# CONVERGENCE OF MARKOV CHAINS IN THE RELATIVE SUPREMUM NORM

LARS HOLDEN
*NORWEGIAN COMPUTING CENTER*

ABSTRACT It is proved that the strong Doeblin condition (i.e., $r^s(y|x) \geq a_s\pi(y)$ for all $x, y$ in the state space) implies convergence in the relative supremum norm for a general Markov chain. The convergence is geometric with rate $(1 - a_s)^{1/s}$. If the detailed balance condition is satisfied, then the strong Doeblin condition is equivalent to convergence in the relative supremum norm. It is proved that the pointwise relative error vanishes and the chain converges in most other norms under weaker assumptions. The results in the paper also give a qualitative understanding of the convergence.

1991 MATHEMATICS SUBJECT CLASSIFICATION 65C05, 65U05

KEY WORDS Markov chain, geometric convergence, relative supremum norm.

## 1. INTRODUCTION

Markov chains are widely used as models and computational devices in areas ranging from statistics to physics. The theory and applications of Markov chains are very active fields of research: see, for example, Meyn and Tweedie (1993) and Gilks, Richardson, and Spiegelhalter (1996).

This paper shows that the strong Doeblin condition implies geometric convergence in the relative supremum norm i.e. $\sup_x\{(p(x) - \pi(x))/\pi(x)\}$ if the initial relative error is bounded. The strong Doeblin condition is that the $s$-step transition density $r^s(y|x)$ satisfies $r^s(y|x) \geq a_s\pi(y)$ for all $x, y$ in the state space. The convergence rate is $(1 - a_s)^{1/s}$. For Markov chains satisfying the detailed balance condition, the strong Doeblin condition is equivalent with convergence in the relative supremum norm. Similarly, Meyn and Tweedie (1993), Theorem 16.2.3, prove that the Doeblin condition is equivalent to uniform ergodicity, i.e. uniform geometric convergence in total variation norm. The strong Doeblin condition implies the Doeblin condition.

The relative supremum norm is used in this paper since there is a very simple expression for propagation of the relative error as stated in the Proposition. This simple expression leads to a bound on the relative supremum error that easily may be used to prove convergence in other norms and bounds on the eigenvalues. In many applications it is easier to estimate the coefficient $a_s$ in the strong Doeblin condition than other convergence criteria like e.g. eigenvalues of the transition density $r^s(y|x)$. For Markov chains

*Date*: April 7th 1997, revised versionApril 7th 1999.
Research supported by Research Council of Norway.

that do not satisfy the strong Doeblin condition, convergence in other norms may be proved using Theorem 2 in this paper.

The relative supremum norm is widely used in approximation theory, differential equations and linear equation systems. It is the most appropriate norm for some applied problems. In order to bound the expected value of a function of a stochastic variable, it is necessary to bound the relative error of the stochastic variable. This is illustrated by the following: Let $x$ be a stochastic variable and $p$ an approximation to the distribution $\pi$ of $x$. Let further $f_p = \int f(x)p(x)dx$ and $f_\pi = \int f(x)\pi(x)dx$. Then

$$|f_p - f_\pi| \leq \int f(x)|p(x) - \pi(x)|dx \leq f_\pi sup_x \frac{|p(x) - \pi(x)|}{\pi(x)}.$$

In Stewart (1994) p. 158, this norm is recommended in checking convergence of Markov chains.

This paper is a generalisation of the results in Holden (1998) which proves similar theorems for the Metropolis–Hastings algorithm. It also generalises the results in Mengersen and Tweedie (1996) from an independent to a general proposal function in the Metropolis–Hastings algorithm.

## 2. A general Markov chain

Let $\Omega \subset \mathbb{R}^n$ be a Borel measurable state space or, alternatively, let $\Omega$ be a discrete state space, and let $\mu$ be a measure on $\Omega$. Let further the transition density $r$ be integrable with respect to $\mu$, including point mass distributions. The results are also valid for more general state spaces $\Omega$. A Markov chain is defined as follows.

Markov chain.

1. Generate an initial state $x^0 \in \Omega$ from the density $p^0$
2. For $i = 0, \ldots, n$:
   Generate a new state $x^{i+1}$ from the density $r(.|x^i)$.

The $s$-step transition density $r^s$ and the density after $j$ iterations, $p^j$ are implicitly defined by

$$p^{i+s}(y) = \int_\Omega r^s(y|x)p^i(x)\, d\mu(x) \tag{1}$$

where $r^1(y|x) = r(y|x)$. According to Meyn and Tweedie (1993) Theorem 13.0.1, there exists a limiting density $\pi$ and the Markov chain converges in total variation norm to this limiting density if the chain is aperiodic, Harris recurrent and has an invariant measure. In this paper it is assumed that there exists a density $\pi$ that is integrable with respect to $\mu$, and $\pi(x) > 0$ for $x \in \Omega$, and $\int_\Omega \pi(x)d\mu(x) = 1$, and

$$\pi(y) = \int_\Omega r^s(y|x)\pi(x)\, d\mu(x) \qquad \text{for all integers } s > 0. \tag{2}$$

The strong Doeblin condition requires that there exist an integer $s > 0$ and a constant $a_s \in (0, 1]$, such that

$$r^s(y|x) \geq a_s \pi(y) \qquad \text{for all } x, y \in \Omega. \tag{3}$$

The strong Doeblin condition implies the Doeblin condition as defined in Meyn and Tweedie (1993) p. 391; there exists a probability measure $\phi$ with the property that for some $s$, $\varepsilon < 1$, $\delta > 0$

$$\phi(A) > \varepsilon \Rightarrow r^s(A|x) \geq \delta$$

for every $x \in \Omega$ and $A \subseteq \Omega$. We use the notation that $\phi(A) = \int_A \phi(x)d\mu(x)$ and $r^s(A|x) = \int_A r^s(y|x)d\mu(y)$. See also Doob (1953), p. 197. The Doeblin condition has weaker assumptions in the tails of the distribution.

The detailed balance condition is

$$r^s(y|x)\pi(x) = r^s(x|y)\pi(y) \qquad \text{for all } x, y \in \Omega. \tag{4}$$

It is trivial to prove that if the detailed balance condition is satisfied for $s = 1$, it is also satisfied for $s > 1$. This is satisfied for a wide class of Markov chains including the Metropolis-Hastings algorithm, see Example 2. The detailed balance condition is discussed in e.g. Ripley (1987) and Geyer (1993).

A set $C \subseteq \Omega$ is $v_s$-small if there exist $s > 0$ and a non-trivial measure $\nu_s$ such that for $x \in C$ and $y \in \Omega$

$$r^s(y|x) \geq \nu_s(y).$$

It is convenient to define the pointwise relative error $R^i(x) = (p^i(x) - \pi(x))/\pi(x) = p^i(x)/\pi(x) - 1$ and the relative supremum norm $R^i_M = \sup_{x\in\Omega} \left|R^i(x)\right|$. The following proposition is used in several of the proofs.

Proposition The pointwise relative error satisfies for $i \geq 0$

$$R^{i+s}(y) = \int_\Omega \frac{r^s(y|x)}{\pi(y)} R^i(x)\pi(x)\, d\mu(x).$$

The Proposition states that the relative error at step $i + s$ is the average of the relative error at step $i$ weighted by $\pi(x)r^s(y|x)/\pi(y)$. In many applications $r^s(y|x)$ is a smooth density which is large only for $|x - y|$ small (i.e. small steps are more likely). Assume for a moment that the error function $R^i(x)$ may be approximated by $\sum_i c_i \sin(2\pi i x)$ and that $r(y|x) = b\exp((x - y)/\sigma)^2$. Then the Proposition may be used to show that high-frequency errors ( i.e. $i$ large) decreases faster than low-frequency errors. This effect is larger the smaller $\sigma$ is. This property is also illustrated in Example 1.

Proof Combining (1) and (2) gives

$$p^{i+s}(y) - \pi(y) = \int_\Omega r^s(y|x)(p^i(x) - \pi(x))\, d\mu(x)$$

which proves the Proposition by

$$\frac{p^{i+s}(y)}{\pi(y)} - 1 = \int_\Omega \frac{r^s(y|x)}{\pi(y)}(p^i(x) - \pi(x))\ d\mu(x).$$

□

## 3. Assuming the strong Doeblin condition

Theorem 1. Assume that $\pi(x) > 0$ for all $x \in \Omega$ and that the strong Doeblin condition (3) is satisfied including $a_s = 0$. Then the probability density of the Markov chain satisfies for $i \geq 0$

$$\left|\frac{p^{i+s}(y)}{\pi(y)} - 1\right| \leq (1 - a_s) \sup_{x\in\Omega} \left\{\left|\frac{p^i(x)}{\pi(x)} - 1\right|\right\} \text{ for all } y \in \Omega. \tag{5}$$

Assume that the detailed balance condition (4) is satisfied and that the strong Doeblin condition is not satisfied, then there exist $p^0$ with $R^0_M$ finite, and a constant $c > 0$ such that $\sup_{x\in\Omega}\{|p^i(x)/\pi(x) - 1|\} > c$ for all $i > 0$.

The strong Doeblin condition (3) with $a_s > 0$ is a necessary and sufficient condition for convergence in the relative supremum norm for chains satisfying the detailed balance condition. In Example 5 the transition density satisfies neither the strong Doeblin condition nor the detailed balance condition, but the Markov chain converges in the relative supremum norm. Note that the convergence depends on $p^0$ unlike similar theorems for total variation norms, see Mengersen and Tweedie (1996). If the right hand side of (5) is not bounded, we can not expect to get convergence.

The theorem implies that $R^i_M = \sup_{x\in\Omega} |(p^i(x)/\pi(x)) - 1|$ does not increase and that the Markov chain converges geometrically if $a_s > 0$ and $R^0_M$ finite. The convergence is fast if $r^s(y|x) \approx \pi(y)$ for all $x, y \in \Omega$ and in s steps if $r^s(y|x) \equiv \pi(y)$. Often $r(y|x) = 0$ for combinations of $x, y \in \Omega$. Then it is necessary to use several steps (i.e. $s > 1$) such that the strong Doeblin condition (3) is satisfied as shown in Holden (1998). In that method it is necessary to specify a path between any two states. The size of $a_s$ depends on how probable this path is. The decrease in the relative supremum norm per iteration is $(1 - a_s)^{1/s}$. In some cases increasing $s$ gives better estimates for the convergence per step. This is illustrated in Example 1.

This theorem may be used for comparison between different transition densities. This is also possible if these transition densities have different computational cost, such that the number of iterations differs in the comparison.

In later sections this result is used for proving geometric convergence in other norms and bounds on the eigenvalues. Convergence in the relative supremum norm implies convergence in most other norms.

Proof The Proposition gives

$$\begin{aligned}
R^{i+s}(y) &= \int_\Omega \frac{r^s(y|x)}{\pi(y)} R^i(x)\pi(x)\, d\mu(x) \\
&= R^i_M \int_\Omega \frac{r^s(y|x)}{\pi(y)} \pi(x)\, d\mu(x) - \int_\Omega \frac{r^s(y|x)}{\pi(y)} (R^i_M - R^i(x))\pi(x)\, d\mu(x) \\
&\leq R^i_M - a_s \int_\Omega (R^i_M - R^i(x))\pi(x)\, d\mu(x) \\
&= R^i_M(1 - a_s) + a_s \int_\Omega R^i(x)\pi(x)\, d\mu(x) \\
&= R^i_M(1 - a_s).
\end{aligned}$$

Define $\tilde{p}^i$ such that the corresponding $\tilde{R}^i = -R^i$. Note that $\tilde{p}^i$ may be negative and thus not a density. Perform the same calculation as above with $\tilde{R}$ replacing $R$. This gives

$$-R^{i+s}(y) = \tilde{R}^{i+s}(y) \le R^i_M(1-a_s)$$

which combined with the inequality above gives (5). It remains to prove the implication in the other direction.

First it is needed to prove that there exist Borel measurable sets $A, B \subseteq \Omega$ satisfying

$$r^{s_2}(y|x) \ge \pi(y)/2 \qquad \text{for all } x \in A \text{ and all } y \in B \tag{6}$$

where $\pi(A) > 0$ and $\pi(B) > 0$. Define the function $g(y,x) = \max\{0, r^{s_2}(y|x) - \pi(y)/2\}$. This function is non-negative and for all values of $x$, $\int g(y,x)dy \ge .5$. Then the Tonelli theorem in (Royden 1968), p 270, implies that the function $g$ is integrable for $(y,x)$ in $\Omega \times D$ where $D$ is a subset of $\Omega$ of finite measure. Then $\int_{\Omega\times D} g(y,x)d(\mu\times\mu)(y,x) > 0$ implies the existence of $A \times B \subseteq \Omega \times D \subseteq \Omega \times \Omega$ such that $g(y,x) > 0$ for $(x,y) \in A \times B$. This implies (6).

Assume that the strong Doeblin condition is not satisfied. Then the ratio $r^s(v|w)/\pi(v)$ for any $s > 0$ may be made arbitrary small for some choice of $v, w \in \Omega$. The jump from $w$ to $v$ may be made using $s_1$ steps to $x \in A$, and $s_2$ steps to $y \in B$, and then $s_3$ steps to $v$ where $s_1 + s_2 + s_3 = s$. This gives the following bound on the ratio

$$\begin{aligned}
\frac{r^s(v|w)}{\pi(v)} &\ge \frac{1}{\pi(v)} \int_A \int_B r^{s_1}(x|w) r^{s_2}(y|x) r^{s_3}(v|y) d\mu(y) d\mu(x) \\
&\ge \frac{1}{2\pi(v)} \int_A \int_B r^{s_1}(x|w) \pi(y) r^{s_3}(v|y) d\mu(y) d\mu(x) \\
&= \frac{1}{2\pi(v)} \int_A r^{s_1}(x|w) d\mu(x) \int_B \pi(y) r^{s_3}(v|y) d\mu(y) \\
&= \frac{1}{2\pi(w)} \int_A r^{s_1}(w|x)\pi(x) d\mu(x) \frac{1}{\pi(v)} \int_B r^{s_3}(v|y)\pi(y) d\mu(y)
\end{aligned}$$

where the detailed balance equation is used in the last equation. Since $r^s(v|w)/\pi(v)$ can be arbitrary small for any value of $s > 0$, then either $H_A(w,s_1) = \frac{1}{\pi(w)}\int_A r^{s_1}(w|x)\pi(x)d\mu(x)$ or $H_B(v,s_3) = \frac{1}{\pi(v)}\int_B r^{s_3}(v|x)\pi(x)d\mu(x)$ can be made arbitrary small. Define

$$C := \begin{cases} A & \text{if for all } \varepsilon > 0 \text{ and all } s, \text{ there exists } w \in \Omega \text{ such that } H_A(w,s) < \varepsilon \\ B & \text{otherwise.} \end{cases}$$

Then $H_C(z,s') = \frac{1}{\pi(z)}\int_B r^{s'}(z|x)\pi(x)d\mu(x)$ can be arbitrary small for any value of $s'$. Define

$$p^0(x) := \begin{cases} (1+c_1)\,\pi(x) & \text{for } x \in C, \\ (1-c_2)\,\pi(x) & \text{for } x \in \Omega \setminus C, \end{cases} \tag{7}$$

for $c_1, c_2 > 0$. The definition of $p^0(x)$ gives for any value of $s' > 0$

$$\begin{aligned} p^{s'}(z) - \pi(z) &= \int_\Omega r^{s'}(z|x)(p^0(x) - \pi(x))\, d\mu(x) \\ &= \int_C r^{s'}(z|x)(p^0(x) - \pi(x))\, d\mu(x) \\ &\quad + \int_{\Omega\setminus C} r^{s'}(z|x)(p^0(x) - \pi(x))\, d\mu(x) \\ &= c_1 \int_C r^{s'}(z|x)\pi(x)\, d\mu(x) - c_2 \int_{\Omega\setminus C} r^{s'}(z|x)\pi(x)\, d\mu(x) \\ &= -c_2\pi(z) + (c_1 + c_2) \int_C r^{s'}(z|x)\pi(x)\, d\mu(x) \end{aligned}$$

Since $\frac{1}{\pi(z)} \int_C r^{s'}(z|x)\pi(x)\, d\mu(x)$ may be made arbitrary small for any value of $s'$

$$\sup_{x\in\Omega} |\frac{p^{s'}(x)}{\pi(x)} - 1| \geq c_2$$

for all values of $s'$. The relative supremum norm does not increase according to the first part of the theorem. Hence, the last part of the theorem is at least satisfied for $p^0$ defined in (7). □

## 4. Without assuming the strong Doeblin condition

In this section we prove that the pointwise relative error vanishes under weaker assumption than the strong Doeblin condition. This weaker assumption may for some applications be easier to verify than other convergence criteria. It is also useful for proving convergence in other norms than the relative supremum norm.

Theorem 2. Assume there exist $B \subseteq A \subseteq \Omega$, and $a > 0$ such that

$$r^s(y|x) \geq a_s\pi(y) \qquad \text{for all } x \in B \text{ and } y \in A \tag{8}$$

Assume further that the initial relative error is bounded i.e. $\left|p^0(x)/\pi(x) - 1\right| \leq R_M^0$ for all $x \in \Omega$. Then the probability density of the Markov chain satisfies

$$\left|\frac{p^{ns}(y)}{\pi(y)} - 1\right| \leq R_M^0 \left((1 - a_s b)^n + 2\frac{1-b}{b}\right) \quad \text{for } n \geq 0 \text{ and all } y \in A \tag{9}$$

where $b = \pi(B)$ and $0 < a_s b \leq 1$.

If there exists a sequence $\{(A_i, B_i)\}_i$ that satisfies (8) such that $b_i = \pi(B_i) \to 1$ when $i \to \infty$ and $A_i \supseteq B_i$ satisfy $\cup_{j>i} A_j = \Omega$ for all $i \in \mathbb{N}$, then the relative error vanishes pointwise, i.e. for all $y \in \Omega$ we have $|1 - p^i(y)/\pi(y)| \to 0$ when $i \to \infty$. If $A_i = \Omega$ for all $i > 0$, then the Markov chain converges in relative supremum norm i.e. $\sup_{y\in\Omega} |1 - p^i(y)/\pi(y)| \to 0$ when $i \to \infty$.

The theorem states that if the strong Doeblin condition is satisfied in sub-spaces $B \subseteq A \subseteq \Omega$, then the error in the relative supremum norm at least decreases geometrically to $(1 - b)/b = 1/\pi(B) - 1$ in $A$ relative to the initial error. Property (8) is weaker than that $B$ is a small set with measure $\pi$, since the equation is only satisfied for $y \in A$, not for all $y \in \Omega$.

If the Markov chain is $\phi$-irreducible and aperiodic, then according to Meyn and Tweedie (1993) Proposition 5.5.5 (iii) and Theorem 5.5.7, there exists a sequence of small sets $B_i$ such that $\cup_i B_i = \Omega$. However, these small sets will in general not satisfy (8) with $\pi$ as minorizing measure. This is shown in example 4. In that example the Markov chain is $\phi$-irreducible and aperiodic but there does not exist a sequence of small sets $B_i$ such that $\cup_i B_i = \Omega$. Theorem 2 may be used to prove that the pointwise relative error vanishes in that example.

PROOF Define $R^i_B = \sup_{x\in B} |R^i(x)|$ and

$$C_y := \{x \in \Omega : r^s(y|x) \geq a_s \pi(y)\}$$

for $y \in A$. Note that $B \subseteq C_y$. It follows from Theorem 1 that $|R^i(x)| \leq R^0_M$ for all $i \geq 0$ and $x \in \Omega$. Further calculation using the Proposition gives for $y \in A$

$$\begin{aligned}
R^{i+s}(y) &= \int_\Omega \frac{r^s(y|x)}{\pi(y)} R^i(x)\pi(x)\, d\mu(x) \\
&= R^i_B \int_\Omega \frac{r^s(y|x)}{\pi(y)}\pi(x)\, d\mu(x) - \int_\Omega \frac{r^s(y|x)}{\pi(y)}(R^i_B - R^i(x))\pi(x)\, d\mu(x) \\
&= R^i_B - \int_{C_y} \frac{r^s(y|x)}{\pi(y)}(R^i_B - R^i(x))\pi(x)\, d\mu(x) \\
&\quad - \int_{\Omega\setminus C_y} \frac{r^s(y|x)}{\pi(y)}(R^i_B - R^i(x))\pi(x)\, d\mu(x) \\
&\leq R^i_B - a_s \int_{C_y} (R^i_B - R^i(x))\pi(x)\, d\mu(x) + R^0_M a_s \int_{\Omega\setminus C_y} \pi(x)\, d\mu(x) \\
&= R^i_B(1 - a_s \int_{C_y} \pi(x)\, d\mu(x)) - a_s \int_{\Omega\setminus C_y} R^i(x)\pi(x)\, d\mu(x) \\
&\quad + R^0_M a_s \int_{\Omega\setminus C_y} \pi(x)\, d\mu(x) \\
&\leq R^i_B(1 - a_s \int_B \pi(x)\, d\mu(x)) + 2R^0_M a_s \int_{\Omega\setminus B} \pi(x)\, d\mu(x) \\
&= R^i_B(1 - a_s b) + 2R^0_M a_s(1-b).
\end{aligned}$$

Define $\tilde{p}^i$ such that the corresponding $\tilde{R}^i = -R^i$. Note that $\tilde{p}^i$ may be negative and thus not a density. Perform the same calculation as above with $\tilde{R}^i$ replacing $R^i$. This gives

$$\tilde{R}^{i+s}(y) \leq R^i_B(1 - a_s b) + 2R^0_M a_s(1-b)$$

which implies

$$|R^{i+s}(y)| \leq R^i_B(1 - a_s b) + 2R^0_M a_s(1-b)$$

for $y \in A \supseteq B$. Induction gives for $y \in A$

$$\begin{aligned}
|R^{ns}(y)| &\le R_M^0(1-a_s b)^n + 2R_M^0 a_s(1-b)\sum_{j=0}^{n-1}(1-a_s b)^j \\
&= R_M^0(1-a_s b)^n + 2R_M^0 a_s(1-b)\frac{1-(1-a_s b)^n}{1-(1-a_s b)} \\
&= R_M^0(1-a_s b)^n + 2R_M^0\frac{1-b}{b}(1-(1-a_s b)^n) \\
&\le R_M^0\left((1-a_s b)^n + 2\frac{1-b}{b}\right).
\end{aligned}$$

It is trivial to prove that $0 < 1 - a_s b \le 1$.

If there exists a sequence $\{(A_i, B_i)\}_i$ that satisfies (8) such that $b_i = \pi(B_i) \to 1$ when $i \to \infty$ and $A_i \supseteq B_i$ satisfy $\cup_{j>i} A_j = \Omega$, then for all $\varepsilon > 0$, there exists $I$ such that for $i > I$ implies $2(1-b_i)/b_i \le \varepsilon/2$. For all $y \in \Omega$, there exists $j > I$ such that $y \in A_j$ and there exists $N$ such that $n > N$ implies $(1 - a_{s,j} b_j)^{ns} < \varepsilon/2$. This implies that $R^{ns}(y) < \varepsilon R_M^0$. If $A_j = \Omega$, then $R^{ns}(y)$ may be replaced by $R_M^{ns}$ in the above expression.

□

## 5. Other norms

The theorems may be generalised to convergence results in other norms. So far we have used the relative supremum norm, $L_{\pi,\infty}$:

$$\|f\|_{\pi,\infty} = \sup_{x\in\Omega}\left|\frac{f(x)}{\pi(x)}\right|.$$

Define the $L_q$ norm, $q \in (0,\infty)$

$$\|f\|_q = \left(\int_\Omega f^q(x)\,d\mu(x)\right)^{1/q},$$

the supremum norm, $L_\infty$, as

$$\|f\|_\infty = \sup_{x\in\Omega}|f(x)|$$

and the total variation norm as

$$\|f\|_{\mathrm{TV}} = \sup_{C\subseteq\Omega}\left|\int_C f(x)\,d\mu(x)\right|.$$

It is well known and easy to prove that if $H = L_q, L_\infty$ or the total variation norm, then

$$\|f\|_H \le \|f\|_{\pi,\infty}\|\pi\|_H. \tag{10}$$

Hence, Theorem 1 may be used to prove convergence in other norms.

Athreya, Doss and Sethuraman (1996) give an example where there is not geometric convergence in the total variation norm. The example satisfies the detailed balance condition (4). According to Theorem 1 there is either geometric convergence or no convergence in relative supremum norm. Since the total variation norm is less than the relative supremum norm, there is at least one $p^0$ for which there is not convergence in the relative supremum norm in their example.

COROLLARY 1. Assume there exists a sequence $\{(A_i, B_i)\}_i$ that satisfies (8) such that $b_i = \pi(B_i) \to 1$ when $i \to \infty$ and $A_i$ satisfy $\cup_{j>i} A_j = \Omega$ for all $i \in \mathbb{N}$ and $\sup_{x\in\Omega} |1 - p^0(x)/\pi(x)|$ is bounded.

The Markov chain converges in $L_q$ norm if $\int_\Omega \pi^q(x)\, d\mu(x)$ is finite.

The Markov chain converges in $L_\infty$ norm if $\pi(x)$ is bounded in $\Omega$ and $\sup_{x\in\Omega\setminus B_i}\{\pi(x)\} \to 0$ when $i \to \infty$.

The Markov chain converges in total variation norm.

PROOF Convergence in $L_q$ norm is proved by

$$\begin{aligned}\|p^{ns} - \pi\|_q^q &= \int_\Omega |p^{ns}(x) - \pi(x)|^q\, d\mu(x) = \int_\Omega |R^{ns}(x)|^q \pi^q(x)\, d\mu(x) \\ &\leq \left(R_M^0\right)^q \left( ((1 - a_{s,i} b_i)^n + 2\frac{1-b_i}{b_i})^q \int_{B_i} \pi^q(x)\, d\mu(x) + \int_{\Omega\setminus B_i} \pi^q(x)\, d\mu(x)\right)\end{aligned}$$

which may be made arbitrarily small by choosing $b_i$ close to 1 and $n$ large.

The $L_\infty$ norm is expressed as

$$\|p^{ns} - \pi\|_\infty = \sup_{x\in\Omega}\{R^{ns}(x)\pi(x)\} \leq R_M^0 \max\{4\frac{1-b_i}{b_i} \sup_{x\in\Omega}\{\pi(x)\}, \sup_{x\in\Omega\setminus B_i}\{\pi(x)\}\},$$

for sufficient large $n$ dependent on $b_i$. When $i \to \infty$, the expression above vanishes which implies convergence in $L_\infty$.

The total variation norm may be expressed as

$$\begin{aligned}\|p^{ns} - \pi\|_{\mathrm{TV}} &= \sup_{C\subseteq\Omega} \int_C R^{ns}(x)\pi(x)\, d\mu(x) \\ &= \sup_{C\subseteq\Omega} \left\{ \int_{C\cap B_i} R^{ns}(x)\pi(x)\, d\mu(x) + \int_{C\cap(\Omega\setminus B_i)} R^{ns}(x)\pi(x)\, d\mu(x) \right\}.\end{aligned}$$

The Markov chain converges in total variation by the same argument as for the $L_\infty$ norm given above. □

## 6. EIGENVALUES

It is also possible to express the convergence in the form of eigenvalues of the transition density $r(y|x)$. Since the algorithm converges, the absolute value of all eigenvalues except for the eigenvalue which correspond to the limiting distribution $\pi(x)$ is less than 1. The eigenvalue with the next largest absolute value determines the convergence rate. Note however that in the infinite dimensional case it is natural to use spectrum instead of eigenvalues. In the non-degenerate case the spectrum is typically an interval with 1 as the right hand end of the interval, see Example 4. Hence, it will have values arbitrary close to 1 and corresponding slow convergence. If the strong Doeblin condition is satisfied, it is possible to bound the convergence in terms of eigenvalues/spectrum. The following corollary is proved. Example 3 shows however that the bound is not necessarily optimal.

COROLLARY 2. If the strong Doeblin condition (3) is satisfied, then any solution $\lambda$ and $v$ of the equation

$$\lambda v(y) = \int_\Omega r(y|x) v(x)\, d\mu(x) \tag{11}$$

satisfies either $\lambda = 1$ and $v = \pi$, or $|\lambda| \le (1-a_s)^{1/s}$.

Proof Let us first observe that any solution $(\lambda, v)$ of (11) with $\lambda \ne 1$ satisfies $\int_\Omega v(x)\,d\mu(x) = 0$, since

$$\lambda \int_\Omega v(y) d\mu(y) = \int_\Omega \left(\int_\Omega r(y|x) d\mu(y)\right) v(x)\,d\mu(x) = \int_\Omega v(x)\,d\mu(x).$$

This implies $\int_\Omega v(x)\,d\mu(x) = 0$.

Assume $p^0 = \pi + v$ where $v$ satisfies (11). Then Theorem 1 states that for all $y \in \Omega$

$$|\lambda^s| \left|\frac{v(y)}{\pi(y)}\right| \le (1-a_s) \sup_{x\in\Omega} \left|\frac{v(x)}{\pi(x)}\right|.$$

Hence $|\lambda| \le (1-a_s)^{1/s}$. □

## 7. Some examples

Example 1. Let $\Omega = (0,1)$, $\pi(x) \equiv 1$ and

$$r(y|x) = \begin{cases} (1-d)/c & \text{if } |x-y|_1 < c/2, \\ d/(1-c) & \text{otherwise,} \end{cases}$$

where $0 \le d < 1$, $0 < c < 1$ and $|x|_1 = \min_{i\in\mathbb{Z}}\{|x+i|\}$. For $d > 0$, the strong Doeblin condition is satisfied with $s = 1$ and $a = d/(1-c)$ for $c \le 1-d$. If $d$ is small, this gives slow convergence. For $d = 0$ or $d$ small it may be better to choose $s > 1$. In order to find $a_s$ for $s > 1$, it is natural to use the approach in Holden (1998). That is, for given $x, y \in \Omega$, define possible sequences for jumping from $x$ to $y$. Define the sequence $\{D_i\}_{i=0}^s$ such that $D_0 = \{x\}$ $D_s = \{y\}$ and for any $u \in D_i$, $v \in D_{i+1}$, $|u-v|_1 < c/2$. This gives for $s > 1/c$ that $a_s = ((1-d)/c)((1-d)(cs-1)/(2cs))^{s-1}$. This is not the optimal choice for all values of $s$, but it combines a reasonably good choice with simplicity of calculation.

Assume

$$p_k^0(x) = \begin{cases} 2 & \text{for } 2j2^{-k} < x \le (2j+1)2^{-k} \text{ for } j \in \mathbb{N}_0 \\ 0 & \text{otherwise} \end{cases}$$

for $k \in \mathbb{N}$. The initial error $\sup_{0<x<1} |p_k^0(x) - \pi(x)| = 2$ independent of $k$. However, the error $\sup_{0<x<1} |p_k^i(x) - \pi(x)|$ decreases faster as a function of $i$ the larger $k$ is, since high frequency errors decreases faster than low frequency errors. The dependency on $k$ is larger the smaller $c$ and $d$ are.

Example 2. The Metropolis–Hastings algorithm generates a sample from the density $\pi$ by:

1. Generate an initial state $x^0 \in \Omega$ from the density $p^0$.
2. For $i = 0, \ldots, n$:
   (a) Generate a state $y$ from the density $q(\cdot|x^i)$.
   (b) Calculate $\alpha(y, x^i) = \min\left\{1, \frac{\pi(y)q(x^i|y)}{\pi(x^i)q(y|x^i)}\right\}$.

(c) Set $x^{i+1} = \begin{cases} y & \text{with probability } \alpha(y, x^i) \\ x^i & \text{with probability } 1-\alpha(y, x^i). \end{cases}$

The Metropolis–Hastings algorithm satisfies always the detailed balance condition. It satisfies the strong Doeblin condition (3) with the same $a$ with $s = 1$ if $q(y|x) \geq a\pi(y)$ for all $x, y \in \Omega$ since

$$r(y|x) \geq \alpha(y,x)q(y|x) = \min\left\{q(y|x), \frac{\pi(y)}{\pi(x)}q(x|y)\right\} \geq a\pi(y).$$

It is also possible to use weaker assumptions. Given $x, y \in \Omega$, define possible sequences for jumping from $x$ to $y$ by defining the sequence $\{D_i\}_{i=0}^{s}$ such that $D_0 = \{x\}$, $D_s = \{y\}$ and for any $u \in D_i$ and $v \in D_{i+1}$, satisfy $q(v|u) \geq a_i\pi(v)$ and $q(u|v) \geq a_i\pi(u)$. This gives

$$r^s(y|x) \geq \pi(y)\prod_{i=1}^{s}\left(a_i \int_{D_i} \pi(x)\, d\mu(x)\right)$$

which satisfies the strong Doeblin condition for sufficient large values of $s$. This is discussed in more detail in Holden (1998).

Example 3. This example shows that the bound on the eigenvalue in Corollary 3 is not always optimal. Assume the state space consists of $n$ points with limiting distribution $(\pi_1, \pi_2, ..., \pi_n)$ and with transition matrix

$$Q = \begin{bmatrix} \pi_1 + (1-a)\pi_2 & \pi_2 - (1-a)\pi_2 & \pi_3 & \ldots & \pi_n \\ \pi_1 - (1-a)\pi_1 & \pi_2 + (1-a)\pi_1 & \pi_3 & \ldots & \pi_n \\ \pi_1 & \pi_2 & \pi_3 & \ldots & \pi_n \\ \ldots. & \ldots & \ldots & \ldots & \ldots \\ \pi_1 & \pi_2 & \pi_3 & \ldots & \pi_n \end{bmatrix}$$

where $0 \leq a < 1$. This transition matrix satisfies the strong Doeblin condition $Q_{i,j} = r(i|j) \geq a\pi_i$ and has eigenvectors $(\pi_1, \pi_2, ..., \pi_n)$ and $(1, -1, 0, 0, ..., 0)$ with eigenvalues 1 and $(1-a)(\pi_1 + \pi_2)$, respectively and $n-2$ eigenvectors with eigenvalue 0. In this example the upper bound given in Corollary 2 is optimal for $n = 2$ but not optimal for $n > 2$.

Example 4 Define the following random walk on $\mathbb{Z}$

$$r^1(j|i) = \begin{cases} 0 & \text{for } |j-i| > 1 \\ 1/4 & \text{for } j = i \text{ if } i \neq 0 \\ 1/4 & \text{for } j = i \pm 1 \text{ and } (j-i)i \geq 0 \\ 1/2 & \text{otherwise} \end{cases}$$

The limiting distribution for the random walk is $\pi^i = p^i = 2^{-|i|}/3$. This Markov chain satisfies the detailed balance condition (4) but does not satisfy the strong Doeblin condition and hence does not converge in the relative supremum norm. Theorem 2 and Corollary 1 may be used to prove convergence in other norms. Let $s = 2^i$, $A_i = B_i = \{j \in \mathbb{Z}; |j| \leq i\}$ and $a_i = 2^{-3i}$. Then $b_i = 1 - 2^{-i+2}/3$ for $i > 0$. This gives

$$|R^{ns}(y)| \leq R_M^0((1-2^{-3i+1})^n + 2^{-i+3}) \qquad \text{for } |y| \leq i$$

i.e. the pointwise relative error vanishes. Corollary 1 gives convergence in $L_q$, $L_\infty$, and total variation norm.

Any values of $p_{-1}, p_0, p_1$ different from zero defines a solution $\lambda$ and $v = (..., p_{-1}, p_0, p_1, ...)$ of (11) defined by the equations

$$\begin{aligned}\lambda &= (1 + p_{-1}/p_0 + p_1/p_0)/2 \\ p_{i+1} &= -p_{i-1}/2 + (2\lambda - 1/2)p_i \text{ for } i > 0 \\ p_{i-1} &= -p_{i+1}/2 + (2\lambda - 1/2)p_i \text{ for } i < 0\end{aligned}$$

$p_i$, $i \in \mathbb{Z}$ bounded for $-1/2 \leq \lambda \leq 1$. Note that $(-1/2, 1]$ is part of the spectrum of the transition density $r(j|i)$.

EXAMPLE 5 Let $\Omega = \mathbb{R}$ and

$$r(y|x) = \begin{cases} \exp(-|x-y|) & \text{for } x(y-x) < 0 \\ 0 & \text{otherwise} \end{cases}$$

The limiting distribution is $\pi(x) = \frac{1}{2}\exp(-|x|)$. This Markov chain does not satisfy the detailed balance condition (4) since

$$\exp(-|x-y| - |y|) >> \exp(-|x-y| - |x|) \qquad \text{for } |y| << |x|$$

nor the strong Doeblin condition since

$$\frac{r^s(0|x)}{\pi(0)} \to 0 \qquad \text{when } |x| \to \infty.$$

This is proved by

$$\begin{aligned} r^s(x_s|x_0) &= \int \cdots \int \prod_{i=1}^{s} r(x_i|x_{i-1}) dx_1 \ldots dx_{s-1} \\ &= \int \cdots \int \exp(-\sum_{i=1}^{s} |x_i - x_{i-1}|) dx_1 \ldots dx_{s-1} \end{aligned}$$

The integration area is divided into a disjoint union $\{I_j\}_{j=0}^\infty$ where

$$I_j = \{\{x_i\}_{i=1}^{s-1}; \max |x_i| \in [jx_0, (j+1)x_0)\}$$

The area of the integration in $I_j$ is bounded by $((j+1)x_0)^{s-1}$ and for $\{x_i\}_{i=1}^{s-1} \in I_j$ and $x_s = 0$, we have $\exp(-\sum_{i=1}^{s} |x_i - x_{i-1}|) \leq \exp(-2sjx_0)$.

This gives

$$r^s(0|x_0) \leq \sum_{j=0}^{\infty} ((j+1)x_0)^{s-1} \exp(-2sjx_0) \to 0$$

when $x_0 \to \infty$ for all $s \in \mathbb{N}$.

Theorem 2 may be used to show that the Markov chain converges in the relative supremum norm by choosing $s = 2$, $A = \Omega$ and $B_i = \{x \in \Omega; |x| < i\}$. Then $b_i \to 1$ and $R_M^i \to 0$ when $i \to \infty$. This example shows that the detailed balance condition is a necessary condition in order to have equivalence between the strong Doeblin condition and convergence of Markov chains in the relative supremum norm. The Markov chain converges also in $L_q, L_\infty$ and total variation norm according to (10).

**Acknowledgement.** The author thanks the referees and Fred Espen Benth for valuable comments.

## References


Athreya, K. B., Doss, H., and Sethuraman, J. (1996) On the convergence of the Markov chain simulation method. *Annals of Statistics* **24**, 69–100

Doob, J. L. (1953) *Stochastic Processes.* Wiley, New York.

Geyer, C.J. and Møller, J. (1993) Simulation procedures and likelihood inference for spatial point processes, Technical Report 260, Dept. of Theor. Stat., Institute of Mathematics, University of Aarhus.

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

Holden, L. (1998) Geometric convergence of the Metropolis–Hastings simulation algorithm. Statistics and Probability Letters **39**, 371-377.

Mengersen, K. L. and Tweedie, R. L. (1996) Rates of Convergence of the Hastings and Metropolis Algorithms. *The Annals of Statistics* **24**, 101–121.

Meyn, S. P. and Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability.* Springer-Verlag, London.

Ripley, B. D., (1987) *Stochastic Simulation,* John Wiley & Sons, New York.

Royden H.L., (1968) *Real Analysis* , Macmillian, New York.

Stewart, W. J., (1994) Introduction to the Numerical Solution of Markov Chains, Princeton University Press.



Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway

E-mail address: Lars.Holden@nr.no