Convergence times for parallel Markov chains

Béatrice Lachaud[1] and Bernard Ycart[2]

[1] Departament de Matemàtiques, Universitat Autònoma de Barcelona, lachaud@mat.uab.es
[2] IMAG, Université Grenoble 1, bernard.ycart@imag.fr

Summary. For the numerous applications of Markov chains (in particular MCMC methods), the problem of detecting an instant at which the convergence takes place is crucial. The 'cut-off phenomenon', or abrupt convergence, provides an answer to this problem. When a sample of Markov chains, or more generally of exponentially converging processes, is simulated in parallel, it remains far from its stationary distribution until a deterministic instant, and approaches it exponentially fast afterwards. The cut-off instant is explicitly known, and can be detected algorithmically using appropriate stopping times. The technique is illustrated on the Ornstein-Uhlenbeck diffusion.

1 Introduction

The Monte Carlo Markov Chains methods are widely used nowadays in a huge area of applications fields (see [6, 11] for general references). They are adapted to the simulation of a given probability distribution ν, when the state space is too large for a direct simulation, even when ν is only known up to a proportionality constant. They consist in expressing the target distribution ν as the asymptotic distribution of a certain Markov chain, which is then run for long enough until it 'reaches its equilibrium'. The crucial question is to define a stopping time. Among the different possibilities, we will focus on the *parallelization method*. It consists in running n independent copies of the chain, until that sample has reached stationarity. Indeed, contrarily to a single copy, the sample converges in a very abrupt way, for n large enough: before a certain deterministic time, the so-called 'cut-off instant', the sample remains far from its equilibrium, then it converges exponentially fast. Thus it is natural to stop the algorithm at the cut-off instant, provided this instant can be detected algorithmically. It turns out that, in many cases of practical interest, the cut-off instant is asymptotically equivalent to the 'hitting time', defined as the instant at which an empirical mean of the sample reaches its expected value for the first time. The hitting time can be computed on line.

Thus a natural method consists of running the sample until the empirical mean of some function reaches its expected value for the first time. At that instant, the distribution of the sample should be close to equilibrium, thus an n-sample of the distribution ν has been simulated.

A similar method had been proposed by Ycart [16], but the cut-off phenomenon on which it is based could be justified only for finite state space Markov chains at the time of that article. Recent results show that the cut-off phenomenon is much more general than previously thought. Thus it is reasonable to expect that the stopping time method can be used to detect the convergence of samples of processes, under very general hypotheses.

This article generalizes the result of [16] to stochastic processes on an infinite state space. The theoretical basis of the method has been developed in the article [3], where the cut-off phenomenon for samples of exponentially converging processes is studied. Here we will discuss more practical issues.

Since its identification by Aldous and Diaconis [1], the cut-off phenomenon of steep convergence to equilibrium has been observed on many Markov chains [2, 12, 4, 10, 13, 15, 9, 14, 5]. Recently in [3], the cut-off phenomenon has been exhibited for n-samples of independent processes, identically distributed or not, and for different distances. The main result of [3] gives conditions under which such a n-tuple has a cut-off in the sense of total variation, Hellinger, chi-square, and Kullback distances. For each $i = 1, \ldots, n$, the i-th coordinate is assumed to converge with exponential rate ρ_i to its equilibrium measure, in the sense that the logarithm of the distance at time t is equivalent to $-\rho_i t$. Under technical conditions on the ρ_i's, it is proved that a cut-off occurs for the n-tuple at time

$$t_n = \max\left\{ \frac{\log i}{2\rho_{(i,n)}} \; ; \; i = 1, \ldots, n \right\} \; ,$$

where $\rho_{(1,n)}, \ldots, \rho_{(n,n)}$ are the values of $\rho_1, \ldots, \rho_n$ ranked in increasing order.

Sharper results are proved for i.i.d. coordinates: if ρ is the common rate of exponential convergence for the coordinates, then not only does a cut-off occur for the n-tuple at time $\log n/(2\rho)$, but for any fixed u precise estimates show that the distance to equilibrium at time $\log n/(2\rho) + u$ converges to a positive value. In other words, the cut-off occurs over an interval of time of length $O(1)$ around $\log n/(2\rho)$.

In Section 2, theoretical results about the cut-off phenomenon will be stated more precisely. We will also introduce an appropriate hitting time, similar to the one defined in [16]. In the particular case of the Ornstein-Uhlenbeck process, we will recall the results of Lachaud in [8] that prove the asymptotic equivalence between the hitting and the cut-off times. This is precisely the type of result needed to justify that stopping the algorithm at the hitting time will indeed output a sample of the asymptotic distribution as desired. The method, and its practical illustration, will be presented in Section 3, together with experimental results for the Ornstein-Uhlenbeck process.

2 The theoretical results

The known theoretical results about the cut-off phenomenon and the link with appropriate hitting times are recalled in this section.

Let X be a stochastic process with state space E, converging to its asymptotic distribution, denoted by ν. Distances between probability distributions can be measured in many ways: see Gibbs and Su [7] for a useful review. We shall use here mainly the Kullback distance (denoted by d_K):

$$d_K(\mu,\nu) = \left(\int_{S_\mu} f \, \log(f/g) \, d\lambda \right)^{1/2} ,$$

where λ is a measure that dominates both μ and ν, f and g are the respective densities of μ and ν with respect to λ, and S_μ is the support of μ. The Kullback distance is used here only because it gives the simplest statements; it could be replaced by other distances.

We suppose that the convergence to equilibrium is exponentially fast. More precisely, denoting by $\mathcal{L}X(t)$ the distribution of the process X at time t, we assume that there exist two positive reals R and ρ (the exponential rate of convergence) such that,

$$\lim_{t\to+\infty} d_K(\mathcal{L}X(t),\nu)e^{\rho t} = R \, . \tag{1}$$

Let $X^{(n)} = (X_1, \ldots, X_n)$ be a n-sample of the process X, i.e. a n-tuple of independent copies of X. Like its coordinates, the process $X^{(n)}$ converges at exponential rate ρ to its asymptotic distribution, which is the tensor product of n copies of ν. We denote by $d_n(t)$ the Kullback distance between the distribution of $X^{(n)}$ at time t and the equilibrium distribution $\nu^{\otimes n}$. As a consequence of the results proved in [3], one can state the following theorem.

Theorem 1. *Assume* (1) *holds, and let u be a fixed real. Then*

$$\lim_{n\to\infty} d_n \left(\frac{\log n}{2\rho} + u \right) = Re^{-\rho u} \, . \tag{2}$$

The limit (2) is a simple way to express the cut-off phenomenon. If u is a large negative real, then $Re^{-\rho u}$ is large: before the cut-off instant $\frac{\log n}{2\rho}$ the sample is far from equilibrium. If u is a large positive real, then $Re^{-\rho u}$ is small: after the cut-off instant $\frac{\log n}{2\rho}$ the sample converges exponentially fast to equilibrium. Similar results can be proved for the Hellinger, chi-square and total variation distances: see in particular Theorems 15 and 16 of [3].

Consider now a 'test function', i.e. a mapping f from the state space E into $\mathbf{R}$. Let $\{X^{(n)}(t)\} = \{(X_1(t), \ldots, X_n(t))\}$ be the sample process, starting with all coordinates equal to x_0. The empirical mean of f at time t is:

$$M_{x_0}^{(n)}(f)(t) = \frac{1}{n} \sum_{i=1}^{n} f(X_i(t)) \, . \tag{3}$$

For all $t \geq 0$, the random variable $M_{x_0}^{(n)}(f)(t)$ has expectation:

$$E\left(M_{x_0}^{(n)}(f)(t)\right) = E\left(f(X(t))\right) \ . \tag{4}$$

At time 0, it is equal to $f(x_0)$. By the law of large numbers, if n is large, $M_{x_0}^{(n)}(f)(t)$ should stay close to $E\left(f(X(t))\right)$ at all times. If X converges exponentially fast to ν in the sense of the total variation distance, then as t tends to infinity, $E\left(f(X(t))\right)$ converges exponentially fast to $\nu f = \int f \, d\nu$. It seems reasonable to stop the sample when $M_{x_0}^{(n)}(f)(t)$ first meets νf. Let us define the appropriate hitting time by:

$$T_{x_0}^{(n)}(f) = \inf\{\, t \geq 0 \,;\, M_{x_0}^{(n)}(f)(t) = \nu f \,\} \ . \tag{5}$$

It is intuitively clear that $T_{x_0}^{(n)}(f)$ should be asymptotically equivalent to the cut-off instant, at least in probability:

$$\forall \epsilon > 0 \ , \quad \lim_{n \to \infty} Prob\left(\left| \frac{T_{x_0}^{(n)}(f)}{\frac{\log n}{2\rho}} - 1 \right| < \epsilon \right) = 1 \ . \tag{6}$$

This result has already been proved in [16] in the case where the sampled process is a reversible Markov chain with a finite state space. It has been checked for a number of processes with infinite state space, but unfortunately we are not able at this point to prove it in full generality. This is the reason why we shall restrict our study to the case of the Ornstein-Uhlenbeck process, which has been studied by Lachaud. In [8], the function f is the identity. A central limit theorem for the random variable $T_{x_0}^{(n)}(f)$ is proved, which describes the behaviour of the variable around the cut-off instant. Thus not only do we know that (6) holds, but also the fluctuations of $T_{x_0}^{(n)}(f)$ around $\frac{\log n}{2\rho}$ can be explicitly computed.

The Ornstein-Uhlenbeck process is the solution X of the following stochastic differential equation:

$$\begin{cases} dX(t) = -\rho X(t)\, dt + \sigma\sqrt{2\rho}\, dB(t) \\ X(0) \ = x_0 \end{cases} \ , \tag{7}$$

where σ and ρ are positive reals, x_0 is real, and B is the standard Brownian motion.

The process X is Gaussian, with mean $x_0 e^{-\rho t}$ and variance $\sigma^2(1 - e^{-2\rho t})$. Its stationary distribution ν is the centered Gaussian distribution with variance σ^2, and the convergence is exponentially fast, with rate ρ. The distance between the distribution of X at time t and ν can be explicitly computed. In particular, it is easy to check that (1) holds, with

$$R = \frac{|x_0|}{\sigma\sqrt{2}} \ .$$

Consequently, Theorem 1 shows that the cut-off phenomenon occurs. As in [8], we shall choose the simplest possible function f: the identity, and denote $T_{x_0}^{(n)} = T_{x_0}^{(n)}(f)$. The following proposition is proved in [8]:

Proposition 1. *Let U_n be defined by:*

$$U_n = \rho \left(T_{x_0}^{(n)} - \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left(\frac{2\sigma^2}{x_0^2} \right) \right) . \tag{8}$$

As n tends to infinity, the random variable U_n converges in distribution to the distribution with density g on $\mathbf{R}$, where:

$$g(t) = \frac{2}{\sqrt{\pi}} \exp\left(-t - e^{-2t}\right) . \tag{9}$$

This proposition implies in particular (6), or else that the random variable

$$\frac{\log(n)}{2T_{x_0}^{(n)}} \tag{10}$$

is a consistent estimator of ρ.

In the next section, we shall discuss the practical consequences of these results for the detection of convergence.

3 Applications

Summarizing the ideas and the results given in the previous sections, we suggest the following method to detect the instant of convergence of a stochastic process to its asymptotic distribution:

1. choose an integer n and build a n-sample of the process;
2. run the n-sample, and compute the mean process at each time;
3. stop when the mean process reaches for the first time the expectation of the asymptotic distribution.

Let us illustrate this method in practice. The sampled process is the Ornstein-Uhlenbeck diffusion described in the previous section, with the following parameters:

$$\rho = 1, \quad \sigma = 1 \quad \text{and} \quad x_0 = 10 .$$

We will use two sizes of sample: $n = 100$, and $n = 1000$. In order to follow the evolution of the process and its convergence, we use the following methodology. Fix a sequence of regularly spaced instants $t_1, \ldots, t_k$ (here $t_i = i/100$). The experience consists in simulating a n-sample of Ornstein-Uhlenbeck processes $(X_1, \ldots, X_n)$ with the parameters given above. We test, at each instant t_i, if the n-tuple $(X_1(t_i), \ldots, X_n(t_i))$ is a Gaussian sample with distribution

$\mathcal{N}(0, \sigma^2)$. We use the classical Kolmogorov-Smirnov test. At each instant t_i the p-value of the test is computed. On the same experiment, the first instant where the mean of the n-sample reaches zero (the expectation of the asymptotic distribution) is computed: it is an observation of the r.v. $T_{x_0}^{(n)}$. The experiment is continued after the hitting time, to check whether the equilibrium is indeed obtained.

The above experiment is repeated 1000 times. For each t_i a 1000-sample of p-values is obtained; also, 1000 observations of the r.v. $T_{x_0}^{(n)}$ have been computed. In Figure 1 we have associated to each t_i the average of the 1000 corresponding p-values; the first figure is drawn for $n = 100$ and the second one for $n = 1000$. There is little difference between both; the larger the size of the sample, the steeper the transition to equilibrium, as suggested by Theorem 1.

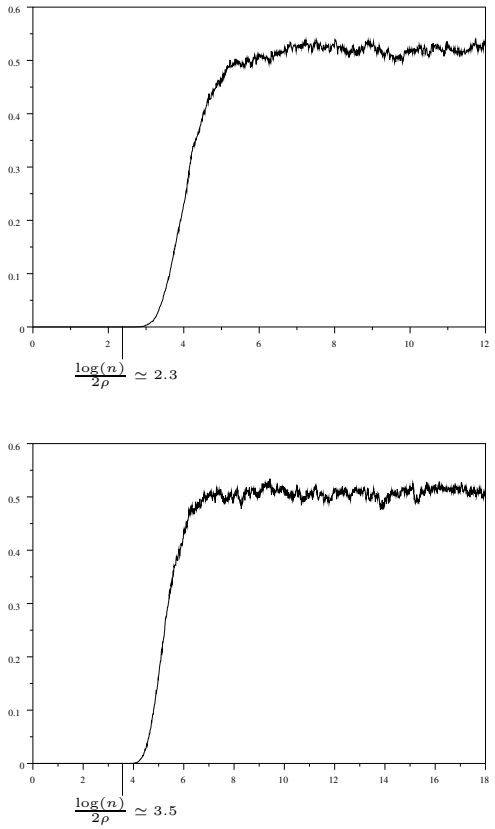

Fig. 1. Average of the p-values of the Kolmogorov-Smirnov test, for $n = 100$ and $n = 1000$.

We can see that the averaged p-value of the test becomes positive from the cut-off instant on. This means that the null hypothesis (i.e. that the n-sample is distributed according to the asymptotic distribution) is systematically rejected before the cut-off instant. To make sure that the convergence is indeed well established, Figure 1 suggests that one should wait until approximately twice the cut-off instant (the stationary regime is established when the averaged p-value stabilizes around 50%).

Using the 1000-sample of observations of the r.v. $T_{x_0}^{(n)}$, the result of Proposition 1 can also be illustrated. Let us compute the following normalization of the sample:

$$U_n = \rho \left(T_{x_0}^{(n)} - \frac{\log(n)}{2\rho} + \frac{1}{2\rho} \log \left(\frac{2\sigma^2}{x_0^2} \right) \right) .$$

The Kolmogorov-Smirnov test allows us to decide whether this sample is distributed with the limit distribution with density g. Table 1 gives the computed p-values, in the cases $n = 100$ and $n = 1000$. The small p-values suggest that the convergence in Proposition 1 is rather slow.

n	p-value
100	0.0048
1000	0.029

Table 1. p-value of the Kolmogorov-Smirnov test between the 1000-sample of U_n and the limit distribution with density g.

In conclusion we can say that the particular case of the Ornstein-Uhlenbeck process illustrates the close link between the cut-off phenomenon and the detection of convergence. The empirical method that we have proposed in this article is efficient to detect the convergence of a stochastic process to its asymptotic distribution. Experiments show that the convergence cannot be established before the hitting time of the asymptotic expectation by an empirical mean. By waiting for a conservative estimate of twice the hitting time, one can make sure that stationarity is reached.

References

1. D. Aldous, P. Diaconis, Shuffling cards and stopping times, Amer. Math. Monthly 93 (5) (1986) 333–348.
2. D. Aldous, P. Diaconis, Strong uniform times and finite random walks, Adv. Appl. Math. 8 (1987) 69–97.
3. J. Barrera, B. Lachaud, B. Ycart, Cut-off for n-tuples of exponentially converging processes, Stoch. Proc. Appl., to appear (2006).
4. P. Diaconis, The cut-off phenomenon in finite Markov chains, Proc. Nat. Acad. Sci. U.S.A. 93 (4) (1996) 1659–1664.

5. P. Diaconis, L. Saloff-Coste, Separation cut-offs for Birth and Death chains model, Tech. Rep. 2005-14, Dept. of Statistics, Stanford Univ. (2005).
6. G. S. Fishman, Monte-Carlo concepts algorithms and applications, Springer-Verlag (1996).
7. A. Gibbs, F. Su, On choosing and bounding probability metrics, Internat. Statist. Review 70 (3) (2002) 419–435.
8. B. Lachaud, Cut-off and hitting times for a sample of Ornstein-Uhlenbeck processes and its average, J. Appl. Probab. 42 (4) (2005) 1069–1080.
9. S. Martínez, B. Ycart, Decay rates and cut-off for convergence and hitting times of Markov chains with countably infinite state space, Adv. in Appl. Probab. 33 (1) (2001) 188–205.
10. U. Porod, The cut-off phenomenon for random reflections, Ann. Probab. 24 (1) (1996) 74–96.
11. C. P. Robert, Méthodes de Monte-Carlo par chaînes de Markov, Economica (1996).
12. J. S. Rosenthal, On generalizing the cut-off phenomenon for random walks on groups, Adv. in Appl. Math. 16 (3) (1995) 306–320.
13. L. Saloff-Coste, Lectures on finite Markov chains, in: Lectures on probability theory and statistics (Saint-Flour, 1996), Vol. 1665 of Lecture Notes in Math., Springer, Berlin, 1997, pp. 301–413.
14. B. Ycart, Cutoff for Markov chains: some examples and applications, in: Complex systems (Santiago, 1998), Vol. 6 of Nonlinear Phenom. Complex Systems, Kluwer Acad. Publ., Dordrecht, 2001, pp. 261–300.
15. B. Ycart, Cutoff for samples of Markov chains, ESAIM Probab. Statist. 3 (1999) 89–106.
16. B. Ycart, Stopping tests for Markov chain Monte-Carlo methods, Methodol. Comput. Appl. Probab. 2 (1) (2000) 23–36.