Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 116 (2006) 1433–1446
stochastic processes and their applications
www.elsevier.com/locate/spa

# Cut-off for $n$-tuples of exponentially converging processes

Javiera Barrera[b], Béatrice Lachaud[b], Bernard Ycart[a,*]

[a] LMC-IMAG, B.P. 53, 38041 Grenoble Cedex 09, France
[b] MAP5, CNRS UMR 8145, Université Paris 5, 45 rue des Saints-Pères, 75270 Paris Cedex 06, France

Received 19 December 2005; received in revised form 3 March 2006; accepted 10 March 2006
Available online 29 March 2006

**Abstract**

Given an $n$-tuple of independent processes, each converging at an exponential rate, conditions are given under which a cut-off occurs for the $n$-tuple, when the convergence is measured by different distances between probability distributions. More precise estimates and explicit examples are given for the case of i.i.d. coordinates.


*MSC:* primary 60F05; secondary 60J25

*Keywords:* Cut-off; Total variation distance; Hellinger distance; Chi-square distance; Kullback distance; Exponential convergence

## 1. Introduction

Since its identification by Aldous and Diaconis [1], the cut-off phenomenon of steep convergence to equilibrium has been observed on many Markov chains [2,17,5,15,18,13,22]. In [19] Saloff-Coste gives an extensive list of random walks for which the phenomenon occurs. Before a certain 'cut-off time' those chains stay far from equilibrium in the sense that the total variation distance between the distribution at time $t$ and the equilibrium measure is close to 1; after that instant, the total variation distance decays exponentially to 0. There exist many other

* Corresponding author. Tel.: +33 476 51 4995; fax: +33 476 63 12 63.
*E-mail addresses:* barrera@math-info.univ-paris5.fr (J. Barrera), lachaud@math-info.univ-paris5.fr (B. Lachaud), Bernard.Ycart@imag.fr (B. Ycart).

0304-4149/$ - see front matter 
doi:10.1016/j.spa.2006.03.003

ways of characterizing convergence than the total variation distance: see Gibbs and Su [11] for a useful review. We shall use the following definition of cut-off, that holds for any 'distance' (in the sense of [11]) between probability distributions. In order to simplify notation, all processes are indexed by continuous time; switching to discrete time requires very little adaptation.

**Definition 1.** For each $n \geq 0$, let $X^{(n)}$ be a stochastic process on a measurable space $E^{(n)}$, converging in distribution to some probability law $\nu^{(n)}$. Let $d$ be a distance between probability distributions, taking values between 0 and $M$, where $M$ is either a positive real or $+\infty$. For $t \geq 0$, denote by $d^{(n)}(t)$ the distance between the distribution of $X^{(n)}$ at time $t$ and $\nu^{(n)}$:

$$d^{(n)}(t) = d(\mathcal{L}X^{(n)}(t), \nu^{(n)}).$$

Let $(t_n)$ be a sequence of positive reals. The sequence of processes $(X^{(n)})$ is said to have a cut-off at time $(t_n)$ in the sense of distance $d$ if for $c > 0$:

$$c < 1 \Longrightarrow \lim_{n\to\infty} d^{(n)}(ct_n) = M,$$
$$c > 1 \Longrightarrow \lim_{n\to\infty} d^{(n)}(ct_n) = 0.$$

Several other definitions of cut-off have been proposed. In [19] Saloff-Coste introduces the notion of $L^p$-cut-off, which is very close to Definition 1 when $d$ is the $L^p$ distance. In the same article the author also discusses the notion of pre-cut-off (Definition 3.8, p. 280 in [19]) as a means to capture the order of magnitude of a possible cut-off.

The purpose of this article is to study the cut-off phenomenon for $n$-tuples of independent processes, identically distributed or not, and for different distances. Notice that the processes need not be Markovian. For each $i = 1, \ldots, n$, we assume that the $i$-th coordinate converges with exponential rate $\rho_i$ to its equilibrium measure. Under technical conditions on the $\rho_i$'s, Theorem 3 states that a cut-off occurs for the $n$-tuple at any time equivalent to

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

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

According to Diaconis and Saloff-Coste [7], Y. Peres recently conjectured a sufficient condition for a Markov chain to have cut-off, bearing on the gap and the mixing time, which is the time at which the distance to stationarity drops below 1/4. In [7] Diaconis and Saloff-Coste proved Peres' conjecture for birth and death chains and the separation distance. According to [7], Chen and Saloff-Coste proved it in a more general setting for the $L^2$ distance. If our results are applied to a Markovian $n$-tuple, the gap is the smallest rate of convergence $\rho_{(1,n)}$. Our cut-off time is equivalent to the mixing time. Peres' condition corresponds to our condition (4). Thus, not only do we prove Peres' conjecture for $n$-tuples of independent processes, but we also give an explicit expression for the mixing (cut-off) time.

When the coordinates of the $n$-tuple are i.i.d., sharper results can be proved (Theorems 9 and 10) for the total variation, Hellinger, chi-square and Kullback distances: if $\rho$ is the common rate of exponential convergence for the coordinates, then not only does a cut-off occur for the $n$-tuple at time $\log n/(2\rho)$, but for any fixed $u$ the distance to equilibrium at time $\log n/(2\rho) + u$ converges to a positive value. In other words, the cut-off occurs over an interval of time of length $O(1)$ around $\log n/(2\rho)$.

Particular cases of Theorems 3 and 10 have already appeared in the literature. One of the first examples of cut-off was studied by Diaconis et al. [8,6] for the random walk on the hypercube. As remarked by Ycart [23], p. 91, that random walk can be interpreted as the discrete time version of an $n$-tuple of i.i.d. continuous time binary Markov chains. The cut-off for an $n$-tuple of i.i.d. random walks was treated by Aldous and Diaconis [2] (Proposition 7.7, p. 89). The case of $n$-tuples of i.i.d. reversible Markov chains with finite state space was treated by Ycart [23], and applications to stopping tests for MCMC methods were described in [24]. In [4] Bon and Păltănea considered the case of independent, but not identically distributed continuous time binary Markov chains, in the context of reliability theory.

The original proofs of cut-off for random walks by Diaconis et al. [8,6], as well as for $n$-tuples of reversible Markov chains by Ycart [23], mainly relied on the spectral analysis of the transition matrix. The approach chosen here relates the cut-off phenomenon to the way distances account for the concentration of product measures. A comparison of the use of different distances for measuring cut-offs of random walks can be found in Su's thesis [21].

The paper is organized as follows. Section 2 contains our main results, Lemma 2 and Theorem 3. Section 3 studies the particular cases of the total variation, Hellinger, chi-square and Kullback distances. Section 4 discusses explicit examples of cut-off times. In Section 5, the i.i.d. case is studied and two more examples are given: M/M/$\infty$ birth–death processes, and Ornstein–Uhlenbeck diffusions.

## 2. Main result

Some distances behave better than others with respect to the product of measures. By 'behaving' we mean that the distance between two measure products is controlled by the sum of some power of distances between coordinates. Here is a more precise statement.

Let $n$ be a positive integer. For $i = 1, \dots, n$, let $(E_i, \mathcal{F}_i)$ be a measurable space, $\mu_i$ and $\nu_i$ be two probability distributions on $E_i$. Let $E^{(n)}$ denote the Cartesian product $E_1 \times \cdots \times E_n$, endowed with the product $\sigma$-algebra. Let $\mu^{(n)}$ and $\nu^{(n)}$ denote the tensor products of the $\mu_i$'s and $\nu_i$'s respectively:

$$\mu^{(n)} = \mu_1 \otimes \cdots \otimes \mu_n \qquad \text{and} \qquad \nu^{(n)} = \nu_1 \otimes \cdots \otimes \nu_n.$$

We will assume:

$$\phi\left(\sum_{i=1}^{n}(d(\mu_i, \nu_i))^k\right) \le d(\mu^{(n)}, \nu^{(n)}) \le \psi\left(\sum_{i=1}^{n}(d(\mu_i, \nu_i))^k\right), \tag{1}$$

for some positive integer $k$ and two functions $\phi$, $\psi$ such that $\phi(x)$ tends to $M$ (maximal value of $d$) as $x$ tends to infinity and $\psi(x)$ tends to 0 as $x$ tends to 0 (examples will be given in Section 3).

The cut-off phenomenon for $n$-tuples of independent processes is explained by the following lemma.

**Lemma 2.** *For $i = 1, 2, \dots$, let $d_i$ be a positive function defined on $\mathbb{R}^+$, and $\rho_i$ a positive real. For $n \ge 1$, denote by $\rho_{(1,n)}, \dots, \rho_{(n,n)}$ the values of $\rho_1, \dots, \rho_n$ ranked in increasing order, and by $\tau_n$ the following real:*

$$\tau_n = \max\left\{\frac{\log i}{\rho_{(i,n)}}, i = 1, \dots, n\right\}. \tag{2}$$

*Assume the following hypotheses hold.*

(1) *There exists a positive function g, decreasing and tending to* 0 *as t tends to infinity, and a positive real $t_0$ such that for all $t \geq t_0$ and all $i \geq 1$,*

$$\left| \frac{\log d_i(t)}{t} + \rho_i \right| \leq g(t). \tag{3}$$

(2)

$$\lim_{n \to \infty} \rho_{(1,n)}\, \tau_n = +\infty. \tag{4}$$

(3) *For any positive real c,*

$$\lim_{n \to \infty} \frac{g(c\tau_n)}{\rho_{(1,n)}} = 0. \tag{5}$$

*Then for any positive integer k, any positive real c and any sequence $(\tau_n')$ such that $\lim \tau_n'/\tau_n = 1$,*

$$c < 1 \Longrightarrow \lim_{n \to \infty} \sum_{i=1}^{n} (d_i(c\tau_n'/k))^k = +\infty,$$

$$c > 1 \Longrightarrow \lim_{n \to \infty} \sum_{i=1}^{n} (d_i(c\tau_n'/k))^k = 0.$$

Of the three hypotheses, the first one is obviously the most important. It says that not only should the $d_i(t)$ converge to zero at exponential rate $\rho_i$, but also they should do so uniformly in $i$. The other hypotheses involve $\rho_{(1,n)}$, which is the minimum of $\rho_1, \ldots, \rho_n$. As soon as the sequence $(\rho_i)$ does not tend to $+\infty$, $\rho_{(1,n)}$ is bounded and $\tau_n$ tends to infinity. If $(\rho_i)$ is bounded away from 0 and does not tend to infinity, then both (4) and (5) are trivially satisfied. But it may also happen that some subsequence tends to 0, in which case, $\tau_n$ should tend to $+\infty$, and $g(c\tau_n)$ to 0, fast enough to compensate. In the Markovian case, condition (4) corresponds to Peres' condition (see [7] and references therein).

Rephrased in terms of distance to equilibrium for an $n$-tuple of processes, Lemma 2 becomes:

**Theorem 3.** *Let d be a distance between probability distributions satisfying (1). Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of independent processes. Denote by $d_i(t)$ the distance to equilibrium of $X_i$ at time $t$. Assume that the functions $d_i(t)$ satisfy the hypotheses (3), (4) and (5) of Lemma 2. Let $X^{(n)}$ denote the n-tuple of processes $(X_1, \ldots, X_n)$.*

*The sequence of processes $(X^{(n)})$ has a cut-off in the sense of distance d at any sequence of times equivalent to $\tau_n/k$, where $\tau_n$ is defined by (2).*

Here is the proof of Lemma 2.

**Proof.** We first prove the result for $\tau_n$. Define

$$g_i(t) = \frac{\log d_i(t)}{t} + \rho_i.$$

Thus:

$$\sum_{i=1}^{n} (d_i(c\tau_n/k))^k = \sum_{i=1}^{n} \exp(-\rho_i c\tau_n + c\tau_n g_i(c\tau_n/k)).$$

By (3), the $g_i$'s are uniformly bounded:

$$\forall t \geq t_0,\ \forall i, \quad |g_i(t)| \leq g(t).$$

Therefore for $n$ large enough:

$$S_n \exp\left(-c\tau_n g\left(c\frac{\tau_n}{k}\right)\right) \leq \sum_{i=1}^{n} d_i \left(c\frac{\tau_n}{k}\right)^k \leq S_n \exp\left(c\tau_n g\left(c\frac{\tau_n}{k}\right)\right),$$

with

$$S_n = \sum_{i=1}^{n} \exp(-\rho_i c\tau_n) = \sum_{i=1}^{n} \exp(-\rho_{(i,n)} c\tau_n).$$

We first treat the upper bound, for $c > 1$. Observe that for all $i = 1, \dots, n$, $\exp(-\rho_{(i,n)} c\tau_n) \leq i^{-c}$, since $\tau_n \geq \log i/\rho_{(i,n)}$. For all $l = 1, \dots, n-1$, one can write:

$$\begin{aligned} S_n &\leq l\mathrm{e}^{-\rho_{(1,n)} c\tau_n} + \sum_{i=l+1}^{n} i^{-c} \\ &\leq l\mathrm{e}^{-\rho_{(1,n)} c\tau_n} + \int_l^n x^{-c}\,\mathrm{d}x \\ &= l\mathrm{e}^{-\rho_{(1,n)} c\tau_n} + \frac{1}{c-1}\left(l^{-(c-1)} - n^{-(c-1)}\right). \end{aligned}$$

This bound also holds for $l = n$. Define now $l_n = \lfloor \mathrm{e}^{\rho_{(1,n)}\tau_n} \rfloor$, where $\lfloor\cdot\rfloor$ denotes the integer part; $l_n$ is no larger than $n$, by definition of $\tau_n$. One has:

$$\begin{aligned} S_n &\leq l_n\mathrm{e}^{-\rho_{(1,n)} c\tau_n} + \frac{l_n^{-(c-1)}}{c-1} \\ &\leq \mathrm{e}^{-\rho_{(1,n)}(c-1)\tau_n} + \frac{(\mathrm{e}^{\rho_{(1,n)}\tau_n} - 1)^{-(c-1)}}{c-1} \\ &= \mathrm{e}^{-\rho_{(1,n)}(c-1)\tau_n}\left(1 + \frac{(1 - \mathrm{e}^{-\rho_{(1,n)}\tau_n})^{-(c-1)}}{c-1}\right). \end{aligned}$$

Therefore:

$$\begin{aligned} \sum_{i=1}^{n} d_i\left(c\frac{\tau_n}{k}\right)^k &\leq \mathrm{e}^{-\rho_{(1,n)}(c-1)\tau_n}\left(1 + \frac{(1 - \mathrm{e}^{-\rho_{(1,n)}\tau_n})^{-(c-1)}}{c-1}\right)\exp\left(c\tau_n g\left(c\frac{\tau_n}{k}\right)\right) \\ &= \left(1 + \frac{(1 - \mathrm{e}^{-\rho_{(1,n)}\tau_n})^{-(c-1)}}{c-1}\right)\exp\left(-\rho_{(1,n)}\tau_n\left((c-1) - c\frac{g(\frac{c\tau_n}{k})}{\rho_{(1,n)}}\right)\right), \end{aligned}$$

which tends to 0 as $n$ tends to infinity, using (4) and (5).

Let us now treat the lower bound, for $0 < c < 1$. For each $n$, choose $i_n^*$ such that $\tau_n = \log i_n^*/\rho_{(i_n^*,n)}$, i.e. $i_n^* = \exp(\tau_n \rho_{(i_n^*,n)}) \geq \exp(\tau_n \rho_{(1,n)})$. One has:

$$\begin{aligned} S_n &\geq \sum_{i=1}^{i_n^*} \exp(-c\rho_{(i,n)}\tau_n) \\ &\geq \exp((1-c)\rho_{(i_n^*,n)}\tau_n) \\ &\geq \exp((1-c)\rho_{(1,n)}\tau_n). \end{aligned}$$

Hence:

$$\sum_{i=1}^{n} d_i \left( c\frac{\tau_n}{k} \right)^k \geq \exp\left( \rho_{(1,n)}\tau_n \left( (1-c) - c\frac{g(\frac{c\tau_n}{k})}{\rho_{(1,n)}} \right) \right),$$

which tends to $+\infty$ as $n$ tends to infinity, using (4) and (5).

Consider now another sequence $(\tau_n')$, equivalent to $(\tau_n)$. The new sum can be bounded as before:

$$S_n' \exp\left( -c\tau_n' g\left( c\frac{\tau_n'}{k} \right) \right) \leq \sum_{i=1}^{n} d_i \left( c\frac{\tau_n'}{k} \right)^k \leq S_n' \exp\left( c\tau_n' g\left( c\frac{\tau_n'}{k} \right) \right),$$

with

$$S_n' = \sum_{i=1}^{n} \exp(-\rho_i c\tau_n').$$

Let us treat the upper bound. Fix $c' < 1$ such that $cc' > 1$. For $n$ large enough, $c' \leq \tau_n'/\tau_n \leq 1/c'$. Therefore:

$$S_n' \leq \sum_{i=1}^{n} \exp(-\rho_i cc'\tau_n),$$

and

$$\exp\left( c\tau_n' g\left( c\frac{\tau_n'}{k} \right) \right) \leq \exp\left( \frac{c}{c'}\tau_n g\left( cc'\frac{\tau_n}{k} \right) \right),$$

since $g$ is decreasing. The upper bound of $S_n$ can be applied to $S_n'$, replacing $c$ by $cc'$. One gets:

$$\begin{aligned} &\sum_{i=1}^{n} d_i \left( c\frac{\tau_n'}{k} \right)^k \\ &\quad \leq \mathrm{e}^{-\rho_{(1,n)}(cc'-1)\tau_n} \left( 1 + \frac{(1-\mathrm{e}^{-\rho_{(1,n)}\tau_n})^{-(cc'-1)}}{cc'-1} \right) \exp\left( \frac{c}{c'}\tau_n g\left( cc'\frac{\tau_n}{k} \right) \right) \\ &\quad = \left( 1 + \frac{(1-\mathrm{e}^{-\rho_{(1,n)}\tau_n})^{-(cc'-1)}}{cc'-1} \right) \exp\left( -\rho_{(1,n)}\tau_n \left( (cc'-1) - \frac{c}{c'}\frac{g(\frac{cc'\tau_n}{k})}{\rho_{(1,n)}} \right) \right), \end{aligned}$$

which tends to 0 as $n$ tends to infinity. For the lower bound, the proof is similar and will be omitted. □

## 3. Examples of distances

We will discuss here the behavior of some classical distances with respect to products. We follow the presentation by Gibbs and Su [11] (see also Pollard, [14] Section 3.3, p. 59 and Reiss [16], Chapter 3). The definition of some distances may change from one textbook to another. The choices made here ensure that, for a large class of processes converging at an exponential rate for all four distances, the rates are equal.

Let $E$ be a measurable space, with $\sigma$-algebra $\mathcal{F}$. Let $\mu$ and $\nu$ be two measures on $E$. We will denote by $\lambda$ any dominating measure (for instance $(\mu+\nu)/2$), and by $f$, $g$ the densities of $\mu$, $\nu$ with respect to $\lambda$.

**Definition 4.** (1) The total variation distance between $\mu$ and $\nu$ is

$$\begin{aligned} d_{TV}(\mu,\nu) &= \sup_{A\in\mathcal{F}} |\mu(A)-\nu(A)| \\ &= \frac{1}{2}\sup_{\|\varphi\|_\infty\le 1}\left|\int_E \varphi\,\mathrm{d}\mu - \int_E \varphi\,\mathrm{d}\nu\right| \\ &= \frac{1}{2}\int_E |f-g|\,\mathrm{d}\lambda. \end{aligned} \tag{6}$$

(2) The Hellinger distance between $\mu$ and $\nu$ is

$$d_H(\mu,\nu) = \frac{1}{\sqrt{2}}\left(\int_E (\sqrt{f}-\sqrt{g})^2\,\mathrm{d}\lambda\right)^{1/2} = \left(1-\int_E \sqrt{fg}\mathrm{d}\lambda\right)^{1/2}. \tag{7}$$

(3) The $L^2(\lambda)$ distance between $\mu$ and $\nu$ is

$$d_{L^2(\lambda)}(\mu,\nu) = \left(\int_E (f-g)^2\,\mathrm{d}\lambda\right)^{1/2}. \tag{8}$$

(4) The Kullback distance between $\mu$ and $\nu$ is

$$d_K(\mu,\nu) = \left(\int_{S_\mu} f\log(f/g)\,\mathrm{d}\lambda\right)^{1/2}, \tag{9}$$

where $S_\mu$ denotes the support of $\mu$.

Notice that the expressions of $d_{TV}$, $d_H$ and $d_K$ do not depend on the choice of the dominating measure $\lambda$, whereas $d_{L^2(\lambda)}$ is not intrinsic. It is customary to overlook the fact that among these quantities, some of them are not distances in the topological sense (the Kullback 'distance' is not even symmetric). If $\mu$ is absolutely continuous with respect to $\nu$, then $d^2_{L^2(\nu)}(\mu,\nu)$ is the usual chi-square distance of $\mu$ with respect to $\nu$. Since we shall mainly deal with that case, and in order to ensure homogeneity, we will call chi-square distance of $\mu$ and $\nu$ and denote by $d_{\chi^2}(\mu,\nu)$ the $L^2(\nu)$ distance:

$$d_{\chi^2}(\mu,\nu) = d_{L^2(\nu)}(\mu,\nu).$$

The following inequalities between distances are classical (see figure 1 in the article of Gibbs and Su [11], together with references and historical remarks therein).

**Proposition 5.** (1) $d_H^2(\mu,\nu) \le d_{TV}(\mu,\nu)$.
(2) $d_{TV}(\mu,\nu) \le d_H(\mu,\nu)\sqrt{2-d_H^2(\mu,\nu)}$.
(3) $d_{TV}(\mu,\nu) \le d_{\chi^2}(\mu,\nu)/2$.
(4) $d_{TV}(\mu,\nu) \le d_K(\mu,\nu)/\sqrt{2}$.
(5) $d_H(\mu,\nu) \le \sqrt{2}\,d_K(\mu,\nu)$.
(6) $d_K(\mu,\nu) \le d_{\chi^2}(\mu,\nu)$.

Recall that we are interested in distances between measure products. Let $n$ be a positive integer. For $i = 1,\ldots,n$, let $\mu_i$ and $\nu_i$ be two probability distributions. Let $\mu^{(n)}$ and $\nu^{(n)}$ denote the tensor products of the $\mu_i$'s and $\nu_i$'s respectively. If $\lambda_i$ is a dominating measure for $\mu_i$ and $\nu_i$, then $\lambda^{(n)} = \lambda_1\otimes\cdots\otimes\lambda_n$ will be taken as the dominating measure for $\mu^{(n)}$ and $\nu^{(n)}$. The following proposition summarizes the relations between distances of measure products and distances of their coordinates.

**Proposition 6.** (1) *Total variation:*

$$1 - \exp\left(-\frac{1}{2}\sum_{i=1}^{n} d_{TV}^2(\mu_i, \nu_i)\right) \le d_{TV}(\mu^{(n)}, \nu^{(n)}) \le \sum_{i=1}^{n} d_{TV}(\mu_i, \nu_i). \tag{10}$$

(2) *Hellinger:*

$$d_H^2(\mu^{(n)}, \nu^{(n)}) = 1 - \prod_{i=1}^{n}\left(1 - d_H^2(\mu_i, \nu_i)\right).$$

$$1 - \exp\left(-\sum_{i=1}^{n} d_H^2(\mu_i, \nu_i)\right) \le d_H^2(\mu^{(n)}, \nu^{(n)}) \le \sum_{i=1}^{n} d_H^2(\mu_i, \nu_i). \tag{11}$$

(3) *Chi-square:*

$$d_{\chi^2}^2(\mu^{(n)}, \nu^{(n)}) = \prod_{i=1}^{n}\left(1 + d_{\chi^2}^2(\mu_i, \nu_i)\right) - 1$$

$$\sum_{i=1}^{n} d_{\chi^2}^2(\mu_i, \nu_i) \le d_{\chi^2}^2(\mu^{(n)}, \nu^{(n)}) \le \exp\left(\sum_{i=1}^{n} d_{\chi^2}^2(\mu_i, \nu_i)\right) - 1. \tag{12}$$

(4) *Kullback:*

$$d_K^2(\mu^{(n)}, \nu^{(n)}) = \sum_{i=1}^{n} d_K^2(\mu_i, \nu_i). \tag{13}$$

**Proof.** For the Hellinger, chi-square and Kullback distances, the relations are well known (see Lemma 3.3.10, p. 100 in [16]). For the total variation distance, the upper bound is also classical. The lower bound is an easy consequence of (11), together with points (1) and (2) of Proposition 5. □

Proposition 6 shows that the Hellinger, the chi-square and the Kullback distances satisfy hypothesis (1): Theorem 3 applies with $k = 2$. Observe that the value of $k$, and hence the location of the cut-off instant, may depend on the definition of the distance; if the classical definition of the chi-square and Kullback distances had been used instead of (8) and (9) the value of $k$ would have been 1 instead of 2.

The total variation distance does not behave well in the sense of (1), since no upper bound in $\sum d^2(\mu_i, \nu_i)$ is available. However the inequalities of Proposition 5 show that, if there is a cut-off at time $t_n$ for the Hellinger distance then there is a cut-off at the same time for the total variation distance. Moreover, if each process $X_i$ has the same rate of exponential convergence for the total variation and either the chi-square or the Kullback distance then there is a cut-off in the sense of the total variation distance if the hypotheses of Theorem 3 hold for both distances. The cut-off phenomenon in the sense of the total variation distance can be related to Kakutani's theorem (e.g. Shiryaev [20], p. 528), which states that the distributions of two sequences of independent random variables are either mutually singular (their total variation distance is 1) or absolutely continuous. Our results can be seen as a finite dimensional version which describes where and how the transition from singularity takes place in exponentially converging $n$-tuples.

## 4. Cut-off times

In order to illustrate Theorem 3, let us consider a sequence of independent binary Markov jump processes. For $i \geq 1$, let $\alpha_i$ and $\rho_i$ be positive reals such that $0 < \alpha_i < \rho_i$. The process $X_i$ takes its values in $\{0, 1\}$. It jumps from 0 to 1 with rate $\alpha_i$, from 1 to 0 with rate $\rho_i - \alpha_i$. It is well known (see e.g. Section 7.5 in Bhat [3]) that the distribution of $X_i(t)$, starting from 0 at time 0 is Bernoulli with parameter

$$p_i(t) = \frac{\alpha_i}{\rho_i}\left(1 - \mathrm{e}^{-\rho_i t}\right).$$

The distances to equilibrium are easy to compute.

$$\text{Total variation:} \quad d_i(t) = \frac{\alpha_i}{\rho_i}\,\mathrm{e}^{-\rho_i t}$$

$$\text{Hellinger:} \quad d_i(t) = \sqrt{\frac{\alpha_i}{8(\rho_i - \alpha_i)}}\,\mathrm{e}^{-\rho_i t}\,(1 + o(1))$$

$$\text{Chi-square:} \quad d_i(t) = \sqrt{\frac{\alpha_i}{\rho_i - \alpha_i}}\,\mathrm{e}^{-\rho_i t}$$

$$\text{Kullback:} \quad d_i(t) = \sqrt{\frac{\alpha_i}{2(\rho_i - \alpha_i)}}\mathrm{e}^{-\rho_i t}(1 + o(1)).$$

What follows holds for any distance satisfying hypothesis (1), and also for the total variation distance, using Proposition 5. In order to simplify statements, we will describe as the 'cut-off time' the instant $\tau_n$ defined by (2), overlooking the fact that for the distances of Definition 4 the actual cut-off time will be $\tau_n/2$.

Take for instance $\alpha_i = \rho_i/2$. Then the uniform convergence hypothesis (3) is trivially satisfied, since $\log(d_i(t)/t) + \rho_i$ can be bounded by $g(t) = K/t$, with a suitable constant $K$. Hence $g(c\tau_n)/\rho_{(1,n)} = K/(c\tau_n\rho_{(1,n)})$ and hypotheses (4) and (5) are equivalent. Whether they are satisfied or not only depends on the sequence $(\rho_i)$. As already observed, if $0 < \liminf \rho_i < +\infty$, then $\tau_n$ tends to infinity and $\rho_{(1,n)}$ remains bounded away from 0, so Theorem 3 applies. If both $\rho_i$ and $\log i/\rho_i$ increase to infinity (e.g. $\rho_i = \log(\log(i + 2))$) then $\tau_n = \log n/\rho_n$ is a cut-off time. The sequence $(\rho_i)$ may also tend to 0. Take for instance $\rho_i = 1/\log(i + 1)$: again Theorem 3 applies; in this case the cut-off time $\tau_n$ is equivalent to $(\log(n))^2$.

If (3) holds and if the convergence rates $\rho_i$ converge to $\rho > 0$, then Theorem 3 applies. As we shall see, the cut-off time $\tau_n$ is equivalent to $\log n/\rho$, as if all rates were equal to $\rho$. It is natural to look for more general conditions under which $\log n/\rho$ is a cut-off time. In the setting of binary processes, Bon and Păltănea [4] propose sufficient conditions for a cut-off to occur at time $\log n/(2\liminf \rho_i)$ in the sense of the total variation distance. Their result can be seen as a particular case of Theorem 3 and Proposition 7 below.

**Proposition 7.** *For any positive $\rho$, denote by $N(\rho, n)$ the number of rates no larger than $\rho$ among $\rho_1, \ldots, \rho_n$:*

$$N(\rho, n) = \sum_{i=1}^{n} \mathbb{I}_{[0,\rho]}(\rho_i),$$

*where $\mathbb{I}_A$ denotes the indicator function of a set $A$. For $n \geq 1$, define $\rho_n^*$ as:*

$$\rho_n^* = \min\left\{\frac{\rho_i \log n}{\log N(\rho_i, n)}; i = 1, \ldots, n\right\},$$

*with* $1/\log(1) = +\infty$. *The instant* $\tau_n$ *defined by* (2) *is asymptotically equivalent to* $\tau'_n = \log n/\rho$ *if and only if the sequence* $(\rho^*_n)$ *converges to* $\rho > 0$.

**Proof.** Observe that $t_n$ can be expressed using $N(\rho_i, n)$ as follows.

$$\tau_n = \max\left\{\frac{\log i}{\rho_{(i,n)}}; i = 1, \dots, n\right\} = \max\left\{\frac{\log N(\rho_i, n)}{\rho_i}; i = 1, \dots, n\right\}.$$

The ratio $\tau'_n/\tau_n$ tends to 1 as $n$ tends to infinity if and only if

$$\begin{aligned}\rho &= \lim_{n\to\infty} \frac{\log n}{\tau_n}\\ &= \lim_{n\to\infty} \frac{\log n}{\max\left\{\frac{\log N(\rho_i,n)}{\rho_i};\ i = 1, \dots, n\right\}}\\ &= \lim_{n\to\infty} \min\left\{\frac{\rho_i \log n}{\log N(\rho_i, n)}; i = 1, \dots, n\right\}. \qquad \square\end{aligned}$$

Proposition 7 can be understood as follows. For $\rho > 0$, $N(\rho, n)$ is the number of coordinates in the $n$-tuple which converge more slowly than $\mathrm{e}^{-\rho t}$. If that number is large enough (in the sense that $\log n/\log N(\rho, n)$ remains bounded), then the sub-tuple of corresponding coordinates will converge only after $\log N(\rho, n)/\rho$. This will be the cut-off time for the full $n$-tuple, provided it is the latest convergence time of all sizeable sub-tuples. We think it interesting to further illustrate the idea of a cut-off for sub-tuples by the following proposition, which treats the case where the sequence $(\rho_i)$ has a finite number of accumulation values.

**Proposition 8.** *Let* $A$ *be a fixed integer. For* $a = 1, \dots, A$, *let* $k \mapsto \varphi_a(k)$ *be an increasing integer valued function. Denote by* $m_a(n)$ *the number of values of* $\varphi_a(k)$ *between 1 and* $n$:

$$m_a(n) = \sum_k \mathbb{I}_{[1,n]}(\varphi_a(k)).$$

*Assume that* $m_1(n) + \cdots + m_A(n) = n$ *and the* $\varphi_a(k)$ *are pairwise distinct. Assume moreover that for* $a = 1, \dots, A$, *the subsequence* $(\rho_{\varphi_a(k)})$ *converges to* $\varrho_a > 0$. *Denote by* $\sigma_n$ *the following real:*

$$\sigma_n = \max\left\{\frac{\log m_a(n)}{\varrho_a}; a = 1, \dots, A\right\},$$

*with* $\log 0 = -\infty$. *Then* $(\sigma_n)$ *and* $(\tau_n)$ *(defined by* (2)*) are asymptotically equivalent.*

**Proof.** The hypotheses imply that any value $\rho_i$ belongs to only one of the subsequences $(\rho_{\varphi_1(n)}), \dots, (\rho_{\varphi_A(n)})$. Without loss of generality, we will assume that $\varrho_1, \dots, \varrho_A$ are all distinct and ranked in increasing order. For $a = 1, \dots, A$, let $m^*_a(n) = m_1(n) + \cdots + m_a(n)$. Let

$$\sigma^*_n = \max\left\{\frac{\log m^*_a(n)}{\varrho_a}; a = 1, \dots, A\right\}.$$

We will prove first that $(\tau_n)$ and $(\sigma^*_n)$ are equivalent. We use the same expression for $\tau_n$ as in the proof of Proposition 7.

$$\tau_n = \max\left\{\frac{\log N(\rho_i, n)}{\rho_i}; i = 1, \ldots, n\right\}$$
$$= \max_{a=1,\ldots,A} \max\left\{\frac{\log N(\rho_{\varphi_a(k)}, n)}{2\rho_{\varphi_a(k)}}; k = 1, \ldots, m_a(n)\right\}.$$

Fix $\epsilon > 0$, small enough to ensure that all intervals $(\varrho_a - \epsilon\,,\ \varrho_a + \epsilon)$ are disjoint. For $i$ large enough, if $i = \varphi_a(k)$ then $\rho_i \in (\varrho_a - \epsilon\,,\ \varrho_a + \epsilon)$. Thus there exists an integer $K$ such that for $n$ large enough,

$$\max\left\{\frac{\log N(\rho_{\varphi_a(k)}, n)}{\rho_{\varphi_a(k)}}; k = 1, \ldots, m_a(n)\right\} \le \frac{\log(m_a^*(n) + K)}{(\varrho_a - \epsilon)}. \tag{14}$$

Take now $n$ such that $\rho_{\varphi_a(m_a(n)/2)}, \ldots, \rho_{\varphi_a(m_a(n))}$ are all smaller than $\varrho_a + \epsilon$, and consider the largest among these $m_a(n)/2$ values. This yields:

$$\max\left\{\frac{\log N(\rho_{\varphi_a(k)}, n)}{\rho_{\varphi_a(k)}}; k = 1, \ldots, m_a(n)\right\} \ge \frac{\log(m_{a-1}^*(n) + m_a(n)/2 - K')}{\varrho_a + \epsilon}, \tag{15}$$

for some fixed integer $K'$. It follows from (14) and (15) that $\tau_n$ is equivalent to $\sigma_n^*$.

It remains to prove that $\sigma_n^* = \sigma_n(1 + o(1))$. Obviously, $\sigma_n \le \sigma_n^*$. In the definition of $\sigma_n^*$, the maximum is reached either for $a = 1$, or for some $a > 1$ such that:

$$\frac{\log m_a^*(n)}{\varrho_a} \ge \frac{\log m_{a-1}^*(n)}{\varrho_{a-1}} \iff \frac{\log m_a^*(n)}{\log m_{a-1}^*(n)} \ge \frac{\varrho_a}{\varrho_{a-1}} > 1.$$

If $n$ is large enough, this implies:

$$\begin{aligned}
&\log m_a^*(n) > \log m_{a-1}^*(n) + \log 2 \\
&\quad\iff m_a^*(n) > 2m_{a-1}^*(n) \\
&\quad\iff m_a(n) > m_{a-1}^*(n) \\
&\quad\iff 2m_a(n) > m_a^*(n).
\end{aligned}$$

Therefore:

$$\sigma_n^* \le \max\left\{\frac{\log(2m_a(n))}{\varrho_a}; a = 1, \ldots, A\right\},$$

hence the result. □

Proposition 8 can be understood as follows. For $A = 1$, the sequence of rates converges to $\varrho_1$, and the cut-off occurs at the same time as if all rates were equal to $\varrho_1$. For $A > 1$, the $n$-tuple under consideration is made of $A$ independent sub-tuples, with respective cardinalities $m_1(n), \ldots, m_A(n)$. The $a$-th sub-tuple has a cut-off at time $\log m_a(n)/\varrho_a$. The cut-off time $\sigma_n$ for the full $n$-tuple is the latest among these times. This may have somewhat unexpected consequences. Take for instance $A = 2$ and $\varphi_1(k) = k^2$. One has $m_1(n) = \lfloor\sqrt{n}\rfloor$ and $m_2(n) = n - m_1(n) = n(1 + o(1))$. Take $\varrho_1 = 1$ and $\varrho_2 = 3$. The cut-off for the $n$-tuple occurs at time $s_n = \log n/2$, and not $\log n$ or $\log n/3$ as one could have thought.

## 5. The i.i.d. case

In the particular case where all coordinates converge at the same exponential rate $\rho$, the cut-off conditions (4) and (5) are naturally fulfilled and condition (3) only says that the sequences

$(\log d_i(t)/t)$ must converge uniformly in $i$ to $-\rho_i$. This is trivially the case when $d_i(t)$ is the same for all $i$ (in particular if the coordinates are i.i.d. processes). We will assume moreover that the $d_i(t)$ converge exponentially in a stronger sense than (3): there exist two positive reals $R$ and $\rho$ such that for all $i$,

$$\lim_{t\to+\infty} d_i(t)\mathrm{e}^{\rho t} = R. \tag{16}$$

In what follows, $\rho$ remains constant, but $R$ may depend on the distance. Different values will be denoted by $R_{TV}$, $R_H$, $R_{\chi^2}$, and $R_K$. Under the hypothesis (16), Proposition 6 yields more precise estimates of the distance $d^{(n)}(t)$ for $t$ around the cut-off instant. Here are the results for the Hellinger, chi-square and Kullback distances (the proofs are easy and will be omitted).

**Theorem 9.** (1) *Assume $d$ is the Hellinger distance and (16) holds.*

$$\lim_{n\to\infty} d^{(n)}\left(\frac{\log n}{2\rho}+u\right) = \left(1-\exp\left(-R_H^2\mathrm{e}^{-2\rho u}\right)\right)^{1/2}.$$

(2) *Assume $d$ is the chi-square distance and (16) holds.*

$$\lim_{n\to\infty} d^{(n)}\left(\frac{\log n}{2\rho}+u\right) = \left(\exp\left(R_{\chi^2}^2\mathrm{e}^{-2\rho u}\right)-1\right)^{1/2}.$$

(3) *Assume $d$ is the Kullback distance and (16) holds.*

$$\lim_{n\to\infty} d^{(n)}\left(\frac{\log n}{2\rho}+u\right) = R_K\mathrm{e}^{-\rho u}.$$

The total variation distance is particular, as already remarked. Even if we assume that (16) holds both for the total variation and another distance, Proposition 6 does not imply the convergence of $d^{(n)}(\log n/(2\rho)+u)$. Only bounds are obtained, which are easy consequences of (10) and (11) in Proposition 6, together with point (2) in Proposition 5.

**Theorem 10.** *Denote by $d_{TV,i}(t)$ and $d_{H,i}(t)$ the distances to equilibrium of the $i$-th component, measured as total variation and Hellinger distances respectively. Assume that there exist positive reals $R_{TV}$, $R_H$ and $\rho$ such that for all $i$,*

$$\lim_{t\to+\infty} d_{TV,i}(t)\mathrm{e}^{\rho t} = R_{TV} \qquad \text{and} \qquad \lim_{t\to+\infty} d_{H,i}(t)\mathrm{e}^{\rho t} = R_H.$$

*Let $d_{TV}^{(n)}(t)$ denote the total variation distance to equilibrium of the $n$-tuple $X^{(n)}(t)$. Then the following inequalities hold:*

$$1-\exp\left(-\frac{1}{2}R_{TV}^2\mathrm{e}^{-2\rho u}\right) \le \liminf_{n\to\infty} d_{TV}^{(n)}(\log(n)/(2\rho)+u),$$

*and:*

$$\limsup_{n\to\infty} d_{TV}^{(n)}(\log(n)/(2\rho)+u) \le \left(1-\exp\left(-2R_H^2\mathrm{e}^{-2\rho u}\right)\right)^{1/2}.$$

Theorem 10 suggests that the total variation distance to equilibrium of the $n$-tuple behaves as a double exponential when $u$ tends to $-\infty$,

$$\liminf_{n\to\infty} 1-d_{TV}^{(n)}(t_n+u) \le \exp\left(-\frac{1}{2}R_{TV}^2\mathrm{e}^{-2\rho u}\right)+o(u),$$

and a simple exponential when $u$ tends to $+\infty$,

$$\limsup_{n\to\infty} d_{TV}^{(n)}(t_n+u) \le \sqrt{2}R_H \mathrm{e}^{-\rho u} + o(u).$$

Applying Theorem 10 to binary processes as in the previous section yields bounds which are coherent with those obtained for the random walk on the $n$-dimensional hypercube by Diaconis et al. [8,6], and for finite state space reversible Markov chains by Ycart [23].

Further illustration is given by the M/M/$\infty$ birth–death process, and the Ornstein–Uhlenbeck diffusion.

M/M/$\infty$ *birth–death process*

The process $X$ is a birth–death process with constant birth rate $\alpha$ (from $k$ to $k+1$) and linear death rate $k\rho$, from $k$ to $k-1$ (see e.g. [9], Section 7a, Ch. XVII). If $X(0)=0$, the distribution of $X(t)$ is Poisson with parameter

$$\alpha(t) = \frac{\alpha}{\rho}\left(1-\mathrm{e}^{-\rho t}\right),$$

and the asymptotic distribution $\nu$ is also Poisson, with parameter $\alpha/\rho$. Theorems 9 and 10 apply, with

$$R_{TV} = R\left(\frac{\alpha}{\rho}\right), \qquad R_H = \sqrt{\frac{\alpha}{8\rho}}, \qquad R_{\chi^2} = \sqrt{\frac{\alpha}{\rho}}, \qquad R_K = \sqrt{\frac{\alpha}{2\rho}},$$

where

$$R(a) = \frac{\mathrm{e}^{-a}}{\lfloor a\rfloor!}\, a^{\lfloor a\rfloor+1}.$$

The cut-off phenomenon for the family of M/M/$\infty$ processes indexed by the initial state $n$ was studied by Martínez and Ycart [13] in the context of birth–death processes on trees. In Proposition 6.1 they found bounds analogous to those given above for the total variation distance (see also p. 293 of [22]).

*Ornstein–Uhlenbeck diffusion*

The process $X$ is a solution of the following stochastic differential equation (see e.g. [10], example 4(b), Ch. X):

$$\begin{cases} \mathrm{d}X(t) = \alpha\sqrt{2\rho}\,\mathrm{d}B_t - \rho X(t)\mathrm{d}t, \\ X(0) = x_0, \end{cases}$$

where $\alpha, \rho > 0$ and $\{B_t\,, t \ge 0\}$ is the standard Brownian motion. The distribution of $X(t)$ is Gaussian with parameters

$$m(t) = x_0 \mathrm{e}^{-\rho t} \qquad \text{and} \qquad v(t) = \alpha^2\left(1-\mathrm{e}^{-2\rho t}\right),$$

and the asymptotic distribution $\nu$ is also Gaussian, with parameters 0 and $\alpha^2$. Theorems 9 and 10 apply with

$$R_{TV} = \frac{|x_0|}{\alpha\sqrt{2\pi}}, \qquad R_H = \frac{|x_0|}{\alpha\sqrt{8}}, \qquad R_{\chi^2} = \frac{|x_0|}{\alpha}, \qquad R_K = \frac{|x_0|}{\alpha\sqrt{2}}.$$

The cut-off for Ornstein–Uhlenbeck diffusions has been studied by Lachaud [12], who relates it to the asymptotic distribution of the hitting time of 0 via the empirical mean of the $n$-tuple.

## Acknowledgments

J. Barrera acknowledges support from Programa Cientifico Milenio P01-005, and a CONICYT Ph.D. fellowship.

## References

[1] D. Aldous, P. Diaconis, Shuffling cards and stopping times, Amer. Math. Monthly 93 (5) (1986) 333–348.
[2] D. Aldous, P. Diaconis, Strong uniform times and finite random walks, Adv. Appl. Math. 8 (1987) 69–97.
[3] U.N. Bhat, Elements of Applied Stochastic Processes, John Wiley & Sons Inc., New York, 1972.
[4] J.-L. Bon, E. Păltănea, Convergence of the number of failed components in a Markov system with nonidentical components, J. Appl. Probab. 38 (4) (2001) 882–897.
[5] P. Diaconis, The cut-off phenomenon in finite Markov chains, Proc. Natl. Acad. Sci. USA 93 (4) (1996) 1659–1664.
[6] P. Diaconis, R.L. Graham, J.A. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, Random Structures Algorithms 1 (1) (1990) 51–72.
[7] P. Diaconis, L. Saloff-Coste, Separation cut-offs for Birth and Death chains model, Tech. Rep. 2005-14. Dept. of Statistics, Stanford Univ., 2005.
[8] P. Diaconis, M. Shahshahani, Time to reach stationarity in the Bernoulli–Laplace diffusion model, SIAM J. Math. Anal. 18 (1) (1987) 208–218.
[9] W. Feller, An Introduction to Probability Theory and its Applications. Vol. I, 3rd edn, John Wiley & Sons Inc., New York, 1968.
[10] W. Feller, An Introduction to Probability Theory and its Applications. Vol. II, 2nd edn, John Wiley & Sons Inc., New York, 1971.
[11] A. Gibbs, F. Su, On choosing and bounding probability metrics, Internat. Statist. Rev. 70 (3) (2002) 419–435.
[12] B. Lachaud, Cut-off and hitting times for a sample of Ornstein–Uhlenbeck processes and its average, J. Appl. Probab. 42 (4) (2005) 1069–1080.
[13] S. Martínez, B. Ycart, Decay rates and cut-off for convergence and hitting times of Markov chains with countably infinite state space, Adv. in Appl. Probab. 33 (1) (2001) 188–205.
[14] D. Pollard, A User's Guide to Measure Theoretic Probability, Cambridge University Press, 2001.
[15] U. Porod, The cut-off phenomenon for random reflections, Ann. Probab. 24 (1) (1996) 74–96.
[16] R.-D. Reiss, Approximate Distributions of Order Statistics, in: Springer Series in Statistics, Springer-Verlag, New York, 1989.
[17] J.S. Rosenthal, On generalizing the cut-off phenomenon for random walks on groups, Adv. Appl. Math. 16 (3) (1995) 306–320.
[18] L. Saloff-Coste, Lectures on finite Markov chains, in: Lectures on Probability Theory and Statistics (Saint-Flour, 1996), in: Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.
[19] L. Saloff-Coste, Random walks on finite groups, in: Probability on Discrete Structures, in: Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 263–346.
[20] A.N. Shiryaev, Probability, in: Graduate Texts in Math., vol. 95, Springer-Verlag, New York, 1996.
[21] F. Su, Methods for quantifying rates of convergence for random walks on groups, Ph.D. Thesis, Harvard University, 1995.
[22] B. Ycart, Cutoff for Markov chains: some examples and applications, in: Complex Systems (Santiago, 1998), in: Nonlinear Phenom. Complex Systems, vol. 6, Kluwer Acad. Publ., Dordrecht, 2001, pp. 261–300.
[23] B. Ycart, Cutoff for samples of Markov chains, ESAIM Probab. Stat. 3 (1999) 89–106 (electronic).
[24] B. Ycart, Stopping tests for Markov chain Monte-Carlo methods, Methodol. Comput. Appl. Probab. 2 (1) (2000) 23–36.