Raport Badawczy
Research Report

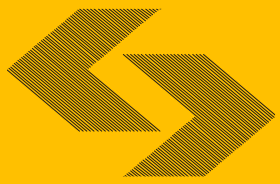
RB/94/2007

On some method for diagnosing convergence in MCMC setups via atoms and renewal sets

M. Romaniuk

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences



Control and Cybernetics

vol. **36** (2007) No. 4

On some method for diagnosing convergence in MCMC setups via atoms and renewal sets

by

Maciej Romaniuk

Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01–447 Warszawa, Poland
e-mail: mroman@ibspan.waw.pl

Abstract: MCMC setups are among the best known methods for conducting computer simulations necessary in statistics, physics, biology, etc. However, to obtain appropriate solutions, additional convergence diagnosis must be applied for trajectory generated by Markov Chain. In the paper we present the method for dealing with this problem, based on features of so called "secondary" chain (the chain with specially selected state space). The secondary chain is created from the initial chain by picking only some observations connected with atoms or renewal sets. The discussed method has some appealing properties, like high degree of diagnosis automation. Apart from theoretical lemmas, the example of application is also provided.

Keywords: convergence diagnosis, Markov Chain Monte Carlo, Markov Property, atom, renewal set, renewal theory, automated diagnosis of simulations.

1. Introduction

The end of the previous century brought a colossal improvement in speed of calculations. Because of computer development, researchers could build more complex, more "real-life" models. The same applies to mathematics, statistics, physics and biology, where computer simulations are widely used.

One of the best known methods in computer simulations are MCMC (*Markov Chain Monte Carlo*) algorithms, successors of MC (*Monte Carlo*) approach (see Metropolis et al., 1953, Metropolis and Ulam, 1949). They are commonly used in many practical areas (see, e.g., Boos, Zhang, 2000; Booth, Sarkar, 1998; Bremaud, 1999, Doucet et al., 2000; Gelfand et al., 1990; Gilks et al., 1997; Kass et al., 1998; Koronacki et al., 2005; Lasota, Niemiro, 2003; Li et al., 2000; Mehta et al., 2000; Robert, Casella, 2004; Romaniuk, 2003).

The MCMC method is based on a simple observation. In order to find the expected value $\mathbb{E}_{\pi_X} h(X)$ for some function $h(.)$ and probability distribution $\pi_X(.)$,

we could generate Markov Chain $X_0, X_1, X_2, \ldots$ with the stationary distribution π_X. The convergence of estimator derived from the simulated samples is guaranteed by ergodic theorems (see, e.g., Robert, Casella, 2004, for additional details). Hence, we do not have to generate values directly from $\pi_X(.)$, but we may use more general algorithms like the Gibbs sampler or the Metropolis-Hastings algorithm.

But during the conduct of simulations two questions arise all the time. The first one is connected with choosing appropriate number of steps n_{stat} for simulated trajectory, when the sampled transition probability $\Pr_{x_0}^{n_{\text{stat}}}(.)$ is close enough to the assumed stationary probability $\pi_X(.)$ regardless of the starting point x_0. The second one is related to finding the number of steps n_2, when the estimator of $\mathbb{E}_{\pi_X} h(X)$, derived from sample $X_{n_{\text{stat}}+1}, X_{n_{\text{stat}}+2}, \ldots, X_{n_2}$, has sufficiently small *error*, measured e.g. by variance. These two questions are covered by *convergence diagnosis* and are one of main aspects for MCMC simulations.

There is a lot of various convergence diagnosis methods (see, e.g., Robert, Casella, 2004, for comparative review). But we have to say that it is not so easy to compare them and find "the best one" or even "the best ones". Firstly, these methods very often base on different *features* of the underlying Markov Chains, e.g. specific structure of state space for Markov Chain. Secondly, the two questions mentioned before are used to be written in mathematical formulas not corresponding to one another, i.e. not directly comparable. Thirdly, it is not even possible to draw a comparison between the *heuristic* and *theoretical* (i.e. based on mathematical proofs) methods. Therefore, each new convergence diagnosis method may be seen as additional tool for experimenters, which gives them a new possibility to check the obtained simulations.

In this paper we discuss the method based on the concept of the so called *"secondary chain"*. Such a chain is derived from the original trajectory by observing the samples only in moments determined by special, probability rules. These rules are connected with notion of *atoms* and *renewal sets*, which are a specific example of more general *renewal moments* and are part of *renewal theory*.

The method presented has three main advantages. Firstly, it is supported by strong *mathematical reasoning*. Therefore, it is far less influenced by observer's intuition and his experience than the entirely heuristic methods. Secondly, the obtained solutions are *strict*, i.e. they are not asymptotic. Hence, this method is not biased by additional error provided by limit theorems. Thirdly, the discussed solutions may be used in *highly automated manner*. This gives the possibility to prepare general diagnosis algorithms for a wide class of MCMC problems.

The paper is organized as follows. In Section 2 we present the necessary basic definitions and theorems. Then, in Section 3.1 we introduce the notion of secondary chain and some fundamental facts about it. In Section 3.2 we formulate two inequalities which are directly connected to the convergence diagnosis questions mentioned before. Then, we find answers for these questions in Sections 3.3 and 3.4 in a few lemmas and more heuristic remarks. These are based

on the properties of secondary chain. In Section 4 we present how the derived results may be applied in case of a simple example. The concluding remarks are contained in Section 5.

2. Basic definitions and theorems

In this section we introduce fundamental definitions and theorems. Additional necessary definitions may be found in, e.g., Bremaud (1999), Fishman (1996), Robert and Casella (2004).

Let $(X_i)_{i=0} = (X_0 = x_0, X_1, \ldots)$ denote a Markov Chain (abbreviated further MC), and $\mathbb{B}(\mathcal{X})$ be the σ–field of Borel sets for space $\mathcal{X}$.

The chain $(X_i)_{i=0}$ has its values in a space $\mathcal{X}$, where $\mathcal{X} \subset \mathbb{N}$ or $\mathcal{X} \in \mathbb{B}(\mathbb{R}^k)$. In the first case such MC is called a *discrete MC*, and in the second – an MC on *continuous state space*.

Suppose that the chain $(X_i)_{i=0}$ is ergodic and has an adequate stationary probability distribution $\pi_X(.)$. In this paper the term "ergodicity" means that the chain is recurrent (or Harris recurrent in case of MC on continuous state space $\mathcal{X}$), aperiodic and irreducible.

If $(X_i)_{i=0}$ is a *discrete* Markov Chain, we define its transition matrix $\mathbb{P}_X$ as

$$\mathbb{P}_X = (\Pr(X_{k+1} = j | X_k = i))_{i,j=1}^{s_X} , \tag{1}$$

where s_X is the power of $\mathcal{X}$. In case of continuous state space $\mathcal{X}$, let us denote by $\mathcal{K}_X(.,.)$ the transition kernel of this chain

$$\Pr(X_{k+1} \in \mathcal{B} | X_k = x) = \int_{\mathcal{B}} \mathcal{K}_X(x, y)\, dy . \tag{2}$$

Definition 1 *The set $\mathcal{A}$ is called an* atom *if there exists a probability distribution $\nu(.)$ such that*

$$\Pr(X_{k+1} \in \mathcal{B} | X_k = x) = \nu(\mathcal{B}) \tag{3}$$

for every $x \in \mathcal{A}$ and every $\mathcal{B} \in \mathbb{B}(\mathcal{X})$

Definition 2 *The set $\mathcal{A}$ is called* renewal set *if there exist a real $0 < \epsilon < 1$ and a probability measure $\nu(.)$ such that*

$$\Pr(X_{k+1} \in \mathcal{B} | X_k = x) \geq \epsilon \nu(\mathcal{B}) \tag{4}$$

for every $x \in \mathcal{A}$ and every $\mathcal{B} \in \mathbb{B}(\mathcal{X})$.

These two definitions may be found in, e.g., Asmussen (1979), Robert and Casella (2004).

If $\mathcal{A}$ is a renewal set, it is advantageous to slightly change the used MCMC algorithm which generates the values of $(X_i)_{i=0}$. It is easily seen that

$$\Pr(X_{k+1} | X_k) = \epsilon \nu(X_{k+1}) + (1 - \epsilon) \frac{\Pr(X_{k+1} | X_k) - \epsilon \nu(X_{k+1})}{1 - \epsilon} \tag{5}$$

in case of discrete MC, or

$$\mathcal{K}(x_k, x_{k+1}) = \epsilon\nu(x_{k+1}) + (1-\epsilon)\frac{\mathcal{K}(x_k, x_{k+1}) - \epsilon\nu(x_{k+1})}{1-\epsilon} \tag{6}$$

for MC on continuous state space $\mathcal{X}$. Hence, we have the following modification of the algorithm: when $X_k \in \mathcal{A}$, generate X_{k+1} according to

$$X_{k+1} = \begin{cases} X_{k+1} \sim \nu(.) & \text{if } U_{k+1} \leq \epsilon \\ X_{k+1} \sim \frac{\mathcal{K}(x_k,.) - \epsilon\nu(.)}{1-\epsilon} & \text{if } U_{k+1} > \epsilon \end{cases}, \tag{7}$$

where U_i are *iid* random variables from a uniform distribution on $[0,1]$, independent on $(X_i)_{i=0}$. Since (5) or (6), the modification (7) of the MCMC algorithm does not change the properties of the chain. Also its stationary distribution is still the same, i.e. $\pi_X(.)$. This modification for MCMC algorithms was introduced in Athreya and Ney (1978), Nummelin (1978). The generation according to (7) may be difficult because of the complex structure of the "remainder" kernel. A way around this problem was shown in Mykland, Tierney and Yu (1995).

DEFINITION 3 *The atom (or renewal set) $\mathcal{A}$ is called* geometrically ergodic atom *(or renewal set) if there exist $r > 1$ and $M > 0$ such that*

$$|\Pr_x^n(y) - \pi_X(y)| \leq Mr^{-n} , \tag{8}$$

for any $x, y \in \mathcal{A}$, where $\Pr_x^n(.)$ denotes $\Pr(X_n = . \,|X_0 = x)$.

Let us denote by $\mathbb{E}_{\pi_X} h(X)$ the expected value of the function $h : \mathcal{X} \to \mathbb{R}$ calculated according to the stationary distribution π_X. Appropriate symbols – $\mathrm{Cov}_{\pi_X}(g,h)$ and $\mathrm{Var}_{\pi_X}(h)$ – are used for covariance and variance.

3. Proposal for the of convergence diagnosis method

In this section we present a convergence diagnosis method for the MCMC output. This proposal uses the notions of *atoms* and *renewal sets* and also some properties derived for *discrete* Markov Chains (see Section 2).

3.1. Introducing secondary chain

Suppose that we are interested in diagnosing convergence of some ergodic Markov Chain $(X_i)_{i=0} = (X_0 = x_0, X_1, \ldots)$. We denote a stationary probability measure for this chain by $\pi_X(.)$, its transition matrix by $\mathbb{P}_X$ (or transition kernel by $\mathcal{K}_X(.,.)$ in case of MC on continuous state space) and the space of its values by $\mathcal{X}$. Suppose also that we know two atoms (or renewal sets) $\mathcal{A}_1, \mathcal{A}_2$ for this chain.

Therefore, we can create the *secondary* chain $(Y_i)_{i=1}$ based on our initial chain $(X_i)_{i=0}$. If $\mathcal{A}_1, \mathcal{A}_2$ are atoms, then we can define

$$\zeta_1 := \min\{i = 1, \ldots : X_i \in \mathcal{A}_1 \cup \mathcal{A}_2\} \ , \tag{9}$$

$$\zeta_{k+1} := \min\{i > \zeta_k : X_i \in \mathcal{A}_1 \cup \mathcal{A}_2\} \ , \tag{10}$$

$$Y_k = X_{\zeta_k} \ . \tag{11}$$

It is seen that the chain $(Y_i)_{i=1}$ has Markov Property for the *truncated* (reduced) space $\mathcal{Y}' := \{\mathcal{A}_1, \mathcal{A}_2\}$ — see Lemma 1 for proof.

If these two sets are renewal sets, we should introduce the modification (7) and change the definition of the chain $(Y_i)_{i=1}$ according to

$$\zeta_1 := \min\{i = 1, \ldots : (X_i \in \mathcal{A}_1 \wedge U_i \leq \epsilon_{\mathcal{A}_1}) \vee (X_i \in \mathcal{A}_2 \wedge U_i \leq \epsilon_{\mathcal{A}_2})\} \ , \tag{12}$$

$$\zeta_{k+1} := \min\{i > \zeta_k : (X_i \in \mathcal{A}_1 \wedge U_i \leq \epsilon_{\mathcal{A}_1}) \vee (X_i \in \mathcal{A}_2 \wedge U_i \leq \epsilon_{\mathcal{A}_2})\} \ , \tag{13}$$

$$Y_k = X_{\zeta_k} \ , \tag{14}$$

where $\epsilon_{\mathcal{A}_j}$ denotes the parameter ϵ for appropriate renewal set $\mathcal{A}_j$ in condition (7). Also in this case the secondary chain $(Y_i)_{i=1}$ has Markov Property for the space $\mathcal{Y}'$.

We may conclude previous observations in a simple lemma:

LEMMA 1 *If $\mathcal{A}_1, \mathcal{A}_2$ are atoms (or renewal sets), the chain $(Y_i)_{i=1}$ defined by conditions* (9) — (11) *(or* (12) — (14)*, respectively) is a Markov Chain for the space $\mathcal{Y}' := \{\mathcal{A}_1, \mathcal{A}_2\}$. This chain is ergodic.*

Proof. The chain $(Y_i)_{i=1}$ has Markov property for reduced space $\{x : x \in \mathcal{A}_1 \cup \mathcal{A}_2\}$ from Strong Markov Property. If $\mathcal{A}_j$ is a small set, then from (3) probability $\Pr(Y_{k+1} \in \mathcal{B} | Y_k = y)$ is constant for all $y \in \mathcal{A}_j$. Hence

$$\Pr(Y_{k+1} \in \mathcal{A}_{j'} | Y_k \in \mathcal{A}_j) = \Pr(Y_{k+1} \in \mathcal{A}_{j'} | Y_k = y) \tag{15}$$

for all $y \in \mathcal{A}_j$.

If $\mathcal{A}_j$ is a renewal set, the argument is similar. The modification (7) introduces independent generation from probability measure $\nu_{\mathcal{A}_j}(.)$ with probability $\epsilon_{\mathcal{A}_j}$. And this measure is constant for all $y \in \mathcal{A}_j$.

The ergodicity of $(Y_i)_{i=1}$ follows directly from ergodicity of $(X_i)_{i=0}$. ■

The proof similar to the one given above may be found in Guihenneuc-Jouyaux and Robert (1998), but it is more complicated there than the straightforward reasoning presented in Lemma 1.

For simplicity of notation, we continue to call atoms or renewal sets $\mathcal{A}_j$ as *special sets*, keeping in mind different definitions of the secondary chain $(Y_i)_{i=1}$ for these two cases.

The moments ζ_i defined previously, may be additionally *partitioned* between corresponding special sets. Hence, we obtain the following definition of $\zeta_i^{(j)}$ for the fixed atom $\mathcal{A}_j$:

$$\zeta_1^{(j)} := \min\{i = 1, \ldots : X_i \in \mathcal{A}_j\} \ , \tag{16}$$

$$\zeta_{k+1}^{(j)} := \min\{i > \zeta_k^{(j)} : X_i \in \mathcal{A}_j\} \ . \tag{17}$$

For the renewal set $\mathcal{A}_j$ the definition of $\zeta_i^{(j)}$ is an equivalent change of the above formulas, i.e.:

$$\zeta_1^{(j)} := \min\{i = 1, \ldots : X_i \in \mathcal{A}_j \wedge U_i \leq \epsilon_{\mathcal{A}_j}\} \ , \tag{18}$$

$$\zeta_{k+1}^{(j)} := \min\{i > \zeta_k^{(j)} : X_i \in \mathcal{A}_j \wedge U_i \leq \epsilon_{\mathcal{A}_j}\} \ . \tag{19}$$

Therefore, $\zeta_1^{(j)}$ may be considered as the moment of *first visit* in the set $\mathcal{A}_j$.

LEMMA 2 *For the fixed $j = 1, 2$, the sums of the form $\sum_{k=\zeta_i^{(j)}+1}^{\zeta_{i+1}^{(j)}} X_k$ are conditionally iid in the stationary regime for $i = 1, \ldots$. The same applies for the sums $\sum_{k=\zeta_i+1}^{\zeta_{i+1}} X_k$ for $i = 1, 2, \ldots$.*

Proof. The variables $\zeta_k^{(j)}$ and ζ_k are stopping times. Therefore, the sequences $\left(X_{\zeta_i^{(j)}+1}, \ldots, X_{\zeta_{i+1}^{(j)}}\right)$ (or their equivalents for ζ_k) are conditionally iid in the stationary regime from the Strong Markov Property. Hence, the appropriate sums are also conditionally iid (for additional remarks see, e.g., Bremaud, 1999, Chapter 2.7). ■

3.2. Diagnosis of the initial chain

As we have noted in Section 3.1, for chain $(X_i)_{i=0}$ with two known special sets $\mathcal{A}_j$ $(j = 1, 2)$ we may introduce additional chain $(Y_i)_{i=1}$. The chain $(Y_i)_{i=1}$ is a discrete MC with only two states, regardless of cardinality and power of the space $\mathcal{X}$.

During diagnosis of the initial chain, we are interested in two values – n_{stat} and n_{Var}. The first value – n_{stat} – is the time moment when we are *close* enough to stationary distribution π_X, i.e.

$$\left\| P_{x_0}^{n_{\text{stat}}} - \pi_X \right\| \leq \varepsilon_1 \ , \tag{20}$$

where $\|.\|$ indicates some determined norm for space $\mathcal{X}$, e.g. total variation norm which is used in the rest of this paper, $\Pr_{x_0}^{n_{\text{stat}}}(.) = \Pr(X_{n_{\text{stat}}} = . \ | X_0 = x_0)$. When the number of simulations n_{stat} in the MCMC algorithm is attained, in the light of (20) we may treat $(X_i)_{i \geq n_{\text{stat}}}$ as being distributed *almost* from a *stationary distribution* π_X.

Suppose that we are interested in obtaining an estimator of the expected value $\mathbb{E}_{\pi_X} h(X)$ based on the average of the initial chain. It is easy to see that we would like to achieve sufficiently *small* variance of this estimator and find the quantity n_{Var} fulfilling the condition

$$\mathrm{Var}\left(\frac{1}{s}\sum_{k=n_{\mathrm{stat}}+1}^{n_{\mathrm{Var}}} h(X_k) - \mathbb{E}_{\pi_X} h(X)\right) \leq \varepsilon_2 \ , \tag{21}$$

where $s = n_{\mathrm{Var}} - n_{\mathrm{stat}}$.

We conclude observations concerning the problems given by (20) and (21) in the following lemmas and remarks.

3.3. Finding n_{stat} value

Let us start from the the classical case, when $\mathcal{X}$ has only two states.

LEMMA 3 *Suppose that* $\mathcal{X} = \{\mathcal{A}_1, \mathcal{A}_2\} = \{1, 2\}$. *Then, inequality*

$$\left\|\mathrm{Pr}_{x_0}^{n_{\mathrm{stat}}} - \pi_X\right\|_{\sup} \leq \varepsilon_1 \tag{22}$$

is fulfilled for

$$n_{stat} \geq \frac{\ln \frac{\varepsilon_1(\alpha+\beta)}{\min\{\alpha,\beta\}}}{\ln \gamma} \ , \tag{23}$$

where α *and* β *are derived for transition matrix of* $(X_i)_{i=0}$

$$\mathbb{P}_X = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \ , \tag{24}$$

and $\gamma = 1 - \alpha - \beta$.

Proof. In this case, it is known that stationary distribution is

$$\pi_X^{\mathrm{T}} = (\pi_X(1), \pi_X(2)) = \frac{1}{\alpha+\beta}\,(\beta, \alpha) \tag{25}$$

and the k–th step transition matrix is

$$\mathbb{P}_X^k = \begin{pmatrix} \pi_X(1) & \pi_X(2) \\ \pi_X(1) & \pi_X(2) \end{pmatrix} + \frac{\gamma^k}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \ . \tag{26}$$

If we start our chain in the state $\mathcal{A}_1 = 1$, then the k–th step probability will be

$$\left(\pi_Y(1) + \frac{\alpha\gamma^k}{\alpha+\beta}, \pi_Y(2) + \frac{-\alpha\gamma^k}{\alpha+\beta}\right) \ . \tag{27}$$

Hence, (22) is fulfilled for such n_{stat} that $\frac{\alpha\gamma^{n_{\text{stat}}}}{\alpha+\beta} \leq \varepsilon_1$. If we start the chain from the state $\mathcal{A}_2$, we obtain $\frac{\beta\gamma^{n_{\text{stat}}}}{\alpha+\beta} \leq \varepsilon_1$. Joining these two results and knowing that $\gamma < 1$, we establish (23). Another approach to this result with some faults (e.g. the chain considered there is not MC) may be found in Raftery and Lewis (1999). ■

Then, we can turn to a more general case, when $\mathcal{X}$ has more than only two states.

LEMMA 4 *Suppose that $\mathcal{X}$ is a finite space and $\mathcal{A}_1$ is a known atom for $\mathcal{X}$. Then*

$$\sum_{y\in\mathcal{X}} |\Pr_x^n(y) - \pi_X(y)| \leq 2\Pr_x(\zeta_1^{(1)} \geq n) + \sum_{j=0}^{n-1} \Pr_x(\zeta_1^{(1)} = j)\cdot$$
$$\left(\sum_{k=1}^{n-j-1} \left|\Pr_{\mathcal{A}_1}^k(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)\right| \Pr_{\mathcal{A}_1}(\zeta_1^{(1)} \geq n-k-j) + \right.$$
$$\left. + \pi_X(\mathcal{A}_1)\mathbb{E}_{\mathcal{A}_1}\left(\zeta_1^{(1)} - (n-j)\right)_+\right) . \quad (28)$$

Proof. Let us remind that $\zeta_1^{(1)}$ may be treated as the moment of the first visit in the set $\mathcal{A}_1$.

If we know the atom $\mathcal{A}_1$, then for any $y \in \mathcal{X}$ we have

$$\pi_X(y) = \pi_X(\mathcal{A}_1) \sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n = y, \zeta_1^{(1)} \geq n) , \quad (29)$$

where $\Pr_x(.)$, as usually, denotes $\Pr(.|X_0 = x)$. The proof of (29) may be found in Robert and Casella (2004, see Theorem 4.5.3).

We have

$$\Pr_x^n(y) = \Pr_x(X_n = y, \zeta_1^{(1)} \geq n) + \sum_{j=0}^{n-1} \Pr_x(X_j \in \mathcal{A}_1, \zeta_1^{(1)} = j)\cdot$$
$$\left(\sum_{k=0}^{n-j-1} \Pr_{\mathcal{A}_1}^k(\mathcal{A}_1)\Pr_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \geq n-k-j)\right) . \quad (30)$$

The notation $P_{\mathcal{A}_1}^k(\mathcal{A}_1)$ and $\Pr_{\mathcal{A}_1}(.)$ is validated because of the thesis of Lemma 1.

Using expansion (30) we have

$$|\Pr_x^n(y) - \pi_X(y)| \leq \Pr_x(X_n = y, \zeta_1^{(1)} \geq n) + \left|\sum_{j=0}^{n-1} \Pr_x(\zeta_1^{(1)} = j)\cdot\right.$$
$$\left.\left(\sum_{k=0}^{n-j-1} \Pr_{\mathcal{A}_1}^k(\mathcal{A}_1)\Pr_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \geq n-k-j\right) - \pi_X(y)\right| . \quad (31)$$

Hence

$$|\mathrm{Pr}_x^n(y) - \pi_X(y)| \le \mathrm{Pr}_x(X_n = y, \zeta_1^{(1)} \ge n) + \Bigg| \sum_{j=0}^{n-1} \mathrm{Pr}_x(\zeta_1^{(1)} = j) \cdot \Bigg(\sum_{k=0}^{n-j-1} \mathrm{Pr}_{\mathcal{A}_1}^k(\mathcal{A}_1) \mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \ge n-k-j) - \pi_X(y) \Bigg) - \pi_X(y) \sum_{j=n}^{\infty} \mathrm{Pr}_x(\zeta_1^{(1)} = j) \Bigg| \ . \quad (32)$$

From (29) for any $j \le n-1$ we have

$$\pi_X(y) = \pi_X(\mathcal{A}_1) \sum_{k=0}^{n-j-1} \mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \ge n-k-j) + + \pi_X(\mathcal{A}_1) \sum_{l=n-j+1}^{\infty} \mathrm{Pr}_{\mathcal{A}_1}(X_l = y, \zeta_1^{(1)} \ge l) \ . \quad (33)$$

After applying (33) to (32) we have

$$|\mathrm{Pr}_x^n(y) - \pi_X(y)| \le \mathrm{Pr}_x(X_n = y, \zeta_1^{(1)} \ge n) + \Bigg| \sum_{j=0}^{n-1} \mathrm{Pr}_x(\zeta_1^{(1)} = j) \cdot \Bigg(\sum_{k=0}^{n-j-1} \Big(\mathrm{Pr}_{\mathcal{A}_1}^k(\mathcal{A}_1) - \pi_X(\mathcal{A}_1) \Big) \mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \ge n-k-j) - \pi_X(\mathcal{A}_1) \sum_{l=n-j+1}^{\infty} \mathrm{Pr}_{\mathcal{A}_1}(X_l = y, \zeta_1^{(1)} \ge l) \Bigg) - \pi_X(y) \mathrm{Pr}_x(\zeta_1^{(1)} \ge n) \Bigg| \ . \quad (34)$$

Straightforwardly

$$|\mathrm{Pr}_x^n(y) - \pi_X(y)| \le \mathrm{Pr}_x(X_n = y, \zeta_1^{(1)} \ge n) + \sum_{j=0}^{n-1} \mathrm{Pr}_x(\zeta_1^{(1)} = j) \cdot \Bigg(\sum_{k=0}^{n-j-1} \Big| \mathrm{Pr}_{\mathcal{A}_1}^k(\mathcal{A}_1) - \pi_X(\mathcal{A}_1) \Big| \mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \ge n-k-j) + + \pi_X(\mathcal{A}_1) \sum_{l=n-j+1}^{\infty} \mathrm{Pr}_{\mathcal{A}_1}(X_l = y, \zeta_1^{(1)} \ge l) \Bigg) + \pi_X(y) \mathrm{Pr}_x(\zeta_1^{(1)} \ge n) \ , \quad (35)$$

which constitutes (28). ■

The equations (28) and (35) may be used to establish further dependencies between the initial and the secondary chain. Now we present a simple lemma, which may be helpful in the practice of MCMC setups.

LEMMA 5 *Suppose that $\mathcal{A}_1$ is a geometrically ergodic atom with constant M_1 and coefficient r_1, and there exist $M_2 > 0$, $r_2 > 1$, $M_3 > 0$, $r_3 > 1$ such that*

$$\Pr_{\mathcal{A}_1}(\zeta_1^{(1)} \geq n) \leq M_2 r_2^{-n} \ , \tag{36}$$

and

$$\Pr_x(\zeta_1^{(1)} = n) \leq M_3 r_3^{-n} \tag{37}$$

are fulfilled. Then, inequality

$$\sum_{y \in \mathcal{X}} |\Pr_x^n(y) - \pi_X(y)| \leq \varepsilon_1 \tag{38}$$

is satisfied for n given as the solution of formula

$$2\frac{M_3 r_3^{1-n}}{r_3 - 1} + \frac{M_2 M_3 r_3 (r_3^{-n} - r_2^{-n})}{(r_2 - 1)(r_2 - r_3)} + \\ + \frac{M_1 M_2 M_3}{(r_2 - r_1)} \left(\frac{r_1 r_3 (r_3^{-n} - r_1^{-n})}{(r_1 - r_3)} + \frac{r_2 r_3 (r_3^{-n} - r_2^{-n})}{(r_3 - r_2)} \right) \leq \varepsilon_1 \ . \tag{39}$$

Proof. After applying conditions (8), (36), (37) to inequality (28) we can straightforwardly prove (39). ∎

It is worth noting that it is possible to improve the inequality (39). If we know the value of stationary probability $\pi_X(\mathcal{A}_1)$, then we have a more detailed condition

$$2\frac{M_3 r_3^{1-n}}{r_3 - 1} + \frac{\pi_X(\mathcal{A}_1) M_2 M_3 r_3 (r_3^{-n} - r_2^{-n})}{(r_2 - 1)(r_2 - r_3)} + \\ + \frac{M_1 M_2 M_3}{(r_2 - r_1)} \left(\frac{r_1 r_3 (r_3^{-n} - r_1^{-n})}{(r_1 - r_3)} + \frac{r_2 r_3 (r_3^{-n} - r_2^{-n})}{(r_3 - r_2)} \right) \leq \varepsilon_1 \ . \tag{40}$$

3.4. Finding n_{Var} value

After establishing n_{stat} we may turn to finding evaluation for n_{Var}. To simplify the notation in the rest of this section we denote by $\zeta_1^{(j)}$ the first visit in set $\mathcal{A}_j$ for the chain $(X_k)_{k \geq n_{\mathrm{stat}}}$, by $\zeta_2^{(j)}$ – the second one, etc.

As previously, we start from the case of the two-state space $\mathcal{X}$.

LEMMA 6 *Suppose that* $\mathcal{X} = \{\mathcal{A}_1, \mathcal{A}_2\} = \{1, 2\}$ *and* $X_{n_{\text{stat}}} \sim \pi_X$. *Then condition* (21) *is fulfilled for any* $s = n - n_{\text{stat}}$ *given by inequality*

$$\frac{\alpha\beta(h(1) - h(2))^2}{(\alpha + \beta)^3} \left(\frac{2\gamma(\gamma^s - 1)}{s^2} + \frac{(\alpha + \beta)(1 - 2\gamma)}{s} \right) \leq \varepsilon_2 \ , \tag{41}$$

where $\gamma = 1 - \alpha - \beta$.

Proof. For simplicity of notation, as in Lemma 3 we suppose that state i denotes the atom $\mathcal{A}_i$.

We have

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{s} \sum_{k=n_{\text{stat}}+1}^{n} h(X_k) - \mathbb{E}_{\pi_X} h(X) \right) = \\
&= \frac{1}{s^2} \left(\operatorname{Var}\left(\sum_{k=n_{\text{stat}}+1}^{n-1} h(X_k) - (s-1)\mathbb{E}_{\pi_X} h(X) \right) + \right. \\
&+ 2 \operatorname{Cov}\left(\left(\sum_{k=n_{\text{stat}}+1}^{n-1} h(X_k) - (s-1)\mathbb{E}_{\pi_X} h(X) \right) (h(X_n) - \mathbb{E}_{\pi_X} h(X)) \right) + \\
&\left. + \operatorname{Var}\left(h(X_n) - \mathbb{E}_{\pi_X} h(X) \right) \right) \ . \qquad (42)
\end{aligned}$$

In this case we can write the covariance as

$$\begin{aligned}
&\operatorname{Cov}\left(\left(\sum_{k=n_{\text{stat}}+1}^{n-1} h(X_k) - (s-1)\mathbb{E}_{\pi_X} h(X) \right) (h(X_n) - \mathbb{E}_{\pi_X} h(X)) \right) = \\
&= \sum_{k=n_{\text{stat}}+1}^{n-1} \mathbb{E}\left((h(X_k) - \mathbb{E}_{\pi_X} h(X)) (h(X_n) - \mathbb{E}_{\pi_X} h(X)) \right) = \\
&= \sum_{k=1}^{s-1} \mathbb{E}\left((h(X_{n-k}) - \mathbb{E}_{\pi_X} h(X)) (h(X_n) - \mathbb{E}_{\pi_X} h(X)) \right) \ . \qquad (43)
\end{aligned}$$

From the assumption that $X_{n_{\text{stat}}} \sim \pi_X$, for $n \geq n_{\text{stat}}$ the variables X_n may be treated as derived from stationary distribution π_X. Therefore

$$\begin{aligned}
&\mathbb{E}\left((h(X_{n-k}) - \mathbb{E}_{\pi_X} h(X)) (h(X_n) - \mathbb{E}_{\pi_X} h(X)) \right) = \\
&= \mathbb{E}\left(h(X_{n-k}) h(X_n) \right) - \mathbb{E}^2_{\pi_X} h(X) \ . \qquad (44)
\end{aligned}$$

From properties (25) and (26) for the two-state MC we have

$$\begin{aligned}\mathbb{E}\left(h(X_{n-k})h(X_n)\right) - \mathbb{E}^2_{\pi_X} h(X) &= h^2(1)\pi_X(1)\Pr_1(X_k=1)+\\ &+ h(1)h(2)\pi_X(1)\Pr_1(X_k=2) + h(2)h(1)\pi_X(2)\Pr_2(X_k=1)+\\ &+ h^2(2)\pi_X(2)\Pr_2(X_k=2) - \left(h(1)\pi_X(1)+h(2)\pi_X(2)\right)^2 =\\ &= \frac{\alpha\beta}{(\alpha+\beta)^2}\gamma^k(h(1)-h(2))^2 \ . \quad (45)\end{aligned}$$

Applying (42) and simple recurrence to the above formula we obtain

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{s}\sum_{k=n_{\mathrm{stat}}+1}^{n} h(X_k) - \mathbb{E}_{\pi_X} h(X)\right) =\\ &= \frac{1}{s^2}\left(\mathrm{Var}\left(\sum_{k=n_{\mathrm{stat}}+1}^{n-1} X_k - (s-1)\mathbb{E}_{\pi_X} h(X)\right) +\right.\\ &\left.+2\frac{\alpha\beta}{(\alpha+\beta)^2}(h(1)-h(2))^2\gamma\frac{1-\gamma^{s-1}}{1-\gamma} + \mathbb{E}_{\pi_X} h^2(X) - \mathbb{E}^2_{\pi_X} h(X)\right) =\\ &= \frac{1}{s^2}\left(\frac{2\alpha\beta\gamma}{(\alpha+\beta)^3}(h(1)-h(2))^2\left(s-1-\gamma\frac{1-\gamma^{s-1}}{1-\gamma}\right)+\right.\\ &\left.\quad + s\left(h^2(1)\frac{\beta}{\alpha+\beta} + h^2(2)\frac{\alpha}{\alpha+\beta} - \left(h(1)\frac{\beta}{\alpha+\beta} + h(2)\frac{\alpha}{\alpha+\beta}\right)^2\right)\right) \ , \quad (46)\end{aligned}$$

which constitutes the thesis (41). ■

Now we turn to the general case, i.e. when $\mathcal{X}$ has more than only two states. In practice, an experimenter, after generation of some length of the chain, is interested in knowing whether an appropriately small error measured by (21) is attained. Therefore, it is possible in practice that an observer could choose such values of n_{stat} and n_{Var} that they are also moments of visits in special sets $\mathcal{A}_1$ and $\mathcal{A}_2$. This procedure is helpful in elimination of "tails", i.e. two fragments of chain: between n_{stat} and first visit in the special set, and last visit in the special set and n_{Var}. The estimation of these tails is very complicated in case of n_{Var} evaluation.

Let n_{stat} and n_{Var} be values preliminarily chosen by the experimenter. For these deterministic parameters, suppose that $n_{\mathrm{s}} \geq n_{\mathrm{stat}}$ and $n_{\mathrm{V}} \leq n_{\mathrm{Var}}$ are moments of visit in a special set $\mathcal{A}_j$, where n_{s} is first such moment, and n_{V} is the last one. Let

$$M_{(j)} = \#\{k : n_{\mathrm{s}} \leq \zeta_k^{(j)} \leq n_{\mathrm{V}}\} \ . \quad (47)$$

Obviously, $M_{(j)}$ is the random number of visits in $\mathcal{A}_j$ between n_{s} and n_{V}. For

the determined j we have

$$\frac{1}{n_{\mathrm{V}}-n_{\mathrm{s}}}\sum_{k=n_{\mathrm{s}}+1}^{n_{\mathrm{V}}}h(X_k)-\mathbb{E}_{\pi_x}h(X)=$$
$$=\frac{1}{n_{\mathrm{V}}-n_{\mathrm{s}}}\sum_{i=1}^{M_{(j)}-1}\left(\sum_{k=\zeta_i^{(j)}+1}^{\zeta_{i+1}^{(j)}}h(X_k)-\left(\zeta_{i+1}^{(j)}-\zeta_i^{(j)}\right)\mathbb{E}_{\pi_X}h(X)\right)\ , \quad (48)$$

which constitutes the following remark:

REMARK 1 *Suppose that $n_{\mathrm{s}}\geq n_{\mathrm{stat}}$ and $n_{\mathrm{V}}\leq n_{\mathrm{Var}}$ are moments of visits in the special set $\mathcal{A}_j$. Then*

$$\mathrm{Var}_{\pi_X}\left(\frac{1}{n_{\mathrm{V}}-n_{\mathrm{s}}}\sum_{k=n_{\mathrm{s}}+1}^{n_{\mathrm{V}}}h(X_k)-\mathbb{E}_{\pi_x}h(X)\right)=$$
$$=\mathrm{Var}_{\pi_X}\left(\frac{1}{n_{\mathrm{V}}-n_{\mathrm{s}}}\sum_{i=1}^{M_{(j)}-1}\left(\sum_{k=\zeta_i^{(j)}+1}^{\zeta_{i+1}^{(j)}}h(X_k)\right.\right.$$
$$\left.\left.-\left(\zeta_{i+1}^{(j)}-\zeta_i^{(j)}\right)\ \mathbb{E}_{\pi_X}h(X)\right)\right)\ . \quad (49)$$

In order to achieve appropriate evaluation of (49) we have to find variance estimator of single "fragment" in the trajectory. Let $S_i^{(j)}=\sum_{k=\zeta_i^{(j)}+1}^{\zeta_{i+1}^{(j)}}h(X_k)$. Then the value

$$\sigma_{(j)}^2=\mathrm{Var}_{\pi_X}\left(\sum_{k=\zeta_1^{(j)}+1}^{\zeta_2^{(j)}}h(X_k)-\left(\zeta_2^{(j)}-\zeta_1^{(j)}\right)\ \mathbb{E}_{\pi_X}h(X)\right) \quad (50)$$

may be estimated by the usual sum of squares estimator

$$\hat{\sigma}_{(j)}^2=\frac{1}{m_{(j)}-1}\sum_{i=1}^{m_{(j)}-1}\left(S_i^{(j)}-\frac{1}{m_{(j)}-1}\sum_{l=1}^{m_{(j)}-1}S_l^{(j)}\right)^2\ , \quad (51)$$

where $m_{(j)}$ is the number of visits in $\mathcal{A}_j$ (see Lemma 2). A similar estimator for the case of $M_{(j)}\to\infty$ was introduced in Robert (1995).

We could generalize our considerations for the case of more than only one special set. Let n_{s} and n_{V} be moments of visits in some special sets $\mathcal{A}_1$ and / or $\mathcal{A}_2$, not in only one determined $\mathcal{A}_j$. We use additional notation

$$T^{(j,l)}=(\zeta_2 \text{ and } Y_2\in\mathcal{A}_l)-(\zeta_1 \text{ and } Y_1\in\mathcal{A}_j)\ , \quad (52)$$

$$M_{(j,l)} = \#\{k : n_{\mathrm{s}} \leq Y_k \in \mathcal{A}_j, Y_{k+1} \in \mathcal{A}_l, \zeta_{k+1} \leq n_{\mathrm{V}}\} \ . \tag{53}$$

In such a case we have from Strong Markov Property

$$\frac{1}{n_{\mathrm{V}} - n_{\mathrm{s}}} \sum_{k=n_{\mathrm{s}}+1}^{n_{\mathrm{V}}} h(X_k) - \mathbb{E}_{\pi_x} h(X) =$$
$$= \frac{1}{n_{\mathrm{V}} - n_{\mathrm{s}}} \sum_{i=1}^{m-1} \left(\sum_{k=\zeta_i+1}^{\zeta_{i+1}} h(X_k) - (\zeta_{i+1} - \zeta_i)\, \mathbb{E}_{\pi_X} h(X) \right) \ . \tag{54}$$

In other words, we divide sequence $(X_i)_{i=n_{\mathrm{s}}}^{n_{\mathrm{V}}}$ into moments determined by ζ_i. In the right-hand side sum of (54) we can distinguish $M_{(1,1)}$ fragments which start and finish in $\mathcal{A}_1$, $M_{(1,2)}$ fragments which start in $\mathcal{A}_1$ and finish in $\mathcal{A}_2$, etc. Therefore we have the following remark:

REMARK 2 *Let $n_{\mathrm{s}} \geq n_{\mathrm{stat}}$ and $n_{\mathrm{V}} \leq n_{\mathrm{Var}}$ be moments of visits in special sets. Then*

$$\mathrm{Var}_{\pi_X} \left(\frac{1}{n_{\mathrm{V}} - n_{\mathrm{s}}} \sum_{k=n_{\mathrm{s}}+1}^{n_{\mathrm{V}}} h(X_k) - \mathbb{E}_{\pi_x} h(X) \right) =$$
$$= \mathrm{Var}_{\pi_X} \left(\frac{1}{n_{\mathrm{V}} - n_{\mathrm{s}}} \sum_{j,l=1}^{2} \sum_{i=1}^{M_{(j,l)}-1} \left(\sum_{k=\zeta_i^{(j)}+1 \ and \ Y_i \in \mathcal{A}_j}^{\zeta_{i+1}^{(l)} \ and \ Y_{i+1} \in \mathcal{A}_l} h(X_k) - \right. \right.$$
$$\left. \left. T^{(j,l)} \mathbb{E}_{\pi_X} h(X) \right) \right) \ . \tag{55}$$

As previously, we need appropriate variance estimator of the trajectory "fragment". Let

$$S_i^{(j,l)} = \sum_{k=\zeta_i+1 \text{ and } Y_i \in \mathcal{A}_j}^{\zeta_{i+1} \text{ and } Y_{i+1} \in \mathcal{A}_l} h(X_k) \ . \tag{56}$$

Then, variance

$$\sigma^2_{(j,l)} = \mathrm{Var}_{\pi_X} \left(\sum_{k=\zeta_1+1 \text{ and } Y_1 \in \mathcal{A}_j}^{\zeta_2 \text{ and } Y_2 \in \mathcal{A}_l} h(X_k) - T^{(j,l)} \mathbb{E}_{\pi_X} h(X) \right) \tag{57}$$

may be estimated by the usual sum of squares

$$\hat{\sigma}^2_{(j,l)} = \frac{1}{m_{(j,l)} - 1} \sum_{i=1}^{m_{(j,l)}-1} \left(S_i^{(j,l)} - \frac{1}{m_{(j,l)}} \sum_{k=1}^{m_{(j,l)}-1} S_k^{(j,l)} \right)^2 , \tag{58}$$

where $m_{(j,l)}$ is the number of transitions between special sets $\mathcal{A}_j$ and $\mathcal{A}_l$.

As previously stated, in the method presented we would like to eliminate the problems caused by "tails", which are very hard to estimate for the experimenter. Therefore, based on Remark 1 and formula (51), we could postulate that condition (21) is fulfilled for any $n \geq n_{\mathrm{Var}}$ if

$$n_{\mathrm{stat}} + \hat{\sigma}_{(j)} \sqrt{\frac{m_{(j)} - 1}{\varepsilon_2}} \leq n_{\mathrm{Var}} \ . \tag{59}$$

And from Remark 2 and estimator (58) we could postulate the generalization of condition (59). Then, (21) is fulfilled for any $n \geq n_{\mathrm{Var}}$ if

$$n_{\mathrm{stat}} + \sqrt{\frac{m_{(1,1)}\hat{\sigma}^2_{(1,1)} + m_{(1,2)}\hat{\sigma}^2_{(1,2)} + m_{(2,1)}\hat{\sigma}^2_{(2,1)} + m_{(2,2)}\hat{\sigma}^2_{(2,2)}}{\varepsilon_2}} \leq n_{\mathrm{Var}} \ . \tag{60}$$

We have to note that the presented considerations are based on theoretical foundations, but also include strong heuristic "flavour". Our initial remarks use random variables n_{s} and n_{V}, which are also moments of visits in special sets. In (59) and (60) we "overinterpreted" these results as if $s = n_{\mathrm{Var}} - n_{\mathrm{stat}}$ were a deterministic variable and both n_{stat} and n_{Var} are moments of visits in special sets. Because of this, it is possible to formulate inequalities for meeting the condition (21) in a relatively simple way. But the lack of direct connection between (59), (60) and the previous results is a disadvantage of such approach. However, in practice it is always possible to start or finish observations of trajectory in appropriate moments, i.e. visits in the special sets.

Results for an estimator similar to (58) may be found in Robert (1995). But in this reference only the asymptotic features of distances between the values

$$\frac{m_{(j)}\hat{\sigma}^2_{(j)}}{n} \tag{61}$$

for $n \to \infty$ and various $\mathcal{A}_j$ are used.

In this paper we use non-asymptotic features and we show direct connection between the discussed method and condition (21), which is the basis for MCMC methodology.

4. Example of application

After introducing the methodology appropriate for finding values n_{stat} and n_{Var}, we present now the example of their application. For simplicity of notation and in order to derive conclusions, we use state space $\mathcal{X}$ with a few atoms.

We should emphasize that the solutions established in lemmas in Section 3 give *exact* (i.e. demonstrated by mathematical reasoning, not only *heuristic* approach) and *precise* (i.e. *non-asymptotic*) values. Therefore, we may focus only on the problem of transition of the acquired results from theoretical formulas to the practical example.

Let us suppose that we are interested in finding the value $\mathbb{E}_f h(X)$, where $f(.)$ describes the state space $\mathcal{X}$ with eight atoms and stationary probabilities

$$f(.) = (1/20, 1/20, 2/20, 2/20, 3/20, 3/20, 4/20, 4/20) \,, \tag{62}$$

i.e. first atom has stationary probability 1/20, the second one – 1/20, etc., and $h(.)$ is a uniform function on $\mathcal{X}$, i.e.

$$h(.) = (1, 1, 1, 1, 1, 1, 1, 1) \,. \tag{63}$$

Because of this special form of the function $h(.)$, all the states of space $\mathcal{X}$ have the same "weight" and "importance" during MCMC simulations.

In order to calculate $\mathbb{E}_f h(X)$ we use *independent Metropolis-Hastings algorithm* (see e.g. Robert and Casella, 2004). Our main trajectory has million elements and is initiated from state one. We also assume that $\mathcal{A}_1 = 3$ and $\mathcal{A}_2 = 7$. Therefore, we may compare values n_{stat} and n_{Var} on the basis of states with various stationary probabilities.

Firstly we would like to find n_{stat}. To apply lemmas from Section 3.3, we have to evaluate necessary parameters $r_1, M_1, r_2, M_2, r_3, M_3$ (see assumptions for Lemma 5). Normally, experimenter may have some additional knowledge about these values, but we use additional simulations in order to determine $r_1, M_1, r_2, M_2, r_3, M_3$. Hence, we generate additional sets of 50000 trajectories with 100 steps in each trajectory and appropriate starting points – states one, three and seven. Then, we apply the "pessimistic optimization" approach.

Namely, if we suppose that for the optimal parameters r_1 and M_1 we have

$$|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)| \approx M_1 r_1^{-n} \,, \tag{64}$$

then

$$\frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \approx r_1^{-n+1} \,. \tag{65}$$

Therefore, we can find a "pessimistic" evaluation of $\hat{r}_1$ in the sense of satisfying the condition

$$\hat{r}_1 = \min_{r \in \mathbb{R}_+} \left\{ \forall n = 2, 3, \ldots : r^{-n+1} - \frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \geq 0 \right\} \,. \tag{66}$$

As easily seen, (66) gives us the "maximal pessimistic" guess for $\hat{r}_1$, because in this light $\hat{r}_1$ has to be the upper limit for all steps in strictly *deterministic* sense. In case of any numerical errors or even for greater values for n (note exponential decrease in conditions for Lemma 5), this method may give larger values of $\hat{r}_1$ than they are in reality. However, other methods, like meeting the weaker condition

$$r^{-n+1} - \frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \geq 0 \quad \vee \quad \left| r^{-n+1} - \frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \right| \leq \delta \tag{67}$$

for some small enough δ, may be easily criticized because of unknown error generated by the selection of the value of δ.

After fixing the value $\hat{r}_1$, like in (66), we may find $\hat{M}_1$ in the same manner, as satisfying the condition

$$\hat{M}_1 = \min_{M \in \mathbb{R}_+} \left\{ \forall n = 1, 2, \ldots : M \hat{r}_1^{-n} - |\Pr^n_{\mathcal{A}_1}(y) - \pi_X(\mathcal{A}_1)| \geq 0 \right\} . \tag{68}$$

The analogous formulas may be derived for parameters r_2, M_2, r_3, M_3.

Then, from the "pessimistic optimization" for $\mathcal{A}_1$ we have

$$\hat{r}_1 = 1.04 \ , \ \hat{M}_1 = 0.0268 \ , \hat{r}_2 = 1.0941 \ , \ \hat{M}_2 = 1.0888 \ ,$$
$$\hat{r}_3 = 1.0904 \ , \ \hat{M}_3 = 0.1372 \ . \tag{69}$$

We can substitute these values into the formula (39) in order to find the number of steps n_{stat} for the given value ε_1 (see Table 1). In this table, the column "True value of ε_1" give the exact value of the left-hand side for (39) and the determined number of steps n_{stat} from the second column. The graph of the left-hand side of (39) as a function of steps number n is shown in Fig. 1.

Table 1. Evaluation of n_{stat} for the third state

Assumed value of ε_1	Number of steps n_{stat}	True value of ε_1
0.1	90	0.0978145
0.02	120	0.0196767
0.01	135	0.00974242
0.001	190	0.000981598

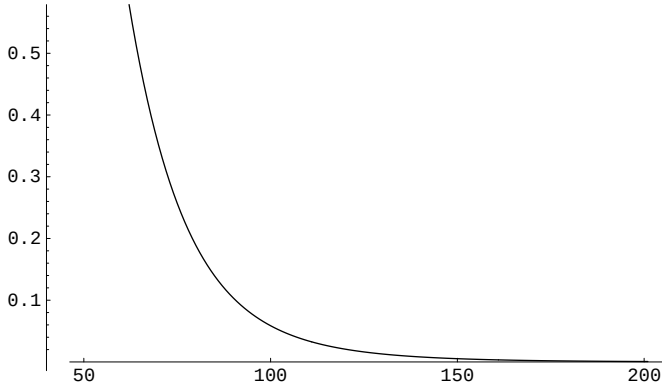


Figure 1. Error level ε_1 as a function of n for the third state

If we use the improved inequality (40) instead of (39), we may observe the reduction of the necessary number of steps n_{stat}, especially for larger ε_1 (see Table 2). This phenomenon is even easier to trace in Fig. 2, where curve is much steeper at the beginning of the graph.

Table 2. Evaluation of n_{stat} for the third state based on inequality (40)

Assumed value of ε_1	Number of steps n_{stat}	True value of ε_1
0.1	75	0.0981865
0.02	114	0.0195048
0.01	131	0.00989127
0.001	190	0.000967164

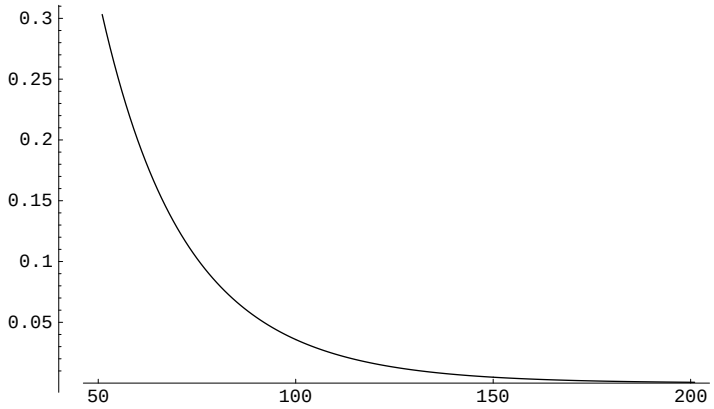


Figure 2. Error level ε_1 as a function of n for the third state based on inequality (40)

We may perform the same analysis for the seventh state, i.e. the special set $\mathcal{A}_2$. In this case the necessary parameters may be evaluated as

$$\hat{r}_1 = 1.0438\ ,\ \hat{M}_1 = 0.0793\ ,\ \hat{r}_2 = 1.14385\ ,\ \hat{M}_2 = 1.1439\ ,$$
$$\hat{r}_3 = 1.1231\ ,\ \hat{M}_3 = 0.1394\ . \quad (70)$$

Because the atom $\mathcal{A}_2$ has higher stationary probability than $\mathcal{A}_1$, we obtain lesser numbers of steps n_{stat} (see Table 3 and Fig. 3).

We can also apply improved inequality (40) for the set $\mathcal{A}_2$, but due to the faster exponential convergence, guaranteed by higher values of $\hat{r}_i$, the profit of reduction of n_{stat} value is not so visible as in the previous case.

Table 3. Evaluation of n_{stat} for the seventh state

Assumed value of ε_1	Number of steps n_{stat}	True value ofε_1
0.1	71	0.0992184
0.02	107	0.0192124
0.01	123	0.00961369
0.001	176	0.000988225

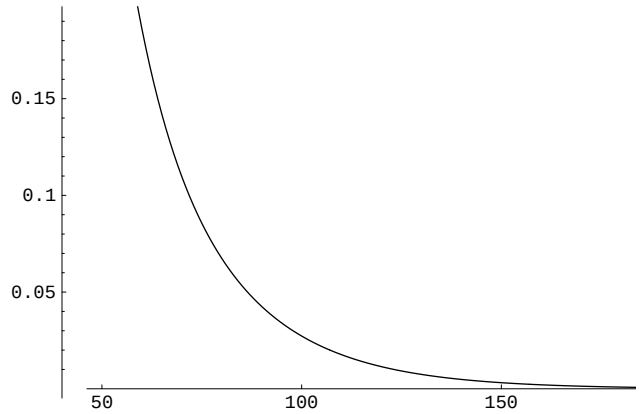


Figure 3. Error level ε_1 as a function of n for the seventh state

Now we turn to finding the value of n_{Var}. Therefore, we use the methodology described in Section 3.4. Let us start from the conclusions based on inequality (59). The solutions derived for $\mathcal{A}_2$ may be found in Table 4. In the second column there is the appropriate number of steps n_{stat} calculated for $\epsilon_1 = \epsilon_2$ (compare with Table 1). Because, according to Remark 1, the beginning and the ending moments should be the visits in $\mathcal{A}_1$, the obtained values are put in the third and fourth column. In order to minimize the possible influence of n_{stat} on n_{Var} evaluation, we always multiply the necessary number of steps n_{stat} by two before further analysis. In the last column, the theoretical length of necessary trajectory

$$\sigma_{(1)}\sqrt{\frac{m_{(1)}-1}{\varepsilon_2}} \tag{71}$$

(compare with (59)) is calculated.

The same calculation could be done for $\mathcal{A}_2$ (see Table 5).

Table 4. Evaluation of n_{Var} for the third state

Assumed value of ε_2	Number of steps n_{stat}	Start	Stop	Theoretical length
0.1	90	180	306	125.294
0.02	120	278	858	575.985
0.01	135	278	1282	995.365
0.001	190	388	13226	12835.7

Table 5. Evaluation of n_{Var} for the seventh state

Assumed value of ε_2	Number of steps n_{stat}	Start	Stop	Theoretical length
0.1	71	144	222	75.3279
0.02	107	222	622	398.957
0.01	123	246	1074	823.829
0.001	176	360	8886	8525.97

As it was observed during simulations, the value $\sigma_{(j)}$ may not be correctly estimated. This can be seen, e.g., in Fig. 4, where the value of $\hat{\sigma}_{(2)}$ as a function of steps n for the seventh state is plotted.

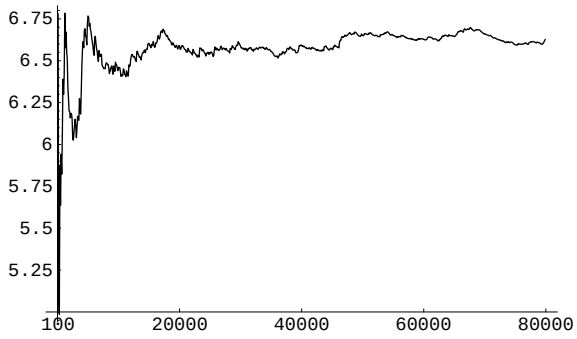


Figure 4. Value of $\hat{\sigma}_{(j)}$ as a function of n

Because of this, we may calculate the estimator $\hat{\sigma}_{(j)}$ based on a larger part of the main trajectory (or even the whole of it) and then put the estimated value into the inequality (59). For example, if we do this for the whole one million steps trajectory, we obtain new evaluations for n_{Var}. These values for the seventh state are shown in Table 6.

Table 6. Evaluation of n_{Var} for the seventh state with corrected value of $\hat{\sigma}_{(2)}$

Assumed value of ε_2	Number of steps n_{stat}	Start	Stop	Theoretical length
0.1	71	144	230	80.8683
0.02	107	222	727	500.687
0.01	123	246	1139	878.454
0.001	176	360	9260	8861.14

If we compare Tables 5 and 6, we can note that values for n_{Var} are $10-20\%$ bigger if we take into account the whole trajectory for the purpose of estimation of $\sigma_{(2)}$.

Next, we may apply the inequality (60) based Remark 2. As previously, in order to minimize the bias of estimators, we multiply by two the maximum values from Tables 1 and 3 for the fixed $\varepsilon_2 = \varepsilon_1$. This gives us the starting values for evaluation of n_{Var}.

The results derived for both sets $\mathcal{A}_1$ and $\mathcal{A}_2$ simultaneously are given in Table 7.

Table 7. Evaluation of n_{Var} for both atoms simultaneously

Assumed value of ε_2	Number of steps n_{stat}	Start	Stop	Theoretical length
0.1	90	180	222	41.5832
0.02	120	244	323	78.518
0.01	135	274	522	245.208
0.001	190	382	3783	3400.62

If we compare outcomes from Table 7 and the previous ones, we can easily see that taking into account both of the atoms dramatically reduces the necessary number of steps n_{Var}. This can be explained by additional information provided by considering two sets instead of only one.

As previously, random behaviour of $\hat{\sigma}_{(j,l)}$ may be observed. Therefore, we also calculated the estimators $\hat{\sigma}_{(j,l)}$ for the whole trajectory. The results for this case may be found in Table 8. The evaluations of n_{Var} reveal a bigger value, similarly as in the case for only one special set.

It is worth noting that despite the simple structure of the state space $\mathcal{X}$, the presented application has deep connections with more complex problems, e.g. similar atom state space may be found in *analysis and restoration of images degraded by noise* (see, e.g., Koronacki et al., 2005, Lasota, Niemiro 2003).

Table 8. Evaluation of n_{Var} for both atoms simultaneously with corrected value of $\hat{\sigma}_{(j,l)}$

Assumed value of ε_2	Number of steps n_{stat}	Start	Stop	Theoretical length
0.1	90	180	230	46.7273
0.02	120	244	502	252.75
0.01	135	274	751	469.959
0.001	190	382	4571	4181.49

5. Concluding remarks

We start from formulating two inequalities which correspond to standard questions in MCMC setups, i.e. *when the sampled transition probability is close to the determined stationary probability of Markov Chain?* and *how many iterations should be used in order to minimize the error of estimator?* These inequalities correspond to finding two values – number of steps n_{stat} and n_{Var} for the trajectory generated by some MCMC method. Then we use the features of secondary chain in order to appoint these values in a series of lemmas. Thereby we obtain a useful set of conditions which could be used for checking the convergence in the MCMC setup. The example of application of theoretical lemmas and reasoning based on them for the case of state space with atoms is also provided. It should be mentioned that this paper comprises some results from the Ph.D. dissertation (see Romaniuk, 2007), where additional remarks may be found.

We should emphasize the usefulness of the presented method, which could be used in a highly automated manner and provide strict results for the experimenter. However, we should note that not just one, but a whole set of various algorithms and methods should be applied in order to control the MCMC output and guarantee the convergence of the simulated trajectory at a satisfactory level.

The possibilities of complementing the discussed method might also be considered. For example, the conditions obtained might be improved, like in (40). However, additional information about the structure of state space or underlying Markov Chain may be necessary in such a case. The dependencies among the number of special sets, their allocation, possible modes in state space and obtained solutions may be examined. The lemmas may be also generalized for other cases of state space structure and numbers of special sets.

References

Asmussen, S. (1979) *Applied Probability and Queues.* J. Wiley, New York.
Athreya, K.B. and Ney, P. (1978) A new approach to the limit theory of

recurrent Markov chains. *Trans. Amer. Math. Soc.* **245**, 493–501.

Boos, D. and Zhang, J. (2000) Monte Carlo Evaluation of Resampling – Based Hypothesis Tests. *Journal of the American Statistical Association* **95**, No. 450.

Booth, J.G. and Sarkar, S. (1998) Monte Carlo Approximation of Bootstrap Variances. *The American Statistician* **52**, 4.

Bremaud, P. (1999) *Markov Chains — Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer Verlag, New York.

Brooks, S.P. and Roberts, G.O. (1998) Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing* **8**, 319–335.

Cox, D.R. and Miller, H.D. (1965) *The Theory of Stochastic Processes*. Chapman and Hall, London.

Doucet, A., Godsill, S. and Andrieu, Ch. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* **10**.

Fishman, G.S. (1996) *Monte Carlo — Concepts, Algorithms and Applications*. Springer Verlag, New York.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990) Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association* **85**, 412.

Geyer, C.J. (1992) Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7**, 473–511.

Gilks, W.R., Richardson, S. and Spiegelhalter, D. J. (1997) *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Guihenneuc-Jouyaux, Ch. and Robert, Ch.P. (1998) Discretization of Continuous Markov Chains and Markov Chain Monte Carlo Convergence Assessment. *Jour. of American Stat. Assoc.* **93**, 443.

Iosifescu, M. (1980) *Finite Markov Processes and Their Applications*. Wiley, New York.

Kass, R.E., Carlin, B.P., Gelman, A. and Neal, R.M. (1998) Markov Chain Monte Carlo in Practice: A Roundtable Discussion *The American Statistician* **52**, 2.

Kipnis, C. and Varadhan, S.R. (1986) Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions *Comm. Math. Phys.* **104**, 1–19.

Koronacki, J., Lasota, S. and Niemiro, W. (2005) Positron emission tomography by Markov chain Monte Carlo with auxiliary variables. *Pattern Recognition* **38**, 241–250.

Lasota, S. and Niemiro, W. (2003) A version of the Swendsen-Wand algorithm for restoration of images degraded by Poisson noise. *Pattern Recognition* 36, 931 – 941.

Li, S., Pearl, D.K. and Doss, H. (2000) Phylogenetic Tree Construction Using Markov Chain Monte Carlo. *Journal of the American Statistical Association* **95**, 450

MEHTA, C.R., PATEL, N.R. and SENCHAUDHURI, P. (2000) Efficient Monte Carlo Methods for Conditional Logistic Regression. *Journal of the American Statistical Association* **95**, 449

MENGERSEN, K.L., ROBERT, CH.P. and GUIHENNEUC–JOUYAUX, CH. (1999) MCMC Convergence Diagnostics: A Review. In: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith, eds., *Bayesian Statistics 6*, Oxford University Press, 415–440.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. and TELLER, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**.

METROPOLIS, N. and ULAM, S. (1949) The Monte Carlo Method. *Journal of American Statistical Association* **44**.

MYKLAND, P., TIERNEY, L. and YU, B. (1995) Regeneration in Markov Chain Samplers. *JASA* **90**, 233 – 241.

NUMMELIN, E. (1978) A splitting technique for Harris recurrent Markov Chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwändte Gebiete* **43**, 309 – 318.

RAFTERY, A.E. and LEWIS, S.M. (1999) How many iterations in the Gibbs Sampler? In: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., *Bayesian Statistics 4*. Oxford University Press, 763–773.

ROBERT, CH.P. (1995) Convergence Control Methods for Markov Chain Monte Carlo Algorithm. *Statistical Science* **10**, 3.

ROBERT, CH.P. and CASELLA, G. (2004) *Monte Carlo Statistical Methods*, 2nd ed., Springer Verlag, New York.

ROMANIUK, M. (2003) Pricing the Risk-Transfer Financial Instruments via Monte Carlo Methods. *Systems Analysis Modelling Simulation* **43**, 8, 1043 - 1064.

ROMANIUK, M. (2007) *Application of renewal sets for convergence diagnosis of MCMC setups* (in Polish). Ph.D. dissertation, Systems Research Institute, Polish Academy of Sciences.