Chapter 10: Markov Chain Monte Carlo (MCMC)

강의 목표

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

Introduction

- Objectives: Estimate θ from posterior distribution.
- 사후분포의 유도는 베이지안 추론의 핵심이다.
- 하지만, 모수가 한개가 아닌 경우에는 사후분포의 유도가 어렵다.
- e.g.: $Normal(\mu, \sigma^2)$.

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 간단치 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.
- If some parameters μ, σ^2 is provided, 사후분포로부터의 표본생성이 용이하다.
- 어떤 모수의 완전 조건부 사후분포(full conditional posterior distribution)란 그 모수를 제외한 다른 모든 모수가 주어졌을 때 해당 모수의 조건부 사후분포를 말한다.

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 깁스표본기법 (Gibbs Sampling Method)의 알고리즘을 적용 시킬 수 있다.

Gibbs Sampling Method

- 겁스 표본기법 알고리즘은 조건에 들어가는 모수 값을 가장 최근의 값들로 대제하면서 차례대로 $(\theta_1, \theta_2, \theta_3)$ 을 완전 조건부 사후분포로부터 추출하는 것이다.

Step 0): $(\theta_1, \theta_2, \theta_3)$의 초기값 $(\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)})$를 정한다.

Step 1): $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)})$의 값 생성:

$$\theta_1^{(1)} \sim \pi(\theta_1 \mid x, \theta_2^{(0)}, \theta_3^{(0)})$$

$$\theta_2^{(1)} \sim \pi(\theta_2 \mid x, \theta_1^{(1)}, \theta_3^{(0)})$$

$$\theta_3^{(1)} \sim \pi(\theta_3 \mid x, \theta_2^{(1)}, \theta_3^{(1)})$$

Gibbs Sampling Method

Step 2): $\theta^{(2)} = (\theta_1^{(2)}, \theta_2^{(2)}, \theta_3^{(2)})$의 값 생성:

$$\theta_1^{(2)} \sim \pi(\theta_1 \mid x, \theta_2^{(1)}, \theta_3^{(1)})$$
$$\theta_2^{(2)} \sim \pi(\theta_2 \mid x, \theta_1^{(2)}, \theta_3^{(1)})$$
$$\theta_3^{(2)} \sim \pi(\theta_3 \mid x, \theta_2^{(2)}, \theta_3^{(2)})$$

Step m): $\theta^{(m)} = (\theta_1^{(m)}, \theta_2^{(m)}, \theta_3^{(m)})$의 값 생성:

$$\theta_1^{(m)} \sim \pi(\theta_1 \mid x, \theta_2^{(m-1)}, \theta_3^{(m-1)})$$
$$\theta_2^{(m)} \sim \pi(\theta_2 \mid x, \theta_1^{(m)}, \theta_3^{(m-1)})$$
$$\theta_3^{(m)} \sim \pi(\theta_3 \mid x, \theta_2^{(m)}, \theta_3^{(m)})$$

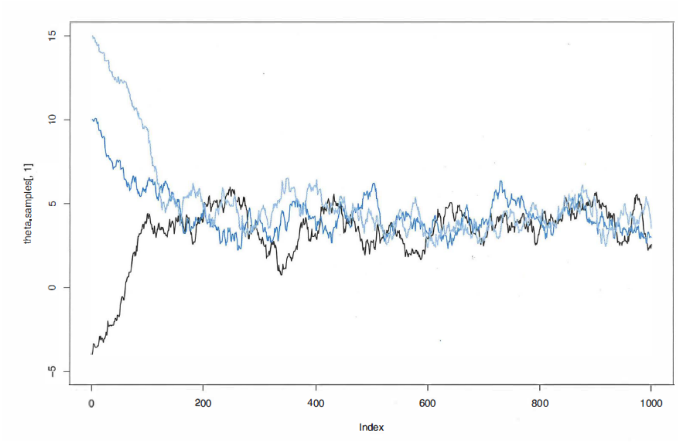
Gibbs Sampling Method

- Gibbs Sampling Method은 마코브 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.
- m이 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.

Gibbs Sampling Trace Plot)



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 표본의값

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

Propeties 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.

Propeties 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

- Suppose that

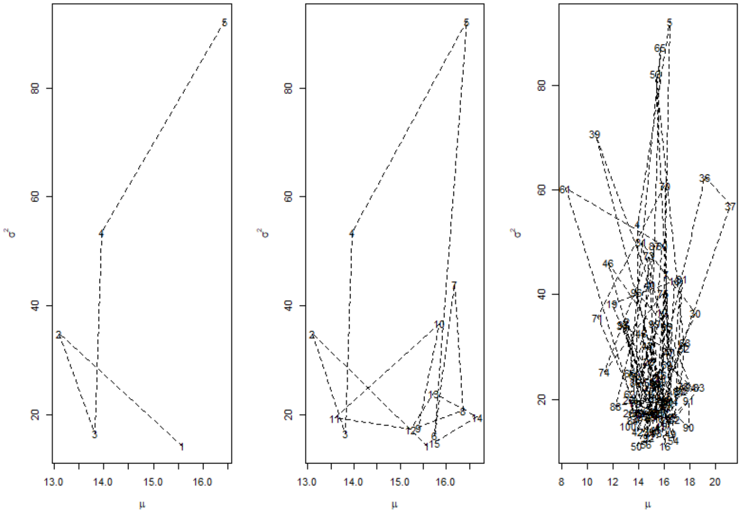
$$X_1, \ldots, X_{10} \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

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

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

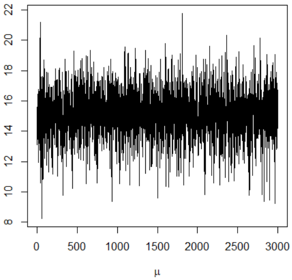
Example

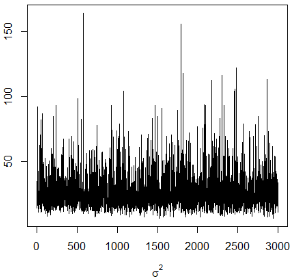
- We first draw the trace plots (m = 5, 15, 100):



Example

- We first draw the trace plots given other parameters (m = 5, 15, 100):





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
#####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################################
```

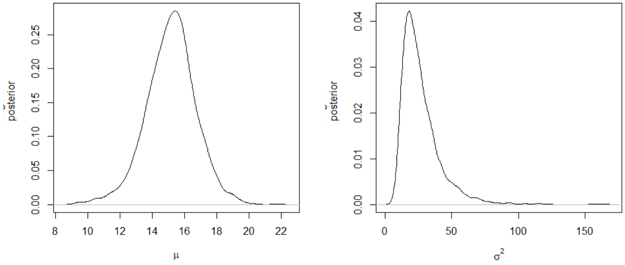
Example

```
#### 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="")
```

Another Example: Why Gibbs Sampling

- $\mathbf{Y} = (Y_1, \ldots, Y_5)$가 다음과 같은 모수를 갖는 다항분포를 따른다.

$$\mathbf{Y} \sim Multi(n, a_1\theta + b_1, a_2\theta + b_2, a_3\eta + b_3, a_4\eta + b_4, c(1-\theta-\eta)),$$
$$a_i, b_i \leq 0, \quad 0 < c = 1 - \sum b_i = a_1 + a_2 = a_3 + a_4 < 1.$$

- 이 모형에서 미지의 모수는 (θ, η)이며 $0 < \theta < 1$, $0 < \eta < 1$, $0 < \theta + \eta < 1$의 제한조건을 갖는다.
- 모수(θ, η)에 $Dirichlet(\alpha_1, \alpha_2, \alpha_3)$ 사전분포를 가정하면

$$\pi(\theta, \eta) \propto \theta^{\alpha_1 - 1}\eta^{\alpha_2 - 1}(1-\theta-\eta)^{\alpha_3 - 1}$$
$$0 < \theta < 1, 0 < \eta < 1, 0 < \theta + \eta < 1$$

Another Example

- 사후밀도함수 (Posterior distribution):

$$\begin{aligned}\pi(\theta, \eta \mid \mathbf{y}) &\propto (a_1\theta + b_1)^{y_1}(a_2\theta + b_2)^{y_2}(a_3\theta + b_3)^{y_3}(a_4\theta + b_4)^{y_4} \\ &\times (c(1 - \theta - \eta))^{y_5}\theta^{\alpha_1 - 1}\eta^{\alpha_2 - 1}(1 - \theta - \eta)^{\alpha_3 - 1}\end{aligned}$$

- 사후밀도함수가 θ와 η의 복잡한 다항식이다.
- 그러므로 이 사후분포로부터 θ와 η의 사후기대치, 사후분산, 사후주변분포를 수리적으로 구하기는 어려움이 따른다.

Another Example

- 이 문제를 깁스 표본기법으로 해결하는 방법을 살펴보자.
- Let $\mathbf{X} = (X_1, \ldots, X_9)$ and

$$X \sim Multi(n, a_1\theta, b_1, a_2\theta, b_2, a_3\eta, b_3, a_4\eta, b_4, c(1-\theta-\eta)).$$

- 만약 $\mathbf{X}$의 관측치 $\mathbf{x}$가 주어진다면

$$\begin{aligned}
\pi(\theta, \eta \mid \mathbf{x}) &\propto (a_1\theta)^{x_1} b_1^{x_2} (a_2\theta)^{x_3} (a_3\eta)^{x_5} b_3^{x_6} (a_4\eta)^{x_7} b_4^{x_8} (c(1-\theta-\eta))^{x_9} \\
&\times \theta^{\alpha_1 - 1} \eta^{\alpha_2 - 1} (1-\theta-\eta)^{\alpha_3 - 1} \\
&\propto \theta^{x_1 + x_3 + \alpha_1 - 1} \eta^{x_5 + x_7 + \alpha_2 - 1} (1-\theta-\eta)^{x_9 + \alpha_3 - 1}
\end{aligned}$$

- 따라서 $\pi(\theta, \eta \mid x)$는 디리슈레(Dirichlet) 분포를 따르므로 사후추론이 매우 간편해진다.

Another Example

- X와 Y의 관계를 표로 나타내면 아래표와 같다.
- 5개의 관측가능한 범주 중 Y_1은 첫 번째 범주의 도수이다.
- 범주의 확률이 $a_1\theta + b_1$이므로 이 사건은 두 가지 요인 A_1, B_1에 의해 발생한다고 볼 수 있으며 $P(A_1) = a_1\theta$, $P(B_1) = b_q$.

키(cm)	Y_1		Y_2		Y_3		Y_4		Y_5	n
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	n
체중(kg)	$a_1\theta$	b_1	$a_2\theta$	b_2	$a_3\eta$	b_3	$a_4\eta$	b_4	$c(1-\theta-\eta)$	1

Table: X와 Y의 관계

Another Example

- A_1, B_1에 대응하는 도수는 X_1, X_2이다.
- X_1과 X_2의 확률이 모수에 의존하는 구조가 다르기 때문에 X_1과 X_2가 관측된다면 모수추정이 용이하겠지만 X_1, X_2는 관측 불가능하고 단지 통합 도수 Y_1만 가능할 뿐이다. 다른 범주에 대해서도 마찬가지이다.
- X와 Y의 관계를 살펴보면

$$
\begin{aligned}
X_1 + X_2 &= Y_1, \quad X_3 + X_4 = Y_2, \\
X_5 + X_6 &= Y_3, \quad X_7 + X_8 = Y_4, \\
X_9 &= Y_5.
\end{aligned}
$$

Another Example

- $X_2 = Y_1 - X_1$으로부터 X_1과 Y_1을 알면 X_2를 알 수 있기 때문에 (X_1, X_2)를 $(X_1 Y_1)$으로 대체할 수 있다.
- 마찬가지로, (X_3, X_4)를 (X_3, Y_2)로, (X_5, X_6)를 (X_5, Y_3)로, (X_7, X_8)을 (X_7, Y_4)로 대체할 수 있다.
- 단, X_1, X_3, X_5, X_7은 관측이 되지 않는 잠재변수(latent variables)이다. 잠재변수와 Y를 합하여 $X = (X_1, X_3, X_5, X_7, Y)$를 새로운 관측변수로 간주하면 (θ, η)의 완전 조건부 분포는 디레슈레 분포로 주어진다.
- 잠재변수 X_1, X_3, X_5, X_7은 미지의 변수이므로 이들을 모수와 같이 취급하여 깁스 과정에서 표본을추출한다.

Another Example

- 종합하면, (θ, η)의 완전 조건부 사후분포를 구해 보면

$$\theta, \eta \mid \mathbf{x} \sim Dirichlet(x_1 + x_3 + \alpha_1, x_5 + x_7 + \alpha_2, x_9 + \alpha_3)$$

- 또한 표를 보면 $Y_1 = y_1$이 주어졌을때 X_1은 두요인 중 요인 A_1의 도수이므로 θ와 y_1이 주어졌을 때 X_1의 조건부 분포는 $B(y_1, \frac{a_1\theta}{a_1\theta + b_1})$이다.
- 마찬가지로 X_3, X_5, X_7의 조건부분포는

$$\begin{aligned}
X_3 \mid \theta, y_2 &\sim B(y_2, \frac{a_2\theta}{a_2\theta + b_2}) \\
X_5 \mid \eta, y_3 &\sim B(y_3, \frac{a_3\eta}{a_3\eta + b_3}) \\
X_7 \mid \eta, y_4 &\sim B(y_4, \frac{a_4\eta}{a_4\eta + b_4}).
\end{aligned}$$

Another Example

- 이들 조건부 분포 모두 난수생성이 쉬운 분포이므로 깁스 표본기법을 이용하여 (X_1, X_3, X_5, X_7)과 (θ, η)의 표본을 생성할 수 있고, 이 중 (θ, η)의 표본만 취하면 이들은 $\pi(\theta, \eta \mid \mathbf{y})$를 따르는사후표본이다.
- 따라서 깁스 표본기법으로부터 생성된 (θ, η)의 표본을 사용하여 (θ, η)에 대한 사후추론을 수행할 수 있다.
- 다음은 $\mathbf{y} = (14, 1, 1, 1, 5)$가
$$Multi\left(22, \frac{1}{4}\theta + \frac{1}{8}, \frac{1}{4}\theta, \frac{1}{4}\eta, \frac{1}{4}\eta + \frac{3}{8}, \frac{1}{2}(1 - \theta - \eta)\right)$$
로부터의 관측치일 때 $\pi(\theta, \eta) \sim Dirichlet(1, 1, 1)$ 사전분포를 사용하여 (θ, η)에 대한 사후추론을 시행하는 깁스표본기법을 적용하자.

Another Example

```
Y=c(14,1,1,1,5); n=22; a=c(0.25,0.25,0.25,0.25)
b=c(0.125,0,0,0.375); c=0.5; alpha=c(1,1,1);
m=1000; N=3000
THETA= matrix(0,m+N,3); X=c(1:9)*0
#Dirichelet generator、
rDirich=function(n,p,a){
x=matrix(0,n,p)
for(i in 1:p) x[,i]=rgamma(n,a[i])
x=x/apply(x,1,sum)
}
#initialize
theta= 0.5; eta=0.5; THETA[1,]= c(theta,eta, 1-theta-eta)
X[1]= 0.5*Y[1]; X[3]= 0.5*Y[2]; X[5]= 0.5*Y[3]
X[7]=0.5*Y[4]; X[9]=Y[5]
```

Another Example

```
##################simulate#############################
for(i in 2:(m+N)){
p=3
aa=c(X[1]+X[3]+alpha[1],X[5]+X[7]+alpha[2],X[9]+alpha[3])
THETA[i,]=rDirich(1,p,aa)
theta=THETA[i,1]; eta=THETA[i,2]
X[1]=rbinom(1,Y[1], a[1]*theta/(a[1]*theta + b[1]))
X[3]=rbinom(1,Y[2], a[2]*theta/(a[2]*theta + b[2]))
X[5]=rbinom(1,Y[3], a[3]*eta/(a[3]*eta + b[3]))
X[7]=rbinom(1,Y[4], a[4]*eta/(a[4]*eta + b[4]))
}
#################End simulation #########################
theta.sim=THETA[ (m+1): (m+N), 1]; eta.sim=THETA[ (m+1): (m+N), 2]
par(mfrow=c(1,2))
plot(density(theta.sim), type="l",xlab=expression(theta),
ylab="posterior",main="")
plot(density(eta.sim), type="l", xlab=expression(eta), main="")
```

Another Example

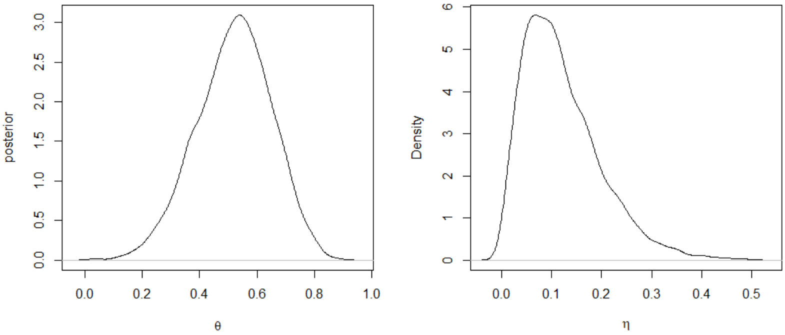


Figure: 깁스 표본기법으로부터 추정된 θ, η의 주변 사후분포

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:

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

$\alpha = \min\{1, \alpha^*\}.$

3. θ^*의 채택 또는 기각

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

Metropolis Algorithm

- 후보추출함수 (Sample Generating Density $q(.)$): θ^*를 추출하기 위해 임의로 선택된 밀도함수.
- 후보추출함수 $q(.)$ 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.

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

- δ값은 랜덤워크의 보폭에 해당
- δ가 너무 작으면 높은 자기상관관계로 인해 수렴이 늦어질 수 있다.
- δ가 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)

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

(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^*:

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

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 1 이다.

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.