8. 마르코프 연쇄 몬테 카를로 방법

(Markov Chain Monte Carlo Method; MCMC)

MCMC 방법이란 다변량 분포에서 표본을 근사적으로 추출하는 여러 가지 반복적인 방법들을 통틀어 일컫는 말이다. 그러면, 이러한 몇 가지 방법론과 그 기본 성질들을 설명하기에 앞서, 마르코프 연쇄의 기본적인 개념을 알아 보자.

8.1 마르코프 연쇄

이산 변량인 경우의 마르코프 연쇄를 가지고 그 성질을 살펴 보겠다. 마르코프 연쇄란 일련의 임의 변량, $X_0,\ldots,X_{n+1}$ 을 일컬으며, 이들 변량들이 다음과 같은 마르코프 특성을 만족함을 의미한다.

$$P(X_{n+1}=s_i \mid X_0=s_j,\ldots,X_n=s_k)=P(X_{n+1}=s_i \mid X_n=s_k)$$

위의 표현이 말하는 것은, 어떠한 계(system)의 움직임을 예측하고자 할 때 그 계의 과거 이력을 모두 알아야 할 필요 없이 단지 그 계의 현재 상태만을 알면 된다는 것이다. X_n 에 표기된 아래 첨자는 일련의 마르코프 과정에서의 단계 또는 시간이라고 볼 수 있다. 각각의 X_n 은 공통된 마르코프 연쇄상에서의 사상 범위를 가진다. 그리고 이 범위는 유한일 수도 무한일 수도 있다. 여기서는 유한이면서도 이산인 사상 $S=\{s_1,\ldots,s_r\}$ 만을 다루고자 한다. 마르코프 과정은 이들 사상에서 시간이 0 인 사건에서 출발하여 단위 시간이 흐름에 따라 한 단계씩 다음 단계로 이동한다. 그리고 이러한 단계별 전환을 *이동*(*step*)이라고 한다. 그리고 i 상태에서 j 로 전이될 확률 $P(i,j)$ 는 오직 이전 단계의 사건 s_i 에 따라서만 결정된다. 이때 확률 $P(i,j)$ 를 전이 확률(또는 전이 커늘, transition probability or transition kernel)이라고 하는데 이 확률은 조건부 확률이다. 그리고 행렬 $\mathbf{P}=\{P(i,j)\}$ 는

열의 합이 1 이 되는 전이 확률이다.

X_{n+1} 이라는 임의 변량이 $n+1$ 단계에서 $s_k\ (k=1,\ldots,r)$ 라는 값을 갖는다고 하자. 그러면, 마르코프 과정을 시작하기에 앞서 먼저 X_0에 대한 주변 분포, 즉, $P(X_0 = s_k) = p_k(X_0)$ 와 X_0가 주어진 조건하에서 X_1 의 조건부 분포인 $P(X_1 = s_j \mid X_0 = s_k) = P(k, j)$ 를 먼저 알아 봐야 할 것이다.

$n+1$ 단계에서 X_{n+1} 이 s_j 일 확률을 $p_j(X_{n+1})$ 로 나타내면 이는 다음과 같다.

$$\begin{aligned} p_j(X_{n+1}) &= P(X_{n+1} = s_j) \\ &= \sum_k P(X_{n+1} = s_j \mid X_n = s_k) P(X_n = s_k) \\ &= \sum_k P(k, j) P(X_n = s_k) \\ &= \sum_k P(k, j) p_k(X_n) \end{aligned} \qquad 8.1$$

위의 식 (8.1)에서 마지막 줄을 보면 n 단계에 대한 확률을 이용하여 $n+1$ 단계의 확률을 나타내고 있다. 이를 행렬의 형태로 나타내면 다음과 같이 쓸 수 있다.

$$p(X_{n+1}) = p(X_n)\mathbf{P} \qquad 8.2$$

이를 풀어 써보면, $p(X_1) = p(X_0)\mathbf{P}$, $p(X_2) = p(X_1)\mathbf{P} = p(X_0)\mathbf{PP}$ 와 같이 전개되며, 결국에는 다음과 같은 일반식을 유도할 수 있다.

$$p(X_n) = p(X_0)\mathbf{P}^n \qquad 8.3$$

마르코프 과정이 ***평형***(***또는 불변의***) ***분포***, p^* 에 이르게 되면, 즉, $p(X_n)=p(X_{n+1})$일 때, p^*은 다음과 같은 관계를 만족시킨다.

$$p^*P=p^* \tag{8.4}$$

다시 말해서, 일단 식(8.4)가 만족되면, p^*는 불변 분포를 한다(마르코프 연쇄에서는 몇 개의 불변하는 분포가 존재할 수 있다). 즉, 어느 단계에서의 주변 분포가 p^*이면 그 다음 단계의 분포는 $p^*P=p^*$가 된다. 그리고 연쇄가 평형 분포 p^*에 이르게 되면, 그 다음 이동에서도 계속 이 분포가 유지된다.

만약에 연쇄 반응이 다음과 같은 ***상세 균형***(*detailed balance*) (그리고 연쇄가 ***가역적*** (*reversible*)이면)의 성질을 만족한다면, 불변 분포에 수렴한다는 것은 자동적으로 이루어질 것이다.

모든 j와 k에 대하여

$$P(j,k)p_j^* = P(k,j)p_k^* \tag{8.5}$$

이 성립한다. 왜냐하면,

$$\begin{aligned}\sum_j P(j,k)p_j^* &= \sum_j P(k,j)p_j^* \\ &= p_j^* \sum_j P(k,j) \\ &= p_k^*\end{aligned} \tag{8.6}$$

처럼 식 (8.4)의 형태를 보이기 때문이다. 또한 식 (8.5)와 같은 표현을 ***상세***

균형 방정식이라고 한다. 이와 같이 평형 분포 p^*로부터 표본을 추출하기 위하여, 식 (8.5)와 같은 가역적인 조건을 만족하는 전이 행렬 $\mathbf{P}$를 이용하여 마르코프 연쇄를 평형 분포에 도달할 때까지 돌리게 된다. 그러나 그 평형 분포가 과연 유일한 것인지 또 그 평형에 얼마나 수렴을 하는지에 대한 의문은 남는다. $\mathbf{P}$가 ***비주기적***(*aperiodic*)일 때 $\mathbf{P}$가 더 이상 나눌 수 없는 ***기약 행렬***(*irreducible*)이고 연쇄가 p^*로 수렴한다면, 그 평형 분포는 유일하다고 할 수 있다. (연쇄가 정기적인 간격으로 어떤 상태로 되돌아가게 될 때, $\mathbf{P}$를 ***주기적***(*periodic*)이라고 한다. 그리고, ***기약 연쇄***(*irreducible chain*) 란 모든 단계 간에 서로 연결되어 $\mathbf{P}$에 요약된 확률하에서 어떤 단계 간에 전이가 가능하다는 말이다.)

예제 8.1.1

가상적으로 비(R)나 맑음(N)과 구름(C)과 같은 세 가지만의 기상 상태만 있다고 하자. 그러면 이러한 기상 조건의 사상은 $S = \{R, N, C\}$가 되고 다음과 같은 전이 행렬 $\mathbf{P}$를 가정해 보자.

$$\mathbf{P} = \begin{matrix} R \\ N \\ C \end{matrix} \begin{bmatrix} \overset{R}{0.5} & \overset{N}{0.25} & \overset{C}{0.25} \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \qquad 8.1.1.1$$

가령, $n = 0$일 때 날씨가 맑다(N)고 하면, 1 단계(날)에서 세 가지 상태에 대한 확률 분포는 행렬 (8.1.1.1)로부터 다음과 같이 바로 구할 수 있다.

$$\begin{aligned} &(P(X_1 = R), P(X_1 = N), P(X_1 = C)) \\ &= (p_R(X_1), p_N(X_1), p_C(X_1)) \\ &= (0.5, 0, 0.5) \\ &= p(X_1) \end{aligned}$$

다른 방법으로는 $p(X_0) = (0, 1, 0)$ 라고 정의하고서 식 (8.3)을 적용하면 $p(X_1)$ 을 구할 수 있다.

그리고 2 단계에서의 세 가지 상태에 대한 확률을 구할 때에도 식 (8.3)을 적용하여 다음과 같이 계산하게 된다.

$$p(X_2) = p(X_0)\, P^2 = (0.375, 0.25, 0.375)$$

그리하여, $n = 11$이 되면 이 연쇄는 다음과 같은 평형 상태로 수렴한다.

$$p^* = (0.4, 0.2, 0.4)$$

이러한 평형은 시발점이 $p(X_0) = (1, 0, 0)$ 이거나 $p(X_0) = (0, 1, 0)$ 나 $p(X_0) = (0, 0, 1)$ 이든가에 상관없이 같다. 그리고 이것이 과연 평형 상태인가 하는 것은 다음 관계식으로써 입증이 된다.

$$p^*\mathbf{P} = (0.4, 0.2, 0.4) = p^*$$

더욱이 이 연쇄 반응은 가역적이라는 조건을 만족시킨다. 이를 증명하기 위해서 먼저 $R = 1, N = 2, C = 3$ 로 각각 표시하면 $p_R^* = p_1^*$ 과 $p_N^* = p_2^*$ 및 $p_C^* = p_3^*$ 로 나타낼 수 있으므로 다음과 같은 상세 평형 방정식으로 점검할 수 있다.

$$p_1^* P(1,2) = (0.4)(0.25) = 0.1 = p_2^* P(2,1) = (0.2)(0.5)$$

$$p_1^* P(1,3) = (0.4)(0.25) = 0.1 = p_3^* P(3,1) = (0.4)(0.25)$$

$$p_2^* P(2,3) = (0.2)(0.5) = 0.1 = p_3^* P(3,2) = (0.4)(0.25)$$

예제 8.1.2

A와 a라는 두 개의 대립 유전자를 가지는 한 유전자좌의 유전 형태를 알아 보자. 가령, 0 단계에서 이형 접합체 Aa를 집단으로부터 임의로 개체를 선택하여 교배시켰다고 하자. 그리고 여기서 나온 자손을 제 1 세대에서 다시 이형 접합인 개체에 교배시키는 과정을 매 세대마다 같은 방법으로 반복하였을 때 평형 분포 p^*를 계산해 보자. 이 연쇄 반응에서는 상태가 세 가지이며 전이 확률 행렬은 다음과 같다.

$$\mathbf{P} = \begin{matrix} AA \\ Aa \\ aa \end{matrix} \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix}$$

0 단계에서 개체가 AA, Aa 및 aa 개체 중 어떤 유전자형에 교배 되느냐에 따라 0 단계의 확률 분포는 $p(X_0) = (1,0,0)$ 나 $p(X_0) = (0,1,0)$ 또는 $p(X_0) = (0,0,1)$이 된다. 물론 이때 이 이형 접합체가 두 가지 동형 접합체에 같은 비율로 교배 된다고 하면 $p(X_0) = (0.5, 0, 0.5)$이 된다. 그러면, 초기에 이 이형 접합체에 교배 되는 유전자형에 상관없이 평형 상태에서 가질 수 있는 세 가지 유전자형에 대한 확률 분포(집단의 크기가 충분히 크다고 가정할 때)는 충분히 큰 n을 식 (8.3)에 적용하여 계산하면 다음과 같다.

$$p^* = (0.25, 0.5, 0.25)$$

그리고 여기서 p^*가 식 (8.4)를 만족시키고 마르코프 연쇄가 가역적이라는 것을 다음과 같이 점검해 볼 수 있다.

$$p_1^* P(1, 2) = (0.25)(0.5) = 0.125 = p_2^* P(2, 1) = (0.5)(0.25)$$

$$p_1^* P(1, 3) = (0.25)(0) = 0 = p_3^* P(3, 1) = (0.25)(0)$$

$$p_2^* P(2, 3) = (0.5)(0.25) = 0.125 = p_3^* P(3, 2) = (0.25)(0.5)$$

이제 마르코프 연쇄의 수렴과 관련하여 중요한 몇 가지 결과를 정리해 보자. 마르코프 과정 $X_1, X_2, \ldots$에서 $X_1, X_2, \ldots$들이 연속 확률 변수이고 그 조건부 확률이 $X_{n-1} = x$ 라는 조건하에서 $X_n = y$ 이고 그 전이 확률이 $P(x, y)$ 라고 할 때 다음이 성립한다.

$$\int_{-\infty}^{\infty} P(x, y) dy = 1$$

여기서 X_n 의 pdf, $p_n(y)$ 를 계산하려면

$$p_n(y) = \int_{-\infty}^{\infty} p_{n-1}(x) P(x, y) dx$$

과 같이 할 수 있고 이 연쇄가 평형 분포 $p^*(y)$ 에 도달하면 이때 다음 관계식이 성립한다.

$$p^*(y) = \int_{-\infty}^{\infty} p^*(x)P(x,y)dx \qquad 8.1.2.1$$

이러한 결과는 식 (8.6)과 같은 이산의 경우와 동일하다. 이어서 이야기하고자 하는 것은 MCMC 방법에 있어서 중심이 되는 문제가 바로 우리가 표본을 추출하고자 하는 평형 분포의 확률 밀도 함수 즉, p^*에 도달하게 되는 전이 커늘, $P(x,y)$ 를 알아내는 것이라는 점이다. 예제 **8.1.1** 과 **8.1.2** 에서는 전이 확률 행렬 **P** 를 알고서 평형 상태의 분포를 알고자 하였다. 그러나, 실제로 MCMC 를 적용하는 경우에 있어서는 그 반대의 경우에 해당한다. 즉, 평형 분포는 (대개 그 비율의 형태로) 이미 알려져 있으나 전이 커늘을 모르는 경우가 일반적이다. 그러면, 마르코프 연쇄 반응이 수렴하게 되는 유일한 정적인 분포만 알고 있을 때, 어떻게 거기서 **P** 를 끄집어 낼 수 있을까?

8.2 메트로폴리스-헤이스팅스 알고리듬 (Metropolis-Hastings Algorithm)

메트로폴리스-헤이스팅스 알고리즘이란, 먼저 아무 분포에서든 표본 추출을 하고, 그 다음 이를 수정하는 절차를 거쳐 얻은 표본이 근사적으로 원래 표본을 추출하고자 하는 평형 분포에서 얻은 임의 관측치처럼 만드는 것을 말한다. n 단계에서 다음 단계로 갈 때, 우선 어떤 *후보 생성*(*proposal* or *candidate generating*) 분포 $q(x,y)$ 로부터 *후보점*(*candidate*) y 를 표본 추출한다. 이 밀도 함수는 아무 것이나 가능하고, 해석해 보면, n 단계에서 연쇄 과정은 x 상태에 있고, *후보 생성* 밀도로 인하여 $n+1$ 단계로 가는 사건 y 가 생성된다라고 할 수 있다. 여기서 밀도 $q(x,y)$ 는 현재의 상태인 x 에 따라 다르게 된다. 예를 들어 $q(x,y)$ 가 평균이 x 이고 고정적인 공분산 행렬을 가지는 다변량 분포를 따른다고 하자. 그리고 $q(x,y)$ 가 식 (8.5)의 가역 조건을 모든 (x,y) 에서 만족시킨다고 하면,

$$q(x,y)p^*(x) = q(y,x)p^*(y) \quad 8.2.1$$

이 성립하고, 만족할 만한 전이 커늘이 나타나면서 추출한 y **가 옳은 표본으로 받아들여질 것이다. 그러나 어떤** (x,y) **에서는 다음과 같은 관계를 보일 수 있다.**

$$q(x,y)p^*(x) > q(y,x)p^*(y) \quad 8.2.2$$

이런 경우에는 x **에서** y **로 전이하는 확률이 매우 높다(** y **에서** x **로 전이하는 확률은 매우 낮다). 이러한 상황을 없애기 위해,** $a(x,y) < 1$ **라는 이동에 대한 확률을 도입함으로써** x **에서** y **로 이동하는 횟수를 줄일 수 있다. 우리는 이것을 *이동 확률*(*probability of move*)이라고 부르기로 하자. 일단 이동이 일어 나면** $x_{n+1} = y$ **가 되고, 이동이 일어나지 않으면** $x_{n+1} = x_n$ **이 된다. 그리고** $a(y,x)$ **는 1 이고** $a(x,y) < 1$ **라고 정의해야 할 것이다. 또 가역성을 충족시키려면 다음 관계식이 성립해야 한다.**

$$\begin{aligned} q(x,y)p^*(x)a(x,y) &= q(y,x)p^*(y)a(y,x) \\ &= q(y,x)p^*(y) \end{aligned} \quad 8.2.3$$

따라서,

$a(x,y) = p^*(y)q(y,x)/p^*(x)q(x,y)$ **가 도출된다.**

이때 오직 x **에서** y **로 이동하는 경우만을 생각해 보면, 위 식이 1 이하 또는 클 경우를 생각할 수 있다. 그런데 뒤의 두 가지 경우에는 후보점** y **가 항상 수락된다.**

반면에, 부등식 (8.2.2)의 부등호가 거꾸로 될 경우에는 $a(x,y) = 1$ **라 두고 위와 같은 방식으로** $a(y,x)$ **를 유도할 수 있다. 여기서** $a(x,y)$ **와**

$a(y,x)$ 라는 확률을 이용한 것은 부등식 (8.2.2)가 두 가지 상황에 대해 균형을 이룸으로써 이들 확률로 가역 조건을 만족시키게 하는 것이다.

위와 같은 논리에서 이동 확률은 다음과 같이 결정한다.

$$a(x,y) = \begin{cases} \min\left[\dfrac{p^*(y)q(y,x)}{p^*(x)q(x,y)}, 1\right], & p^*(x)q(x,y) > 0 \text{ 일 경우} \\ 1, & \text{기타의 경우} \end{cases} \qquad 8.2.4$$

여기서 한 가지 새겨둘 것은, 조건식 (8.2.1)가 성립할 경우에는 항상 $a(x,y)=1$ 이 되고 후보 추출 y 가 항상 수락 된다는 점이다. 그러면 위의 식 (8.2.4)가 가역적인가를 알아 보자. x 와 y 를 평형 분포 $p^*(.)$ 에서 추출한 두 표본이라고 하고, 이에 대해 $p^*(y)q(y,x) \geq p^*(x)q(x,y)$ 가 성립하면서 따라서 $a(x,y)=1$ 라고 하자. 이런 경우에 x 에서 y 로의 전이 확률 밀도는 다음과 같이 된다.

$$\begin{aligned} &p^*(x)P(x,y) \\ &= p^*(x)q(x,y)a(x,y) \\ &= p^*(x)q(x,y) \end{aligned}$$

반면에 y 에서 x 로의 전이 확률 밀도는 다음과 같다.

$$\begin{aligned} &p^*(y)P(y,x) \\ &= p^*(y)q(y,x)a(y,x) \\ &= p^*(y)q(y,x)\frac{p^*(x)q(x,y)}{p^*(y)q(y,x)} \\ &= p^*(x)q(x,y) \end{aligned}$$

따라서 $p^*(x)P(x,y)=p^*(y)P(y,x)$ 가 성립함을 알 수 있고, 가역적이라고 결론지을 수 있다. 여기서 중요한 것은 $p^*(\cdot)$ 에 대해 정규화 상수까지만이라도 알아야 한다는 것이다. 왜냐하면, 상수가 $p^*(y)/p^*(x)$ 비율에서 상쇄될 수 있기 때문이다.

메트로폴리스-헤이스팅스 알고리듬은 다음과 같이 매우 간단하다.

```
INITIALIZE X0; SET n=0
DO I=1 TO M
    SAMPLE Y FROM q(.|Xn)
    SAMPLE U FROM Un(0,1)
    IF(U. LE. a(Xn, Y)) SET Xn+1 =Y
    OTHERWISE SET Xn+1 =Xn
    n=n+1
ENDDO
```

8.3 깁스 샘플러

메트로폴리스-헤이스팅스 알고리듬에서는 이동 확률(the probability of move)이 식 (8.2.4)와 같이 주어 졌다. 깁스 샘플러는 메트로폴리스-헤이스팅스 알고리듬의 특수한 형태이다. 주기 n 에, X 가 r 개의 원소를 가진 벡터 즉 $X=(X_1,X_2,\ldots,X_i,\ldots,X_r)$라고 하고, X_{-i} 를 다음과 같이 정의하자.

$$X_{-i}=(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_r) \quad 8.3.1$$

즉 X_i 원소가 빠진 벡터를 의미한다. 그리고 일반적으로 다음과 같이 나타낼 수 있다.

$$Y=(Y_i,X_{-i});\ Y_{-i}=X_{-i} \text{ 그리고 } X=(X_i,X_{-i})$$

Y 는 X 벡터에서 X_i 가 갱신된 벡터를 의미한다. 식 (8.2.4)에서 후보 밀도 함수로서 완전 조건부 분포를 이용한다면 다음과 같다.

$$q(Y, X) = p^*(X_i \mid Y_{-i});\ q(X, Y) = p^*(Y_i \mid X_{-i})$$

식 (8.2.4)에 있는 식을 치환하면 $a(x, y)$의 항은 다음과 같이 된다.

$$\begin{aligned} \frac{p^*(Y)}{p^*(X)} \frac{p^*(X_i \mid Y_{-i})}{p^*(Y_i \mid X_{-i})} &= \frac{p^*(Y)}{p^*(X)} \frac{p^*(X_i \mid X_{-i})}{p^*(Y_i \mid X_{-i})} \\ &= \frac{p^*(Y)}{p^*(X)} \frac{p^*(X_i, X_{-i}) p^*(X_{-i})}{p^*(X_{-i}) p^*(Y_i, X_{-i})} = 1 \end{aligned} \qquad 8.3.2$$

깁스 샘플러에서 후보 밀도 함수들은 완전 조건부 분포 $p^*(Y_i \mid X_{-i})$ 에서 추출되며, 항상 수락된다. 즉 항상 $X_i^{n+1} = Y_i$ 이다. 분명하게 표기하면, 깁스 샘플러의 전이 커늘(transition kernel)은 다음과 같다.

$$\prod_{i=1}^{r} p^*\left(X_i^{n+1} \mid X_j^n, j > i, X_j^{n+1}, j < i\right) \qquad 8.3.3$$

깁스 샘플러가 (8.1.2.1)을 만족시키며 완전 조건부 분포로부터 표본 추출을 반복하면, 수렴했을 때 결합 분포로부터의 표본을 생성한다는 것을 증명할 것이다.

$X = (X_1, X_2)$ 인 경우를 예로 들어 보자. 전과 같이 Y_i 를 $X_i (i = 1,2)$의 갱신된 값이라고 하자. 이동은 $X = (X_1, X_2)$ 에서 $Y = (Y_1, Y_2)$ 로 한다. 이동은 두 단계를 거친다. 따라서 전이 커늘은 $P(X, Y) = p(Y_1 \mid X_2) p(Y_2 \mid Y_1)$ 이다. 먼저 $p(Y_1 \mid X_2)$에서 표본 추출을 하고 이어서 $p(Y_2 \mid Y_1)$ 에서 표본

추출을 한다면 Y_1과 Y_2의 결합 분포로부터의 추출을 얻는다는 것을 증명할 것이다. $p(Y_1 | X_2)$에서 그 다음 $p(Y_2 | Y_1)$에서 표본 추출하는 것은 "완전 조건부 분포(fully conditional distribution)"에서 표본 추출을 하는 것과 같다.

깁스 샘플러가 결합 분포로부터의 표본을 생성한다는 것을 증명하기 위하여, 연속적으로 표본 추출할 때 (8.1.2.1)이 만족된다는 것을 증명해야만 한다. 즉 다음과 같다.

$$\int\int p(Y_1 | X_2) p(Y_2 | Y_1) p(X_1, X_2) dX_1 dX_2 = p(Y_1, Y_2) \qquad 8.3.4$$

(적분 기호 안의 처음 두 항은 전이 커늘이다) 우선 X_1에 관하여 적분하고, 다시 X_2에 관하여 적분하면 다음과 같다.

$$\begin{aligned}\int p(Y_1 | X_2) p(Y_2 | Y_1) p(X_2) dX_2 &= \int p(Y_1, X_2) p(Y_2 | Y_1) dX_2 \\ &= p(Y_1) p(Y_2 | Y_1) \\ &= p(Y_1, Y_2)\end{aligned}$$

이것으로 (8.3.4)가 성립한다는 것을 증명했다.

8.3.1 조건부 분포가 결합 분포를 결정

이전 절에서 평형 분포가 존재한다는 가정 하에, 완전 조건부 사후 분포에서 표본 추출을 하면, 수렴했을 때 결합 사후 분포로부터의 추출을 생성한다는 것을 증명했다. 이 결과는 *완전 조건부 분포의 형태가 결합 분포의 형태를 유일하게 결정한다*는 것을 의미한다. 이것에 관한 자세한 점을 알아보기 위하여 X와 Y의 분포 $p(x, y)$를 예로 들어 보자. $p(x, y)$가 양의 조건을 만족한다고 가정하자. 즉 $X = x$와 $Y = y$에 대하여 $p(x) > 0$

이고 $p(y) > 0$ 일 경우, $p(x,y) > 0$ 이다. 이것은 ***기약성**(irreducibility)*에 대한 충분 조건이다. x_0 와 y_0 를 $p(x,y)$ 의 한 점이라고 하자. $p(x,y)$ 의 모든 점에 대하여 다음과 같다는 것을 쉽게 증명할 수 있다.

$$\frac{p(x,y)}{p(x_0,y_0)} = \frac{p(x \mid y_0)p(y \mid x)}{p(x_0 \mid y_0)p(y_0 \mid x)} \tag{8.3.1.1}$$

그러므로 다음과 같이 된다.

$$p(x,y) = p(x_0,y_0)\frac{p(x \mid y_0)p(y \mid x)}{p(x_0 \mid y_0)p(y_0 \mid x)} \tag{8.3.1.2}$$

$p(x \mid y)$ 와 $p(y \mid x)$ 가 모든 X 와 Y 에 대하여 알려져 있다고 가정하면, $p(x,y)$ 를 결정하기 위하여 필요한 것은 $p(x_0,y_0)$ 뿐이다. $p(x,y)$ *가* $\int\int p(x,y)dx\,dy = 1$ *를 만족하는 적절한 확률 밀도 함수라면* 다음과 같다.

$$p(x_0,y_0) = \frac{1}{\int\int \frac{p(x \mid y_0)p(y \mid x)}{p(x_0 \mid y_0)p(y_0 \mid x)}dx\,dy} \tag{8.3.1.3}$$

즉, $p(x_0,y_0)$ 또한 완전 조건부 분포에 의해 결정되며 그러므로 (비록 결합 분포를 명확히 나타내는 것이 불가능 하더라도) 완전 조건부 분포가 어떻게 결합 분포의 형태를 결정하는지 증명했다.

이 결과는 결합 사후 분포가 적절하고 기약성(irreducibility)에 대한 충분 조건인 양수 조건(positivity condition)이 지켜져야 한다는 것이 필요하다. $p(x \mid y)$ *와* $p(y \mid x)$ *가 적절한 분포라도 결합 사후 분포* $p(x,y)$ *가 적절한 분포가 아닐 수 있다.*

8.3.2 깁스 샘플링 알고리듬

모수 벡터 $q = (q_1, q_2, q_3)$를 예로 들어 보자. 사후 분포를 $p(q_1, q_2, q_3 \mid \mathbf{y})$라 하고, 초기값이 (q_1^0, q_2^0, q_3^0)라고 하자. 깁스 샘플러를 적용하는 법은 다음 루프를 반복시키는 것이다.

For i=1 to n

$p(q_1 \mid q_2^{i-1}, q_3^{i-1}, \mathbf{y})$에서 q_1^i를 표본 추출

$p(q_2 \mid q_1^{i}, q_3^{i-1}, \mathbf{y})$에서 q_2^i를 표본 추출

$p(q_3 \mid q_1^{i}, q_2^{i}, \mathbf{y})$에서 q_3^i를 표본 추출

샘플러가 어느 정도 궤도에 오를 때까지 얻은 초기 표본들은 다 버리게 된다. 이런 초기 단계(burn-in period)가 끝난 후 반복수 n을 크게 할 때 얻어지는 세 원소 (q_1^i, q_2^i, q_3^i) 는 정규화된 사후 분포 $p(q_1, q_2, q_3 \mid \mathbf{y}) \big/ \int p(q_1, q_2, q_3 \mid \mathbf{y})\, dq_1 dq_2 dq_3$ 로부터의 추출로 간주한다. 그리고 q_1^i 같은 개개의 원소는 주변 사후 분포 $p(q_1 \mid \mathbf{y}) \big/ \int p(q_1 \mid \mathbf{y}) dq_1$ 로부터의 표본으로 간주한다.

8.4 완전 조건부 사후 분포

8.3 절에서 언급했듯이, 깁스 샘플러를 적용하려면 완전 조건부 분포로부터 추출하는 것이 필요하다. 이 절에서는 어떻게 완전 조건부 분포를 유도하는지 증명할 것이다.

정상 분포(stationary distribution) $p(X)$(별표를 생략하기로 한다)를 하는 모

수 벡터 X 가 있다고 하자. $X_{-i}^{n} = \left(X_1^{n}, \ldots, X_{i-1}^{n}, X_{i+1}^{n-1}, \ldots X_r^{n-1}\right)$를 n 번째 반복에서 $r-1$개 원소를 갖는 벡터라 하자. X_{-i}^{n}의 각 원소는 스칼라 또는 벡터가 될 수 있다. n 번째 반복에서 X_i^n의 완전 조건부 분포는 $p\left(X_i^n \mid X_{-i}^n\right)$으로 표시할 수 있고 다음과 같다.

$$p\left(X_i^n | X_{-i}^n\right) = \frac{p\left(X_i^n, X_{-i}^n\right)}{\int p\left(X_i^n, X_{-i}^n\right) dX_i^n} \tag{8.4.1}$$

완전 조건부 분포는 X_i^n의 함수이고, (8.4.1)의 분모는 X_i^n의 함수가 아니므로 다음과 같이 쓸 수 있다.

$$p\left(X_i^n | X_{-i}^n\right) \propto p\left(X_i^n, X_{-i}^n\right) \tag{8.4.2}$$

그러므로 완전 조건부 분포를 구성하기 위하여 $p\left(X_i^n, X_{-i}^n\right)$ 함수에서 X_i^n를 포함하는 항만을 고르기만 하면 된다. 출발점은 항상 결합 사후 분포이며, 비례값만이라도 알 필요가 있다.

완전 조건부 분포가 표준형을 갖는 경우가 있는데, 이런 경우 깁스 샘플러를 직접 적용할 수 있다. 이 경우가 아닐 때는 다른 샘플링 기법을 적용한다.

예제 8.4.1

간단한 예로, 이변량 정규 분포로부터 나온 단일 관측치 $\mathbf{y} = (y_1, y_2)$로 구성된 자료를 예로 들어 보자. 이것의 평균은 $q = (q_1, q_2)$로서 알려져 있지 않고, 분산(공분산) 행렬은 $\mathbf{V} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$ 로 알려져 있다. 그러므로 $p(\mathbf{y} \mid q) \sim N(q, \mathbf{V})$이고, $p(q \mid \mathbf{y})$에서 표본을 추출하려고 한다. q 에 대한 사전 분포를 '$p(q) \propto$ 상수'라고 하면 q 의 사후 분포는 다음과 같다.

$$p(q \mid \mathbf{y}) \propto p(q)p(\mathbf{y} \mid q) \propto p(\mathbf{y} \mid q) \qquad 8.4.1.1$$

식 (8.4.1.1)이 의미하는 바는 $p(q \mid \mathbf{y})$가 이변량 정규 분포에 비례한다는 것이다. 이 분포는 고정된 자료 $\mathbf{y}$(및 고정된 분산 $\mathbf{V}$, 이 경우 분산이 알려져 있다고 가정했으므로)에 대하여 $q = (q_1, q_2)$의 함수로서 나타나야만 한다. 식 (8.4.1.1)을 q 의 함수로 간주했을 때, $(q \mid \mathbf{y}) \sim N(\mathbf{y}, \mathbf{V})$가 된다. 스칼라 형태에 깁스 샘플러를 적용하기 위하여 $p(q_1 \mid q_2, \mathbf{y})$ 와 $p(q_2 \mid q_1, \mathbf{y})$가 필요하다. 정규 분포의 특성에 비추어 보면 다음과 같다(q_1과 q_2가 결합 이변량 정규 분포라면 조건부 분포 또한 정규 분포이다. - 정의 1.2.1).

$$q_1 \mid q_2, \mathbf{y} \sim N\left(y_1 + r(q_2 - y_2), 1 - r^2\right) \qquad 8.4.1.2$$

$$q_2 \mid q_1, \mathbf{y} \sim N\left(y_2 + r(q_1 - y_1), 1 - r^2\right) \qquad 8.4.1.3$$

깁스 샘플러는 다변량인 경우에도 적용할 수 있다. 이 경우 다음 분포에서 q_1과 q_2를 결합하여 표본 추출한다.

$$q \mid \mathbf{y} \sim N(\mathbf{y}, \mathbf{V}) \qquad 8.4.1.4$$

깁스 샘플러는 식 (8.4.1.2)와 (8.4.1.3)에서 번갈아 표본 추출하며 진행한다. 수 많은 반복 후에 그 계(system)는 정상 분포(stationary distribution)에 수렴

한다. 이 때, 식 (8.4.1.2)로부터 구한 깁스 표본들은 $p(q_1 | \mathbf{y})$에서 추출한 것이고 식 (8.4.1.3)에서 구한 깁스 표본들은 $p(q_2 | \mathbf{y})$에서 추출한 것이다. 이것들은 모수의 주변 사후 분포들이다. 결합하여 취한 표본 (q_1, q_2)(아마 상관되었을 것이다)는 $p(q | \mathbf{y})$에서 표본 추출한 것이다.

깁스 샘플러를 식 (8.4.1.4)에서 동시에 추출하는 것과 같이 이변량에 적용한다면 계(system)는 1 회에 수렴한다. q의 모든 쌍들은 $p(q | \mathbf{y})$로부터 독립적으로 추출한 것이다.

이변량에 적용하면 단변량에 적용한 것보다 정상 분포(stationary distribution)에 보다 빨리 수렴한다. 식 (8.4.1.2)와 식 (8.4.1.3)에서 단변량으로 표본 추출할 때 n번째 표본 쌍 $(q_1, q_2)^{(n)}$은 $n+1$번째 표본 쌍 $(q_1, q_2)^{(n+1)}$과 상관되어 있다. 또한 연속되는 표본과도 상관되어 있다. 이 때문에 천천히 수렴한다.

예제 8.4.2

다음과 같이 모수가 두 개인 베이지안 모형을 생각해 보자.

$$y_i | m, s^2 \sim N(m, s^2), \quad i = 1, \ldots, n \qquad 8.4.2.1$$

$$m \sim N(0,1) \qquad 8.4.2.2$$

$$s^2 | S, n \sim Sc_n^{-2} \qquad 8.4.2.3$$

m의 사전 분포 $p(m)$는 평균이 0 이고 분산이 1 인 정규 분포라고 가정하자. 그리고 분산의 사전 분포 $p(a^2)$은 자유도가 n 인 척도화된 역카이 제곱 분포라고 가정하자(역카이 제곱 분포의 정의에 대해서는 예제 2.1.5 를, 척도화된 역카이 제곱 분포의 정의에 대해서는 예제 2.1.6 을 보라). 모수

S 와 n 는 사전 분포의 모수이고, 그것들은 hyperparameter 들이다. 여기서는 그것들을 알고 있다고 가정하자.

> **역자 주:**
> hyperparameter 에 대한 Daniel Sorensen 의 설명이다. 원문을 그대로 싣는다.
> "When one specifies a prior distribution for a parameter of interest, this prior distribution usually has some parameters itself. These parameters from the prior distribution are known as hyperparameters. These hyperparameters can be assumed known, can be estimated from the data at hand and then used as if known (empirical Bayesian inference) or they can be modelled via hierarchical modelling procedures."

m 와 s^2 가 주어졌을 때 자료가 조건부 독립이라고 가정하면 다음과 같다.

$$p(\mathbf{y} \mid m, s^2) = \prod_{i=1}^{n} (2p\,s^2)^{-\frac{1}{2}} \exp\left[-\frac{(y_i - m)^2}{2s^2}\right]$$

$$= (2p\,s^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(y_i - m)^2}{2s^2}\right] \qquad 8.4.2.4$$

모든 모수의 사후 분포는 다음과 같이 주어진다.

$$p(m, s^2 \mid \mathbf{y}) \propto p(m)p(s^2)p(\mathbf{y} \mid m, s^2) \qquad 8.4.2.5$$

예제 2.1.5 로부터 역카이 제곱 분포는 다음 식에 비례한다는 것을 상기하자.

$$p(c^{-2}) \propto (c^{-2})^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2c^{-2}}\right], \quad c^{-2} > 0$$

s^2 에 대한 척도화된 역카이 제곱 사전 분포는 다음과 같다.

$$p\left(s^2 \mid S,n\right) \propto \exp\left(s^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{nS}{2s^2}\right], \quad s^2 > 0 \qquad 8.4.2.6$$

그러므로 사후 분포 (8.4.2.5)를 다음과 같이 쓸 수 있다(이제부터 S 와 n 에 관한 조건을 없애기로 한다).

$$p\left(m, s^2 \mid \mathbf{y}\right) \propto$$

$$\exp\left[-\frac{m^2}{2}\right]\left(s^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{nS}{2s^2}\right]\left(s^2\right)^{-\frac{n}{2}} \exp\left[-\frac{\sum_i (y_i - m)^2}{2s^2}\right] \qquad 8.4.2.7$$

그리고 이것은 s^2 가 양인 경우만 정의된다(그렇지 않은 경우 $p\left(m, s^2 \mid \mathbf{y}\right) = 0$ 이다).

이제 완전 조건부 분포를 유도하자. $p\left(m \mid s^2, \mathbf{y}\right)$에 대하여 식 (8.4.2.7)로부터 m 를 포함한 항들만을 취한다.

$$\begin{aligned} p\left(m \mid s^2, \mathbf{y}\right) &\propto \exp\left[-\frac{\sum_i (y_i - m)^2}{2s^2}\right] \exp\left[-\frac{m^2}{2}\right] \\ &= \exp\left[-\frac{\sum_i \left[(y_i - \hat{m}) + (\hat{m} - m)^2\right]}{2s^2}\right] \exp\left[-\frac{m^2}{2}\right] \end{aligned} \qquad 8.4.2.8$$

여기서 $\hat{m} = n^{-1} \sum_{i=1}^{n} y_i$ 이다. 첫째 항의 제곱 부분을 전개하면 다음과 같다.

$$\sum_i (y_i - \hat{m})(\hat{m} - m) = (\hat{m} - m)\sum_i (y_i - \hat{m}) = 0$$

그러므로 다음과 같다.

$$p(m \mid s^2, \mathbf{y}) \propto \exp\left[-\frac{s^2 m^2 + n(m - \hat{m})^2}{2s^2}\right] \qquad 8.4.2.9$$

$c = \dfrac{1}{A+B}(Aa + Bb)$일 경우 다음과 같은 항등식이 성립한다.

$$A(z-a)^2 + B(z-b)^2 = (A+B)(z-c)^2 + \frac{AB}{A+B}(a-b)^2$$

s^2를 A, n을 B, m를 z, 0을 a, $\hat{m}$을 b라고 하면 (8.4.2.9)는 다음과 같다.

$$p(m \mid s^2, \mathbf{y}) \propto \exp\left[-\frac{(s^2 + n)(m - c)^2}{2s^2}\right] \qquad 8.4.2.10$$

여기서 $c = \dfrac{n\hat{m}}{s^2 + n}$이다.

이것을 보면 평균이 c이고 분산이 $s^2/(s^2 + n)$인 정규 분포의 커늘 (kernel)임을 알 수 있다. 그러므로 다음과 같다.

$$m \mid s^2, \mathbf{y} \sim N\left(\frac{n\hat{m}}{s^2 + n}, \frac{s^2}{s^2 + n}\right) \qquad 8.4.2.11$$

s^2의 완전 조건부 사후 분포를 구하기 위하여, 식 (8.4.2.7)에서 s^2를 포함하는 항을 선택한다.

$$
\begin{aligned}
p\left(s^2 \mid m, \mathbf{y}\right) &\propto \left(s^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{nS}{2s^2}\right] \left(s^2\right)^{-\frac{n}{2}} \exp\left[-\frac{\sum_i \left(y_i - m\right)^2}{2s^2}\right] \\
&= \left(s^2\right)^{-\left(\frac{n+n}{2}+1\right)} \exp\left[-\frac{nS + \sum_i \left(y_i - m\right)^2}{2s^2}\right] \qquad 8.4.2.12 \\
&= \left(s^2\right)^{-\left(\frac{\tilde{n}}{2}+1\right)} \exp\left[-\frac{\tilde{n}\tilde{S}}{2s^2}\right]
\end{aligned}
$$

여기서 $\tilde{S} = \frac{nS + \sum_i \left(y_i - m\right)^2}{\tilde{n}}$ 이고 $\tilde{n} = n + n$ 이다. 식 (8.4.2.12)는 척도화된 역카이 제곱 분포이다. 이때 자유도는 $\tilde{n}$ 이고 척도 모수는 $nS + \sum_i \left(y_i - m\right)^2$ 이다. 식 (8.4.2.12)로부터 표본을 추출하기 위하여 자유도가 $\tilde{n}$ 인 카이 제곱 분포에서 추출하고 이 수의 역수를 취한 후 $\tilde{n}\tilde{S} = nS + \sum_i \left(y_i - m\right)^2$ 를 곱하면 된다.

이 모형에 깁스 샘플러를 적용하기 위하여 식 (8.4.2.11)로부터 m를 표본 추출하고, 식 (8.4.2.12)에 있는 m를 갱신하고 s^2를 표본 추출한 후, 식 (8.4.2.11)의 s^2를 갱신하고 m를 다시 추출하는 것을 반복한다. 수렴했을 때, 표본 m^i (i =1,…,깁스 표본의 수)를 주변 사후 분포 $p(m \mid \mathbf{y})$에서 나온 표본으로 간주하고 표본 $\left(s^2\right)^i$를 주변 사후 분포 $p\left(s^2 \mid \mathbf{y}\right)$에서 나온 표

본으로 간주한다.

예제 8.4.3

평균이 $m=(m_1, m_2)$이고 분산이 $\mathbf{V}=\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$인 이변량 정규 분포에서 나온 자료 $\mathbf{y}_i=(y_{1i}, y_{2i})$ ($i=1,\ldots,n$)를 예로 들어 보자. 관측치들의 조건부 분포는 다음과 같다.

$$\begin{aligned} p(\mathbf{y} \mid m, \mathbf{V}) &= |2p\mathbf{V}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i - m)' \mathbf{V}^{-1}(y_i - m)\right] \\ &= |2p|^{-n}|\mathbf{V}|^{-\frac{n}{2}}\left[-\frac{1}{2}tr(\mathbf{V}^{-1}\mathbf{S})\right] \end{aligned} \qquad 8.4.3.1$$

여기서 $\mathbf{S}=\begin{bmatrix} \sum_i (y_{1i}-m_1)^2 & \sum_i (y_{1i}-m_1)(y_{2i}-m_2) \\ \sum_i (y_{1i}-m_1)(y_{2i}-m_2) & \sum_i (y_{2i}-m_2)^2 \end{bmatrix}$이다.

(예제 1.2.4 참조)

m의 사전 분포가 균일 분포이고 $\mathbf{V}$의 사전 분포가 다음과 같은 역 위샤트 분포라고 가정하자(정의 1.2.2).

$$p(\mathbf{V} \mid \mathbf{V}_0, n_0) \propto |\mathbf{V}|^{-\frac{1}{2}(n_0+3)} \exp\left[-\frac{1}{2}tr(\mathbf{V}^{-1}\mathbf{V}_0^{-1})\right] \qquad 8.4.3.2$$

여기서 $\mathbf{V}_0$ 와 n_0 는 hyperparameter 이며 알려져 있다고 가정한다. Hyperparameter $\mathbf{V}_0$는 $(n_0-3)\mathbf{V}_0^{-1}=\tilde{E}(\mathbf{V} \mid \mathbf{V}_0, n_0)$를 만족하는 것으로 선택

한다. 여기서 $\tilde{E}(\mathbf{V} \mid \mathbf{V}_0, n_0)$ 는 과거 정보에 기초하여 $\mathbf{V}$ 의 선택된 값이다 (정의 1.2.2 참조). 그러면 결합 사후 분포는 다음과 같다($\mathbf{V}_0$ 와 n_0 에 관한 표기는 생략한다).

$$p(m, \mathbf{V} \mid \mathbf{y}) \propto p(\mathbf{y} \mid m, \mathbf{V}) p(\mathbf{V})$$
$$= |2p|^{-n} |\mathbf{V}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} tr\left(\mathbf{V}^{-1} S\right)\right] |\mathbf{V}|^{-\frac{1}{2}(n_0+3)} \exp\left[-\frac{1}{2} tr\left(\mathbf{V}^{-1} \mathbf{V}_0^{-1}\right)\right] \qquad 8.4.3.3$$

m 의 조건부 사후 분포를 유도하기 위하여 m 를 포함하고 있는 결합 사후 분포의 항을 선택한다. 그러면 다음과 같다.

$$p(m \mid \mathbf{V}, \mathbf{y}) \propto \exp\left[-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{y}_i - m)' \mathbf{V}^{-1} (\mathbf{y}_i - m)\right]$$
$$= \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left[(\mathbf{y}_i - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - m)\right]' \mathbf{V}^{-1} \left[(\mathbf{y}_i - \bar{\mathbf{y}}) + (\bar{\mathbf{y}} - m)\right]\right]$$
$$\propto \exp\left[-\frac{1}{2} \sum_{i=1}^{n} (\bar{\mathbf{y}} - m)' \mathbf{V}^{-1} (\bar{\mathbf{y}} - m)\right]$$
$$= \exp\left[-\frac{1}{2} n (\bar{\mathbf{y}} - m)' \mathbf{V}^{-1} (\bar{\mathbf{y}} - m)\right]$$

여기서 $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2) = \left(n^{-1} \sum_{i=1}^{n} y_{1i}, n^{-1} \sum_{i=1}^{n} y_{2i}\right)$ 이다. 결국 다음과 같다.

$$m \mid \mathbf{V}, \mathbf{y} \sim N\left(\bar{\mathbf{y}}, n^{-1} \mathbf{V}\right) \qquad 8.4.3.4$$

정규 분포의 특성을 이용하여, $m_1 \mid m_2, \mathbf{V}, \mathbf{y}$ 와 $m_2 \mid m_1, \mathbf{V}, \mathbf{y}$ 의 형태를 쉽게 유도할 수 있다.

식 (8.4.3.3)으로부터 구한 $\mathbf{V}$의 조건부 사후 분포는 다음과 같다.

$$p(\mathbf{V} \mid m, \mathbf{y}) \propto |\mathbf{V}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} tr\left(\mathbf{V}^{-1} S\right)\right] |\mathbf{V}|^{-\frac{1}{2}(n_0+3)} \exp\left[-\frac{1}{2} tr\left(\mathbf{V}^{-1}\mathbf{V}_0^{-1}\right)\right]$$
$$= |\mathbf{V}|^{-\frac{1}{2}(n_0+3+n)} \exp\left[-\frac{1}{2} tr\left[\mathbf{V}^{-1}\left(S + \mathbf{V}_0^{-1}\right)\right]\right]$$

이것은 척도 행렬이 $\left(S + \mathbf{V}_0^{-1}\right)^{-1}$이고 자유도가 $n_0 + n$인 역위샤트 분포임을 알 수 있다. 즉 다음과 같다.

$$\mathbf{V} \mid m, \mathbf{y} \sim IW_2\left(\left(S + \mathbf{V}_0^{-1}\right)^{-1}, n_0 + n\right) \qquad 8.4.3.5$$

이산적인 상황에서 깁스 샘플러를 적용하는 것과 Reducible Markov Chain 의 경우를 예로 들어 설명하겠다.

> 역자 주:
> reducible Markov chain 에 대한 Daniel Sorensen 의 설명이다. 원문을 그대로 싣는다.
> “A reducible Markov chain is a chain where not all state-spaces communicate. This has the consequence that if we generate such a reducible chain, the chain may converge to either one or another subspace of the posterior distribution. The Gibbs sampler can be viewed as a Monte Carlo generated Markov chain; the draws obtained from reducible Markov chains will only represent a subspace of the entire distribution, and inferences - which should be based on samples from the WHOLE posterior distribution, would be faulty.”

예제 8.4.4

아비(f), 어미(m) 및 자손으로 구성된 간단한 가계를 예로 들자. 자료는 한 자손의 유전자형으로 구성되어 있다고 가정하자. 단일 유전자좌에 두 개의 대립 유전자(A , a)가 있고, y (자손의 유전자형)로 표시한 자료는 Aa라고 가정한다. 또한 A의 빈도는 0.5 이고 집단은 Hardy-Weinberg 평형

상태에 있다고 가정한다. 자손의 유전자형이 주어졌을 때, 아비의 유전자형에 대한 확률 분포를 알고 싶다.

유전자 빈도가 알려져 있고 Hardy-Weinberg 평형 상태에 있다는 가정에 비추어 보면, 아비의 유전자형에 대한 확률 분포는 바로 쓸 수 있다(아비의 유전자형은 AA, Aa, aa 이고 확률은 각각 0.25, 0.5, 0.25 이다). 그러나 형식적으로 확률 분포를 구하고, 깁스 샘플러를 이용하여 평형 분포를 어떻게 구하는지 보이고, 마지막으로 마르코프 연쇄식에서의 문제를 공식화하고 평형 분포를 구할 것이다.

다음 표기법을 필요할 때 이용할 것이다. 유전자형 AA, Aa, aa 에 대한 값을 x_1, x_2, x_3 로 하는 임의 변량을 X 로 하자. 주어진 자료와 유전자 빈도하에서 부모 유전자형의 결합 분포(즉, 결합 사후 분포)를 다음과 같이 베이즈 이론을 이용하여 구할 수 있다(표기법상 유전자 빈도에 관한 조건은 생략한다).

$$P(f = x_i, m = x_j \mid y) = \frac{P(f = x_i, m = x_j)P(y \mid f = x_i, m = x_j)}{\sum_{k,l} P(f = x_k, m = x_l)P(y \mid f = x_k, m = x_l)} \quad 8.4.4.1$$

부모는 상관이 없는 개체로 간주하므로 다음과 같다.

$$P(f = x_i, m = x_j) = P(f = x_i)P(m = x_j)$$

약간만 계산을 하면 식 (8.4.4.1)의 분모가 0.5 가 되는 것을 증명할 수 있다(주어진 y 하에서 7 가지의 교배 형태가 가능한데 (아비, 어미)로 표시할 경우 (AA, Aa), (AA, aa), (Aa, AA), (Aa, Aa), (Aa, aa), (aa, AA), (aa, Aa) 등이다). 예를 들어 y 와 유전자 빈도가 주어졌을 때 $f = AA$ 이고 $m = Aa$ 인 결합 사후 확률은 다음과 같다.

$$P(f=x_1, m=x_2 \mid y) = \frac{P(f=x_1)P(m=x_2)P(y \mid f=x_1, m=x_2)}{\sum_{i,j} P(f=x_i, m=x_j)P(y \mid f=x_i, m=x_j)}$$
$$= \frac{(1/4)(1/2)(1/2)}{(1/2)} = \frac{1}{8}$$

마찬가지로 $P(f=x_1, m=x_1 \mid y)=0$ 이고 $P(f=x_1, m=x_3 \mid y)=1/8$ 이다. 마지막으로 이러한 결합 사후 확률로부터 (유전자 빈도에 조건적인) 아비 유전자형의 주변 사후 확률을 계산한다. 예를 들어 앞서 계산한 것을 이용하여 다음과 같이 계산한다.

$$P(f=x_1 \mid y) = \sum_{i=1}^{i=3} P(f=x_1, m=x_i \mid y)$$
$$= 0 + \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

두 가지 다른 주변 사후 확률도 구할 수 있다.

$$P(f=x_2 \mid y) = \frac{1}{2} \text{ 그리고 } P(f=x_3 \mid y) = \frac{1}{4}$$

한 번에 하나씩 갱신하는 방법으로 깁스 샘플러를 적용하기 위하여 조건부 사후 분포로부터 추출한다. 이들 조건부 사후 분포는 결합 사후 분포에 비례한다(8.4.2 참조). 예를 들어 보자.

$$P(f=AA \mid m=AA, y) \propto P(f=AA)P(y \mid f=AA, m=AA) = \frac{1}{4}0 = 0$$
$$P(f=Aa \mid m=AA, y) \propto P(f=Aa)P(y \mid f=Aa, m=AA) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

$$P(f=aa\,|\,m=AA,y)\propto P(f=aa)P(y\,|\,f=aa,m=AA)=\frac{1}{4}1=\frac{1}{4}$$

척도화(scaling) 시킨 후 이들 확률은 다음과 같아 진다.

$P(f=Aa\,|\,m=AA,y)=1/2$ 그리고 $P(f=aa\,|\,m=AA,y)=1/2$

다음과 같이 추출을 시작할 수도 있다.

$$P(f=AA\,|\,m=Aa,y)=1/4$$
$$P(f=Aa\,|\,m=Aa,y)=1/2$$
$$P(f=aa\,|\,m=Aa,y)=1/4$$

이 분포로부터 표본 추출하기 위하여, 균일 분포 $Un(0,1)$에서 추출한다. 아비의 결과적인 유전자형은 이 추출의 결과가 0.25($f=AA$)이하인가, 0.25 와 0.75($f=Aa$) 사이에 있는가, 0.75($f=aa$)를 초과하는가에 달려 있다. 아비의 유전자형으로 $f=aa$ 를 추출하였다면 적절한 조건부 사후 분포로부터 추출한다.

$$P(m=AA\,|\,f=aa,y)=1/2$$
$$P(m=Aa\,|\,f=aa,y)=1/2$$

표본 추출하고 갱신하는 이 과정을 계속한다. Burn-in 기간을 지난 후, 각 주기내의 표본은 $p(f=x_i,m=x_j\,|\,y)$으로부터 추출한 것이고 그리고 주변 분포로 보면 $p(f=x_i\,|\,y)$와 $p(m=x_i\,|\,y)$으로부터 추출한 것이다. 추출된 유전자형을 세어 보면 추정된 주변 사후 빈도들이 아비와 어미 모두 유전자형 AA, Aa, aa에 대하여 각각 0.25, 0.5, 0.25에 근접하게 된다.

마르코프 연쇄에 관한 문제점을 공식화하기 위하여, 먼저 확률적 진행 W

를 정의하자. 이것은 r 상태의 어느 것($w_1, \ldots, w_r$)도 가정할 수 있다. 이 진행에 대한 전이 행렬 $\mathbf{P} = \{P(i, j)\}$은 원소 $P(i, j) = P(W = w_j \mid W = w_i)$ 를 갖는다. 이 원소는 i 번째 상태에 있을 때 j 번째 상태로 이동할 확률이다. 각 상태들의 확률을 평형에 도달하게 하는 r 차원의 벡터 p^* 를 찾아내고자 한다. 이번 예에서는 $r = 7$ 이고 상태 벡터의 원소들은 7 가지 가능한 교배 형태(주어진 y)들이다. 즉(아비, 어미)로 표시할 경우 (AA, Aa), (AA, aa), (Aa, AA), (Aa, Aa), (Aa, aa), (aa, AA), (aa, Aa) 등이다. 이들 가능한 교배 형태들을 $w_i(i = 1, \ldots, 7)$로 표시할 수 있다. 이제 위에서 적용했던 깁스 샘플러를 위한 전이 확률을 정의하자. 샘플러의 이런 적용은 한 번에 한 부모를 갱신하여 다음 상태로 이동하면서 마르코프 연쇄를 생성한다. 예를 들어 $\left(f = x_1^{(0)}, m = x_2^{(0)}\right)$에서 $\left(f = x_1^{(1)}, m = x_3^{(1)}\right)$로의 이동을 생각해 보자. 여기서 위 첨자는 표본 추출 주기를 의미한다(모든 확률들은 자료 y 에 관하여 조건적이다). 한 번에 하나씩 갱신하는 방식에 따라 다음과 같이 된다.

$$
\begin{aligned}
P\left(w_2^{(1)} \mid w_1^{(0)}\right) &= P\left(f = x_1^{(1)}, m = x_3^{(1)} \mid f = x_1^{(0)}, m = x_2^{(0)}\right) \\
&= P\left(f = x_1^{(1)} \mid m = x_2^{(0)}\right) P\left(m = x_3^{(1)} \mid f = x_1^{(1)}\right) \\
&= \frac{1}{4}\frac{1}{2} = \frac{1}{8}
\end{aligned}
$$

여기서 우변의 항들은 식 (8.4.4.1)에서 구하고 적절한 주변 확률 분포로 나눈 것이다. 마르코프 연쇄인 7 × 7 차수의 전이 확률 행렬 P 는 다음과 같다(상태 n 에서의 교배 형태가 행에 있고, 상태 $n+1$ 의 교배 형태가 열에 있다. 각 원소는 상태 n 의 교배 형태가 상태 $n+1$ 의 교배 형태가 될 확률이다).

$$
\begin{matrix}
w_1 = AA, Aa \\
w_2 = AA, aa \\
w_3 = Aa, AA \\
w_4 = Aa, Aa \\
w_5 = Aa, aa \\
w_6 = aa, AA \\
w_7 = aa, Aa
\end{matrix}
\begin{bmatrix}
1/8 & 1/8 & 1/8 & 1/4 & 1/8 & 1/8 & 1/8 \\
1/4 & 1/4 & 1/8 & 1/4 & 1/8 & 0 & 0 \\
0 & 0 & 1/8 & 1/4 & 1/8 & 1/4 & 1/4 \\
1/8 & 1/8 & 1/8 & 1/4 & 1/8 & 1/8 & 1/8 \\
1/4 & 1/4 & 1/8 & 1/4 & 1/8 & 0 & 0 \\
0 & 0 & 1/8 & 1/4 & 1/8 & 1/4 & 1/4 \\
1/8 & 1/8 & 1/8 & 1/4 & 1/8 & 1/8 & 1/8
\end{bmatrix}
$$

원소 $P(i, j)$를 계산하는 법을 알아 보기 위하여 $P(1,1)$을 예로 들어 보자.

$$
\begin{aligned}
&P\left(f = x_1^{(1)}, m = x_2^{(1)} \mid f = x_1^{(0)}, m = x_2^{(0)}\right) \\
&\quad = P\left(f = x_1^{(1)} \mid m = x_2^{(0)}\right) P\left(m = x_2^{(1)} \mid f = x_1^{(1)}\right) \\
&\quad = \frac{1}{4}\frac{1}{2} = \frac{1}{8}
\end{aligned}
$$

0 단계에서 마르코프 진행의 확률 분포를 $p(W_0)$이라 하자. 예를 들어 다음과 같은 형태를 취할 수 있다.

$$p(W_0) = (0 \quad 1/4 \quad 0 \quad 1/4 \quad 0 \quad 1/2 \quad 0)$$

이 초기 구성을 이용하여 1 단계에서는 다음을 얻을 수 있다.

$$p(W_1) = p(W_0)\mathbf{P} = (0.094 \quad 0.094 \quad 0.125 \quad 0.250 \quad 0.125 \quad 0.156 \quad 0.156)$$

6 단계에서는 다음과 같은 형태를 갖는 평형 분포에 도달하게 된다.

$$p^* = (1/8 \quad 1/8 \quad 1/8 \quad 1/4 \quad 1/8 \quad 1/8 \quad 1/8)$$

이 분포로부터 쉽게 아비 유전자형 AA, Aa, aa의 비율을 쉽게 계산할 수 있다. 계산해 보면 각각 0.25, 0.5 및 0.25 이다. 이 연쇄가 가역성 조건 (8.5)

를 만족시키며 연쇄가 irreducible 하다는 것을 쉽게 증명할 수 있다. 그러므로 평형 분포 p^* 는 유일하다.

예제 8.4.5

이번 예제에서는 마르코프 연쇄의 reducibility 경우를 예로 들어 보자. 한 유전자좌에 세 개의 대립 유전자가 있는 경우를 생각해 보자. 혈액형의 유전자형이 좋은 예가 될 것이다. 자료 $\mathbf{y}$ 가 두 자손의 유전자형 $\mathbf{y}=(AB,OO)$ 으로 구성되어 있는 경우를 생각해 보자. 부모 유전자형의 확률 분포를 추론하고자 한다면, 이 경우 확률 분포를 다음과 같이 직접 구할 수도 있다.

$$\begin{aligned}\Pr(f=AO, m=BO\mid\mathbf{y})&=1/2\\ \Pr(f=BO, m=AO\mid\mathbf{y})&=1/2\end{aligned} \qquad 8.4.5.1$$

그럼에도 불구하고 깁스 샘플러를 이용하여 이 분포의 몬테 카를로 추정치를 구하고자 한다. 더구나 샘플러는 완전 조건부 사후 분포로부터 한 번에 한 유전자형을 추출하며 스칼라적으로 적용될 수 있다. 초기 구성이 $m=BO$ 라고 하자. 다음 확률 분포로부터 추출하고 싶다.

$$\Pr(f=AO\mid m=BO,\mathbf{y}) \qquad 8.4.5.2$$

$$\Pr(f=BO\mid m=BO,\mathbf{y}) \qquad 8.4.5.3$$

그러나 (8.4.5.2)와 관련된 확률이 1 이고 (8.4.5.3)과 관련된 확률은 0 이라는 것에 주의하자. 마찬가지로 다른 적절한 구성으로 시작한다고 해도, 결과적인 확률은 다음과 같이 1 아니면 0 이다.

$$\begin{aligned}\Pr(f=BO\mid m=AO,\mathbf{y})&=1\\ \Pr(f=AO\mid m=AO,\mathbf{y})&=0\end{aligned}$$

어떤 적절한 상태에서 시작하더라도, 이 상태를 결코 벗어날 수가 없다. 이것이 *주어진 자료하에서 샘플러를 적용하더라도* 연쇄 상태가 변하지 않는 경우이다. 결과적인 마르코프 연쇄가 reducible 한 것이다. 이번 예는 결합 분포 (8.4.5.1)로부터 표본 추출하여 문제를 해결할 수 있다. 이 결합 표본 추출은 부모 유전자형의 결합 사후 분포의 몬테 카를로 실현(realization)이 된다.

덧붙여 말하자면 이 예제에 대한 전이 확률 행렬은 다음과 같다.

$$\mathbf{P} = \begin{matrix} AO \\ BO \end{matrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$\mathbf{P}$는 reducible 마르코프 연쇄를 생성한다. 어떤 n에 대해서도 $\mathbf{P}^n$은 0 원소가 있는 행렬을 생성한다. 이것은 각 상태들이 communication 하지 않는다는 것을 증명한다.