ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

E. NUMMELIN
R. L. TWEEDIE

Geometric ergodicity for a class of Markov chains

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 61, série *Mathématiques*, nº 14 (1976), p. 145-154
<http://www.numdam.org/item?id=ASCFM_1976__61_14_145_0>

GEOMETRIC ERGODICITY FOR A CLASS OF MARKOV CHAINS

NUMMELIN, E. AND TWEEDIE, R.L.
UNIVERSITE DE CANBERRA (ENGLAND)

Suppose $\{X_n\}$ is a Markov chain on a countable state space $S = \{0, 1, \ldots\}$, with transition probabilities $P^n(i,j) = Pr(X_n=j \mid X_0 = i)$. If $\{X_n\}$ is irreducible and aperiodic, then it is known (cf. [1]) that for every pair (i,j) the transition probabilities $P^n(i,j)$ have limits $\pi(j)$ independent of i as $n \to \infty$. In [2], Kendall showed that if, for some state i, this convergence is geometrically fast in the sense that, for some $\rho_{ii} < 1$,

$$|P^n(i,i) - \pi(i)| = O(\rho_{ii}^n),$$

then for every pair (k,j), there exists $\rho_{kj} < 1$ such that

$$|P^n(k,j) - \pi(j)| = O(\rho_{kj}^n).$$

Hence this property of "geometric ergodicity" is a solidarity property of the chain, rather than of individual states.

In [11], Vere-Jones significantly improved the result of Kendall by exhibiting the existence of a common value $\rho < 1$ such that $\rho_{kj} \leqq \rho$ for every pair (k,j). In the transient case, where $\pi_j \equiv 0, \rho$ can be chosen as the best-possible rate of convergence for all the pairs (k,j) ; but this is not the case in the positive recurrent case, when $\pi_j > 0$. Vere-Jones [11] gives an example of a 3-state chain where ρ cannot be chosen as a common rate of convergence, and in [8] Teugels gives a chain on $n+1$ states with n different values for the optimal convergence rates ρ_i of the states i.

In this note we consider a Markov chain $\{X_n\}$ on a general state space $(X,\mathcal{F})$, with transition probabilities

$$P^n(x,A) = Pr(X_n \varepsilon A \mid X_0 = x)$$

which are assumed to be measures on $\mathcal{F}$ for each x, and measurable functions on X for each fixed $A \varepsilon \mathcal{F}$. The basic assumption we shall make throughout is that there exists an <u>atom</u> $\alpha \varepsilon X$, which can be reached with positive probability from every point $x \varepsilon X$; that is, $\sum_n P^n(x,\alpha) > 0$ for all $x \varepsilon X$. Such a chain is, in the nomenclature of Revuz ([7], p.71), δ_α -irreducible, where δ_α is the Dirac measure at α ; and so from [10], there is a "maximal" irreducibility measure M_α (which can be taken in this case as $M(.) = \sum_n P^n(\alpha,.)2^{-n}$) such that $M_\alpha(A) > 0$ implies $\sum_n P^n(x,A) > 0$ for every x ; and such that $M_\alpha(A) = 0$ implies $M_\alpha\{x : \sum_n P^n(x, A) > 0\} = 0$.

From [9] we know that $\{X_n\}$ can be classified as either 1-recurrent, if $\sum_n P^n(x, A) \equiv \infty$ for all $x \varepsilon X$ and all $A \varepsilon \mathcal{F}$ with $M_\alpha(A) > 0$; or as 1-transient, when there exists a sequence of sets $A(j) \uparrow X$ and constants $b_j \uparrow \infty$ with $\sum_n P^n(x, A(j)) \leqq b_j$ for all $x \varepsilon X$. In the latter, 1-transient, case an analogue of Vere-Jones result is already known : from Theorem 1 of [9] we have

<u>Theorem 1</u>. If, for some $\rho_\alpha < 1$,

$$P^n(\alpha,\alpha) = O(\rho_\alpha^n),$$

then there is a sequence $R(j) \uparrow X$, and a number $\rho<1$ such that for M_α-almost all $x \varepsilon X$, and each j,

$$P^n(x, R(j)) = O(\rho^n).$$

We now turn to the situation where $\{X_n\}$ is 1-recurrent, and $P^n(\alpha,\alpha)$ tends to a positive limit (clearly geometric ergodicity cannot hold when $\{X_n\}$ is 1-recurrent with $P^n(\alpha,\alpha) \to 0$, i.e. in the null-recurrent case). We then know that there is a unique probability measure π on $\mathcal{F}$, which is equivalent to M_α , satisfies $\pi = \pi P$, and is such that $P^n(\alpha,\alpha) \to \pi(\alpha)$ (cf. [9] , or in the more usual case where $\mathcal{F}$ is countably generated, Section 3.2 of [7]). We now prove :

<u>Theorem 2</u>. Suppose that for some $\rho_\alpha < 1$ and $\pi(\alpha) > 0$,

$$| P^n(\alpha,\alpha) - \pi(\alpha) | = O(\rho_\alpha^n). \tag{1}$$

Then there exists $\rho < 1$ sucht that for π-almost all x,

$$||P^n(x,.) - \pi(.)|| = O(\rho^n), \tag{2}$$

where $||.||$ denotes total variation on $\mathcal{F}$.

<u>Proof</u>. Let $\tau_\alpha = \inf (n \geq 1 : X_n = \alpha)$, and write

$$\Gamma_{x,A}(n) = Pr(X_n \in A,\ \tau_\alpha \geq n | X_0 = x),$$

so that in particular, $\Gamma_{x,\alpha}(n) = Pr(\tau_\alpha = n \mid X_0 = x)$, and

$$\Gamma_{x,X}(n) = Pr(\tau_\alpha \geq n \mid X_0 = x). \tag{3}$$

We first note two facts :

(a) from the original work of Kendall [2], our assumption (1) is equivalent to existence of r_0 with $1 < r_0 < \rho_\alpha^{-1}$ such that

$$\sum_{n=1}^{\infty} \Gamma_{\alpha,\alpha}(n)\ r_0^n < \infty; \tag{4}$$

and

(b) from [9] , Theorem 4, the measure π can be given as

$$\pi(A) = \pi(\alpha) \sum_{n=1}^{\infty} \Gamma_{\alpha,A}(n), \quad A \in \mathcal{F}. \tag{5}$$

Now by considering the times of first and last entrance to α , we have the decomposition

$$P^n(x,A) = \Gamma_{x,A}(n) + \Gamma_{x,\alpha} * u * \Gamma_{\alpha\ ,\ A}(n), \tag{6}$$

where * denotes convolution, and $u(n) = P^n(\alpha,\alpha)$. Hence

$$|P^n(x,A) - \pi(A)| \leqq \Gamma_{x,A}(n) + |\Gamma_{x,\alpha} * u * \Gamma_{\alpha,A}(n) - \pi(\alpha)\Gamma_{\alpha,A} * 1(n)|$$

$$+ |\pi\ (\alpha)\Gamma_{\alpha,A} * 1(n) - \pi(A)|$$

$$\leqq \Gamma_{x,X}(n) + \Gamma_{\alpha,X} * |\Gamma_{x,\alpha} * u(n) - \pi(\alpha)|$$

$$+ \pi(\alpha) \sum_{j=n+1}^{\infty} \Gamma_{\alpha,X}(j), \tag{7}$$

from (5) and the fact that $\Gamma_{x,A}(n) \leqq \Gamma_{x,X}(n)$. For any $r > 1$, then

$$\sum_{n=1}^{\infty} r^n \ ||P^n(x,.) - \pi(.)|| \leqq \sum_{n=1}^{\infty} r^n \ \Gamma_{x,X}(n)$$

$$+ \sum_{m=1}^{\infty} r^m \Gamma_{\alpha,X}(m) \sum_{n=1}^{\infty} r^n \ |P^n(x,\alpha) - \pi(\alpha)| + \pi(\alpha) \sum_{n=1}^{\infty} r^n \sum_{j=n+1}^{\infty} \Gamma_{\alpha,X}(j). \tag{8}$$

Now from (4) and (3), we have that for $r \leqq r_0$,

$$\infty > \sum_{n=1}^{\infty} r^n \sum_{j=n}^{\infty} \Gamma_{\alpha,\alpha}\ (j) = \sum_{n=1}^{\infty} r^n \ \Gamma_{\alpha,X}(n) \ ; \tag{9}$$

and this in turn implies that for $r \leqq r_0$

$$\infty > \sum_{n=1}^{\infty} r^n \sum_{j=n}^{\infty} \Gamma_{\alpha,X}(j). \tag{10}$$

Moreover, we have for any $k < n$

$$\Gamma_{\alpha,X}(n) = \int_{\{\alpha\}^c} \Gamma_{\alpha,dy}(k)\ \Gamma_{y,X}(n-k) \ :$$

hence from (9), we have for each k and $r \leqq r_0$

$$\infty > \int_{\{\alpha\}^c} \Gamma_{\alpha,dy}(k) \sum_{n=1}^{\infty} r^n \Gamma_{y,X}(n),$$

and so using (5), for π-almost all x

$$\infty > \sum_{n=1}^{\infty} r^n \Gamma_{x,X}(n). \tag{11}$$

Next, we have from (6) with $A = \{\alpha\}$ that for π-almost all x,

$$|P^n(x,\alpha) - \pi(\alpha)| \leqq |\Gamma_{x,\alpha} * u(n) - \pi(\alpha)\Gamma_{x,\alpha} * 1(n)| + \pi(\alpha) \sum_{j=n+1}^{\infty} \Gamma_{x,\alpha}(j),$$

since $\sum_{j=1}^{\infty} \Gamma_{x,\alpha}(j)=1$ for π-almost all x in this recurrent case. Hence for π-almost all x and all $r \leqq r_0$,

$$\sum_{n=1}^{\infty} r^n|P^n(x,\alpha) - \pi(\alpha)| \leqq \sum_{m=1}^{\infty} r^m \Gamma_{x,\alpha}(m) \sum_{n=1}^{\infty} r^n | u(n) - \pi(\alpha)|$$

$$+ \pi(\alpha) \sum_{n=1}^{\infty} r^n \sum_{j=n+1}^{\infty} \Gamma_{x,\alpha}(j) < \infty, \tag{12}$$

from (11) and (1). But now (9) - (12) show that all the terms on the right hand side of (8) converge for $r \leqq r_0$ and π-almost all x ; and so we have (2) as required.

Let us now turn to the situation where X_0 has an arbitrary initial distribution λ on $\mathcal{F}$. If (1) holds, then from the proof above, we will have for some $\rho_\lambda < 1$

$$||\lambda P^n(.) - \pi(.)|| = O(\rho_\lambda^n),$$

provided for some $r > 1$

$$\sum_{n=1}^{\infty} r^n \Gamma_{\lambda,\alpha}(n) < \infty , \tag{13}$$

$$\sum_{n=1}^{\infty} \Gamma_{\lambda,\alpha}(n) = 1, \tag{14}$$

where $\Gamma_{\lambda,\alpha}(n) = \Pr(\tau_\alpha = n \mid X_0 \sim \lambda)$.

But (13), (14) and (1) not only imply but are in fact equivalent to (1), together with

$$|\lambda P^n(\alpha) - \pi(\alpha)| = O(\rho_1^n) \tag{15}$$

for some $\rho_1 < 1$. To see this, we note firstly that since $\lambda P^n(\alpha) \to \pi(\alpha) \sum_{n=1}^{\infty} \Gamma_{\lambda,\alpha}(n)$, (15) clearly implies (14) ; and secondly (1) and (15) are equivalent, from [2] , to the existence of a constant $s > 1$ such that the functions

$$(1 - z) \sum_{n=1}^{\infty} z^n u(n) = U(z)$$

$$(1 - z) \sum_{n=1}^{\infty} z^n \lambda P^n(\alpha) = A(z)$$

are analytic and zero-free except at $z = 1$, in $|z| \leqq s$. Now from (6) again, we have that for $r \leqq s$,

$$A(r) = \sum_{n=1}^{\infty} \Gamma_{\lambda,\alpha}(n) r^n U(r),$$

from which (13) follows.

We can thus see :

Theorem 3. Suppose λ is any initial distribution and that for some $\rho_0 < 1$,

$$|P^n(\alpha,\alpha) - \pi(\alpha)| = O(\rho_0^n) \text{ and}$$

$$|\lambda P^n(\alpha) - \pi(\alpha)| = O(\rho_0^n).$$

Then for some $\rho < 1$,

$$||\lambda P^n(.) - \pi(.)|| = O(\rho^n).$$

It is perhaps of interest to remark that the method used to prove Theorem 2 shows that when there is an atom α in the space, and for some $\pi(\alpha) > 0$

$$| P^n(\alpha,\alpha) - \pi(\alpha) | \to 0,$$

then there is a measure π such that for π- almost all x,

$$||P^n(x, .) - \pi(.)|| \to 0. \tag{16}$$

The existence of a single π-null set on which (16) holds can only be shown in general provided $\mathcal{F}$ is countably generated, but the techniques above are all independent of such an assumption.

Finally, we give an application of our results to random walk on a half-line. Let $\{Y_1,Y_2,...\}$ be a sequence of independently and identically distributed random variables on $(-\infty,\infty)$, with $Pr(Y_i < 0) > 0$. If we write

$$X_n = (X_{n-1} + Y_n)^+, \quad n=1,2,...,$$

then clearly 0 is an atom for $\{X_n\}$. In [3], Miller investigates conditions for geometric ergodicity of the state zero, and then applies the results of [11] to deduce geometric ergodicity of the whole chain $\{X_n\}$ when distributions of Y_i, and so defining $E(Y_i)$ in the extended form of Miller ([3] p. 356) we can give :

Theorem 4. (i) A necessary and sufficient condition for the existence of $\rho_0 < 1$ such that $P^n(0,0) = O(\rho_0^n)$ is

$$E(Y_i) > 0, \quad Pr(Y_i \leqq -x) \leqq De^{-\eta|x|} \tag{17}$$

for some $0 < D < \infty$ and $\eta > 0$. If (17)holds, then for some $\rho < 1$ and M_0-almost all x and all relativtly compact K,

$$P^n(x, K) = O(\rho^n). \tag{18}$$

(ii) A necessary and sufficient condition for the existence of $\rho_0 < 1$ and $\pi(0) > 0$ such that $|P^n(0, 0) - \pi(0)| = O(\rho_0^n)$ is

$$E(Y_i) < 0, \quad Pr(Y_i > x) \leqq De^{-\eta x} \tag{19}$$

for some $0 < D < \infty$ and $\eta > 0$. If (19) holds, then there exists $\rho < 1$ and a probability measure π on $[0,\infty)$ such that for π-almost all x,

$$||P^n(x, .) - \pi(.)|| = O(\rho^n). \tag{20}$$

Proof. Miller [3] shows that (17) or (19) are, in their respective cases, conditions such that the state zero has the required geometric ergodicity properties. The rest of the results then come immediately from our Theorems 1 and 2, except for the statement that (18) holds for all relatively compact K. This last comes from Theorem 1 of [6] , as a consequence of which we know that relatively compact sets on $[0,\infty)$ are R-sets (i.e. have the "correct" geometric rate of convergence), provided the transition probability kernel P maps continuous bounded functions to continuous bounded functions. Since this is true for all random walks, the theorem holds.

In conclusion, we should state that the results given here can be extended to the more general analytic context of R-positive Markov chains, described in [9] ; and that by using the splitting technique of [4] , the need for an atom in the state space can be eliminated. This extension of our results is described in [5] .

REFERENCES

[1] Chung. K.L. : Markov Chains with Stationary Transition Probabilities. (2nd Ed.) Springer-Verlag, Berlin, 1967.

[2] Kendall, D.G. : Unitary dilations of Markov transition operators and the corresponding integral representations for transition-probability matrices, pp. 139-161 in U. Grenander (Ed.), Probability and statistics. Stockholm : Almqvist and Wiksell, 1959.

[3] Miller, H.D. : Geometric ergodicity in a class of denumerable Markov chains. Z. Wahrscheinlichkeitstheorie verw. Geb. 4 (1965) , 354-373.

[4] Nummelin, E. : A splitting technique for φ-recurrent Markov chains, (submitted).

[5] Nummelin, E. and Tweedie, R.L. : Geometric ergodicity and R-positivity for general Markov chains. (submitted).

[6] Pollard, D.B. and Tweedie, R.L. : R-theory for Markov chains on a topological state space II. Z. Wahrscheinlichkeitstheorie verw. Geb. 34 (1976), 269-278.

[7] Revuz, D. : Markov Chains. North-Holland, Amsterdam, 1975.

[8] Teugels, J.L. : An example of geometric ergodicity in a finite Markov chain. J. Appl. Prob. 9 (1972) , 466-469.

[9] Tweedie, R.L. : R-theory for Markov chains on a general state space I : solidarity properties and R-recurrent chains. Ann. Probability 2 (1974), 840-864.

[10] Tweedie, R.L. : Criteria for classifying general Markov chains. Adv. Appl. Prob. 8 (1976) (to appear).

[11] Vere-Jones, D. : Geometric ergodicity in denumerable Markov chains. Quart. J. Math. (Oxford 2nd series) 13 (1962), 7-28.