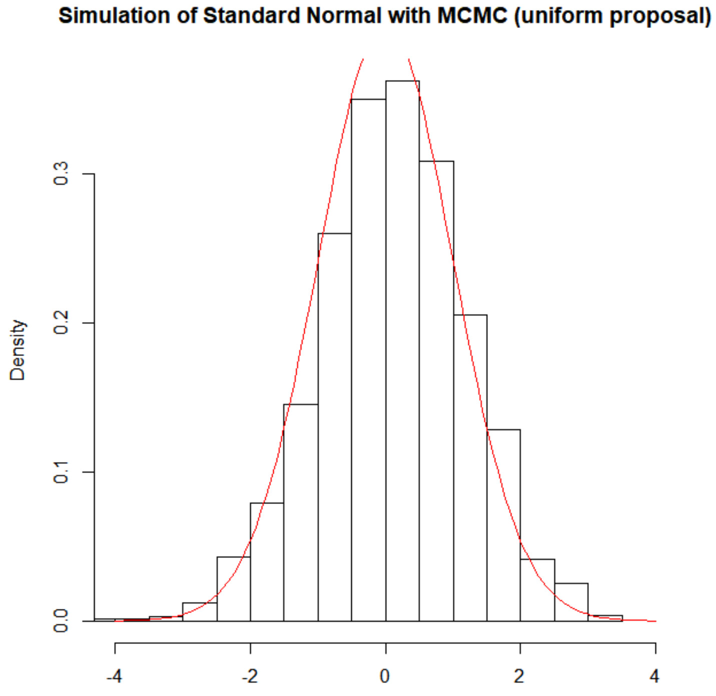
▶ $\boxed{X_{t+1} = \begin{cases} y, & if\ U \le \alpha(x_t, y) \\ x_t, & otherwise \end{cases}}$

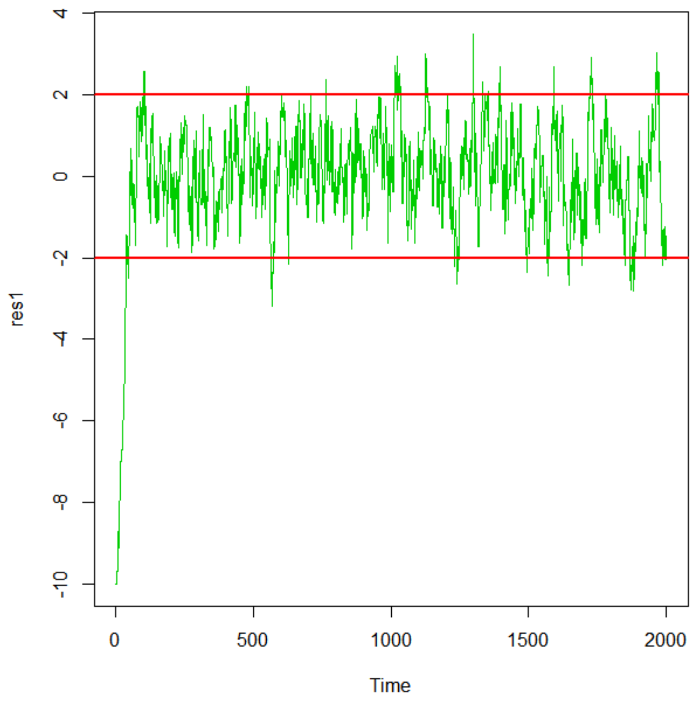
– Example

▶ $X_0 = -10$

▶ 목표 분포(정상 분포) : $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

▶ 제안 분포 : $g(y|X_t) = \begin{cases} \frac{1}{2}, & X_t - 1 \le y \le X_t + 1 \\ 0, & \text{그 외.} \end{cases}$





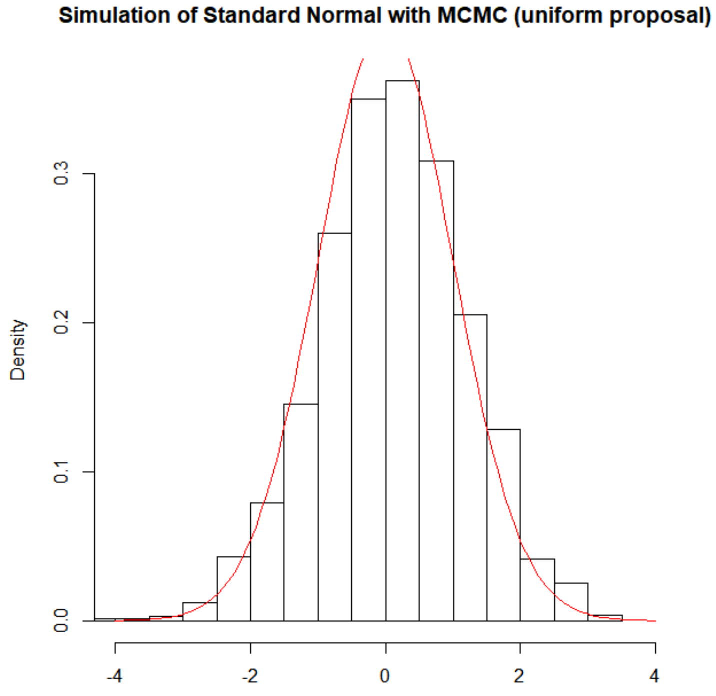
Burn-in period?

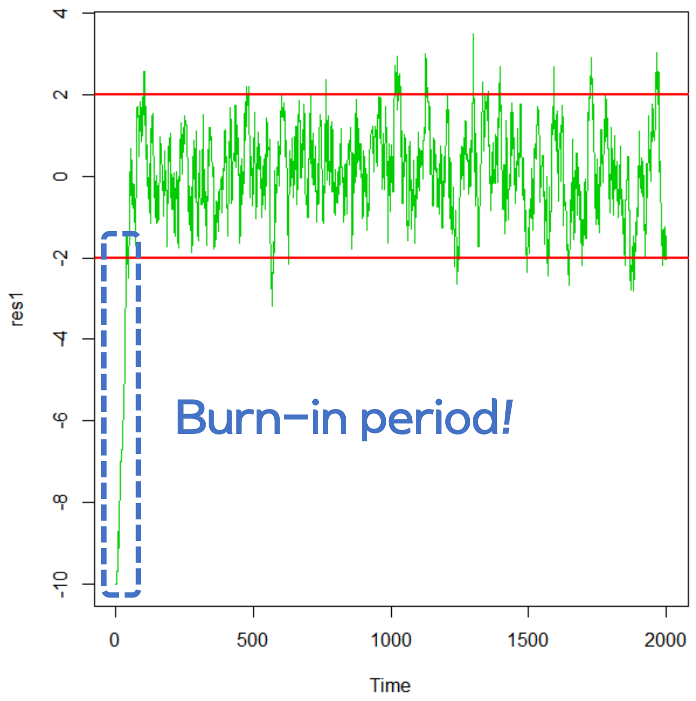
– Example

▶ $X_0 = -10$

▶ **목표 분포(정상 분포) :** $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

▶ **제안 분포 :** $g(y|X_t) = \begin{cases} \frac{1}{2}, & X_t - 1 \leq y \leq X_t + 1 \\ 0, & \text{그 외.} \end{cases}$





– Traits

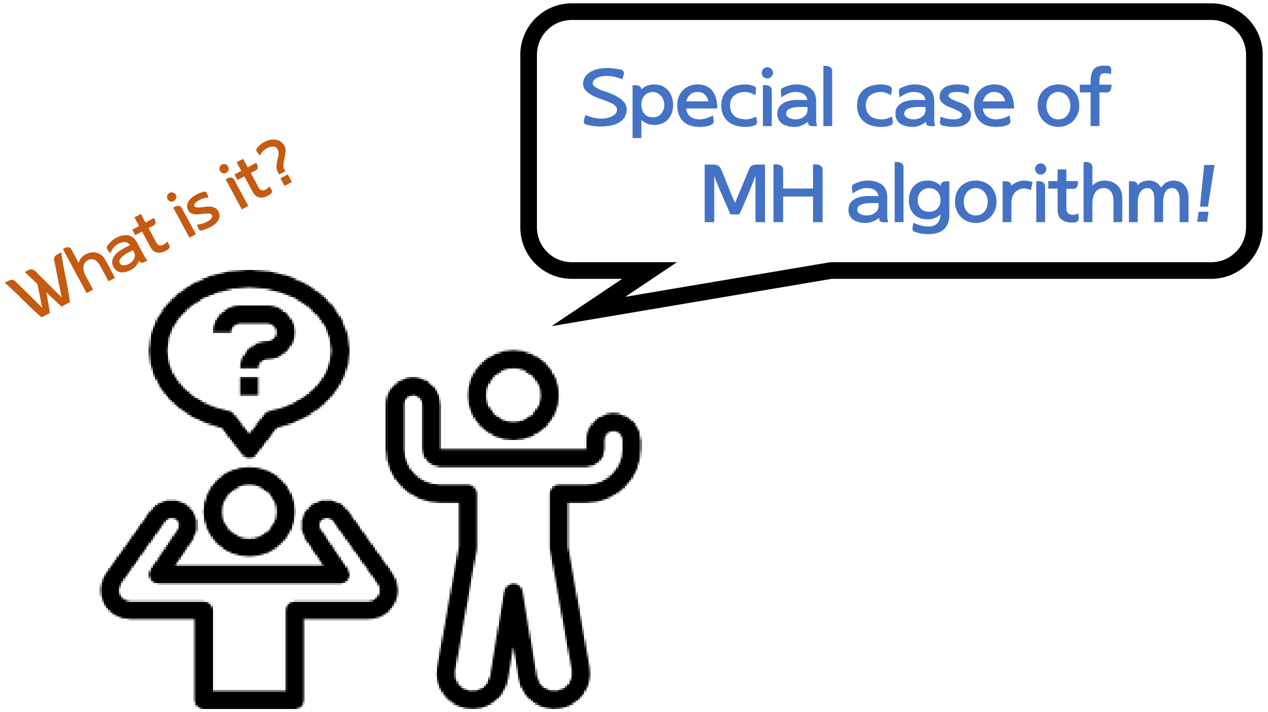
– 제안 분포와 목표 분포의 관계에 따라 목표 분포로의 수렴 속도가 달라질 수 있다.

– 생성된 표본들은 목표 분포를 따르지만 서로 독립이 아니고 상관관계가 존재한다.
–> 상관관계를 줄이기 위해 예를 들어 매 세 번째마다 난수를 표본으로 추출하는 Thinning 작업을 하기도 함(표집시차 = 3)

– 마코프 체인이 목표 분포로 도달해 표본을 얻기 위해 얼마나 많이 반복해야 하는지에 대해 정확히 알 수는 없지만,
흔히 trace plot이나 자기상관계수 확인 등의 시각적인 방법을 사용하며, 수렴 여부를 확인하기 위한 통계량 또한 이용 가능

– 일반적으로 full conditional probability를 구할 수 없을 때 사용한다
(구할 수 있을 때는 Gibbs sampling이 효율적!)

Gibbs Sampling



Gibbs Sampling

–> 목표 분포가 다변량 분포일 때!

–> 모든 단변량 조건부 p.d.f가 알려진 형태, 난수 추출이 쉬울 때!

–> k 차원 다변량 확률변수 $\mathbf{X} = (X_1, \ldots, X_k)$와

k–1차원 확률변수 $\mathbf{X}_{(-\mathrm{j})} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k)$,

j=1,⋯,k에 대해 완전 조건부분포인 $\mathbf{X}_{(-\mathrm{j})}$ 가 주어졌을 때

X_j 의 단변량 조건부 p.d.f $f(X_j \mid \mathbf{X}_{(-\mathrm{j})})$를 고려

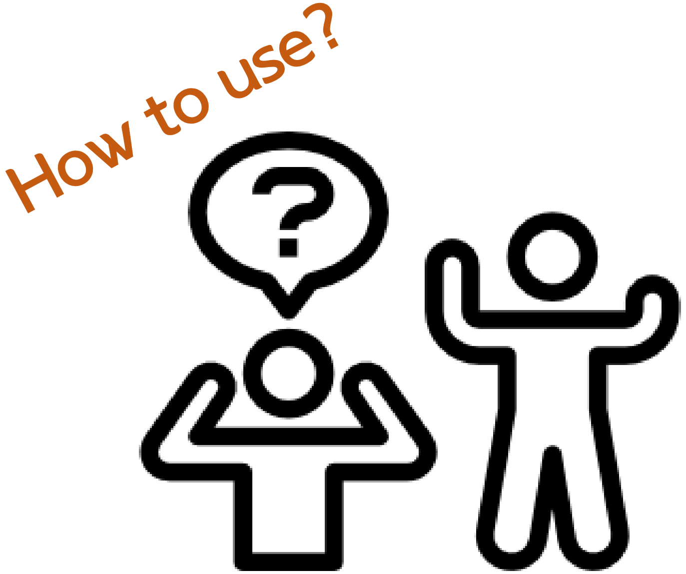
Gibbs Sampling

Gibbs Sampling

$\mathbf{X}^{(0)} = (X_1^{(0)}, \ldots, X_k^{(0)})$

초기값 설정, t = 0,1,2,⋯에 대해 다음을 반복

Gibbs Sampling



$$\mathbf{X}^{(0)} = (X_1^{(0)}, \ldots, X_k^{(0)})$$

초기값 설정, t = 0,1,2,⋯에 대해 다음을 반복

$$X_1^{(t+1)} \mid \mathbf{x}_{(-1)} \sim f(x_1 \mid x_2^{(t)}, \ldots, x_k^{(t)}),$$

Gibbs Sampling

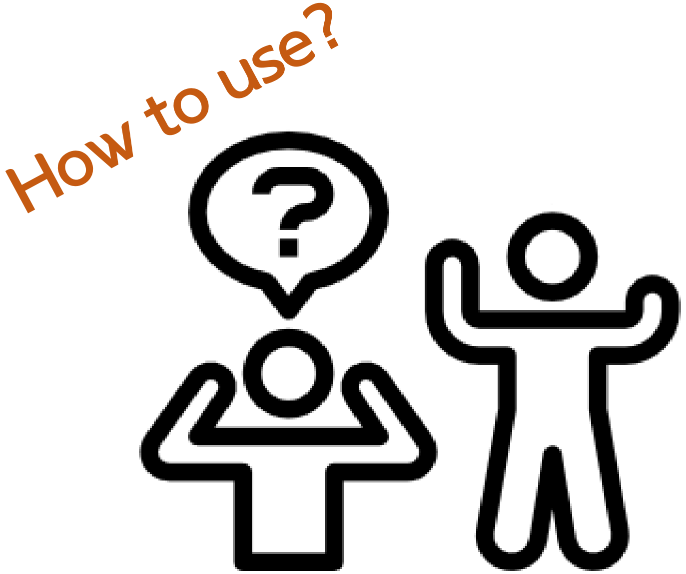
$$\mathbf{X}^{(0)} = (X_1^{(0)}, \ldots, X_k^{(0)})$$

초기값 설정, t = 0,1,2,…에 대해 다음을 반복

$$X_1^{(t+1)} \mid \mathbf{x}_{(-1)} \sim f(x_1 \mid x_2^{(t)}, \ldots, x_k^{(t)}),$$

$$X_2^{(t+1)} \mid \mathbf{x}_{(-2)} \sim f(x_2 \mid x_1^{(t+1)}, x_3^{(t)}, \ldots, x_k^{(t)}),$$

Gibbs Sampling



$$\mathbf{X}^{(0)} = (X_1^{(0)}, \ldots, X_k^{(0)})$$

초기값 설정, t = 0,1,2,⋯에 대해 다음을 반복

$$X_1^{(t+1)} \mid \mathbf{x}_{(-1)} \sim f(x_1 \mid x_2^{(t)}, \ldots, x_k^{(t)}),$$

$$X_2^{(t+1)} \mid \mathbf{x}_{(-2)} \sim f(x_2 \mid x_1^{(t+1)}, x_3^{(t)}, \ldots, x_k^{(t)}),$$

$$\vdots$$

$$X_{k-1}^{(t+1)} \mid \mathbf{x}_{(-\mathbf{k}+\mathbf{1})} \sim f(x_{k-1} \mid x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{k-2}^{(t+1)}, x_k^{(t)}),$$

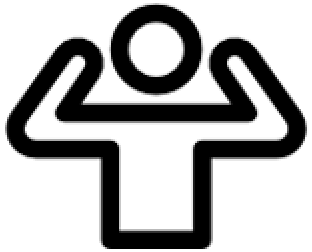
$$X_k^{(t+1)} \mid \mathbf{x}_{(-\mathbf{k})} \sim f(x_k \mid x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{k-1}^{(t+1)})$$

– Example

＊ 이변량 정규분포

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$

분포의 p.d.f도 모르고, 난수 추출 방법도 몰라요!



Unit 03 | MCMC : Gibbs Sampling

– Example

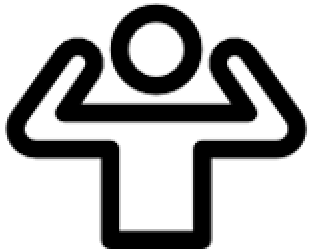
＊ 이변량 정규분포

<Full conditional prob.>

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$

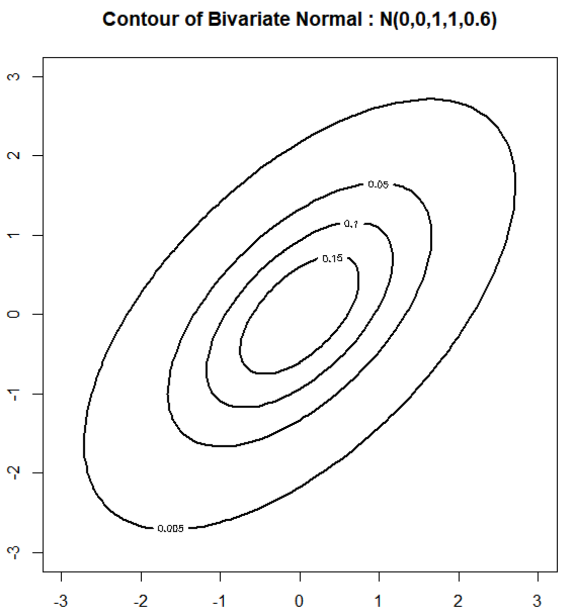
$$\begin{aligned} X_1 | x_2 &\sim N(\theta_1 + \rho(x_2 - \theta_2), 1 - \rho^2), \\ X_2 | x_1 &\sim N(\theta_2 + \rho(x_1 - \theta_1), 1 - \rho^2). \end{aligned}$$

분포의 p.d.f도 모르고, 난수 추출 방법도 몰라요!



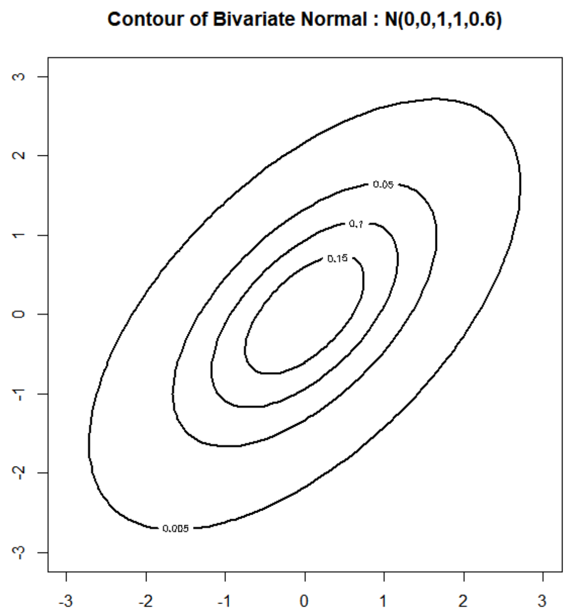
– Example

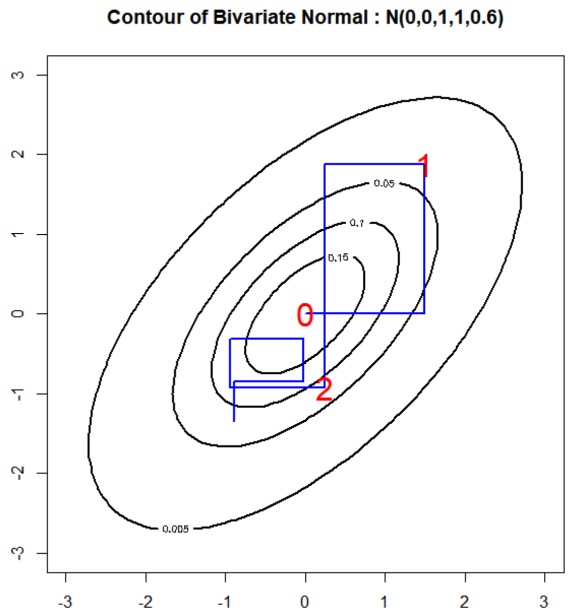
*** 이변량 정규분포**



– Example

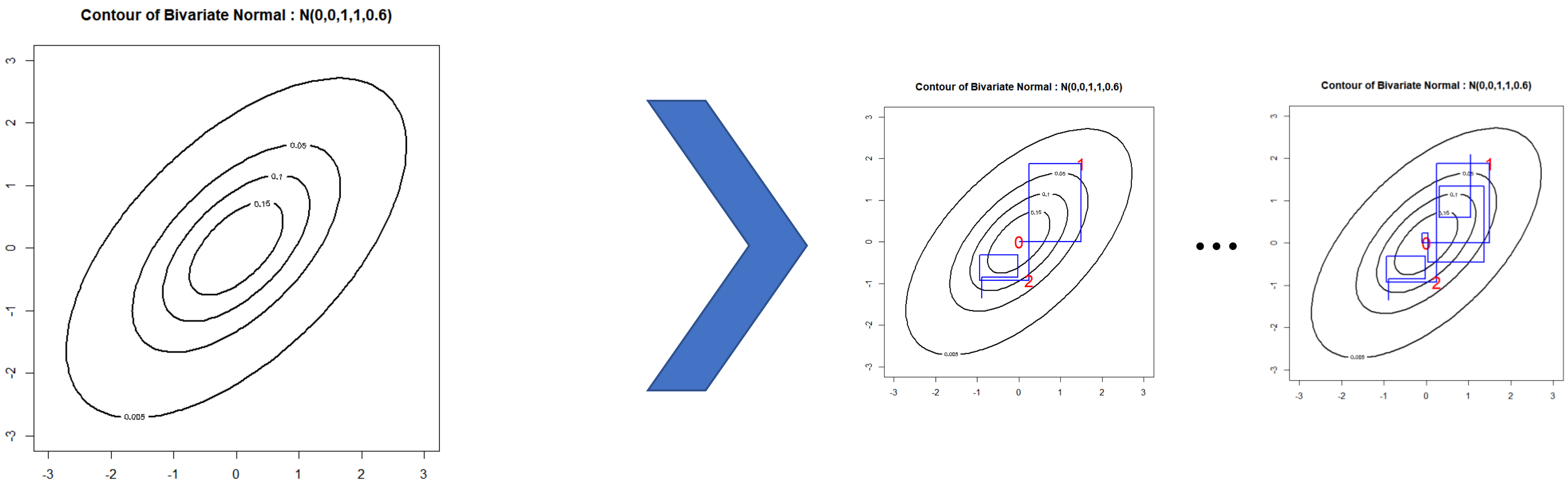
* 이변량 정규분포





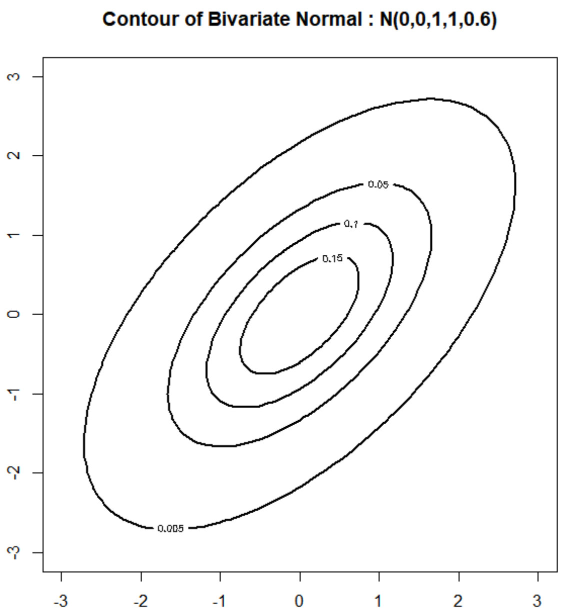
– Example

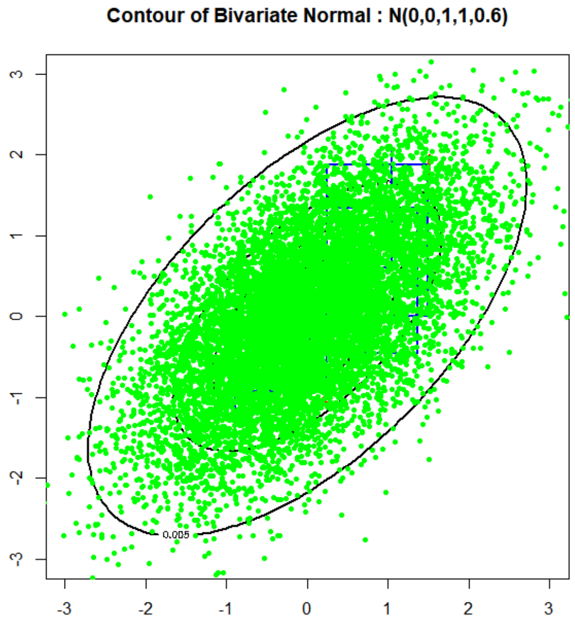
＊ 이변량 정규분포



– Example

＊ 이변량 정규분포





Q & A

감사합니다!

감사합니다?? ㅎㅎ 아직!

과제

– Easy!

이변량 분포의 two-dimentional p.d.f가 다음과 같이 있다.

$$f(x_1, x_2) = 2\exp\left(-x_1 x_2 - x_1 - x_2\right),\ x_1 > 0,\ x_2 > 0$$

깁스 샘플링을 사용해서, 사이즈 $n = 10^3$의 random vector $\{(X_{1i}, X_{2i})\}_{i=1}^{n}$를 생성하라. (burn-in period 꼭 고려하기!)
생성한 두 확률변수 샘플은 joint p.d.f $f(x_1, x_2)$를 따를 것이다.

<꼭 해주셔야 할 것!>

1. x_1, x_2 의 marginal p.d.f 구하기

2. 2차원의 coutour(등고선) 그래프 시각화하고 확인하기

3. 2차원 산점도를 그리고 목표 분포의 모양새 추정해보기