J Stat Phys (2011) 143: 413–419
DOI 10.1007/s10955-011-0194-y

Exact Solution for a Class of Random Walk on the Hypercube

Benedetto Scoppola

Received: 26 October 2010 / Accepted: 4 April 2011 / Published online: 16 April 2011

Abstract A class of families of Markov chains defined on the vertices of the n-dimensional hypercube, $\Omega_n = \{0, 1\}^n$, is studied. The single-step transition probabilities $P_{n,ij}$, with $i, j \in \Omega_n$, are given by $P_{n,ij} = \frac{(1-\alpha)^{d_{ij}}}{(2-\alpha)^n}$, where $\alpha \in (0, 1)$ and d_{ij} is the Hamming distance between i and j. This corresponds to flip independently each component of the vertex with probability $\frac{1-\alpha}{2-\alpha}$. The m-step transition matrix $P^m_{n,ij}$ is explicitly computed in a close form. The class is proved to exhibit cutoff. A model-independent result about the vanishing of the first m terms of the expansion in α of $P^m_{n,ij}$ is also proved.

Keywords Finite Markov chain · Random walk on the hypercube · Cutoff

1 Introduction

In recent times a considerable effort has been spent to study the phenomenon of *cutoff* for Markov chain, i.e the abrupt convergence behavior of the measure to the stationarity regime.

There are many ways to define the occurrence of the cutoff phenomenon. The general idea is the following: we define a family of ergodic Markov Chain $\{(X^t_n), \Omega_n, P_n, \mu^t_n, \pi_n\}$, where Ω_n is the state space of (X^t_n), P_n is its transition matrix, μ^t_n is the evolution at time t of the initial measure μ^0_n and π_n is the stationary measure. We say that the family exhibits cutoff if there exist a notion $d(\mu^t_n, \pi_n)$ of distance between μ^t_n and π_n, $d(\mu^t_n, \pi_n) \le 1$, and some initial measures μ^0_n such that, for $n \to \infty$, $d(\mu^t_n, \pi_n)$ is very close to 1 for $t < a_n - o(a_n)$, while $d(\mu^t_n, \pi_n)$ is close to zero for $t > a_n + o(a_n)$, where a_n is a family of deterministic times.

One can be interested in a certain degree of accuracy in the description of the detail of this abrupt behavior, and this gives various possible definitions of the cutoff. Here we use as

B. Scoppola (✉)
Dipartimento di Matematica, Università di Roma "Tor Vergata", Roma, Italy
e-mail: scoppola@mat.uniroma2.it

distance $d(\mu_n^t, \pi_n)$ the total variation distance

$$d_{TV}(\mu_n^t, \pi_n) = \frac{1}{2} \sum_{x \in \Omega_n} |\mu_n^t(x) - \pi_n(x)| \tag{1.1}$$

and we define the cutoff phenomenon by means of the Diaconis' paradigm (see [4]): given a_n, b_n two sequences such that

$$\lim_{n\to\infty} \frac{b_n}{a_n} = 0 \tag{1.2}$$

a family of Markov chains $\{(X_n^t), \Omega_n, P_n, \mu_n^t, \pi_n\}$ is said to exhibit cutoff, or more precisely an (a_n, b_n)-cutoff with respect to the total variation distance, if for some initial measure μ_n^0

$$\begin{aligned} \lim_{\theta\to\infty} \limsup_{n\to\infty} d_{TV}(\mu_n^{a_n+\theta b_n}, \pi_n) = 0 \\ \lim_{\theta\to\infty} \liminf_{n\to\infty} d_{TV}(\mu_n^{a_n-\theta b_n}, \pi_n) = 1 \end{aligned} \tag{1.3}$$

This notion of (a_n, b_n)-cutoff is quite detailed, in the sense that it states also the width, negligible in terms of a_n but possibly diverging in n, of the time window in which the cutoff takes place.

The existence of this or other forms of cutoff has been proved for a variety of different models. The phenomenon has been discovered in the early 1980, when Diaconis and Shahshahani (see [5]) proved the cutoff in a problem of card shuffling, namely the random transposition model. Then a considerable effort has been spent in proving the cutoff for random walks on groups (see e.g. [1, 3]). In all these papers the relations between the structure of the group and the spectral properties of the Markov chain are the key ingredient of the proof of the cutoff behavior. It is particularly relevant in the context of this paper the case of the lazy random walk on the n-dimensional hypercube. This has been studied in [6], and alternative proofs of the cutoff are available (see for instance [9] and references therein). The system is defined as follows: the space state is the set of vertices of an n dimensional hypercube. At each step the state remains the same with probability $1/2$, while with probability $1/2$ a component of the vertex of the hypercube is chosen with uniform probability, and it is flipped. The system exhibits (a_n, b_n)-cutoff with respect to the total variation distance with $a_n = (1/2)n \log n$ and $b_n = n$. Another important system to be quoted in the context of this paper is a class of random walks on the hypercube in which some particular subsets of components (chosen randomly) may be flipped together (see [14]). In all the cases mentioned above the stationary measure is uniform. More recently the cutoff behavior has been proved for systems with a non uniform measure having anyway a large support (for the important case of the Ising model, see [10–12]). It exists a second class of systems for which it is possible to prove the cutoff phenomenon: they are systems having the stationary measure concentrated in some small region of the state space; typically they are birth and death processes with some kind of drift towards a small region (see e.g. [7, 9] and references therein); in the latter case the cutoff is in general easier to establish, and it has been proven to be related to certain properties of the hitting times of the chains (see [2, 13]). Even if some common features between the two classes of systems mentioned above is proposed in [8], the proof of the cutoff behavior for systems with uniform stationary measure needs in general a quite detailed knowledge of the evolution of the measure.

In this paper a class of families of discrete time Markov chains with uniform stationary measure is proved to exhibit cutoff by an easy direct computation of the evolution of the transition matrix. For particular values of the parameters appearing in the expression of the transition probabilities the behavior of the system turns out to be very similar to the "classical" lazy random walk on the hypercube.

2 The Model

The family of Markov chains studied here, indexed by n, is defined as follows: let the state space Ω_n of the discrete time Markov chain (X_n^t) be the set of the binary nth-ple, or, in other words, the set of the vertices of the n dimensional unit hypercube, $\Omega_n = \{0,1\}^n$. The transition probability matrix of the Markov chain (X_n^t) is given by

$$P_{n,ij} = \frac{(1-\alpha)^{d_{ij}}}{(2-\alpha)^n} \tag{2.1}$$

where $\alpha \in (0,1)$ and d_{ij} is the Hamming distance between i and j, i.e. the number of different bits between i and j. Note that the parameter α may in general depend on n.

The system hence has the possibility to evolve in a single step from any vertex to any vertex, but the transition probability becomes exponentially small with the distance. This reminds the mean-field approximation via Kac potentials that is used in statistical mechanics, where here the pure mean field ($\alpha = 0$) corresponds to a sequence of independent trials. As in the case of Kac potential, the case $\alpha \to 1$ tends to be a short range (nearest neighbors random walk) system.

Note first that the transition probabilities defined in (2.1) give after a single step a distribution of d_{ij} which is a binomial of parameter $\frac{1-\alpha}{2-\alpha}$, since the number of j at a distance d from i is evidently $\binom{n}{d}$, and hence

$$P_n(d_{ij} = d) = \binom{n}{d}\left(\frac{1-\alpha}{2-\alpha}\right)^d \left(\frac{1}{2-\alpha}\right)^{n-d} \tag{2.2}$$

This means that the chain may be interpreted as follows: at each step each component of the nth-ple has an independent probability $\frac{1-\alpha}{2-\alpha}$ to flip its value. Another alternative interpretation of this model is the following: define a continuous time Markov chain in which each component flips independently, and look at it only for integer times.

The main results of this paper are listed in the following theorem.

Theorem

(i) *The transition probabilities after a time m of the family of Markov chains defined by* (2.1) *are*

$$P_{n,ij}^m = \left(\frac{1}{2}\right)^n \left(1 + \left(\frac{\alpha}{2-\alpha}\right)^m\right)^{n-d_{ij}} \left(1 - \left(\frac{\alpha}{2-\alpha}\right)^m\right)^{d_{ij}} \tag{2.3}$$

(ii) *The family of Markov chains (X_n^t) exhibits (a_n, b_n)-cutoff with respect to the total variation distance at time $a_n = \frac{1}{2\log\frac{2-\alpha}{\alpha}}\log n$ with a window of the order $b_n = \frac{1}{\log\frac{2-\alpha}{\alpha}}$.*

Proof of (i) Consider the matrix element $P_{n,ij}^m$. At each step each component is flipped independently with probability $\frac{1-\alpha}{2-\alpha}$, while the probability at each times that the component remains unchanged is $\frac{1}{2-\alpha}$. After a time m each component is equal to its original value if it is flipped an even number of times, while it is different if the number of independent flips is odd.

Hence we can write

$$P^m_{n,ij} = (\mathcal{E}(m,\alpha))^{n-d_{ij}} (\mathcal{O}(m,\alpha))^{d_{ij}} \tag{2.4}$$

where

$$\mathcal{E}(m,\alpha) = \sum_{l \text{ even}} \binom{m}{l} \frac{(1-\alpha)^l}{(2-\alpha)^m} \tag{2.5}$$

and

$$\mathcal{O}(m,\alpha) = \sum_{l \text{ odd}} \binom{m}{l} \frac{(1-\alpha)^l}{(2-\alpha)^m} \tag{2.6}$$

Now it is obvious that

$$\sum_l \binom{m}{l} \frac{(1-\alpha)^l}{(2-\alpha)^m} = 1 = \mathcal{E}(m,\alpha) + \mathcal{O}(m,\alpha) \tag{2.7}$$

while

$$\sum_l \binom{m}{l} \frac{[-(1-\alpha)]^l}{(2-\alpha)^m} = \left(\frac{\alpha}{2-\alpha}\right)^m = \mathcal{E}(m,\alpha) - \mathcal{O}(m,\alpha) \tag{2.8}$$

Hence, solving (2.7) and (2.8) we have

$$\mathcal{E}(m,\alpha) = \frac{1}{2}\left(1 + \left(\frac{\alpha}{2-\alpha}\right)^m\right) \tag{2.9}$$

and

$$\mathcal{O}(m,\alpha) = \frac{1}{2}\left(1 - \left(\frac{\alpha}{2-\alpha}\right)^m\right) \tag{2.10}$$

Putting (2.9) and (2.10) in (2.4) we have the proof of (i). Note that the parameter α may be n-dependent without affecting the proof. □

Proof of (ii) Note first of all that the stationary measure of this families of random walk is uniform, because the transition matrix (2.1) is doubly Markov. Let us assume then that the initial measure μ^0_n is concentrated in a single vertex, say the origin. Hence $\mu^0_n(0) = 1$. Let us compute the total variation distance between the uniform measure π_n and the measure $\mu^m_n = \mu^0_n P^m_n$. To compute the total variation distance, since all the vertices at distance d from the origin 0 have the same probability, it is natural to consider the probability distribution of d:

$$\begin{aligned} d_{TV}(\pi_n, \mu^m_n) &= d_{TV}(\pi_n, \mu^0_n P^m_n) \\ &= \sum_{d=0}^{n} \left| \binom{n}{d} \left(\frac{1}{2}\right)^n \left[1 - \left(1 + \left(\frac{\alpha}{2-\alpha}\right)^m\right)^{n-d} \left(1 - \left(\frac{\alpha}{2-\alpha}\right)^m\right)^d\right] \right| \end{aligned} \tag{2.11}$$

We are therefore considering the total variation distance between two binomial distributions. The stationary distribution has average value $\frac{n}{2}$ and variance equal to $\frac{n}{4}$, while the distribution μ_n^m has average value $\frac{n}{2}(1-(\frac{\alpha}{2-\alpha})^m)$ and variance equal to $\frac{n}{4}(1-(\frac{\alpha}{2-\alpha})^{2m})$. The difference $\Delta(m)$ between the average values of the two distributions is hence

$$\Delta(\alpha, m) = \frac{n}{2}\left(\frac{\alpha}{2-\alpha}\right)^m \tag{2.12}$$

Since the width scale of both distributions is $\sqrt{n}$, we have that if

$$\Delta(\alpha, m) \geq \sqrt{n}e^{\theta} \tag{2.13}$$

then the total variation distance between the two binomials tends to one for θ tending to infinity. This corresponds to consider m such that

$$m \log \frac{\alpha}{2-\alpha} \geq \log 2 - \frac{1}{2}\log n + \theta \tag{2.14}$$

and hence

$$m \leq \frac{1}{2\log\frac{2-\alpha}{\alpha}}\log n - \frac{1}{\log\frac{2-\alpha}{\alpha}}\theta \tag{2.15}$$

On the other side, for m such that

$$\Delta(\alpha, m) \leq \sqrt{n}e^{-\theta} \tag{2.16}$$

and hence

$$m \geq \frac{1}{2\log\frac{2-\alpha}{\alpha}}\log n + \frac{1}{\log\frac{2-\alpha}{\alpha}}\theta \tag{2.17}$$

we have that the difference between the variances of the two distribution is proportional to $e^{-\theta}$ and then the total variation distance between the two binomial tends to zero for θ tending to infinity. This corresponds to say that, taking as $a_n = \frac{1}{2\log\frac{2-\alpha}{\alpha}}\log n$ and $b_n = \frac{1}{\log\frac{2-\alpha}{\alpha}}$ our family of Markov chains (X_n^t) exhibit an (a_n, b_n)-cutoff. □

Remark As soon as the parameter α is a fixed quantity, independent from n, the system tends to flip a finite fraction of sites at each step, according to a binomial distribution, and it has cutoff at time $\frac{1}{2\log\frac{2-\alpha}{\alpha}}\log n$ with constant window.

Since the expression of $P_{n,ij}^m$ is exact, we can also use the results above for n-dependent α. In particular for $1-\alpha = \frac{\beta}{n}$, being $\beta > 0$ a fixed parameter, the system tends to flip at each step a number of sites according to a Poisson distribution of parameter $\frac{\beta}{1+\beta/n}$. Since for large n we have $\log\frac{2-\alpha}{\alpha} \approx \frac{2\beta}{n}$, the system exhibits cutoff at time $\frac{n\log n}{4\beta}$ with window proportional to n. Hence in this case the cutoff is on the same time scale than the cutoff for the nearest neighbors random walk on the hypercube (see discussion in the section above), as one expects.

For $1-\alpha$ tending to zero with n with a rate slower that $1/n$ the cutoff is at an intermediate time between $\log n$ and $n\log n$.

3 The Expansion in α of $P^m_{n,ij}$

It is easy to see, by the explicit expression (2.3) of $P^m_{n,ij}$, that its series expansion around $\alpha = 0$ vanishes up to the $(m-1)$-th term. This result is model-independent, in the sense that it can be proved in a much more general setup.

Theorem *Consider an ergodic Markov chain (X^t) defined on the state space Ω, with $|\Omega| = k$. Let the transition matrix P be of the form*

$$P_{ij} = \frac{1}{k} + O(\alpha) \tag{3.1}$$

and let the stationary measure π be of the form

$$\pi_j = \frac{1}{k} + O(\alpha) \tag{3.2}$$

Then

$$P^m_{ij} - \pi_j = O(\alpha^m) \tag{3.3}$$

Proof by induction (3.3) is evidently true for $m = 1$. Hence we have to prove that assuming $P^m_{ij} - \pi_j = O(\alpha^m)$ we have that $P^{m+1}_{ij} - \pi_j = O(\alpha^{m+1})$. We write

$$\begin{aligned} P^{m+1}_{ij} - \pi_j &= \sum_l (P^m_{il} - \pi_l) P_{lj} \\ &= \sum_l (P^m_{il} - \pi_l)\left(\frac{1}{k} + O(\alpha)\right) = \sum_l (P^m_{il} - \pi_l) O(\alpha) \\ &= O(\alpha^m) O(\alpha) = O(\alpha^{m+1}) \end{aligned} \tag{3.4}$$

Note that the relation above does not implies necessarily the presence of cutoff. For instance, define a random walk on the discrete n-circle, define $d(i,j) = \min\{|i-j|, n-1-|i-j|\}$ and let the transition matrix be $P_{ij} = \frac{e^{-\alpha d(i,j)}}{Z}$ with $Z = \sum_j e^{-\alpha d(i,j)}$. The system clearly exhibits a simple diffusion on the circle, and no cutoff is present. However the conditions of the theorem above are satisfied. □

Acknowledgements Many thanks to Fabio Martinelli, Eyal Lubetzky, and two anonymous referees.

References

1. Aldous, D., Diaconis, P.: Shuffling cards and stopping times. Am. Math. Mon. **93**, 333–348 (1986)
2. Barrera, J., Bertoncini, O., Fernández, R.: Abrupt convergence and escape behavior for birth and death chains. J. Stat. Phys. **137**, 595–623 (2009)
3. Bayer, D., Diaconis, P.: Trailing the dovetail shuffle to its lair. Ann. Appl. Probab. **2**, 294–313 (1992)
4. Diaconis, P.: The cutoff phenomenon in finite Markov chains. Proc. Natl. Acad. Sci. USA **93**, 1659–1664 (1996)
5. Diaconis, P., Shahshahani, M.: Generating a random permutation with random transpositions. Z. Wahrscheinlichkeitstheor. Verw. Geb. **57**(2), 159–179 (1981)
6. Diaconis, P., Graham, R.L., Morrison, J.A.: Asymptotic analysis of a random walk on a hypercube with many dimensions. Random Struct. Algorithms **1**, 51–72 (1990)

7. Ding, J., Lubetzky, E., Peres, Y.: Total-variation cutoff in birth-and-death chains. Arxiv preprint. arXiv:0801.2625 (2008)
8. Lancia, C., Scoppola, B.: Entropy driven cutoff phenomenon, in preparation
9. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. AMS, Providence (2009)
10. Levin, D.A., Luczak, M.J., Peres, Y.: Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. Probab. Theory Relat. Fields **146**, 223–265 (2010)
11. Lubetzky, E., Sly, A.: Cutoff for the Ising model on the lattice. Arxiv preprint. arXiv:0909.4320 (2009)
12. Lubetzky, E., Sly, A.: Critical Ising on the square lattice mixes in polynomial time. Arxiv preprint. arXiv:1001.1613 (2010)
13. Martinez, S., Ycart, B.: Decaying rates and cutoff for convergence and hitting times of Markov chains. Adv. Appl. Probab. **33**, 188–205 (2001)
14. Wilson, D.B.: Random walks on $\mathbf{Z}_2^d$. Probab. Theory Relat. Fields **108**, 441–457 (1997)