CUTOFF FOR SAMPLES OF MARKOV CHAINS

BERNARD YCART[1]

Abstract. We study the convergence to equilibrium of $n-$samples of independent Markov chains in discrete and continuous time. They are defined as Markov chains on the $n-$fold Cartesian product of the initial state space by itself, and they converge to the direct product of n copies of the initial stationary distribution. Sharp estimates for the convergence speed are given in terms of the spectrum of the initial chain. A cutoff phenomenon occurs in the sense that as n tends to infinity, the total variation distance between the distribution of the chain and the asymptotic distribution tends to 1 or 0 at all times. As an application, an algorithm is proposed for producing an $n-$sample of the asymptotic distribution of the initial chain, with an explicit stopping test.

Résumé. Nous étudions la convergence vers l'équilibre d'échantillons de taille n de chaînes de Markov en temps discret et continu. Ces échantillons sont des chaînes de Markov sur le produit cartésien de n copies de l'espace initial par lui-même, et ils convergent vers le produit de n copies de la mesure stationnaire de la chaîne initiale. Des estimations précises de la vitesse de convergence sont données en termes du spectre de la chaîne initiale. Un phénomène de transition rapide se produit, au sens où la distance en variation totale entre la loi de la chaîne et la mesure d'équilibre converge vers 1 ou 0 quand n tend vers l'infini. Comme application, nous proposons un algorithme, muni d'un test d'arrêt explicite, pour produire un échantillon de taille n de la mesure d'équilibre de la chaîne initiale.

1. INTRODUCTION

The word *cutoff* was coined by Diaconis and Aldous to characterize the property of many Markov chains, usually with a high degree of symmetry, to converge abruptly to their stationary distribution. Diaconis [4] and Section 2.4 of Saloff-Coste [16] are two general references on this phenomenon. One of the best known examples is the random walk on the $n-$dimensional hypercube [5,6]. Let $P^{(t)}$ denote the distribution after t steps of that random walk starting with equal coordinates, and π its stationary distribution. If the dimension n is large, the total variation distance between $P^{(t)}$ and π, $\|P^{(t)} - \pi\|$, stays close to 1 for a while, then drops suddenly to a small value and converges exponentially fast to 0 thereafter. The time at which the fall occurs is $(1/4)n \log n$ (see Diaconis and Shahshahani [6]). A very precise description of the evolution in time of $\|P^{(t)} - \pi\|$ is given by Diaconis *et al.* [5].

In this paper an analogous phenomenon is shown to occur for samples of Markov chains. Let $P = (p_{ij})$ be the transition matrix of a reversible and irreducible discrete time Markov chain on a finite state space $E = \{i, j, \dots\}$.

A natural way to build a sample of size n of that chain is to have n copies of it evolve independently. The result is a Markov chain on the product space E^n, for which the transition probability from $(i_1, \ldots, i_n)$ to $(j_1, \ldots, j_n)$ is:

$$p_{i_1 j_1} \ldots p_{i_n j_n}.$$

It will be called the *parallel sample chain*, and its transition matrix will be denoted by $\tilde{P}$. If π is the reversible measure for the initial matrix P, then the product measure $\tilde{\pi} = \pi^{\otimes n}$ is reversible for $\tilde{P}$. If n is large, it will be shown that the parallel sample chain reaches its equilibrium measure $\tilde{\pi}$ at a cutoff time equal to

$$\frac{\log n}{2 \log(1/\alpha)},$$

where α is the closest to 1 among absolute values of eigenvalues of P different from 1 (Props. 3.1 and 3.2).

Let us turn now to continuous time and consider a Markov generator $\Lambda = (\lambda_{ij})$ on E, admitting π as its reversible measure. Coupling together n independent continuous time Markov chains with generator Λ leads to a chain on E^n, for which the only possible transitions change one single coordinate at a time. If $j_m \neq i_m$, the transition rate from $(i_1, \ldots, i_m, \ldots, i_n)$ to $(i_1, \ldots, j_m, \ldots, i_n)$ is $\lambda_{i_m j_m}$. All other transition rates are null. The corresponding generator will be denoted by $\tilde{\Lambda}$. Again, $\tilde{\pi} = \pi^{\otimes n}$ is the reversible measure of $\tilde{\Lambda}$. As will be shown in Propositions 3.4 and 3.5, a cutoff occurs for the chain with generator $\tilde{\Lambda}$, at time

$$\frac{\log n}{2\beta},$$

where β is the smallest among absolute values of non null eigenvalues of Λ.

Discrete and continuous time Markov chains are related through harmonization, also called uniformization in some references (see [3, 11, 16]). Let $\Lambda = (\lambda_{ij})$ be a Markov generator on E and set

$$\lambda = \max_{i \in E} \sum_{j \neq i} \lambda_{ij}.$$

Let I_E be the identity matrix indexed by E. Then $P = I_E + (1/\lambda)\Lambda$ is a transition matrix on E. Conversely, if P is a transition matrix and λ a positive real, then $\Lambda = \lambda(P - I_E)$ is a Markov generator on E. Let $\{X_k, \ k \in \mathbb{N}\}$ be a Markov chain with transition matrix P, and $\{K_t, \ t \geq 0\}$ be a Poisson process with intensity λ, independent from the chain $\{X_k, \ k \in \mathbb{N}\}$. Define for each $t \geq 0$:

$$Z_t = X_{K_t}.$$

Then $\{Z_t, \ t \geq 0\}$ is a continuous time Markov chain with generator Λ. Thus the discrete time chain $\{X_k\}$ and its continuous time version $\{Z_t\}$ differ only by a Poissonian change of scale.

Let us consider now the discrete time counterpart of the generator $\tilde{\Lambda}$ defined above. It is a transition matrix on E^n for which at most one coordinate is changed at each step. If $j_m \neq i_m$, the transition probability from $(i_1, \ldots, i_m, \ldots, i_n)$ to $(i_1, \ldots, j_m, \ldots, i_n)$ is $\lambda_{i_m j_m}/(n\lambda)$. That transition matrix will be denoted by:

$$\tilde{Q} = I_{E^n} + \frac{1}{n\lambda}\tilde{\Lambda}.$$

A Markov chain on E^n with transition matrix $\tilde{Q}$ will be referred to as *sequential sample chain*. That chain has a cutoff at time $(1/2\beta)n\lambda \log n$.

In order to relate the parallel and the sequential sample chains, we shall set $\lambda = 1$ and $\Lambda = P - I_E$. For the parallel sample chain, all coordinates evolve simultaneously according to the transition matrix P. For the sequential sample chain, one coordinate is picked up with probability $1/n$, then it is changed according to P.

Both versions have a cutoff at a time which is of order $n \log n$ (if counted in number of changes of coordinates). However the constant differs. For $\tilde{P}$ (parallel sample), it is

$$\frac{1}{2 \log(1/\alpha)},$$

whereas for $\tilde{Q}$ (sequential sample), it is

$$\frac{1}{2(1 - \alpha')}.$$

Here α is the highest absolute value among eigenvalues of P different from 1, and α' is the closest to 1 among the same eigenvalues. If $\alpha = \alpha'$ then the parallel sample chain will reach equilibrium faster than the sequential one (since $\log(1/\alpha) > (1 - \alpha)$). But it may be the case that P has an eigenvalue very close to -1, the chain being almost periodic. Then $\log(1/\alpha)$ will be small, whereas $1 - \alpha'$ may stay reasonably large. Not surprisingly, the sequential sample chain will smooth out the effect of periodicity much faster than the parallel sample chain.

Finally, take as a particular case $E = \{0, 1\}$ and let P be the deterministic transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then the sequential sample chain with transition matrix $\tilde{Q}$ is the symmetric random walk on the hypercube of dimension n. To avoid periodicity problems, Diaconis and Shahshahani ([6] p. 126) consider a slightly different matrix, namely $(1/(n+1))I_{\tilde{E}} + (n/(n+1))\tilde{Q}$. This does not change essentially the cutoff phenomenon and our result is coherent with theirs for that particular case (cutoff at $(1/4)n \log n$).

Due to their growing number of applications, Markov Chain Monte-Carlo (MCMC) methods have received a lot of attention in the past ten years (see Robert [15] or Fishman [9] as general references). If π is a probability distribution to be simulated, the idea is to express it as the reversible measure of a Markov chain. If that Markov chain is simulated for long enough, and sampled at n regular instants, one can consider that the output is an independent $n-$sample of the distribution π. One major problem is to decide what "long enough" means, *i.e.* when exactly the algorithm should be stopped. Several heuristics have been proposed, in particular by Raftery and Lewis [14] (see also Chap. 6 of Robert [15]). The cutoff phenomenon investigated here may provide an answer to this question. An independent sample of size n of π is a realization of a random variable on E^n, distributed according to the product measure $\pi^{\otimes n}$, which is reversible for both the parallel and the sequential sample chain defined above. Thus, the desired sample can be obtained by simulating either sample chain up to equilibrium, that is up to their cutoff times. Indeed, our results show that the sample chains are far from equilibrium before their cutoff times, and that it is basically useless to run them for longer after. Cutoff times are expressed in terms of the spectrum of the transition matrix P, which is not known in general. For the parallel sample chain, we show that the cutoff can be algorithmically detected. Let F be some subset of E, and let $\{X^{(t)}\} = \{(X_1^{(t)}, \dots, X_n^{(t)})\}$ be the parallel sample chain, starting from coordinates equal to i at time 0. Consider the empirical measure of F at time t:

$$S_i^{(t)}(F) = \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}_F(X_m^{(t)}),$$

where $\mathbb{1}_F(x)$ is the indicator function for x being in F. As n goes to infinity, and for t larger than the cutoff time, this estimate tends to $\pi(F)$. Propositions 4.1 and 4.2 show that in general, $S_i^{(t)}(F)$ stays far from the target value $\pi(F)$ before cutoff, and will reach that value exactly at the cutoff time. Thus, by properly choosing the initial state i and the subset F, one gets not only a way to estimate the cutoff time without knowing the spectrum, but also a rigorous stopping test for MCMC methods.

The paper is organized as follows. Section 2 contains the algebraic results on the spectral decomposition of $\tilde{P}$ and $\tilde{\Lambda}$, using Kronecker products and sums. Section 3 contains the main results. After recalling classical

estimates of the distance $\|P^{(t)} - \pi\|$ of a Markov chain at time t to its reversible distribution, we prove the cutoff results in discrete time (Props. 3.1 and 3.2), then in continuous time (Props. 3.4 and 3.5). Applications to MCMC methods, and simulation results for some random walks on the line are presented in Section 4.

2. Kronecker products and sums

Coupling two independent discrete time Markov chains corresponds to a well known algebraic operation on their transition matrices: the *Kronecker product*. Coupling together continuous time Markov chains leads to a Markov chain on the product space, the generator of which is the *Kronecker sum* of the initial generators. For the following classical definitions and results on Kronecker products and sums, we refer to Chapter 12 of Bellman [1].

Definition 2.1. Let A and B be two square matrices indexed respectively by the elements of the finite sets E and F:

$$A = (a_{ij})_{i,j \in E} \quad \text{and} \quad B = (b_{kl})_{k,l \in F}.$$

The Kronecker product of A by B is the square matrix $A \otimes B$, indexed by $E \times F$, defined as:

$$A \otimes B = (a_{ij}\, b_{kl})_{(i,k),(j,l) \in E \times F}.$$

The Kronecker sum of A and B is the square matrix defined as:

$$A \oplus B = A \otimes I_F + I_E \otimes B,$$

where I_E and I_F denote the identity matrices indexed by E and F respectively.

The eigenvectors of the Kronecker product and sum are the same. They are easily deduced from those of A and B.

Proposition 2.1. *Let $v = (v_i)_{i \in E}$ (respectively: $w = (w_k)_{k \in F}$) be an eigenvector of A (respectively: B) associated to the eigenvalue α (respectively: β). Then the vector*

$$V = (v_i w_k)_{(i,k) \in E \times F}$$

is an eigenvector associated to $\alpha\beta$ for $A \otimes B$, and to $\alpha + \beta$ for $A \oplus B$.

For the rest of this section n is a fixed positive integer. We shall consider $n-$fold products and sums of the same matrix. Let us start with discrete time.

Definition 2.2. Let $P = (p_{ij})$, $i,j \in E$, be the transition matrix of a homogeneous discrete time Markov chain on the finite set E. We call *parallel sample chain* associated to P, a homogeneous, discrete time Markov chain on the product set E^n, with transition matrix:

$$\tilde{P} = P^{\otimes n} = \underbrace{P \otimes \cdots \otimes P}_{n \ times}.$$

It is an immediate consequence of Definition 2.1 and Proposition 2.1 that $\tilde{P}$ is a transition matrix on E^n. Let $(i_1, \ldots, i_n)$ and $(j_1, \ldots, j_n)$ be two states in E^n. The coefficient of order $((i_1, \ldots, i_n), (j_1, \ldots, j_n))$ of $\tilde{P}$ is:

$$p_{i_1 j_1} \cdots p_{i_n j_n}.$$

So for the parallel sample chain, all coordinates change simultaneously at each step and they do so independently, according to the transition matrix P.

Definition 2.3. Let $\Lambda = (\lambda_{ij})$, $i, j \in E$, be the generator of a homogeneous continuous time Markov chain on the finite set E. We call *sequential sample* chain associated to Λ, a homogeneous, continuous time Markov chain on the product set E^n, with generator:

$$\tilde{\Lambda} = \Lambda^{\oplus n} = \underbrace{\Lambda \oplus \cdots \oplus \Lambda}_{n \; times}.$$

Again, it is immediate to check that $\tilde{\Lambda}$ is a Markov generator on E^n. The coefficient of order $((i_1, \ldots, i_n)$, $(j_1, \ldots, j_n))$ of $\tilde{\Lambda}$ can be non null only if the two $n-$tuples differ at one coordinate at most. If $j_m \neq i_m$, the transition rate from $(i_1, \ldots, i_{m-1}, i_m, i_{m+1}, \ldots, i_n)$ to $(i_1, \ldots, i_{m-1}, j_m, i_{m+1}, \ldots, i_n)$ is $\lambda_{i_m j_m}$. So the sequential sample chain is the concatenation of n independent copies of the chain with generator Λ.

From now on, we shall assume that the transition matrix P and the generator Λ are related by $P = I_E + \Lambda$. This will simplify notations without restricting generality (see [16] p. 322). It is quite natural to consider also the harmonized transition matrix on E^n associated to $\tilde{\Lambda}$.

$$\tilde{Q} = I_{E^n} + \frac{1}{n}\tilde{\Lambda}.$$

This matrix $\tilde{Q}$ is related to P by:

$$\tilde{Q} = \frac{1}{n}P^{\oplus n} = \frac{1}{n}\left(\underbrace{P \oplus \cdots \oplus P}_{n \; times}\right).$$

A discrete time Markov chain on E with transition matrix $\tilde{Q}$ will still be called *sequential sample chain* associated to P. The probabilistic interpretation of that chain is that at each step, a coordinate is chosen at random between 1 and n, then its value is changed according to the transition matrix P, the two successive choices being independent. Thus a discrete time sequential sample chain is a particular case of a product Markov chain, in the sense of Mathé [13]. Notice that even if all coordinates are independent at time 0, it is no longer the case at finite times, contrarily to what happens for the parallel sample chain. However asymptotic independence is true for all sample chains. Whatever the initial state, the n coordinates of the parallel and sequential sample chains will be asymptotically independent, and thus the limit distribution will be a direct product of stationary measures for P. It is also an easy consequence of Proposition 2.1.

Proposition 2.2. *Let P be the transition matrix of a Markov chain on the finite set E, and $\Lambda = P - I_E$ be the associated generator. For each n, let π_n be a stationary distribution for P and Λ. Then the product distribution*

$$\pi_1 \otimes \cdots \otimes \pi_n = \left(\prod_{m=1}^{n} \pi_m(i_m)\right)_{(i_1, \ldots, i_n) \in E^n}$$

is a stationary distribution for the parallel and sequential sample chains, in discrete and continuous time. If the initial chain is ergodic, then so are the sample chains.

Our goal is to get sharp estimates on the convergence speed at which the product measure is reached for sample chains. The most precise estimates on that convergence speed for a Markov chain can be obtained by making use of the full spectrum of the transition matrix (see [16]). In our case, the spectrums of $\tilde{P}$, $\tilde{\Lambda}$ and $\tilde{Q}$ are easily deduced from those of P and Λ, using iteratively Proposition 2.1. In order to simplify notations, we shall assume that the matrices P and Λ can be diagonalized. The general case could be treated in a similar way, at the expense of a tedious discussion and heavier notations. It will not be developed here.

Proposition 2.3. *Let A be a square matrix indexed by the elements of a finite set E, with cardinality γ. Let $\{\alpha_1, \ldots, \alpha_\gamma\}$ be the eigenvalues of the matrix A with possible multiplicities, and let $\{v_1, \ldots, v_\gamma\}$ be a base of eigenvectors such that v_ℓ is associated to α_ℓ for all $\ell = 1, \ldots, \gamma$. Denote by $\mathcal{E}$ the set of all mappings from*

$\{1, \dots, n\}$ to $\{1, \dots, \gamma\}$. To each element $\eta = (\eta(m))_{m \leq n}$ of $\mathcal{E}$, associate the vector V_η, indexed by the elements of E^n, defined as follows. For each $(i_1, \dots, i_n) \in E^n$, its coordinate of order $(i_1, \dots, i_n)$ is

$$V_\eta(i_1, \dots, i_n) = \prod_{m=1}^{n} v_{\eta(m)}(i_m).$$

Then V_η is an eigenvector for the matrices $A^{\otimes n}$ and $A^{\oplus n}$. It is associated to the eigenvalue

$$\prod_{m=1}^{n} \alpha_{\eta(m)} \quad for \quad A^{\otimes n},$$

and to the eigenvalue

$$\sum_{m=1}^{n} \alpha_{\eta(m)} \quad for \quad A^{\oplus n}.$$

The spectrums of the matrices $\tilde{P}$, $\tilde{\Lambda}$ and $\tilde{Q}$ can be described as follows if we assume, in order to simplify notations, that the eigenvalues of P are all distinct.

Corollary 2.1. *Let* $\{\alpha_1, \dots, \alpha_\gamma\}$ *be the eigenvalues of the transition matrix* P. *Let* $n_1, \dots, n_\gamma$ *be non negative integers such that*

$$n_1 + \cdots + n_\gamma = n.$$

Then

$$\alpha_1^{n_1} \dots \alpha_\gamma^{n_\gamma}$$

is an eigenvalue for $\tilde{P}$ *with multiplicity*

$$\binom{n}{n_1 \dots n_\gamma} = \frac{n!}{n_1! \dots n_\gamma!}.$$

The values

$$(n_1(\alpha_1 - 1) + \cdots + n_\gamma(\alpha_\gamma - 1))$$

and

$$\frac{1}{n}(n_1\alpha_1 + \cdots + n_\gamma\alpha_\gamma)$$

are eigenvalues for $\tilde{\Lambda}$ *and* $\tilde{Q}$ *respectively, with the same multiplicity.*

All eigenvalues have been obtained since

$$\sum_{\substack{n_1, \dots, n_\gamma \\ n_1 + \cdots + n_\gamma = n}} \binom{n}{n_1 \dots n_\gamma} = \gamma^n,$$

which is the dimension of the matrices.

3. Cutoff times for sample chains

The finite state space is still denoted by E, and its cardinality by γ. From now on, we shall assume that the matrices P and $\Lambda = P - I_E$ are irreducible and reversible with respect to the (strictly) positive measure $\pi = (\pi(i))$, $i \in E$. Although reversibility may seem a very restrictive hypothesis, it is by far the most frequent case in applications to MCMC methods. We believe that our results can be extended to the non reversible case at the expense of heavier notations and techniques. Under the reversibility hypothesis, all eigenvalues of P and

Λ are real. We shall denote by α_ℓ, $\ell = 1, \ldots, \gamma$ the eigenvalues of P, and assume they are ranked in decreasing order.

$$1 = \alpha_1 > \alpha_2 \geq \cdots \geq \alpha_\gamma \geq -1.$$

We shall assume moreover that P is aperiodic, which implies that $\alpha_\gamma > -1$. The eigenvalues of Λ will be denoted by $\beta_\ell = \alpha_\ell - 1$, $\ell = 1, \ldots, \gamma$. The cutoff times will be expressed in terms of the two following quantities:

$$\alpha = \max\{|\alpha_2|, |\alpha_\gamma|\} \quad \text{and} \quad \alpha' = \alpha_2.$$

Reversibility means that P and Λ are self-adjoint in $\ell^2(\pi)$. Let D be the diagonal matrix with diagonal coefficients equal to $\sqrt{\pi(i)}$.

$$D = Diag\left(\sqrt{\pi(i)},\ i \in E\right).$$

Then DPD^{-1} and $D\Lambda D^{-1}$ are symmetric matrices. We shall denote by $\{v_1, \ldots, v_\gamma\}$ an orthonormal base of eigenvectors of DPD^{-1}, such that v_ℓ is associated to α_ℓ for all $\ell = 1, \ldots, \gamma$. The first vector will be chosen such that

$$v_1(i) = \sqrt{\pi(i)}, \quad \forall i \in E.$$

For $t = 0, 1, \ldots$ we shall denote by $P_i^{(t)}$ the distribution at (discrete) time t of a Markov chain with transition matrix P, starting from state i at time 0. This distribution is the $i-$th row of the matrix P^t. Through diagonalization, it can be expressed in terms of the eigenvalues and eigenvectors of DPD^{-1} as:

$$P_i^{(t)}(j) = \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} \sum_{\ell=1}^{\gamma} v_\ell(i)\, v_\ell(j)\, \alpha_\ell^t, \quad \forall i, j \in E. \tag{3.1}$$

In continuous time, we shall denote by $Q_i^{(t)}$ the distribution at time $t \geq 0$ of a Markov chain with generator Λ, starting from state i at time 0. It is the $i-$th row of the matrix $\exp(t\Lambda)$. Its expression is the following:

$$Q_i^{(t)}(j) = \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} \sum_{\ell=1}^{\gamma} v_\ell(i)\, v_\ell(j)\, \exp(t\beta_\ell), \quad \forall i, j \in E. \tag{3.2}$$

The distance between the distribution at time t of a Markov chain and its stationary distribution can be measured in many ways (see Sect. 6 of [4]). We shall use only the *total variation* distance and the *chi-square* distance. The total variation distance between $P_i^{(t)}$ and π is:

$$\left\|P_i^{(t)} - \pi\right\| = \max_{F \subset E}\left|P_i^{(t)}(F) - \pi(F)\right| = \frac{1}{2}\sum_{j \in E}\left|P_i^{(t)}(j) - \pi(j)\right|. \tag{3.3}$$

The chi-square distance is:

$$\chi\left(P_i^{(t)}, \pi\right) = \sum_{j \in E}\frac{\left(P_i^{(t)}(j) - \pi(j)\right)^2}{\pi(j)}. \tag{3.4}$$

By the Cauchy-Schwarz inequality, one has:

$$\left\|P_i^{(t)} - \pi\right\| \leq \frac{1}{2}\left(\chi(P_i^{(t)}, \pi)\right)^{1/2}. \tag{3.5}$$

The chi-square distance is particularly adapted to the reversible case. The following expressions are easily deduced from (3.1) and (3.2).

$$\chi\left(P_i^{(t)}, \pi\right) = -1 + \frac{1}{\pi(i)} \sum_{\ell=1}^{\gamma} v_\ell^2(i) \, \alpha_\ell^{2t} \ , \ \forall t = 0, 1, \ldots \tag{3.6}$$

$$\chi\left(Q_i^{(t)}, \pi\right) = -1 + \frac{1}{\pi(i)} \sum_{\ell=1}^{\gamma} v_\ell^2(i) \, \exp(2t\beta_\ell) \ , \ \forall t \geq 0. \tag{3.7}$$

Of course the choice of the initial state is crucial in the cutoff phenomenon. All our sample chains will be initialized deterministically by the state $\tilde{\imath}$ of E^n, for which all coordinates are equal to $i \in E$.

$$\tilde{\imath} = \left(\underbrace{i, \ldots, i}_{n \ times}\right).$$

We shall first treat the parallel sample chain, with transition matrix $\tilde{P} = P^{\otimes n}$ on E^n. The cutoff time for that chain is

$$\frac{\log n}{2 \log(1/\alpha)}.$$

We want to prove that the total variation distance between the distribution at time t, $\tilde{P}_{\tilde{\imath}}^{(t)}$ and the product distribution $\tilde{\pi} = \pi^{\otimes n}$ tends to 1 before cutoff time and to 0 after. We shall first prove the latter.

Proposition 3.1. *Let ε be a positive real. Assume*

$$t > \frac{\log n}{2 \log(1/\alpha)} - \frac{\log \pi(i)}{2 \log(1/\alpha)} - \frac{\log(\log(1 + 4\varepsilon^2))}{2 \log(1/\alpha)} \ . \tag{3.8}$$

Then,

$$\left\| \tilde{P}_{\tilde{\imath}}^{(t)} - \tilde{\pi} \right\| < \varepsilon.$$

Remark: The result could have been presented in a more usual way by saying that if $t > (\log n + c)/(2 \log(1/\alpha))$ for some $c > 0$ then,

$$\left\| \tilde{P}_{\tilde{\imath}}^{(t)} - \tilde{\pi} \right\| < \frac{1}{2} \left(-1 + \exp\left(\frac{e^{-c}}{\pi(i)}\right) \right)^{1/2} .$$

Formula (3.8) clearly shows the importance of the initial state i in the cutoff phenomenon. It can be expected that starting from a state i with a low asymptotic probability will delay and attenuate the cutoff effect.

Proof. It suffices to prove the inequality for the chi-square distance $\chi(\tilde{P}_{\tilde{\imath}}^{(t)}, \tilde{\pi})$. We shall use the exact expression (3.6) of that distance, adapting it to the eigenvalues of $\tilde{P}$. Let $\tilde{D}$ be the diagonal matrix, indexed by $\tilde{E}$, whose diagonal coefficient of order $(i_1, \ldots, i_n)$ is $\sqrt{\pi(i_1) \ldots \pi(i_n)}$. It is easy to check that

$$\tilde{D}\tilde{P}\tilde{D}^{-1} = (DPD^{-1})^{\otimes n}.$$

Thus the eigenvectors of $\tilde{D}\tilde{P}\tilde{D}^{-1}$ can be deduced from those of DPD^{-1} by Proposition 2.3. Let η be a mapping from $\{1, \ldots, n\}$ to $\{1, \ldots, \gamma\}$, and V_η be the vector indexed by $\tilde{E}$ and defined by

$$V_\eta(i_1, \ldots, i_n) = \prod_{m=1}^{n} v_{\eta(m)}(i_m).$$

For all $\ell = 1, \ldots, \gamma$ denote by n_ℓ the cardinality of $\eta^{-1}(\ell)$. The vector V_η is an eigenvector of $\tilde{D}\tilde{P}\tilde{D}^{-1}$, associated to the eigenvalue:

$$\alpha_1^{n_1} \ldots \alpha_\gamma^{n_\gamma}.$$

Its coordinate of order $\tilde{\imath}$ is

$$V_\eta(\tilde{\imath}) = v_1^{n_1}(i) \ldots v_\gamma^{n_\gamma}(i).$$

Thus the chi-square distance can be expressed as:

$$\begin{aligned}
\chi(\tilde{P}_{\tilde{\imath}}^{(t)}, \tilde{\pi}) &= -1 + \frac{1}{\pi^n(i)} \sum_{\substack{n_1, \ldots, n_\gamma \\ n_1 + \cdots + n_\gamma = n}} \frac{n!}{n_1! \ldots n_\gamma!} (v_1^2(i)\, \alpha_1^{2t})^{n_1} \ldots (v_\gamma^2(i)\, \alpha_\gamma^{2t})^{n_\gamma} \\
&= -1 + \left(\frac{v_1^2(i)}{\pi(i)} \alpha_1^{2t} + \cdots + \frac{v_\gamma^2(i)}{\pi(i)} \alpha_\gamma^{2t} \right)^n.
\end{aligned}$$

In the right hand side of the last equality, the first term in the n−th power is equal to 1. Using $(1+x) \le \exp(x)$, one gets:

$$\chi(\tilde{P}_{\tilde{\imath}}^{(t)}, \tilde{\pi}) \le -1 + \exp\left(n \sum_{\ell=2}^{\gamma} \frac{v_\ell^2(i)}{\pi(i)} \alpha_\ell^{2t} \right) \le -1 + \exp\left(\frac{n}{\pi(i)} \alpha^{2t} \right).$$

To obtain the last inequality, we have used the definition of α and the fact that $\sum_\ell v_\ell^2(i) = 1$, the base $\{v_\ell\}$ being orthonormal.

Now

$$\chi(\tilde{P}_{\tilde{\imath}}^{(t)}, \tilde{\pi}) < \varepsilon \quad \Longleftarrow \quad \frac{n}{\pi(i)}\alpha^{2t} < \log(1+\varepsilon)$$

$$\Longleftarrow \quad t > \frac{\log n}{2\log(1/\alpha)} - \frac{\log \pi(i)}{2\log(1/\alpha)} - \frac{\log(\log(1+\varepsilon))}{2\log(1/\alpha)}.$$

The result follows by (3.5). $\qquad \square$

Proposition 3.2 proves the convergence to 1 of $\|\tilde{P}_{\tilde{\imath}}^{(t)} - \tilde{\pi}\|$ before the cutoff time. The proof will use classical Chernov bounds for the tail probabilities of the binomial distribution (see for instance [2] p. 55).

Lemma 3.1. *Let B be a binomial random variable with parameters n and p. For all $b \in]0,1[$, define*

$$h(p,b) = \left(\frac{1-p}{1-b} \right)^{1-b} \left(\frac{p}{b} \right)^b.$$

Then

$$P[B > nb] < h^n(p,b) \quad \text{if} \quad b > p,$$

and

$$P[B < nb] < h^n(p,b) \quad \text{if} \quad b < p.$$

Proposition 3.2. *Let i be an element of E such that*

$$w(i) = \sum_{\ell:\, |\alpha_\ell|=\alpha} v_\ell^2(i) > 0.$$

Let c be a positive real. There exists $n_0(c) > 0$ such that if $n > n_0(c)$ and

$$t < \frac{\log n - c}{2\log(1/\alpha)},$$

then

$$\left\| \tilde{P}_{\tilde{i}}^{(t)} - \tilde{\pi} \right\| > 1 - 4 \exp\left(\frac{-e^c w^2(i)}{8\pi(i)(1 - \pi(i))} \right). \tag{3.9}$$

Proof. Notice first that the total variation distance is a decreasing function of time, so that if we can prove (3.9) for some t, it will also be true for all smaller values of time. We are going to exhibit a set $\tilde{F} \subset E^n$ depending on t, such that for n large enough, and $t < (\log n - c)/(2\log(1/\alpha))$,

$$\tilde{\pi}(\tilde{F}) > 1 - 2 \exp\left(\frac{-e^c w^2(i)}{8\pi(i)(1-\pi(i))} \right) \text{ and } \tilde{P}_{\tilde{i}}^{(t)}(\tilde{F}) < 2 \exp\left(\frac{-e^c w^2(i)}{8\pi(i)(1-\pi(i))} \right).$$

This will prove the result, by definition of the total variation distance (3.3). Notice that (3.9) is trivially true if the righthand side is negative. From now on, we assume c is large enough to make it positive. The construction of the set $\tilde{F}$ will make use of the number of coordinates equal to i in the sample:

$$N_i(i_1, \dots, i_n) = \sum_{m=1}^{n} \mathbb{1}_{\{i\}}(i_m).$$

Under $\tilde{P}_{\tilde{i}}^{(t)}$, the distribution of N_i is binomial with parameters n and $P_i^{(t)}(i)$, whereas under $\tilde{\pi}$, it is binomial with parameters n and $\pi(i)$. Notice that $P_i^{(2t)}(i)$ is a decreasing function of time. The idea is that for t even, and below cutoff, the number of coordinates equal to i in a sample chain, starting from $\tilde{i}$, is significantly higher than $n\pi(i)$. Thus we define $\tilde{F}$ as follows:

$$\tilde{F} = \left\{ (i_1, \dots, i_n) \in E^n \ : \ N_i(i_1, \dots, i_n) < n(\pi(i) + (P_i^{(t)}(i) - \pi(i))/2) \right\}.$$

Recall the identity (3.1):

$$P_i^{(t)}(i) = \sum_{\ell=1}^{\gamma} v_\ell^2(i)\, \alpha_\ell^t.$$

Assume $t = t(n)$ remains even and $c = c(n)$ is a bounded function such that $t = (\log n - c)/(2\log(1/\alpha))$. Then

$$P_i^{(t)}(i) = \pi(i) + n^{-1/2} e^{c/2} w(i) + o(n^{-1/2}).$$

Applying Lemma 3.1 with $p = \pi(i)$ and $b = \pi(i) + (P_i^{(t)}(i) - \pi(i))/2$ yields

$$\tilde{\pi}(\tilde{E} \setminus \tilde{F}) < \exp\left(\frac{-e^c w^2(i)}{8\pi(i)(1 - \pi(i))} + o(1) \right).$$

If Lemma 3.1 is applied to $p = P_i^{(t)}(i)$ and the same value of b, the same bound holds for $\tilde{P}_{\tilde{i}}^{(t)}(\tilde{F})$. The result follows. $\qquad\square$

At this point it is natural to ask whether the cutoff result could hold uniformly whatever the initial state of the sample chain. In Proposition 3.2, the hypothesis clearly indicates that convergence can happen earlier than $\log n/(2\log(1/\alpha))$, depending on the initial state. But it does not happen later and Proposition 3.1 easily yieds a uniform upperbound.

Proposition 3.3. *Let ε be a positive real. Assume*

$$t > \frac{\log n}{2\log(1/\alpha)} - \frac{\log \underline{\pi}}{2\log(1/\alpha)} - \frac{\log(\log(1 + 4\varepsilon^2))}{2\log(1/\alpha)}, \tag{3.10}$$

with $\underline{\pi} = \min_{i \in E} \pi(i)$. Then,

$$\left\| \tilde{P}_{\bar{\imath}}^{(t)} - \tilde{\pi} \right\| < \gamma \varepsilon,$$

for all $\bar{\imath} = (i_1, \dots, i_n) \in \tilde{E}$.

Proof. For $i \in E$, denote by n_i the number of coordinates of $\bar{\imath}$, and by $\{\tilde{Y}_i^{(t)}\}$ the sample chain reduced to those coordinates. Proposition 3.1 applies to $\{\tilde{Y}_i^{(t)}\}$, replacing n by n_i. But the $\{\tilde{Y}_i^{(t)}\}$'s are independent and the distribution of the sample chain $\tilde{X}$ is the product of their distributions, through appropriate renumbering of coordinates. Now if P and π (respectively P' and π') are probability distributions on $\tilde{E}$ (respectively $\tilde{E}'$), then

$$\| P \otimes P' - \pi \otimes \pi' \| \leq \| P - \pi \| + \| P' - \pi' \|.$$

The result follows immediately.

$\square$

Let us now turn to continuous time. The sequential sample chain has generator $\tilde{\Lambda} = \Lambda^{\oplus n}$ on E^n. As before, the sample will be initialized with identical coordinates. The distribution at time $t \geq 0$ will be denoted by $\tilde{Q}_{\bar{\imath}}^{(t)}$. The cutoff time is

$$\frac{\log n}{2(1 - \alpha')}.$$

Proposition 3.4. *Let ε be a positive real. Assume*

$$t > \frac{\log n}{2(1 - \alpha')} - \frac{\log \pi(i)}{2(1 - \alpha')} - \frac{\log(\log(1 + 4\varepsilon^2))}{2(1 - \alpha')} . \tag{3.11}$$

Then,

$$\| \tilde{Q}_{\bar{\imath}}^{(t)} - \tilde{\pi} \| < \varepsilon.$$

Proof. The proof follows exactly the same lines as that of Proposition 3.1. It relies on an upper bound for the chi-square distance $\chi(\tilde{Q}_{\bar{\imath}}^{(t)}, \tilde{\pi})$. The eigenvalues of Λ are:

$$0 = \beta_1 > \beta_2 \geq \cdots \geq \beta_\gamma,$$

where $\beta_\ell = \alpha_\ell - 1$. With the notations of Propositions 2.3 and 3.1, the vector V_η is an eigenvector of $\tilde{D}\tilde{\Lambda}\tilde{D}^{-1}$, associated to the eigenvalue

$$n_1 \beta_1 + \cdots + n_\gamma \beta_\gamma.$$

Using (3.7), the chi-square distance can be expressed as follows.

$$\begin{aligned}
\chi(\tilde{Q}_{\bar{\imath}}^{(t)}, \tilde{\pi}) &= -1 + \frac{1}{\pi^n(i)} \sum_{\substack{n_1, \dots, n_\gamma \\ n_1 + \cdots + n_\gamma = n}} \frac{n!}{n_1! \dots n_\gamma!} \left(v_1^{2n_1}(i) \dots v_\gamma^{2n_\gamma}(i) \right) \exp\left(2t(n_1 \beta_1 + \cdots + n_\gamma \beta_\gamma) \right) \\
&= -1 + \left(\frac{v_1^2(i)}{\pi(i)} \exp(2t\beta_1) + \cdots + \frac{v_\gamma^2(i)}{\pi(i)} \exp(2t\beta_\gamma) \right)^n \\
&\leq -1 + \exp\left(n \sum_{\ell=2}^{\gamma} \frac{v_\ell^2(i)}{\pi(i)} \exp(2t\beta_\ell) \right) \leq -1 + \exp\left(\frac{n}{\pi(i)} \exp(2t\beta_2) \right).
\end{aligned}$$

The result follows easily from the last inequality (recall that $1 - \alpha' = -\beta_2$).

$\square$

Proposition 3.4 yields immediately a uniform result analogous to Proposition 3.3 which will not be written here.

Proposition 3.5. *Let i be an element of E such that*

$$w'(i) = \sum_{\ell:\, \alpha_\ell = \alpha'} v_\ell^2(i) > 0.$$

Let c be a positive real. There exists $n_0(c) > 0$ such that if $n > n_0(c)$ and

$$t < \frac{\log n - c}{2(1 - \alpha')},$$

then

$$\|\tilde{Q}_i^{(t)} - \tilde{\pi}\| > 1 - 4\exp\left(\frac{-e^c w'^2(i)}{8\pi(i)(1 - \pi(i))}\right). \tag{3.12}$$

Proof. The proof is similar to that of Proposition 3.2. There are only two differences. The first one is that $Q_i^{(t)}(i)$ is a decreasing function of time in all cases. The second one is the estimate of $Q_i^{(t)}(i)$, for $t < (\log n - c)/(2(1-\alpha'))$. By the identity (3.2):

$$Q_i^{(t)}(i) = \pi(i) + n^{-1/2} e^{c/2} w'(i) + o(n^{-1/2}).$$

The rest of the proof is identical, and we shall omit the details. □

The discrete time sequential sample chain has transition matrix $\tilde{Q} = (1/n)P^{\oplus n}$. It differs from the continuous time sample chain by a Poissonian change of scale, with intensity n. Its cutoff time is

$$\frac{n \log(n)}{2(1 - \alpha')}.$$

The analogue of Proposition 3.4 is obtained straightforwardly using the spectral decomposition of $\tilde{Q}$ and standard inequalities. For the lowerbound, the technique of Propositions 3.1 and 3.2 does not apply since the coordinates of the chain are not independent anymore. One can deduce from Proposition 3.5 a weaker result by using the large number properties of the Poissonian change of time (see also [5] p. 59). We shall not detail these results.

Diaconis *et al.* [5] give a very precise description of the total variation distance $\|\tilde{Q}_{\tilde{i}}^{(t)} - \tilde{\pi}\|$ for the random walk on the hypercube. We believe that their results can be generalized to the present situation. Let $\{X^{(t)}\} = \{(X_1^{(t)}, \ldots, X_n^{(t)})\}$ be the parallel sample chain, starting from $\tilde{i}$ (the argument is the same for in continuous time). Consider the empirical measure associated to $\{X^{(t)}\}$:

$$S_i^{(t)}(F) = \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}_F(X_m^{(t)}), \ \forall F \subset E.$$

For all $F \subset E$, the random variable $S_i^{(t)}(F)$ is binomial, with parameters n and $P_i^{(t)}(F)$. As t tends to infinity, it converges in distribution to a binomial with parameters n and $\pi(i)$. For n large, these binomial distributions have normal approximations. The maximum over F of the total variation distance between the two normal approximations is an asymptotic lower bound for $\|\tilde{P}_{\tilde{i}}^{(t)} - \tilde{\pi}\|$. In view of Remark A (p. 60 of [5]), it seems natural to conjecture that it is actually a good approximation. The idea of using the empirical measure for cutoff detection will be implemented in Section 4.

4. Application to MCMC convergence

In a growing number of applications, sampling from a given distribution π is done through the simulation of a Markov chain, admitting π as its reversible measure. Many methods for constructing the transition matrix P of such a chain have been proposed, in particular the famous Metropolis algorithm (see Chap. 4 of Robert [15]). The goal being to obtain a sample of size n of the distribution π, one can run a single trajectory of the chain with transition matrix P, and sample it at n regular instants, far enough from each other. The problem is to choose those instants, optimizing the running time of the algorithm while guaranteeing the distribution of the sample (see [15], Chap. 6). An alternative is to run n independent copies of the chain, starting all from the same initial state, *i.e.*, simulate the parallel sample chain studied in the previous section. The advantage of the latter approach is that the cutoff time is a natural stopping time for the algorithm: the distribution of the sample is far from the goal before cutoff time, and very close after. So the simulation should be run at least until cutoff time, while it is useless to run it for much longer after. Since the cutoff time grows only as $\log(n)$, this method seems reasonable from the point of view of computing time. Here the sequential sample chain will not be considered, since it is algorithmically less efficient.

Of course, one should observe that after the term in $O(\log(n))$ of formula (3.8), comes a term in $O(\log(\pi(i))$, which is an indication of the fact that the cutoff effect will be delayed and damped in the case of a very large state space. A clear cutoff can be expected in practice only if the sample size n is much larger than the cardinality of the state space γ. This is undoubtedly a serious limitation to the method proposed here.

In general the state space E has a large number of elements, and computing the spectrum and eigenvectors of P is not feasible. We will show that the cutoff time can be detected algorithmically without knowing a priori its value. The idea is an extension of the method of proof of Propositions 3.2 and 3.5. There, it was shown that the proportion of the sample coordinates equal to i should be far from $\pi(i)$ before cutoff, and reach that value exactly at cutoff. However, if $\pi(i)$ is small, testing the proportion of coordinates equal to i can lead to very imprecise results. But the same reasoning can be applied to the probability of any subset F of E. Let $\{X^{(t)}\} = \{(X_1^{(t)}, \dots, X_n^{(t)})\}$ be the parallel sample chain, starting with all coordinates equal to i. The empirical measure of F at time t is

$$S_i^{(t)}(F) = \frac{1}{n} \sum_{m=1}^n \mathbb{1}_F(X_m^{(t)}).$$

Assume that $i \in F$ (this does not restrict generality, since F can be replaced by $E \setminus F$). Thus $S_i^{(0)}(F) = 1$. We define the *hitting time* of the asymptotic value $\pi(F)$ as the following random variable $T_i(F)$:

$$T_i(F) = \inf \left\{ t \geq 0 \ : \ S_i^{(t)}(F) < \pi(F) \right\}.$$

It is natural to expect that $T_i(F)$ should be close to the cutoff time. This is not always true.

Proposition 4.1. *Let $\{X^{(t)}; t \in \mathbb{N}\}$ be the parallel sample chain with transition matrix $\tilde{P}$, starting from $X^{(0)} = \tilde{\imath}$. Assume i and F are such that:*

- $w_i(F) = \displaystyle\sum_{\ell: |\alpha_\ell| = \alpha} \sum_{j \in F} \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} v_\ell(i) \, v_\ell(j) > 0,$
- $P_i^{(t)}(F)$ *is a decreasing function of t.*

Then

$$T_i(F) \left(\frac{\log(n)}{2 \log(1/\alpha)} \right)^{-1}$$

converges to 1 in probability as n tends to infinity.

In other terms, $\log(n)/(2T_i(F))$ is a consistent estimator of $\log(1/\alpha)$.

Proof. The probability distribution of the parallel sample chain $\{X^{(t)}\,;\,t \in \mathbb{N}\}$, starting from $X^{(0)} = \tilde{\imath}$ will be denoted by *Prob*. Let c be a constant. We need to prove that, with a probability tending to 1 as n tends to infinity, $T_i(F)$ is larger than $c\log(n)/(2\log(1/\alpha))$ if $c < 1$ and smaller if $c > 1$. The former will be proved first.

Denote by t_0 the integer part of $c\log(n)/(2\log(1/\alpha))$. The probability that $T_i(F) \le t_0$ is the probability that at least one of the $S_i^{(t)}(F)$ is smaller than $\pi(F)$, for $t = 0, \ldots, t_0$. Hence:

$$Prob[T_i(F) \le t_0] \le \sum_{t=0}^{t_0} Prob\left[S_i^{(t)}(F) < \pi(F)\right] \le t_0 Prob\left[S_i^{(t_0)}(F) < \pi(F)\right].$$

The last inequality comes from the fact that $P_i^{(t)}(F)$ is decreasing. Notice that the distribution of $nS_i^{(t)}(F)$ is binomial, with parameters n and $P_i^{(t)}(F)$. As in the proof of Proposition 3.2, we may assume that t_0 is even. From (3.1) comes:

$$P_i^{(t_0)}(F) = \pi(F) + w_i(F)n^{-c/2} + o(n^{-c/2}).$$

Chernov bound (Lem. 3.1), applied to $p = P_i^{(t_0)}(F)$ and $b = \pi(F)$ gives:

$$Prob\left[S_i^{(t_0)}(F) < \pi(F)\right] \le \exp\left(-\frac{w_i^2(F)}{2\pi(F)(1-\pi(F))}n^{1-c} + o(n^{1-c})\right).$$

If $c < 1$ then $t_0 Prob[S_i^{(t_0)}(F) < \pi(F)]$ tends to 0 as n tends to infinity. Let now c be larger than 1. Let d be an integer, t and s two positive reals. Consider the d regularly spaced instants $t+s, \ldots, t+ds$, and define S as the random vector

$$S = \left(S_i^{t+s}(F), \ldots, S_i^{t+ds}(F)\right).$$

Our goal is to show that

$$Prob\left[S \in [\pi(F), +\infty)^d\right]$$

can be made arbitrarily small, for a particular choice of t, s and d such that:

$$\frac{\log n}{2\log(1/\alpha)} < t + s < t + ds < \frac{c\log n}{2\log(1/\alpha)}.$$

Some estimates on the distribution of the vector S are needed here. For $m = 1, \ldots, n$ denote by J_m the random vector

$$J_m = \left(\mathbb{1}_F(X_m^{(t+s)}), \ldots, \mathbb{1}_F\left(X_m^{(t+ds)}\right)\right).$$

The vectors J_m are i.i.d. and their sum is nS. Their expectation is

$$\mathbb{E}[J_m] = \left(P_i^{(t+s)}(F), \ldots, P_i^{(t+ds)}(F)\right).$$

Denote by $C = (C_{kh})$ their covariance matrix. For $1 \le k < h \le d$ one has:

$$C_{kk} = P_i^{(t+ks)}(F)(1 - P_i^{(t+ks)}(F)),$$

$$C_{kh} = \sum_{j\in F} P_i^{(t+ks)}(j)P_j^{((h-k)s)}(F) - P_i^{(t+ks)}(F)P_i^{(t+hs)}(F).$$

Let $v = \sqrt{\pi(F)(1-\pi(F))}$. From (3.1), as t and s tend to infinity, C tends to the diagonal matrix $v^2 I_d$. More precisely, let $\|.\|_\infty$ be the maximum absolute value of all coefficients of a matrix. One has:

$$\left\|C - v^2 I_d\right\|_\infty \le (k_1\alpha^t + k_2\alpha^s),$$

where k_1 and k_2 are two constants, not depending on d.

Fix n_0 and c_0 such that for $n > n_0$ and $t > (\log(n) + c_0)/(2\log(1/\alpha))$:

$$P_i^{(t+ks)}(F) < \pi(F) + \frac{1}{\sqrt{n}}.$$

Then

$$Prob\left[S_i^{(t+ks)}(F) \geq \pi(F)\,,\ 1 \leq k \leq d\right] \leq Prob\left[\sqrt{n}(S_i^{(t+ks)}(F) - P_i^{(t+ks)}(F)) \geq -1\,,\ 1 \leq k \leq d\right].$$

The idea is to bound the above probability by the probability of $[-1, +\infty)^d$ under the $d-$dimensional Gaussian distributions $\mathcal{N}_d(0, C)$ (mean 0, covariance matrix C) and $\mathcal{N}_d(0, v^2 I_d)$. For this, two successive bounds are needed, first a Berry-Esséen estimate, then a bound for the distance between the two normal distributions. This is a standard procedure for empirical processes, and can be found in several references, such as Doukhan ([7] p. 51) and Giné [10]. Using the inequalities given in the proof of Lemma 5.6 of [10] (p. 132), one gets:

$$
\begin{aligned}
Prob[S \in]\pi(F), +\infty[^d] \quad &\leq \quad \mathcal{N}_d(0, C)(]-1, +\infty[^d) + \frac{k_3 dv^2}{\sqrt{n}} \leq \mathcal{N}_d(0, v^2 I_d)(]-1, +\infty[^d) \\
&\quad + \frac{k_3 dv^2}{\sqrt{n}} + k_4(d)(\alpha^t + \alpha^s) = \rho^d + \frac{k_3 dv^2}{\sqrt{n}} + k_4(d)(\alpha^t + \alpha^s),
\end{aligned}
$$

where $\rho = \Phi(1/v) < 1$, denoting by Φ the standard Gaussian distribution function. For $\varepsilon > 0$, one can fix d such that $\rho^d < \varepsilon/3$, then $n_1 > n_0$ such that for $n > n_1$, $k_3 dv^2/\sqrt{n} < \varepsilon/3$. It remains to choose $t > (\log n/(2\log(1/\alpha))$ and s such that

$$Prob[S \in]\pi(F), +\infty[^d] < \varepsilon.$$

Thus the probability that $T_i(F)$ is smaller or equal to $t + ds$ is larger than $1 - \varepsilon$. Moreover, for n large enough, the choices can be made to satisfy:

$$t + ds < \frac{c \log n}{2 \log(1/\alpha)}.$$

Hence the result.   □

As already remarked in the proof of Proposition 3.2, the hypothesis that $P_i^{(t)}(F)$ is a decreasing function of time is fulfilled for $F = \{i\}$, if all eigenvalues of P are non negative. This can be ensured replacing P by P^2, *i.e.* simulating two steps of the chain in a row. The problem is that the hypotheses of Proposition 4.1 cannot be checked in practice, except in very particular cases. Notice though that if the state space E is endowed with a partial order and if the Markov chain with matrix P is stochastically monotone for that order (see for instance Massey [12]) then there are some natural choices for i and F. Indeed, let i be a maximal element, and F be an increasing subset of E:

$$j \in F \Longrightarrow k \in F\,,\ \forall k \geq j.$$

Then $P_i^{(t)}(F)$ (or $Q_i^{(t)}(F)$) is a decreasing function of time. This can be used for instance for birth and death processes or attractive spin systems.

In general, it can be expected that $P_i^{(t)}(F) - \pi(F)$ will change sign several times, so that the hitting time $T_i(F)$ will be significantly lower than cutoff. This can be arranged by considering a new family of hitting times. We shall define $T_i^{(M)}(F)$ as the first time at which $S_i^{(t)}(F)$ enters a confidence interval around $\pi(F)$.

Definition 4.1. Let $i \in E$, $F \subset E$ and $M > 0$. Let $\{X^{(t)}\,;\ t \in \mathbb{N}\}$ be the parallel sample chain with transition matrix $\tilde{P}$, starting from $X^{(0)} = \tilde{\imath}$. We call hitting time of $\pi(F)$ starting from i at level M, the random variable

$$T_i^{(M)}(F) = \inf\left\{t\ :\ \frac{\sqrt{n}|S_i^{(t)}(F) - \pi(F)|}{\sqrt{\pi(F)(1 - \pi(F))}} < M\right\}.$$

As can be expected, the analogue of Proposition 4.1 holds for $T_i^{(M)}(F)$, under weaker hypotheses.

Proposition 4.2. *Assume i and F are such that*

- $w_i(F) = \displaystyle\sum_{\ell:\,|\alpha_\ell|=\alpha} \sum_{j\in F} \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} v_\ell(i)\, v_\ell(j) \neq 0,$
- $P_i^{(t)}(F) - \pi(F)$ *is non null for all $t \in \mathbb{N}$.*

Then

$$T_i^{(M)}(F) \left(\frac{\log(n)}{2\log(1/\alpha)}\right)^{-1}$$

converges to 1 in probability as n tends to infinity.

Proof. As in the proof of Proposition 4.1, we must prove that, with a probability tending to 1 as n tends to infinity, the hitting time $T_i^{(M)}(F)$ is larger than $c\log(n)/(2\log(1/\alpha))$ if $c < 1$ and smaller if $c > 1$. For $c > 1$, the argument is comparable, and we shall not reproduce it here. For $c < 1$, let t_0 be the integer part of $c\log(n)/(2\log(1/\alpha))$. We need to prove that for all $t = 0,\dots,t_0$, $P_i^{(t)}(F)$ stays away from the confidence interval

$$\left[\pi(F) \pm \frac{M}{\sqrt{n}}\sqrt{\pi(F)(1-\pi(F))}\right].$$

There exists a fixed t_1 such that $|P_i^{(t)}(F) - \pi(F)|$ is strictly decreasing for $t > t_1$. The equation $P_i^{(t)}(F) = \pi(F)$ has at most γ solutions in $\mathbb{R}^+$, all between 0 and t_1. Since it is assumed that none of these solutions is an integer, one gets

$$\min_{t \leq t_1} |P_i^{(t)}(F) - \pi(F)| > 0.$$

By the already used argument (Chernov bound), with probability tending to 1 with n, the hitting time $T_i^{(M)}(F)$ will be larger than t_0. $\square$

We believe that the cases where both $T_i(F)$ and $T_i^{(M)}(F)$ fail to detect the cutoff are rarely encountered in practice. Running a parallel sample chain up to a hitting time can reasonably be expected to output a fair sample of the target distribution π. We present below two examples of Markov chains, for which simulations experiments have been run. Let us start with an example where $T_i(F)$ fails to estimate the cutoff time, whatever i and F. Consider the following transition matrix P on $E = \{1,2,3\}$.

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}. \tag{4.1}$$

Its reversible measure is:

$$\pi(1) = 0.4 \qquad \pi(2) = 0.4 \qquad \pi(3) = 0.2.$$

The eigenvalues are $\{-0.809,\, 0.309,\, 1\}$, thus $\alpha = 0.809$ and $\alpha' = 0.309$. Since $P_3^{(1)}(3) = 0$, the hitting time $T_3(\{3\})$ is always equal to 1. As n tends to infinity, the hitting time $T_1(\{1\})$ converges to 3, since $P_1^{(3)}(1) = 0.375$. All other hitting times tend to 1, 2 or 3. It is clear on this example that the problem comes from the oscillatory behavior of $P_i^{(t)}(F)$, which is alternatively smaller and larger than the target value $\pi(F)$. This is due to the negative eigenvalue. Replacing P by P^2 solves the problem. Table 1 presents estimated values for the expectation and standard deviation of $T_1(\{1\})$, computed over 1000 independent experiments, using P^2 instead of P. The theoretical value $\log(n)/(2\log(1/\alpha))$ is given for comparison.

Simulation estimates for the parallel sample chain based on P, for hitting times of $\pi(1)$, starting from $i = 1$, at different levels M are presented in Table 2. In order to ease the interpretation, the levels M where chosen

TABLE 1.   Estimated mean and standard deviation of $T_1(\{1\})$, compared with cutoff time, for the parallel sample chain based on P^2, where P is given in (4.1).

n	estimated mean	standard deviation	$\dfrac{\log(n)}{2\log(1/\alpha)}$
1000	12.668	5.167	16.297
10000	18.056	4.896	21.729

by reference to the standard Gaussian distribution:

$$M = \Phi^{-1}\left(\frac{1+\epsilon}{2}\right).$$

With that choice, $T_i^{(M)}(F)$ appears as the first entry time in a bilateral confidence interval with asymptotic level of confidence ϵ. The sample size is $n = 10000$, and the expectations and standard deviations were estimated over 1000 independent simulations. The expectations should be compared with the theoretical value of the cutoff time: $\log(n)/(2\log(1/\alpha)) = 21.729$. The expectation and standard deviation of $T_1^{(M)}(\{1\})$ are decreasing functions of ϵ.

TABLE 2.   Estimated mean and standard deviations of $T_1^{(M)}(\{1\})$, for the parallel sample chain based on the matrix P of (4.1).

ϵ	0.1	0.2	0.3	0.4	0.5
estimated mean	21.201	16.540	14.367	13.516	12.766
standard deviation	9.048	4.909	3.790	2.978	2.453

In order to test the method with different sizes of state space, the natural random walk on $E = \{1,\dots,\gamma\}$ was considered (Ex. 2.3.1 p. 348 of Saloff-Coste [16]). The transition matrix P is the following.

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & \dots & 0 \\ 0.5 & 0 & 0.5 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0.5 & 0 & 0.5 \\ 0 & \dots & 0 & 0.5 & 0.5 \end{pmatrix}. \tag{4.2}$$

The uniform measure is reversible: $\pi(i) = 1/\gamma$ for all i. The spectral decomposition of P is given explicitly in Feller [8], XVI.3. It turns out in this case that $\alpha = \alpha'$. We chose $i = 1$ and $F = \{1,\dots,j\}$, for j ranging from

1 to $\gamma - 1$. Table 3 presents estimates of the expectation $T_1(\{1, \ldots, j\})$, each computed over 1000 independent experiments, for the parallel sample chain of size $n = 10000$. For each value of γ, the theoretical value of the cutoff time, $\log(n)/(2\log(1/\alpha))$ is given for comparison. For each value of i and j, a relative error was computed as the quotient of the amplitude of the 0.95 confidence interval by the mean. The maximum of these relative errors was found to be 0.0168.

TABLE 3.   Estimated mean of $T_1(\{1, \ldots, j\})$, compared with cutoff time, for the parallel sample chain based on the matrix P of (4.2).

$\gamma \backslash j$	1	2	3	4	5	6	7	8	9	$\frac{\log(n)}{2\log(1/\alpha)}$
3	9.0	9.1								6.6
4	16.1	17.4	16.2							13.3
5	24.4	27.0	27.0	24.6						21.7
6	34.1	38.2	39.4	38.1	34.0					32.0
7	44.4	49.9	52.6	53.0	50.1	44.9				44.3
8	56.3	63.9	68.1	69.5	68.5	64.6	56.8			58.3
9	69.1	78.7	84.1	87.1	86.7	84.7	79.3	70.2		74.3
10	82.8	95.0	102.3	106.1	108.1	107.0	103.0	96.0	84.0	92.1

REFERENCES

[1] R. Bellman, *Introduction to matrix analysis*. McGraw-Hill, London (1960).

[2] N. Bouleau and D. Lépingle, *Numerical methods for stochastic processes*. Wiley, New York (1994).

[3] E. Çinlar, *Introduction to stochastic processes*. Prentice Hall, New York (1975).

[4] P. Diaconis, The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* **93** (1996) 1659–1664.

[5] P. Diaconis, R. Graham and J. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions. *Rand. Struct. Algorithms* **1** (1990) 51–72.

[6] P. Diaconis and M. Shahshahani, Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.* **18** (1987) 208–218.

[7] P. Doukhan, *Mixing, properties and examples*. Springer-Verlag, New York, *Lecture Notes en Statist.* **85** (1994).

[8] W. Feller, *An introduction to probability theory and its applications*, Vol. I. Wiley, London (1968).

[9] G.S. Fishman, *Monte-Carlo concepts algorithms and applications*. Springer-Verlag, New York (1996).

[10] E. Giné, Lectures on some aspects of the bootstrap, P. Bernard, Ed., *École d'été de probabilités de Saint-Flour XXVI*, Springer-Verlag, New York, *Lectures Notes in Math.* **1664** (1997) 37–151.

[11] J. Keilson, Markov chain models - rarity and exponentiality. Springer-Verlag, New York. *Appl. Math. Sci.* **28** (1979).

[12] A.W. Massey, Stochastic orderings for Markov processes on partially ordered spaces. *Math. Oper. Research* **12** (1987) 350–367.

[13] P. Mathé, Relaxation of product Markov chains on product spaces. Preprint WIAS, Berlin (1997).

[14] A.E. Raftery and S. Lewis, Implementing MCMC, W.R. Gilks, S.T. Richardson and D.J. Spiegelhalter, Eds., *Markov Chain Monte-Carlo in practice*, Chapman and Hall, London (1992) 115–130.

[15] C.P. Robert, *Méthodes de Monte-Carlo par chaînes de Markov*. Economica, Paris (1996).

[16] L. Saloff-Coste, Lectures on finite Markov chains, P. Bernard, Ed., *École d'été de probabilités de Saint-Flour XXVI*, Springer-Verlag, New York, *Lecture Notes in Math.* **1664** (1997) 301–413.