Alea 7, 65–77 (2010)

Separation and coupling cutoffs for tuples of independent Markov processes

Stephen B. Connor

Department of Mathematics, University of York, York YO10 5DD, UK
URL: http://maths.york.ac.uk/www/sbc502
E-mail address: sbc502@york.ac.uk

Abstract. We consider an n-tuple of independent ergodic Markov processes, each of which converges (in the sense of separation distance) at an exponential rate, and obtain a necessary and sufficient condition for the n-tuple to exhibit a separation cutoff. We also provide general bounds on the (asymmetric) window size of the cutoff, and indicate links to classical extreme value theory.

1. Introduction

It is well known that a large number of Markov chains exhibit cutoff phenomena when converging to stationarity. This phenomenon occurs when the distance of the chain from equilibrium (measured using, for example, the total-variation metric or separation distance) stays close to its maximum value for some time, before dropping relatively fast and tending quickly to zero. Such behaviour was first identified for the transposition shuffle on the symmetric group (Diaconis and Shahshahani, 1981), and has since been shown to hold for many natural sequences of random walks on groups (see Saloff-Coste, 2004 for a review).

In a recent paper, Barrera et al. (2006) consider n-tuples of independent processes, and give sufficient conditions for cutoffs to hold when distance from stationarity is measured using total-variation, Hellinger, chi-square and Kullback distances, under the assumption that each coordinate process converges exponentially fast. In the particular case when all coordinates converge at the same rate, the window size of the cutoff (to be defined below) is also determined.

In this paper we consider the separation distance of such n-tuples from stationarity and give conditions (very similar to those in Barrera et al., 2006) guaranteeing the existence of a separation cutoff. Our approach is slightly different from that of Barrera *et al.*, however: instead of working with a set of ordered exponential rates

Received by the editors August 12, 2009; accepted March 19, 2010.
2000 *Mathematics Subject Classification.* 60B10, 60J27.
Key words and phrases. Cutoff phenomenon, cutoff window, strong stationary time, Lambert W-function.

we choose to work with discrete probability measures. This enables us to relate cutoff to convergence of (suitably scaled versions of) these measures. Furthermore, we are able to provide general bounds on the window size of the cutoff (not only when all coordinates converge at the same rate). In particular, we show that in general the right-hand side of the cutoff window may be of significantly larger order than the left.

The paper is organised as follows. In Section 2 we recall the definitions of total-variation and separation distance, and make formal the notion of cutoff time and window size. In Section 3 we present our main result concerning the existence of a separation cutoff, and prove general bounds on the window-size of such a cutoff. We then apply this to the example of a continuous-time random walk on the hypercube $\mathbb{Z}_2^n$, where each coordinate may move at a different rate, and present a specific case which shows that our general window-size bounds are tight. Some links to classical extreme value theory are also highlighted. Finally, in Section 4, we briefly consider the notion of a *coupling cutoff* for two such n-tuples.

2. The cutoff phenomenon

In keeping with the notation of Diaconis and Saloff-Coste (2006), for two probability measures μ and ν on a finite space $(E, \mathcal{E})$ we shall write $D(\mu, \nu)$ for a general notion of distance between them. One example is *total-variation distance*

$$D(\mu,\nu) = \|\mu - \nu\|_{\mathrm{TV}} = \sup_{A\in\mathcal{E}} |\mu(A) - \nu(A)| \ ,$$

while the *separation distance* is defined to be

$$D(\mu,\nu) = \mathrm{sep}\,(\mu,\nu) = \max_{x\in E} \left\{ 1 - \frac{\mu(x)}{\nu(x)} \right\} \ .$$

Note that separation is not a metric due to its asymmetry. Both of these distances take values in $[0, 1]$, and it is simple to show (Aldous and Diaconis, 1987) that

$$\|\mu - \nu\|_{\mathrm{TV}} \le \mathrm{sep}\,(\mu,\nu) \, .$$

Separation distance is intimately linked with the notion of strong stationary times. Let X be a Markov chain with time-t distribution P^t and stationary distribution π.

Definition 2.1. A *strong stationary time (SST)* T is a randomized stopping time for X such that

$$\mathbb{P}(X_t = k \,|\, T \le t) = \pi(k) \, , \qquad \text{for all } 0 \le t < \infty, \ k \in E.$$

If T is a SST for X, then (Aldous and Diaconis, 1987)

$$\mathrm{sep}\,(P^t, \pi) \le \mathbb{P}(T > t) \, , \qquad \text{for all } t \ge 0. \tag{2.1}$$

An optimal SST is one which achieves equality in (2.1) for all $t \ge 0$: existence is demonstrated in Aldous and Diaconis (1987) (in discrete-time).

We may now define the notion of a cutoff phenomenon for a given distance function D (including, but not restricted to, those distances defined above).

Definition 2.2. For $n \ge 1$, let X_n be a stochastic process taking values on a finite space $(E_n, \mathcal{E}_n)$, with time-t distribution P_n^t and stationary distribution π_n.

We say that the sequence $\{E_n, X_n, \pi_n\, ;\, n = 1, 2, \dots\}$ exhibits a (τ_n, b_n)-*D-cutoff* if $\tau_n, b_n > 0$ satisfy $b_n = o(\tau_n)$ and

$$d_-(c) = \liminf_{n\to\infty} D(P_n^{\tau_n + cb_n}, \pi_n) \quad \text{satisfies} \lim_{c\to-\infty} d_-(c) = 1\,,$$

$$d_+(c) = \limsup_{n\to\infty} D(P_n^{\tau_n + cb_n}, \pi_n) \quad \text{satisfies} \lim_{c\to\infty} d_+(c) = 0\,.$$

Here τ_n is called the *cutoff time*, and b_n will be referred to as the *window* of the cutoff. (We may simply say that the sequence X_n exhibits a τ_n-D-cutoff when we are not concerned with the window size b_n.)

Furthermore, it is possible to analyse the window size in more detail by considering separately the windows either side of the cutoff time τ_n. That is, instead of using a single sequence b_n to establish convergence in equations (4.2) and (4.3), we can consider each convergence statement separately.

Definition 2.3. Suppose the sequence $\{E_n, X_n, \pi_n\}$ exhibits a τ_n-D-cutoff. If there exist sequences b_n^L, b_n^R with $\max\{b_n^L, b_n^R\} = o(\tau_n)$, such that

$$d_-^L(c) = \liminf_{n\to\infty} D(P_n^{\tau_n + cb_n^L}, \pi_n) \quad \text{satisfies} \lim_{c\to-\infty} d_-^L(c) = 1,$$

$$\text{and} \quad d_+^R(c) = \limsup_{n\to\infty} D(P_n^{\tau_n + cb_n^R}, \pi_n) \quad \text{satisfies} \lim_{c\to\infty} d_+^R(c) = 0,$$

then b_n^L will be called a *left-window* and b_n^R a *right-window* of the cutoff.

To the best of the author's knowledge, the only published article to identify a difference between the left and right windows of a cutoff phenomenon is that of Chen and Saloff-Coste (2008). For the processes considered in this paper however, such a distinction will prove to be rather important.

3. Separation cutoff for n-tuples of independent processes

Let $X_n = (X_n^1, X_n^2, \dots, X_n^n)$ be an n-tuple of independent, continuous-time Markov chains on a finite space $(E_n, \mathcal{E}_n)$, with initial state $x_n = (x_n^1, \dots, x_n^n)$ and stationary distribution $\pi_n = \pi_n^1 \times \dots \times \pi_n^n$. Let

$$\text{sep}_n(t) = \text{sep}(P_n^t, \pi_n) \quad \text{and} \quad \text{sep}_n^i(t) = \text{sep}(P_{n,i}^t, \pi_n^i)\,,$$

where $P_{n,i}^t$ denotes the distribution of X_n^i at time t.

Proposition 3.1. *For all $t \geq 0$,*

$$\text{sep}_n(t) = 1 - \prod_{i=1}^n (1 - \text{sep}_n^i(t))\,.$$

Proof: The independence of the chains implies that

$$1 - \text{sep}_n(t) = \min_{y_n^1, \dots, y_n^n} \prod_{i=1}^n \frac{P_{n,i}^t(x_n^i, y_n^i)}{\pi_n^i(y_n^i)} = \prod_{i=1}^n (1 - \text{sep}_n^i(t))\,,$$

since each term in the product may be minimised individually. □

If T_n^i is an optimal SST for X_n^i $(1 \leq i \leq n)$, then letting $T_n = \max T_n^i$ one can check that T_n is a SST for the n-tuple. Proposition 3.1 shows that

$$\text{sep}_n(t) = \mathbb{P}(T_n > t) \quad \text{for all } t \geq 0,$$

and it follows that T_n is an optimal SST for X_n.

As in Barrera et al. (2006), we are interested in processes for which each component X_n^i converges at an exponential rate λ_n^i, although now this convergence is to be measured using separation distance. Rather than following the route of Barrera et al. (2006) and using an ordered set of rates $\{\lambda_{(i,n)}\}$, we prefer to work instead with discrete probability measures μ_n on $(0,\infty)$, where

$$\mu_n(\{\lambda\}) = \frac{1}{n}\,\#\{\lambda_n^i \,:\, \lambda_n^i = \lambda\}\,.$$

(This is similar to the use of *design measures* in design theory, see e.g. St. John and Draper, 1975.) The result of this will be that the existence of a separation cutoff can be directly related to the convergence of appropriately scaled versions of μ_n as $n \to \infty$. We define κ_n by

$$\kappa_n = \min\left\{\lambda > 0 \,:\, \mu_n(0,\lambda] > 0\right\}\,.$$

The main result of this paper is the following:

Theorem 3.2. *Let X_n be an n-tuple of independent ergodic Markov processes, each of whose components satisfies $|g_{\lambda_n^i}(t)| \le g(t)$ for all $t \ge 0$, where $g_{\lambda_n^i}$ is defined by*

$$\frac{\log \operatorname{sep}_n^i(t)}{t} + \lambda_n^i = g_{\lambda_n^i}(t)\,,$$

and where g is a bounded continuous function satisfying $g(t) \le O(t^{-1})$. As above, let μ_n be the discrete probability measure describing the set $\{\lambda_n^i\}$, with support $[\kappa_n,\infty)$.

(1) *The sequence of n-tuples X_n exhibits a separation cutoff at time*

$$\tau_n = \max_{\lambda \ge \kappa_n}\left\{\frac{\log(n\mu_n(0,\lambda])}{\lambda}\right\}$$

if and only if $\tau_n\kappa_n \to \infty$;

(2) *The window of the separation cutoff is in general asymmetric: the left side is at most $O(1/\kappa_n)$, and the right side is bounded above by $W(\tau_n\kappa_n)/\kappa_n$, where W is the Lambert W-function.*

As remarked in Barrera et al. (2006), under the conditions of Theorem 3.2 the spectral gap of X_n is equal to κ_n and the separation-mixing time equivalent to τ_n. Thus Theorem 3.2(i) shows that the conjecture of Peres (reported in Diaconis and Saloff-Coste, 2006; Chen and Saloff-Coste, 2008) holds true for separation cutoff for the processes considered here.

Consider an n-tuple X_n satisfying the conditions of Theorem 3.2. Using μ_n and Proposition 3.1, the separation distance at time t may be written as

$$\operatorname{sep}_n(t) = 1 - \exp\left(n\int_{\kappa_n}^{\infty}\log(1 - e^{-t(\lambda - g_\lambda(t))})\mu_n(d\lambda)\right)\,. \tag{3.1}$$

One benefit of working with separation distance in this setting is that equation (3.1) holds for any μ_n, whereas there is no longer a simple exact expression for the total-variation distance between P_n^t and π_n when the rates λ_n^i are not identical (Barrera et al., 2006).

The proof of Theorem 3.2 will be established by the results of Proposition 3.3, Lemma 3.6 and Theorem 3.7 below.

Proposition 3.3. *For the sequence $\{X_n\}$ to exhibit a τ_n-separation cutoff, it is necessary for $\tau_n \kappa_n \to \infty$.*

Proof: Restricting attention to the mass at κ_n in equation (3.1) immediately implies that, for any $c > 1$,

$$\begin{aligned}\operatorname{sep}_n(c\tau_n) &\geq \exp(-c\tau_n(\kappa_n - g_{\kappa_n}(c\tau_n))) \\ &\geq \exp(-c\tau_n\kappa_n)\exp(-c\tau_n g(c\tau_n)).\end{aligned}$$

For a separation cutoff to hold at τ_n, we require that $\operatorname{sep}_n(c\tau_n) \to 0$ for all fixed $c > 1$: this fails, however, if $\tau_n\kappa_n \not\to \infty$ (since the final exponential term above is bounded away from zero due to our conditions on g). □

Now, given a measure μ_n, define τ_n by

$$\tau_n = \max_{\lambda \geq \kappa_n}\left\{\frac{\log(n\mu_n(0,\lambda])}{\lambda}\right\} = \frac{\log(n\mu_n(0,\lambda_n^*])}{\lambda_n^*}, \tag{3.2}$$

where $\lambda_n^* \in [\kappa_n, \infty)$ is defined by this last equality. (If there are two or more values of λ achieving the maximum in equation (3.2) then we shall (arbitrarily) always take λ_n^* to be the minimum of these values.) Given λ_n^*, we may define a new measure ν_n on $(0,\infty)$ as follows:

$$\nu_n(\{x\}) = \frac{\mu_n(\{\lambda_n^* x\})}{\mu_n(0,\lambda_n^*]}. \tag{3.3}$$

This measure has total mass $(\mu_n(0,\lambda_n^*])^{-1} \in [1,n]$ and satisfies $\nu_n(0,1] = 1$. The idea behind this scaling is as follows: λ_n^* describes in some sense the 'critical point' of μ_n – it will be shown that if $\tau_n\kappa_n \to \infty$ then any mass μ_n places to the left of λ_n^* will not influence the separation cutoff time. For ease of notation we define

$$\beta_n = n\mu_n(0,\lambda_n^*] \in [1,n].$$

Lemma 3.4. *If $\tau_n\kappa_n \to \infty$ then:*

(i) $\beta_n \to \infty$;

(ii) $\nu_n(0,1] \xrightarrow{w} \delta_1$ *(where $\xrightarrow{w}$ denotes weak convergence).*

Proof: (i) $\beta_n = \exp(\tau_n\lambda_n^*) \geq \exp(\tau_n\kappa_n) \to \infty$ by assumption.

(ii) By definition of τ_n (3.2),

$$\frac{\log(n\mu_n(0,\lambda])}{\lambda} \leq \frac{\log\beta_n}{\lambda_n^*} \quad \text{for all } \lambda \geq \kappa_n.$$

Thus for all $x \geq \kappa_n/\lambda_n^*$,

$$\frac{\log(n\mu_n(0,x\lambda_n^*])}{x} \leq \log\beta_n.$$

This yields

$$n\mu_n(0,x\lambda_n^*] \leq \beta_n^x \quad \text{for all } x \geq \kappa_n/\lambda_n^*. \tag{3.4}$$

Hence

$$\nu_n(0,x] = \frac{\mu_n(0,x\lambda_n^*]}{\mu_n(0,\lambda_n^*]} = \frac{n\mu_n(0,x\lambda_n^*]}{\beta_n} \leq \beta_n^{x-1}, \tag{3.5}$$

where the inequality follows from (3.4). Thus for all $\varepsilon \in (0,1)$,

$$\nu_n(0,1-\varepsilon] \leq \beta_n^{-\varepsilon} \xrightarrow[n\to\infty]{} 0.$$

Since $\nu_n(0,1] = 1$ for all n, this proves the required convergence. □

This makes more precise what is meant by λ_n^* describing the 'critical point' of μ_n. Under the assumption that $\tau_n \kappa_n \to \infty$, the measures ν_n converge weakly to δ_1 on $(0,1]$: this is exactly the sort of behaviour to be expected if the sequence $\{\lambda_n^*\}$ captures information about the cutoff time. Lemma 3.6 and Theorem 3.7 make this observation exact: their proofs rely on Proposition 3.5, which describes the behaviour of the function θ_n defined by

$$\theta_n(t) = \beta_n \int_{\kappa_n/\lambda_n^*}^{\infty} \exp(-t\lambda_n^* \lambda)\nu_n(d\lambda)\,. \tag{3.6}$$

Proposition 3.5. *The following inequalities hold for all $t \geq \log 2/\kappa_n$:*

$$1 - \exp(-e^{-tg(t)}\theta_n(t)) \leq \mathrm{sep}_n\,(t) \leq 1 - \exp(-2e^{2tg(t)}\theta_n(t))\,. \tag{3.7}$$

Note that if $\tau_n\kappa_n \to \infty$, Proposition 3.5 implies that the behaviour of $\mathrm{sep}^{(n)}$ around τ_n is determined by that of θ_n.

Proof: Using the measure ν_n, the separation in equation (3.1) may be rewritten as follows:

$$\mathrm{sep}_n\,(t) = 1 - \exp\left(\beta_n \int_{\kappa_n/\lambda_n^*}^{\infty} \log\left(1 - e^{-t(\lambda_n^*\lambda - g_{\lambda_n^*\lambda}(t))}\right)\nu_n(d\lambda)\right)\,. \tag{3.8}$$

Now note that the following simple inequality holds for $0 \leq x \leq 1/2$:

$$-x - x^2 \leq \log(1-x) \leq -x\,.$$

Applying this inequality to the log term in equation (3.8), and bounding $g_{\lambda_n^*\lambda}(t)$ by $\pm g(t)$, shows that for all $t \geq \log 2/\kappa_n$:

$$1 - \exp(-e^{-tg(t)}\theta_n(t)) \leq \mathrm{sep}_n\,(t) \leq 1 - \exp(-e^{tg(t)}\theta_n(t) - e^{2tg(t)}\theta_n(2t))\,.$$

Finally, observe from (3.6) that $\theta_n(2t) \leq \theta_n(t)$ for all $t \geq 0$: the result follows immediately. □

We are now in a position to prove the existence of the left-hand side of the cutoff in Theorem 3.2.

Lemma 3.6. *Suppose that $\tau_n\kappa_n \to \infty$, with τ_n defined as in (3.2). Let $b_n^L = 1/\lambda_n^* \leq O(1/\kappa_n)$. Then*

$$\mathrm{sep}_-^L(c) = \liminf_{n\to\infty} \mathrm{sep}_n\,(\tau_n + cb_n^L) \quad \textit{satisfies} \quad \lim_{c\to-\infty} \mathrm{sep}_-^L(c) = 1.$$

(Note that since $\tau_n\kappa_n \to \infty$, $b_n^L = o(\tau_n)$, as is required for any candidate window-size.)

Proof: Consider $\theta_n(\tau_n + c/\lambda_n^*)$, for fixed $c \in \mathbb{R}$. Since $\tau_n\kappa_n \to \infty$ it follows from Lemma 3.4(i) that for any fixed $c \in \mathbb{R}$,

$$\tau_n + \frac{c}{\lambda_n^*} = \frac{\log\beta_n + c}{\lambda_n^*} \geq 0$$

for large enough n. By definition of τ_n, with $\tau_n + c/\lambda_n^* \geq 0$:

$$\begin{aligned}\theta_n(\tau_n + c/\lambda_n^*) &= \beta_n \int_{\kappa_n/\lambda_n^*}^{\infty} \exp(-\left[\tau_n + c/\lambda_n^*\right] \lambda_n^* \lambda) \nu_n(d\lambda) \\ &\geq \beta_n \int_{\kappa_n/\lambda_n^*}^{1} \exp(-\left[\tau_n + c/\lambda_n^*\right] \lambda_n^* \lambda) \nu_n(d\lambda) \\ &\geq \beta_n \nu_n(0,1] \left(\frac{e^{-c}}{\beta_n}\right) = e^{-c}\,. \end{aligned} \tag{3.9}$$

Combining Proposition 3.5 and inequality (3.9) shows that for all $c \in \mathbb{R}$,

$$\mathrm{sep}_-^L(c) \geq 1 - \limsup_{n\to\infty} \exp(-e^{-\gamma_n^L(c)} \theta_n(\tau_n + c/\lambda_n^*))\,,$$

where

$$\gamma_n^L(c) = (\tau_n + cb_n^L) g(\tau_n + cb_n^L) \underset{n\to\infty}{\sim} \tau_n g(\tau_n) = O(1)\,. \tag{3.10}$$

Hence

$$\mathrm{sep}_-^L(c) \geq 1 - \exp(-Me^{-c})\,,$$

for some finite constant $M > 0$, and thus $\mathrm{sep}_-^L(c) \to 1$ as $c \to -\infty$, as claimed. □

It turns out that the general bound for the right-window of the cutoff is significantly larger than that for the left. Theorem 3.7, which completes the proof of Theorem 3.2, makes use of the Lambert W-function (see Corless et al., 1996). This is the function defined for all $x \in \mathbb{C}$ by

$$W(x)e^{W(x)} = x\,.$$

$W(x)$ is positive and increasing for $x \in \mathbb{R}^+$, with $W(x) \sim \log(x/\log x)$ as $x \to \infty$.

Theorem 3.7. *Suppose that $\tau_n\kappa_n \to \infty$, with τ_n defined as in (3.2). Then*

$$\mathrm{sep}_+^R(c) = \limsup_{n\to\infty} \mathrm{sep}_n\left(\tau_n + cW(\tau_n\kappa_n)/\kappa_n\right) \quad \textit{satisfies} \lim_{c\to\infty} \mathrm{sep}_+^R(c) = 0.$$

Proof: In order for a sequence b_n^R to be a right-window for the separation cutoff, it is sufficient to show that $\theta_n(\tau_n + cb_n^R) \leq h(c)$ for sufficiently large n, where $h(c) \to 0$ as $c \to \infty$. For then, using inequality (3.7) it follows that

$$\begin{aligned}\mathrm{sep}_+^R(c) &= \limsup_{n\to\infty} \mathrm{sep}_n\left(\tau_n + cb_n^R\right) \\ &\leq 1 - \liminf_{n\to\infty} \exp(-2e^{2\gamma_n^R(c)} \theta_n(\tau_n + cb_n^R))\,,\end{aligned}$$

where $\gamma_n^R(c)$ is defined analogously to (3.10). Thus, for some finite M,

$$\mathrm{sep}_+^R(c) \leq 1 - \exp(-Mh(c)) \xrightarrow[c\to\infty]{} 0\,.$$

We therefore search for an upper bound on the function $\theta_n(\tau_n + cb_n^R)$ for fixed $c > 0$. The form of τ_n, and use of integration by parts, yield the following:

$$\begin{aligned}\theta_n(\tau_n + cb_n^R) &= \beta_n \int_{\kappa_n/\lambda_n^*}^{\infty} \left(\frac{e^{-cb_n^R\lambda_n^*}}{\beta_n}\right)^{\lambda} \nu_n(d\lambda) \\ &= \beta_n \left[\left(\frac{e^{-cb_n^R\lambda_n^*}}{\beta_n}\right)^{\lambda} \nu_n(0,\lambda]\right]_{\kappa_n/\lambda_n^*}^{\infty} \\ &\quad + \beta_n \log(\beta_n e^{cb_n^R\lambda_n^*}) \int_{\kappa_n/\lambda_n^*}^{\infty} \left(\frac{e^{-cb_n^R\lambda_n^*}}{\beta_n}\right)^{\lambda} \nu_n(0,\lambda]\, d\lambda\,. \end{aligned} \tag{3.11}$$

Now, for $c > 0$, this first term is negative for all n. Discarding this, and using inequality (3.5) to bound $\nu_n(0,\lambda]$ in the second term, we see that

$$\begin{aligned}\theta_n(\tau_n + cb_n^R) &\le \beta_n \log(\beta_n e^{cb_n^R\lambda_n^*}) \int_{\kappa_n/\lambda_n^*}^{\infty} \left(\frac{e^{-cb_n^R\lambda_n^*}}{\beta_n}\right)^{\lambda} \beta_n^{\lambda-1}\, d\lambda \\ &= \log(\beta_n e^{cb_n^R\lambda_n^*}) \frac{e^{-cb_n^R\kappa_n}}{cb_n^R\lambda_n^*} \\ &= e^{-cb_n^R\kappa_n}\left(\frac{\tau_n}{cb_n^R} + 1\right), \quad \text{by definition of } \tau_n. \end{aligned} \tag{3.12}$$

Since b_n^R must satisfy $b_n^R = o(\tau_n)$, this upper bound tends to infinity with n unless $cb_n^R\kappa_n \ge W(\tau_n\kappa_n)$, by definition of the Lambert W-function. Thus, with $b_n^R = W(\tau_n\kappa_n)/\kappa_n$, (3.12) satisfies

$$e^{-cb_n^R\kappa_n}\left(\frac{\tau_n}{cb_n^R} + 1\right) \xrightarrow[n\to\infty]{} h(c) = \begin{cases}\infty & 0 < c < 1\\ 1 & c = 1\\ 0 & c > 1\,.\end{cases}$$

It follows that for $c > 1$, $\theta_n(\tau_n + cW(\tau_n\kappa_n)/\kappa_n) \to 0$ as $n \to \infty$, and so

$$\mathrm{sep}_+^R(c) = \limsup_{n\to\infty} \mathrm{sep}_n\left(\tau_n + cW(\tau_n\kappa_n)/\kappa_n\right) = 0\,.$$

Therefore $b_n^R = W(\tau_n\kappa_n)/\kappa_n$ is a right-window of the cutoff, as claimed. □

This bound on the right-window is significantly larger than that for the left-window. Since $\tau_n\kappa_n$ necessarily tends to infinity when a separation cutoff holds, it follows that

$$O(1/\kappa_n) < \frac{W(\tau_n\kappa_n)}{\kappa_n} = o(\tau_n)\,.$$

3.1. *Random walks on* $\mathbb{Z}_2^n$. Let $\mathbb{Z}_2^n$ be the group of binary n-tuples under coordinate-wise addition modulo 2: this can be viewed as the vertices of an n-dimensional hypercube. A continuous-time random walk X_n on $\mathbb{Z}_2^n$ may be defined as follows. Let $\{\Lambda_n^i : 1 \le i \le n\}$ be a set of independent Poisson processes, with the rate of Λ_n^i equal to $2\rho_n^i$: whenever there is an incident on Λ_n^i, with probability $1/2$ the i^{th} coordinate, $X_{n,i}$, is flipped to its opposite value. The unique equilibrium distribution of X_n is the uniform distribution on $\mathbb{Z}_2^n$, U_n.

Let T_n^i be the time of the first incident on Λ_n^i. It is simple to show that T_n^i is an optimal SST for $X_{n,i}$, with

$$\mathbb{P}(T_n^i > t) = e^{-2t\rho_n^i}\,.$$

Thus $T_n = \max T_n^i$ is an optimal SST for X_n. (This is similar in flavour to the optimal SST for the continuous-time birth-death processes of Diaconis and Saloff-Coste, 2006: there the SST is given by a sum of exponential random variables of varying rates, rather than their maximum.)

It follows that X_n satisfies the conditions of Theorem 3.2, with

$$\lambda_n^i = 2\rho_n^i\,, \quad \text{and} \quad g \equiv 0\,.$$

Writing $\rho_n^* = \min\left\{\rho_n^i\right\}$, the sequence X_n therefore exhibits a separation cutoff at time

$$\tau_n = \max_{\rho \geq \rho_n^*}\left\{\frac{\log(n\mu_n(0,2\rho])}{2\rho}\right\}$$

if and only if $\tau_n\rho_n^* \to \infty$. In this case, $\tau_n = 2\hat{\tau}_n$, where $\hat{\tau}_n$ is the total-variation cutoff time according to Theorem 12 of Barrera et al. (2006).

For many simple examples, such as the symmetric random walk for which all coordinates jump at rate 1, the result of Theorem 3.7 gives an extremely conservative bound for the right-window. (Simple direct calculations show that a $(\log n/2, 1)$-separation cutoff holds, whereas the bound on b_n^R from Theorem 3.7 tends to infinity with n.) However, the following example shows that the bound of Theorem 3.7 can be achieved, and so cannot be improved upon in general.

Example 3.8. Consider the sequence of random walks on $\mathbb{Z}_2^n$ with $\rho_n^i = \max\{1, 2\log_n(i)\}$. The associated probability measure for X_n is

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{\max\{2,\,4\log_n(i)\}}\,.$$

The measure μ_n places all its mass in the interval $[2,4]$, with $\kappa_n = 2$ and

$$\mu_n[2,\lambda] = \frac{\lfloor n^{\lambda/4}\rfloor}{n} \sim n^{\lambda/4-1}, \quad \text{for all } \lambda \in [2,4].$$

For this sequence,

$$\tau_n = \max_{2\leq\lambda\leq 4}\left\{\frac{\log(n\mu_n[2,\lambda])}{\lambda}\right\} = \max_{2\leq\lambda\leq 4}\left\{\frac{\log(n^{\lambda/4})}{\lambda}\right\} = \frac{\log n}{4}\,.$$

Note that this maximum is attained at all $\lambda \in [2,4]$: we arbitrarily take $\lambda_n^* = 2$ to be the minimum of these values. This gives $\beta_n = \sqrt{n}$, and hence $\nu_n[1,x] = n^{(x-1)/2}$ for $x \in [1,2]$. Since $\tau_n \to \infty$ as $n \to \infty$, this random walk exhibits a τ_n-separation cutoff. Now, by Lemma 3.6, the left-window of this separation cutoff is bounded above by $1/\lambda_n^* = 1/2$. However, for fixed $c > 0$ and some sequence $b_n^R = o(\tau_n)$, integration by parts as in equation (3.11) yields the following:

$$\theta_n\left(\frac{\log n}{4}+cb_n^R\right)\sim(e^{-4cb_n^R}-e^{-2cb_n^R}) + \sqrt{n}\log(\sqrt{n}\,e^{2cb_n^R})\int_1^2\left(\frac{e^{-2cb_n^R}}{\sqrt{n}}\right)^{\lambda}n^{(\lambda-1)/2}\,d\lambda \sim e^{-2cb_n^R}\frac{\tau_n}{cb_n^R}\,.$$

Arguing as in the proof of Theorem 3.7, a (τ_n, b_n^R)-separation cutoff does not hold for any sequence $b_n^R = o(W(\tau_n))$ (see Connor, 2007 for further details).

3.2. *Links to extreme value theory.* Looking back to the discussion following Proposition 3.1, where the separation distance is identified as the tail distribution of the maximum of a set of independent random variables T_n^i, it is reasonable to ask how the above results relate to the theory of extreme values. If the random variables $\{T_n^i\}$ are i.i.d. for all i and n then the Fisher-Tippet-Gnedenko Theorem guarantees convergence in distribution of a renormalized T_n to one of three possible distributions. For example, if X_n is a random walk on $\mathbb{Z}_2^n$ for which the rate of each coordinate is chosen at random, with

$$\mathbb{P}(\rho_n^i = p_k) = q_k\,, \qquad k=1,\ldots,m,$$

for all i and n, Theorem 2.7.2 of Galambos (1978) shows that a renormalized T_n has a limiting Gumbel distribution. Indeed, writing $p^* = \min\{p_k\}$, direct calculation shows that

$$\mathrm{sep}_n\left(\frac{\log n + c}{2p^*}\right) = 1-\left(1-\sum_{j=1}^{m}q_k\left[\frac{e^{-c}}{n}\right]^{p_k/p^*}\right)^n \sim 1-\left(1-q^*\frac{e^{-c}}{n}\right)^n \xrightarrow[n\to\infty]{} 1-\exp(-q^*e^{-c})\,.$$

In this case we see that both right- and left-hand windows are $O(1)$.

More generally, the function θ_n defined in equation (3.6) may be interpreted as follows. Let $\{V_n^i : 1\le i\le n\}$ be independent, identically distributed random variables, whose distribution is a mixture over λ of *Exp*(λ) distributions, with mixture probability distribution μ_n. Then, for $t\ge 0$,

$$\mathbb{P}(V_n^i > t) = \int_0^\infty e^{-\lambda t}\mu_n(d\lambda)\,,$$

and so

$$\mathbb{E}\left[\sum_{i=1}^n \mathbf{1}_{[V_n^i>t]}\right] = n\int_0^\infty e^{-\lambda t}\mu_n(d\lambda) = \theta_n(t)\,.$$

Thus $\theta_n(t)$ describes the mean number of exceedances of level t by the set of random variables $\{V_n^i\}$. In particular, Proposition 3.5 implies that the set of n-tuples driven by μ_n exhibits a τ_n-separation cutoff if and only if

$$\mathbb{E}\left[\sum_{i=1}^n \mathbf{1}_{[V_n^i>c\tau_n]}\right] \xrightarrow[n\to\infty]{} \begin{cases}\infty & 0<c<1\\ 0 & c>1\,.\end{cases}$$

4. Coupling cutoffs

It is well known that the coupling method can be used to bound the rate of convergence to equilibrium of a Markov chain, via the coupling inequality (see Lindvall, 2002). Let X_n and Y_n be two copies of a Markov process on E_n with equilibrium distribution π_n.

Definition 4.1. A *coupling* of X_n and Y_n is a process $(\hat{X}_n, \hat{Y}_n)$ on $E_n \times E_n$ such that

$$\hat{X}_n \stackrel{\mathcal{D}}{=} X_n \quad \text{and} \quad \hat{Y}_n \stackrel{\mathcal{D}}{=} Y_n \,,$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution.

The *coupling time* T_n^c of $\hat{X}_n$ and $\hat{Y}_n$ is defined by

$$T_n^c = \inf \left\{ t \geq 0 \,:\, \hat{X}_n^t = \hat{Y}_n^t \right\} .$$

For a given coupling of X_n and Y_n, define

$$\bar{F}_n(t) = \mathbb{P}(T_n^c > t), \quad t \geq 0 \,, \tag{4.1}$$

to be the tail probability of T_n^c. Suppose now that $X_n^0 = x_n^0$ is fixed, and that $Y_n^0 \sim \pi_n^0$. We then define the following behaviour, in analogy with Definition 2.2:

Definition 4.2. For $n \geq 1$, let T_n^c and $\bar{F}_n$ be defined as above. We say that the sequence $\{E_n, X_n, \pi_n, T_n^c\}$ exhibits a (τ_n, b_n)-*coupling-cutoff* if $\tau_n, b_n > 0$ satisfy $b_n = o(\tau_n)$ and

$$\bar{F}_-(c) = \liminf_{n\to\infty} \bar{F}_n(\tau_n + cb_n) \quad \text{satisfies} \lim_{c\to-\infty} \bar{F}_-(c) = 1 \,, \tag{4.2}$$

$$\bar{F}_+(c) = \limsup_{n\to\infty} \bar{F}_n(\tau_n + cb_n) \quad \text{satisfies} \lim_{c\to\infty} \bar{F}_+(c) = 0 \,. \tag{4.3}$$

Thus a coupling cutoff occurs when the distance between the two processes, measured using the tail probability of the coupling time T_n^c, asymptotically exhibits an abrupt change from one to zero at time τ_n. (Note that if T_n^c is a maximal coupling time for all n (Griffeath, 1974/75) then a coupling-cutoff is equivalent to a total-variation cutoff.) As with the optimal SST of Section 3, if (X_n, Y_n) is a pair of n-tuples whose i^{th} coordinates may be independently coupled at an exponential rate λ_n^i, then T_n^c is the maximum of n coupling times and this yields an analogous version of Theorem 3.2 for coupling cutoffs.

For the random walks on $\mathbb{Z}_2^n$ considered in Section 3.1, no intuitive maximal coupling is known in general; for the *symmetric* random walk a (nearly) maximal solution is presented in Matthews (1987), and a stochastically optimal *co-adapted* coupling is described in Connor and Jacka (2008). However, X_n and Y_n may be simply coupled by allowing their i^{th} coordinates to evolve independently until the time that they first agree, whereafter they move synchronously. If X_n^0 and Y_n^0 do not agree on the i^{th} coordinate (which happens with probability $1/2$), then it follows that the time taken for agreement on this coordinate is equal to the time of the first incident on a Poisson process of rate $2p_n^i$, and so this coupling takes place at an exponential rate. Thus a random walk on $\mathbb{Z}_2^n$ exhibits a coupling cutoff if and only if it exhibits a separation cutoff (with the same values of τ_n and b_n).

In general, the assumption that each component of the n-tuples may be co-adaptedly coupled at an exponential rate is not restrictive: indeed, this is a reasonable assumption for many Markov processes of interest (Burdzy and Kendall, 2000). There is also a possibility that the coupling variant of Theorem 3.2 outlined above could have interesting consequences for coupling-based perfect simulation algorithms (such as CFTP and variants) for high-dimensional distributions.

Acknowledgements

Thanks to Persi Diaconis for making time to discuss this topic during a visit by the author to Université Nice Sophia Antipolis in June 2007, and to Pierre Del Moral for making this visit possible. Thanks also to Wilfrid Kendall for motivating and improving the work in this paper while the author was working at the University of Warwick, and to a helpful referee.

References

David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** (1), 69–97 (1987). ISSN 0196-8858. doi:10.1016/0196-8858(87)90006-6. MR876954.

Javiera Barrera, Béatrice Lachaud and Bernard Ycart. Cut-off for n-tuples of exponentially converging processes. *Stochastic Process. Appl.* **116** (10), 1433–1446 (2006). ISSN 0304-4149. doi:10.1016/j.spa.2006.03.003. MR2260742.

Krzysztof Burdzy and Wilfrid S. Kendall. Efficient Markovian couplings: examples and counterexamples. *Ann. Appl. Probab.* **10** (2), 362–409 (2000). ISSN 1050-5164. doi:10.1214/aoap/1019487348. MR1768241.

Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.* **13**, no. 3, 26–78 (2008). ISSN 1083-6489. MR2375599.

S.B. Connor. *Coupling: Cutoffs, CFTP and Tameness*. Ph.D. thesis, University of Warwick (2007).

Stephen Connor and Saul Jacka. Optimal co-adapted coupling for the symmetric random walk on the hypercube. *J. Appl. Probab.* **45** (3), 703–713 (2008). ISSN 0021-9002. doi:10.1239/jap/1222441824. MR2455179.

R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth. On the Lambert W function. *Adv. Comput. Math.* **5** (4), 329–359 (1996). ISSN 1019-7168. doi:10.1007/BF02124750. MR1414285.

Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* **16** (4), 2098–2122 (2006). ISSN 1050-5164. doi:10.1214/105051606000000501. MR2288715.

Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** (2), 159–179 (1981). ISSN 0044-3719. doi:10.1007/BF00535487. MR626813.

Janos Galambos. *The asymptotic theory of extreme order statistics*. John Wiley & Sons, New York-Chichester-Brisbane (1978). ISBN 0-471-02148-2. Wiley Series in Probability and Mathematical Statistics; MR489334.

David Griffeath. A maximal coupling for Markov chains. *Z. Wahrsch. und Verw. Gebiete* **31**, 95–106 (1974/75). MR0370771.

Torgny Lindvall. *Lectures on the coupling method*. Dover Publications Inc., Mineola, NY (2002). ISBN 0-486-42145-7. Corrected reprint of the 1992 original; MR1924231.

Peter Matthews. Mixing rates for a random walk on the cube. *SIAM J. Algebraic Discrete Methods* **8** (4), 746–752 (1987). ISSN 0196-5212. doi:10.1137/0608060. MR918073.

Laurent 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.

R. C. St. John and N. R. Draper. *D*-optimality for regression designs: a review. *Technometrics* **17**, 15–23 (1975). ISSN 0040-1706. MR0373188.