Probability and Stochastic Processes II - Lecture 4c

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC632024.html

2024

IV.4 Markov Chain Monte Carlo

- suppose it is required to approximate the following, where $h : S \to \mathbb{R}^1$, π is a pdf on S and μ is counting measure,

$$E_\pi(h) = \int_S h(x)\pi(x)\mu(dx) = \sum_{i \in S} h(i)\pi(i)$$

- if we can generate a sample from π, say $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \pi$, then we can estimate $E_\pi(h)$ by

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k)$$

- if we can't generate from π but have a good importance sampler g that we can sample from, then generate $X_1, \ldots, X_n \overset{i.i.d.}{\sim} g$ and estimate $E_\pi(h)$ by

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k)\frac{\pi(X_k)}{g(X_k)}$$

and note that this is the first case when $g = \pi$ and also recall that we discussed how to assess the error in this estimate

- but what if we can't generate from π and can't find a good importance sampler

- another possibility is to construct an irreducible, aperiodic MC $\{X_n : n \in \mathbb{N}_0\}$ with state space S that has stationary distribution π

- then, under conditions

$$\frac{1}{n}\sum_{k=1}^{n} h(X_k) \to E_\pi(h)$$

- intuitively

$$E_{i_0}(h(X_k)) = \sum_{i\in S} h(i)P_{i_0}(X_k = i) \approx \sum_{i\in S} h(i)\pi(i) \text{ for large } k \text{ by MCCT}$$

- given π (possibly unnormalized), how do we construct an irreducible, aperiodic MC $\{X_n : n \in \mathbb{N}_0\}$ with stationary distribution π

Metropolis Algorithm

- for $S \subset \mathbb{Z}$ consisting of consecutive integers (could be finite or infinite) with $\pi_i > 0$ for every $i \in S$

- define transition probabilities as follows

$$\begin{aligned} p_{ii+1} &= \frac{1}{2}\min\left\{1, \frac{\pi_{i+1}}{\pi_i}\right\} \\ p_{ii-1} &= \frac{1}{2}\min\left\{1, \frac{\pi_{i-1}}{\pi_i}\right\} \\ p_{ii} &= 1 - p_{ii+1} - p_{ii-1} \\ p_{ij} &= 0 \text{ otherwise} \end{aligned}$$

so just three possibilities, the state increases or decreases by one unit or stays the same

- **note** - the distribution π_i need not be normalized for this algorithm since $c\pi_i / c\pi_j = \pi_i / \pi_j$

- as an algorithm: at step n, so we know X_{n-1}

1. generate $U_n \sim U(0,1)$ and independently generate $Y_n \in \{X_{n-1}-1, X_{n-1}+1\}$ with prob. 1/2 each

2. put
$$X_n = \begin{cases} Y_n & \text{if } U_n \le \frac{\pi_{Y_n}}{\pi_{X_{n-1}}} \text{ (accept new state)} \\ X_{n-1} & \text{otherwise (reject new state)} \end{cases}$$

- note that, if $\pi_{Y_n}/\pi_{X_{n-1}} > 1$, then the new state is accepted

Proposition IV.12 *(MCMC Convergence Theorem)* The Metropolis algorithm produces an irreducible, aperiodic MC with stationary dist. π.

Proof: Clearly $\{X_n : n \in \mathbb{N}_0\}$ is a MC for any initial distribution ν. Also there is positive probability of the transition $i \to i+1$ when $i, i+1 \in S$ and $i \to i-1$ when $i, i-1 \in S$ and this implies the chain is irreducible. Now

$$\begin{aligned}
\pi_i p_{ij} &= \pi_j p_{ji} = 0 \text{ when } |i-j| > 2 \\
\pi_i p_{ii+1} &= \pi_i \frac{1}{2} \min\left\{1, \frac{\pi_{i+1}}{\pi_i}\right\} = \frac{1}{2} \min\{\pi_i, \pi_{i+1}\} \\
&= \pi_{i+1} \frac{1}{2} \min\left\{1, \frac{\pi_i}{\pi_{i+1}}\right\} = \pi_{i+1} p_{i+1i} \\
\pi_i p_{ii-1} &= \pi_i \frac{1}{2} \min\left\{1, \frac{\pi_{i-1}}{\pi_i}\right\} = \frac{1}{2} \min\{\pi_i, \pi_{i-1}\} \\
&= \pi_{i-1} \frac{1}{2} \min\left\{1, \frac{\pi_i}{\pi_{i-1}}\right\} = \pi_{i-1} p_{i-1i}
\end{aligned}$$

and so the detailed balance equations are satisfied so the chain is time reversible. This implies that π is a stationary distribution for the chain which implies that that the chain is recurrent and so every state has the same period. If S is infinite, then there exists i_0 s.t. $\pi_{i_0} > \pi_{i_0+1}$ or $\pi_{i_0} > \pi_{i_0-1}$ and so $p_{i_0 i_0} > 0$ and when S is finite, then for i_0 a boundary state we have $p_{i_0 i_0} > 0$. This implies that $p_{i_0 i_0} > 0$ so state i_0 has period 1 and, since the chain is recurrent, all states have period 1 which implies the chain is aperiodic. ■

Corollary IV.13 (i) $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ for every $i, j \in \mathcal{S}$, and (ii) $\lim_{n\to\infty} P(X_n = j) = \pi_j$ for any initial distribution ν.

Proof: MCCT. ■

Example IV.10

- $S = \{0, 1, \ldots, 10\}$ and $\pi_i \propto \cos^2(i)\binom{10}{i}\left(\frac{1}{2}\right)^i$ then with r denoting the binomial$(10, 1/2)$ pdf

$$
\begin{aligned}
p_{1011} &= 0 \\
p_{ii+1} &= \frac{1}{2}\min\left\{1, \frac{\cos^2(i+1)r(i+1)}{\cos^2(i)r(i)}\right\} \text{ for } i \in \{0, \ldots, 9\} \\
p_{ii-1} &= \frac{1}{2}\min\left\{1, \frac{\cos^2(i-1)r(i-1)}{\cos^2(i)r(i)}\right\} \text{ for } i \in \{1, \ldots, 10\} \\
p_{0-1} &= 0
\end{aligned}
$$

```
# plot the unnormalized density
x=c(0:10)
p=((cos(x))**2)*dbinom(x,10,1/2)
plot(x,p,type="l")
```

```
#normalize and get table of probabilities
norm=sum(p)
probx=p/norm
tab=rbind(x,probx)
t(tab)
       x       probx
 [1,] 0 0.001956606
 [2,] 1 0.005711852
 [3,] 2 0.015247865
 [4,] 3 0.230116825
 [5,] 4 0.175551551
 [6,] 5 0.039674070
 [7,] 6 0.378807998
 [8,] 7 0.133448793
 [9,] 8 0.001863983
[10,] 9 0.016242926
[11,] 10 0.001377531
```

```
# generate the Metropolis chain in w
n=100000
w=c(0:n)
# starting value
w[1]=5
wcurrent=w[1]
for (i in 2:(n+1)){
v=rbinom(1,1,1/2)
if(v==0){y=wcurrent-1}
else{y=wcurrent+1}
w[i]=wcurrent
if((y != -1) & (y != 11)){
u=runif(1,0,1)
pdfnum=((cos(y))**2)*dbinom(y,10,1/2)
pdfdenom=((cos(wcurrent))**2)*dbinom(wcurrent,10,1/2)
if(u <= pdfnum/pdfdenom){w[i]=y}
wcurrent=w[i]
}
}
```

```
# record proportion of times each state is visited by the
chain
prop=rep(0,11)
for(i in 1:11){
for (j in 1:n){
if(w[j] == (i-1)){
prop[i]=prop[i]+1}
}
}
prop=prop/n
tab2=rbind(c(0:10),prop)
cat("Chain length = ",n,"\n")
t(tab2)
```

```
Chain length = 100
> t(tab2)
 prop
 [1,] 0 0.00
 [2,] 1 0.00
 [3,] 2 0.00
 [4,] 3 0.17
 [5,] 4 0.16
 [6,] 5 0.07
 [7,] 6 0.45
 [8,] 7 0.15
 [9,] 8 0.00
[10,] 9 0.00
[11,] 10 0.00
```

```
Chain length = 10000
> t(tab2)
 prop
 [1,] 0 0.0005
 [2,] 1 0.0046
 [3,] 2 0.0153
 [4,] 3 0.2139
 [5,] 4 0.1577
 [6,] 5 0.0387
 [7,] 6 0.3749
 [8,] 7 0.1414
 [9,] 8 0.0034
[10,] 9 0.0467
[11,] 10 0.0029
```

```
Chain length = 1e+05
> t(tab2)
 prop
 [1,] 0 0.00171
 [2,] 1 0.00564
 [3,] 2 0.01461
 [4,] 3 0.22386
 [5,] 4 0.17243
 [6,] 5 0.03889
 [7,] 6 0.38075
 [8,] 7 0.13807
 [9,] 8 0.00200
[10,] 9 0.02033
[11,] 10 0.00171
```
■

- on disadvantage with MCMC is that the length of the chain needs to be large enough to ensure stationarity to be achieved as well as a desirable accuracy of the approximation

- also error analysis is more complicated because the estimate of $E_\pi(h)$, namely, $\frac{1}{n}\sum_{k=1}^{n} h(X_k)$ has correlated summands (discussed in a time series course)

- there are a number of MCMC algorithms related to Metropolis

Metropolis-Hastings

- for $S \subset \mathbb{Z}$ consisting of consecutive integers (could be finite or infinite) with $\pi_i > 0$ for every $i \in S$

- let $q : S \times S \to (0, \infty)$ be s.t. $\sum_{j \in S} q(i,j) = 1$ for all $i \in S$ and $q(i,j) > 0$ iff $q(j,i) > 0$ called a *proposal distribution*

- define transition probabilities

$$p_{ij} = \begin{cases} q(i,j) \min\left\{1, \frac{\pi_j q(j,i)}{\pi_i q(i,j)}\right\} & \text{when } i \neq j \\ 1 - \sum_{k \neq i} p_{ik} & \text{when } i = j \end{cases}$$

- clearly $0 \leq p_{ij} \leq 1$ when $i \neq j$ and note that

$$\begin{aligned} \sum_{k \neq i} p_{ik} &= \sum_{k \neq i} q(i,k) \min\left\{1, \frac{\pi_k q(k,i)}{\pi_i q(i,k)}\right\} \\ &= \frac{1}{\pi_i} \sum_{k \neq i} \min\left\{\pi_i q(i,k), \pi_k q(k,i)\right\} \leq \frac{1}{\pi_i} \sum_{k \neq i} \pi_i q(i,k) \\ &= \sum_{k \neq i} q(i,k) \leq 1 \end{aligned}$$

and so the p_{ij} define a valid transition matrix

- also when $i \neq j$

$$\begin{aligned}\pi_i p_{ij} &= \pi_i q(i,j) \min\left\{1, \frac{\pi_j q(j,i)}{\pi_i q(i,j)}\right\} = \min\left\{\pi_i q(i,j), \pi_j q(j,i)\right\} \\ &= \pi_j q(j,i) \min\left\{1, \frac{\pi_i q(i,j)}{\pi_{ij} q(j,i)}\right\} = \pi_j p_{ji}\end{aligned}$$

and so the chain is time reversible with stationary distribution π

- if $q(i,j) > 0$ for all i,j, then the chain is irreducible and also (as above)

$$\sum_{k \neq i} p_{ik} \leq \sum_{k \neq i} q(i,k) < 1$$

since $q(i,i) > 0$ which implies $p_{ii} > 0$ and so the chain is aperiodic

- therefore by MCCT (i) $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$ for every $i, j \in \mathcal{S}$, and (ii) $\lim_{n \to \infty} P(X_n = j) = \pi_j$ for any initial distribution ν.

- the Metropolis-Hastings algorithm: at step n, so we know X_{n-1}

1. generate $U_n \sim U(0,1)$ and independently generate $Y_n \sim q(X_{n-1}, \cdot)$
2. put

$$X_n = \begin{cases} Y_n & \text{if } U_n \leq \frac{\pi_{Y_n} q(Y_n, X_{n-1})}{\pi_{X_{n-1}} q(X_{n-1}, Y_n)} \text{ (accept new state)} \\ X_{n-1} & \text{otherwise (reject new state)} \end{cases}$$

- note that, if $\pi_{Y_n} q(Y_n, X_{n-1}) / \pi_{X_{n-1}} q(X_{n-1}, Y_n) > 1$, then the new state is accepted

- note that for the standard Metropolis algorithm $q(i,j) = 1/2$ when $|i-j| = 1$ and is 0 otherwise
■

Gibbs Sampling

- suppose $f : S \times S \to (0, \infty)$ is such that $c = \sum_{(i,j) \in S \times S} f(i,j) < \infty$ so $g = c^{-1} f$ is a pdf on $S \times S$

- then $g_1(i) = c^{-1} \sum_{j \in S} f(i,j)$ is the marginal pdf of the 1st coordinate and $g_2(j) = c^{-1} \sum_{i \in S} f(i,j)$ is the marginal pdf of the 2nd coordinate

- also

$$\begin{aligned} g_1(i \,|\, j) &= \frac{g(i,j)}{g_2(j)} = \text{conditional pdf of 1st coordinate given 2nd} \\ g_2(j \,|\, i) &= \frac{g(i,j)}{g_1(i)} = \text{conditional pdf of 2ndt coordinate given 1st} \end{aligned}$$

and note both are independent of c

- suppose that it is easy to generate from $g_1(\cdot \,|\, j)$ and $g_2(\cdot \,|\, i)$ but hard to generate from g (1-dimensional as opposed 2-dimensional generating problem)

- Gibbs sampling algorithm for n-th step given (X_{n-1}, Y_{n-1})

1. generate $V \sim$ binomial$(1, 1/2)$
2. if $V = 0$ put $Y_n = Y_{n-1}$ and generate $X_n \sim g_1(\cdot \mid Y_{n-1})$ and if $V = 1$ put $X_n = X_{n-1}$ and generate $Y_n \sim g_2(\cdot \mid X_{n-1})$ (possibly changing one coordinate at each step)

- this is a 2-dimensional MC with transition probabilities

$$p_{(i,j)(k,l)} = \begin{cases} \frac{1}{2}g_1(i \mid j) + \frac{1}{2}g_2(j \mid i) & i = k, j = l \\ \frac{1}{2}g_1(k \mid j) & i \neq k, j = l \\ \frac{1}{2}g_2(l \mid i) & i = k, j \neq l \\ 0 & \text{otherwise} \end{cases}$$

- we have

$$\begin{aligned} \sum_{(k,l)} p_{(i,j)(k,l)} &= p_{(i,j)(i,j)} + \sum_{k \neq i} p_{(i,j)(k,j)} + \sum_{l \neq j} p_{(i,j)(i,l)} \\ &= \frac{1}{2}g_1(i \mid j) + \frac{1}{2}g_2(j \mid i) + \frac{1}{2}\sum_{k \neq i} g_1(k \mid j) + \frac{1}{2}\sum_{l \neq j} g_2(l \mid i) = 1 \end{aligned}$$

and so this defines a valid transition matrix

- also putting

$$\pi_{(i,j)} = g(i,j)$$

and using $g(i,j) = g_1(i \mid j) g_2(j) = g_2(j \mid i) g_1(i)$

$$\pi_{(i,j)} p_{(i,j)(k,l)} = \begin{cases} g^2(i,j)\left(\frac{1}{2g_2(j)} + \frac{1}{2g_1(i)}\right) & i = k, j = l \\ \frac{g(i,j)g(k,,j)}{2g_2(j)} & i \neq k, j = l \\ \frac{g(i,j)g(i,,l)}{2g_1(i)} & i = k, j \neq l \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_{(k,l)} p_{(k,l)(i,j)} = \begin{cases} g^2(i,j)\left(\frac{1}{2g_2(j)} + \frac{1}{2g_1(i)}\right) & i = k, j = l \\ \frac{g(k,j)g(i,,j)}{2g_2(j)} & i \neq k, j = l \\ \frac{g(i,l)g(i,j)}{2g_1(i)} & i = k, j \neq l \\ 0 & \text{otherwise} \end{cases}$$

we have that the chain is time reversible and so has stationary distribution given by g

- also if $g(i,j) > 0$ for all (i,j) then $g_1(i \mid j) > 0$ and $g_2(j \mid i) > 0$ for all (i,j) which implies the chain is irreducible and aperiodic

- therefore by MCCT (i) $\lim_{n\to\infty} p^{(n)}_{(k,l)(i,j)} = g(i,j)$ for every $(i,j) \in \mathcal{S} \times \mathcal{S}$, and (ii) $\lim_{n\to\infty} P((X_n, Y_n) = (i,j)) = \pi_j$ for any initial distribution ν.

■

Exercises

IV.6 Text 2.6.3

IV.7 Text 2.6.4 (like the Metropolis-Hastings algorithm)