ALEA, Lat. Am. J. Probab. Math. Stat. 11 (2), 445–458 (2014)

Bounds for left and right window cutoffs

Javiera Barrera and Bernard Ycart

Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez.
Av. Diagonal las Torres 2640
Peñalolén, Santiago, Chile.
E-mail address: javiera.barrera@uai.cl
URL: http://jbarrera.uai.cl/

Laboratoire Jean Kuntzmann, Univ. Grenoble-Alpes
51 rue des Mathématiques 38041
Grenoble cedex 9, France.
E-mail address: Bernard.Ycart@imag.fr
URL: http://ljk.imag.fr/membres/Bernard.Ycart

Dedicated to the memory of Béatrice Lachaud

Abstract. The location and width of the time window in which a sequence of processes converges to equilibrum are given under conditions of exponential convergence. The location depends on the side: the left-window and right-window cutoffs may have different locations. Bounds on the distance to equilibrium are given for both sides. Examples prove that the bounds are tight.

1. Introduction

The term "cutoff" was introduced by Aldous and Diaconis (1986), to describe the phenomenon of abrupt convergence of shuffling Markov chains. Many families of stochastic processes have since been shown to have similar properties: see Levin et al. (2009, Chap. 8) for an introduction to the subject, Saloff-Coste (2004) for a review of random walk models in which the phenomenon occurs, and Chen and Saloff-Coste (2008) for an overview of the theory. Consider a sequence of stochastic processes in continuous time, each converging to a stationary distribution. Denote by $d_n(t)$ the distance between the distribution at time t of the n-th process and its stationary distribution, the 'distance' having one of the usual definitions (total variation, separation, Hellinger, relative entropy, L^p, etc.). The phenomenon can be

Received by the editors December 27, 2013; accepted August 3, 2014.
1991 *Mathematics Subject Classification.* 60J25.
Key words and phrases. cutoff; exponential ergodicity.
J. Barrera was partially supported by grants Anillo ACT88, Fondecyt n°1100618, and Basal project CMM (Universidad de Chile). B. Ycart was supported by Laboratoire d'Excellence TOUCAN (Toulouse Cancer).

expressed at three increasingly sharp levels (more precise definitions will be given in section 2).

(1) The sequence has a cutoff at (t_n) if $d_n(ct_n)$ tends to the maximum M of the distance if $c < 1$, to 0 if $c > 1$.
(2) The sequence has a window cutoff at (t_n, w_n) if $\liminf d_n(t_n + cw_n)$ tends to M as c tends to $-\infty$, and $\limsup d_n(t_n + cw_n)$ tends to 0 as c tends to $+\infty$.
(3) The sequence has a profile cutoff at (t_n, w_n) with profile F if $F(c) = \lim d_n(t_n + cw_n)$ exists for all c, and F tends to M at $-\infty$, to 0 at $+\infty$.

There are essentially two ways to interpret the cutoff time t_n: as a mixing time Levin et al. (2009, Chap. 18), or as a hitting time Martínez and Ycart (2001). These two interpretations are equivalent for the total variation and separation distances where couplings times can characterize both distances. Also the cutoff definition can be applied directly to hitting times, see Connor (2010). For samples of Markov chains, the latter interpretation can be used to determine explicit online stopping times for MCMC algorithms Ycart (2000); Lachaud (2005); Lachaud and Ycart (2006); Diédhiou and Ngom (2009).

Sequences of processes for which an explicit profile can be determined are scarce. The first example of a window cutoff concerned the random walk on the hypercube for the total variation distance; it was treated by Diaconis and Shahshahani shortly after the introduction of the notion in 1987. The profile cutoff result was proved by Diaconis et al. (1990). Cutoffs for random walks on more general products or sums of graphs have been investigated in Ycart (2007), and more recently by Miller and Peres (2012). Random walks on the hypercube can be interpreted as samples of binary Markov chains. Diaconis et al.'s results were generalized to samples of continuous and discrete time finite state Markov chains for the chi-squared and total variation distance in Ycart (1999), then to samples of more general processes, for four different distances in Barrera et al. (2006, section 5) and for separation distance in Connor (2010) (see also Levin et al. (2009, Chap. 20)). Other examples of profile cutoffs include the riffle shuffle for the total variation distance by Bayer and Diaconis (1992), and birth and death chains for the separation distance by Diaconis and Saloff-Coste (2006) or the total variation distance by Ding et al. (2010). When the maximum M of the distance is 1 (total variation, separation), the profile F decreases from 1 to 0. Thus it can be seen as the survival function of some probability distribution on the real line. A Gaussian distribution has been found for the riffle shuffle with the total variation distance in Bayer and Diaconis (1992, Theorem 2) or for some birth and death chain with the separation distance in Diaconis and Saloff-Coste (2006, Theorem 6.1). A Gumbel distribution has been found for samples of finite Markov chains and the total variation distance in Diaconis et al. (1990); Ycart (1999). For the Hellinger, chi-squared, or relative entropy distances, other profiles were obtained in Barrera et al. (2006).

Explicit profiles are usually out of reach, in particular for the total variation distance: only a window cutoff can be hoped for. However the definition above, which is usually agreed upon (Chen and Saloff-Coste (2008, Definition 2.1) or Levin et al. (2009, p. 218)), may not capture the variety of all possible situations. As will be shown here, the location of a left-window cutoff should be distinguished from that of a right-window cutoff: see Figure 18.2, p. 256 of Levin et al. (2009). The main result of this note, Theorem 2.2, expresses the characteristics of the left and right windows in terms of a decomposition into exponentials of the distances $d_n(t)$. It refines some of the results in Chen and Saloff-Coste (2010), in particular Theorem 3.8. Explicit bounds on the distance to equilibrium are given. They are proved to be tight, using examples of cutoffs for Ornstein-Uhlenbeck processes (see Lachaud (2005)).

The paper is organized as follows. Section 2 contains formal definitions and statements. Examples are given in section 3. Theorem 2.2 is proved in section 4.

2. Definitions and statements

For each positive integer n a stochastic process $X_n = \{X_n(t)\,;\ t \geqslant 0\}$ is given. We assume that $X_n(t)$ converges in distribution to ν_n as t tends to infinity. The convergence is measured by one of the usual distances (total variation, separation, Hellinger, relative entropy, L^p, etc.), the maximum of which is denoted by M ($M = 1$ for total variation and separation, $M = +\infty$ for relative entropy, chi-squared…). The distance between the distribution of $X_n(t)$ and ν_n is denoted by $d_n(t)$.

Definition 2.1. Denote by (t_n) and (w_n) two sequences of positive reals, such that $w_n = o(t_n)$. They will be referred to as respectively as *location* and *width*. The sequence (X_n) has:

(1) a left-window cutoff at (t_n, w_n) if:

$$\lim_{c\to-\infty} \liminf_{n\to\infty} \inf_{t<t_n+cw_n} d_n(t) = M\ ;$$

(2) a right-window cutoff at (t_n, w_n) if:

$$\lim_{c\to+\infty} \limsup_{n\to\infty} \sup_{t>t_n+cw_n} d_n(t) = 0\ ;$$

(3) a profile cutoff at (t_n, w_n) with profile F if:

$$\forall c \in \mathbb{R}\,,\ F(c) = \lim_{n\to\infty} d_n(t_n + cw_n)$$

exists and satisfies:

$$\forall c \in \mathbb{R}\,,\ 0 < F(c) < M \quad \text{and} \quad \lim_{c\to-\infty} F(c) = M\ , \quad \lim_{c\to+\infty} F(c) = 0\ .$$

If both left- and right-window cutoffs hold for the same location t_n and width w_n, then a (t_n, w_n)-cutoff holds in the sense of Definition 2.1 in Chen and Saloff-Coste (2008). The location and width are not uniquely determined. If a left-window cutoff holds at location t_n, it also holds at any location t'_n such that $t'_n \leqslant t_n$. Symmetrically, if a right-window cutoff holds at location t_n, it also holds at any location t'_n such that $t'_n \geqslant t_n$. Moreover, if a cutoff holds for width w_n, it also holds for any width w'_n such that $w'_n \geqslant w_n$. The location and width of a left-window cutoff will be said to be optimal if for any $c < 0$:

$$\liminf_{n\to\infty} \inf_{t<t_n+cw_n} d_n(t) < M\ .$$

Those of a right-window cutoff are optimal if for any $c > 0$:

$$\limsup_{n\to\infty} \sup_{t>t_n+cw_n} d_n(t) > 0\ .$$

This corresponds to strong optimality in the sense of Chen and Saloff-Coste (2008, Definition 2.2). Of course, if a profile cutoff holds, then the left- and right-window cutoffs hold at the same location and width, which are optimal for both. Examples will be given in section 3.

Our definition can be useful to describe the convergence of chains that do not show a cutoff but instead have a multiple step convergence. An example due to D. Aldous, is illustrated on Figure 18.2, p. 256 of Levin et al. (2009). For that example, one can check that there is a $((15+5/3)n, \sqrt{n})$ left-window cutoff and a $((15+6)n, \sqrt{n})$ right-window cutoff. So our definition can be seen as a refined version of pre-cutoff (compare with Section 18.1 of that same book). Furthermore, when there is a cutoff, definition 2.1 can capture different asymptotic behaviors to the left and right of the cutoff instant. More importantly, distinguishing locations and widths on both sides may permit to capture the exact profile of convergence. This will be illustrated by the following example (Figure 2.1), which is a modification of the one studied by Lacoin (2014). Consider a continuous Markov chain (X_n)

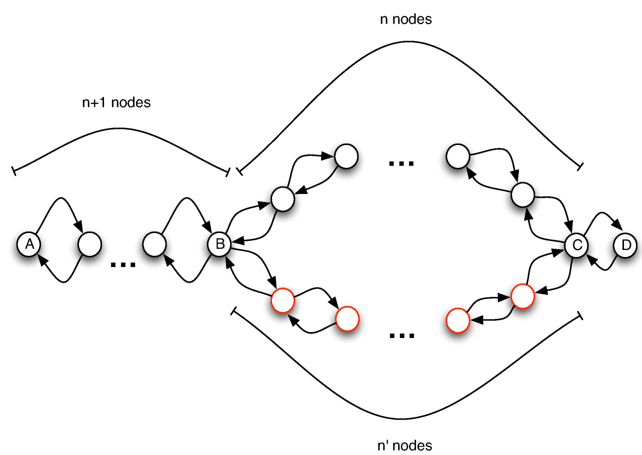

FIGURE 2.1. The Markov chain is composed of three birth and death chains and an extra node as in the figure. The first birth and death chain goes from node A to node B: it has $n+1$ nodes, it has birth rates 1 and death rates 2^{-n}. From B to C, two birth and death chains are considered, each starting from B, and leading to C. The upper branch has n nodes (including B) with birth rate 1 except for node B which has birth rate $1/2$ and with death rate $2^{-n-\lceil n^{\beta}\rceil}$. The lower branch (nodes in red) has $n' = n+\lceil n^{\beta}\rceil$ nodes and also has birth rate 1 except for node B which has birth rate $1/2$ and death rate 2^{-n}. Finally, at the right of the chain there is node D and there we have birth rate 1 from C to D and death rate 2^{-n^3} from D to C.

the transitions of which are those of Figure 2.1. The chain is reversible and the equilibrium measure is easily calculated: it gives weight $1 - O(2^{-n^3})$ to state D. When the chain starts from state A, hitting D is almost equivalent to reaching stationarity in total variation distance. More precisely if T_n is the hitting time of D starting from A, then $d_n(t) = \mathcal{P}(T_n \geqslant t) + o(1)$. Arguing as Lacoin (2014), as death rates are small, with probability tending to one, (X_n) does not backtrack before $t = 4n$. So (X_n) reaches D for the first time using either the upper branch or the lower branch, each with probability 1/2. In the first case T_n is the sum of $2n$ exponential random variables and in the second case T_n is the sum of $2n + \lceil n^\beta \rceil$ exponential random variables. From Chebychev's inequality, it can be concluded that there is cutoff at time $(T_n) = 2n + n^\beta$. Yet more precise estimates are obtained using the central limit theorem:

$$\begin{array}{lcl} \lim_{n\to\infty} \mathbb{P}(T_n > 2n + n^\beta + c(2n + n^\beta)^{1/2}) & = & \frac{1}{2}(1 - \Phi(c)) \text{ for } c > 0, \\ \lim_{n\to\infty} \mathbb{P}(T_n > 2n - c(2n)^{1/2}) & = & \frac{1}{2}(1 - \Phi(c)) + \frac{1}{2} \text{ for } c > 0. \end{array}$$

where Φ is the cumulative distribution of the standard normal distribution. For $1/2 < \beta < 1$, and using the total variation distance, the right-window cutoff is at $(t_n^+, w_n^+) = (2n + n^\beta, (n + n^\beta/2)^{1/2})$ and the left-window cutoff is at $(t_n^-, w_n^-) = (2n, n^{1/2})$. For $0 < \beta \leqslant 1/2$, left and right window are the same, but for $\beta > 1/2$, they are different.

Our main result relates the location and width of the left- and right-window cutoffs to the terms of a decomposition into exponentials of the functions $d_n(t)$. From now on, we assume $M = +\infty$: the distance is relative entropy, L^p for $p > 1$, etc. The result is expressed for a sequence of continuous time processes, it could be written in discrete time, at the expense of heavier notations.

Theorem 2.2. *Assume that for each n, there exist an increasing sequence of positive reals $(\rho_{i,n})$, and a sequence of non negative reals $(a_{i,n})$ with $a_{1,n} > 0$, such that:*

$$d_n(t) = \sum_{i=1}^{+\infty} a_{i,n}\, \mathrm{e}^{-\rho_{i,n} t} \ . \tag{2.1}$$

Denote by $A_{i,n}$ the cumulated sums of $(a_{i,n})$, truncated to values no smaller than 1.

$$A_{i,n} = \max\{1, a_{1,n} + \cdots + a_{i,n}\} \ .$$

For each n, define:

$$t_n = \sup_i \frac{\log(A_{i,n})}{\rho_{i,n}} \ , \tag{2.2}$$

$$w_n = \frac{1}{\rho_{1,n}} \ , \tag{2.3}$$

$$r_n = w_n \left(\log(\rho_{1,n} t_n) - \log(\log(\rho_{1,n} t_n))\right) \ . \tag{2.4}$$

Assume that:

(1) *for n large enough,*

$$0 < t_n < +\infty \ , \tag{2.5}$$

(2)

$$\lim_{n\to\infty} \rho_{1,n} t_n = +\infty \ , \tag{2.6}$$

(3) *there exists a positive real α such that for n large enough, and for all $i \geqslant 2$,*

$$a_{i,n} \leqslant \alpha A_{i-1,n} \ . \tag{2.7}$$

Then (X_n) has a left-window cutoff at (t_n, w_n), a right-window cutoff at $(t_n + r_n, w_n)$. More precisely:

$$\forall c < 0 \ , \quad \liminf_{n\to\infty} d_n(t_n + c w_n) \geqslant \mathrm{e}^{-c} \ , \tag{2.8}$$

$$\forall c > 0 \ , \quad \limsup_{n\to\infty} d_n(t_n + r_n + c w_n) \leqslant \mathrm{e}^{-c} \ . \tag{2.9}$$

Conditions (2.5) and (2.7) are technical. Condition (2.6) is known as Peres criterion: Chen and Saloff-Coste (2008) have proved that it implies cutoff for L^p distances with $p > 1$, and given a counterexample for the L^1 distance. A consequence is that $w_n = o(t_n)$ as requested by Definition 2.1, and more precisely that $w_n = o(r_n)$ and $r_n = o(t_n)$.

A decomposition into exponentials of the distance to equilibrium such as (2.1) holds for many processes: functions of finite state space Markov chains, functions of exponentially ergodic Markov processes, etc. Assuming that the decomposition only has non-negative terms is a stronger requirement: see Chen and Saloff-Coste (2010, section 4). It implies that $d_n(t)$ is a decreasing function of t. We do not view it as a limitation. Indeed, if (2.1) has negative terms, it can be decomposed as $d_n(t) = d_n^+(t) - d_n^-(t)$, with:

$$d_n^+(t) = \sum_{i=1}^{+\infty} \max\{a_{i,n}, 0\}\, \mathrm{e}^{-\rho_{i,n} t} \quad \text{and} \quad d_n^-(t) = -\sum_{i=1}^{+\infty} \min\{a_{i,n}, 0\}\, \mathrm{e}^{-\rho_{i,n} t} \,. \tag{2.10}$$

and apply our main result to each term to obtain the behavior of d_n as in the following corollary:

Corollary 2.3. *If $d_n(t)$ has negative terms, let $d_n^+(t)$ and $d_n^-(t)$ be as in (2.10). Assume that Theorem 2.2 applies to both $d_n^+(t)$ and $d_n^-(t)$, leading to left-window cutoffs at (t_n^+, w_n^+) and (t_n^-, w_n^-), right-window cutoffs at $(t_n^+ + r_n^+, w_n^+)$ and $(t_n^- + r_n^-, w_n^-)$. Then for all n, $w_n^- < w_n^+$, and*

$$\limsup_{n\to\infty} \frac{t_n^-}{t_n^+} \leqslant 1 \,.$$

Moreover if $\limsup_{n\to\infty} t_n^-/t_n^+ < 1$ then equations (2.8) and (2.9) remain valid for $d_n(t)$ with $(t_n, r_n, w_n) = (t_n^+, r_n^+, w_n^+)$.

Proof: The fact that $d_n(t)$ is nonnegative implies that the leading term of $d_n^+(t)$ is larger than the leading term of $d_n^-(t)$. Therefore for all n, $\rho_{1,n}^+ < \rho_{1,n}^-$, hence $w_n^+ > w_n^-$. For the same reason, (2.8) and (2.9) imply:

$$\forall c > 0 \lim_{n\to\infty} d_n^+(ct_n^+) = 0 \Longrightarrow \lim_{n\to\infty} d_n^-(ct_n^+) = 0 \,,$$

hence $\limsup_{n\to\infty} t_n^-/t_n^+ \leqslant 1$.

If $\limsup_{n\to\infty} t_n^-/t_n^+ < 1$ then for n large enough, $t_n^- < t_n^+$, and $t_n^- + r_n^- < t_n^+ + r_n^+$. Hence,

$$\lim_{n\to\infty} d_n^-(t_n^+ + cw_n^+) = \lim_{n\to\infty} d_n^-(t_n^+ + r_n^+ + cw_n^+) = 0.$$

As a consequence,

$$\begin{aligned} \liminf_{n\to\infty} d_n(t_n^+ + cw_n^+) &= \liminf_{n\to\infty} d_n^+(t_n^+ + cw_n^+) \,, \\ \limsup_{n\to\infty} d_n(t_n^+ + r_n^+ + cw_n^+) &= \limsup_{n\to\infty} d_n^+(t_n^+ + r_n^+ + cw_n^+) \,. \end{aligned}$$

□

Theorem 3.8 of Chen and Saloff-Coste (2010) is related to Theorem 2.2 above: it describes a (t_n, r_n)-cutoff but, as we shall see in the following result, r_n should be seen as a correction on the location of the window. A combination of the hypotheses of the two theorems leads to the next result where a more detailed description of the convergence is obtained:

Corollary 2.4. *Consider $d_n(t)$ as defined in equation 2.1. For $C>0$ let $j_n^C := \inf\{i : A_{i,n} > C\}$, and let $d_n^C(t)$ be defined by*

$$d_n^C(t) := A_{j_n,n} e^{-t\rho_{j_n,n}} + \sum_{i=j_n+1}^{m_n} a_{i,n} e^{-t\rho_{i,n}}.$$

Assume Theorem 2.2 applies to $d_n^C(t)$, and denote by $t_n(C)$, $r_n(C)$, and $w_n(C)$ the corresponding times. Assume moreover that there exist $C>0$ and $\epsilon>0$ such that:

$$\lim_{n\to\infty} \sum_{i=1}^{j_n^C-1} e^{-\epsilon t_n \rho_{i,n}} a_{i,n} = 0 \tag{2.11}$$

Then d_n also satisfies inequalities (2.9) and (2.8) for $t_n(C)$, $r_n(C)$ and $w_n(C)$.

Proof: Let $S_1(t) = \sum_{i=1}^{j_n^C-1} e^{-t\rho_{i,n}} a_{i,n}$. It is easy to see that for any $C>0$

$$d_n^C(t) \le d_n(t) \le S_1(t) + d_n^C(t).$$

The left-window inequality (2.8) is straightforward. To obtain the right-window inequality (2.8) we need to prove that

$$\lim_{n\to\infty} S_1(t_n(C) + r_n(C) + cw_n(C)) = 0.$$

Corollary 3.3 in Chen and Saloff-Coste (2010) says that if there exists $\epsilon>0$ for which equation (2.11) is valid, then it is valid for any $\epsilon>0$, in particular for $\epsilon=1$. Therefore

$$\lim_{n\to\infty} S_1(t_n(C) + r_n(C) + cw_n(C)) \le \lim_{n\to\infty} \sum_{i=1}^{j_n^C-1} e^{-t_n(C)\rho_{i,n}} a_{i,n} = 0.$$

□

Equation 2.11 is the same as equation (b) in Theorem 3.8 of Chen and Saloff-Coste (2010); equation (a) in Theorem 3.8 is the same as condition (2) in Theorem 2.2 when we apply it to d_n^C. Finally the existence of α stated in condition (3) in Theorem 2.2 is not requested in Theorem 3.8 of Chen and Saloff-Coste (2010). Therefore, as already said, this Corollary can be viewed as a combination of both results. Theorem 3.8 states that there is $(t_n(C), r_n(C))$-cutoff. Here, a $(t_n(C), w_n(C))$ left-cutoff is obtained, which is more precise because $w_n(C) = o(r_n(C))$. For the other side, we obtain a $(t_n(C)+r_n(C), w_n(C))$ right-cutoff. This result gives a more precise description of the profile of convergence than our Theorem 2.2, because if $j_n^C > 1$ and the location of the cutoff does not change, that is $t_n(C) = t_n$, the width of the windows will be improved from $1/\rho_{1,n}$ to $1/\rho_{j_n^C,n}$.

In the next section, sequences of processes having a profile cutoff at (t_n, w_n) or $(t_n + r_n, w_n)$, with profile $F(c) = \mathrm{e}^{-c}$ will be constructed, thus proving that (2.8) and (2.9) are tight.

3. Examples

Several examples from the existing literature could be written as particular cases of Theorem 2.2: reversible Markov chains for the L^2 distance Ycart (1999); Chen and Saloff-Coste (2010), n-tuples of independent processes for the relative entropy distance Barrera et al. (2006), random walks on sums or products of graphs Ycart (2007), samples of Ornstein-Uhlenbeck processes Lachaud (2005). The objective of this section is not an extensive review of possible applications, but rather the explicit construction of some sequences illustrating the tightness of (2.8) and (2.9), and the possible locations of window cutoffs. We shall use here the relative entropy distance, also called Kullback-Leibler divergence: if μ and ν are two probability measures with densities f and g with respect to λ, then:

$$d(\mu,\nu) = \int_{S_\mu} f \log(f/g)\,\mathrm{d}\lambda\;,$$

where S_μ denotes the support of μ. The main advantage of choosing that distance is its simplicity for dealing with tensor products:

$$d(\mu_1\otimes\mu_2, \nu_1\otimes\nu_2) = d(\mu_1,\nu_1) + d(\mu_2,\nu_2)\;.$$

Let a and ρ be two positive reals. Our building block will be a one-dimensional Ornstein-Uhlenbeck process, denoted by $X_{a,\rho}$ (see Lachaud (2005) on cutoff for samples of Ornstein-Uhlenbeck processes). The process $X_{a,\rho}$ is a solution of the equation:

$$\mathrm{d}X(t) = -\frac{\rho}{2}X(t)\,\mathrm{d}t + \sqrt{\rho}\,\mathrm{d}W(t)\;,$$

where W is the standard Brownian motion. The distribution of $X_{a,\rho}(0)$ is normal with expectation $\sqrt{2a}$ and variance 1. It can be easily checked that the distribution of $X_{a,\rho}(t)$ is normal with expectation $\sqrt{2a}\,\mathrm{e}^{-\rho t/2}$ and variance 1. Therefore the (relative entropy) distance to equilibrium is:

$$d(t) = a\,\mathrm{e}^{-\rho t}\;.$$

Consider now two sequences (a_n) and (ρ_n) of positive reals, and assume that (a_n) tends to infinity. Theorem 2.2 applies to the sequence of processes (X_{a_n,ρ_n}) with $a_{1,n} = a_n$, $\rho_{1,n} = \rho_n$, and $a_{i,n} = 0$ for $i > 1$. The location and width are:

$$t_n = \frac{\log(a_n)}{\rho_n} \quad\text{and}\quad w_n = \frac{1}{\rho_n}\;.$$

The sequence has a profile cutoff at (t_n, w_n) with profile $F(c) = \mathrm{e}^{-c}$. Indeed:

$$d_n(t_n + cw_n) = a_n\mathrm{e}^{-(\rho_n t_n + c)} = \mathrm{e}^{-c}\;.$$

Hence (2.8) is tight. For $\rho_n \equiv \rho$, $X_{a_n,\rho}$ is a Markov process with a fixed semigroup, and an increasingly remote starting point: cutoff for such sequences were studied in Martínez and Ycart (2001).

Using tuples of independent Ornstein-Uhlenbeck processes, one can construct sequences X_n for which the distance to equilibrium is any finite sum of exponentials. Let m_n be an integer. For $i = 1,\dots,m_n$, let $a_{i,n}$ and $\rho_{i,n}$ be two positive reals. Define the process X_n as:

$$X_n = \left(X_{a_{1,n},\rho_{1,n}},\dots,X_{a_{m_n,n},\rho_{m_n,n}}\right)\;,$$

where the coordinates are independent, each being an Ornstein-Uhlenbeck process as defined above. The distance to equilibrium of X_n is:

$$d_n(t) = \sum_{i=1}^{m_n} a_{i,n}\,\mathrm{e}^{-\rho_{i,n}t}\;. \tag{3.1}$$

Let n be an integer larger than 1. Let β_n be a real such that $0 \leqslant \beta_n \leqslant 1$. Define:

$$a_{1,n} = \mathrm{e}^n\;,\quad \rho_{1,n} = \frac{n}{1 + \frac{\beta_n}{n}\log\left(\frac{n}{\log(n)}\right)}\;, \tag{3.2}$$

and for $i = 2, \ldots, m_n = 9^n$,

$$a_{i,n} = \mathrm{e}^{-n} \ , \quad \rho_{i,n} = \log(\mathrm{e}^n + (i-1)\mathrm{e}^{-n}) \ . \tag{3.3}$$

The following notation is introduced for clarity:

$$\ell_n = \log\left(\frac{n}{\log(n)}\right) \ .$$

Using (2.2), (2.3), and (2.4), one gets:

$$t_n = 1 + \frac{\ell_n \beta_n}{n} = \frac{n}{\rho_{1,n}} \ , \quad w_n = \frac{t_n}{n} \ , \quad r_n = \frac{t_n \ell_n}{n} = \ell_n w_n \ . \tag{3.4}$$

Lemma 3.1. *Let d_n be defined by (3.1), with $a_{i,n}$ and $\rho_{i,n}$ given by (3.2) and (3.3). Assume the following limit (possibly equal to $+\infty$) exists:*

$$\gamma = \lim_{n\to\infty} (1-\beta_n)\ell_n \ . \tag{3.5}$$

Then:

$$\forall c \in \mathbb{R} \ , \quad \lim_{n\to\infty} d_n\left(t_n + (1-\beta_n)r_n + cw_n\right) = \mathrm{e}^{-c}(1+\mathrm{e}^{-\gamma}) \ . \tag{3.6}$$

A few cases that are obtained applying Lemma 3.1 are listed below. They illustrate the variety of possible behaviors.

- $\beta_n \equiv 1$: a cutoff with profile $2\mathrm{e}^{-c}$ occurs at (t_n, w_n).
- $\beta_n \equiv \beta \in [0,1)$: a cutoff with profile e^{-c} occurs at $(t_n + (1-\beta)r_n, w_n)$. For $\beta = 0$, this proves that (2.9) is tight.
- $\beta_n = (1+(-1)^n)/2$: a left-window cutoff occurs at (t_n, w_n), a right-window cutoff at $(t_n + r_n, w_n)$. The locations and width are optimal.
- $\beta_n = 1 - \gamma/\ell_n$, with $\gamma > 0$: a cutoff with profile $\mathrm{e}^{-c}(1+\mathrm{e}^{\gamma})$ occurs at (t_n, w_n).
- $\beta_n = 1 - (2+(-1)^n)/\ell_n$: a (t_n, w_n)-cutoff occurs, t_n and w_n are optimal. Yet no value of c is such that $d_n(t_n + cw_n)$ converges: there is no profile.

Proof: The main step is the following limit.

$$\lim_{n\to\infty} d_n\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right) = \mathrm{e}^{-c}(1+\mathrm{e}^{-\gamma}) \ . \tag{3.7}$$

In the sum defining d_n, let us isolate the first term: $d_n\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right) = D_1 + D_2$, with

$$D_1 = a_{1,n} \exp\left(-\rho_{1,n}\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)\right)$$

and

$$D_2 = \sum_{i=2}^{m_n} a_{i,n} \exp\left(-\rho_{i,n}\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)\right) \ .$$

The first term is:

$$D_1 = \exp\left(-\frac{(1-\beta_n)\ell_n + c}{t_n}\right) \ .$$

Its limit is $\mathrm{e}^{-(\gamma+c)}$ because $(1-\beta_n)\ell_n$ tends to γ and t_n tends to 1. The second term is:

$$D_2 = \sum_{i=2}^{+\infty} \mathrm{e}^{-n} \left(\mathrm{e}^n + (i-1)\mathrm{e}^{-n}\right)^{-\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)} \ .$$

Thus D_2 is a Riemann sum for the decreasing function $x \mapsto x^{-\left(1+\frac{\ell_n}{n}+\frac{c}{n}\right)}$. Therefore,

$$\int_{\mathrm{e}^n+\mathrm{e}^{-n}}^{\mathrm{e}^n+m_n\mathrm{e}^{-n}} x^{-\left(1+\frac{\ell_n}{n}+\frac{c}{n}\right)}\,\mathrm{d}x < D_2 < \int_{\mathrm{e}^n}^{\mathrm{e}^n+(m_n-1)\mathrm{e}^{-n}} x^{-\left(1+\frac{\ell_n}{n}+\frac{c}{n}\right)}\,\mathrm{d}x\,. \tag{3.8}$$

Now:

$$\frac{(\mathrm{e}^n)^{-\left(\frac{\ell_n}{n}+\frac{c}{n}\right)}}{\frac{\ell_n}{n}+\frac{c}{n}} = \mathrm{e}^{-c}\frac{\log(n)}{\ell_n+c}\,,$$

which tends to e^{-c}. Moreover,

$$\frac{(\mathrm{e}^n+(m_n-1)\mathrm{e}^{-n})^{-\left(\frac{\ell_n}{n}+\frac{c}{n}\right)}}{\frac{\ell_n}{n}+\frac{c}{n}} \leqslant \frac{n}{\ell_n+c}\left(\frac{m_n^{1/n}}{\mathrm{e}}\right)^{-(\ell_n+c)}\,,$$

which tends to 0 for $m_n = 9^n > \mathrm{e}^{2n}$. So the upper bound in (3.8) tends to e^{-c}. There remains to prove that the difference between the two integrals tends to 0. That difference is smaller than:

$$\int_{\mathrm{e}^n}^{\mathrm{e}^n+\mathrm{e}^{-n}} x^{-\left(1+\frac{\ell_n}{n}+\frac{c}{n}\right)}\,\mathrm{d}x = \left(\frac{(\mathrm{e}^n)^{-\left(\frac{\ell_n}{n}+\frac{c}{n}\right)}}{\frac{\ell_n}{n}+\frac{c}{n}}\right)\left(1-(1+\mathrm{e}^{-2n})^{-\left(\frac{\ell_n}{n}+\frac{c}{n}\right)}\right)\,.$$

We have seen that the first factor tends to e^{-c}. The second factor tends to 0, hence the result.

Let us now deduce (3.6) from (3.7). Using (3.4),

$$1+\frac{\ell_n}{n}+\frac{c}{n} = t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\,.$$

Hence:

$$\lim_{n\to\infty} d_n\left(t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\right) = \mathrm{e}^{-c}(1+\mathrm{e}^{-\gamma})\,. \tag{3.9}$$

Let us write:

$$t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n} = t_n + (1-\beta_n)r_n + cw_n - ((1-\beta_n)r_n + cw_n)\left(\frac{\ell_n\beta_n}{nt_n}\right)\,.$$

For $\beta_n = 1$ and $c < 0$ we have:

$$\begin{aligned}
0 &\leqslant d_n\left(t_n + cw_n\right) - d_n\left(t_n + c\frac{w_n}{t_n}\right) \\
&\leqslant \left(\exp\left(\rho_{m_n,n}(-c)w_n\frac{\ell_n}{n}\right) - 1\right) d_n\left(t_n + c\frac{w_n}{t_n}\right) \\
&\leqslant \left(\exp\left((-c)c'\frac{\ell_n}{n}\right) - 1\right) d_n\left(t_n + c\frac{w_n}{t_n}\right)
\end{aligned}$$

.

where the last inequality is obtained from $1 \leqslant \frac{\rho_{m_n,n}}{\rho_{1,n}} \leqslant c'$ with $c' = 2(1+6/\mathrm{e})$. On the other case:

$$\begin{aligned}
0 &\leqslant d_n\left(t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\right) - d_n\left(t_n + (1-\beta_n)r_n + cw_n\right) \\
&\leqslant \left(\exp\left(\rho_{1,n}\left(((1-\beta_n)r_n + cw_n)\frac{\ell_n\beta_n}{n}\right)\right) - 1\right) d_n\left(t_n + (1-\beta_n)r_n + cw_n\right) \\
&= \left(\exp\left(\frac{\ell_n^2(1-\beta_n)\beta_n + c\ell_n\beta_n}{n}\right) - 1\right) d_n\left(t_n + (1-\beta_n)r_n + cw_n\right)\,.
\end{aligned}$$

Therefore if $M_n = \max\left\{\exp\left((-c)c'\frac{\ell_n}{n}\right), \exp\left(\frac{\ell_n^2(1-\beta_n)\beta_n + c\ell_n\beta_n}{n}\right)\right\}$ then :

$$\begin{aligned}|d_n\left(t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\right) - d_n\left(t_n + (1-\beta_n)r_n + cw_n\right)| \\ \leq d_n\left(t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\right)(M_n - 1)\end{aligned}$$

Hence the difference tends to 0 because $M_n - 1$ tends to 0, since $\frac{\ell_n^2}{n}$ tends to 0. □

4. Proof of Theorem 2.2

Proofs of inequalities (2.8) and (2.9) are given below.

Proof of (2.8): Let c be a negative real. Fix ϵ such that $0 < \epsilon < -c$. Using (2.2), define i_n^* as:

$$i_n^* = \min\left\{ i\,,\ t_n - \epsilon w_n \leqslant \frac{\log(A_{i,n})}{\rho_{i,n}} \leqslant t_n \right\}\ . \tag{4.1}$$

From (2.6), $t_n + cw_n$ is positive for n large enough. Then:

$$\begin{aligned}d_n(t_n + cw_n) &= \sum_{i=1}^{+\infty} a_{i,n}\exp(-\rho_{i,n}(t_n + cw_n)) \\ &\geqslant \sum_{i=1}^{i_n^*} a_{i,n}\exp(-\rho_{i,n}(t_n + cw_n)) \\ &\geqslant A_{i_n^*,n}\exp(-\rho_{i_n^*,n}(t_n + cw_n)) \\ &\geqslant \exp((-\epsilon w_n - cw_n)\rho_{i_n^*}) \\ &\geqslant \exp((-\epsilon w_n - cw_n)\rho_{1,n}) \\ &= \mathrm{e}^{-c-\epsilon}\ .\end{aligned}$$

Since the inequality holds for all $\epsilon > 0$, the result follows. □

Proof of (2.9): Let c be a positive real. Our goal is to prove the following inequality.

$$d_n(t_n + r_n + cw_n) \leqslant \mathrm{e}^{-(r_n + cw_n)\rho_{1,n}}\frac{t_n}{r_n + cw_n}\left(\frac{r_n + cw_n}{t_n} + \mathrm{e}^{C_n}\right)\ , \tag{4.2}$$

where C_n tends to 0 as n tends to infinity. Let us first check that (4.2) implies (2.9). Observe that $\frac{r_n + cw_n}{t_n}$ tends to 0. Using (2.3) and (2.4):

$$\mathrm{e}^{-(r_n + cw_n)\rho_{1,n}}\frac{t_n}{r_n + cw_n} = \mathrm{e}^{-c}\frac{1}{1 - \frac{\log(\log(t_n\rho_{1,n})) - c}{\log(t_n\rho_{1,n})}}\ .$$

By (2.6) the right-hand side tends to e^{-c}, hence the result.

To prove (4.2), split the sum defining $d_n(t_n + r_n + cw_n)$ into two parts S_1 and S_2, with:

$$S_1 = \sum_{i=1}^{l} a_{i,n}\exp(-\rho_{i,n}(t_n + r_n + cw_n))$$

and

$$S_2 = \sum_{i=l+1}^{+\infty} a_{i,n} \exp(-\rho_{i,n}(t_n + r_n + cw_n)) \,.$$

Using the fact that the $\rho_{i,n}$ are increasing,

$$S_1 \leqslant A_{l,n} \exp(-\rho_{1,n}(t_n + r_n + cw_n)) \,. \tag{4.3}$$

To bound S_2, the idea is the same as in the proof of (3.6). From (2.2), we obtain $\exp(-\rho_{i,n}t_n) \leqslant A_{i,n}^{-1}$. Therefore:

$$S_2 \leqslant \sum_{i=l+1}^{+\infty} a_{i,n} A_{i,n}^{-(1+(r_n+cw_n)/t_n)} \,. \tag{4.4}$$

The function $x \mapsto x^{-(1+(r_n+cw_n))/t_n}$ is decreasing, and its integral from l to $+\infty$ converges. The right-hand side of (4.4) is a Riemann sum for that integral. Therefore:

$$S_2 \leqslant \frac{t_n}{r_n + cw_n} A_{l,n}^{-(r_n+cw_n)/t_n} \,. \tag{4.5}$$

If $t_n = \frac{\log(A_{1,n})}{\rho_{1,n}}$, or equivalently $A_{1,n} = \exp(t_n\rho_{1,n})$, the application of (4.3) and (4.5) for $l = 1$ yields:

$$d_n(t_n + r_n + cw_n) \leqslant \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} \frac{t_n}{r_n + cw_n} \left(\frac{r_n + cw_n}{t_n} + 1 \right) \,, \tag{4.6}$$

which is (4.2) for $C_n = 0$. If $t_n \neq \frac{\log(A_{1,n})}{\rho_{1,n}}$, then $A_{1,n} < \exp(t_n\rho_{1,n})$. Moreover

$$t_n \leq \frac{1}{\rho_{1,n}} \sup_i \log(A_{i,n}).$$

Then there are two cases, $t_n = \frac{1}{\rho_{1,n}} \sup_i \log(A_{i,n})$ and $t_n < \frac{1}{\rho_{1,n}} \sup_i \log(A_{i,n})$. We can prove that the first one lead to a contradiction, from equation (2.5) we have that $t_n < \infty$ and therefore $\sup_i \log A_{i,n} < \infty$. Let

$$\epsilon_n = \frac{1}{2} t_n \left(1 - \frac{\rho_{1,n}}{\rho_{2,n}} \right) \,.$$

Let $\hat{\imath}_n > 1$ be such that $\epsilon_n \geq t_n - \frac{\log A_{\hat{\imath}_n,n}}{\rho_{\hat{\imath}_n,n}}$ then the contradiction follows from

$$\epsilon_n \geq \frac{1}{\rho_{1,n}} \sup_i \log(A_{i,n}) - \frac{\log A_{\hat{\imath}_n,n}}{\rho_{\hat{\imath}_n,n}} \geq \sup_i \log(A_{i,n}) \left(\frac{1}{\rho_{1,n}} - \frac{1}{\rho_{\hat{\imath}_n,n}} \right) \geq 2\epsilon_n.$$

In the second case it is clear that there exist l_n such that

$$A_{l_n-1,n} < \mathrm{e}^{\rho_{1,n}t_n} \leqslant A_{l_n,n} \,. \tag{4.7}$$

Applying (4.3) and (4.5) to $l = l_n - 1$ yields:

$$\begin{aligned} & d_n(t_n + r_n + cw_n) && (4.8) \\ & \quad \leqslant \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} + \frac{t_n}{r_n + cw_n} \exp\left(-\frac{r_n + cw_n}{t_n} \log A_{l_n-1,n} \right) \\ & \quad = \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} + \frac{t_n}{r_n + cw_n} \exp\left(-(r_n + cw_n)\rho_{1,n} \frac{\log A_{l_n-1,n}}{\rho_{1,n}t_n} \right) \\ & \quad = \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} \frac{t_n}{r_n + cw_n} \left(\frac{r_n + cw_n}{t_n} + \mathrm{e}^{C_n} \right) \,. && (4.9) \end{aligned}$$

with

$$C_n = (r_n + cw_n)\rho_{1,n}\left(1 - \frac{\log A_{l_n-1,n}}{\rho_{1,n}t_n}\right) . \tag{4.10}$$

We must prove that C_n tends to 0. By (2.3) and (2.4):

$$(r_n + cw_n)\rho_{1,n} = \log(\rho_{1,n}t_n) - \log\log(\rho_{1,n}t_n) + c . \tag{4.11}$$

From (4.7):

$$0 < 1 - \frac{\log(A_{l_n-1,n})}{\rho_{1,n}t_n} \leqslant \frac{1}{\rho_{1,n}t_n}\log\left(1 + \frac{a_{l_n,n}}{A_{l_n-1,n}}\right) . \tag{4.12}$$

Plugging (4.11) and (4.12) into (4.10), for n large enough:

$$0 < C_n \leqslant \left(\frac{\log(\rho_{1,n}t_n) - \log\log(\rho_{1,n}t_n) + c}{\rho_{1,n}t_n}\right)\log\left(1 + \frac{a_{l_n,n}}{A_{l_n-1,n}}\right) .$$

By (2.6), the first factor of the right-hand side tends to 0. Moreover, condition (2.7) entails that for n large enough:

$$\log\left(1 + \frac{a_{l_n,n}}{A_{l_n-1,n}}\right) < \log(1+\alpha) .$$

Hence the result. $\square$

Acknowledgement

We would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly* **93** (5), 333–348 (1986). MR841111.

J. Barrera, B. Lachaud and B. Ycart. Cut-off for n-tuples of exponentially converging processes. *Stochastic Process. Appl.* **116** (10), 1433–1446 (2006). MR2260742.

D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** (2), 294–313 (1992). MR1161056.

G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.* **13**, no. 3, 26–78 (2008). MR2375599.

G.-Y. Chen and L. Saloff-Coste. The L^2-cutoff for reversible Markov processes. *J. Funct. Anal.* **258** (7), 2246–2315 (2010). MR2584746.

S.B. Connor. Separation and coupling cutoffs for tuples of independent Markov processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **7**, 65–77 (2010). MR2644042.

P. Diaconis, R.L. Graham and J.A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1** (1), 51–72 (1990). MR1068491.

P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* **16** (4), 2098–2122 (2006). MR2288715.

P. Diaconis and M. Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.* **18** (1), 208–218 (1987). MR871832.

A. Diédhiou and P. Ngom. Cutoff time based on generalized divergence measure. *Statist. Probab. Lett.* **79** (10), 1343–1350 (2009). MR2522607.

J. Ding, E. Lubetzky and Y. Peres. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields* **146** (1-2), 61–85 (2010). MR2550359.

B. Lachaud. Cut-off and hitting times of a sample of Ornstein-Uhlenbeck processes and its average. *J. Appl. Probab.* **42** (4), 1069–1080 (2005). MR2203823.

B. Lachaud and B. Ycart. Convergence times for parallel Markov chains. In *Positive systems*, volume 341 of *Lecture Notes in Control and Inform. Sci.*, pages 169–176. Springer, Berlin (2006). MR2250253.

H. Lacoin. A product chain without cutoff. *ArXiv Mathematics e-prints* (2014). arXiv: 1407.1754v1.

D.A. Levin, Y. Peres and E.L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI (2009). ISBN 978-0-8218-4739-8. With a chapter by James G. Propp and David B. Wilson. MR2466937.

S. Martínez and B. Ycart. Decay rates and cutoff for convergence and hitting times of Markov chains with countably infinite state space. *Adv. in Appl. Probab.* **33** (1), 188–205 (2001). MR1825322.

J. Miller and Y. Peres. Uniformity of the uncovered set of random walk and cutoff for lamplighter chains. *Ann. Probab.* **40** (2), 535–577 (2012). MR2952084.

L. Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 263–346. Springer, Berlin (2004). MR2023654.

B. Ycart. Cutoff for samples of Markov chains. *ESAIM Probab. Statist.* **3**, 89–106 (electronic) (1999). MR1716128.

B. Ycart. Stopping tests for Markov chain Monte-Carlo methods. *Methodol. Comput. Appl. Probab.* **2** (1), 23–36 (2000). MR1783151.

B. Ycart. Cut-off for large sums of graphs. *Ann. Inst. Fourier (Grenoble)* **57** (7), 2197–2208 (2007). Festival Yves Colin de Verdière. MR2394540.