Electronic Journal of Statistics
Vol. 16 (2022) 4379–4397
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2043

Monte Carlo Markov chains constrained on graphs for a target with disconnected support*

Roy Cerqueti[1,2,3]

[1] *Sapienza University of Rome, Department of Social and Economic Sciences. Piazzale Aldo Moro, 5, I-00185, Rome, Italy.*

[2] *London South Bank University, School of Business. 103 Borough Road, London, SE1 0AA, UK.*

[3] *Université d'Angers, GRANEM. SFR Confluences, F-49000 Angers, France.*
e-mail: roy.cerqueti@uniroma1.it

and

Emilio De Santis[4]

[4] *University of Rome La Sapienza, Department of Mathematics. Piazzale Aldo Moro, 5, I-00185, Rome, Italy.*
e-mail: desantis@mat.uniroma1.it

Abstract: This paper presents a theoretical Monte Carlo Markov chain procedure in the framework of graphs. It specifically deals with the construction of a Markov chain whose empirical distribution converges to a given reference one. The Markov chain is constrained over an underlying graph so that states are viewed as vertices, and the transition between two states can have positive probability only in the presence of an edge connecting them. The analysis focuses on the relevant case of support of the target distribution not connected in the graph. Some general arguments on the speed of convergence are also carried out.

MSC2020 subject classifications: Primary 60J10, 62E25; secondary 60B10.
Keywords and phrases: Markov chain Monte Carlo, graphs, convergence of probability distributions.

Received July 2021.

1. Introduction

Monte Carlo Markov Chain (MCMC) problems represent a challenging research theme not only for their natural practical implications but also for the related methodological advancements.

*E. D.S. was partially supported by Simmetrie e Disuguaglianze in Modelli Stocastici RM118164035FE854 and by Dipendenza tra Variabili Aleatorie e nei Processi Stocastici RM120172B73EEB91

The idea of a MCMC problem is to build a Markov chain with a target stationary distribution (see e.g. [5, 14, 30]). To pursue this scope, several algorithms have been proposed in the literature. Some of them are worthy of being mentioned.

In the Metropolis Hastings algorithm (see [20, 25]), a transition kernel is employed to iteratively generate a value y at time $t+1$ on the basis of the value x observed at time t.

When the states space is huge, the Metropolis Hastings algorithm must be used with great care to avoid that the probabilities of transition become too small and in practice unusable for computer simulation purposes.

The Gibbs sampler, see [18], solves the problem of the huge cardinality in the presence of a multivariate structure for the states space. The strategy is to change state by changing only one of the components of the multivariate state. In so doing, there are few transition probabilities that are different from zero; therefore, they remain not too small in order to be used on a computer. The Gibbs sampler loses meaningfulness when the multivariate structure of the state space is not identified.

The debate on the validity of the Gibbs sampler has been remarkably enriched by [19]. In the quoted paper, the Author elaborates on [30] and deals with a Bayesian choice of a vector of models, whose individual components are selected among a set of countable candidates. Each model has several unknown parameters; such a number depends on the specific model. In this context of not fixed dimension of the parameter set, [19] adapts the Metropolis-Hastings algorithm, by proposing a so-called "reversible jump" version of it (see also [2] for further advancements). In [6], the Authors observe that the convergence issues of the MCMC procedures always arise when the problem involves the selection of one among many different model specifications. [6] proposes a modified Gibbs sampler procedure obtained by introducing a sort of average of the considered models, to solve the convergence matter. In general, the issue of the convergence is a critical aspect, as also acknowledged by Persi Diaconis in his long experience of scientific research and publications in the field. In this respect, we strongly recommend the reading of Diaconis' personal view on the matter, with some relevant insights of the future development of the MCMC in both areas of mathematical advancements and practical applications (see [11, 12]).

Our paper adds to this debate by dealing with a MCMC problem on a connected graph where the target distribution has support not connected in the graph. More specifically, we construct some Markov chains whose empirical distributions converge to such a target one as time goes to infinity. In so doing, a discussion on the speed of *convergence* is also carried out.

Some notation is needed, to present the problem. We will refer hereafter to a *connected graph* $G = (\mathcal{S}, E)$, being $\mathcal{S}$ the set collecting the nodes and E the set of the edges. The nodes $s, t \in \mathcal{S}$ are declared *adjacent* in G if $\{s, t\} \in E$ or $s = t$. The degree of a node $s \in \mathcal{S}$ is $\deg(s)$. We denote the maximum degree of G by $\Delta(G) = \max_{s\in\mathcal{S}} \deg(s)$. The distance between two nodes $s, t \in \mathcal{S}$, say $d(s, t)$, is the number of the edges of the shortest path connecting s and t. The

diameter of G is

$$d = d(G) = \max_{s,t \in \mathcal{S}} d(s,t).$$

We now state a definition linking graphs and stochastic processes.

Definition 1. We say that a stochastic process $X = (X(t) : t \in \mathbb{N})$ on the state space $\mathcal{S}$ is *consistent* with the graph $G = (\mathcal{S}, E)$ if, for each $t \in \mathbb{N}$, $X(t)$ and $X(t+1)$ are adjacent in G with probability one.

Given two graphs $G = (\mathcal{S}, E)$ and $G' = (\mathcal{S}', E')$ we say that G' is a *subgraph* of G if $\mathcal{S}' \subset \mathcal{S}$ and $E' \subset E$, and we write $G' \subset G$.

A particular class of subgraphs will be of interest in the following. Specifically, the subgraph $G' = (\mathcal{S}', E') \subset G = (\mathcal{S}, E)$ is said to be an *induced subgraph* of G if $s, t \in \mathcal{S}'$ and $\{s,t\} \in E$ imply $\{s,t\} \in E'$. In this case we write $G' = G[\mathcal{S}']$ in order to stress the dependence on the set of nodes $\mathcal{S}'$.

We notice that Definition 1 implies that if a process $X = (X(t) : t \in \mathbb{N})$ is consistent with a graph G' then it is also consistent with any graph G such that $G' \subset G$.

From now on we only consider $|\mathcal{S}| < \infty$ and, consequently, a finite graph $G = (\mathcal{S}, E)$. Given a graph $G = (\mathcal{S}, E)$ and a distribution $\mu = (\mu(s) : s \in \mathcal{S})$, we will provide in this paper an answer to the following question:

Q: *Is it possible to construct a (not necessarily homogeneous) Markov chain $X = (X(t) : t \in \mathbb{N})$ which is consistent with G and such that its empirical distribution converges almost surely to μ as t goes to infinity?*

More precisely we aim at constructing a reversible Markov chain $X = (X(t) : t \in \mathbb{N})$ with the following properties: X is consistent with the graph G and

$$\lim_{t \to \infty} \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{1}_{\{X(m)=s\}} = \mu(s), \qquad s \in \mathcal{S} \qquad a.s.. \tag{1}$$

The motivations to pose the question **Q** are basically three:

a) we face the problem of the large cardinality of the states space by controlling the transitions among the states through the edges of a graph;
b) we introduce a clear structure of the states space through the graph so that one can think to get some desired properties such as stochastic monotonicity or fast convergence;
c) the introduction of a graph which constrains the positive transitions of the Markov chain describes several real-life evolution phenomena, where it is possible to move in a single step only from a state to an "adjacent one".

In the following, we provide an answer to question **Q** by showing that it is possible to construct such a (not necessarily time-homogeneous) Markov chain.

The adopted strategy for responding to **Q** is of constructive type. We start from the introduction of a sequence of probability distributions $(\mu_k : k \in \mathbb{N})$ converging to the target μ and such that $supp\,(\mu_k) = \mathcal{S}$. Accordingly, we

construct a suitable sequence of irreducible and aperiodic transition matrices $(P^{(\mu_k,G)} : k \in \mathbb{N})$ such that $(\mu_k, P^{(\mu_k,G)})$ is reversible. Then, we derive a new non-homogeneous Markov chain consistent with G as a concatenation of this sequence of homogeneous ones. The transition matrix $P^{(\mu_k,G)}$ insists over a time interval whose length increases with k. The proper selection of the time intervals – whose length is estimated on the basis of the diameter of G and is independent from the target distribution μ – is shown to lead to the response to question **Q** for the resulting non-homogeneous Markov chain. In this respect, see also Example 2.

In this construction, a not unexpected problem of dependence arises. We control for the dependence in two steps. First, we lump together the events related to the Markov chain at large time-distance; second, the convergence properties of the Markov chain over all the times are obtained through a union bound procedure. At this aim, a coupling of the Markov chain with a sequence of independent random variables is performed (see Theorem 1). Such a method allows an estimation of the speed of convergence to the target probability distribution μ; moreover, it has the relevant advantage of being general and reproducible in other contexts.

The paper also presents an extension of the above-discussed framework for the products of graphs and product measures. The usefulness of such an extension lies in its ability to reduce the cardinality of the states space. Indeed, it moves from a multidimensional Markov chain to several unidimensional ones (see Theorem 2). In this context, we present a discussion on the possible applicability of such a result in the field of game theory (see Remark 3).

The structure of the paper is the following. In the next section we present the main results of our study. Section 3 provides some supporting remarks and examples, by discussing advantages, limitations and applicability of the proposed procedure. Section 4 offers the extension of our findings to the case of the product of graphs and distributions.

2. Main results

For a target probability distribution $(\mu(s) : s \in \mathcal{S})$ and a connected graph G, we give an answer to question **Q** in the interesting case when $G[supp\,(\mu)]$ is not connected.

As we will see, the solution strategy for solving the problem moves from creating a sequence of distributions that converges to the target one and having the states space $\mathcal{S}$ as support.

Without loss of generality, we start from a nonnegative function $\eta : \mathcal{S} \to [0, +\infty)$, which induces a distribution $\mu = (\mu(s) : s \in \mathcal{S})$ such that

$$\mu(s) = \frac{\eta(s)}{Z}, \qquad s \in \mathcal{S}, \tag{2}$$

where $Z = \sum_{s' \in \mathcal{S}} \eta(s')$ is the normalizing constant. We take $M > 0$ such that $\max_{s \in \mathcal{S}} \eta(s) \leq M$.

The distribution μ in (2) will be the target probability one. The derivation of μ from a given η is a standard procedure in statistical mechanics. In this context, the computation of the normalizing constant Z can be computationally complex, mainly when the cardinality of $\mathcal{S}$ is huge. In the following, we will implement a strategy for answering to $\mathbf{Q}$ that will manage this problem by avoiding such a computation.

Let N denote the cardinality of $\mathcal{S}$. Since $G[supp\,(\mu)]$ is not connected, then it contains at least two points. Since $supp\,(\mu) \subset \mathcal{S}$ and G is connected, then $N \geq 3$.

We now consider $k \in \mathbb{N}$ and define

$$\eta_k(s) = \begin{cases} \eta(s), & \text{if } \eta(s) > 0; \\ \frac{1}{k}, & \text{if } \eta(s) = 0. \end{cases} \tag{3}$$

As in (2), the η_k in (3) induces the probability distribution $\mu_k = (\mu_k(s) : s \in \mathcal{S})$, with the normalizing constant $Z_k = \sum_{s' \in \mathcal{S}} \eta_k(s')$.

Denote by $||\cdot||_{TV}$ the total variation norm (see e.g. [24]). Since $Z \leq Z_k$ and $Z_k - Z \leq \frac{N}{k}$, one has

$$||\mu_k - \mu||_{TV} = \frac{1}{2}\sum_{s\in\mathcal{S}} |\mu_k(s) - \mu(s)| = \frac{1}{2}\sum_{s\in\mathcal{S}} \left|\frac{\eta_k(s)}{Z_k} - \frac{\eta(s)}{Z}\right| \leq$$

$$\leq \sum_{s\in\mathcal{S}} \left[\frac{\eta(s)(Z_k - Z)}{2Z_k Z} + \frac{1}{2kZ_k}\right] \leq \frac{N}{2kZ_k} + \frac{N}{2kZ_k} \leq \frac{N}{kZ} = \frac{\Gamma}{k}, \tag{4}$$

where $\Gamma = \frac{N}{Z}$.

By construction, for any

$$k > \bar{k} = \left\lceil \frac{1}{\min\{\eta(s) > 0 : s \in \mathcal{S}\}} \right\rceil,$$

we have $Z_k \leq MN$ so that

$$\mu_k(s) \geq \frac{1}{kMN}, \qquad s \in \mathcal{S}. \tag{5}$$

Let us label the elements of $\mathcal{S} = \{s^1, \ldots, s^N\}$ such that

$$\mu(s^1) \geq \mu(s^2) \geq \cdots \geq \mu(s^N) = 0.$$

According to the definition of μ_k, for $k > \bar{k}$, one also obtains

$$\mu_k(s^1) \geq \mu_k(s^2) \geq \cdots \geq \mu_k(s^N) > 0. \tag{6}$$

We construct the transition matrix $P^{(\mu_k,G)} = (p_{l,m} : l, m = 1, \ldots, N)$ related to the distribution μ_k and to the graph $G = (\mathcal{S}, E)$. The dependence on k of the elements of matrix $P^{(\mu_k,G)}$ is conveniently omitted. For each $l, m = 1, \ldots, N$,

$$p_{l,m} = \begin{cases} p, & \text{if } l < m \text{ and } \{s^l, s^m\} \in E; \\ \frac{\mu_k(s^m)}{\mu_k(s^l)} p, & \text{if } l > m \text{ and } \{s^l, s^m\} \in E; \\ p_l, & \text{if } l = m; \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where

$$p_l = 1 - p\left[\sum_{m':m'>l} \mathbf{1}_{\{\{s^l,s^{m'}\}\in E\}} + \sum_{m':m'<l} \frac{\mu_k(s^{m'})}{\mu_k(s^l)}\mathbf{1}_{\{\{s^l,s^{m'}\}\in E\}}\right] \tag{8}$$

and

$$p \leq \tilde{p} = \min_{l=1,\ldots,N} \frac{1}{2\left(\sum_{m':m'>l} \mathbf{1}_{\{\{s^l,s^{m'}\}\in E\}} + \sum_{m':m'<l} \frac{\mu_k(s^{m'})}{\mu_k(s^l)}\mathbf{1}_{\{\{s^l,s^{m'}\}\in E\}}\right)}. \tag{9}$$

Clearly, $P^{(\mu_k,G)}$ is a transition or stochastic matrix. As already announced above, we notice that the construction of the transition matrix $P^{(\mu_k,G)}$ does not require the knowledge of the normalizing constant Z_k, in that $\frac{\mu_k(s^{m'})}{\mu_k(s^l)} = \frac{\eta_k(s^{m'})}{\eta_k(s^l)}$. This is particularly relevant for the computational applications of the proposed algorithm, in accord to the appropriateness of the Gibbs sampler for computations (see [18]). Moreover, by definition $p \leq \tilde{p} \leq \frac{1}{2}$. In fact, since G is connected, there exists at least an edge $\{s^1, s^m\} \in E$, with $m > 1$; thus the denominator of (9) is at least equal to 2, when $l = 1$.

Definition in (7) assures that the couple $(\mu_k, P^{(\mu_k,G)})$ is reversible. Moreover, $P^{(\mu_k,G)}$ is irreducible, since G is connected; thus, μ_k is the unique invariant distribution of $P^{(\mu_k,G)}$. The transition matrix $P^{(\mu_k,G)}$ is also aperiodic since, by (8) and (9), $p_l \geq \frac{1}{2}$ for $l = 1, \ldots, N$.

Condition (9) is of particular usefulness. Indeed, the estimation of $\tilde{p}$ from below can be obtained more easily than computing its exact value, hence being more appropriate for applications. In this respect, we provide an example where the computation of $\tilde{p}$ is hardly achieved while its estimation from below is simple.

Example 1. *To make the arguments simple, we take a given μ in place of a sequence $(\mu_k : k \in \mathbb{N})$. Let us consider an Ising model on a finite square sublattice Λ of $\mathbb{Z}^2$ with side $L \in \mathbb{N}$, e.g. $\Lambda = \{0, \ldots, L-1\}^2$. Thus, a single state is a vector in $\mathcal{S} = \{-1, +1\}^{L^2}$, and the cardinality of $\mathcal{S}$ is $N = 2^{L^2}$. The probability distribution μ is such that*

$$\mu(s) \propto \eta(s) = e^{-\mathcal{H}(s)}, \quad s \in \mathcal{S}$$

where

$$\mathcal{H}(s) = -\sum_{i,j\in\Lambda:||i-j||_1=1} J_{ij} s_i s_j$$

is the Hamiltonian. The interactions $(J_{ij} : i, j \in \Lambda : ||i - j||_1 = 1)$ can be taken either as assigned in an interval or randomly sampled from a suitable distribution. We here take any single interaction as assigned in $[\frac{1}{16}, \frac{2}{16}]$.

We form the graph $G = (\mathcal{S}, E)$ by imposing

$$\{s, s'\} \in E \iff ||s - s'||_1 = 2.$$

Substantially, $\{s, s'\} \in E$ is equivalent to say that s, s' differ only on one component.

In this case, the computation of $\tilde{p}$ in (9) is rather complex. However, a lower bound for $\tilde{p}$ is almost simple to obtain. Indeed,

$$\min_{l=1,\dots,N} \frac{1}{2\left(\sum_{m':m'>l} \mathbf{1}_{\{\{s^l,s^{m'}\}\in E\}} + \sum_{m':m'<l} \frac{\mu(s^{m'})}{\mu(s^l)} \mathbf{1}_{\{\{s^l,s^{m'}\}\in E\}}\right)} \geq$$

$$\geq \frac{1}{2L^2 \max\left\{\frac{\mu(s)}{\mu(s')} : \{s,s'\}\in E\right\}} \geq \frac{e^{-2}}{2L^2}.$$

We introduce the ergodic coefficient of Dobrushin (see [13] and [4] p. 235), which is defined as

$$\delta(P) = 1 - \inf_{i,j=1,\dots,N} \sum_{h=1}^{N} p_{i,h} \wedge p_{j,h} \tag{10}$$

where $P = (p_{i,j} : i,j = 1,\dots,N)$ is a stochastic matrix.

Lemma 1. *Given the transition matrix $P^{(\mu_k,G)}$ on $\mathcal{S}$ constructed above, with $N = |\mathcal{S}| \geq 3$ and $p \in \left[\frac{c}{k}, \tilde{p}\right]$, the Dobrushin's ergodic coefficient can be bounded from above as follows*

$$\delta((P^{(\mu_k,G)})^d) \leq 1 - \left(\frac{c}{k}\right)^d,$$

for any $k > \bar{k}$, where $c = \frac{1}{2\Delta(G)MN}$.

Proof. For $k > \bar{k}$, inequalities (5) and (6) provide

$$1 \leq \frac{\mu_k(s^m)}{\mu_k(s^l)} \leq \frac{1}{\mu_k(s^l)} \leq kMN, \qquad \text{for } l > m.$$

We now observe that, for $k > \bar{k}$,

$$\tilde{p} \geq \min_{l=1,\dots,N} \frac{1}{2kMN \deg(s^l)} = \frac{1}{2kMN\Delta(G)} = \frac{c}{k}.$$

Therefore, one can select p into the nonempty interval $\left[\frac{c}{k}, \tilde{p}\right]$. Then, for $p \in \left[\frac{c}{k}, \tilde{p}\right]$ and $\{s^l, s^m\} \in E$ one has,

$$p_{l,m} \geq \frac{c}{k}. \tag{11}$$

For $k > \bar{k}$, since the graph G is connected and $p_l \geq \frac{1}{2}$ for each $l = 1,\dots,N$ and by definition of diameter d, then (11) gives that

$$p_{l,m}^{(d)} \geq \left(\frac{c}{k}\right)^d, \qquad \forall\, l,m = 1,\dots,N,$$

where $p_{l,m}^{(d)}$ is the transition probability from s^l to s^m in d steps.

Then, by definition of the ergodic coefficient of Dobrushin in (10), one has the thesis. □

Given an arbitrary distribution over $\mathcal{S}$, namely $\lambda = (\lambda(s) : s \in \mathcal{S})$, we construct a non-homogeneous Markov chain $X = (X(t) : t \in \mathbb{N})$ with λ as initial distribution. The transition matrix of the Markov chain X at time $t \in \mathbb{N}$ will be denoted by $P(t) = (p_{i,j}(t) : i, j = 1, \dots, N)$.

Let us consider an increasing sequence of times $(t_\ell : \ell \in \mathbb{N})$, and let us define

$$P(t) = \sum_{k=\bar{k}+1}^{\infty} P^{(\mu_k, G)} \mathbf{1}_{[t_k, t_{k+1}-1]}(t). \tag{12}$$

Theorem 1. *Consider a connected graph $G = (\mathcal{S}, E)$ and a distribution $\mu = (\mu(s) : s \in \mathcal{S})$. Assume that $G[supp(\mu)]$ is not connected. Then any Markov chain $X = (X(t) : t \in \mathbb{N})$ constructed above with transition matrix given in (12) and with sequence of times $(t_\ell = \ell^{5d} : \ell \in \mathbb{N})$ is consistent with G and (1) holds true, i.e.*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{1}_{\{X(m)=s\}} = \mu(s), \qquad s \in \mathcal{S} \qquad a.s..$$

Proof. The fact that X is consistent with G derives from the construction of P (see (7) and (12)).

To prove the result, we first check that it is true for the sequence of times $(t_\ell = \ell^{5d} : \ell \in \mathbb{N})$, i.e.

$$\lim_{\ell\to\infty} \frac{1}{t_\ell} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}} = \mu(s), \qquad s \in \mathcal{S} \qquad a.s.. \tag{13}$$

By definition of $(t_\ell : \ell \in \mathbb{N})$ one has

$$\lim_{\ell\to\infty} \frac{t_{\ell+1} - t_\ell}{t_\ell} = 0 \text{ and } \lim_{\ell\to\infty} \frac{t_{\ell+1}}{t_\ell} = 1. \tag{14}$$

Therefore, if (13) holds true, also (1) is satisfied. In fact, for $t \in [t_\ell, t_{\ell+1} - 1]$, one has

$$\frac{1}{t_{\ell+1}} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}} \le \frac{1}{t} \sum_{m=0}^{t} \mathbf{1}_{\{X(m)=s\}} \le \frac{1}{t_\ell} \sum_{m=0}^{t_{\ell+1}-1} \mathbf{1}_{\{X(m)=s\}}$$

$$\le \frac{1}{t_\ell} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}} + \frac{1}{t_\ell} \sum_{m=t_\ell}^{t_{\ell+1}-1} 1 = \frac{1}{t_\ell} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}} + \frac{t_{\ell+1} - t_\ell}{t_\ell}.$$

In any case, by (13) and (14)

$$\lim_{\ell\to\infty} \frac{1}{t_{\ell+1}} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}} = \lim_{\ell\to\infty} \frac{1}{t_\ell} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}}$$

$$= \lim_{\ell\to\infty} \left[\frac{1}{t_\ell} \sum_{m=0}^{t_\ell-1} \mathbf{1}_{\{X(m)=s\}} + \frac{t_{\ell+1} - t_\ell}{t_\ell} \right],$$

therefore, by the squeeze theorem,

$$\lim_{t\to\infty}\frac{1}{t}\sum_{m=0}^{t}\mathbf{1}_{\{X(m)=s\}}=\lim_{\ell\to\infty}\frac{1}{t_\ell}\sum_{m=0}^{t_\ell-1}\mathbf{1}_{\{X(m)=s\}}.$$

For $\varepsilon>0$ and $s\in\mathcal{S}$ let us define the sequence of events $(B_\ell(\varepsilon,s):\ell\in\mathbb{N})$ as

$$B_\ell(\varepsilon,s)=\left\{\left|\mu_\ell(s)-\frac{1}{t_{\ell+1}-t_\ell}\sum_{m=t_\ell}^{t_{\ell+1}-1}\mathbf{1}_{\{X(m)=s\}}\right|<\varepsilon\right\}. \tag{15}$$

To obtain (13) it is enough to prove that, for each $\varepsilon>0$ and $s\in\mathcal{S}$

$$\mathbb{P}\left(\liminf_{\ell\to\infty}B_\ell(\varepsilon,s)\right)=1. \tag{16}$$

By (15) and (16), we conveniently restrict the attention to the interval $[t_\ell,t_{\ell+1}-1]$.

Notice that for each initial distribution λ on $\mathcal{S}$, Lemma 1 and Dobrushin's Theorem (see e.g. [4]) give that

$$||\lambda P(t_\ell)^{\ell^{2d}}-\mu_\ell||_{TV}\le\delta(P(t_\ell)^d)^{\lfloor\frac{\ell^{2d}}{d}\rfloor}\le\left(1-\left(\frac{c}{\ell}\right)^d\right)^{\left\lfloor\frac{\ell^{2d}}{d}\right\rfloor}\le\exp\left(-\hat{c}\left\lfloor\frac{\ell^d}{d}\right\rfloor\right), \tag{17}$$

where $\hat{c}=c^d$, for ℓ large enough. Estimation (17) will turn out to be useful soon.

Let us fix $i=0,\dots,\ell^{2d}-1$ and consider the sequence of times

$$\tau_{\ell,i}=(t_\ell+h\ell^{2d}+i:h\ge0,t_\ell+h\ell^{2d}+i<t_{\ell+1}). \tag{18}$$

We observe that each time $t\in[t_\ell,t_{\ell+1}-1]$ can be written uniquely as $t=t_\ell+h\ell^{2d}+i$, for a value of $i=0,\dots,\ell^{2d}-1$ and a value of h.

Now, take some auxiliary sequences of independent random variables

$$Y^{(\ell,i)}=(Y^{(\ell,i)}(t):t\in\tau_{\ell,i}) \tag{19}$$

with values on $\mathcal{S}$ such that $Y^{(\ell,i)}(t)$ has distribution μ_ℓ. We notice that the random variables belonging to different sequences are not in general independent.

The proof of (16) would be simple if the X's were i.i.d. random variables, which is not the case. Then, we proceed by using the maximal coupling (see [24]) between the X's and the $Y^{(\ell,i)}$'s defined above.

Specifically, by inequality (17) one can couple $X(t_\ell+h\ell^{2d}+i)$ with $Y^{(\ell,i)}(t_\ell+h\ell^{2d}+i)$, by assuming the knowledge of the X's at the previous times in $\tau_{\ell,i}$ of the type $t_\ell+h'\ell^{2d}+i$, where $h'=0,\dots,h-1$.

For ℓ large enough, $h\ge1$ and $t_\ell+h\ell^{2d}+i<t_{\ell+1}$, one has

$$\mathbb{P}(X(t_\ell+h\ell^{2d}+i)\ne Y^{(\ell,i)}(t_\ell+h\ell^{2d}+i)\Big|\bigcap_{h'=0}^{h-1}\{X(t_\ell+h'\ell^{2d}+i)=s_{h'}\})=$$

$$= \mathbb{P}(X(t_\ell + h\ell^{2d} + i) \neq Y^{(\ell,i)}(t_\ell + h\ell^{2d} + i) | \{X(t_\ell + (h-1)\ell^{2d} + i) = s_{h-1}\})$$
$$\leq \exp\left(-\hat{c}\left\lfloor \frac{\ell^d}{d} \right\rfloor\right). \tag{20}$$

We notice that the first equality of (20) comes from the Markov property, while the second inequality comes from (17).

For $\ell \in \mathbb{N}$ and $i \in [0, \ell^{2d} - 1]$, let us define the events

$$A_{\ell,i} = \left\{X(t) = Y^{(\ell,i)}(t) : t \in \tau_{\ell,i} \setminus \{t_\ell + i\}\right\} =$$

$$= \left\{X(t_\ell + h\ell^{2d} + i) = Y^{(\ell,i)}(t_\ell + h\ell^{2d} + i) : h \geq 1 \text{ and } t_\ell + h\ell^{2d} + i < t_{\ell+1}\right\}. \tag{21}$$

We notice that $A_{\ell,i}$ is an event which compares the Markov chain X and the independent random variables $Y^{(\ell,i)}$ on times belonging to the sublattice $\tau_{\ell,i}$ of step ℓ^{2d}.

By denoting $A_{\ell,i,t} = \{X(t) = Y^{(\ell,i)}(t)\}$ for $t \in \mathbb{N}$, we can write

$$A_{\ell,i} = \bigcap_{t \in \tau_{\ell,i} \setminus \{t_\ell + i\}} A_{\ell,i,t}.$$

For ℓ large enough (we omit this statement hereafter, being clear its validity from the context), by using the chain rule over the $A_{\ell,i,t}$ ordered by increasing t, the Markov property on the sublattice $\tau_{\ell,i}$ and by (20), one has

$$\mathbb{P}(A_{\ell,i} | X(t_\ell + i) = s_0) \geq \left[1 - \exp\left(-\hat{c}\left\lfloor \frac{\ell^d}{d} \right\rfloor\right)\right]^{(\ell+1)^{5d}}$$
$$\geq 1 - (\ell+1)^{5d} \exp\left(-\hat{c}\left\lfloor \frac{\ell^d}{d} \right\rfloor\right), \tag{22}$$

for each $s_0 \in \mathcal{S}$. It is worth noticing that $(\ell+1)^{5d}$ overestimates the cardinality of $\tau_{\ell,i}$.

Since inequality in (22) holds true irrespectively on s_0, we can remove the conditioning on $X(t_\ell + i) = s_0$, so that

$$\mathbb{P}(A_{\ell,i}) \geq 1 - (\ell+1)^{5d} \exp\left(-\hat{c}\left\lfloor \frac{\ell^d}{d} \right\rfloor\right). \tag{23}$$

We also set $\hat{A}_\ell = \bigcap_{i=0}^{\ell^{2d}-1} A_{\ell,i}$ so that

$$\mathbb{P}(\hat{A}_\ell) = 1 - \mathbb{P}\left(\bigcup_{i=0}^{\ell^{2d}-1} A_{\ell,i}^c\right) \geq 1 - (\ell+1)^{7d} \exp\left(-\hat{c}\left\lfloor \frac{\ell^d}{d} \right\rfloor\right). \tag{24}$$

By (24) and the first Borel-Cantelli lemma, one has that

$$\mathbb{P}(\liminf_{\ell \to \infty} \hat{A}_\ell) = 1. \tag{25}$$

In fact, since

$$\sum_{\ell=1}^{\infty}(\ell+1)^{7d}\exp\left(-\hat{c}\left\lfloor\frac{\ell^d}{d}\right\rfloor\right)<\infty \tag{26}$$

then $\sum_{\ell=1}^{\infty}\mathbb{P}(\hat{A}_\ell^c)<\infty$.

Now, for $\varepsilon>0$ and $s\in\mathcal{S}$, let us define the sequence of events $(\hat{B}_\ell(\varepsilon,s):\ell\in\mathbb{N})$ as

$$\hat{B}_\ell(\varepsilon,s)=\bigcap_{i=0}^{\ell^{2d}-1}\left\{\left|\mu_\ell(s)-\frac{1}{|\tau_{\ell,i}|}\sum_{m\in\tau_{\ell,i}}\mathbf{1}_{\{Y^{(\ell,i)}(m)=s\}}\right|<\frac{\varepsilon}{2}\right\}. \tag{27}$$

Now, the rate function of a Bernoulli r.v. with parameter $\pi\in[0,1]$ is

$$I^{(\pi)}(x)=x\log\left(\frac{x}{\pi}\right)+(1-x)\log\left(\frac{1-x}{1-\pi}\right),\ \text{if } x\in(0,1); \tag{28}$$

while $I^{(\pi)}(x)=+\infty$ if $x\notin(0,1)$. By large deviation theory, one has

$$\mathbb{P}\left(\left|\mu_\ell(s)-\frac{1}{|\tau_{\ell,i}|}\sum_{m\in\tau_{\ell,i}}\mathbf{1}_{\{Y^{(\ell,i)}(m)=s\}}\right|\geq\frac{\varepsilon}{2}\right)\leq e^{-g(\varepsilon)|\tau_{\ell,i}|}\leq e^{-g(\varepsilon)\ell^{3d-1}}, \tag{29}$$

where

$$g(\varepsilon)=\min_{\pi\in[0,1]}\left\{I^{(\pi)}\left(\pi-\frac{\varepsilon}{2}\right),I^{(\pi)}\left(\pi+\frac{\varepsilon}{2}\right)\right\}>0.$$

Therefore, by the union bound,

$$\mathbb{P}(\hat{B}_\ell(\varepsilon,s))\geq 1-\ell^{2d}e^{-g(\varepsilon)\ell^{3d-1}}. \tag{30}$$

Therefore

$$\mathbb{P}(\liminf_{\ell\to\infty}\hat{B}_\ell(\varepsilon,s))=1. \tag{31}$$

Moreover,

$$\hat{B}_\ell(\varepsilon,s)\cap\hat{A}_\ell\subset B_\ell(\varepsilon,s),$$

thus

$$\liminf_{\ell\to\infty}(\hat{B}_\ell(\varepsilon,s)\cap\hat{A}_\ell)\subset\liminf_{\ell\to\infty}B_\ell(\varepsilon,s). \tag{32}$$

By (25), (31) and (32) we have the thesis. □

3. A discussion of the main results

This section presents some remarks about Theorem 1, along with some relevant examples.

As preliminary remark, we notice that a target distribution μ with disconnected support has to be perturbated to be simulated in the MCMC context (see formula (3)). The properties of our Markov chain allows us to control Dobrushin's coefficient – hence, leading to solving the question **Q**. Such a control cannot be obtained by using the standard Metropolis-Hastings algorithm.

Remark 1. *The definition of* $(t_\ell : \ell \in \mathbb{N})$ *provided in Theorem 1 represents only one of the possible choices. In this respect, it is interesting to note that the proof of Theorem 1 can be adapted to other sequences of times. For example, one can take* $t_{\ell+1} - t_\ell \geq c\ell^{5d-1}$*, with* $c > 0$*. One can reproduce the proof of Theorem 1 by suitably decomposing the intervals in subintervals with some useful properties. In particular, for any* $\ell \in \mathbb{N}$*, we can select* $I_\ell \in \mathbb{N}$ *and an increasing sequence*

$$t_\ell^{(0)}, t_\ell^{(1)}, \ldots, t_\ell^{(I_\ell)}$$

such that $t_\ell = t_\ell^{(0)}$*,* $t_\ell^{(I_\ell)} = t_{\ell+1}$ *and the following property holds*

$$\lim_{\ell\to\infty} \sup_{i\in\{0,1,\ldots,I_\ell-1\}} \frac{t_\ell^{(i+1)} - t_\ell^{(i)}}{t_\ell^{(i)}} = 0; \qquad \lim_{\ell\to\infty} \frac{t_\ell^{(0)} - t_{\ell-1}^{(I_\ell-1)}}{t_{\ell-1}^{(I_\ell-1)}} = 0.$$

Then, one can take the sequence $t_\ell^{(0)}, t_\ell^{(1)}, \ldots, t_\ell^{(I_\ell)}$ *to obtain that the Markov chain with transition matrix as in* (12) *satisfies* (1)*.*

Next example shows that the convergence of the distribution μ_k to the distribution μ should not be taken *too fast* and $t_{\ell+1} - t_\ell$ should be not taken *too small* in order to hold (1).

Example 2. *Let us consider a graph* $G = (\mathcal{S}, E)$ *with* $\mathcal{S} = \{s^1, s^2, s^3, s^4\}$ *and* $E = \{\{s^1, s^3\}, \{s^3, s^4\}, \{s^2, s^4\}\}$*.*

Let us take the distribution $\mu = (\mu(s) : s \in \mathcal{S})$ *having* $\mu(s^1) = \mu(s^2) = \eta(s^1) = \eta(s^2) = \frac{1}{2}$*, and define* $t_\ell = \ell$*, for each* $\ell \in \mathbb{N}$*, and the sequence of distributions* $(\hat{\mu}_\ell : \ell \in \mathbb{N})$ *where* $\hat{\mu}_\ell = \mu_{2^\ell}$*. In particular,* $||\hat{\mu}_\ell - \mu||_{TV} \leq \frac{\Gamma}{2^\ell}$ *(see* (4)*).*

We take a non-homogeneous Markov chain $X = (X(t) : t \in \mathbb{N})$ *with transition matrix* $P(\ell) = (p_{m,n}(\ell) : m, n = 1, 2, 3, 4)$*, at time* ℓ*, given by*

$$P(\ell) = P^{(\hat{\mu}_\ell, G)}, \qquad \ell \in \mathbb{N}.$$

Accordingly to the definition of $\tilde{p}$ *given in* (9) *and highlighting the dependence of* $\tilde{p}$ *on the index* ℓ*, one has*

$$\tilde{p}(\ell) = \frac{1}{2(1 + 2^{\ell-1})}, \qquad \ell \in \mathbb{N}. \tag{33}$$

Thus, (33) *gives that* $p_{1,1}(\ell) = 1 - \frac{1}{2(1+2^{\ell-1})}$ *at time* ℓ *(see* (8)*). Therefore, the Borel-Cantelli's Lemma guarantees that*

$$|\{\ell \in \mathbb{N} : X(\ell) = s^1, X(\ell+1) \neq s^1\}| < \infty \qquad a.s., \tag{34}$$

and therefore

$$\mathbb{P}\left(\bigcap_{s\in\mathcal{S}} \{\lim_{t\to\infty} \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{1}_{\{X(m)=s\}} = \mu(s)\}\right) = 0. \tag{35}$$

In fact, formula (34) *allows to consider only* $\omega \in \Omega$ *such that condition* $|\{\ell \in \mathbb{N} : X(\ell) = s^1, X(\ell+1) \neq s^1\}| < \infty$ *is satisfied. If* $X(\ell) = s^1$ *for a finite number of times* ℓ*, then*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{1}_{\{X(m)=s^1\}} = 0 \qquad a.s..$$

Notice that Example 2 gives a natural comparison between our setting and the simulated annealing (see [22, 26]). The goal of the quoted references was to obtain convergence in law, while here we obtain convergence almost surely of the empirical distribution to the target one.

In both cases the hope is that the rate of convergence is fast but, if one tries to have an *excessively high* rate of convergence, it leads to local minima (case of simulated annealing) or not convergence of the empirical measure to the target distribution μ in our framework. In the case of excessively fast convergence rate, the response to question **Q** might be wrong.

We now give an illustrative example based on the existing literature to highlight the usefulness of Theorem 1.

Example 3. *The situation of connected graph* G *and* $G[supp\,(\mu)]$ *not connected can be deduced from previous contributions in the literature. In this respect, we mention the hard-core model with two kinds of particles (see e.g., [21, 27]). In particular, [27] presents a gas model on a lattice with particles of two kinds: one is small-sized (type 1) and the other is large (type 2). Each type 1-particle occupies only one site, leaving free the others; the particles of type 2 are so large that they prevent to occupy their sites and also their adjacent ones. At each site of the lattice, one can find no more than one particle. In doing so, we assign to each site* $x \in \Lambda$ *a vector* $(n_1(x), n_2(x)) \in \{(0,0),(0,1),(1,0)\}$*, where* $n_j(x)$ *denotes the number of particles of type* j *on site* x*. We define the collection of such vectors – under the constraints related to particles of type 1 and 2 described above – simply as the* relevant states $\tilde{\mathcal{S}}$*.*

In this framework, we consider the hard-core model with a canonical ensemble condition – i.e., with a fixed number of particles of type 1 and 2, say N_1 *and* N_2 *respectively – on a finite lattice* Λ*. We take a graph* $G = (\mathcal{S}, E)$ *as follows. The state space* $\mathcal{S}$ *is the collection of two sequences of nonnegative integers over the lattice* Λ *describing the number of particles of type 1 and 2 over all the sites of* Λ*. By using the notation above, we say that* $n_j(x) = 0, 1, \ldots, N_j$*, with the only constraint that the total number of type 1 and type 2 particles is given by* N_1 *and* N_2*, respectively. Given two states* $s, t \in \mathcal{S}$*, there is an edge* $\{s,t\} \in E$ *if and only if* t *is obtained from* s *by moving only one particle from a site to an adjacent one on the lattice* Λ*. Therefore,* G *is a connected graph.*

The hard-core model introduces a distribution μ *on* $\mathcal{S}$ *with support* $\tilde{\mathcal{S}}$*. We do not need the detailed definition of* μ *but only that* μ *has support* $\tilde{\mathcal{S}}$*. If the lattice is one-dimensional – so that two sites are adjacent when they have unitary distance – then the particles types are ordered over the lattice and the order is preserved over time. Therefore, if both* $N_1, N_2 > 0$*, then* $G[supp\,(\mu)]$ *is not connected.*

We now treat some cases that are not covered by Theorem 1.

In the case that $G[supp\,(\mu)]$ is connected one can construct a standard MCMC algorithm using a reversible homogeneous Markov chain $X = (X(t) : t \in \mathbb{N})$ having μ as invariant distribution. A possible choice for the transition matrix of X is e.g. $P^{(\mu, G[supp\,(\mu)])}$.

We point out that if G is not connected and $supp\,(\mu)$ is contained in more than one connected components of G, then one cannot pass from a component to another one. This assures that it does not exist a stochastic process consistent with G which satisfies (1).

Finally, assume that G is connected and $G[supp\,(\mu)]$ is not connected. Then, we can apply Theorem 1 that shows a non-homogeneous Markov chain responding to question **Q**. Nevertheless, each homogeneous Markov chain X consistent with G does not satisfy (1). Indeed, if X would be consistent with G and satisfying (1), then the ergodic theorem would imply that some states out of $supp\,(\mu)$ should be visited with frequency greater than zero.

Remark 2. *Theorem 1 can be used also in some circumstances in which $G[supp\,(\mu)]$ is connected. Suppose that the set of the vertices $\mathcal{S}$ is given by the union of three disjoint components: $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ such that $\mu(\mathcal{S}_1) = 1-\epsilon-\epsilon^2$, $\mu(\mathcal{S}_3) = \epsilon$, while $\mu(\mathcal{S}_2) = \epsilon^2$, being $0 < \epsilon \ll 1$ very small. Moreover, assume that it does not exist an edge of G connecting $\mathcal{S}_1$ and $\mathcal{S}_3$. Therefore the Markov chain needs to visit the states in $\mathcal{S}_2$ to achieve $\mathcal{S}_1$ from $\mathcal{S}_3$. If the starting state of the Markov chain belongs to $\mathcal{S}_3$, then it will remain in $\mathcal{S}_3$ for a long time being $\mu(\mathcal{S}_2)$ particularly small. Evidently, this leads to a slow convergence of the empirical distribution to μ; the introduction of the μ_k's and of the transition matrix in (12) might be of usefulness for letting the convergence be faster. In doing so, our approach is close to the principle of the annealing.*

In this context, we can also implement a different strategy for raising the speed of convergence of the MCMC algorithm. Specifically, one can fix $k \in \mathbf{N}$ and consider the perturbated distribution μ_k instead of the target distribution μ. One can conjecture that the speed of convergence of a homogeneous Markov chain $X^{(k)} = (X^{(k)}(t) : t \in \mathbb{N})$ with transition matrix $P^{(\mu_k, G)}$ is higher than X with transition matrix $P^{(\mu, G[supp\,(\mu)])}$. Following (4), we bound the error from above as follows:

$$\lim_{t\to\infty} \left| \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{1}_{\{X^{(k)}(m)=s\}} - \mu(s) \right| \leq 2||\mu_k - \mu||_{TV} \leq \frac{2\Gamma}{k}, \qquad s \in \mathcal{S} \qquad a.s.. \tag{36}$$

Let us consider $f : \mathcal{S} \to \mathbb{R}$ and $\mathbb{E}_\mu(f) = \sum_{s\in\mathcal{S}} f(s)\mu(s)$, by ergodic theory one has

$$\begin{aligned} \lim_{t\to\infty} \left| \frac{1}{t} \sum_{m=0}^{t-1} f(X^{(k)}(m)) - \mathbb{E}_\mu(f) \right| &\leq 2 \max_{s\in\mathcal{S}} |f(s)| \cdot ||\mu_k - \mu||_{TV} \\ &\leq 2 \max_{s\in\mathcal{S}} |f(s)| \cdot \frac{\Gamma}{k}, \qquad a.s.. \end{aligned} \tag{37}$$

Thus, accepting the error $2\max_{s\in\mathcal{S}} |f(s)|\Gamma/k$, that can be taken arbitrarily small,

one can always use an homogeneous Markov chain to numerically estimate $\mathbb{E}_\mu(f)$.

We point out that a proper selection of the graph may lead to a more efficient MCMC procedure. In particular, graphs can contribute to the reduction of the number of possible transitions among states, as also Gibbs sampler proposes (see e.g. [18]). In fact, when the number of the states is extremely large, then the unconstrained transition probabilities involving all the pairs of states may be too small, hence too difficult to simulate. In this respect, a proper choice of the graph should ensure the connections among highly probable states, thus avoiding the creation of metastable states (sometimes called *wells*, see [3, 23]). Indeed, wells are states in which the Markov chain is expected to spend an extremely long time before being able to visit other high-probability ones. This would increase dramatically the mixing time and the convergence speed of the MCMC algorithm (see e.g. [1, 16, 17]).

In this context, a very useful reading are [9, 15, 28], where the (stochastically) monotone MCMC is explored. In details, a Markov chain is said to be stochastically monotone when the states space is endowed with a partial order and there exists a coupling of the chain with itself that maintains the partial order of the states space at any time. Stochastically monotone Markov chains are particularly simple in the simulation procedures (see [15] and [28] for connections with the perfect simulation literature). Now, let us assume that the states space $\mathcal{S}$ is endowed with a partial order and consider the target distribution μ on $\mathcal{S}$. Naturally, there are infinite Markov chains satisfying (1). Some of them might be stochastically monotone, i.e. simple in the simulation process. The role of the graph in obtaining stochastically monotone Markov chains might then be crucial.

As a paradigmatic example, we can take the classical ferromagnetic Ising model assigning a spin $\sigma(i) \in \{-1, +1\}$ to each vertex $i \in V$ and assume that the set $\mathcal{S} = \{-1, +1\}^V$ is endowed with a partial order such that $\sigma' \preceq \sigma''$ if and only if $\sigma'(i) \leq \sigma''(i)$ for each $i \in V$. In this situation, we have that the Markov chain identified by the Gibbs sampler is stochastically monotone, and this property leads to affordable simulation exercises for the convergence towards the Gibbs measure of the ferromagnetic Ising model (see [28] and, more recently, [10]). There are also other Markov chains converging to the Gibbs measure which do not maintain the ordering of the states space (see e.g. [7, 28]).

It is not difficult to construct other examples for non-ferromagnetic Ising models (where the Gibbs sampler is not stochastically monotone) such that Markov chains consistent with suitably defined graphs are stochastically monotone.

4. Product of graphs and product distributions

We now introduce the standard definition of product of graphs, as in [29]. It leads to a simplification of the MCMC simulations.

Definition 2. Consider two graphs $G_1 = (\mathcal{S}_1, E_1), G_2 = (\mathcal{S}_2, E_2)$. The strong product $G_1 \boxtimes G_2$ is a graph $G = (\mathcal{S}, E)$, where $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ and E collects the couples $\{(s_1, s_2), (\bar{s}_1, \bar{s}_2)\}$, with $(s_1, s_2), (\bar{s}_1, \bar{s}_2) \in \mathcal{S}$, such that one of the following condition is verified

- $\{s_1, \bar{s}_1\} \in E_1$ and $s_2 = \bar{s}_2$;
- $s_1 = \bar{s}_1$ and $\{s_2, \bar{s}_2\} \in E_2$;
- $\{s_1, \bar{s}_1\} \in E_1$ and $\{s_2, \bar{s}_2\} \in E_2$.

Since the strong product of graphs is associative (see [29]), then Definition 2 can be extended to any collection of $r > 2$ graphs obtaining $G = G_1 \boxtimes \cdots \boxtimes G_r$.

Let us consider now r finite sets $\mathcal{S}_1, \ldots, \mathcal{S}_r$ and take a product distribution $\mu = \prod_{h=1}^{r} \mu_h$, where μ_h is a distribution on the space $\mathcal{S}_h$. We construct r independent Markov chains $X_1 = (X_1(t) : t \in \mathbb{N}), \ldots, X_r = (X_r(t) : t \in \mathbb{N})$ such that the h-th Markov chain X_h has state space $\mathcal{S}_h$ and an arbitrary initial distribution $\lambda_h = (\lambda_h(s_h) : s_h \in \mathcal{S}_h)$, for each $h = 1, \ldots, r$.

Moreover, by replacing $\mathcal{S}$ with $\mathcal{S}_h$ and μ with μ_h, we replicate the construction provided before Theorem 1. In so doing, we take $\eta_h : \mathcal{S}_h \to [0, M_h]$ to define the distribution μ_h, for $h = 1, \ldots, r$. Then, we take $k \in \mathbb{N}$ to define the perturbed distribution $(\mu_h)_k = ((\mu_h)_k(s_h) : s_h \in \mathcal{S}_h)$.

Now, take a sequence of increasing times $(t_\ell^{(h)} : \ell \in \mathbb{N})$, such that

$$\min_{h=1,\ldots,r} t_{\ell+1}^{(h)} - t_\ell^{(h)} \geq \zeta \ell^{5d_h - 1}, \tag{38}$$

with ζ a positive constant and d_h is the diameter of graph G_h.

The transition matrices of X_h are $(P_h(t) : t \in \mathbb{N})$ as in (12):

$$P_h(t) = \sum_{k=1}^{\infty} P^{((\mu_h)_k, G_h)} \mathbf{1}_{[t_k^{(h)}, t_{k+1}^{(h)} - 1]}(t). \tag{39}$$

We introduce the Markov chain

$$X = \Big(X(t) = (X_1(t), \ldots, X_r(t)) \in \mathcal{S} : t \in \mathbb{N} \Big). \tag{40}$$

Next result is similar to Theorem 1 but it is based on the independent Markov chains constructed above.

Theorem 2. *Let $\mathcal{S} = \prod_{h=1}^{r} \mathcal{S}_h$ and $G(\mathcal{S}, E) = G_1(\mathcal{S}_1, E_1) \boxtimes \cdots \boxtimes G_r(\mathcal{S}_r, E_r)$. Let us consider $\mu_1, \ldots, \mu_r$ constructed as above, the product distribution $\mu = \prod_{h=1}^{r} \mu_h$ and consider the Markov chains X as in (40).*

Then

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{1}_{\{X(m)=s\}} = \lim_{t\to\infty} \frac{1}{t} \sum_{m=0}^{t-1} \prod_{h=1}^{r} \mathbf{1}_{\{X_h(m)=s_h\}} = \prod_{h=1}^{r} \mu_h(s_h) = \mu(s), \tag{41}$$

for each $s = (s_1, \ldots, s_r) \in \mathcal{S}$.

Proof. We follow the arguments in the proof of Theorem 1 and proceed in a componentwise form. We take a sequence of times $\tau_{\ell,i,h}$ as in (18) for the h-th component, with step ℓ^{2d_h}. Accordingly, we take the variables $Y^{(\ell,i,h)}$ as in (19), that are independent and with distribution $(\mu_h)_\ell$.

On the ground of what done for deriving the constant Γ in Section 2, we can find a positive $\hat{\Gamma}$ such that

$$\left|\left|\mu - \prod_{h=1}^{r}(\mu_h)_\ell\right|\right|_{TV} \leq \sum_{h=1}^{r}||\mu_h - (\mu_h)_\ell||_{TV} \leq \frac{r\hat{\Gamma}}{\ell}, \tag{42}$$

so that the total variation distance in (42) goes to zero as $\ell \to \infty$.

Similarly to (26), we can employ Borel-Cantelli Lemma to obtain

$$\sum_{h=1}^{r}\sum_{\ell=1}^{\infty}(\ell+1)^{7d_h}\exp\left(-\hat{c}_h\left\lfloor\frac{\ell^{d_h}}{d_h}\right\rfloor\right) < \infty, \tag{43}$$

where $\hat{c}_h$ is a suitable positive constant. By following the final steps of the proof of Theorem 1, condition (43) guarantees the convergence in (41). □

Remark 3. *Theorem 2 is grounded on the strong assumption that the target distribution is a product one. However, such a condition leads to the possibility of estimating the target distribution in a parallel way – hence, leading to a remarkable reduction of the computational complexity of the procedure. Interestingly, the case of weakly dependent components might be also explored through the procedure presented in Theorem 2. Moreover, Theorem 2 might present relevant rooms for applications – for instance, in game theory. Specifically, one can build a dynamic game where each player can move from her/his strategy to another one only if such strategies are adjacent on a prefixed graph G. In this context, we point out that Theorem 2 can be useful for identifying some mixed Nash equilibria, that can be seen as product measures (see [8]).*

Acknowledgments

We are deeply grateful to two anonymous referees and to the Editor-in-Chief for the remarkably pertinent suggestions, that greatly improved the article.

References

[1] P. Baldi, A. Frigessi, and M. Piccioni. Importance sampling for Gibbs random fields. *Ann. Appl. Probab.*, 3(3):914–933, 1993. MR1233634

[2] F. Bartolucci, L. Scaccia, and A. Mira. Efficient Bayes factor estimation from the reversible jump output. *Biometrika*, 93(1):41–52, 2006. MR2277738

[3] J. Beltrán and C. Landim. Tunneling and metastability of continuous time Markov chains. *J. Stat. Phys.*, 140(6):1065–1114, 2010. MR2684500

[4] P. Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues. MR1689633
[5] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors. *Handbook of Markov chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2011. MR2742422
[6] B. P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. *J. R. Stat. Soc. Series B*, 57(3):473–484, 1995.
[7] R. Cerqueti and E. De Santis. Stochastic Ising model with flipping sets of spins and fast decreasing temperature. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(2):757–789, 2018. MR3795065
[8] R. Cerqueti and E. De Santis. G-games with coalitions. *Submitted*, 2021.
[9] D. J. Daley. Stochastically monotone Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 10:305–317, 1968. MR0242270
[10] E. De Santis and A. Lissandrelli. Developments in perfect simulation of Gibbs measures through a new result for the extinction of Galton-Watson-like processes. *J. Stat. Phys.*, 147(2):231–251, 2012. MR2922768
[11] P. Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009. MR2476411
[12] P. Diaconis. Some things we've learned (about Markov chain Monte Carlo). *Bernoulli*, 19(4):1294–1305, 2013. MR3102552
[13] R. Dobrushin. Central limit theorem for non-stationary Markov chains. I. *Teor. Veroyatnost. i Primenen.*, 1:72–89, 1956. MR0086436
[14] P. Fearnhead. Markov chain Monte Carlo, sufficient statistics, and particle filters. *J. Comput. Graph. Statist.*, 11(4):848–862, 2002. MR1951601
[15] J. A. Fill and J. Kahn. Comparison inequalities and fastest-mixing Markov chains. *Ann. Appl. Probab.*, 23(5):1778–1816, 2013. MR3114917
[16] A. Frigessi, P. di Stefano, C.-R. Hwang, and S. J. Sheu. Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B*, 55(1):205–219, 1993. MR1210432
[17] A. Frigessi, F. Martinelli, and J. Stander. Computational complexity of Markov chain Monte Carlo methods for finite Markov random fields. *Biometrika*, 84(1):1–18, 1997. MR1450188
[18] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI(6):721–741, 1984.
[19] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995. MR1380810
[20] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970. MR3363437
[21] I. Jauslin and J. L. Lebowitz. High-fugacity expansion, Lee-Yang zeros, and order-disorder transitions in hard-core lattice systems. *Comm. Math. Phys.*, 364(2):655–682, 2018. MR3869440
[22] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. MR0702485

[23] C. Landim, M. Loulakis, and M. Mourragui. Metastable Markov chains: from the convergence of the trace to the convergence of the finite-dimensional distributions. *Electron. J. Probab.*, 23:Paper No. 95, 34, 2018. MR3858923

[24] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1992. MR1180522

[25] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[26] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli. Convergence and finite-time behavior of simulated annealing. *Adv. in Appl. Probab.*, 18(3):747–771, 1986. MR0857328

[27] T. J. Oliveira and J. F. Stilck. Solution on the bethe lattice of a hard core athermal gas with two kinds of particles. *The Journal of Chemical Physics*, 135(18):184502, 2011.

[28] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 223–252, 1996. MR1611693

[29] G. Sabidussi. Graph multiplication. *Math. Z.*, 72:446–457, 1959/1960. MR0209177

[30] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Stat.*, 22(4):1701–1762, 1994. MR1329166