Markov Chain Monte Carlo (MCMC)

Gunwoong Park

Lecture Note

University of Seoul

강의 목표

- 사후분포의 근사적 추론(MCMC) 방법의 이해
    1. Gibbs Sampling
    2. Metropolis-Hastings
    3. Random Walk Metropolis
    4. Metropolis within Gibbs

Introduction

Introduction

- Objectives: Estimate θ from posterior distribution.
- 사후분포의 유도는 베이지안 추론의 핵심.
- 하지만, 모수가 여러개인 경우 사후분포의 유도가 매우 어려움.
- e.g.: *Normal*(μ, σ^2).

Two-Parameters Example

- Suppose that
$$X_1, \ldots, X_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2),$$
$$\mu \sim N(\mu_0, \sigma_0^2), \quad \sigma^2 \sim IG(a, b).$$
- In this case with $\mathbf{x} = (x_1, \ldots, x_n)$, posterior distribution for (μ, σ^2) is following:
$$\sigma^{-n} exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 + \frac{1}{\sigma_0^2}(\mu-\mu_0)^2\right\}\right](\sigma^2)^{-a-1} exp(-b/\sigma^2)$$
- 위 분포는 흔히 알려진 분포가 아니며, 결합 사후분포로부터의 표본생성도 쉽지 않다.

Example

- Posterior for σ^2 given μ is:

$$\pi(\sigma^2 \mid \mathbf{x}, \mu) \sim IG\left(a + \frac{n}{2}, b + \frac{1}{2}\sum(x_i - \mu)^2\right).$$

- Posterior for μ given σ^2 is Normal Distribution.
- μ 또는 σ^2 중 하나의 모수가 주어진다면, 사후분포로부터의 표본생성이 용이하다.
- 어떤 모수의 완전 조건부 사후분포(full conditional posterior distribution)란 그 모수를 제외한 다른 모든 모수가 주어졌을 때 해당 모수의 조건부 사후분포.

Gibbs Sampling Method

Full Conditional Posterior Distribution

- $\theta = (\theta_1, \theta_2, \theta_3)$의 각 원소모수의 완전 조건부 사후분포

$$\pi(\theta_1, | \, x, \theta_2, \theta_3), \quad \pi(\theta_2, | \, x, \theta_1, \theta_3), \quad \pi(\theta_3, | \, x, \theta_1, \theta_2)$$

  로부터의 표본추출이 용이한 경우를 생각해 보자.
- 이 경우 깁스 표본 기법 (Gibbs Sampling Method)의 알고리즘을 적용 시킬 수 있다.

Gibbs Sampling Method

Gibbs Sampling Method

Gibbs Sampling Method

- Gibbs Sampling Method은 Markov 성질을 가짐.
- $\theta_1^{(m)}$은 가장 최근의 $(\theta_2^{(m-1)}, \theta_3^{(m-1)})$에만 의존할 뿐 그 이전의 (θ_2, θ_3) 값에는 의존하지 않는다.
- m 이 커짐에 따라 $\theta^{(m)}$은 $\pi(\theta \mid x)$로 수렴한다.

$$\theta^{(m)} \sim \pi(\theta \mid x)$$

- 따라서 m 시점 이후 N개의 표본 $\theta^{(m+1)}, ..., \theta^{(m+N)}$를 θ의 사후표본으로 사용한다.

Burn-in Time

- m: $\theta^{(i)}$의 분포가 $\pi(\theta \mid x)$로 수렴하기까지 걸리는 시간 (warm-up time, burn-in time)을 의미.
- N: 표본크기.

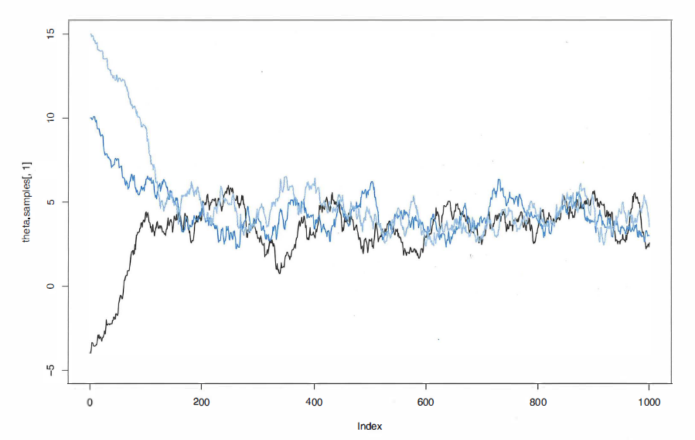
How to choose burn-in time m?

- 표본의 경로그림(trace plot):

  반복수(iteration number) vs 생성된 표본의 값

- 경로그림이 어느 시점 이후에 초기치의 영향을 벗어나 서로 잘 겹쳐지는지, 그리고 이들 경로그림이 어느 시점 이후에 증가하거나 감소하지 않고 안정적인 패턴을 보이는지 확인하는 것.

Gibbs Sampling Trace Plot



Properties of MCMC

- $\theta^{(i)}$ 들이 서로 마코브체인으로 관련되어 있으므로 독립된 표본이 아니다.
- Convergence.

$$\widehat{E}[\theta \mid x] = \frac{1}{N}\sum_{i=1}^{N} \theta^{(i)} \to E[\theta \mid x].$$

- Converging Speed: 수렴속도가 독립표본에 비하여 매우 느릴 수 있다.
- N개의 연관성 있는 표본이 갖는 정보는 독립된 표본들의 정보양이 적기 때문.

Properties of MCMC

- 수렴속도를 개선 하기 위해서 깁스 표본 알고리즘을 N번 독립적으로수행때 어 $\theta_1^{(m)}, ..., \theta_N^{(m)}$을 얻어 사용.
- 하지만 샘플을 얻는 시간 많이 걸리는 단점이 있음.
- 절충안: 깁스 표본 알고리즘을 충분히 길게 돌린 후 m번 이후에 k 간격을 두고 표본을 취하는 방법.
- $(\theta^{(m)}, \theta^{(m+k)}, \theta^{(m+2k)}, ..., \theta^{(m+(N-1)k)})$을 추정에 사용하는 것이다.
- 왜냐하면 샘플들의 연관성이 약해져 대략 독립표본으로 간주할 수 있기 때문.

Example for Normal Case

- Suppose that

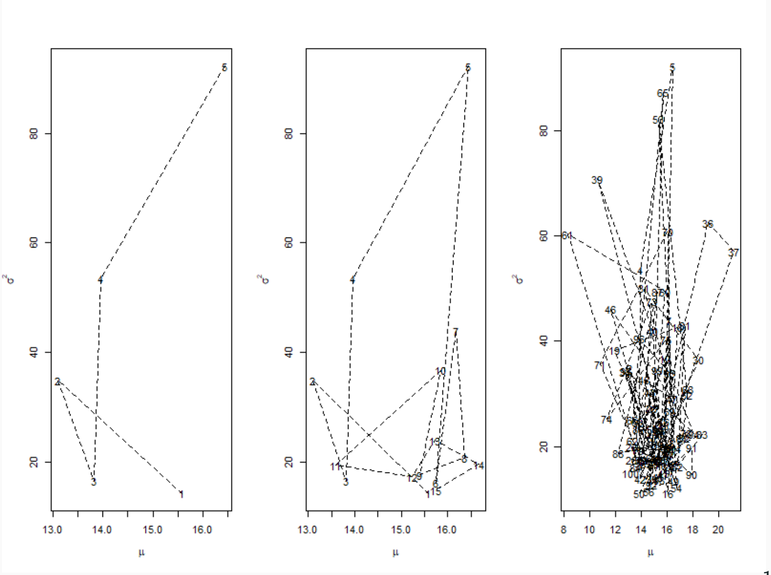
$$X_1, \ldots, X_{10} \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

$$\mu \sim N(10, 5^2), \quad \sigma^2 \sim IG(0.5, 1).$$

- Let $x_1, \ldots, x_{10} = (10, 13, 15, 11, 9, 18, 20, 17, 23, 21)$.
- Apply Gibbs sampling method for $\theta = (\mu, \sigma^2)$.

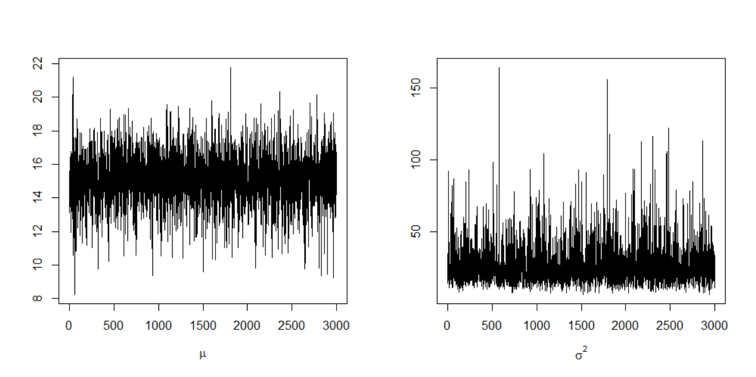
Trace Plots for μ and σ^2

- The trace plots (m = 5):



Trace Plots for each μ and σ^2

- The trace plots given other parameters (m = 5):



Example

```
M=3000; m=500; mu0=10; sigsq0=25; a=0.5; b=1
x=c(10,13,15,11,9,18,20,17,23,21)
n=length(x); xbar=mean(x); var.x=var(x)
THETA=matrix(nrow=M ,ncol=2)
sigsq=var.x  #initial value of sigsq

#####Start Simulation##########################
for(nsim in 1:M){
 #generate mu
 condpost.mu=(sigsq/n * mu0 +sigsq0*xbar)/(sigsq/n +sigsq0)
 condpost.var=1/(1/sigsq0 +n/sigsq)
 mu=rnorm(1, condpost.mu,sqrt(condpost.var))
 #generate sigsq
 condpost.a=a+n/2
 condpost.b =b+1/2*((n-1) *var.x +n *(xbar -mu)^2)
 sigsq =1/rgamma(1,condpost.a , condpost.b)
 #save
 THETA[nsim, ] = c(mu,sigsq)
}
#######End Simulation ################################
```

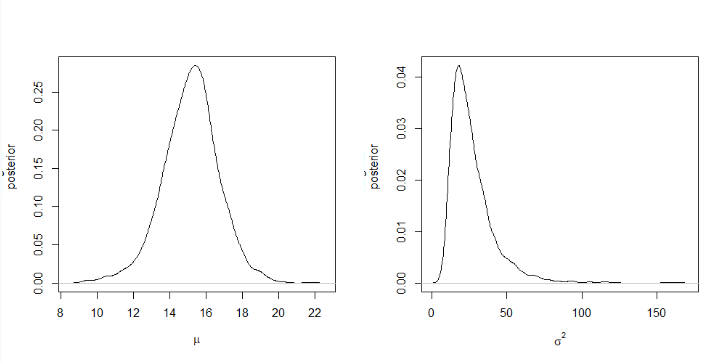
Example

```
#### Trace Plot 1 ######
par(mfrow=c(1,3))
plot(THETA[1:5, ], type="n",
xlab=expression(mu),ylab=expression(sigma^2))
lines(THETA[1:5, ],lty=2)
for(i in 1:5) text(THETA[i,1 ],THETA[i,2],i)
plot(THETA[1: 15, ] , type="n",
xlab=expression(mu),ylab=expression(sigma^2))
lines(THETA[1:15, ], lty=2)
for(i in 1:15) text(THETA[i,1 ],THETA[i,2],i)
plot(THETA[1:100, ], type="n",
xlab=expression(mu),ylab=expression(sigma^2))
lines(THETA[1:100, ], lty=2)
for(i in 1:100) text(THETA[i,1 ],THETA[i,2],i)

#### Trace Plot 2 ######
par(mfrow=c(1,2))
plot(THETA[,1], type="l", xlab=expression(mu), ylab="")
plot(THETA[,2], type="l", xlab=expression(sigma^2), ylab="")
```

Example

- 깁스 표본기법에서 μ, σ^2의 주변 사후분포:




```
par(mfrow=c(1,2))
plot(density(THETA[m:M,1]),xlab=expression(mu),ylab="marginal
posterior",main="")
plot(density(THETA[m:M,2]),xlab=expression(sigma^2),ylab="marginal
posterior", main="")
```

Metropolis-Hastings

Problem of Gibbs Sampling

- 깁스표본기법은 각 원소모수의 완전 조건부 사후분포가 표본생성이 용이한 분포여야 한다
- 즉 각 원소 θ_i들의 완전 조건부 사후분포 $\pi(\theta_i \mid x, \theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_p)$로부터 직접적인 표본생성이 가능해야 한다.
- 사전분포가 conjugate prior가 아닌 경우에, 완전 조건부 사후분포로부터 표본추출이 용이하지 않아, 깁스표본기법의 사용이 **불가능**.

Metropolis-Hastings Algorithm

1. 후보 표본 θ^*의 추출:

$$\theta^* \sim q(\theta \mid \theta^{(k-1)}).$$

2. 채택확률 계산:

$$\alpha^* = \frac{\pi(\theta^* \mid x)}{\pi(\theta^{(k-1)} \mid x)} \frac{q(\theta^{(k-1)} \mid \theta^*)}{q(\theta^* \mid \theta^{(k-1)})}, \quad \alpha = \min\{1, \alpha^*\}$$

3. θ^*의 채택 또는 기각

$$\begin{aligned} u &\sim U(0,1), \\ \theta^{(k)} &= \begin{cases} \theta^* & \textit{if} \quad u \le \alpha, \\ \theta^{(k-1)} & \textit{if} \quad u > \alpha. \end{cases} \end{aligned}$$

Metropolis Algorithm

- 후보추출함수 (Sample Generating Density $q(\cdot)$): θ^*를 추출하기 위해 임의로 선택된 밀도함수.
- 후보추출함수 $q(\cdot)$ 가 대칭인 경우 메트로폴리스 알고리즘이라고 한다.

$$q(\theta^* \mid \theta^{(k-1)}) = q(\theta^{(k-1)} \mid \theta^*)$$

- 채택확률 α^*이 사후밀도함수의 비가 된다.

$$\alpha^* = \frac{\pi(\theta^* \mid x)}{\pi(\theta^{(k-1)} \mid x)}$$

Random Walk Metropolis Algorithm

- 많은 경우 $q(.)$를 정하기 쉽지 않다.
- 메트로폴리스 알고리즘에서 k단계의 후보 표본을 현재 표본 $\theta^{(k-1)}$을 평균으로 하는 정규분포에서 추출할 경우 이를 랜덤워크(random walk) 메트로폴리스 기법이라고 한다.

$$q(\cdot \mid \theta^{(k-1)}) = N(\theta^{(k-1)}, \delta^2 I).$$

- δ 값은 랜덤워크의 보폭에 해당.
- δ 가 너무 작으면 높은 자기상관관계로 인해 수렴이 늦어질 수 있다.
- δ 가 너무 크면 확률이 낮은 영역에서 후보 표본을 추출 할 가능성이 높다.

Properties of Metropolis-Hastings Algorithm

- 용이한 계산: 채택확률 α^*의 계산에서 사후밀도함수의 비는 사후밀도함수식의 정규화 상수를 몰라도 계산이 가능.
- 사후밀도함수의 비:

$$\begin{aligned}\frac{\pi(\theta^* \mid \mathbf{x})}{\pi(\theta^{(k-1)} \mid \mathbf{x})} &= \frac{f(\mathbf{x} \mid \theta^*)\pi(\theta^*)/f(\mathbf{x})}{f(\mathbf{x} \mid \theta^{(k-1)})\pi(\theta^{(k-1)})/f(\mathbf{x})} \\ &= \frac{f(\mathbf{x} \mid \theta^*)\pi(\theta^*)}{f(\mathbf{x} \mid \theta^{(k-1)})\pi(\theta^{(k-1)})}\end{aligned}$$

since

$$\pi(\theta \mid \mathbf{x}) = f(\mathbf{x} \mid \theta)\pi(\theta)/f(\mathbf{x})$$

Metropolis-Hastings Algorithm for Multiple Parameters

- Suppose that $\theta = (\theta_1, \theta_2)$. For step k):

  (1) $\theta_1^* \sim q_1(\theta_1 \mid \mathbf{x}, \theta_1^{(k-1)}, \theta_2^{(k-1)})$

  (2)

$$\begin{aligned}
\alpha_1^* &= \frac{\pi(\theta_1^*, \theta_2^{(k-1)} \mid \mathbf{x})}{\pi(\theta_1^{(k-1)}, \theta_2^{(k-1)} \mid \mathbf{x})} \times \frac{q_1(\theta_1^{(k-1)} \mid \theta_1^*, \theta_2^{(k-1)})}{q_1(\theta_1^* \mid \theta_1^{(k-1)}, \theta_2^{(k-1)})} \\
\alpha_1 &= min\{1, \alpha_1^*\}
\end{aligned}$$

  (3) $u_1 \sim U(0,1)$

$$\theta_1^{(k)} = \begin{cases} \theta_1^* & \text{if } u_1 \leq \alpha_1 \\ \theta_1^{(k-1)} & \text{if } u_1 > \alpha_1 \end{cases}$$

Metropolis-Hastings Algorithm for Multiple Parameters

- $\theta_2^{(k)}$의 추출:

  (1) $\theta_2^* \sim q_2(\theta_2 \mid \mathbf{x}, \theta_1^{(k)}, \theta_2^{(k-1)})$

  (2)

$$\begin{aligned} \alpha_2^* &= \frac{\pi(\theta_1^{(k)}, \theta_2^* \mid \mathbf{x})}{\pi(\theta_1^{(k)}, \theta_2^{(k-1)} \mid \mathbf{x})} \times \frac{q_2(\theta_2^{(k-1)} \mid \theta_1^{(k)}, \theta_2^*)}{q_2(\theta_2^* \mid \theta_1^{(k)}, \theta_2^{(k-1)})} \\ \alpha_2 &= min\{1, \alpha_2^*\} \end{aligned}$$

  (3) $u_2 \sim U(0,1)$

$$\theta_2^{(k)} = \begin{cases} \theta_2^* & \text{if } u_2 \leq \alpha_2 \\ \theta_2^{(k-1)} & \text{if } u_2 > \alpha_2 \end{cases}$$

Gibbs Sampling vs Metropolis-Hastings Algorithm

- 깁스 표본기법은 메트로폴리스-헤스팅스 알고리즘의 특수한 경우.
- 깁스 표본기법에서는 후보추출함수 $q(.)$가 각 원소 모수의 완전 조건부 사후밀도함수이다.

$$q_1(\theta_1 \mid \mathbf{x}, \theta_1^{(k-1)}, \theta_2^{(k-1)}) = \pi(\theta_1 \mid \mathbf{x}, \theta_2^{(k-1)})$$
$$q_2(\theta_2 \mid \mathbf{x}, \theta_1^{(k)}, \theta_2^{(k-1)}) = \pi(\theta_2 \mid \mathbf{x}, \theta_1^{(k)})$$

- 채택확률 α_1^*:

$$\begin{aligned}
& \frac{\pi(\theta_1^*, \theta_2^{(k-1)} \mid \mathbf{x})}{\pi(\theta_1^{(k-1)}, \theta_2^{(k-1)} \mid \mathbf{x})} \times \frac{\pi(\theta_1^{(k-1)} \mid \mathbf{x}, \theta_2^{(k-1)})}{\pi(\theta_1^* \mid \mathbf{x}, \theta_2^{(k-1)})} \\
= \quad & \frac{\pi(\theta_1^* \mid \mathbf{x}, \theta_2^{(k-1)})\pi(\theta_2^{(k-1)} \mid \mathbf{x})}{\pi(\theta_1^{(k-1)} \mid \mathbf{x}, \theta_2^{(k-1)})\pi(\theta_2^{(k-1)} \mid \mathbf{x})} \times \frac{\pi(\theta_1^{(k-1)} \mid \mathbf{x}, \theta_2^{(k-1)})}{\pi(\theta_1^* \mid \mathbf{x}, \theta_2^{(k-1)})} \\
= \quad & 1
\end{aligned}$$

Gibbs Sampling vs Metropolis-Hastings Algorithm

- 마찬가지로 $\alpha_2^* = I$ 이 되어 항상 $mathbb\theta_i^{(k)} = mathbb\theta_i^*$가 성립한다.
- 따라서 깁스 표본기법은 메트로폴리스-헤스팅스 기법에서 매번 후보표본 θ^*를 $\theta^{(k)}$로 받아들이며 후보의 채택확률은 1이다.

Metropolis within Gibbs

- 깁스 표본기법이 메트로폴리스-헤스팅스 기법의 특수한 경우이므로 θ를 부분적으로 나누어 메트로폴리스-헤스팅스 기법을 적용할 때 θ_2에 대하여 깁스 표본기법 적용이 가능하다면 θ_2는 깁스 표본기법을 적용하고 θ_1은 메트로폴리스-헤스팅스 기법을 적용할 수 있다.
- 역으로 깁스 표본기법을 적용하는데 일부 원소에서는 완전 조건부 사후분포로부터의 표본생성이 용이하지 않다면 이들 원소에 대해서는 메트로폴리스-헤스팅스 기법을 적용하고 나머지 원소들에 대해서는 깁스 기법을 적용해도 된다.