GEOMETRIC CONVERGENCE OF A GENERAL MARKOV CHAIN

LARS HOLDEN
NORWEGIAN COMPUTING CENTER AND UNIVERSITY OF OSLO

ABSTRACT. It is proved that a general Markov chain converges in the relative supremum norm if and only if the Doeblin property (i.e., $r^s(y|x) \geq a_s\pi(y)$ for all x, y in the state space) is satisfied, where $r^s(y|x)$ is the s-step transition probability density. The convergence is geometric with ratio $(1-a_s)^{1/s}$. This result is used to prove convergence in other norms under weaker assumptions. The state space may be either continuous or discrete. The results in the paper also give a qualitative understanding of the convergence.

1. Introduction

Markov chains are widely used as models and computational devices in areas ranging from statistics to physics. A chain starts in an initial state x and at each iteration it moves to another state y according to the transition function $r(y|x)$. Under mild conditions the probability distribution after i iterations, $p^i(x)$, approaches a limiting function $\pi(x)$.

This paper shows that the Doeblin property for one value of s, i.e., $r^s(y|x) \geq a_s\pi(y)$ for all x, y in the state space, is a necessary and sufficient criterion for the convergence of the chain in the relative supremum norm. The function $r^s(y|x)$ is the s-step transition function.

We also show that this norm is a natural norm for proving convergence. Convergence in the relative supremum norm is always geometric, and it may be used to prove geometric convergence in other norms also. Other norms may, however, converge when the relative supremum norm does not. The relative supremum norm emphasizes the areas where $\pi(x)$ is small. If the tails of the distribution are not important, other norms may be better.

This paper is a generalization of the results in Holden (1996) which proves similar theorems for the Metropolis–Hastings simulation algorithm. Mengersen and Tweedie (1996) give another proof for geometric convergence of this algorithm.

The theory and applications of Markov chains are very active fields of research: see, for example, Meyn and Tweedie (1993) and Geyer (1992).

2. A general Markov chain

Let $\Omega \subset \mathbb{R}^n$ be a Borel measurable state space and $\pi(x)$ a probability density such that $\int_\Omega \pi(x)\,dx = 1$ or, alternatively, let Ω be a discrete state

Date: April 7, 1997.
1991 *Mathematics Subject Classification.* 60J27;
Key words and phrases. Markov chain, convergence.
Research supported by Research Council of Norway.

space and $\pi(x)$ probabilities on this state space. All the results and proofs are valid both for absolute continuous densities and for probabilities on a discrete state space. The continuous state space terminology will be used in this paper. If the state space is discrete, the integral signs should be replaced by summation signs. The initial probability densities $p^0(x)$ and the transition function $r(x|y)$, $x, y \in \Omega$ are positive in Ω or in a subset of Ω. The integrals of these densities over Ω or subspace of Ω are assumed equal to 1. All the densities are assumed to be absolutely continuous or to be probabilities on a discrete state space. A Markov chain is defined as follows.

MARKOV CHAIN. To generate a sample from the probability density $\pi(x)$:

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

Let us define the s-step transition function by

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

The 1-step transition function is the ordinary transition function $r^1(y|x) = r(y|x)$. In this paper it is assumed that there exists a density $\pi(x)$ which satisfies

$$\pi(y) = \int_\Omega r^s(y|x)\pi(x)\,dx \tag{2}$$

for all integers $s > 0$. This is satisfied if the chain is recurrent; see Meyn and Tweedie (1993). Since our estimate is based on the relative supremum norm it is necessary to assume that $\pi(x) > 0$ for $x \in \Omega$.

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

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

This condition is central for the present paper. See Doob (1953), p. 197, for references.

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 $L_{\pi,\infty}$, i.e., $R^i_M = \sup_{x\in\Omega} \left|R^i(x)\right|$. The following proposition is used in several of the proofs.

PROPOSITION. *The pointwise relative error satisfies*

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

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 function which is large only for $|x - y|$ small. This Proposition may then be used to prove that high-frequency error in $R^i(x)$ decreases much faster than low-frequency errors. The proof of the Proposition is short.

Proof. Combining (1) and (2) gives

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

which implies

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

which proves the Proposition. □

3. Positive generation function

In order to ensure convergence we must formulate the theorem for sufficiently large number of steps that the transition function is always positive. The convergence result is then as follows.

Theorem 1. *Assume the Doeblin property (3) is satisfied and that $\pi(x) > 0$ for $x \in \Omega$. Then the probability density of the Markov chain satisfies for $y \in \Omega$*

$$\left|\frac{p^{i+s}(y)}{\pi(y)} - 1\right| \le (1 - a_s) \sup_{x\in\Omega}\left\{\left|\frac{p^i(x)}{\pi(x)} - 1\right|\right\}. \tag{4}$$

If there do not exist an integer $s > 0$ and a constant $a_s \in (0,1]$ such that the Doeblin property (3) is satisfied for $x, y \in \Omega$, then there exists $\epsilon > 0$ and $p^0(x)$ such that

$$\sup_{x\in\Omega}\left|\frac{p^j(x)}{\pi(x)} - 1\right| = \epsilon \tag{5}$$

for all $j \ge 0$.

The theorem states that $R^i(y) = |(p^i(y)/\pi(y)) - 1|$ does not increase and that the Markov chain converges geometrically if $a_s > 0$. The convergence is fast if $r^s(y|x) \approx \pi(y)$ and immediate if $r^s(y|x) = \pi(y)$. The Doeblin property (3) with $a_s > 0$ is a necessary and sufficient condition for convergence in the relative supremum norm.

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

Very often $r(y|x) = 0$ for a particular state x and for y in a large set. Then it is necessary to use several steps (i.e., $s > 1$) such that the Doeblin property (3) is satisfied with $a_s > 0$. As stated in Theorem 1, the existence of s and $a_s > 0$ is a necessary condition for convergence. A method for finding a_s and s is shown in Holden (1996). In that method it is necessary to specify a possible chain between any two states. The size of a_s depends on how probable this chain is. The decrease in the relative supremum norm per iteration is $(1-a_s)^{1/s}$. If s is increased, it is often possible to increase a_s. In some cases this increase is sufficiently large that this also gives a better estimate for the convergence even if $a_s > 0$ for the lower value of s. This is illustrated in Example 1.

In Doob (1953) it is proved that the Doeblin property implies convergence in the total variation norm, i.e., $\sup_{A\subset\Omega}\left|\int_A p^i(x) - \pi(x)\,dx\right|$ where $p^0(x)$ is assumed to be a $\delta(x)$ function. A $\delta(x)$ function cannot be handled in the relative supremum norm. One way to get around this problem is to go through a few iterations before applying the theorem.

In Section 5 it is shown how this result may be used for proving geometric convergence in other norms too. Convergence in the relative supremum norm implies convergence in most other norms.

In Diaconis (1996), cutoff phenomena for finite Markov chains are studied. It is assumed that the Markov chain starts in one state, i.e., $p^0(x)$ is a δ function. In the observed chains the total variation norm stays close to 1 for a certain number of iterations, then suddenly drops and then tends to zero exponentially fast. This phenomena may be explained using the above theorem. Naturally there is no reduction in the relative supremum norm and only a small reduction in the total variation norm before a sufficient number of iterations have been performed that it is possible to reach a significant amount of the state space. Then the Doeblin property (3) is satisfied with $a_s > 0$ and we get geometric convergence in the relative supremum norm. This implies that there is also geometric convergence in the total variation norm, according to Corollary 1 in Section 5.

Proof. The Proposition gives

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

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

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

which is equivalent to

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

This proves that (3) implies (4). It remains to prove the implication in the other direction.

Choose $y \in \Omega$, $a \in (0,1)$ and $s > 0$. Define $A^{a,s}_y = \{x \in \Omega; r^s(x|y) \ge a\pi(x)\}$. Assume $\int_{A^{a,s}_y} \pi(x)\,dx > 0$. If this is not the case, it is trivial to show that the theorem is satisfied. Let $A_y = \sup_{a,s} A^{a,s}_y$ and

$$p^0(x) = \begin{cases} (1+\epsilon)\,\pi(x) & \text{for } x \in \Omega \setminus A_y, \\ (1-\beta\epsilon)\,\pi(x) & \text{else,} \end{cases}$$

where $\epsilon, \beta > 0$ are determined such that $\int_\Omega p^0(x)\,dx = 1$ and $1 - \beta\epsilon > 0$.

Assume first $A_y \neq \Omega$. A chain with $x^0 \in A_y$ does not join $\Omega \setminus A_y$ for any s. Then $p^j(x) = (1+\epsilon)\,\pi(x)$ for $x \in \Omega \setminus A_y$ and all $j \geq 0$, which implies (4).

Assume then $A_y = \Omega$. Choose $a > b > 0$, $s > 0$ and $v \in \Omega \setminus A_y^{b,2s}$. Then for sufficiently small a

$$\begin{aligned} b\pi(v) \geq r^{2s}(v|y) &= \int_\Omega r^s(v|x) r^s(x|y)\,dx \\ &\geq \int_{A_y^{a,s}} r^s(v|x) r^s(x|y)\,dx \\ &\geq a \int_{A_y^{a,s}} r^s(v|x)\pi(x)\,dx. \end{aligned}$$

It is possible to bound the same integrand in the domain $\Omega \setminus A_y^{a,s}$ using the calculation

$$\begin{aligned} \pi(v) &= \int_\Omega r^s(v|x)\pi(x)\,dx \\ &= \int_{A_y^{a,s}} r^s(v|x)\pi(x)\,dx + \int_{\Omega\setminus A_y^{a,s}} r^s(v|x)\pi(x)\,dx \\ &\leq \frac{b}{a}\pi(v) + \int_{\Omega\setminus A_y^{a,s}} r^s(v|x)\pi(x)\,dx \end{aligned}$$

which implies that

$$\pi(v)\left(1 - \frac{b}{a}\right) \leq \int_{\Omega\setminus A_y^{a,s}} r^s(v|x)\pi(x)\,dx.$$

The definition of $p^0(x)$ gives

$$\begin{aligned} p^s(v) - \pi(v) &= \int_\Omega r^s(v|x)(p^0(x) - \pi(x))\,dx \\ &= \int_{A_y^{a,s}} r^s(v|x)(p^0(x) - \pi(x))\,dx \\ &\quad + \int_{\Omega\setminus A_y^{a,s}} r^s(v|x)(p^0(x) - \pi(x))\,dx \\ &= -\beta\epsilon \int_{A_y^{a,s}} r^s(v|x)\pi(x)\,dx + \epsilon \int_{\Omega\setminus A_y^{a,s}} r^s(v|x)\pi(x)\,dx \\ &\geq -\frac{\beta\epsilon b}{a}\pi(v) + \epsilon\pi(v)\left(1 - \frac{b}{a}\right) \\ &= \pi(v)\epsilon\left(1 - \frac{\beta b}{a} - \frac{b}{a}\right). \end{aligned}$$

Since we may chose b arbitrarily small, this implies

$$\sup_{x\in\Omega} \frac{p^s(v)}{\pi(v)} \geq 1 + \epsilon.$$

Using the first part of the theorem with $a = 0$ gives

$$\sup_{x\in\Omega} \left|\frac{p^j(x)}{\pi(x)} - 1\right| = \epsilon.$$

□

4. Vanishing generation function

In the previous section it was proved that if the Doeblin property is satisfied with $a > 0$, then the Markov chain converges geometrically in the relative supremum norm. In this section the relative error is bounded when the Doeblin property is satisfied in only part of Ω.

Theorem 2. *Let $z \in \mathbb{R}$, $z > 0$ and define $B_z \subset \Omega$ as $B_z = \{x \in \Omega; |x| \leq z\}$. Assume that for all $z > 0$*

$$r^s(x|y) \geq a_z \pi(x) \qquad \text{for all } x, y \in B_z \text{ with } a_z > 0,$$

$$\int_{B_z} \pi(x)\, dx \geq b_z > 0$$

and

$$\left| \frac{p^0(x)}{\pi(x)} - 1 \right| \leq R_M^0 \qquad \text{for all } x \in \Omega.$$

Then the probability density of the Markov chain satisfies for $y \in B_z$

$$\left| \frac{p^{ns}(y)}{\pi(y)} - 1 \right| \leq R_M^0 \left((1 - a_z b_z)^n + 2 \frac{1 - b_z}{b_z} \right)$$

where $0 < 1 - a_z b_z < 1$.

This theorem proves that if the Doeblin property is satisfied in a subspace $B_z \subset \Omega$, then the error in the relative supremum norm at least decreases geometrically to $(1-b_z)/b_z \leq -1+1/\int_{B_z} \pi(x)\, dx$ in B_z relative to the initial error.

Proof. Define $R_{z,M}^i(x) = \sup_{x \in B_z} |R^i(x)|$ and $B_z \subset C_{z,y}$ where

$$C_{z,y} = \{x \in \Omega; r^s(x|y) \geq a_z \pi(x)\}.$$

It follows from the previous theorem that $|R^i(x)| \leq R_M^0$ for all $i \geq 0$ and $x \in \Omega$. Further calculation using the Proposition gives

$$\begin{aligned}
R^{i+s}(y) &= \int_\Omega \frac{r^s(y|x)}{\pi(y)} R^i(x) \pi(x)\, dx \\
&= R_M^i \int_\Omega \frac{r^s(y|x)}{\pi(y)} \pi(x)\, dx - \int_\Omega \frac{r^s(y|x)}{\pi(y)} (R_M^i - R^i(x)) \pi(x)\, dx \\
&= R_M^i - \int_{C_{z,y}} \frac{r^s(y|x)}{\pi(y)} (R_M^i - R^i(x)) \pi(x)\, dx \\
&\quad - \int_{\Omega \setminus C_{z,y}} \frac{r^s(y|x)}{\pi(y)} (R_M^i - R^i(x)) \pi(x)\, dx \\
&\leq R_M^i - a_z \int_{C_{z,y}} (R_M^i - R^i(x)) \pi(x)\, dx + R_M^0 a_z \int_{\Omega \setminus C_{z,y}} \pi(x)\, dx \\
&= R_M^i (1 - a_z \int_{C_{z,y}} \pi(x)\, dx) - a_z \int_{\Omega \setminus C_{z,y}} R^i(x) \pi(x)\, dx \\
&\quad + R_M^0 a_z \int_{\Omega \setminus C_{z,y}} \pi(x)\, dx \\
&\leq R_M^i (1 - a_z \int_{B_z} \pi(x)\, dx) + 2 R_M^0 a_z \int_{\Omega \setminus B_z} \pi(x)\, dx \\
&= R_M^i (1 - a_z b_z) + 2 R_M^0 a_z (1 - b_z).
\end{aligned}$$

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

$$\tilde{R}^{i+s}(y) \leq R^i_M(1 - a_z b_z) + 2R^0_M a_z(1 - b_z)$$

which implies

$$|R^{i+s}(y)| \leq R^i_M(1 - a_z b_z) + 2R^0_M a_z(1 - b_z).$$

Induction gives

$$\begin{aligned}
|R^{ns}(y)| &\leq R^0_M(1 - a_z b_z)^n + 2R^0_M a_z(1 - b_z)\sum_{j=0}^{n-1}(1 - a_z b_z)^j \\
&= R^0_M(1 - a_z b_z)^n + 2R^0_M a_z(1 - b_z)\frac{1 - (1 - a_z b_z)^n}{1 - (1 - a_z b_z)} \\
&= R^0_M(1 - a_z b_z)^n + 2R^0_M \frac{1 - b_z}{b_z}(1 - (1 - a_z b_z)^n) \\
&< R^0_M\left((1 - a_z b_z)^n + 2\frac{1 - b_z}{b_z}\right).
\end{aligned}$$

□

5. Other norms

The theorem may be generalized to a convergence result in other norms also. So far we have used the relative supremum norm, $L_{\pi,\infty}$:

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

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

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

The supremum norm, $p = \infty$, is expressed as

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

The total variation norm is defined as

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

We can now formulate the following corollary to the first theorem.

Corollary 1. *Assume that $\pi(x) > 0$ is satisfied for $x \in \Omega$. Then the Doeblin property (3) is a necessary and sufficient condition for convergence in the relative supremum norm. The convergence is geometric with ratio $(1 - a_s)^{1/s}$.*

The Doeblin property (3) and the boundedness of $\int_\Omega \pi^q(x)\,dx$ imply geometric convergence in L_q norm, $q \in (0, \infty)$ with ratio $(1 - a_s)^{1/s}$.

The Doeblin property (3) and the boundedness of $\pi(x)$ for $x \in \Omega$ imply geometric convergence in L_∞ norm with ratio $(1 - a_s)^{1/s}$.

The Doeblin property (3) implies geometric convergence in total variation norm, for any $p^0(x)$ with ratio $(1 - a_s)^{1/s}$.

Proof. The theorem implies that $R_M^{i+s} \leq (1-a_s)R_M^i$, which gives

$$R_M^i \leq (1-a_s)^{(i-s+1)/s} R_M^0$$

for $i \geq s$.

Convergence in L_q norm is proved by

$$\int_\Omega |p^i(x) - \pi(x)|^q \, dx = \int_\Omega \left| \frac{p^i(x) - \pi(x)}{\pi(x)} \right|^q \pi^q(x) \, dx \leq \left(R_M^i\right)^q \int_\Omega \pi^q(x) \, dx.$$

Combined with the previous equation this gives

$$\|p^i - \pi\|_q \leq (1-a_s)^{(i-s+1)/s} R_M^0 \Big(\int_\Omega \pi^q(x) \, dx \Big)^{1/q}.$$

Convergence in L_∞ is handled by

$$\|p^i - \pi\|_\infty \leq (1-a_s)^{(i-s+1)/s} R_M^0 \sup_{x \in \Omega} \{\pi(x)\}.$$

Convergence in total variation norm is proved by

$$\left| \int_C p^i(x) - \pi(x) \, dx \right| = \left| \int_C \frac{p^i(x) - \pi(x)}{\pi(x)} \pi(x) \, dx \right| \leq R_M^i \left| \int_C \pi(x) \, dx \right| \leq R_M^i$$

which implies

$$\|p^i - \pi\|_{\mathrm{TV}} = \sup_{C \subset \Omega} \left| \int_C p^i(x) - \pi(x) \, dx \right| \leq R_M^i \leq (1-a_s)^{(i-s+1)/s} R_M^0.$$

□

Athreya, Doss and Sethuraman (1996) give an example where there is not geometric convergence in the total variation norm. According to the above theorem 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 not convergence in the relative supremum norm in their example.

There is also a similar corollary to Theorem 2.

Corollary 2. *Make the same assumptions as in Theorem 2.*

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

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

The Markov chain converges in total variation norm.

Proof. Convergence in L_q norm is proved by

$$\begin{aligned} \|p^{ns} - \pi\|_q^q &= \int_\Omega |p^{ns}(x) - \pi(x)|^q \, dx = \int_\Omega |R^{ns}(x)|^q \pi^q(x) \, dx \\ &\leq \left(R_M^0\right)^q \left((1 - a_z b_z)^n + 2 \frac{1-b_z}{b_z} \int_{B_z} \pi^q(x) \, dx + \int_{\Omega \setminus B_z} \pi^q(x) \, dx \right) \end{aligned}$$

which may be made arbitrarily small by choosing z and n large.

The L_∞ norm is expressed as

$$\|p^{ns} - \pi\|_\infty = \sup_{x \in \Omega} \{R^{ns}(x) \pi(x)\}.$$

The Markov chain convergences in L_∞ since inside any B_z, $R^{ns}(x)$ approaches zero uniformly and $\pi(x)$ is uniformly bounded. In $\Omega \setminus B_z$, $R^{ns}(x)$ is bounded and $\pi(x)$ approaches zero as z increases by the assumption in the corollary.

The total variation norm may be expressed as

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

The Markov chain convergences in total variation by the same argument as for the L_∞ norm given above. □

6. Eigenvectors

It is also possible to express the convergence in the form of eigenvalues of the operator Q defined by one step of the Markov chain with $r(y|x)$ as the transition probability. 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. The following corollaries are proved.

Corollary 3. *If the Doeblin property (3) is satisfied for all $x, y \in \Omega$, the next largest eigenvalue in absolute value of the operator Q is less than or equal to $(1-a_s)^{1/s}$.*

Proof. Let us first prove that any other eigenvector v with eigenvalue λ of Q must satisfy $\int_\Omega v(x)\,dx = 0$ since

$$\lambda v(y) = \int_\Omega r(y|x)v(x)\,dx$$

and hence

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

This implies that we must have either $v(x) = \pi(x)$ and $\lambda = 1$ or $\int_\Omega v(x)\,dx = 0$.

Assume $p(x) = \pi(x) + v(x)$ for an eigenvector $v(x)$. Then we get from the Theorem 1 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 $\lambda \le (1-a_s)^{1/s}$. □

Example 4 shows that this is not necessarily an optimal bound of λ.

Corollary 4. *If there do not exist s and $a_s > 0$ such that the Doeblin property (3) is satisfied for all $x, y \in \Omega$ and all probability densities in a space $H(\Omega)$ may be written in the form*

$$p(x) = \pi(x) + \sum_i c_i v_i(x)$$

where v_i are eigenvectors of Q, and scaled such that $\sup_{x\in\Omega}|v_i(x)/\pi(x)| = 1$ and $\sum_i c_i$ bounded, then there are eigenvalues of Q arbitrarily close to 1 in absolute value.

Proof. Let

$$p(x) = \pi(x) + \sum_i c_i v_i(x)$$

and assume $|\lambda_i| < b < 1$. Then

$$\sup_{x\in\Omega}\left|\frac{p^j(x)}{\pi(x)} - 1\right| = \sup_{x\in\Omega}\left|\sum_i c_i\lambda_i^j \frac{v_i(x)}{\pi(x)}\right| \le \sum_i c_i\lambda_i^j \sup_{x\in\Omega}\left|\frac{v_i(x)}{\pi(x)}\right| \le b^j \sum_i c_i$$

which may be made arbitrarily small by choosing j sufficient large.

According to Theorem 1 there exists a $p^0(x)$ such that $\sup_{x\in\Omega}|p^j/\pi(x) - 1| = \epsilon$ for all j. Hence the assumption that $|\lambda_i| < b < 1$ cannot be correct, which proves the corollary. □

7. Some examples

Example 1. Let $\Omega = (0, c)$, $\pi(x) = 1/c$ where $c > 0$ and

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

where $0 \le d < 1$, $e < c$ and $|x|_c = \min_{i\in N}\{|x + ic|\}$. For $d > 0$, the Doeblin property is satisfied with $s = 1$ and $a = dc/(c-e)$. If d is small, this gives very 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 (1996). That is, for given $x, y \in \Omega$, define possible sequences for jumping from x to y. Define the sequence $\{D\}_{i=0}^s$ such that $D_0 = \{x\}$ $D_s = \{y\}$ and for any $u \in D_i$ $v \in D_{i+1}$, $|u - v|_c < 1/(2e)$. This gives for $s > c/e$ that $a_s = ((1-d)/e)((1-d)(es-c)/(2ces))^{s-1}$. This is not the optimal choice for all values of s, but it combines a reasonably good choice with simplicity of calculation.

Example 2. Let $\Omega = R$, $\pi(x) \le M$ for all $x \in \Omega$ and

$$r(y|x) = \begin{cases} \frac{1}{2} & \text{if } |x-y| < 1, \\ 0 & \text{else.} \end{cases}$$

Using Example 1 we get $r^s(y|x) \ge a_s\pi(y)$ for $y \le z$, $s > 2z$ and $a_s = 1/(8zM)((s-2z)/8zs)^{s-1}$. Theorem 2 gives for $y \le z$

$$\left|\frac{p^{ns}(y)}{\pi(y)} - 1\right| \le R_M^0(1 - a_z b_z)^n + 2\frac{1-b_z}{b_z}.$$

Hence this error may be made arbitrarily small. In this example we do not get convergence in $L_{\pi,\infty}$ norm since $r^s(y|x) = 0$ when $|y - x| > s$. We do, however, get convergence in L_q, L_∞ and total variation norm from Corollary 2.

Example 3. The Metropolis–Hastings algorithm is as follows.

Metropolis–Hastings. To generate a sample from the probability density $\pi(x)$:

1. Generate an initial state $x^0 \in \Omega$ from the density $p^0(x)$.
2. For $i = 1, \ldots, n$:

(a) Generate an alternative state y from the density $q(y|x^i)$.
(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 the Doeblin property (3) with $s = 1$ if

$$q(y|x) \geq a\pi(y) \qquad \text{for all } x, y \in \Omega \tag{6}$$

since

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

Hence (6) implies the Doeblin property. 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 $\{D\}_{i=0}^s$ such that $D_0 = \{x\}$, $D_s = \{y\}$ and for any $u \in D_i$ $v \in D_{i+1}$, $q(v|u) \geq a_i\pi(v)$ and $q(u|v) \geq a_i\pi(u)$. This gives

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

which satisfies the Doeblin property for sufficient large values of s. This was discussed in more detail in Holden (1996).

Example 4. In this example we want to show that the bound in the eigenvalue is not necessarily an upper bound. Assume the state space consists of n points with limiting distribution $(\pi_1, \pi_2, ..., \pi_n)$ and with transition matrix

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

This transition matrix satisfies the Doeblin property $Q_{i,j} = r(i|j) \geq a\pi_i$ 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. This example shows that the upper bound given in Corollary 3 is optimal for $n = 2$ but not optimal for $n > 2$.

8. Closing remarks

This paper discussed the convergence of a general Markov chain. It shows that the Doeblin property is critical for convergence. The Doeblin property is satisfied with $a_s > 0$ if it is possible to jump between any two states in the state space in s jumps with a positive density. If Doeblin property is satisfied, we will get geometric convergence in the most generally used norms. The geometric convergence ratio is $(1 - a_s)^{1/s}$. Hence it is critical that a_s is as large as possible and s as small as possible.

If an arbitrarily large number of jumps is necessary in order to reach the tail of the limiting distribution, then the Doeblin property is not satisfied and there is no convergence in the relative supremum norm. Under some additional weak assumptions, the chain converges in most other norms.

The results in the paper also gives a good qualitative understanding of the convergence. In particular, the Proposition shows that the high-frequency error is reduced faster than the low-frequency errors.

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

Diaconis, P. (1996) The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences* **93**, 1659–1664.

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

Geyer, C. J. (1992) Practical Markov chain Monte Carlo. *Statistical Science* **7**, 473–483.

Holden, L. (1996) Geometric convergence of the Metropolis–Hastings simulation algorithm. Preprint, University of Oslo.

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.