Raport Badawczy
Research Report

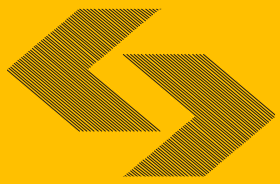
RB/69/2008

Convergence diagnosis to stationary distribution in MCMC methods via atoms and renewal sets

M. Romaniuk

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences



Control and Cybernetics

vol. **37** (2008) No. 1

Convergence diagnosis to stationary distribution in MCMC methods 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 one of the best known methods for conducting computer simulations useful in such areas as statistics, physics, biology, etc. However, to obtain appropriate solutions, the additional convergence diagnosis must be applied for Markov Chain trajectory generated by the algorithm. 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. In this paper we focus on finding the moment when the simulated chain is close enough to the stationary distribution of the Markov chain. The discussed method has some appealing properties, like high degree of diagnosis automation. Apart from theoretical lemmas and a more heuristic approach, the examples of application are 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, the researchers could build more complex, more "real-life" models. The same applies for 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 but brilliant idea. In order to find the expected value $\mathbb{E}_{\pi_X} h(X)$ for some function $h(.) : X \to \mathbb{R}^p$ and probability distribution $\pi_X(.)$, we could generate Markov Chain $X_0, X_1, X_2, \ldots$ with π_X as the stationary distribution. The convergence of the estimator, derived from the simulated sample is guaranteed by the ergodic theorems (see, e.g., Robert, Casella, 2004 for additional details). Therefore, we do not have to generate values directly from $\pi_X(.)$ as in the MC method, but we may use more general algorithms like Gibbs sampler or the Metropolis–Hastings algorithm.

Yet, 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 starting point x_0. The second one is related to finding the number of steps n_{Var}, when the estimator of $\mathbb{E}_{\pi_X} h(X)$, derived from the sample $X_{n_{\text{stat}}+1}, X_{n_{\text{stat}}+2}, \ldots, X_{n_{\text{Var}}}$ has *error* small enough, as measured e.g. by variance. These two questions are covered by *convergence diagnosis* and are one of the main issues in MCMC simulations. However, in this paper we focus only on the first problem, i.e. finding the value n_{stat}. Some answers for the second problem may be found e.g. in Romaniuk (2007b).

There is a lot of various convergence diagnosis methods (see, e.g., Robert, Casella, 2004; El Adlouni et al., 2006, 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, very often these methods make use of different *features* of the underlying Markov Chains, e.g. specific probability structure of the state space. 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 *heuristic* and *theoretical* (i.e. based on mathematical proofs) methods. Therefore, each new convergence diagnosis method may be seen as an additional tool for experimenters, which gives them a new possibility to check the obtained simulations.

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

The methods described cover both theoretical and heuristic approaches.

The presented theoretical method 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 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 lemmas may be used in a *highly automated manner*. This gives the possibility for preparing general diagnosis algorithms for a wide class of MCMC

problems.

The heuristic approach is also based on mathematical lemma, but involves subjective graph checking.

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. Next, in Section 3.3 we present some theoretical lemmas which constitute the foundation of the introduced method and provide the answers for the question about n_{stat}. In Section 3.4 we discuss a more heuristic approach. In Section 4 we present how the derived results may be applied in two examples. The concluding remarks are contained in Section 5.

Some of the solutions presented in this paper are based on ideas from Romaniuk (2007a and 2007b). As it was mentioned before, in Romaniuk (2007b) the methods for finding both values n_{stat} and n_{Var} were presented. However, for n_{stat} appropriate lemmas only in atom case were proved. In this paper we focus only on the problem of n_{stat} value, but the generalized lemmas for the case of renewal sets are added. Additionally, a new heuristic approach for both – atom and renewal – cases is presented.

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 $\mathbb{B}(\mathcal{X})$ is 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 as *discrete MC*, and in the second – as 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 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 exists 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 convenient 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 $(X_i)_{i=0}$. In view of (5) and (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.

LEMMA 1. *Let* $(X_i)_{i=0}$ *be Harris recurrent Markov Chain and*

$$\mathbb{E}_{\pi_X} |f(X)| = \int_{\mathcal{X}} |f(x)| d\pi_X(x) < \infty \tag{9}$$

for some function $f(.) : \mathcal{X} \to \mathbb{R}$ *and*

$$\mathbb{E}_{\pi_X} |l(X)| = \int_{\mathcal{X}} |l(x)| d\pi_X(x) < \infty \ , \ \mathbb{E}_{\pi_X} l(X) \neq 0 \tag{10}$$

for some function $l(.) : \mathcal{X} \to \mathbb{R}$. *Then we have*

$$\frac{\frac{1}{n+1} \sum_{k=0}^{n} f(X_k)}{\frac{1}{n+1} \sum_{k=0}^{n} l(X_k)} \xrightarrow[n \to \infty]{p.n.} \frac{\int_{\mathcal{X}} f(x) \, d\pi_X(x)}{\int_{\mathcal{X}} l(x) \, d\pi_X(x)} \ . \tag{11}$$

For proof of this lemma see Robert and Casella (2004).

3. Proposal of a convergence diagnosis method

In this section we present a convergence diagnosis method for MCMC output. This proposal uses notions of *atoms* and *renewal sets* (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 distribution 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

$$\zeta_1 := \min\{i = 1, \ldots : X_i \in \mathcal{A}_1 \cup \mathcal{A}_2\} \ , \tag{12}$$

$$\zeta_{k+1} := \min\{i > \zeta_k : X_i \in \mathcal{A}_1 \cup \mathcal{A}_2\} \ , \tag{13}$$

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

It is seen that the chain $(Y_i)_{i=1}$ has Markov Property for the *truncated* space $\mathcal{Y}' := \{\mathcal{A}_1, \mathcal{A}_2\}$ — see Lemma 2.

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

$$\zeta_1 := \min\{i = 1, \ldots : (X_i \in \mathcal{A}_1 \wedge U_i \leq \epsilon_{\mathcal{A}_1}) \vee (X_i \in \mathcal{A}_2 \wedge U_i \leq \epsilon_{\mathcal{A}_2})\} \ , \tag{15}$$

$$\zeta_{k+1} := \min\{i > \zeta_k : (X_i \in \mathcal{A}_1 \wedge U_i \leq \epsilon_{\mathcal{A}_1}) \vee (X_i \in \mathcal{A}_2 \wedge U_i \leq \epsilon_{\mathcal{A}_2})\} \ , \tag{16}$$

$$Y_k = X_{\zeta_k} \ , \tag{17}$$

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 $\mathcal{Y}'$. As it was mentioned before, a special method to simulate from the "remainder" kernel may be necessary (see Mykland, Tierney and Yu, 1995).

We may summarise previous observations in a simple lemma:

LEMMA 2. *If $\mathcal{A}_1, \mathcal{A}_2$ are atoms (or renewal sets), the chain $(Y_i)_{i=1}$ defined by conditions (12) — (14) (or (15) — (17), respectively) is a Markov Chain for the space $\mathcal{Y}' := \{\mathcal{A}_1, \mathcal{A}_2\}$. This chain is ergodic.*

The proof may be found in Romaniuk (2007b).

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 both cases.

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

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

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

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

$$\zeta_1^{(j)} := \min\{i = 1, \ldots : X_i \in \mathcal{A}_j \wedge U_i \leq \epsilon_{\mathcal{A}_j}\} \ , \tag{20}$$

$$\zeta_{k+1}^{(j)} := \min\{i > \zeta_k^{(j)} : X_i \in \mathcal{A}_j \wedge U_i \leq \epsilon_{\mathcal{A}_j}\} \ . \tag{21}$$

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

Next lemma is used as justification for a heuristic method described further.

LEMMA 3. *If sets $\mathcal{A}_1, \mathcal{A}_2$ are atoms, then stationary distribution of $\pi_Y(.)$ is given by*

$$\pi_Y(\mathcal{A}_j) = \frac{\int_{x \in \mathcal{A}_j} d\pi_X(x)}{\int_{x \in \mathcal{A}_1} d\pi_X(x) + \int_{x \in \mathcal{A}_2} d\pi_X(x)} \ , \tag{22}$$

for $j = 1, 2$.

If sets $\mathcal{A}_1, \mathcal{A}_2$ *are renewal sets, then stationary distribution of* $\pi_Y(.)$ *is given by*

$$\pi_Y(\mathcal{A}_j) = \frac{\epsilon_{\mathcal{A}_j} \int_{x \in \mathcal{A}_j} d\pi_X(x)}{\epsilon_{\mathcal{A}_1} \int_{x \in \mathcal{A}_1} d\pi_X(x) + \epsilon_{\mathcal{A}_2} \int_{x \in \mathcal{A}_2} d\pi_X(x)} \ , \tag{23}$$

for $j = 1, 2$.

Proof. Because $(Y_i)_{i=1}$ is MC, then from the strong ergodic theorem for Markov chains we have

$$\frac{\sum_{i=1}^{m} \mathbb{1}(Y_i \in \mathcal{A}_j)}{m} \xrightarrow[m \to \infty]{p.n.} \pi_Y(\mathcal{A}_j) \ , \tag{24}$$

for $j = 1, 2$, where

$$m = \#\{i \leq n : X_i \in \mathcal{A}_1 \cup \mathcal{A}_2\} \ . \tag{25}$$

If $\mathcal{A}_1, \mathcal{A}_2$ are atoms, then let

$$m(n) = \#\{i \leq n : X_i \in \mathcal{A}_1 \cup \mathcal{A}_2\} \ , \tag{26}$$

i.e. $m(n)$ is the random number of visits into $\mathcal{A}_1$ and $\mathcal{A}_2$. Because the initial chain is Harris recurrent, then for $n \to \infty$, we have $m(n) \to \infty$ (see Nummelin, 2001).

From (12) – (14) and Lemma 1 we have

$$\begin{aligned} \frac{\sum_{i=1}^{m(n)} \mathbb{1}(Y_i \in \mathcal{A}_j)}{m(n)} &= \frac{\sum_{i=1}^{m(n)} \mathbb{1}(Y_i \in \mathcal{A}_j)}{\sum_{i=1}^{m(n)} \mathbb{1}(Y_i \in \mathcal{A}_1 \cup \mathcal{A}_2)} = \\ &= \frac{\frac{1}{n+1} \sum_{i=0}^{n} \mathbb{1}(X_i \in \mathcal{A}_j)}{\frac{1}{n+1} \sum_{i=0}^{n} \mathbb{1}(X_i \in \mathcal{A}_1 \cup \mathcal{A}_2)} \xrightarrow[n \to \infty]{p.n.} \frac{\int_{x \in \mathcal{A}_j} d\pi_X(x)}{\int_{x \in \mathcal{A}_1} d\pi_X(x) + \int_{x \in \mathcal{A}_2} d\pi_X(x)} \ . \end{aligned} \tag{27}$$

Comparing (24) with (27), we obtain (22) (see also Nummelin, 2001 for similar inference).

If $\mathcal{A}_1, \mathcal{A}_2$ are renewal sets, then let

$$m(n) = \#\{i \leq n : X_i \in (\mathcal{A}_1, U_i \leq \epsilon_{\mathcal{A}_1}) \cup (\mathcal{A}_2, U_i \leq \epsilon_{\mathcal{A}_2})\} \ . \tag{28}$$

From (15) – (17) and Lemma 1 we have

$$\begin{aligned} \frac{\sum_{i=1}^{m(n)} \mathbb{1}(Y_i \in \mathcal{A}_j)}{m(n)} &= \frac{\sum_{i=1}^{m(n)} \mathbb{1}(Y_i \in \mathcal{A}_j)}{\sum_{i=1}^{m(n)} \mathbb{1}(Y_i \in \mathcal{A}_1 \cup \mathcal{A}_2)} = \\ &= \frac{\frac{1}{n+1} \sum_{i=0}^{n} \mathbb{1}(X_i \in \mathcal{A}_j, U_i \leq \epsilon_{\mathcal{A}_j})}{\frac{1}{n+1} \sum_{i=0}^{n} \mathbb{1}(X_i \in (\mathcal{A}_1, U_i \leq \epsilon_{\mathcal{A}_1}) \cup (\mathcal{A}_2, U_i \leq \epsilon_{\mathcal{A}_2}))} \to \\ &\xrightarrow[n \to \infty]{p.n.} \frac{\epsilon_{\mathcal{A}_j} \int_{x \in \mathcal{A}_j} d\pi_X(x)}{\epsilon_{\mathcal{A}_1} \int_{x \in \mathcal{A}_1} d\pi_X(x) + \epsilon_{\mathcal{A}_2} \int_{x \in \mathcal{A}_2} d\pi_X(x)} \ . \end{aligned} \tag{29}$$

In formula (29) we used the independence property for U_i and X_i (see (7)). As previously, comparing (24) with (29), we prove (23).

■

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.

$$\left\| P_{x_0}^{n_{\text{stat}}} - \pi_X \right\| \leq \varepsilon_1 \ , \tag{30}$$

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 achieved, in the light of (30) we may treat $(X_i)_{i \geq n_{\text{stat}}}$ as being *almost* distributed from *stationary distribution* π_X.

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

$$\text{Var}\left(\frac{1}{s} \sum_{k=n_{\text{stat}}+1}^{n_{\text{Var}}} h(X_k) - \mathbb{E}_{\pi_x} h(X) \right) \leq \varepsilon_2 \ , \tag{31}$$

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

In the following we focus only on problem (30). We deal with the second problem in Romaniuk (2007b). Furthermore, for simplicity of formulation and notation, we limit ourselves to the case when $\mathcal{X}$ is a finite set. However, appropriate proofs may be easily generalized for the case of continuous state space $\mathcal{X}$. It is worth noting that from the computational and numerical point of view, the problem of cardinality of $\mathcal{X}$ is rather academic – in computers all the numbers are represented by the finite set of possibilities.

3.3. Probability constraints

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

$$\sum_{y\in\mathcal{X}} |\Pr_x^n(y) - \pi_X(y)| \le 2\Pr_x(\zeta_1^{(1)} \ge n) + \sum_{j=0}^{n-1} \Pr_x(\zeta_1^{(1)} = j)\cdot$$
$$\left(\sum_{k=1}^{n-j-1} \left|\Pr_{\mathcal{A}_1}^k(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)\right| \Pr_{\mathcal{A}_1}(\zeta_1^{(1)} \ge n-k-j) + \right.$$
$$\left. + \pi_X(\mathcal{A}_1)\mathbb{E}_{\mathcal{A}_1}\left(\zeta_1^{(1)} - (n-j)\right)_+\right) . \quad (32)$$

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

$$\pi_X(y) = \pi_X(\mathcal{A}_1) \sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n = y, \zeta_1^{(1)} \ge n) , \quad (33)$$

where $\Pr_x(.)$, as usually, denotes $\Pr(.|X_0 = x)$. The proof of (33) may be found in Robert, Casella (2004, see Theorem 4.5.3).

We have

$$\Pr_x^n(y) = \Pr_x(X_n = y, \zeta_1^{(1)} \ge n) + \sum_{j=0}^{n-1} \Pr_x(X_j \in \mathcal{A}_1, \zeta_1^{(1)} = j)\cdot$$
$$\left(\sum_{k=0}^{n-j-1} \Pr_{\mathcal{A}_1}^k(\mathcal{A}_1)\Pr_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \ge n-k-j)\right) . \quad (34)$$

The notation $P_{\mathcal{A}_1}^k(\mathcal{A}_1)$ and $\Pr_{\mathcal{A}_1}(.)$ is validated because of the thesis of Lemma 2.

Using expansion (34) we obtain

$$|\Pr_x^n(y) - \pi_X(y)| \le \Pr_x(X_n = y, \zeta_1^{(1)} \ge n) + \left|\sum_{j=0}^{n-1} \Pr_x(\zeta_1^{(1)} = j)\cdot\right.$$
$$\left.\left(\sum_{k=0}^{n-j-1} \Pr_{\mathcal{A}_1}^k(\mathcal{A}_1)\Pr_{\mathcal{A}_1}(X_{n-k-j} = y, \zeta_1^{(1)} \ge n-k-j\right) - \pi_X(y)\right| . \quad (35)$$

Hence

$$|\mathrm{Pr}_x^n(y)-\pi_X(y)|\le \mathrm{Pr}_x(X_n=y,\zeta_1^{(1)}\ge n)+\Bigg|\sum_{j=0}^{n-1}\mathrm{Pr}_x(\zeta_1^{(1)}=j)\cdot$$
$$\left(\sum_{k=0}^{n-j-1}\mathrm{Pr}^k_{\mathcal{A}_1}(\mathcal{A}_1)\mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j}=y,\zeta_1^{(1)}\ge n-k-j)-\pi_X(y)\right)$$
$$-\pi_X(y)\sum_{j=n}^{\infty}\mathrm{Pr}_x(\zeta_1^{(1)}=j)\Bigg| . \quad (36)$$

From (33) for any $j\le n-1$ we have

$$\pi_X(y)=\pi_X(\mathcal{A}_1)\sum_{k=0}^{n-j-1}\mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j}=y,\zeta_1^{(1)}\ge n-k-j)+$$
$$+\pi_X(\mathcal{A}_1)\sum_{l=n-j+1}^{\infty}\mathrm{Pr}_{\mathcal{A}_1}(X_l=y,\zeta_1^{(1)}\ge l) . \quad (37)$$

After applying (37) to (36) we have

$$|\mathrm{Pr}_x^n(y)-\pi_X(y)|\le \mathrm{Pr}_x(X_n=y,\zeta_1^{(1)}\ge n)+\Bigg|\sum_{j=0}^{n-1}\mathrm{Pr}_x(\zeta_1^{(1)}=j)\cdot$$
$$\left(\sum_{k=0}^{n-j-1}\left(\mathrm{Pr}^k_{\mathcal{A}_1}(\mathcal{A}_1)-\pi_X(\mathcal{A}_1)\right)\mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j}=y,\zeta_1^{(1)}\ge n-k-j)-\right.$$
$$\left.\pi_X(\mathcal{A}_1)\sum_{l=n-j+1}^{\infty}\mathrm{Pr}_{\mathcal{A}_1}(X_l=y,\zeta_1^{(1)}\ge l)\right)-\pi_X(y)\mathrm{Pr}_x(\zeta_1^{(1)}\ge n)\Bigg| . \quad (38)$$

Straightforwardly

$$|\mathrm{Pr}_x^n(y)-\pi_X(y)|\le \mathrm{Pr}_x(X_n=y,\zeta_1^{(1)}\ge n)+\sum_{j=0}^{n-1}\mathrm{Pr}_x(\zeta_1^{(1)}=j)\cdot$$
$$\left(\sum_{k=0}^{n-j-1}\left|\mathrm{Pr}^k_{\mathcal{A}_1}(\mathcal{A}_1)-\pi_X(\mathcal{A}_1)\right|\mathrm{Pr}_{\mathcal{A}_1}(X_{n-k-j}=y,\zeta_1^{(1)}\ge n-k-j)+\right.$$
$$\left.+\pi_X(\mathcal{A}_1)\sum_{l=n-j+1}^{\infty}\mathrm{Pr}_{\mathcal{A}_1}(X_l=y,\zeta_1^{(1)}\ge l)\right)+\pi_X(y)\mathrm{Pr}_x(\zeta_1^{(1)}\ge n) , \quad (39)$$

which constitutes (32). ■

The equations (32) and (39) may be used to establish further dependencies between the initial and the secondary chains. Now we present a simple lemma, which may be helpful in 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*

$$\Pr_{\mathcal{A}_1}(\zeta_1^{(1)} \geq n) \leq M_2 r_2^{-n} \ , \tag{40}$$

and

$$\Pr_x(\zeta_1^{(1)} = n) \leq M_3 r_3^{-n} \tag{41}$$

are fulfilled. Then inequality

$$\sum_{y \in \mathcal{X}} |\Pr_x^n(y) - \pi_X(y)| \leq \varepsilon_1 \tag{42}$$

is satisfied for n given as the solution of formula

$$2\frac{M_3 r_3^{1-n}}{r_3 - 1} + \frac{M_2 M_3 r_3 (r_3^{-n} - r_2^{-n})}{(r_2 - 1)(r_2 - r_3)} + \\ + \frac{M_1 M_2 M_3}{(r_2 - r_1)} \left(\frac{r_1 r_3 (r_3^{-n} - r_1^{-n})}{(r_1 - r_3)} + \frac{r_2 r_3 (r_3^{-n} - r_2^{-n})}{(r_3 - r_2)} \right) \leq \varepsilon_1 \ . \tag{43}$$

Proof. After applying conditions (8), (40), (41) to inequality (32) we can straightforwardly prove (43). ■

It is worth noting that it is possible to improve the inequality (43). If we know the value of stationary probability $\pi_X(\mathcal{A}_1)$, then we have a more detailed condition

$$2\frac{M_3 r_3^{1-n}}{r_3 - 1} + \frac{\pi_X(\mathcal{A}_1) M_2 M_3 r_3 (r_3^{-n} - r_2^{-n})}{(r_2 - 1)(r_2 - r_3)} + \\ + \frac{M_1 M_2 M_3}{(r_2 - r_1)} \left(\frac{r_1 r_3 (r_3^{-n} - r_1^{-n})}{(r_1 - r_3)} + \frac{r_2 r_3 (r_3^{-n} - r_2^{-n})}{(r_3 - r_2)} \right) \leq \varepsilon_1 \ . \tag{44}$$

In Lemma 4 there is an important assumption that $\mathcal{A}_1$ is an atom. However, we can relax this requirement using the following result:

LEMMA 6. *Suppose that $\mathcal{A}_1$ is a renewal set. Then we have*

$$\pi_X(y) = \frac{1}{\epsilon_{\mathcal{A}_1}} \pi_X(\mathcal{A}_1) \sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n = y, \zeta_1^{(1)} \geq n) \ . \tag{45}$$

Proof. As it was noted, for simplicity of notation the proof will be conducted for discrete MC. However, it could be easily adapted for continuous state space $\mathcal{X}$.

Let

$$\pi_X^{'}(y) = \sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n = y, \zeta_1^{(1)} \geq n) \ . \tag{46}$$

For any $y \in \mathcal{X}$ we have

$$\sum_{x \in \mathcal{X}} \Pr_x(y)\pi_X^{'}(x) = \sum_{x \in \text{ renewal set } \mathcal{A}_1} \Pr_x(y)\pi_X^{'}(x) + \\ + \sum_{x \notin \text{ renewal set } \mathcal{A}_1} \Pr_x(y)\pi_X^{'}(x) \ . \tag{47}$$

For the first sum, if $x \in$ renewal set $\mathcal{A}_1$, then we apply the formula (7). Therefore the probability of transition to the next state does not depend on a particular state x. For the second sum, we use (46). Hence

$$\sum_{x \in \mathcal{X}} \Pr_x(y)\pi_X^{'}(x) = \pi_X(\mathcal{A}_1)\nu_{\mathcal{A}_1}(y) + \\ + \sum_{x \notin \text{ renewal set } \mathcal{A}_1} \Pr_x(y) \left(\sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n = x, \zeta_1^{(1)} \geq n) \right) = \Pr_{\mathcal{A}_1}(y) + \\ + \sum_{x \notin \text{ renewal set } \mathcal{A}_1} \sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n = x, X_{n+1} = y, \zeta_1^{(1)} \geq n) \ . \tag{48}$$

Formula (48) may be simplified to

$$\sum_{x \in \mathcal{X}} \Pr_x(y)\pi_X^{'}(x) = \Pr_{\mathcal{A}_1}(y) + \sum_{n=1}^{\infty} \Pr_{\mathcal{A}_1}(X_n = y, \zeta_1^{(1)} \geq n) = \pi_X^{'}(y) \ , \tag{49}$$

therefore, (46) is an invariant measure.

From (46) we obtain

$$\pi_X^{'}(\mathcal{X}) = \sum_{n=0}^{\infty} \Pr_{\mathcal{A}_1}(X_n \in \mathcal{X}, \zeta_1^{(1)} \geq n) = \\ = \sum_{m=0}^{\infty} m \Pr_{\mathcal{A}_1}(\zeta_1^{(1)} = m) = \mathbb{E}_{\mathcal{A}_1}(\zeta_1^{(1)}) \ . \tag{50}$$

Hence, this measure is finite. Then from the theorem of invariant measure uniqueness, (46) is probability distribution after normalization.

From Kac's theorem we have

$$\mathbb{E}_{\mathcal{A}_1}(\zeta_1^{(1)}) = \mathbb{E}_{\mathcal{A}_1}(X \in \mathcal{A}_1, U_1 \leq \epsilon_{\mathcal{A}_1}) = \left(\pi_X^{'}(\mathcal{A}_1)\right)^{-1} \epsilon_{\mathcal{A}_1} \ , \tag{51}$$

which gives an appropriate normalizing constant for (45). Therefore

$$\pi_X(y) = \frac{1}{\epsilon_{\mathcal{A}_1}} \pi_X(\mathcal{A}_1) \pi_X^{'}(y) \ , \tag{52}$$

which leads to (45). ■

The technique similar to the above proof was used in Robert and Casella (2004).

Now we can prove the generalization of Lemma 4.

LEMMA 7. *Let $\mathcal{A}_1$ be a renewal set and all other assumptions be the same as in Lemma 4. Then*

$$\sum_{y \in \mathcal{X}} |\Pr_x^n(y) - \pi_X(y)| \leq 2\Pr_x(\zeta_1^{(1)} \geq n) + \sum_{j=0}^{n-1} \Pr_x(\zeta_1^{(1)} = j) \cdot$$
$$\left(\sum_{k=1}^{n-j-1} \left| \Pr_{\mathcal{A}_1}^k(\mathcal{A}_1) - \frac{1}{\epsilon_{\mathcal{A}_1}} \pi_X(\mathcal{A}_1) \right| \Pr_{\mathcal{A}_1}(\zeta_1^{(1)} \geq n - k - j) + \right.$$
$$\left. + \frac{1}{\epsilon_{\mathcal{A}_1}} \pi_X(\mathcal{A}_1) \mathbb{E}_x \left(\zeta_1^{(1)} - (n-j) \right)_+ \right) \ . \tag{53}$$

Proof. Analogously to proof of Lemma 4, we apply the formula (45) to (36), obtaining (53). ■

Having Lemma 7 we can modify the result of Lemma 5.

LEMMA 8. *Suppose that $\mathcal{A}_1$ is a renewal set which fulfils the condition*

$$|\Pr_{\mathcal{A}_1}^n(\mathcal{A}_1) - \frac{1}{\epsilon_{\mathcal{A}_1}} \pi_X(\mathcal{A}_1)| \leq M_1 r_1^{-n} \tag{54}$$

and there exist $M_2 > 0$, $r_2 > 1$, $M_3 > 0$, $r_3 > 1$ such that inequalities (40) *and* (40) *are satisfied. Then condition* (42) *is met for n given as the solution of formula* (43).

Proof. Using Lemma 7 analogously as in proof of Lemma 5, we obtain solution (43). ■

3.4. Heuristic approach

In the heuristic approach we use results from Lemma 3. The method may be generalized for continuous space $\mathcal{X}$.

From Lemma 3 for atoms we have

$$\pi_Y(\mathcal{A}_j) = \frac{\sum_{x\in\mathcal{A}_j}\pi_X(x)}{\sum_{x\in\mathcal{A}_1}\pi_X(x)+\sum_{x\in\mathcal{A}_2}\pi_X(x)}\ , \tag{55}$$

and for renewal sets

$$\pi_Y(\mathcal{A}_j) = \frac{\epsilon_{\mathcal{A}_j}\sum_{x\in\mathcal{A}_j}\pi_X(x)}{\epsilon_{\mathcal{A}_1}\sum_{x\in\mathcal{A}_1}d\pi_X(x)+\epsilon_{\mathcal{A}_2}\sum_{x\in\mathcal{A}_2}d\pi_X(x)}\ . \tag{56}$$

It is easily seen that these equations may be used as indicators of *distance between stationary distributions* $\pi_X(.)$ and $\pi_Y(.)$, if for left and right sides of (55) and (56) we take estimators based on *various information*. We denote these estimators as $\hat{\pi}_{X,n}(.)$ and $\hat{\pi}_{Y,n}(.)$, where n emphasizes the number of steps in the sequence $X_0, X_1, \ldots, X_n$. We are then interested in convergence diagnosis based on difference

$$\left|\hat{\pi}_{Y,n}(\mathcal{A}_j) - \frac{\sum_{x\in\mathcal{A}_j}\hat{\pi}_{X,n}(x)}{\sum_{x\in\mathcal{A}_1}\hat{\pi}_{X,n}(x)+\sum_{x\in\mathcal{A}_2}\hat{\pi}_{X,n}(x)}\right| \le \varepsilon_3 \tag{57}$$

for atoms, and after adequate modification of formula (57) according to (56), for renewal sets. Intuitively, if quantity (57) is small enough, we could diagnose convergence.

Estimator $\hat{\pi}_{Y,n}(.)$ is based on transition probabilities. Let

$$m_{(j,l),n} = \#\{k : Y_k \in \mathcal{A}_j, Y_{k+1} \in \mathcal{A}_l, \zeta_{k+1} \le n\}\ . \tag{58}$$

Then

$$\hat{\alpha}_{Y,n} = \frac{m_{(1,2),n}}{m_{(1,1),n}+m_{(1,2),n}} \tag{59}$$

which is a natural estimator for probability of moving between states $\mathcal{A}_1$ and $\mathcal{A}_2$ for secondary chain Y. Analogously

$$\hat{\beta}_{Y,n} = \frac{m_{(2,1),n}}{m_{(2,1),n}+m_{(2,2),n}}\ , \tag{60}$$

and the estimator of transition matrix for Y is given by

$$\hat{\mathbb{P}}_{Y,n} = \begin{pmatrix} 1-\hat{\alpha}_{Y,n} & \hat{\alpha}_{Y,n} \\ \hat{\beta}_{Y,n} & 1-\hat{\beta}_{Y,n} \end{pmatrix}\ . \tag{61}$$

For two-state discrete MC, the estimator of stationary distribution in this case is

$$\hat{\pi}^{\mathrm{T}}_{Y,n} = (\hat{\pi}_{Y,n}(\mathcal{A}_1), \hat{\pi}_{Y,n}(\mathcal{A}_2)) = \frac{1}{\hat{\alpha}_{Y,n}+\hat{\beta}_{Y,n}}\,(\hat{\beta}_{Y,n}, \hat{\alpha}_{Y,n})\ . \tag{62}$$

For estimator $\hat{\pi}_{X,n}(.)$ we apply the weak ergodic theorem. Let

$$\eta_{X,n}(x) = \frac{\mathbb{1}(X_0 = x) + \ldots + \mathbb{1}(X_n = x)}{n+1} \ . \tag{63}$$

Then natural estimator for unknown parameter is

$$\hat{\pi}_{X,n}(x) = \eta_{X,n}(x) \ . \tag{64}$$

It is worth noting that apart from using the same MC, we create the above estimators based on other kind of information – the frequency of moving between states and calculation of transition probability from transition matrix in case of $\hat{\pi}_{Y,n}(.)$, and direct counting of visits in the appropriate states with application of ergodic theorem for $\hat{\pi}_{X,n}(.)$.

For additional diversification of information used for these estimators, it is possible to construct two separate chains or to divide one chain into two parts.

4. Example of application

After introducing methods appropriate for finding the value n_{stat}, now we present examples of their application. Firstly, we use state space $\mathcal{X}$ with a few atoms. Then we investigate the renewal sets case.

4.1. Atoms case

We should emphasize that the solutions established in lemmas of Section 3.3 give *exact* (i.e. proved by mathematical reasoning, not *heuristic*) and *precise* (i.e. *non-asymptotic*) values. Therefore we may focus only on the problem of transferring the obtained results from theoretical formulas to the practical example.

Let us suppose that we are interested in MCMC algorithm, for which function $f(.)$ describes the state space $\mathcal{X}$ with eight atoms and stationary probabilities

$$f(.) = (1/20, 1/20, 2/20, 2/20, 3/20, 3/20, 4/20, 4/20) \ , \tag{65}$$

i.e. first atom has stationary probability 1/20, the second one – 1/20, etc.

We use *independent Metropolis-Hastings algorithm* (see e.g. Robert and Casella, 2004). Our main trajectory has one million elements and starts from state one. We also assume that $\mathcal{A}_1 = 3$ and $\mathcal{A}_2 = 7$. Therefore, we may compare the values n_{stat} based on states with various stationary probabilities.

In order to apply lemmas from Section 3.3, we have to evaluate the 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 them. Hence, we generate additional sets of 50,000 trajectories with 100 steps in each trajectory and appropriate starting points – states one, three and seven. Then, we apply "pessimistic optimization" approach (see also Romaniuk, 2007b).

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

$$|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)| \approx M_1 r_1^{-n} \ , \tag{66}$$

then

$$\frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \approx r_1^{-n+1} \ . \tag{67}$$

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

$$\hat{r}_1 = \min_{r\in\mathbb{R}_+} \left\{ \forall n = 2,3,\ldots : r^{-n+1} - \frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \geq 0 \right\} \ . \tag{68}$$

It can be easily seen that (68) gives us the "maximal pessimistic" guess of $\hat{r}_1$, because in this light $\hat{r}_1$ has to be the upper limit for all steps in a 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 satisfying the weaker condition

$$r^{-n+1} - \frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \geq 0 \quad \vee \quad \left| r^{-n+1} - \frac{|\Pr^n_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|}{|\Pr_{\mathcal{A}_1}(\mathcal{A}_1) - \pi_X(\mathcal{A}_1)|} \right| \leq \delta \tag{69}$$

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

After fixing the value $\hat{r}_1$ like in (68), we may find $\hat{M}_1$ in the same manner, as satisfying the condition

$$\hat{M}_1 = \min_{M\in\mathbb{R}_+} \left\{ \forall n = 1,2,\ldots : M\hat{r}_1^{-n} - |\Pr^n_{\mathcal{A}_1}(y) - \pi_X(\mathcal{A}_1)| \geq 0 \right\} \ . \tag{70}$$

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

$$\hat{r}_1 = 1.04 \ , \ \hat{M}_1 = 0.0268 \ , \hat{r}_2 = 1.0941 \ , \ \hat{M}_2 = 1.0888 \ , \quad \hat{r}_3 = 1.0904 \ , \ \hat{M}_3 = 0.1372 \ . \tag{71}$$

We can substitute these values into the formula (43) 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

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

Table 1. Evaluation of n_{stat} for $\mathcal{A}_1$

value ε_1" gives the exact value of the left hand side for (43) and number of steps n_{stat} is in the second column.

The graph of the left hand side (43) as a function of the number of steps n is shown in Fig. 1.

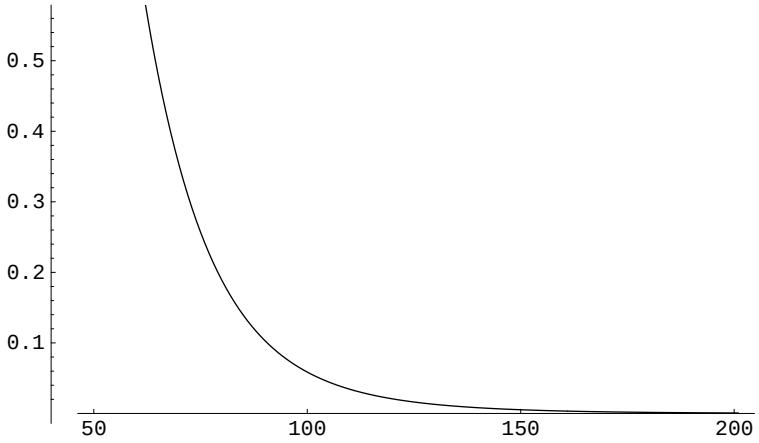


Figure 1. Error level ε_1 as a function of n for $\mathcal{A}_1$

If we use the improved inequality (44) instead of (43), 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 more easily seen in Fig. 2, where curve is much steeper at the beginning of the graph.

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

$$\hat{r}_1 = 1.0438\ ,\ \hat{M}_1 = 0.0793\ ,\ \hat{r}_2 = 1.14385\ ,\ \hat{M}_2 = 1.1439\ ,$$
$$\hat{r}_3 = 1.1231\ ,\ \hat{M}_3 = 0.1394\ . \quad (72)$$

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

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

Table 2. Evaluation of n_{stat} for $\mathcal{A}_1$ based on inequality (44)

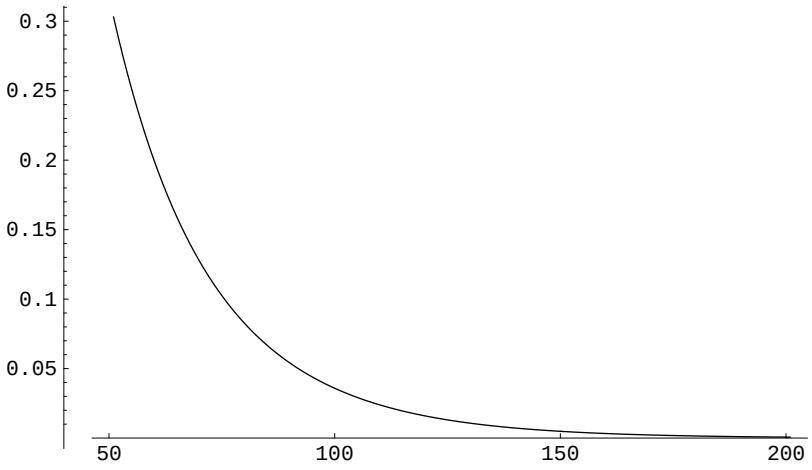


Figure 2. Error level ε_1 as a function of n for $\mathcal{A}_1$ based on inequality (44)

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

Table 3. Evaluation of n_{stat} for $\mathcal{A}_2$

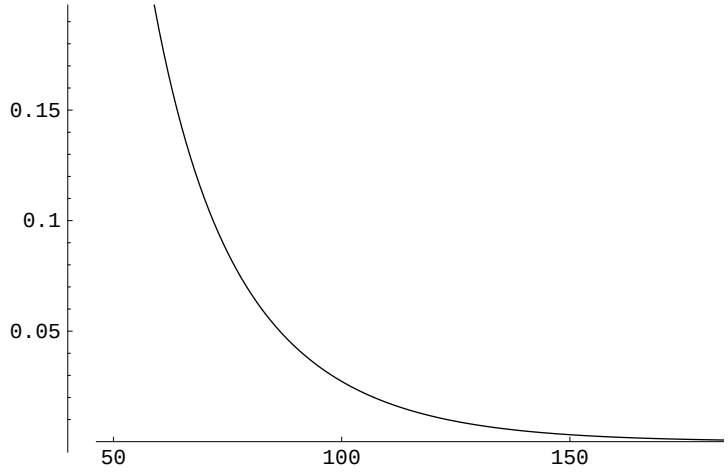


Figure 3. Error level ε_1 as a function of n for $\mathcal{A}_2$

We can also apply improved inequality (44) for the set $\mathcal{A}_2$, but by reason of faster exponential convergence guaranteed by higher values of $\hat{r}_i$, the profit in terms of reduction of the n_{stat} value is not so visible as in previous case.

Now we turn to the heuristic approach (see Section 3.4). Because the difference (57) has the same value for both $\mathcal{A}_1$ and $\mathcal{A}_2$, we may focus only on any of the atoms.

The appropriate graph of (57) as a function of n may be found in Fig. 4. Because of the "jerked" nature of this diagram, it is not so easy to directly point out the moment where the difference is lesser than the fixed value ε_3. However, we may risk hyphotesis, that (57) is lesser than $\varepsilon_3 = 0.2$ from 100th – 150th step, and lesser than $\varepsilon_3 = 0.1$ from 300th – 350th step. Compared with previous results, these numbers of steps are much bigger. It seems that such heuristic approach is rather conservative. However, it could be easily applied and the graphs like Fig. 4 may be used as warnings against possible problems with diagnosis or multimodality of π_X.

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* issues (see e.g. Koronacki, Lasota, Niemiro, 2005, and Lasota, Niemiro, 2003).

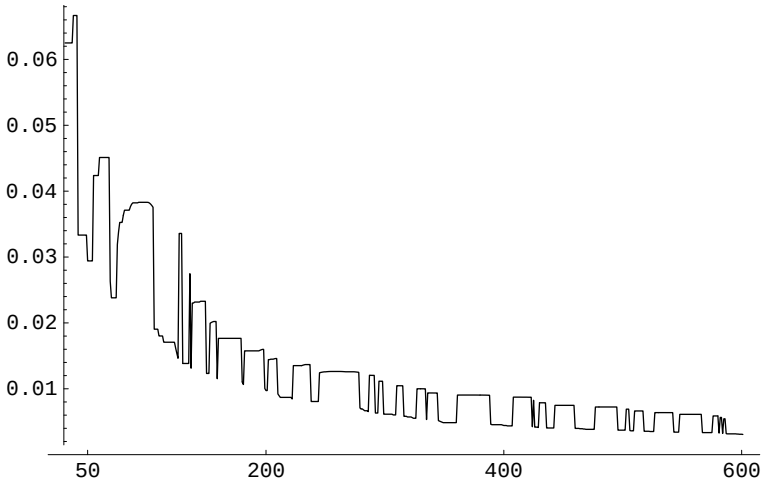


Figure 4. The value of (57) as a function of steps number

4.2. Renewal sets case

Now we apply the methods discussed for the case of renewal sets. Let us suppose that the density $f(.)$ is described by the two-modal mixture of normal distributions

$$f(.) \sim \alpha N(\mu_1, \sigma_1^2) + (1-\alpha) N(\mu_2, \sigma_2^2) \ , \tag{73}$$

where $N(\mu_i, \sigma_i^2)$ is a normal distribution with expected value μ_i and standard deviation σ_i, and α is a mixture parameter. In our example we take

$$\alpha = 0,5 \ , \ \mu_1 = 3 \ , \ \sigma_1 = 1 \ , \ \mu_2 = 7 \ , \ \sigma_2 = 2 \tag{74}$$

(see Fig. 5).

From (4) for continuous state space $\mathcal{X}$ for any $x \in \mathcal{A}_j$ we have (with accuracy to the zero-measure sets)

$$\mathcal{K}_X(x,y) \geq \epsilon_j f_{\nu_j}(y) \ , \tag{75}$$

where $f_{\nu_j}(.)$ is the density function of measure $\nu(.)$ for renewal set $\mathcal{A}_j$.

In this example MC is generated by Metropolis-Hastings algorithm with random walk and symmetric proposal density, i.e. $g(y|x) = g(x|y) = g(|x-y|)$. For simplicity let $g(|.-x|)$ be a uniform distribution on $(x-0,5; x+0,5)$ for fixed x. Let $\mathcal{A}_1 = [2, 2.25]$ and $\mathcal{A}_2 = [7.75, 8]$. In order to use solution provided by Lemma 8, we have to use modification (7). The graph of $f_{\nu_1}(.)$ without the normalizing constant may be found in Fig. 6. It can be easily seen that this

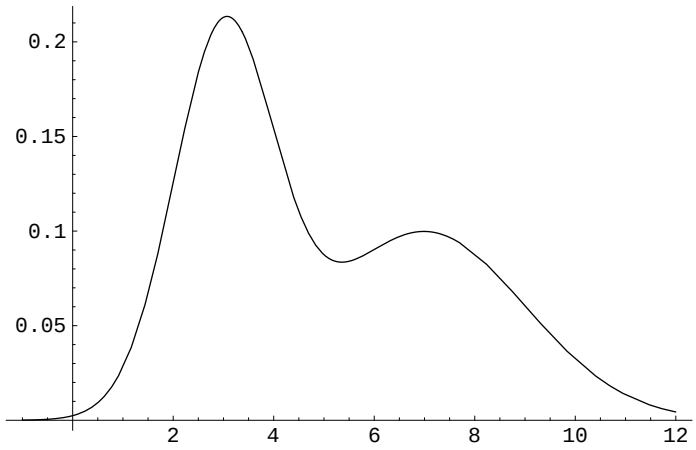


Figure 5. Example of mixture (73)

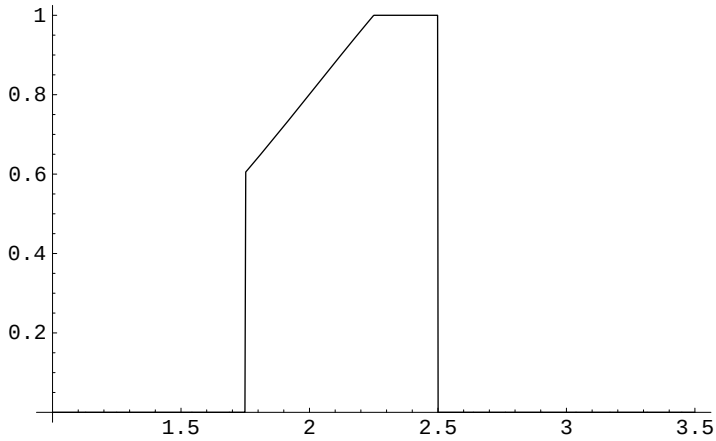


Figure 6. Graph of $f_{\nu_1}(.)$ without the normalizing constant for $\mathcal{A}_1 = [2, 2.25]$

normalizing constant is simultaneously the inverse of maximum ϵ_1 and may be found numerically, which gives $\epsilon_1 = 0.6506717064872144$.

Similary, $\epsilon_2 = 0.722459686557494$ and graph of $f_{\nu_2}(.)$ may be found in Fig. 7.

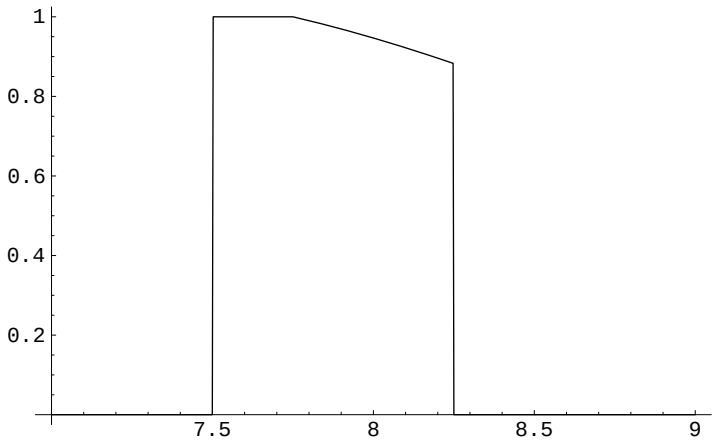


Figure 7. Graph of $f_{\nu_2}(.)$ without the normalizing constant for $\mathcal{A}_2 = [7.75, 8]$

In order to find parameters $r_1, M_1, r_2, M_2, r_3, M_3$, necessary for conditions (40), (41), (54) we use similar approach as in Section 4.1. Therefore, for $\mathcal{A}_1$ we have

$$\hat{r}_1 = 1.034 \ , \ \hat{r}_2 = 1.0345 \ , \ \hat{r}_3 = 1.0131 \ ,$$
$$\hat{M}_1 = 1.05 \ , \ \hat{M}_2 = 1.0069 \ , \ \hat{M}_3 = 0.022 \quad (76)$$

and for $\mathcal{A}_2$

$$\hat{r}_1 = 1.03 \ , \ \hat{r}_2 = 1.0318 \ , \ \hat{r}_3 = 1.0078 \ ,$$
$$\hat{M}_1 = 10.6702 \ , \ \hat{M}_2 = 0.9957 \ , \ \hat{M}_3 = 0.007 \ . \quad (77)$$

These parameters give us the solutions for inequality (43) (see Tables 4 and 5)

As previously, evaluation for the "less frequent" set $\mathcal{A}_2$ increases the n_{stat} value by about $20 - 30\%$.

The heuristic approach may also be applied for this case. The graph is similar as in the previous example with "jerked" character.

5. Concluding remarks

We started from formulation of two inequalities, which correspond to standard questions in MCMC setups, i.e. *when the sampled transition probability*

Assumed value ε_1	Number of steps n_{stat}	True value ε_1
0.1	521	0.0989617
0.02	644	0.0199662
0.01	698	0.0098872
0.001	875	0.000987682

Table 4. Evaluation of n_{stat} for set $\mathcal{A}_1$

Assumed value ε_1	Number of steps n_{stat}	True value ε_1
0,1	682	0.0995218
0,02	889	0.0199263
0,01	979	0.00990229
0,001	1275	0.000992949

Table 5. Evaluation of n_{stat} for set $\mathcal{A}_2$

is close to 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 – the numbers 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 find the n_{stat} estimator. Therefore, we obtain a useful set of conditions which could be used for checking the convergence in MCMC setup. The examples of application of theoretical lemmas and of heuristic approach for the case of state space with atoms and renewal sets are also provided. It has to be mentioned that this paper comprises some contents of doctoral dissertation (see Romaniuk, 2007a), 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 the 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 suitable satisfactory level.

The possibilities of complementing the discussed method might also be considered. For example, the obtained conditions might be improved, like in (44). However, additional information about the structure of state space or underlying Markov Chain may be necessary in such 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 number 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., 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**.

El Adlouni, S., Favre, A.-C. and Bobée, B. (2006) Comparison of methodologies to asses the convergence of Markov chain Monte Carlo methods. *Computational Statistics & Data Analysis* **50**, 2685–2701.

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 Stratistical 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 to-

mography 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 Reviewww, in: BERNARDO J. M., BERGER J. O., DAWID A. P., SMITH A. F. M. (EDS.) *Bayesian Statistics 6*, 415–440, Oxford University Press.

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.

NUMMELIN, E. (2001) *MC's for MCMC'ists.* Preprint 310, December 2001.

RAFTERY, A.E. and LEWIS, S.M. (1999) How many iterations in the Gibbs Sampler? In: BERNARDO, J.M., BERGER, J.O., DAWID, A.P. and SMITH, A.F.M., 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.* Springer–Verlag, 2nd ed., 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. (2007A) *Application of renewal sets for convergence diagnosis of MCMC setups* (in Polish). Ph.D. dissertation, Systems Research Institute Polish Academy of Sciences.

ROMANIUK, M. (2007B) On Some Method for Diagnosing Convergence in MCMC Setups via Atoms and Renewal Sets. *Control and Cybernetics* **36**, 4, 985–1008.