

## FURTHER CRITERIA FOR POSITIVE HARRIS RECURRENCE OF MARKOV CHAINS

ONÉSIMO HERNÁNDEZ-LERMA AND JEAN B. LASSERRE

(Communicated by Claudia Neuhauser)

ABSTRACT. We provide several necessary and sufficient conditions for a Markov chain on a general state space to be positive Harris recurrent. The conditions only concern asymptotic properties of the expected occupation measures.

### 1. INTRODUCTION

This paper is concerned with characterizations of time-homogeneous *positive Harris recurrent* (P.H.R.) Markov chains on a measurable space  $(X, \mathcal{B})$ , with a countably generated, or “separable”,  $\sigma$ -algebra  $\mathcal{B}$ . The P.H.R. chains are, by far, the chains which enjoy the strongest properties. In particular, the so-called Strong Law of Large Numbers (SLLN) for functionals of Markov chains holds (see (1.2) below), a highly desirable property (for instance, for evaluating a functional via simulation). Using standard terminology (see e.g. [4]–[6]), a Markov chain  $\Phi := \{\Phi_t, t = 0, 1, \dots\}$  on a measurable space  $(X, \mathcal{B})$  is said to be P.H.R. if (a) there exists a  $\sigma$ -finite measure  $\psi$  on  $\mathcal{B}$  such that for all initial states  $\Phi_0 = x \in X$ ,

$$P_x(\Phi_t \in B \text{ for some } t < \infty) = 1 \quad \text{whenever } \psi(B) > 0, B \in \mathcal{B},$$

and (b)  $\Phi$  admits a (necessarily unique) invariant probability measure (i.p.m.).

Needless to say, checking the above conditions is not easy, nor are the various sufficient conditions that one may find in the literature.

In this paper, we provide several *necessary and sufficient* conditions that only involve the asymptotic behavior of the sequence of expected occupation measures (or Cesàro averages)

$$S^{(n)}(x, B) := n^{-1} \sum_{t=1}^n P^t(x, B) = n^{-1} E_x \sum_{t=1}^n I_B(\Phi_t)$$

for  $B \in \mathcal{B}$ ,  $x \in X$ , and  $n = 1, 2, \dots$ , where  $I_B$  denotes the indicator function of  $B \in \mathcal{B}$ , and  $P^t$  is the  $t$ -step transition kernel, that is,  $P^t(x, B) = P_x(\Phi_t \in B)$ . Typical results in the literature usually assume *aperiodicity* of the chain under which various sufficient conditions exist for positive Harris recurrence, and properties of

---

Received by the editors March 1, 1999 and, in revised form, August 15, 1999.

1991 *Mathematics Subject Classification*. Primary 60J10, 28A33; Secondary 28C15.

*Key words and phrases*. Probability measures, setwise convergence, Harris (Markov) chains.

This research was partially supported by the CNRS (France)-CONACYT (México) Scientific Cooperation Program, and by the ECOS (France)-ANUIES (Mexico) Educational and Scientific Cooperation Program.

The first author’s research was also supported by CONACYT Grant 3115P-E9608.

P.H.R chains are stated in terms of the convergence of the sequence  $P^n(x, \cdot)$  of  $n$ -step probabilities and not the sequence  $S^{(n)}(x, \cdot)$  (see e.g. Theorem 13.3.3 in Meyn and Tweedie [4] or Corollary 6.7 in Nummelin [5]).

We show that, in fact, the Cesàro sequence  $S^{(n)}(x, \cdot)$  of expected occupation measures is a useful and more appropriate tool for *characterizing* positive Harris recurrence of a Markov chain with no aperiodicity assumption. Namely, we have:

**Theorem 1.1.** *The chain  $\Phi$  is P.H.R. if and only if, for every  $B \in \mathcal{B}$ , there is a nonnegative number  $a_B$  such that*

$$(1.1) \quad S^{(n)}(x, B) \rightarrow a_B \quad \forall x \in X.$$

When  $\Phi$  is known to have a unique i.p.m., one may refine the above result as follows, where  $\|\cdot\|$  denotes the total variation norm.

**Theorem 1.2.** *Suppose that  $\Phi$  admits a unique i.p.m.  $\mu$ . Then the following statements are equivalent:*

- (a)  $\Phi$  is P.H.R.
- (b) For each  $x \in X$ ,  $\|S^{(n)}(x, \cdot) - \mu\| \rightarrow 0$ .
- (c) For each  $x \in X$ ,  $S^{(n)}(x, \cdot) \rightarrow \mu(\cdot)$  setwise.
- (d) For each  $x \in X$ ,  $S^{(n)}(x, B) \rightarrow 0$  for every  $B \in \mathcal{B}$  with  $\mu(B) = 0$ .

The *characterization* (1.1) of P.H.R. chains (also obtained with different arguments in Glynn [1] for continuous-time processes, and for discrete-time processes in his unpublished Ph.D. thesis) is simpler than the SLLN property: *for each  $x \in X$  and  $f \in L_1(\mu) \equiv L_1(X, \mathcal{B}, \mu)$ ,*

$$(1.2) \quad n^{-1} \sum_{t=1}^n f(\Phi_t) \rightarrow \int f d\mu \quad P_x\text{-a.s.},$$

which is another characterization of positive Harris recurrent chains. Indeed, (1.1) is “simpler” than (1.2), because (1.1) is only in *expectation*, not *pathwise*, a significant improvement. In addition, while one needs to check (1.2) for every  $f \in L_1(\mu)$ , only bounded measurable functions are required in (1.1). In addition, one does not need to identify  $a_B$  as  $\mu(B)$ . Finally, observe that Theorem 1.1 does not invoke any irreducibility nor “minorization” hypotheses on the transition kernel, as do most of the results on sufficient conditions for positive Harris recurrence (see [4]–[6]). Condition (1.1) is stated *only* in terms of the chain’s transition kernel  $P(x, B)$ . On the other hand, part (d) in Theorem 1.2 shows that it suffices to have (1.1) for  $\mu$ -null sets if  $\mu$  is the unique i.p.m. In fact, part (d) remains valid if we replace  $\mu$  with any measure  $\nu$  “equivalent” to  $\mu$ . This fact is particularly useful in applications where one easily knows a measure  $\nu$  equivalent to  $\mu$ , but not  $\mu$  itself. This is the case, for instance, for additive-noise systems in  $R^n$

$$(1.3) \quad x_{t+1} = F(x_t) + \xi_t, \quad t = 0, 1, \dots,$$

where  $F : R^n \rightarrow R^n$ ,  $\xi_t$  i.i.d. random vectors, with  $\xi_0$  absolutely continuous with respect to the Lebesgue measure  $\lambda$ , and with a strictly positive density. If the system has an i.p.m.  $\mu$ , then  $\mu$  is equivalent to  $\lambda$ .

The equivalence “(b)  $\Leftrightarrow$  (c)” in Theorem 1.2, that is, the fact that the setwise convergence of  $S^{(n)}(x, \cdot)$  to  $\mu$  implies the much stronger convergence in the total variation norm (i.e., the setwise convergence is *uniform* on the Borel sets), is indeed a very special feature that *characterizes* P.H.R. chains, because of course, this

implication does not hold for arbitrary setwise convergent sequences of probability measures.

From Theorem 1.2 it follows immediately that “noiseless” measure-preserving transformations in  $R^n$  (as in [3], for instance) whose invariant probability has a density w.r.t. the Lebesgue measure  $\lambda$ , cannot be P.H.R. chains. Indeed, the expected occupation measures  $S^{(n)}(x, \cdot)$  are all finitely supported, and, therefore, cannot converge setwise to a p.m. that has a density w.r.t.  $\lambda$ . For the same reasons, random walks on the line with finitely supported noise distribution and invariant measure with a density w.r.t.  $\lambda$  cannot be P.H.R. chains.

Finally, we provide three sufficient conditions to ensure that (d) in Theorem 1.2 holds. Their remarkable feature is that they only involve the one-step transition kernel.

**Theorem 1.3.** *Suppose that  $\Phi$  admits a unique i.p.m.  $\mu$ , and consider the following conditions (i), (ii), (iii):*

- (i) *For every  $x \in X$ ,  $P(x, \cdot) \ll \mu$ .*
- (ii) *For every  $\mu$ -null set  $B \in \mathcal{B}$*

$$\sup_{x \in B} P(x, B) < 1.$$

- (iii) *For every  $\mu$ -null set  $B \in \mathcal{B}$ , there exists a nonnegative measurable function  $f_B : X \rightarrow R$  such that*

$$(1.4) \quad \int P(x, dy) f_B(y) \leq f_B(x) - I_B(x) \quad \forall x \in X.$$

*Then (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii)  $\Rightarrow$  (d) in Theorem 1.2; hence, each of the conditions (i), (ii), (iii) implies that  $\Phi$  is P.H.R.*

The sufficient conditions in Theorem 1.3 are easy to check in some cases (for instance, for an additive-noise system as in (1.3)).

## 2. PROOFS

**2.1. Proof of Theorem 1.1.** The *only if* part is easy: If  $\Phi$  is P.H.R. and  $\mu$  denotes the unique i.p.m., then from the SLLN (1.2) with  $f := I_B$ , and the Bounded Convergence Theorem we obtain

$$(2.1) \quad S^{(n)}(x, B) \rightarrow \mu(B) \quad \forall B \in \mathcal{B}, x \in X.$$

Therefore, (1.1) follows with  $a_B := \mu(B)$ ,  $B \in \mathcal{B}$ .

The *if* part: Suppose that (1.1) holds. Then, by the Vitali-Hahn-Saks Theorem, it follows that (2.1) holds for some p.m.  $\mu$  on  $\mathcal{B}$ . It is straightforward to check that  $\mu$  is the unique i.p.m. Invoking Birkhoff’s Individual Ergodic Theorem [7] and using the uniqueness of  $\mu$ , it follows that for  $f \in L_1(\mu)$  and for  $\mu$ -almost all  $x \in X$

$$\lim_{n \rightarrow \infty} n^{-1} E_x \sum_{t=1}^n f(\Phi_t) = \int f d\mu = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n f(\Phi_t),$$

where the second equality holds  $P_x$ -a.s. Moreover, from the setwise convergence in (2.1), it also follows that the bounded harmonic functions are constant. Then, for an arbitrary  $f \in L_1(\mu)$ , we show that the bounded function  $h(x) := P_x\{n^{-1} \sum_{t=1}^n f(\Phi_t) \rightarrow \int f d\mu\}$  is harmonic. Thus,  $h$  is constant and, since  $h(x) = 1$  for at least one  $x$ , the SLLN (1.2) follows, which in turn implies that  $\Phi$  is P.H.R.

**2.2. Proof of Theorem 1.2.** By Theorem 1.1, (a)  $\Leftrightarrow$  (c), and, on the other hand, it is obvious that (b)  $\Rightarrow$  (c)  $\Rightarrow$  (d). Therefore, to complete the proof it suffices to show that (d)  $\Rightarrow$  (b). To do this, the idea is to use the Lebesgue decomposition of  $P^k$  w.r.t.  $\mu$ , that is,  $P^k(x, \cdot) = Q_k(x, \cdot) + \nu_k(x, \cdot)$  with  $Q_k(x, \cdot) \ll \mu$  and  $\nu_k(x, \cdot) \perp \mu$ . Then, using (d), one may show that  $\liminf_k \nu_k(x, X) = 0$  so that  $\lim_i Q_{k_i}(x, X) = 1$  for some subsequence  $\{k_i\}$ . On the other hand, by the Chacon-Ornstein Theorem we have that  $\|\nu S^{(n)} - \mu\| \rightarrow 0$  for every p.m.  $\nu \ll \mu$  (see e.g. [2]). Therefore, after normalizing  $Q_k$  as  $\hat{Q}_k(x, \cdot) := Q_k(x, \cdot)/Q_k(x, X)$  in the above subsequence, one obtains  $\|\hat{Q}_{k_i}(x, \cdot)S^{(n)} - \mu\| \rightarrow 0$ . Finally, using that

$$P^k S^{(n)}(x, \cdot) = (n+k)S^{(n+k)}/n - kS^{(k)}(x, \cdot)/n,$$

the result follows.

**2.3. Proof of Theorem 1.3.** The implication (i)  $\Rightarrow$  (ii) is trivial. To prove (ii)  $\Rightarrow$  (iii), one first constructs a set  $S \supset B$  such that  $\mu(S) = 0$  and  $P(x, S) = 0$  if  $x \notin S$ . Then, the function  $f_B := \alpha I_S$ , with  $\alpha \geq (1 - \sup_{x \in S} P(x, S))^{-1}$ , satisfies (1.4). Finally, to prove (iii)  $\Rightarrow$  (d) in Theorem 1.2, it suffices to note that iteration of (1.4) yields

$$\sum_{t=1}^n P^t(x, B) \leq f_B(x) - P^{n+1}f_B(x) \leq f_B(x) \quad \forall x \in X,$$

so that  $S^{(n)}(x, B) \rightarrow 0$  for all  $x \in X$ .

#### REFERENCES

1. Glynn P.W. *Some topics in regenerative steady-state simulation*, Acta Appl. Math. 34, pp. 225-236, 1994. MR **94m**:60208
2. Hernández-Lerma O., Lasserre J.B. *Ergodic theorems and ergodic decomposition for Markov chains*, Acta. Appl. Math. 54, pp. 99-119, 1998.
3. Lasota A., Mackey M.C. *Chaos, Fractals and Noise*, 2nd ed., Springer-Verlag, New York, 1994. MR **94j**:58102
4. Meyn S.P., Tweedie R.L. *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993. MR **95j**:60103
5. Nummelin E. *General Irreducible Markov Chains and Non-negative Operators*, Cambridge University Press, Cambridge, 1984. MR **87a**:60074
6. Revuz. D. *Markov Chains*, revised edition, North-Holland, Amsterdam, 1984. MR **86a**:60097
7. Yosida K. *Functional Analysis*, 6th ed., Springer-Verlag, Berlin, 1980. MR **82i**:46002

DEPARTAMENTO DE MATEMÁTICAS, CINVESTAV-IPN, APDO. POSTAL 14-740, MÉXICO D.F. 07000, MEXICO

*E-mail address:* `ohernand@math.cinvestav.mx`

LAAS-CNRS, 7 AVENUE DU COLONEL ROCHE, 31077 TOULOUSE CÉDEX, FRANCE

*E-mail address:* `lasserre@laas.fr`