# Cutoff for Markov chains: some examples and applications

B. Ycart

PRISME, Université René Descartes, Paris
Bernard.Ycart@math-info.univ-paris5.fr

**Abstract**

Some Markov chains converge very abruptly to their equilibrium: the total variation distance between the distribution of the chain at time $t$ and its equilibrium measure is close to 1 until some deterministic 'cutoff time', and close to 0 shortly after. Many examples have been studied by Diaconis and his followers. Our goal is to introduce two families of examples of this phenomenon, focusing mainly on their possible applications. We present firstly samples of Markov chains for which the cutoff depends on the size of the sample. As an application, a new way of implementing Markov chain Monte-Carlo algorithms is proposed, using an explicit stopping rule based on the empirical measure of the sample. Then, we shall study Markov chains on countably many states, where the cutoff phenomenon depends on the starting point of the chain. As a particular case, a criterion of cutoff for birth and death chains on trees will be obtained. Jackson networks will show other applications of both cutoff situations.

Key words: Markov chains, Cutoff, MCMC convergence, Hitting times, Birth and death chains, Jackson networks.

AMS Subject Classification: 60 J 10, 60 J 27.

# 1 Introduction

The word *cutoff* was coined by Diaconis and Aldous in 1983 to characterize the property of many Markov chains, usually with a high degree of symmetry, to converge very abruptly to their stationary distribution. Diaconis [6] and section 2.4 of Saloff-Coste [29] are two general references on this phenomenon. A cutoff phenomenon is said to occur in the convergence to equilibrium of a family of Markov chains $(\{X_a(t)\})$, depending on some parameter $a$, when there exists a sequence of deterministic instants $(t(a))$ such that the total variation distance between the distribution of the chain at time $t$ and its asymptotic distribution tends to 1 if $t$ is smaller than $t(a)$ and to 0 if it is larger. Some notations are needed to make the notion precise.

In the cases studied here, the state spaces will be either finite, and will depend on an integer parameter (section 2), or infinite and fixed, in which case the role of the parameter will be played by an initial state tending to infinity (section 4). On a countable state space $I$, the notion of convergence for a function $f$ defined on $I$ will be the usual one: the limit $\lim_{a\to\infty} f(a) = r$ means that $f(a)$ belongs to a neighborhood $V$ of $r$, as soon as $a$ is outside some finite subset $I_V$ of $I$.

We consider a positive recurrent Markov chain $\{X_a(t)\}$ on a countable state space. Its transition kernel may depend on the parameter $a$. It is denoted by $p_a^{(t)}(i,j)$ and $P_a^{(t)}$ is the associated semi-group; $P_a^{(t)}f$ denotes its action on functions and $\nu P_a^{(t)}$ its action on measures. At this point, $t$ can be either integer or real. Sample chains will be studied in section 2 both in discrete and continuous time. But the application proposed in section 3 concerns discrete time. On infinite state spaces (section 4) the time will be continuous, as for Jackson networks (section 5).

The initial distribution of the chain will in general depend on the parameter, but we shall omit the subscript $a$ for a while. We denote by $\mathbb{P}_\nu$ the probability law of the chain starting from the probability measure $\nu$ and by $\mathbb{E}_\nu$ the associated expectation. With this notation $\nu P^{(t)} f = \mathbb{E}_\nu[f(X(t))]$, and $\nu P^{(t)}$ is the distribution of the chain at time $t$, starting from $\nu$ at time 0. In most cases, the chain will start deterministically from a single state. We put $\mathbb{P}_i \doteq \mathbb{P}_{\delta_i}$, $\mathbb{E}_i \doteq \mathbb{E}_{\delta_i}$, where $\delta_i$ means the Dirac mass at $i$. The invariant probability measure is denoted by $\pi$ and satisfies $\pi P^{(t)} = \pi$ for any $t$. Probability measures on a discrete set $I$ will be compared as usual through the total variation distance:

$$\|\nu - \mu\| = \sup_{A\subset I} |\nu(A) - \mu(A)| = \frac{1}{2}\sum_{a\in I} |\nu(a) - \mu(a)| \,. \tag{1.1}$$

Our definition for cutoff is that of [21]:

**Definition 1.1** *Let $(\{X_a(t)\} : a \in I)$ be a family of Markov chains with distributions $\nu_a P_a^{(t)}$ and equilibrium measures $\pi_a$. Let $(t(a))$ be a family of positive reals, tending to $\infty$ as $a$ tends to $\infty$. We say that the family admits a cutoff at time $(t(a))$ if for any*

*positive constant c:*

$$\lim_{a\to\infty} \|\nu_a P_a^{(ct(a))} - \pi_a\| = 1 \quad \textit{for } c < 1 \ , \tag{1.2}$$

*and*

$$\lim_{a\to\infty} \|\nu_a P_a^{(ct(a))} - \pi_a\| = 0 \quad \textit{for } \ c > 1 \ . \tag{1.3}$$

Notice that cutoff in this sense is defined up to equivalence of sequences: if cutoff occurs at time $(t(a))$ and $\lim(t(a)/s(a)) = 1$ then cutoff occurs also at time $(s(a))$.

One of the best known examples is the random walk on the $a-$dimensional hypercube [8, 7]. Let $p_a^{(t)}$ denote the distribution after $t$ steps of that random walk starting with equal coordinates, and $\pi$ its stationary distribution. If the dimension $a$ is large, the total variation distance between $p_a^{(t)}$ and $\pi$, $\|p_a^{(t)} - \pi\|$, 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 $(1/4)a \log a$ (see Diaconis and Shahshahani [8]). A very precise description of the evolution in time of $\|p_a^{(t)} - \pi\|$ is given by Diaconis, Graham, and Morrison [7]. In that case as in many others, the cutoff occurs in a much sharper way than what is suggested by definition 1.1. Several different definitions of the notion have been given (see [29] p. 362). They are usually much sharper than definition 1.1 above. Indeed we shall obtain sharp results in the cases of sample chains and $M/M/\infty$ queues. But Definition 1.1 is well adapted to a particular case that helps shedding a new light on the notion, namely the equivalence in probability of hitting times to their expectations. We shall introduce hitting times as stopping tests for MCMC algorithms in section 3, as well as tools for detecting cutoff on a countable state space in section 4. There it will be shown that cutoff for convergence to equilibrium is actually equivalent to cutoff for any hitting time of a finite part of the state space.

Many distances between probability measures could replace the total variation distance in Definition 1.1 (see [29] p. 358 and [6] p. 1663). Apart from a technical use of the chi-square distance in section 2, we shall stick to the total variation distance which is well adapted to hitting times, as will be shown in sections 3.1 and 4.1. Also some definitions ask for the cutoff property to hold uniformly in the (deterministic) initial state. Our results will depend in general on the initial distribution, though some uniform results will be given, in particular in sections 2 and 5.

This text is not meant as a review of the fast growing literature on the subject, but rather as a presentation of a few results obtained recently, essentially those of references [11, 21, 31, 32], with a strong bias towards applications and unification of seemingly distinct notions. We shall give full proofs only for some results (Propositions 2.1, 2.2, and 3.1), and outline briefly the proofs of some others, referring to [11, 21, 31, 32] for more details.

Section 2 deals with $n$-samples of i.i.d. finite Markov chains both in discrete (2.1) and in continuous time (2.2). The application to MCMC algorithms is treated in section 3. Several notions of hitting times are defined and their asymptotic equivalence to the cutoff time is proved in 3.1. A first series of simulation experiments on the generation

of random stable sets illustrates the method in 3.2. A more concrete application to the generation of random Aboxes in description logics is described in 3.3. Cutoff phenomenons depending on an initial state tending to infinity are the object of section 4. In general (4.1), we shall prove the equivalence between cutoff for the convergence to equilibrium and a weak law of large numbers for hitting times, which is in general much easier to obtain. This will be illustrated by explicit calculations for birth and death chains on trees and lines in 4.2 and 4.3. The results of 2.2 and 4.1 apply quite naturally to Jackson networks of queues (section 5). For closed networks (5.1) the results of section 2 will be applied whereas open networks (5.2) illustrate those of section 4.

# 2 Cutoff for samples of Markov chains

Here the cutoff phenomenon is shown to occur for $n$-tuples of i.i.d. finite Markov chains. Let $P = (p_{ij})$ be the transition matrix of a reversible and irreducible discrete time Markov chain on a finite state space $E = \{i, j, \ldots\}$. A sample of size $n$ of that chain is built by having $n$ 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 $\pi$ 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 $\alpha$ is the closest to 1 among absolute values of eigenvalues of $P$ different from 1 (Propositions 2.1 and 2.2).

Let us turn now to continuous time and consider a Markov generator $\Lambda = (\lambda_{ij})$ on $E$, admitting $\pi$ as its reversible measure. Coupling together $n$ independent continuous time Markov chains with generator $\Lambda$ 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 2.5 and 2.6, a cutoff occurs for the chain with generator $\tilde{\Lambda}$, at time

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

where $\beta$ is the smallest among absolute values of non null eigenvalues of $\Lambda$.

Discrete and continuous time Markov chains are related through harmonization, also called uniformization in some references (see [5, 19, 29]). 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 $\lambda$ a positive real, then $\Lambda = \lambda(P - I_E)$ is a Markov generator on $E$. Let $\{X_k\,,\ k \in I\!N\}$ be a Markov chain with transition matrix $P$, and $\{K_t\,,\ t \geq 0\}$ be a Poisson process with intensity $\lambda$, independent from the chain $\{X_k\,,\ k \in I\!N\}$. 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 $\Lambda$. 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 $\alpha$ is the highest absolute value among eigenvalues of $P$ different from 1, and $\alpha'$ 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 [8] 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 (cutoff at $(1/4)n\log n$).

## 2.1 Discrete time

The finite state space is still denoted by $E$, and its cardinality by $\gamma$. From now on, we will assume that the matrix $P$ is 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 $P$ are real. We shall denote them by $\alpha_\ell\,,\ \ell=1,\ldots,\gamma$, 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 cutoff time is expressed in terms of the following quantity:

$$\alpha=\max\{|\alpha_2|,|\alpha_\gamma|\}\ .$$

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

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

Then $DPD^{-1}$ is a symmetric matrix. We shall denote by $\{v_1,\ldots,v_\gamma\}$ an orthonormal base of eigenvectors of $DPD^{-1}$, such that $v_\ell$ is associated to $\alpha_\ell$ 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 $t=0,1,\ldots$ 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:

$$p_i^{(t)}(j) = \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} \sum_{\ell=1}^{\gamma} v_\ell(i)\, v_\ell(j)\, \alpha_\ell^t \ , \quad \forall i, j \in E \ . \tag{2.1}$$

The distance between the distribution at time $t$ of a Markov chain and its stationary distribution can be measured in many ways (see section 6 of [6]). We shall use only the *total variation* distance and the *chi-square* distance. The total variation distance has already been defined by (1.1). The chi-square distance is:

$$\chi(p_i^{(t)}, \pi) = \sum_{j \in E} \frac{(p_i^{(t)}(j) - \pi(j))^2}{\pi(j)} \ . \tag{2.2}$$

By the Cauchy-Schwarz inequality, one has:

$$\|p_i^{(t)} - \pi\| \leq \frac{1}{2} (\chi(p_i^{(t)}, \pi))^{1/2} \ . \tag{2.3}$$

The chi-square distance is particularly adapted to the reversible case. The following expression is easily deduced from (2.1).

$$\chi(p_i^{(t)}, \pi) = -1 + \frac{1}{\pi(i)} \sum_{\ell=1}^{\gamma} v_\ell^2(i)\, \alpha_\ell^{2t} \ , \ \forall t = 0, 1, \ldots \tag{2.4}$$

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

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

The cutoff time for the parallel sample chain, with transition matrix $\tilde{P} = P^{\otimes n}$ on $E^n$ 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 2.1** *Let $\varepsilon$ be a positive real. Assume*

$$t > \frac{\log n}{2 \log(1/\alpha)} - \frac{\log \pi(i)}{2 \log(1/\alpha)} - \frac{\log(\log(1 + 4\varepsilon^2))}{2 \log(1/\alpha)} \ . \tag{2.5}$$

*Then,*

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

**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,

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

Formula (2.5) 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 $\chi(\tilde{P}_{\tilde{i}}^{(t)}, \tilde{\pi})$. We shall use the exact expression (2.4) 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 $\sqrt{\pi(i_1)\ldots\pi(i_n)}$. Finding the eigenvectors of $\tilde{D}\tilde{P}\tilde{D}^{-1}$ is standard linear algebra, using the fact that the matrix $\tilde{D}\tilde{P}\tilde{D}^{-1}$ is the Kronecker product of $n$ copies of $DPD^{-1}$ by itself. Let $\eta$ be a mapping from $\{1, \ldots, n\}$ to $\{1, \ldots, \gamma\}$, and $V_\eta$ 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 $n_\ell$ the cardinality of $\eta^{-1}(\ell)$. The vector $V_\eta$ 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:

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

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

$$\begin{aligned}
\chi(\tilde{p}_i^{(t)}, \tilde{\pi}) &\leq -1 + \exp\Big(n \sum_{\ell=2}^{\gamma} \frac{v_\ell^2(i)}{\pi(i)}\,\alpha_\ell^{2t}\Big) \\
&\leq -1 + \exp\Big(\frac{n}{\pi(i)}\alpha^{2t}\Big) .
\end{aligned}$$

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

$$\begin{aligned}\chi(\tilde{p}_i^{(t)}, \tilde{\pi}) \ < \ \eta \ &\Longleftarrow \ \frac{n}{\pi(i)}\alpha^{2t} \ < \ \log(1+\eta) \\ &\Longleftarrow \ t > \frac{\log n}{2\log(1/\alpha)} - \frac{\log \pi(i)}{2\log(1/\alpha)} - \frac{\log(\log(1+\eta))}{2\log(1/\alpha)} \ .\end{aligned}$$

The result follows by (2.3).

□

Proposition 2.2 proves the convergence to 1 of $\|\tilde{P}_i^{(t)} - \tilde{\pi}\|$ before the cutoff time. The proof will use classical Chernov bounds for the tail probabilities of the binomial distribution (see for instance [3] p. 55).

**Lemma 2.1** *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*

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

*and*

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

**Proposition 2.2** *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*

$$\|\tilde{p}_i^{(t)} - \tilde{\pi}\| > 1 - 4\exp\left(\frac{-e^c w^2(i)}{8\pi(i)(1-\pi(i))}\right) \ . \tag{2.6}$$

**Proof.**

Notice first that the total variation distance is a decreasing function of time, so that if we can prove (2.6) 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 $t < (\log n - c)/(2\log(1/\alpha))$,

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

This will prove the result, by definition of the total variation distance (1.1). Notice that (2.6) 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:

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

Under $\tilde{p}_i^{(t)}$, 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 $\tilde{\imath}$, is significantly higher than $n\pi(i)$. Thus we define $\tilde{F}$ as follows:

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

Recall the identity (2.1):

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

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 2.1 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 2.1 is applied to $p = p_i^{(t)}(i)$ and the same value of $b$, the same bound holds for $\tilde{p}_i^{(t)}(\tilde{F})$. The result follows.

□

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 2.2, 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 2.1 easily yieds a uniform upperbound.

**Proposition 2.3** *Let $\varepsilon$ be a positive real. Assume*

$$t > \frac{\log n}{2\log(1/\alpha)} - \frac{\log \underline{\pi}}{2\log(1/\alpha)} - \frac{\log(\log(1+4\varepsilon^2))}{2\log(1/\alpha)} \ , \tag{2.7}$$

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

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

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

Observe that the hypothesis of proposition 2.2 holds for at least one state $i \in E$. This permits to state a uniform cutoff result, provided the initial state of the sample covers a sizable proportion of each element of $E$.

**Proposition 2.4** *Let $(q_1, \ldots, q_\gamma)$ be a strictly positive probability distribution on $E$. Let $n_1(n), \ldots, n_\gamma(n)$ be functions from $\mathbb{N}$ into $\mathbb{N}$ such that:*

$$\sum_{i=1}^{\gamma} n_i(n) = n \quad \textit{and} \quad \lim_{n\to\infty} \frac{n_i(n)}{n} = q_i \ .$$

*For each $n$ let $\bar{\imath} = (i_1, \ldots, i_n)$ be an element of $E^n$ such that:*

$$\sum_{m=1}^{n} \mathbb{1}_{\{i\}}(i_m) = n_i(n) \ , \ i = 1, \ldots, \gamma \ .$$

*Let $c(n)$ be any function from $\mathbb{N}$ into $\mathbb{R}^+$, tending to infinity as $n$ tends to infinity. Define:*

$$t^-(n) = \max\{0 \, , \, \frac{\log n}{2\log(1/\alpha)} - c(n)\} \quad \textit{and} \quad t^+(n) = \frac{\log n}{2\log(1/\alpha)} + c(n) \ .$$

*Then*

$$\lim_{n\to\infty} \|\tilde{p}_{\bar{\imath}}^{(t^-(n))} - \tilde{\pi}\| = 1 \quad \textit{and} \quad \lim_{n\to\infty} \|\tilde{p}_{\bar{\imath}}^{(t^+(n))} - \tilde{\pi}\| = 0 \ .$$

## 2.2 Continuous time

In continuous time, we shall denote by $q_i^{(t)}$ the distribution at time $t \geq 0$ of a Markov chain with generator $\Lambda$, starting from state $i$ at time 0. It is the $i-$th row of the matrix $\exp(t\Lambda)$. Assume the eigenvalues of $\Lambda$ are ranked in decreasing order:

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

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 expressed is terms of $\beta = -\beta_2$:

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

There is very little difference with propositions 2.1 and 2.2. The proofs follow exactly the same lines and we refer to [31] for details. In Proposition 2.6, $\{v_\ell\}$ is an orthonormal basis of eigenvectors of the symmetric matrix $D\Lambda D^{-1}$, just as in Proposition 2.2.

**Proposition 2.5** *Let $\varepsilon$ be a positive real. Assume*

$$t > \frac{\log n}{2\beta} - \frac{\log \pi(i)}{2\beta} - \frac{\log(\log(1+4\varepsilon^2))}{2\beta} \ . \tag{2.8}$$

*Then,*

$$\|\tilde{q}_i^{(t)} - \tilde{\pi}\| < \varepsilon \ .$$

**Proposition 2.6** *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*

$$\|\tilde{q}_i^{(t)} - \tilde{\pi}\| > 1 - 4\exp\left(\frac{-e^c w'^2(i)}{8\pi(i)(1-\pi(i))}\right) \ . \tag{2.9}$$

Proposition 2.5 immediately yields a uniform result analogous to proposition 2.3 which will not be written here. Let us just state the analogue of proposition 2.4.

**Proposition 2.7** *Let $(q_1, \ldots, q_\gamma)$ be a strictly positive probability distribution on $E$. Let $n_1(n), \ldots, n_\gamma(n)$ be functions from $\mathbb{N}$ into $\mathbb{N}$ such that:*

$$\sum_{i=1}^{\gamma} n_i(n) = n \quad \textit{and} \quad \lim_{n\to\infty} \frac{n_i(n)}{n} = q_i \ .$$

*For each $n$ let $\bar{\imath} = (i_1, \ldots, i_n)$ be an element of $E^n$ such that:*

$$\sum_{m=1}^{n} \mathbb{1}_{\{i\}}(i_m) = n_i(n) \ , \ i = 1, \ldots, \gamma \ .$$

*Let $c(n)$ be any function from $\mathbb{N}$ into $\mathbb{R}^+$, tending to infinity as $n$ tends to infinity. Define:*

$$t^-(n) = \max\{0 \, , \, \frac{\log n}{2\beta} - c(n)\} \quad \textit{and} \quad t^+(n) = \frac{\log n}{2\beta} + c(n) \ .$$

*Then*

$$\lim_{n\to\infty} \|\tilde{q}_{\bar{\imath}}^{(t^-(n))} - \tilde{\pi}\| = 1 \quad \textit{and} \quad \lim_{n\to\infty} \|\tilde{q}_{\bar{\imath}}^{(t^+(n))} - \tilde{\pi}\| = 0 \ .$$

Assume that $\lambda = \max_i \sum_j \lambda_{ij} = 1$. The discrete time sequential sample has transition matrix $\tilde{Q} = I_{\tilde{E}} + (1/n)\tilde{\Lambda}$. It differs from the continuous time sample chain by a Poissonian change of scale, with intensity $n$. Its cutoff time is

$$\frac{n \log(n)}{2\beta} .$$

The analogue of Proposition 2.5 is obtained straightforwardly using the spectral decomposition of $\tilde{Q}$ and standard inequalities. For the lowerbound, the technique of propositions 2.2 does not apply since the coordinates of the chain are not independent anymore. One can deduce from proposition 2.6 a weaker result by using the large number properties of the Poissonian change of time (see also [7] p. 59). We shall not detail these results.

Diaconis, Graham, and Morrison [7] give a very precise description of the total variation distance $\|\tilde{q}_{\tilde{\imath}}^{(t)} - \tilde{\pi}\|$ 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 $\tilde{\imath}$ (the argument is the same 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 the 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 $\|\tilde{p}_{\tilde{\imath}}^{(t)} - \tilde{\pi}\|$. In view of Remark A. p. 60 of [7], 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 3.

# 3 Application to MCMC convergence

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 [27] or Fishman [14] as general references). If $\pi$ is a probability distribution to be simulated, the idea is to express it as the reversible measure of a Markov chain. Then two alternatives are possible, the *single chain* and *parallel chain* methods. In the first one, a sample of size $n$ is obtained by extracting $n$ regularly spaced values of a single trajectory of the Markov chain $\{X(t)\}$:

$$(X(t_0), X(t_0 + t_1), \ldots, X(t_0 + (n-1)t_1)) .$$

Even though several heuristics have been proposed, in particular by Raftery and Lewis [26] (see also Chapter 6 of Robert [27]), no rigourous result at this day permits a clear

choice for $t_0$ and $t_1$. The other method consists in running $n$ independent copies of the chain, starting with a preassigned initial distribution. It is believed to be the best way to obtain a good theoretical control on the distribution on the sample (see Chauveau and Diébolt [9]). We believe that the cutoff phenomenon gives a sound theoretical basis to the parallel chain method. Before its cutoff time $\log(n)/(2\log(1/\alpha))$, the parallel sample chain is very far from equilibrium, and it is very close after. The goal being to obtain a $n-$sample of the distribution $\pi$, i.e. a realization of the distribution $\tilde{\pi}$, one should run the $n$ independent copies at least until cutoff, and it is essentially useless to run them for longer after. The cutoff time is expressed in terms of the spectrum of the transition matrix $P$, which is not known in general. However we shall prove that the cutoff can be algorithmically detected using the *empirical measure* associated to the sample.

Let $f$ be any state function, defined on $E$, with values in $\mathbb{R}$. Consider the empirical mean of $f$ at time $t$:

$$S_i^{(t)}(f) = \frac{1}{n}\sum_{m=1}^{n} f(X_m^{(t)}) \; .$$

As $n$ goes to infinity, and for $t$ larger than the cutoff time, this estimate tends to $\langle f, \pi\rangle = \sum_{j\in E} f(j)\pi(j)$. We define three families of *hitting times* (definitions 3.1, 3.2, and 3.3) as first entrance times of $S_i^{(t)}(f)$ in some intervals around the target value $\langle f, \pi\rangle$. Our main results, propositions 3.1 3.2 and 3.3, show that in general, these hitting times are asymptotically equivalent in probability to the cutoff time. Thus running the sample chain up to one of these hitting times is a way to make sure that the cutoff time has been reached.

## 3.1 Hitting times

Let $f$ be a function from $E$ into $\mathbb{R}$. Let $\{X(t)\} = \{(X_1(t), \ldots, X_n(t))\}$ be the parallel sample chain, starting with all coordinates equal to $i$. The empirical mean of $f$ at time $t$ is:

$$S_i^{(t)}(f) = \frac{1}{n}\sum_{m=1}^{n} f(X_m(t)) \; .$$

For all $t \in \mathbb{N}$, the random variable $S_i^{(t)}(f)$ has expectation:

$$\langle f, p_i^{(t)}\rangle = \sum_{j\in E} f(j) p_i^{(t)}(j) \; .$$

At time 0, it is equal to $f(i)$. By the law of large numbers, if $n$ is large, $S_i^{(t)}(f)$ stays close to $\langle f, p_i^{(t)}\rangle$ at all times. As $t$ tends to infinity, $\langle f, p_i^{(t)}\rangle$ tends to $\langle f, \pi\rangle$ at exponential speed, as shows (2.1). It is reasonable to stop the sample when $S_i^{(t)}(f)$ first meets $\langle f, \pi\rangle$.

**Definition 3.1** *Assume $f(i) < \langle f, \pi \rangle$. The* hitting time *associated to $i$ and $f$ is the following random variable $T_i(f)$:*

$$T_i(f) = \inf\{\, t \geq 0 \,:\, S_i^{(t)}(f) \geq \langle f, \pi \rangle \,\} \,.$$

In view of propositions 2.1 and 2.2, it is natural to expect that $T_i(f)$ should be close to the cutoff time. This is true when $\langle f, p_i^{(t)} \rangle$ is a monotone function of time. In definition 3.1, assuming that $f(i) < \langle f, \pi \rangle$ is not a real restriction. If it is not the case, one can replace $f$ by $-f$. The same remark holds true for Proposition 3.1: if $\langle f, p_i^{(t)} \rangle$ is a decreasing function of time, then $\langle -f, p_i^{(t)} \rangle$ is increasing. In what follows, we shall assume that all eigenvalues of $P$ are non negative. This is not a real restriction: if some of them are negative, then one can replace $P$ by $P^2$. This is the same as running two steps of the Markov chain in a row, or else as stopping the algorithm only at even times.

**Proposition 3.1** *Assume $i$ and $f$ are such that:*

- $w_i(f) = \displaystyle\sum_{\ell:\, |\alpha_\ell| = \alpha} \sum_{j \in E} f(j) \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} v_\ell(i)\, v_\ell(j) \neq 0 \,,$
- $\langle f, p_i^{(t)} \rangle$ *is an increasing 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 chain $\{X(t)\,;\, t \in I\!N\}$, starting from $X(0) = \tilde{\imath}$ will be denoted by $I\!P_{\tilde{\imath}}$. 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 $T_i(f) \leq t_0$ is the probability that at least one of the $S_i^{(t)}(f)$ is larger than $\langle f, \pi \rangle$, for $t = 0, \ldots, t_0$. Hence:

$$I\!P_{\tilde{\imath}}[T_i(f) \leq t_0] \leq \sum_{t=0}^{t_0} I\!P_{\tilde{\imath}}[S_i^{(t)}(f) \geq \langle f, \pi \rangle] \,.$$

For all $t$, $nS_i^{(t)}(f)$ is a sum of $n$ i.i.d. bounded random variables. By Hoeffding's inequality ([25] p. 58) one has for all $b > 0$:

$$I\!P_{\tilde{i}}[S_i^{(t)}(f) - \langle f, p_i^{(t)} \rangle \geq b] \leq \exp(-2nb^2/\mu^2) \ ,$$

where $\mu = \max_{j,k \in E} |f(j) - f(k)|$. This inequality is to be applied to

$$b = \langle f, \pi \rangle - \langle f, p_i^{(t)} \rangle \ .$$

For $t \leq t_0 \leq c \log(n)/(2 \log(1/\alpha))$, under the hypothesis that $w_i(f)$ is non null, one gets:

$$b \geq w_i(f) n^{-c/2} + o(n^{-c/2}) \ .$$

Hence

$$I\!P_{\tilde{i}}[S_i^{(t)}(f) \geq \langle f, \pi \rangle] \leq \exp\left( -\frac{w_i^2(f)}{2\mu^2} n^{1-c} + o(n^{1-c}) \right) \ .$$

Then

$$I\!P_{\tilde{i}}[T_i(f) \leq T_0] \leq \frac{\log n}{2 \log(1/\alpha)} \exp\left( -\frac{w_i^2(f)}{2\mu^2} n^{1-c} + o(n^{1-c}) \right) \ ,$$

which tends to 0 as $n$ tends to infinity for $c < 1$.

Let now $c$ be larger than 1. Let $t$, $s$ and $d$ three positive integers. Consider the $d$ regularly spaced instants $t+s, \ldots, t+ds$, and define $S$ to be the random vector

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

Our goal is to show that

$$I\!P_{\tilde{i}}[S \in ]-\infty, \langle f, \pi \rangle[^d]$$

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

$$J_m = (f(X_m^{(t+s)}), \ldots, f(X_m^{(t+ds)})) \ .$$

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

$$I\!E_{\tilde{i}}[J_m] = (\langle f, p_i^{(t+s)} \rangle, \ldots, \langle f, p_i^{(t+ds)} \rangle) \ .$$

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

$$C_{kh} = \sum_{j,j' \in E} f(j) f(j') p_i^{(t+ks)}(j) p_j^{((h-k)s)}(j') - \langle f, p_i^{(t+ks)} \rangle \langle f, p_i^{(t+hs)} \rangle \ .$$

It follows from (2.1) that, as $t$ and $s$ tend to infinity, $C$ tends to the diagonal matrix $v^2(f) I_d$, where $v^2(f)$ is the variance of $f$ under $\pi$.

$$v^2(f) = \sum_{j \in E} f^2(j) \pi(j) - \langle f, \pi \rangle^2 \ . \tag{3.1}$$

More precisely, let $\|.\|_\infty$ denote the maximum absolute value of all coefficients of a matrix. One has:

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

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 $t > (\log(n) + c_0)/(2 \log(1/\alpha))$:

$$\langle f, P_i^{(t+ks)} \rangle > \langle f, \pi \rangle - \frac{1}{\sqrt{n}} \ .$$

Then

$$\begin{aligned} & I\!P_i[S_i^{(t+ks)}(f) < \langle f, \pi \rangle \ , \ 1 \leq k \leq d \,] \\ \leq \ & I\!P_i[\sqrt{n}(S_i^{(t+ks)}(f) - \langle f, P_i^{(t+ks)} \rangle) < 1 \ , \ 1 \leq k \leq d \,] \ . \end{aligned}$$

The idea is to bound the above probability, using the probability of $]-\infty, 1[^d$ under the $d-$dimensional Gaussian distributions $\mathcal{N}_d(0, C)$ (mean 0, covariance matrix $C$) and $\mathcal{N}_d(0, v^2(f) I_d)$. 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 [10], p. 51 and Giné [17]. Using the inequalities given in the proof of lemmas 5.6 and 5.7 of [17], one gets:

$$\begin{aligned} I\!P_i[S \in ]-\infty, \langle f, \pi \rangle[^d] \ &\leq \ \mathcal{N}_d(0, C)(]-\infty, 1[^d) + \frac{k_3 d v^2(f)}{\sqrt{n}} \\ &\leq \ \mathcal{N}_d(0, v^2(f) I_d)(]-\infty, 1[^d) \\ &\qquad + \frac{k_3 d v^2(f)}{\sqrt{n}} + k_4(d)(\alpha^t + \alpha^s) \\ &= \ \rho^d + \frac{k_3 d v^2(f)}{\sqrt{n}} + k_4(d)(\alpha^t + \alpha^s) \ , \end{aligned}$$

where $\rho = \Phi(1/v(f)) < 1$, denoting by $\Phi$ 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 $n > n_1$, $k_3 dv^2(f)/\sqrt{n} < \varepsilon/3$. It remains to choose $t > (\log n + c_0)/(2\log(1/\alpha)$ and $s$ such that

$$I\!P_{\tilde{i}}[S \in ]-\infty, \langle f, \pi\rangle[^d] < \varepsilon \ .$$

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.

□

The crucial hypothesis in proposition 3.1 is that the expectation of $f$ under $p_i^{(t)}$ is an increasing function of $t$. This is the case if $-f$ is the indicator function of state $i$, and if all eigenvalues of $P$ are non-negative. But estimating the proportion of a single state is not sensible if the state space has a very large size.

In some particular cases, stochastic monotonicity can provide natural choices for $i$ and $f$. Assume the state space $E$ is endowed with a partial order and the Markov chain with matrix $P$ is stochastically monotone for that order (see Massey [22]). Let $i$ be a minimal element, and $f$ be an increasing function on $E$. Then $\langle f, p_i^{(t)}\rangle$ is an increasing function of time. This can be used for instance for birth and death chains or attractive spin systems.

Of course, in applications very little is known about $P$ and $\pi$, and the hypotheses of proposition 3.1 cannot be checked. Even worse, in general the expectation of a given function $f$ under $\pi$ cannot be computed. We shall propose two answers to that problem. The first one still uses monotonicity to define a *merging time* for two empirical expectations. The second one uses the notion of confidence interval. Both hitting times are asymptotically equivalent to cutoff time and thus capable of detecting the access to equilibrium. We shall state the corresponding convergence results without proofs (see [32]).

**Definition 3.2** *Let $i_1$ and $i_2$ be two elements of $E$ such that $f(i_1) < \langle f, \pi\rangle$ and $f(i_2) > \langle f, \pi\rangle$. Let $S_{i_1}^{(t)}(f)$ and $S_{i_2}^{(t)}(f)$ be the empirical expectations of $f$ computed over two independent samples of sizes $n_1$ and $n_2$, starting with all coordinates equal to $i_1$ and $i_2$ respectively. The* merging time *associated to $i_1$, $i_2$, and $f$ is the following random variable $T_{i_1 i_2}(f)$:*

$$T_{i_1 i_2}(f) = \inf\{\, t \geq 0 \,:\, S_{i_1}^{(t)}(f) \geq S_{i_2}^{(t)}(f) \,\} \ .$$

The advantage of merging times is that they do not require knowing the value of $\langle f, \pi\rangle$ for detecting the cutoff.

**Proposition 3.2** *Assume $i_1$ $i_2$ and $f$ are such that:*

- *At least one of the two quantities $w_{i_1}(f)$ and $w_{i_2}(f)$ is non null.*
- *$\langle f, p_{i_1}^{(t)} \rangle$ is an increasing function of $t$.*
- *$\langle f, p_{i_2}^{(t)} \rangle$ is a decreasing function of $t$.*

*Then*

$$T_{i_1 i_2}(f) \left( \frac{\log(n_1 + n_2)}{2 \log(1/\alpha)} \right)^{-1}$$

*converges to 1 in probability as $n_1$ and $n_2$ tend to infinity.*

Proposition 3.2 can be applied for instance in a case where stochastic monotonicity holds, if $f$ is increasing with respect to the order of the state space. Then one has to choose $i_1$ minimal and $i_2$ maximal in $E$. As will be shown in section 3.2, the merging time can give good results even when the chain is not stochastically monotone. In any case, it is reasonable to choose $i_1$ and $i_2$ such that $f(i_1)$ is minimal and $f(i_2)$ maximal.

In general, it can be feared that $\langle f, p_i^{(t)} \rangle - \langle f, \pi \rangle$ will change sign several times, so that the hitting time $T_i(f)$ will be significantly lower than cutoff. Of course it can also occur for merging times $T_{i_1 i_2}(f)$. This can be arranged by considering a new family of hitting times. We shall define $T_i^{(\epsilon)}(f)$ as the first time at which $S_i^{(t)}(f)$ enters a bilateral confidence interval of asymptotic level $\epsilon$ around $\langle f, \pi \rangle$.

**Definition 3.3** *Let $i$ be an element of $E$, and $f$ be a function from $E$ into $\mathbb{R}$. Let $\epsilon \in ]0,1[$. Let $\{X(t)\,;\, t \in \mathbb{N}\}$ be the sample chain starting from $X(0) = \tilde{\imath}$. We call* hitting time *associated to $i$ and $f$ at level $\epsilon$, the random variable*

$$T_i^{(\epsilon)}(f) = \inf \left\{ t \,:\, \frac{\sqrt{n} |S_i^{(t)}(f) - \langle f, \pi \rangle|}{v(f)} < \Phi^{-1} \left( \frac{1+\epsilon}{2} \right) \right\} ,$$

*where $v^2(f)$ is the variance of $f$ under $\pi$ and $\Phi$ denotes the standard Gaussian distribution function:*

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \ .$$

Once again, in practice, neither $\langle f, \pi \rangle$ nor $v^2(f)$ can be computed explicitly. But a confidence interval for $\langle f, \pi \rangle$ can be previously estimated over a reduced sample. The entrance time of $S_i^{(t)}(f)$ into that confidence interval is a reasonable stopping time for the full size sample. As can be expected, the analogue of proposition 3.1 holds for $T_i^{(\epsilon)}(f)$, under weaker hypotheses.

**Proposition 3.3** *Assume $i$ and $f$ are such that*

- $w_i(f) = \sum_{\ell:\,|\alpha_\ell|=\alpha} \sum_{j\in E} f(j) \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} v_\ell(i)\, v_\ell(j) \neq 0$ ,
- $\langle f, p_i^{(t)} \rangle \neq \langle f, \pi \rangle$ *for all* $t \in \mathbb{N}$.

*Then*

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

*converges to* 1 *in probability as* $n$ *tends to infinity.*

We believe that the cases where $T_i(f)$, $T_{i_1 i_2}(f)$, and $T_i^{(\epsilon)}(f)$ all fail to detect the cutoff are rarely encountered in practice. Running a sample chain up to a hitting or merging time can reasonably be expected to output a fair sample of the target distribution $\pi$.

Before turning to real implementations of the method in next sections, we shall present two small examples. Here the function $f$ will always be the indicator function of a subset $F$ of the state space $E$, and we shall denote $T_i(F)$ instead of $T_i(\mathbb{1}_F)$. 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 $E = \{1, 2, 3\}$.

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} . \tag{3.2}$$

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 $\epsilon$ are presented in Table 2. 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 $T_1^{(\epsilon)}(\{1\})$ are decreasing functions of $\epsilon$.

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

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 (3.2).

| $\epsilon$ | 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 |

Table 2: Estimated mean and standard deviations of $T_1^{(\epsilon)}(\{1\})$, for the parallel sample chain based on the matrix $P$ of (3.2).

In order to test the method with different sizes of state space, the natural random walk on $E = \{1, \ldots, \gamma\}$ was considered (Example 2.3.1 p. 348 of Saloff-Coste [29]). The transition matrix $P$ is the following.

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & \ldots & 0 \\ 0.5 & 0 & 0.5 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0.5 & 0 & 0.5 \\ 0 & \ldots & 0 & 0.5 & 0.5 \end{pmatrix} . \tag{3.3}$$

The uniform measure is reversible: $\pi(i) = 1/\gamma$ for all $i$. The spectral decomposition of $P$ is given explicitly in Feller [12], XVI.3. It turns out in this case that $\alpha = \alpha'$. We chose $i = 1$ and $F = \{1, \ldots, j\}$, 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 $\gamma$, 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.

| $\gamma\backslash j$ | 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 |

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 (3.3).

## 3.2 Random stable sets

Propositions 3.1, 3.2, and 3.3 are asymptotic results for a fixed state space and a sample size tending to infinity. In applications, the state spaces of interest are usually much larger than the size of the needed sample. In such a situation, no clear cutoff can be expected. One can only say that running the sample until a hitting or merging time as defined in section 3.1 is a minimal precaution, but there is no guaranty that the distribution of the sample will be close to equilibrium shortly after. In order to evaluate the method, we chose to apply it to large state spaces with a small sample size, by simulating samples of size only 100 of uniformly distributed random stable sets on line graphs.

The stable sets of an undirected graph (also called independent sets in many references) are subsets of vertices no two of which are connected (see [4] for a general reference). Let $G = (V, B)$ be a finite undirected graph without loops or multiple edges. If $x \in V$ is a vertex we denote by $N(x)$ the set of its neighbors on the graph.

$$N(x) = \{y \in S \ : \ \{x, y\} \in B\}.$$

For encoding convenience, stable sets can be identified to the corresponding $(0,1)$-incidence vectors, indexed by $V$ . These vectors will be referred to as *stable vectors* of the graph, and their set will be denoted by $I(G)$.

$$I(G) \ = \ \left\{ \eta \in \{0,1\}^V \ : \ \forall x \in V \quad \eta(x) = 1 \Longrightarrow \eta(y) = 0 \ \forall y \in N(x) \right\} \ .$$

It is easy to see that the number of stable sets of a graph with bounded degree grows exponentially with the number of vertices. In [15], it was shown that the generating function of the numbers of stable sets with $k$ elements could be computed in linear time when the graph $G$ is a Cartesian product of a subgraph by a line, a cycle or a tree. Here we shall consider only the case of the line graph with vertex set $V = \{1, \ldots, v\}$ and edge set $B = \{\, \{w, w+1\}\,,\ w = 1, \ldots, v-1\,\}$. Let $g_v(z)$ be the generating function of the number of stable sets with $k$ elements on that graph. The functions $g_v(z)$ can be computed explicitly using the following equation (see [15]).

$$g_v(z) = g_{v-1}(z) + z g_{v-2}(z)\ .$$

The number of elements of $E$ is $g_v(1)$. We shall consider two functions on $E$. The first one is the indicator function of stable sets containing the first vertex.

$$f_1(\eta) = 1 \text{ if } \eta(1) = 1\,,\ 0 \text{ else}\,.$$

The second one is the size of a stable set.

$$f_2(\eta) = \sum_{x=1}^{v} \eta(x)\ .$$

Let $\pi$ be the uniform distribution on $E$. The expectation of $f_1$ under $\pi$ is

$$\langle f_1, \pi \rangle = \frac{g_{v-2}(1)}{g_v(1)}\ .$$

That of $f_2$ is

$$\langle f_2, \pi \rangle = \frac{g'_v(1)}{g_v(1)}\ .$$

Table 4 gives the values of $|E|$ and $\langle f_2, \pi \rangle$ for $v$ ranging from 10 to 100. The value of $\langle f_1, \pi \rangle$ converges exponentially fast to $(3 - \sqrt{5})/2 = 0.381966$.

Let us now describe the simulation procedure for samples of random stable sets. The set $E$ is naturally endowed with a graph structure, inherited from that of the hypercube. Two stable vectors are neighbors in that graph if and only if they differ by exactly one coordinate. The symmetric random walk on that graph admits the uniform measure on $E$ as its reversible measure. It can be defined as the sequence of successive outputs of the following algorithm.

```
Initialize η
Repeat
  choose x at random in V
  If (η(x) = 1)
    then η(x) ⟵ 0
    else
      If (∀y ∈ V , η(y) = 1 ⇒ {x, y} ∉ B)
        then η(x) ⟵ 1
```

| $v$ | $\lvert E\rvert$ | $\langle f_2, \pi\rangle$ |
|---|---|---|
| 10 | 144 | 2.9167 |
| 20 | 17711 | 5.6807 |
| 30 | $2.1783\,10^6$ | 8.4446 |
| 40 | $2.6791\,10^8$ | 11.2085 |
| 50 | $3.2951\,10^{10}$ | 13.9724 |
| 60 | $4.0527\,10^{12}$ | 16.7364 |
| 70 | $4.9845\,10^{14}$ | 19.5003 |
| 80 | $6.1306\,10^{16}$ | 22.2642 |
| 90 | $7.5401\,10^{18}$ | 25.0282 |
| 100 | $9.2737\,10^{20}$ | 27.7921 |

Table 4: Number of stable sets and average size of a random stable set for the line graph with $v$ vertices.

*EndIf*
*EndIf*
*Until* (*stopping rule*)

The set of stable sets is naturally ordered by inclusion and the two functions $f_1$ and $f_2$ are increasing for that order. However, the symmetric random walk on $E$ is not stochastically monotone. We used the algorithm to generate samples of size $n = 100$, using merging times (Definition 3.2) as a stopping test. Apart from numerical evidence, we have no proof that the hypotheses of proposition 3.2 were satisfied in our implementation of the method.

In a first series of experiments, we based the test on $f_1$. The first 50 coordinates of the sample were initialized at the null vector $\eta_1(x) \equiv 0$. The second half of the sample was initialized at the vector $\eta_2$ having only its first coordinate equal to 1. Thus the initial values of $S^{(0)}_{\eta_1}(f_1)$ and $S^{(0)}_{\eta_2}(f_1)$ were 0 and 1 respectively. The simulation was stopped at the merging time $T_{\eta_1\eta_2}(f_1)$. In order to validate the obtained sample, the expectation of $f_2$ under $\pi$ was estimated by computing the mean size of the obtained stable sets at that time. The experiment was repeated 100 times for each $v = 10, 20, \ldots, 100$. Over the 100 repetitions, an estimate for the expectation and standard deviation of the merging time, and of the expected size were computed. The results are given in Table 5. One observes that even though the merging times are quite short, the expected size is correctly estimated, in spite of the very small size of the sample (compare with Table 4).

For the next series of experiments we switched the roles of $f_1$ and $f_2$. The merging time was based on $f_2$, and the sample at merging time was used to estimate the probability for the first vertex to be in the set: $\langle f_1, \pi\rangle$. Half of the sample was initialized with the null vector $\eta_1$. For the other half, we chose to initialize with the vector for

| $v$ | *merging time (mean)* | *merging time (st. dev.)* | *expected size (mean)* | *expected size (st. dev.)* |
|---|---|---|---|---|
| 10 | 21.00 | 8.70 | 2.92 | 0.11 |
| 20 | 39.32 | 18.34 | 5.63 | 0.17 |
| 30 | 60.65 | 25.83 | 8.36 | 0.18 |
| 40 | 81.90 | 32.67 | 11.13 | 0.21 |
| 50 | 102.59 | 39.96 | 13.86 | 0.23 |
| 60 | 129.12 | 49.88 | 16.61 | 0.27 |
| 70 | 136.36 | 58.18 | 19.25 | 0.41 |
| 80 | 148.28 | 56.86 | 22.04 | 0.42 |
| 90 | 187.13 | 67.89 | 24.86 | 0.39 |
| 100 | 203.90 | 77.24 | 27.55 | 0.50 |

Table 5: Experimental results for 100 samples of 100 stable sets on the line graph with $v$ vertices. Merging times and average size of the stable sets.

which the value of $f_2$ is maximal, $\eta_2(x) \equiv 1$, even though it is outside $E$. As before, the experiment was stopped at time $T_{\eta_1\eta_2}(f_2)$, and the proportion of elements of the sample for which the first coordinate was 1 at that time, was computed. This was repeated 100 times and an estimate for the expectation and standard deviation of the merging time, and of the proportion of first coordinates equal to 1, were computed. The results are those of Table 6. The merging times are longer than those of Table 5. The probability for the first vertex to be in the random stable set (theoretical value $\simeq 0.382$) is correctly estimated, with a small standard deviation over the 100 repetitions.

| $v$ | *merging time (mean)* | *merging time (st. dev.)* | *freq. of first vertex (mean)* | *freq. of first vertex (st. dev.)* |
|---|---|---|---|---|
| 10 | 22.40 | 4.74 | 0.382 | 0.047 |
| 20 | 46.87 | 8.89 | 0.381 | 0.051 |
| 30 | 74.93 | 13.64 | 0.381 | 0.046 |
| 40 | 103.15 | 21.10 | 0.381 | 0.049 |
| 50 | 128.58 | 21.75 | 0.379 | 0.046 |
| 60 | 160.87 | 23.66 | 0.381 | 0.051 |
| 70 | 188.19 | 29.03 | 0.377 | 0.046 |
| 80 | 216.77 | 34.82 | 0.375 | 0.047 |
| 90 | 251.17 | 39.18 | 0.380 | 0.048 |
| 100 | 284.23 | 49.59 | 0.377 | 0.049 |

Table 6: Experimental results for 100 samples of 100 stable sets on the line graph with $v$ vertices. Merging times and frequency of stable sets containing the first vertex.

## 3.3 Random Aboxes

This is our first 'real size' application of the method to a true MCMC problem. What is described here is the preliminary part of a joint work with M.C. Rousset ([11]).

The computational complexity of Description Logics (DL) reasoning problems has been extensively studied (see e.g. [24] for a recent survey). Many results on the worst-case time/space complexity for the corresponding inference algorithms are now available. Those results show that, under worst-case conditions, most operations of practical interest can only be performed in a computing time that grows exponentially fast with the size of the structure. However, despite their theoretical intractability, DL systems have been used for real applications. This suggests that worst-case examples are not realistic. In [11], the problem of generating *random benchmarks* for DL systems was addressed. The idea of generating random benchmarks is to perform experimentally probabilistic analyses for DL algorithms (see [18]).

In DL a Tbox defines the logical rules of a given context of application by propositions, called concepts, and binary relations, called roles, together with possible logical interactions between them (implications, exclusions...). It can be seen as the rule of the game. With one Tbox can be associated an Abox, once a set of objects (the players) has been precised. Such an Abox is a set of facts of one of two types, concept-facts and role-facts. If $C$ is a concept in the Tbox and $o$ an object, the concept-fact $C(o)$ belongs to the Abox if $C$ is true for $o$. Similarly a role-fact $R(o, o')$ may belong to the Abox if $o$ and $o'$ are actually related by the role $R$. The Abox is said to be admissible if none of the rules is violated by the conjunction of its facts. The stable sets of a given graph can be seen as particular cases of Aboxes. Here the part of the TBox is played by the initial graph, with no role, concepts placed at the vertices and exclusion relations corresponding to the edges. An Abox for a single object can be seen as a stable set of the graph, if $C(o)$ is interpreted as the vertex $C$ being in the set.

Let $\mathcal{T}$ be a fixed Tbox. Let $\mathcal{A}$ be the set of admissible Aboxes relative to $\mathcal{T}$ and a fixed set of objects. Let $c$ and $r$ be the numbers of concepts and roles in $\mathcal{T}$, and $k$ be the number of objects. The set $\mathcal{A}$ is finite, with a potentially large cardinality, actually bounded above by $2^{ck+rk^2}$. Owing to that combinatorial explosion, the only way of generating uniformly distributed random Aboxes is a MCMC algorithm. As in the case of stable sets, the idea is to define an undirected graph structure, the vertices of which are the elements of the set to be sampled. The symmetric random walk on an undirected connected graph admits the uniform distribution on the vertex set as its unique asymptotic distribution. In the graph structure that we define on $\mathcal{A}$, edges connect pairs of admissible Aboxes differing by a single fact. A priori, the number of possible neighbors of a given Abox is $ck + rk^2$. Due to possible incompatibilities or redundancies, the degree of a general Abox can be notably smaller. This is reflected in the algorithm, through the admissibility test. The algorithm for generating a uniform sample of $n$ admissible Aboxes relative to a given Tbox and a given set of individuals is the following:

```
Initialize [A_1, ..., A_n]
Repeat
  For i = 1 to N:
    Choose concept-fact
        with probability ck/(ck + rk^2)
      or role-fact
        with probability rk^2/(ck + rk^2)
    If concept-fact Then
      Choose a concept expression C
        with probability 1/c, and
      Choose an individual o
        with probability 1/k
      If C(o) ∈ A_i Then A_i ⟵ A_i \ {C(o)}
        Else If A_i ∪ {C(o)} admissible
        Then A_i ⟵ A_i ∪ {C(o)}
    If role-fact Then
      Choose a role R
        with probability 1/r and
      Choose a couple of individuals (o, o')
        with probability 1/k^2
      If R(o, o') ∈ A_i Then A_i ⟵ A_i \ {R(o, o')}
        Else If A_i ∪ {R(o, o')} admissible
        Then A_i ⟵ A_i ∪ {R(o, o')}
  EndFor
Until Stopping test.
```

In a first series of experiments, we initialized the sample with empty Aboxes and we used as a stopping test the hitting time $T_\emptyset(f)$ associated to the indicator function $f$ of the set of Aboxes containing a fixed fact. In a second series of experiments we used the technique of merging times. Half of the sample was initialized with the empty Abox, the other half with identical Aboxes, all containing the target fact. The function $f$ was the same as before. Both series of experiments gave coherent results for running times and statistical properties of the output samples. On a small Tbox, the obtained sample of Aboxes was compared with the output of a single trajectory after a very long running time (single chain method). The results were again coherent.

Many more experiments still have to be run to validate the method on cases of practical interest. We believe that the hitting times method really can help cutting the simulation time and still give satisfactory outputs.

# 4 Cutoff on countable state spaces

This section is part of a joint work with S. Martinez [21], to which we refer for the proofs. Contrarily to the case of sample chains, the state space is a fixed, infinite countable set denoted by $I$. The semigroup $(P^{(t)})$ of the (irreducible positive recurrent) Markov chain is also fixed. We consider exclusively continuous time Markov chains $\{X(t)\,;\, t \geq 0\}$, although we believe that the results of this section can also be written for the discrete case. The generator is still denoted by $\Lambda$. Our chain will start from a fixed state $a \in I$, its distribution is still denoted by $I\!P_a$. The cutoff phenomenon can occur as $a$ tends to infinity.

Here the notion of hitting time appears even more naturally than in sample chains. The hitting time of $b \in I$ is:

$$T_b = \inf\{t \geq 0 \,:\, X(t) = b\}\ .$$

Assume some particular state called 0, is transformed into an absorbing state, so the new generator $\Lambda^{(0)}$ coincides with $\Lambda$ except at row 0, for which $\lambda^{(0)}_{0,b} = 0$ for any $b \in I$. We denote by $P_0^{(t)}$ its semi-group and by $\{X_0(t)\}$ the absorbed chain: $X_0(t) = X(t)$ if $t < T_0$ and $X_0(t) = 0$ if $t \geq T_0$. The total variation distance $\|\delta_a P_0^{(t)} - \delta_0\|$ is equal to the survival function out of 0 starting from $a$,

$$\|\delta_a P_0^{(t)} - \delta_0\| = I\!P_a[T_0 > t]\ .$$

In this particular case, the chain with stationary measure $\delta_0$ has a cutoff at time $(t(a))$ in the sense of Definition 1.1 if and only if the hitting time of 0 is equivalent in probability to $(t(a))$ i.e.

$$\forall \varepsilon > 0, \quad \lim_{a\to\infty} I\!P_a\left[\left|\frac{T_0}{t(a)} - 1\right| > \varepsilon\right] = 0.$$

Observe that this condition implies $\lim_{a\to\infty} I\!E_a\left[\frac{T_0}{t(a)}, \frac{T_0}{t(a)} \leq K\right] = 1$ for any $K > 1$, so cutoff can be expected at time $(t(a)) \sim (I\!E_a[T_0])$, in situations where these expectations exist. In those cases, cutoff means that the hitting time is asymptotically equivalent in probability to its expectation. This can be interpreted as a weak law of large numbers for $T_0$. It is a natural idea to compare the access time to equilibrium with the first hitting time of a given fixed state. It turns out that cutoff occurs for both at about the same time. This is coherent with what had been observed on quite different hitting times in section 3.1.

## 4.1 General result

Proposition 4.1 below shows the equivalence between cutoffs for the access to equilibrium and for the hitting time of a particular state. Actually here, the choice of state 0 is not very important, as we point out in the remark at the end of this section, it can be replaced by any finite subset of the state space.

**Proposition 4.1** *Let $(t(a) : a \in I)$ be a sequence of positive reals tending to infinity as $a$ tends to infinity. The chain $\{X(t)\}$ with stationary measure $\pi$ has a cutoff at time $(t(a))$ if and only if the chain $\{X_0(t)\}$, absorbed at 0, has a cutoff at the same time.*

**Proof.**

In what follows we always take $c > 0$ and we put $t_a = ct(a)$. First, we will prove the following:

$$\forall c < 1, \lim_{a\to\infty} I\!P_a[T_0 < t_a] = 0 \Longleftrightarrow \forall c < 1, \lim_{a\to\infty} \|\delta_a P^{(t_a)} - \pi\| = 1.$$

Assume first the lefthand side. Fix $\varepsilon > 0$ and a finite subset $I_\varepsilon$ of $I$ containing 0, such that $\pi(I_\varepsilon) > 1 - \varepsilon$. We have:

$$\|\delta_a P^{(t)} - \pi\| \geq |I\!P_a[X(t) \in I_\varepsilon] - \pi(I_\varepsilon)| \, .$$

If we can prove that $I\!P_a[X(t_a) \in I_\varepsilon] < \varepsilon$, for $a$ outside some finite subset of $I$, this will finish the proof by implying $\lim_{a\to\infty} \|\delta_a P^{(t_a)} - \pi\| > 1 - 2\varepsilon$. Now

$$I\!P_a[X(t) \in I_\varepsilon] \leq I\!P_a[\inf_{b\in I_\varepsilon} T_b \leq t] \leq \sum_{b\in I_\varepsilon} I\!P_a[T_b \leq t] \, .$$

Then it suffices to show that for any $b$, $I\!P_a[T_b \leq t_a]$ tends to 0 as $a$ tends to infinity. One can write $T_0 \leq T_b + T_0 \circ \theta_{T_b}$ , where $\theta_T$ is the shift operator with respect to (random) time $T$. For $\xi > 0$ take $K = K(\xi)$ such that $I\!P_b[T_0 \geq K] \leq \xi$. Using the strong Markov property we get

$$\begin{aligned}
& I\!P_a[T_b \leq t_a] \\
&= I\!P_a[T_0 \leq (T_0 - T_b) + t_a] \\
&\leq I\!P_a[T_0 \leq T_0 \circ \theta_{T_b} + t_a] \\
&\leq I\!P_a[T_0 \leq K + t_a, T_0 \circ \theta_{T_b} \leq K] + I\!P_a[T_0 \circ \theta_{T_b} \geq K] \\
&\leq I\!P_a[T_0 \leq K + t_a] + I\!P_b[T_0 \geq K] \\
&\leq I\!P_a\left[\frac{T_0}{t(a)} \leq \frac{K + t_a}{t(a)}\right] + \xi \leq 2\xi,
\end{aligned}$$

for $a \notin I_\xi$ for some finite $I_\xi$, because $K/t(a) \underset{a\to\infty}{\longrightarrow} 0$, $t_a/t(a) = c < 1$. We conclude that $\lim_{a\to\infty} \|\delta_a P^{(t_a)} - \pi\| = 1$.

Conversely, assume $\|\delta_a P^{(t_a)} - \pi\|$ tends to 1. Consider $I\!P_0[X(u) = 0]$. It is a continuous function of $u$, starting from 1 at $u = 0$ and tending to $\pi(0) > 0$ as $u$ tends to infinity. It remains bounded away from 0. Let $m > 0$ be its minimum. Denote by $G_a(u) = I\!P_a[T_0 \leq u]$ the distribution function of $T_0$ starting from $a$. We have:

$$\begin{aligned}
I\!P_a[X(t_a) = 0] &= \int_0^\infty I\!P_a[X(t_a) = 0 \,|\, T_0 = u] \, dG_a(u) \\
&\geq \int_0^{t_a} I\!P_a[X(t_a) = 0 \,|\, T_0 = u] \, dG_a(u) \\
&\geq m I\!P_a[T_0 \leq t_a] \, .
\end{aligned}$$

Therefore it is sufficient to prove that $I\!P_a[X(t_a) = 0]$ tends to 0 as $a$ tends to infinity. Fix $\varepsilon$ such that $0 < \varepsilon < \pi(0)/2$. For all $a$ outside a finite subset of $I$, $\|\delta_a P^{(t_a)} - \pi\| > 1 - \varepsilon$. Hence there exists $I(a) \subset I$ such that $|I\!P_a[X(t_a) \in I(a)] - \pi(I(a))| > 1 - 2\varepsilon$. Replacing possibly $I(a)$ by $I \setminus I(a)$, one can assume:

$$I\!P_a[X(t_a) \in I(a)] < 2\varepsilon \quad \text{and} \quad \pi(I(a)) > 1 - 2\varepsilon \ .$$

Since $2\varepsilon < \pi(0)$, $I(a)$ must contain 0, so:

$$I\!P_a[X(t_a) = 0] < 2\varepsilon \ ,$$

hence the result.

We now turn to prove the equivalence

$$\forall c > 1, \lim_{a\to\infty} I\!P_a[T_0 \le t_a] = 1 \Longleftrightarrow \forall c > 1, \lim_{a\to\infty} \|\delta_a P^{(t_a)} - \pi\| = 0.$$

Assume first the lefthand side. Fix $\varepsilon > 0$. From the assumption, there exists a finite set $I_\varepsilon \subset I$ such that for $s_a = \frac{1+c}{2} t(a)$ we have

$$I\!P_a[T_0 \le s_a] \ge 1 - \varepsilon \ , \quad \forall a \notin I_\varepsilon \ .$$

Let $I' \subset I$. Still denoting by $G_a(u)$ the distribution function of $T_0$ starting from $a$, we have:

$$\begin{aligned}
& |I\!P_a[X(t_a) \in I'] - \pi(I')| \\
= \ & |\int_0^\infty I\!P_a[X(t_a) \in I' \,|\, T_0 = u] \, dG_a(u) - \pi(I')| \\
\le \ & \int_0^\infty |I\!P_a[X(t_a) \in I' \,|\, T_0 = u] - \pi(I')| \, dG_a(u) \\
\le \ & \int_0^{s_a} |I\!P_a[X(t_a) \in I' \,|\, T_0 = u] - \pi(I')| \, dG_a(u) + \varepsilon \\
\le \ & \int_0^{s_a} |I\!P_0[X(t_a - u) \in I'] - \pi(I')| \, dG_a(u) + \varepsilon \\
\le \ & \int_0^{s_a} \|\delta_0 P^{(t_a - u)} - \pi\| dG_a(u) + \varepsilon.
\end{aligned}$$

The last bound being independent of $I'$. We have shown:

$$\|\delta_a P^{(t_a)} - \pi\| \le \int_0^{s_a} \|\delta_0 P^{(t_a - u)} - \pi\| dG_a(u) + \varepsilon.$$

Now for $u \in [0, s_a]$ we have $t_a - u \ge t_a - s_a \to \infty$ as $a \to \infty$, then $\|\delta_0 P^{(t_a-u)} - \pi\| \to 0$ as $a \to \infty$ . By Lebesgue's dominated convergence theorem we conclude $\lim_{a\to\infty} \|\delta_a P^{(t_a)} - \pi\| = 0$.

Conversely, assume $\|\delta_a P^{(t_a)} - \pi\|$ tends to 0. Let $\varepsilon > 0$ and fix $I_\varepsilon$ such that $\pi(I_\varepsilon) > 1 - \varepsilon$. Then for $a$ outside some finite subset of $I$,

$$\mathbb{P}_a[X(s_a) \in I_\varepsilon] > 1 - 2\varepsilon \ ,$$

with $s_a = \frac{1+c}{2} t(a)$ as before. Now:

$$\begin{aligned} \mathbb{P}_a[T_0 \leq t_a] &\geq \sum_{b \in I_\varepsilon} \mathbb{P}_a[T_0 \leq t_a \, ; \, X(s_a) = b] \\ &\geq \sum_{b \in I_\varepsilon} \mathbb{P}_b[T_0 \leq t_a - s_a] \mathbb{P}_a[X(s_a) = b] \ . \end{aligned}$$

As $a$ tends to infinity, $t_a - s_a$ tends to infinity and $\mathbb{P}_b[T_0 \leq t_a - s_a]$ tends to 1. Hence, outside a finite subset of $I$,

$$\mathbb{P}_a[T_0 \leq t_a] \geq (1 - \varepsilon) \mathbb{P}_a[X(s_a) \in I_\varepsilon] \geq (1 - \varepsilon)(1 - 2\varepsilon).$$

Hence the result.

□

**Remark 4.1** Let $B \subset I$ be a finite subset of the state space. The hitting time of $B$, $T_B = \min_{b \in B} T_b$, has cutoff at time $(t(a))$ if

$$\forall h > 0, \quad \lim_{a \to \infty} \mathbb{P}_a \left[ \left| \frac{T_B}{t(a)} - 1 \right| > h \right] = 0 \ .$$

From the proof of Proposition 4.1 it follows that $T_B$ has cutoff at the same time $(t(a))$ as $T_0$, and that cutoff for $T_B$ is also equivalent to cutoff for access to equilibrium.

For a particular chain, cutoff for hitting times will in general be easier to check than cutoff for access to equilibrium. In order to prove that $T_0$ is equivalent in probability to its expectation, it is enough to check that the variance of the ratio tends to zero. Denote $Var_a[f] = \mathbb{E}_a[f^2] - (\mathbb{E}_a[f])^2$.

**Corollary 4.1** *Assume*

$$\lim_{a \to \infty} \mathbb{E}_a[T_0] = \infty \ \textit{and} \ \lim_{a \to \infty} Var_a \left[ \frac{T_0}{\mathbb{E}_a[T_0]} \right] = 0 \ . \tag{4.1}$$

*then $X(t)$ admits a cutoff at time $(\mathbb{E}_a[T_0])$.*

In next section, we shall study a situation where both the expectation and the variance of $T_0$ can be explicitly computed.

## 4.2 Birth and death chains on trees

In this section, we consider a particular type of Markov chains that includes birth and death chains on the integers. Their transition graph is defined on a tree. A tree is an unoriented connected graph $G = (I, F)$ with no non trivial closed loop. The choice of one of the vertices as a root, denoted by 0, defines a partial ordering of the set of vertices $I$, for which the root is a minimal element. That ordering is defined as follows. For any $a$ and $b$ in $I$, $a$ is before $b$ ($a \preceq b$) if $a = b$ or if there exists a path without repetition from $b$ to 0 containing $a$. For any vertex $a \neq 0$, there exists a unique vertex, denoted by $p(a)$ (the parent of $a$) such that $p(a) \preceq a$ and $\{a, p(a)\} \in F$. The set of vertices $b$ such that $a = p(b)$ (children of $a$) is denoted by $\sigma(a)$. It is supposed to be finite. There exists a unique path in the tree joining $a$ to 0. Its length will be denoted by $d(a)$ (the depth of $a$) and the set of vertices on that path (except 0) will be denoted by $\ell(a)$.

$$\ell(a) = \{a, p(a), p^2(a), \ldots, p^{d(a)-1}(a)\} .$$

The branch stemming from $a$, denoted by $B_a$ is the set of vertices in the tree after $a$,

$$B_a = \{b \,:\, a \preceq b\} .$$

Given such a tree, a Markov generator is defined by two transition rates for each edge. Let $(\lambda_a : a \in J)$ and $(\mu_a : a \in J)$ be two families of positive reals, indexed by the vertices of $J = I \setminus \{0\}$. The Markov generator $\Lambda$ is defined on $I$ as follows: for all $a \in J$, the transition rate from $p(a)$ to $a$ is $\lambda_a$, that from $a$ to $p(a)$ is $\mu_a$. All other rates are null. If the tree is finite, the probability measure $\pi = (\pi(a) : a \in I)$ defined by:

$$\pi(a) = \pi(0) \prod_{b \in \ell(a)} \frac{\lambda_b}{\mu_b} , \ \forall a \in J , \tag{4.2}$$

is the unique stationary measure for $\Lambda$, which is self-adjoint with respect to $\pi$. In the infinite case, we will assume that the Kolmogorov equations $P_t' = P_t\Lambda$ define a transition semi-group such that $P_t(I) = 1$ for all $t$. We shall not discuss the conditions on $\Lambda$ for this to happen (see for instance [13]). We will also assume positive recurrence, so that (4.2) still defines the unique stationary measure of $\Lambda$. Proposition 4.2 below shows that for chains whose stationary measure $\pi$ has exponential decaying tails, Corollary 4.1 applies and a cutoff occurs at time ($I\!E_a[T_0]$). For this we shall use explicit expressions for the expectation and variance of the travelling time through the edges of the tree. They are given in lemma 4.1 below and we refer to [21] for the proofs.

**Lemma 4.1** *Let $X(t)$ be a positive recurrent Markov chain with generator $\Lambda$. For $a \in J$ we have*

$$I\!E_a[T_{p(a)}] = \frac{\pi(B_a)}{\pi(a)\mu_a} , \tag{4.3}$$

$$I\!E_a[T^2_{p(a)}] = \frac{1}{\pi(a)\mu_a} \sum_{c \in B_a} \frac{2(\pi(B_c))^2}{\pi(c)\mu_c} . \tag{4.4}$$

The first hitting time of 0 starting from $a \in J$ is the sum of the travelling times along the edges of the path $\ell(a)$. Its distribution is the sum of $d(a)$ independent random variables distributed respectively as $T_{p(b)}$, starting from $b$, where $b$ runs in $\ell(a)$. The following result can be seen as a weak law of large numbers for $T_0$ (seen as a sum of travelling times through the edges of the tree), as well as a cutoff result.

**Proposition 4.2** *Consider a positive recurrent chain with generator $\Lambda$ on an infinite tree. Assume that its stationary measure has an exponentially decaying tail, in the sense that $(\pi(B_a)/\pi(a) \,:\, a \in I)$ is bounded. Assume $(1/\mu_a \,:\, a \in I)$ is also bounded, and*

$$\lim_{d(a)\to\infty} \sum_{b\in\ell(a)} \frac{1}{\mu_b} = +\infty \,.$$

*Then*

$$\lim_{d(a)\to\infty} I\!E_a[T_0] = \infty, \quad \lim_{d(a)\to\infty} Var_a\left(\frac{T_0}{I\!E_a[T_0]}\right) = 0.$$

*In particular $X(t)$ admits a cutoff at time $(I\!E_a[T_0])$.*

**Proof.**

Using (4.3) one has:

$$I\!E_a[T_0] = \sum_{b\in\ell(a)} I\!E_b[T_{p(b)}] = \sum_{b\in\ell(a)} \frac{\pi(B_b)}{\pi(b)\mu_b} \geq \sum_{b\in\ell(a)} \frac{1}{\mu_b} \,.$$

Thus $I\!E_a[T_0]$ tends to infinity as $d(a)$ increases.

Denote by $K$ a uniform upperbound of both $(\pi(B_a)/\pi(a) \,:\, a \in I)$ and $(1/\mu_a(a) \,:\, a \in I)$. The variance of $T_0$ starting from $a$ is

$$Var_a[T_0] = \sum_{b\in\ell(a)} I\!E_b[T^2_{p(b)}] - (I\!E_b[T_{p(b)}])^2 \,.$$

Using (4.4) and our hypotheses, one gets:

$$\begin{aligned}
\sum_{b\in\ell(a)} I\!E_b[T^2_{p(b)}] &= \sum_{b\in\ell(a)} \frac{2}{\pi(b)\mu_b} \sum_{c\in B_b} \frac{\pi^2(B_c)}{\pi(c)\mu_c} \\
&\leq \sum_{b\in\ell(a)} \frac{2}{\pi(b)\mu_b} \sum_{c\in B_b} K^2 \frac{\pi(c)}{\mu_c} \\
&\leq \sum_{b\in\ell(a)} 2K^3 \frac{\pi(B_b)}{\pi(b)\mu_b} \\
&= 2K^3 I\!E_a[T_0] \,.
\end{aligned}$$

Notice that $Var_a[T_0]$ is necessarily finite. Actually:

$$Var_a\left[\frac{T_0}{I\!E_a[T_0]}\right] = \frac{Var_a[T_0]}{(I\!E_a[T_0])^2} \leq \frac{2K^3}{I\!E_a[T_0]} \longrightarrow 0 \text{ as } d(a) \longrightarrow \infty \ .$$

The cutoff follows from Corollary 4.1 (observe that $d(a) \to \infty$ is equivalent to $a \to \infty$, since all vertices have finite degree).

□

## 4.3 Birth and death chains on the line

Proposition 4.2 applies in particular to many classical birth and death chains (cf. for instance [5]), among them the $M/M/s$ and $M/M/\infty$ queues.

Consider first the $M/M/s$ queue. It has Poisson input with constant rate $\lambda > 0$, and $s$ servers working independently, each with exponential service times with parameter $\mu > 0$. The condition $\lambda < s\mu$ ensures positive recurrence. The number of customers in the system is a birth and death chain for which the birth rates are constant and equal to $\lambda$, the death rate from $a+1$ to $a$ is $(a+1)\mu$ if $a \leq s-1$, $s\mu$ else. Assume $\lambda < s\mu$, to ensure positive recurrence. From Propositions 4.2 and 4.1, the hitting time of 0 and the access time to equilibrium are equivalent to the expectation of the first, if the initial number of customers in the queue is large. Assuming that the initial number is $a$, by formula (4.3) the expectation of $I\!E_a[T_0]$ is equivalent to $a/(s\mu - \lambda)$. This result can be interpreted as follows. While the number of customers in the file is larger than $s$, the $s$ servers work non stop and perform on average $s\mu$ services by time unit. Meanwhile, $\lambda$ new clients come in on average. Thus it takes about $a/(s\mu - \lambda)$ time units to evacuate the first $a$ customers, or equivalently to reach equilibrium.

We shall examine here in more details the $M/M/\infty$ queue. It is a birth and death chain on $I\!N$ for which the birth rate is constant, the death rate being proportional to the state. More precisely, the transition rate from $a$ to $a+1$ is $\lambda$, and that from $a+1$ to $a$ is $(a+1)\mu$, for all $a \in I\!N$, $\lambda$ and $\mu$ being positive constants. It is positive recurrent for all $\lambda$ and $\mu$, and the stationary measure $\pi$ is the Poisson distribution with parameter $\lambda/\mu$, denoted by $\pi = \mathcal{P}(\lambda/\mu)$. It is a well known fact that if the initial distribution is Poisson with parameter $\rho_0$, then for each $t \geq 0$, the distribution of $X(t)$ is $\mathcal{P}(\rho_t)$ where:

$$\rho_t = \frac{\lambda}{\mu} + (\rho_0 - \frac{\lambda}{\mu})e^{-\mu t} \ .$$

Due to propositions 4.2 and 4.1, the hitting time of 0 and the access time to equilibrium are equivalent to the expectation of the first, if the initial state $a$ is large. Using (4.3), that expectation is easily proved to be equivalent to $\log a/\mu$. Actually in this case a much sharper cutoff result can be obtained, if the initial distribution of the chain is Poisson with a large parameter. (see [21] for the proof).

**Proposition 4.3** *Let $(P^{(t)})$ be the semigroup of the $M/M/\infty$ queue.*

1. *There exists $c_0 > 0$ such that if $c > c_0$, $n > 2\lambda/\mu$, and $t = t(n) < (\log n - c)/\mu$, then*
$$\|\mathcal{P}(n)P_t - \pi\| \geq 1 - 2e^{-\lambda/\mu} \exp(-e^c/4) \; .$$
2. *For all $c > 0$, if $n > \lambda/\mu$ and $t = t(n) > (\log n + c)/\mu$, then*
$$\|\mathcal{P}(n)P_t - \pi\| \leq \left(1 + \frac{\mu}{\lambda}\right) e^{-c} \; .$$

Proposition 4.3 can be naturally related to a particular case of continuous sample chain. With the notations of section 2.2, take $E = \{1, 2\}$ and let $\Lambda$ be the generator:

$$\Lambda = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \; ,$$

with $\lambda$ and $\mu$ strictly positive. Condider the continuous sample chain $\{\tilde{X}(t)\}$. It follows from proposition 2.7 that $\{\tilde{X}(t)\}$ has a cutoff at time $\log n/(2(\lambda + \mu))$. Let $N_2(t)$ be the number of coordinates of $\tilde{X}(t)$ in state 2 at time $t$. Then $\{N_2(t)\}$ is a birth and death chain on $\{0, \ldots, n\}$ with birth rate (from $a$ to $a+1$) $(n - a)\lambda$ and death rate (from $a+1$ to $a$) $(a+1)\mu$. If its distribution at time 0 is binomial $\mathcal{B}(n, p(0))$, then it remains binomial at any time $t$, with parameters $n$ and $p(t)$:

$$p(t) = \frac{\lambda}{\lambda + \mu} + \left(p(0) - \frac{\lambda}{\lambda + \mu}\right) e^{-(\lambda+\mu)t} \; .$$

That birth and death chain also has a cutoff at time $\log n/(2(\lambda + \mu))$. Replace now $\lambda$ by $\lambda/n$, and let $n$ tend to $\infty$. Then $\{N_2(t)\}$ converges in distribution to the $M/M/\infty$ queue, that has a cutoff at time $\log n/\mu$. A closely related birth and death chain on $\{0, \ldots, n\}$ is the $M/M/n/n$ queue. Cutoff for that chain has been studied by Fricker et al. [16].

# 5 Cutoff for Jackson networks

Queueing networks have now found applications in many areas, and particular in the performance analysis of computer systems (see Robertazzi [28]). Due to the complexity of the models, the best that can be searched for in many cases is the stationary measure. Here again, knowing when exactly the stationary regime will be reached has important consequences in applications, let them be Monte-Carlo simulations or statistical studies.

We shall examine here the convergence to equilibrium of closed and open Jackson networks. A Jackson network (see for instance Kelly [20]) is made of $\gamma$ queues with Markovian services. For $i = 1, \ldots, \gamma$, the service rate of the $i$-th queue is $\mu_i(n_i)$ when

$n_i$ customers are present in the file. Customers coming out of queue $i$ are routed immediately to queue $j$ with probability $r_{ij}$.

In a closed Jackson network the total number $n$ of customers in the system is fixed. The routing matrix $(r_{ij})$ is stochastic and assumed to be irreducible and aperiodic. The process under study is the $\gamma$-tuple of numbers of customers in the different queues. Denote it by $\{N(t)\} = \{(N_1(t), \ldots, N_\gamma(t))\}$. Its state space is:

$$I = \{(n_1, \ldots, n_\gamma) \, : \, n_i \in I\!N, \sum_{i=1}^{\gamma} n_i = n\} \; .$$

With appropriate independence assumptions, it is a continuous time Markov chain, for which coordinate $i$ is decreased and coordinate $j$ increased by 1 with rate $\mu_i(n_i)r_{ij}$. The stationary measure has the following product form:

$$\pi(n_1, \ldots, n_\gamma) = c \prod_{i=1}^{\gamma} \prod_{m=1}^{n_i} \frac{\lambda_i}{\mu_i(m)} \; , \tag{5.1}$$

where $c$ is the normalizing constant and the $\lambda_i$'s are a solution of the following flow balance equations.

$$\lambda_i = \sum_{j=1}^{\gamma} \lambda_j r_{ji} \; , \; i = 1, \ldots, \gamma \; .$$

In an open Jackson network, customers enter the system from outside, according to a Poisson process with rate $\lambda$. Incoming customers are routed to queue $i$ with probability $r_{0i}$. Coming out of queue $i$ (with rate $\mu_i(n_i)$ as before) they are routed to queue $j$ with probability $r_{ij}$ or outside the system with probability $r_{i\gamma+1}$. The process is the numbers of customers per file at time $t$, still denoted by $N(t) = (N_1(t) \; \ldots, N_\gamma(t))$. With appropriate independence assumptions, $\{N(t)\}$ is a continuous time Markov chain with state space $I\!N^\gamma$. The transitions are those of a closed network, plus arrivals (increment coordinate $i$ with rate $\lambda r_{0i}$) and departures (decrement coordinate $i$ with rate $\mu_i(n_i)r_{i\gamma+1}$). The stationary distribution, if it exists, has the following product form.

$$\pi(n_1, \ldots, n_\gamma) = \pi(0, \ldots, 0) \prod_{i=1}^{\gamma} \prod_{m=1}^{n_i} \frac{\lambda_i}{\mu_i(m)} \; , \tag{5.2}$$

where the $\lambda_i$'s are the stationary flows through the files, solution of the balance equations:

$$\lambda_i = \lambda r_{0i} + \sum_{j=1}^{\gamma} \lambda_j r_{ji} \; , \; i = 1, \ldots, \gamma \; .$$

Closed as well as open networks will be initialized with a total of $n$ customers, proportionally distributed over the $\gamma$ files (compare with Propositions 2.4 and 2.7). Thus $(q_1, \ldots, q_\gamma)$ will be a fixed positive probability distribution on $\{1, \ldots, \gamma\}$, and $n_1(n), \ldots, n_\gamma(n)$ will be functions from $I\!N$ into $I\!N$ such that:

$$\sum_{i=1}^{\gamma} n_i(n) = n \quad \text{and} \quad \lim_{n \to \infty} \frac{n_i(n)}{n} = q_i \; , \; i = 1, \ldots, \gamma \; . \tag{5.3}$$

With that initialization, we will be interested in cutoff properties as $n$ tends to infinity.

## 5.1 Closed networks

The cutoff for closed Jackson networks was studied in several papers by Anantharam (see [1]). He considered first the case of exponential servers (constant rates $\mu_i(n_i) \equiv \mu_i$), then generalized it to parallel exponential servers in fixed numbers. If the $i$-th file ends with $s_i$ parallel servers with rate $\mu_i$ ($\cdot/M/s_i$ services), the exit rate for that file will be:

$$\begin{aligned} \mu_i(n_i) &= n_i\mu_i \quad \text{if } n_i \leq s_i \\ &= s_i\mu_i \quad \text{if } n_i \geq s_i \ . \end{aligned}$$

Anantharam considered closed networks with a large number $n$ of customers. With bounded service rates it is intuitively clear, and easy to deduce from (5.1), that the stationary distribution tends to saturate the weakest file, namely that for which the throughput ratio $\lambda_i/(s_i\mu_i)$ is the smallest. At equilibrium, most customers will be waiting in that file. Anantharam proves that a cutoff occurs in the sense of definition 1.1 for the convergence to equilibrium, at time $Kn$, where the constant $K$ depends on the network.

The case of many parallel servers is quite different. Assume each customer immediately finds an exponential server with rate $\mu_i$, when entering file $i$ ($\cdot/M/\infty$ services). The exit rate for that file will be $\mu_i(n_i) = n_i\mu_i$. At equilibrium there is no saturation, and a sizable proportion of all customers occupies each file: the asymptotic distribution $\pi$ is multinomial with parameters $n$ and $\lambda_1/\mu_1, \ldots, \lambda_\gamma/\mu_\gamma$, once the $\lambda_i$'s have been properly scaled. But the cutoff phenomenon is still present. No wonder about it: this is a mere re-writing of the continuous sample chain setting of section 2.2. For each couple $(i,j)$, $i \neq j$, define $\lambda_{ij} = \mu_i r_{ij}$. Let $\Lambda$ be the corresponding generator on $E = \{1, \ldots, \gamma\}$. Let $\{\tilde{X}(t)\}$ be the continuous sample chain of size $n$ with generator $\tilde{\Lambda}$ on $E^n$. For each $t \geq 0$ and $i = 1, \ldots, \gamma$, define $N_i(t)$ as the number of coordinates in state $i$:

$$N_i(t) = \sum_{m=1}^{n} \mathbb{1}_{\{i\}}(X_m(t)) \ .$$

Then $\{N(t)\} = \{(N_1(t), \ldots, N_\gamma(t))\}$ is distributed as the Jackson network under study. In other words, since each one of the $n$ customers moves between the files according to generator $\Lambda$, independently of the others, one can see the file each customer is in, as one of the coordinates of the sample chain. Let $\beta$ be the smallest among absolute values of eigenvalues of $\Lambda$. Denote by $p_n^{(t)}$ the distribution of the chain, starting with $n$ customers, distributed according to (5.3). Proposition 2.7 can be rephrased as:

**Proposition 5.1** *Let $c(n)$ be any function from $\mathbb{N}$ into $\mathbb{R}^+$, tending to infinity as $n$ tends to infinity. Define:*

$$t^-(n) = \max\{0\,,\ \frac{\log n}{2\beta} - c(n)\} \quad \textit{and} \quad t^+(n) = \frac{\log n}{2\beta} + c(n) \ .$$

*Then*

$$\lim_{n\to\infty} \|p_n^{(t^-(n))} - \pi\| = 1 \quad \textit{and} \quad \lim_{n\to\infty} \|p_n^{(t^+(n))} - \pi\| = 0 \ .$$

## 5.2 Open networks

The case of open Jackson networks seems quite different since the state space is infinite. However, applying the techniques of section 4, one can see that cutoffs occur at times comparable to the corresponding closed network cases: $O(n)$ for $\cdot/M/s$ servers, and $O(\log n)$ for $\cdot/M/\infty$ servers.

The latter case is the simplest. Assume the exit rate of file $i$ is $n_i\mu_i$, when $n_i$ customers are in it. The system is always positive recurrent, its stationary measure $\pi$ is the product of $\gamma$ Poisson distributions, that of file $i$ having parameter $\lambda_i/\mu_i$ (notations of (5.2)). Denote by $\Lambda$ the generator on $\{1,\dots,\gamma+1\}$, with transition rates $\lambda_{ij} = \mu_i r_{ij}$. Once in the system, each customer is independent of the others, going from file $i$ to file $j$ with rate $\lambda_{ij}$, and being eventually absorbed outside (state $\gamma+1$). Let $\beta$ be the smallest among absolute values of eigenvalues of $\Lambda$. Denote by $I\!P(t)$ the distribution of the chain, starting with $n$ customers, distributed according to (5.3). The cutoff time is the same as that of the closed case, up to a factor 2.

**Proposition 5.2** *The open Jackson network with $\cdot/M/\infty$ servers has a cutoff at time $(t(n))$, with*

$$t(n) = \frac{\log n}{\beta} \ .$$

**Proof.**

As already noticed, all customers have independent trajectories in the system, before they leave. Denote by $F_i(t)$ the distribution function of the sojourn time in the system of a customer initially present in file $i$. These distribution functions can be explicitly computed from the solutions of the Kolmogorov system based on generator $\Lambda$. They have exponentially decaying tails, the decay rate of each one being one of the eigenvalues of $\Lambda$. At least one of them has decay rate $\beta$. Assume it is the case for the $i$-th file. Out of $n_i$ customers initially present in file $i$, the last one leaves the system after a time which is the maximum of $n_i$ i.i.d. random variables with distribution function $F_i$. Let $T_i$ be that time. One can easily check that:

$$\frac{T_i}{\log n_i/\beta}$$

converges to 1 in probability. Recall that $n_i \sim nq_i$. So $\log n/\beta$ is the time at which all $n$ customers initially present in the system have left it. To finish the proof and apply proposition 4.1, one has to prove that the time it

takes to empty the system after that, is negligible compared to $\log n$. This comes immediately from positive recurrence.

□

The case of $\cdot/M/s$ servers is more complicated, but the result is coherent with that of Anantharam [1], p. 84. One major difference with the previous case is that the cutoff time may depend on the initial distribution of customers $(q_i)$.

**Proposition 5.3** *Assume the open Jackson network with $\cdot/M/s$ servers is positive recurrent. There exists a constant $K$, depending on the parameters and the initial distribution of customers $(q_i)$ such that the network has a cutoff at time $(Kn)$.*

Here we shall not give a detailed proof but rather a few indications on the computation of $K$. Initially, all files have $O(n)$ customers, hence all servers are working non-stop. The $i$-th file outputs customers at rate $s_i\mu_i$ per time unit, but inputs them at rate $\lambda r_{0i} + \sum_j s_j\mu_j r_{ji}$. Denote by $\nu_{1,i}$ the drift of file $i$.

$$\nu_{1,i} = \lambda r_{0i} + \sum_{j=1}^{\gamma} s_j\mu_j r_{ji} - s_i\mu_i \ .$$

Since the network is positive recurrent, the global rate $\sum_i \nu_{1,i}$ has to be negative. Let

$$K_1 = -\min_i \frac{q_i}{\nu_{1,i}} > 0 \ .$$

At a time which is equivalent in probability to $K_1 n$, the 'fastest' file, say $i_1$, has emptied down to $O(1)$ customers, the others still having $O(n)$ customers. From that time on, file $i_1$ will be 'quasi-balanced', its output flow being equal to its input flow, and not to $s_i\mu_i$ anymore. The other files will evolve linearly in time, with drifts $\nu_{2,i}$. At least one of them will decrease down to $O(1)$ customers, at a time $K_2 n$. After that, drift rates change again. This goes on until the 'slowest' file has emptied down, after at most $\gamma$ such periods. The constant $K$ of the proposition above is the sum of the constants $K_1$, $K_2$,... determining the asymptotic length of each period.

In both propositions 5.2 and 5.3, the hypothesis that no file should have few customers $(q_i > 0)$ is not essential. Both results state that the cutoff time is essentially what it takes to evacuate all customers initially present in the system. We believe this remains true whatever the initial distribution of customers.

In order to illustrate proposition 5.3, let us consider the example of two tandem $M/M/1$ queues. Customers arrive in the first file at rate $\lambda$. They wait there until they get service at rate $\mu_1$, then they go to the second file, receive another service with rate $\mu_2$ at their turn, then leave the system. Assume positive recurrence, i.e. $\lambda < \min\{\mu_1, \mu_2\}$. Assume moreover that the two servers are different $(\mu_1 \neq \mu_2)$.

Suppose there are initially $n$ customers in the first file, none in the second one. It takes $n/(\mu_1 - \lambda)$ time units to empty the first file. Now two cases are possible.

*Case 1:* $\mu_1 < \mu_2$. Customers can be evacuated by the second file faster than they arrive. The cutoff time is equivalent to that of emptying the first file: $n/(\mu_1 - \lambda)$.

*Case 2:* $\mu_1 > \mu_2$. While the first file empties itself, the second one cannot cope with the flow. At time $n/(\mu_1 - \lambda)$, it has served only $\mu_2 n/(\mu_1 - \lambda)$ customers, thus $(\mu_1 - \mu_2)n/(\mu_1 - \lambda)$ are still there. It will take

$$\frac{\mu_1 - \mu_2}{\mu_1 - \lambda} \frac{n}{\mu_2 - \lambda}$$

time units to evacuate them. The global cutoff time for emptying the system (and reaching equilibrium), will be:

$$\frac{n}{\mu_1 - \lambda} + \frac{\mu_1 - \mu_2}{\mu_1 - \lambda} \frac{n}{\mu_2 - \lambda} = \frac{n}{\mu_2 - \lambda} \, .$$

This can easily be generalized to $\gamma$ tandem files: if $n$ customers are initially present in the first one, the cutoff time is essentially the time it takes to get all of them through the slowest file.

# References


[1] V. Anantharam. Threshold phenomena in the transient behaviour of Markovian models of communication networks and databases. *Queueing Systems*, 5:77-98, 1989.

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

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

[4] G. Chartrand and L. Lesniak. *Graphs and Digraphs. Second ed.* Wadsworth & Brooks/Cole, Mathematics Series, 1986.

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

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

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

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

[9] D. Chauveau and J. Diébolt. MCMC convergence diagnostic via the central limit theorem. *Research report 22/97, Université de Marne la Vallée, 1997.*

[10] P. Doukhan. *Mixing, properties and examples.* L.N. in Stats. **85**. Springer-Verlag, New York, 1994.

[11] Q. Elhaik, M.C. Rousset, and B. Ycart. Generating random benchmarks for description logics. *Proceedings of the 1998 Int. Workshop on Description Logics (DL'98),* 95–99, Trento, 1998.

[12] W. Feller. *An introduction to probability theory and its applications*, volume I. Wiley, London, 1968.

[13] W. Feller. *An introduction to probability theory and its applications*, volume II. Wiley, London, 1971.

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

[15] F. Forbes and B. Ycart. Counting stable sets on Cartesian products of graphs. *Discrete Mathematics*, 186:105–116, 1998.

[16] C. Fricker, P. Robert, and D. Tibi. On the rates of convergence of Erlang's model. *Research Report INRIA 3368*, 1998.

[17] E. Giné. Lectures on some aspects of the bootstrap. In *Ecole d'été de probabilités de Saint-Flour XXVI*, P. Bernard (ed.), L.N. in Math. **1664**, 37–151. Springer-Verlag, New York, 1997.

[18] M. Hofri. *Probabilistic analysis of algorithms.* Springer-Verlag, New York, 1987.

[19] J. Keilson. *Markov chain models - rarity and exponentiality.* Applied Mathematical Sciences **28**. Springer-Verlag, New York, 1979.

[20] F.P. Kelly. *Reversibility and stochastic networks*, Wiley, London, 1979.

[21] S. Martinez and B. Ycart. Decaying rates and cutoff for convergence and hitting times of Markov chains. *Submitted*, 1998.

[22] A.W. Massey. Stochastic orderings for Markov processes on partially ordered spaces. *Math. Oper. Research*, 12(2):350–367, 1987.

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

[24] D. Nardi, F.M. Donini, M. Lenzerini, and A. Schaerf. Reasoning in description logics. In *Principles of Artificial Intelligence.* G. Brewska (ed.), Springer-Verlag, New York, 1995.

[25] V.V. Petrov. *Sums of independent random variables*. Springer-Verlag, Berlin, 1975.

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

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

[28] T.G. Robertazzi. *Computer networks and systems: queuing theory and performance evaluation*. Springer-Verlag, New York, 1990.

[29] L. Saloff-Coste. Lectures on finite Markov chains. In *Ecole d'été de probabilités de Saint-Flour XXVI*, P. Bernard (ed.), L.N. in Math. **1664**, 301–413. Springer-Verlag, New York, 1997.

[30] A. Sinclair. *Algorithms for random generation and counting: a Markov chain approach*. Birkhäuser, Boston, 1993.

[31] B. Ycart. Cutoff for samples of Markov chains. *ESAIM Probability-Statistics* 3:89–107, 1999.

[32] B. Ycart. Stopping tests for Monte-Carlo Markov chain methods. *Meth. and Comp. in Applied Probab.*, 2(1):23-36, 2000.