Homework 3 : Part A

Oussama Ennafii & Sammy Khalife

25/02/2015

Exercise 1

1. Let $\gamma \in \mathbb{N}^*$ and $X = \{1, .., 6\}$. We consider the Hastings-Metropolis algorithm with target π associated to the proposal distribution:

$$q(x,.) = \mathcal{U}(\{X_n - \gamma, X_n - \gamma + 1, ..., X_n - 1, X_n + 1, ..., X_n + \gamma\})$$

Algorithm 1 Metropolis Hastings algorithm

Given one initial point in $x_0 \in X$
for n=1:N
-Generate $Y_n \sim q(x_n, .)$
-$X_{n+1} = Y_n$ with probability $\alpha(X_n, Y_n)$
and $X_{n+1} = X_n$ with probability $1 - \alpha(X_n, Y_n)$.
end

Since $\forall x \in X, \forall y \in X, q(x,y) = \frac{1}{2\gamma}$, the acceptance ratio is equal to

$$\alpha(x,y) = \begin{cases} 1 \wedge \frac{\pi(y)}{\pi(x)} & \text{if} \quad y \in X \\ 0 & \text{otherwise} \end{cases}$$

To avoid technical discussions, we will consider the extended distribution π on $\mathbb{Z}$, with $\pi(k) = 0$ if $k \notin X$ so that :

$$\alpha(x,y) = 1 \wedge \frac{\pi(y)}{\pi(x)}$$

And the transition kernel is equal, for $y \neq x$ to:

$$\begin{aligned} p(x,y) &= q(x,y)\alpha(x,y) \\ &= \frac{1}{2\gamma}(1 \wedge \frac{\pi(y)}{\pi(x)}) \end{aligned}$$

and :

$$p(x,x) = 1 - \sum_{y \neq x} q(x,y)\alpha(x,y)$$

First :

$$
\begin{aligned}
p(x, y \neq x) &= \frac{1}{2\gamma}(1 \wedge \frac{\pi(y)}{\pi(x)}) \\
&\geq \frac{1}{2\gamma}(1 \wedge \frac{\pi(y)}{\max_x \pi(x)}) \\
&= \frac{\pi(y)}{2\gamma \max_x \pi(x)} \\
&\geq \frac{\pi(y)}{2\gamma}
\end{aligned}
$$

The third equality is due to $\frac{\pi(y)}{\max_x \pi(x)} \leq 1$. The fourth inequality is due to $max_x \pi(x) \leq 1$.

Second,

$$
\begin{aligned}
p(x, x) &= 1 - \sum_{y \neq x} q(x, y)\alpha(x, y) \\
&= q(x, x) + \sum_{y \neq x} q(x, y) - \sum_{y \neq x} q(x, y)\alpha(x, y) \\
&= q(x, x) + \sum_{x \neq y} q(x, y)(1 - \alpha(x, y)) \\
&\geq q(x, x) \\
&= \frac{1}{2\gamma} \\
&\geq \frac{\pi(x)}{2\gamma}
\end{aligned}
$$

The second inequality is obtained by upper bounding the sum terms by 1, and using $|X| = 6$. Let $\boxed{\epsilon = \frac{1}{2\gamma}}$, then

$$\forall x, \forall y \quad p(x, y) \geq \epsilon\pi(y)$$

which can be written in terms of subsets B of X :

$$\boxed{p(x, B) \geq \epsilon\pi(B)}$$

2. Given the inequality written above, the algorithm satisfies the Doeblin condition. The Doeblin Lemma gives : $\boxed{\Delta(P) \leq (1 - \epsilon)}$

Given one initial distribution ξ, we have in the general case:

$$||\xi P^n - \pi||_{TV} \leq (\Delta(P))^n ||\xi - \pi||_{TV}$$

which yields here :

$$\boxed{||\xi P^n - \pi||_{TV} \leq (1 - \frac{1}{2\gamma})^n ||\xi - \pi||_{TV}}$$

Where ξP^n is actually the distribution of X_n. We have geometric ergodicity

3. Cf code for implementation.
The following error is computed with comparison of the numerical mean and the theoretical mean : $\pi(1) = \frac{1}{2}$, $\pi(2) = \frac{2}{6}$, $\pi(3) = \frac{1}{24}$, $\pi(4) = \frac{1}{24}$, $\pi(5) = \frac{1}{24}$, $\pi(6) = \frac{1}{24}$, using the average of simulations (Monte Carlo method).

$\gamma = 1$, $error = 0.1440$.
$\gamma = 4$ $error = 0.0641$
$\gamma = 50$ $error = 0.1818$

The large error for $\gamma 50$ is not surprising since small values of γ provide a better theoretical upper bound (cf 2). Here, we see that there is a trade off with regards of the value of γ. Small values will not promote the exploration of the set X which explains the bad convergence of the algorithm.

Exercise 2

1. Let a family of transition matrices

$$P_t = \begin{bmatrix} t & (1-t) \\ (1-t) & t \end{bmatrix}$$

The obvious invariant distribution is $\pi = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$.

Properties of this Markov Chain

Since $t \in]0,1[$, the Markov Chain is obviously irreducible.

Since the coefficients $p_{1,1}$ and $p_{2,2}$ are strictly positive ($t > 0$), all the states are aperiodic and then the chain is aperiodic.

Finally, we will show that the chain is recurrent positive by considering $p_{i,i}^{(k)}$.

The matrix can reduce itself to

$$D = \begin{bmatrix} 2t-1 & 0 \\ 0 & 1 \end{bmatrix}$$

with $Q = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $QDQ^{-1} = P_t$ and $P_t^n = QD^nQ^{-1}$, which gives :

$$P_t^n \begin{bmatrix} (2t-1)^n\frac{1}{2} + \frac{1}{2} & -(2t-1)^n\frac{1}{2} + \frac{1}{2} \\ -(2t-1)^n\frac{1}{2} + \frac{1}{2} & (2t-1)^n\frac{1}{2} + \frac{1}{2} \end{bmatrix}$$

Then, $\forall i \in \{0,1\}, \sum_{k\geq 0} p_{i,i}^{(k)} = +\infty$, and the chain is recurrent positive. There is convergence of the law of X_n towards the invariant distribution independently of the initial distribution.

2. Given the algorithm written X_n depends only on X_{n-1} and,

$$\mathbb{P}(X_{n+1} = x | X_n = y) = P_{t_0}(y, x)\mathbf{1}_{y=0} + P_{t_1}(y, x)\mathbf{1}_{y=1}$$

, then

$$P_* = \begin{bmatrix} t_0 & (1 - t_1) \\ (1 - t_0) & t_1 \end{bmatrix}$$

Solving

$$\pi^* P_* = \pi^*$$

gives $\pi_1^* = \frac{\pi_2(t_1-1)}{t_0-1}$, and $\pi_1^* + \pi_2^* = 1$ yields :

$$\boxed{\pi_1^* = \frac{t_1 - 1}{t_0 + t_1 - 2}}$$

and

$$\boxed{\pi_2^* = \frac{t_0 - 1}{t_0 + t_1 - 2}}$$

The chain is obviously irreducible (each state is available from another since $t_0 \in]0, 1[$ and $t_1 \in]0, 1[$).

Exercise 3

We choose for the numerical simulation the parameters:

$$\begin{aligned} N : &= 1000 \\ c : &= 2 \\ n_0 : &= 100 \end{aligned}$$

1. In figure 1, we see how the sequence X_1^n varies too much and takes a long time to converge to the mean which is 0 in this case. The second plot shows that the acceptance ratio also converges , slowly though, up to around 70%. So there is room to do better.

2. In figure 2, we see that the sequence X_n converges abruptly to the mean after $n0$. The second plot shows that the acceptance ratio also converges rapidly up to around 100%. Further, we can see that the suboptimality factor depend on n in 2 ways. First, it varies a lot and stays strictly over 1 (the optimal case). In the second phase, it starts decreasing very slowly to 1.

3. The suboptimality factor of the first algorithm is constant independent from c:

$$n \mapsto d\frac{trace(\Gamma_\pi^2)}{trace(\Gamma_\pi)^2} > 1$$

The adaptive algorithm yields a sequence that adapts and thus convergences, slowly, to the optimal case. In conclusion, the adaptive random walk is much better than the naive one.

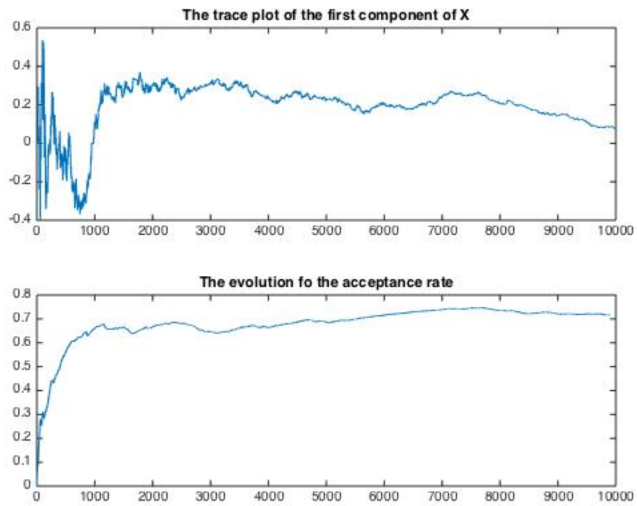

Figure 1: The naive symmetric random walk Hastings-Metropolis chain

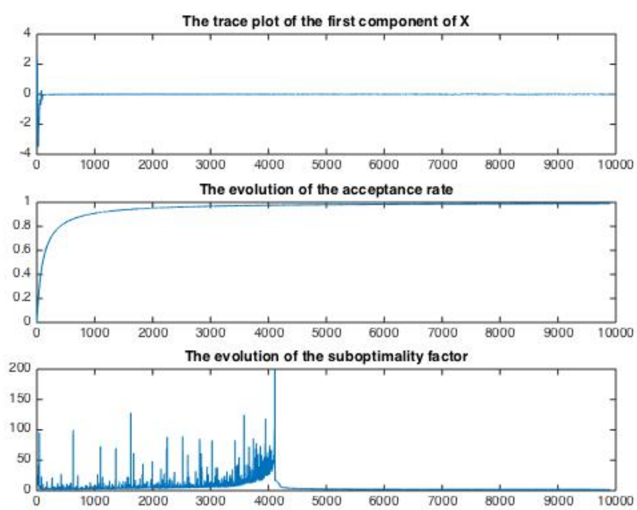

Figure 2: The adaptive symmetric random walk Hastings-Metropolis chain