Markov chain Monte Carlo and Reversible Markov chains

STK2130 Monday February 18

Problem

Want to evaluate (estimate), for some discrete random variable X

$$\theta = \mathrm{E}[h(X)] = \sum_{j=0}^{N} h(j)\pi_j$$

The parameter θ may for instance be an approximation to some high dimensional integral

$$\theta_0 = \int \cdots \int h(x_1, \ldots, x_m) f(x_1, \ldots, x_m) dx_1 \cdots dx_m$$

and N may be a very large number.

We assume, however, that the probabilities π_j are known.

Idea: MCMC

MCMC = Markov chain Monte Carlo

Construct irreducible and ergodic Markov chain $X_0, X_1, X_2, \ldots$ with stationary distribution $\lim_{n\to\infty} \mathrm{P}(X_n = j) = \pi_j$.

Estimate θ by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

It is advisable to run the chain an initial number of times (burnin-period) so that the process is close to the stationary distribution before start taking the average $\hat{\theta}$.

How could such a Markov chain be constructed?

(one good) Solution: Hasting-Metropolis algorithm

Let Y_{n+1} be a candidate for the new state X_{n+1} chosen from a "proposal" distribution

$$q_{ij} = P(Y_{n+1} = j | X_n = i).$$

Given $X_n = i$ and $Y_{n+1} = j$ we accept the candidate value and let $X_{n+1} = j$ with probability

$$\alpha_{ij} = \min(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}})$$

and otherwise reject the candidate, thus $X_{n+1} = X_n = i$.

Then X_n has stationary distribution $\pi_j = \lim_{n\to\infty} \mathrm{P}(X_n = j)$ if the q_{ij} are transition probabilities for an irreducible MC.

Note: This implies that the chain is aperiodic and recurrent

Why does this work? With this construction

the MC satisfies a "detailed balance equation"

$$\pi_j P_{ji} = \pi_i P_{ij}$$

where $P_{ij} = P(X_{n+1} = j | X_n = i)$.

This obviously holds for $j = i$. For $j \neq i$ note

$$\begin{aligned} \pi_i P_{ij} = \pi_i q_{ij} \alpha_{ij} &= \pi_i q_{ij} \min(1, \tfrac{\pi_j q_{ji}}{\pi_i q_{ij}}) = \min(\pi_i q_{ij}, \pi_j q_{ji}) \\ &= \pi_j q_{ji} \min(1, \tfrac{\pi_i q_{ij}}{\pi_j q_{ji}}) = \pi_j q_{ji} \alpha_{ji} = \pi_j P_{ji} \end{aligned}$$

Summing both sides of the detailed balance equation over i gives

$$\sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j$$

determining the stationary distribution π_j (along with $\sum_j \pi_j = 1$).

Reversible Markov chains

With $P(X_n = j) = \pi_j$ the detailed balance equation can be interpreted as

$$\begin{aligned} \mathrm{P}(X_n = i, X_{n+1} = j) &= \mathrm{P}(X_{n+1} = j | X_n = i)\mathrm{P}(X_n = i) \\ &= \pi_i P_{ij} = \pi_j P_{ji} = \mathrm{P}(X_{n+1} = i | X_n = j)\mathrm{P}(X_n = j) \\ &= \mathrm{P}(X_n = j, X_{n+1} = i). \end{aligned}$$

This has the consequence that it is impossible to check if the chain is recorded forwards or backwards, i.e. reversible.

For some MCs, the detailed balanced equations can simplify the derivation of the stationary distribution (a bit), for instance Random Walks in Example 4.35

A worked toy example

We want to simulate from a distribution on states $0, 1, 2, 3$ and 4

$$\pi = (\pi_0, \pi_1, \ldots, \pi_4) = (0.3, 0.2, 0.1, 0.1, 0.3)$$

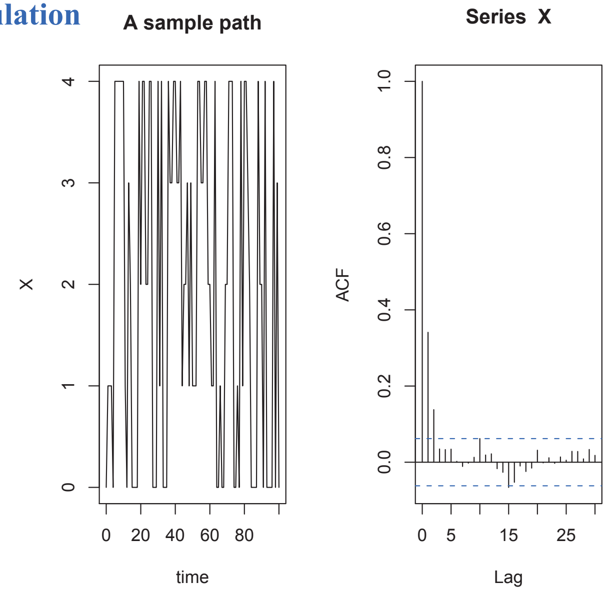
With uniform proposal distrib. $q_{ij} = \mathrm{P}(Y_{n+1} = j | X_n = i) = 0.2$ we get for instance

$$\mathrm{P}_{01} = q_{01} \min(1, \frac{\pi_1 q_{10}}{\pi_0 q_{01}}) = \frac{1}{5}\frac{0.2}{0.3} = 2/15$$

The full transistion matrix becomes

$$\mathrm{P} = \begin{bmatrix} \frac{8}{15} & \frac{2}{15} & \frac{1}{15} & \frac{1}{15} & \frac{3}{15} \\ \frac{2}{10} & \frac{4}{10} & \frac{1}{10} & \frac{1}{10} & \frac{4}{10} \\ \frac{2}{10} & \frac{2}{10} & \frac{2}{10} & \frac{2}{10} & \frac{2}{10} \\ \frac{2}{10} & \frac{2}{10} & \frac{2}{10} & \frac{2}{10} & \frac{2}{10} \\ \frac{3}{15} & \frac{2}{15} & \frac{1}{15} & \frac{1}{15} & \frac{8}{15} \end{bmatrix}$$

A simulation



The left panel shows the autocorrelation corr(X_n, X_{n+k}) which tends fast to 0.

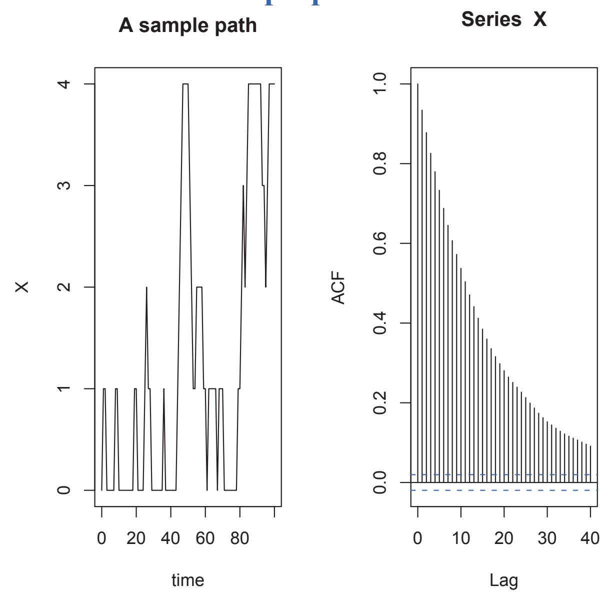
Another proposal distribution

Let the matrix of the q_{ij} equal

$$Q = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

To calculate the transition matrix P for the chain X_n it is only necessary to calculate α_{ij} for which $q_{ij} > 0$.

A simulation with the new prop.distr.



There is larger dependencies in this chain, will need a longer

Choice of q_{ij}

In the preceding example the uniform $q_{ij} = 0.2$ gives smaller dependencies between X_n and X_{n+k} than the second proposal distribution. This shows that there is a need to look for "good" q_{ij}.

It is not, however, always the case that a uniform distribution is best - when the number of states is large and proposal involving just some neighbours can often be preferred.

No need to calculate P

When the number of states $N + 1$ is large it is inconvenient to calculate the transition probabilities P_{ij}. But there is really no need to do so.

The sampling can always be carried out in two steps

(a) Given $X_n = i$ sample $Y_{n+1} = j$ from q_{ij}

(b) Calculate α_{ij} and sample $u_i \sim U[0, 1]$.
If $u_i \leq \alpha_{ij}$ then $X_{n+1} = j$
Otherwise $X_{n+1} = X_n = i$

For instance my R-script

for simulating the second proposal distribution was done without calculating P.

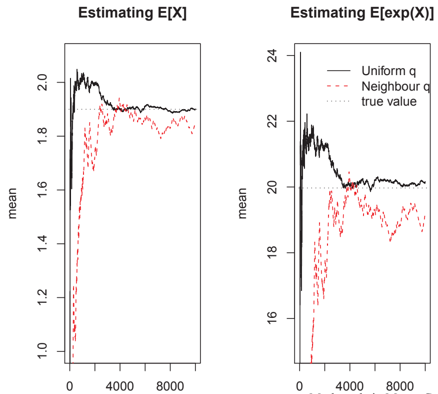
```
X<-0
for (i in 1:10000){
y<-sample(0:4,1,prob=q[X[i]+1,])
alpha<-min(1,px[X[i]+1]*q[X[i]+1,y+1]/(px[y+1]*q[y+1,X[i]+1]))
u<-runif(1)
if (u<alpha) X[i+1]<-y
if (u>alpha) X[i+1]<-X[i]
}
```


Estimation of expectations

Initially we stated that a purpose of MCMC is to evaluate expectations $\theta = \mathrm{E}[h(X)] = \sum_{j=0}^{N} h(j)\pi_j$.

For the toy example we can immediately calculate

$$\theta_1 = \mathrm{E}[X] = \sum_{j=0}^{4} j\pi_j = 1.9 \qquad \theta_2 = \mathrm{E}[e^X] = \sum_{j=0}^{4} e^j\pi_j \approx 19.97$$



Partially determined π_j

In many problems it is only possible to specify the π_j up to a constant, that is we know $\pi_j = b_j/C$ with b_j known and the normalizing constant C unknown.

Hasting-Metropolis is constructed to handle this, because the constant cancels out in

$$\alpha_{ij} = \min(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}) = \min(1, \frac{b_j q_{ji}}{b_i q_{ij}})$$

Such partially determined π_j is for instance often the case in Bayesian statistics (STK4021, H2013) which involve calculating conditional probabilities, using Bayes rule

$$\mathrm{P}(Y = y|X = x) = \frac{\mathrm{P}(X = x|Y = y)\mathrm{P}(Y = y)}{\sum_y \mathrm{P}(X = x|Y = y)\mathrm{P}(Y = y)}$$

Partially determined π_j, contd.

Here we may know $\mathrm{P}(X = x|Y = y)$ and $\mathrm{P}(Y = y)$, but there may be a large number of possible y, thus

$$C = \sum_y \mathrm{P}(X = x|Y = y)\mathrm{P}(Y = y)$$

can be hard to evaluate.

A more complicated example

Suppose we want to simulate from a (slightly silly) distribution

$$\pi_i = \frac{1}{C} f(i) = \frac{1}{C} \exp(-\beta|i| + |i|\log(\alpha) + \gamma \sin(2\pi i))$$

for whole numbers $i \in \mathcal{Z}$ for certain values of $\alpha = 2.222, \beta = 1.111$ and $\gamma = 7.777$.

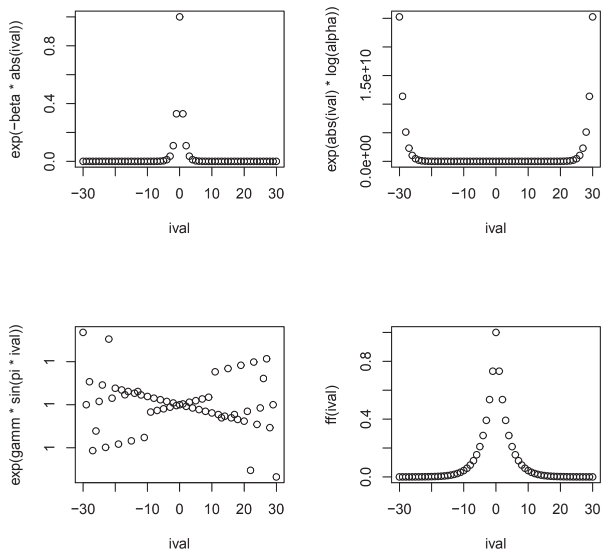
We can not sample candidates uniformly over $\mathcal{Z}$ and simply choose neighbours of $X_n = i$ so the candidates values are

$$Y_{n+1} = \begin{cases} i+1 & \text{with probability} 0.5 \\ i-1 & \text{with probability} 0.5 \end{cases}$$

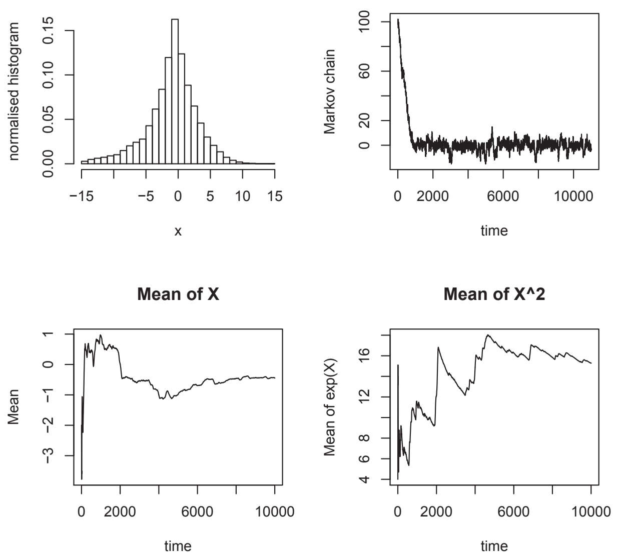
Acceptance probabilities

$$\alpha_{ij} = \min(1, \frac{f(j)}{f(i)}) \text{ for } j = i-1 \text{ and } i+1$$

The function and its components



Results for the "funny" function



Example 4.39 in Ross

$\mathcal{S}$ = the set of all permutations $(x_1, \ldots, x_n)$ of $1, 2, \ldots, n$ that satisfies $\sum_{j=1}^{n} j x_j > \alpha$ for some α.

Want to sample uniformly from $\mathcal{S}$ maybe in order to

- Estimate average $\sum_{j=1}^{n} j x_j$ over $\mathcal{S}$
- Number of elements K in $\mathcal{S}$

Define a neighbour of $s = (x_1, x_2, \ldots, x_n)$ as a permutation where only two elements have changed place. Let $N(s)$ = the number of neighbours of s in $\mathcal{S}$.

Sample candidates uniformly over the neighbours. Since the uniform distribution over $\mathcal{S}$ equals $\pi(s) = \frac{1}{K}$ we get acceptance probabilities

$$\alpha_{s,t} = \min(1, \frac{N(s)}{N(t)})$$

Time reversed Markov chains

Let $\ldots, X_{-n}, X_{-n+1}, \ldots, X_{-1}, X_0, X_1, \ldots, X_n, X_{n+1}, \ldots$ be a Markov chain on times $\mathcal{Z}$ with a stationary distribution $\pi_j = \mathrm{P}(X_n = j)$ for all n.

If we reverse the chain, $Y_n = X_{-n}$ then also Y_n is a MC with stationary distribution $\pi_j = \mathrm{P}(Y_n = j)$

Thus we can work out the transition probabilities $Q_{ij} = \mathrm{P}(Y_{n+1} = j | Y_n = i)$ as

$$\begin{aligned} Q_{ij} &= \mathrm{P}(X_n = j | X_{n+1} = i) = \frac{\mathrm{P}(X_n=j, X_{n+1}=i)}{\mathrm{P}(X_{n+1}=i)} \\ &= \frac{\mathrm{P}(X_n=j)\mathrm{P}(X_{n+1}=i|X_n=j)}{\mathrm{P}(X_{n+1}=i)} = \frac{\pi_j}{\pi_i}\mathrm{P}_{ji} \end{aligned}$$

or $\pi_i \mathrm{Q}_{ij} = \pi_j \mathrm{P}_{ji}$

Time reversible MC's again

Thus we see that if $\mathrm{Q}_{ij} = \mathrm{P}_{ij}$ we obtain

$$\pi_i \mathrm{P}_{ij} = \pi_j \mathrm{P}_{ji}$$

or the detailed balance equation
or the condition for being a time-reversible Markov chain.