

## UNIFORM LARGE DEVIATION PROPERTY OF THE EMPIRICAL PROCESS OF A MARKOV CHAIN

BY RICHARD S. ELLIS<sup>1</sup> AND AARON D. WYNER

*University of Massachusetts and AT & T Bell Laboratories*

This note proves a uniform large deviation property for the empirical process of certain Markov chains that take values in a Polish space. The proof is based on recent results for the empirical measure.

Suppose that one has proved a level-3 large deviation theorem for the empirical process of random vectors  $\{X_j, j \geq 0\}$  taking values in a Polish space. Then according to the contraction principle, the level-3 theorem yields a level-2 large deviation theorem for the empirical measure of  $\{X_j, j \geq 0\}$ . This note considers the reverse implication. We show that a uniform level-3 large deviation theorem for the empirical process of certain Markov chains follows automatically from the uniform level-2 large deviation theorem for the empirical measure of the Markov chain derived in the paper by Ellis (1988).

Let  $X_0, X_1, X_2, \dots$  be a Markov chain on a space  $\Omega$  with stationary transition probabilities  $\pi(x, dy)$ . We assume that the Markov chain takes values in a Polish space  $\mathcal{X}$ . For integers  $j \geq 0$  and  $\alpha \geq 2$ , define the random vector

$$Y_{j,\alpha}(\omega) = (X_j(\omega), X_{j+1}(\omega), \dots, X_{j+\alpha-1}(\omega));$$

i.e.,  $Y_{j,\alpha}(\omega)$  is that element of  $\mathcal{X}^\alpha$  with coordinates  $(Y_{j,\alpha}(\omega))_k = X_{j+k-1}(\omega)$  for  $1 \leq k \leq \alpha$ . The process  $\{Y_{j,\alpha}(\omega), j \geq 0\}$  is a Markov chain that takes values in  $\mathcal{X}^\alpha$  and has stationary transition probabilities  $\pi_\alpha(\mathbf{x}, d\mathbf{y})$  given by

$$\pi_\alpha(x_1, \dots, x_\alpha, dy_1 \times \dots \times dy_\alpha) = \delta_{x_2}(dy_1) \delta_{x_3}(dy_2) \dots \delta_{x_\alpha}(dy_{\alpha-1}) \pi(x_\alpha, dy_\alpha).$$

The methods of this paper lead to the following interesting conclusion. Suppose that one has proved a uniform large deviation property for the empirical measure of the Markov chain  $\{X_j, j \geq 0\}$  under some hypothesis  $H$  on  $\pi(x, dy)$ . Suppose further that for each  $\alpha \geq 2$  the transition probability function  $\pi_\alpha(\mathbf{x}, d\mathbf{y})$  of the Markov chain  $\{Y_{j,\alpha}, j \geq 0\}$  satisfies the same hypothesis  $H$ . Then one automatically obtains a uniform large deviation property for the empirical measure of  $\{Y_{j,\alpha}, j \geq 0\}$  and a uniform large deviation property for the empirical process of  $\{X_j, j \geq 0\}$ . Here is an example of such a hypothesis  $H$ .

**HYPOTHESIS 1.1.** For some  $\beta \in \mathbb{N}$ , some  $M \in [1, \infty)$ , all  $x, x' \in \mathcal{X}$  and all Borel sets  $A$  in  $\mathcal{X}$ ,

$$\pi^\beta(x, A) \leq M\pi^\beta(x', A),$$

---

Received January 1988; revised May 1988.

<sup>1</sup>Supported in part by NSF Grant DMS-85-21536.

AMS 1980 subject classification. 60F10.

Key words and phrases. Markov chain, empirical measure, empirical process, uniform large deviation property, rate function.

where  $\pi^\beta(x, dy)$  denotes the  $\beta$ -step transition probability of the Markov chain [ $\pi^\beta(x, A) = \int_{\mathcal{X}} \pi^{\beta-1}(y, A)\pi(x, dy)$ ].

If  $\pi(x, dy)$  satisfies Hypothesis 1.1 on  $\mathcal{X}$  with  $\beta = \beta_0$ , then for each  $\alpha \geq 2$ ,  $\pi_\alpha(\mathbf{x}, dy)$  satisfies Hypothesis 1.1 on  $\mathcal{X}^\alpha$  with  $\beta = \alpha + \beta_0 - 1$ . Indeed

$$\begin{aligned} \pi_{\alpha+\beta_0-1}(x_1, \dots, x_\alpha, dy_1 \times \dots \times dy_\alpha) &= \pi^{\beta_0}(x_\alpha, dy_1)\pi(y_1, dy_2) \cdots \pi(y_{\alpha-1}, dy_\alpha) \\ &\leq M\pi^{\beta_0}(x'_\alpha, dy_1)\pi(y_1, dy_2) \cdots \pi(y_{\alpha-1}, dy_\alpha) \\ &= M\pi_{\alpha+\beta_0-1}(x'_1, \dots, x'_\alpha, dy \times \dots \times dy_\alpha). \end{aligned}$$

Another example of such a hypothesis  $H$  is given in Hypothesis 1.1(b') in Ellis (1988).

We next define the notion of uniform large deviation property. Theorem 1.3 states the uniform large deviation property for the empirical process under Hypothesis 1.1. The theorem will be proved later.

**DEFINITION 1.2.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be Polish spaces,  $\{Q_{n,x}, n \in \mathbb{N}\}$  a sequence of Borel probability measures on  $\mathcal{Y}$  indexed by  $x \in \mathcal{X}$  and  $I$  a function that maps  $X$  into  $[0, \infty]$ .  $\{Q_{n,x}\}$  is said to have a uniform large deviation property with rate function  $I$  if the following hold:

- (a) For any  $L \geq 0$ , the set  $\{y \in \mathcal{Y}: I(y) \leq L\}$  is compact.
- (b) For each closed set  $F$  in  $\mathcal{Y}$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \sup_{x \in \mathcal{X}} Q_{n,x}\{F\} \right) \leq - \inf_{y \in F} I(y).$$

- (c) For each open set  $G$  in  $\mathcal{Y}$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \left( \inf_{x \in \mathcal{X}} Q_{n,x}\{G\} \right) \geq - \inf_{y \in G} I(y).$$

In order to state our main theorem, Theorem 1.3, we need some notation. For each starting point  $X_0 = x$ ,  $\pi(x, dy)$  induces a probability measure  $P_x$  on  $\Omega$ . For each integer  $j$ , let  $(\mathcal{X}_j, \mathcal{B}_j) = (\mathcal{X}, \mathcal{B})$ , where  $\mathcal{B}$  denotes the Borel  $\sigma$ -field of  $\mathcal{X}$ , and define  $(\mathcal{X}^{\mathbb{Z}}, \Sigma)$  to be the product space  $\prod_{j \in \mathbb{Z}} (\mathcal{X}_j, \mathcal{B}_j)$ , endowed with the product topology. We define  $\mathcal{M}(\mathcal{X}^{\mathbb{Z}})$  to be the space of probability measures on  $(\mathcal{X}^{\mathbb{Z}}, \Sigma)$  with the topology of weak convergence and  $\mathcal{M}_T(\mathcal{X}^{\mathbb{Z}})$  to be the subset of  $\mathcal{M}(\mathcal{X}^{\mathbb{Z}})$  consisting of all measures  $P$  that satisfy  $P \circ T^{-1} = P$ .  $T$  denotes the shift operator on  $\mathcal{X}^{\mathbb{Z}}$ .

For each  $n \in \mathbb{N}$  and each point  $\omega \in \Omega$ , repeat the sequence  $(X_0(\omega), X_1(\omega), \dots, X_{n-1}(\omega))$  periodically into a doubly infinite sequence, obtaining a point  $X(n, \omega) \in \mathcal{X}^{\mathbb{Z}}$ . For each  $n \in \mathbb{N}$ , point  $\omega \in \Omega$  and set  $A \in \Sigma$ , we define the empirical process

$$R_n(\omega, A) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j X(n, \omega)}\{A\}.$$

For each  $\omega \in \Omega$ ,  $R_n(\omega, \cdot)$  is an element of  $\mathcal{M}_T(\mathcal{X}^{\mathbb{Z}})$ . A similar result as in Theorem 1.3 holds for a nonperiodized empirical process.

For  $x = (\dots, x_{-2}, x_{-1}, x_0, x_1, \dots)$  a point in  $\mathcal{X}^{\mathbb{Z}}$ , let  $x_{-1}^{-\infty} = (\dots, x_{-2}, x_{-1})$ . We denote by  $\hat{X}_j$ ,  $j \in \mathbb{Z}$ , the mapping taking  $x$  to  $x_j$  and by  $\hat{X}_{-1}^{-\infty}$  the mapping taking  $x$  to  $x_{-1}^{-\infty}$  [ $\hat{X}_{-1}^{-\infty} = (\dots, \hat{X}_{-1}, \hat{X}_0)$ ]. For  $P$  a measure in  $\mathcal{M}_T(\mathcal{X}^{\mathbb{Z}})$ , we define  $P^*(x_{-1}^{-\infty}, dx_0)$  to be a regular conditional distribution of  $\hat{X}_0$  given  $\hat{X}_{-1}^{-\infty} = x_{-1}^{-\infty}$  and we write  $P(dx_{-1}^{-\infty})$  for the  $P$ -distribution of  $\hat{X}_{-1}^{-\infty}$ . In the next theorem,  $I(\cdot, \cdot)$  denotes relative entropy.

**THEOREM 1.3.** *We assume that  $\pi(x, dy)$  satisfies Hypothesis 1.1. For  $n \in \mathbb{N}$ ,  $x \in \mathcal{X}$  and  $A$  a Borel set in  $\mathcal{M}_T(\mathcal{X}^{\mathbb{Z}})$ , we define*

$$Q_{n,x}\{A\} = P_x\{\omega: R_n(\omega, \cdot) \in A\}.$$

*Then the sequence  $\{Q_{n,x}, n \in \mathbb{N}\}$  has a uniform large deviation property with rate function*

$$I_{\pi, \infty}(P) = \int_{x_{-1}^{-\infty}} I(P^*(x_{-1}^{-\infty}, dx_0), \pi(x_{-1}, dx_0))P(dx_{-1}^{-\infty}), \quad P \in \mathcal{M}_T(\mathcal{X}^{\mathbb{Z}}),$$

where  $\mathcal{X}_{-1}^{-\infty} = \prod_{j=-1}^{-\infty} \mathcal{X}_j$ .

The basic idea of this paper may also be applied in a related situation. For any Borel probability measure  $\mu$  on  $\mathcal{X}$ , we consider the Markov family on  $\Omega$ ,

$$P_\mu\{\cdot\} = \int_{\mathcal{X}} P_x\{\cdot\} \mu(dx),$$

corresponding to the transition probability function  $\pi(x, dy)$  and the initial distribution  $\mu$ . Suppose that one has proved a large deviation property for the  $P_\mu$ -distributions of the empirical measure of the Markov chain  $\{X_j, j \geq 0\}$ , valid for all initial distributions  $\mu$ , under some hypothesis  $H_0$  on  $\pi(x, dy)$ . Suppose further that for each  $\alpha \geq 2$  the transition probability function  $\pi_\alpha(x, dy)$  of the Markov chain  $\{Y_{j,\alpha}, j \geq 0\}$  satisfies the same hypothesis  $H_0$ . Then one automatically obtains a large deviation property for the  $P_\mu$ -distributions of the empirical measure of  $\{Y_{j,\alpha}, j \geq 0\}$  and a large deviation property for the  $P_\mu$ -distributions of the empirical process of  $\{X_j, j \geq 0\}$ . The paper by de Acosta (1988) considers examples of such hypotheses  $H_0$ .

The pioneering work on large deviations for the empirical process is the paper by Donsker and Varadhan (1983). Other large deviations results for the empirical process have been obtained by Donsker and Varadhan (1985), Ellis (1985), Orey (1985), de Acosta (1988), Orey and Pelikan (1988) and Stroock and Deuschel (1989).

**PROOF OF THEOREM 1.3.** We define  $(\mathcal{X}^\alpha, \Sigma^\alpha)$  to be the product space  $\prod_{j \neq -\alpha+1}^0 (\mathcal{X}_j, \mathcal{B}_j)$ , endowed with the product topology and  $\mathcal{M}(\mathcal{X}^\alpha)$  to be the space of probability measures on  $(\mathcal{X}^\alpha, \Sigma^\alpha)$  with the topology of weak convergence. For  $P$  a measure in  $\mathcal{M}_T(\mathcal{X}^{\mathbb{Z}})$ ,  $m_\alpha P$  denotes the measure in  $\mathcal{M}(\mathcal{X}^\alpha)$  which is the restriction of  $P$  to the  $\sigma$ -field  $\Sigma^\alpha$ . For each  $\omega \in \Omega$ , positive integers  $n$  and

$\alpha$  and set  $A \in \Sigma^\alpha$ , we define the process

$$M_{n,\alpha}(\omega, A) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{Y_{j,\alpha}(\omega)}\{A\},$$

which is the empirical measure of  $\{Y_{j,\alpha}, j \geq 0\}$ .  $M_{n,\alpha}(\omega, \cdot)$  is an element of  $\mathcal{M}(\mathcal{X}^\alpha)$ .  $M_{n,1}(\omega, \cdot)$  equals the empirical measure  $n^{-1} \sum_{j=0}^{n-1} \delta_{X_j(\omega)}\{\cdot\}$ .

The paper by Ellis (1988) proves that under Hypothesis 1.1 the  $P_x$ -distributions of  $\{M_{n,1}(\omega, \cdot)\}$  on  $\mathcal{M}(\mathcal{X})$  have a uniform large deviation property with an explicitly given rate function  $I_{\pi,1}$  (see Theorem 1.2 and Remarks 1.3 and 2.1 in that paper). We have seen that for each  $\alpha \geq 2$  the transition probability function  $\pi_\alpha(\mathbf{x}, d\mathbf{y})$  of the Markov chain  $\{Y_{j,\alpha}, j \geq 0\}$  on  $\mathcal{X}^\alpha$  also satisfies Hypothesis 1.1. It follows that the  $P_x$ -distributions of  $\{M_{n,\alpha}(\omega, \cdot)\}$  on  $\mathcal{M}(\mathcal{X}^\alpha)$  have a uniform large deviation property with an explicitly given rate function  $I_{\pi,\alpha}$  [see Theorem 1.4 and Remarks 1.5 and 3.5 in Ellis (1988)].

We now show that the  $P_x$ -distributions of  $\{R_n(\omega, \cdot)\}$  on  $\mathcal{M}_T(\mathcal{X}^\mathbb{Z})$  have a uniform large deviation property with some rate function  $J_{\pi,\infty}$  defined in terms of the rate functions  $\{I_{\pi,\alpha}, \alpha \in \mathbb{N}\}$ . We compare the process  $M_{n,\alpha}(\omega, \cdot)$  with the process  $m_\alpha R_n(\omega, \cdot)$ . For each  $\omega \in \Omega$ ,  $m_\alpha R_n(\omega, \cdot)$  is an element of the space

$$\mathcal{M}_T(\mathcal{X}^\alpha) = m_\alpha \mathcal{M}_T(\mathcal{X}^\mathbb{Z}) = \{\mu \in \mathcal{M}(\mathcal{X}^\alpha) : \mu = m_\alpha P \text{ for some } P \in \mathcal{M}_T(\mathcal{X}^\mathbb{Z})\}.$$

Let  $\Gamma_\alpha$  denote the Lévy-Prohorov metric on  $\mathcal{M}(\mathcal{X}^\alpha)$ . Since

$$\sup_{\omega \in \Omega} \Gamma_\alpha(m_\alpha R_n(\omega, \cdot), M_{n,\alpha}(\omega, \cdot)) \leq 2(\alpha - 1)/n,$$

it follows that for each  $\alpha \in \mathbb{N}$  the  $P_x$ -distributions of  $\{m_\alpha R_n(\omega, \cdot)\}$  on  $\mathcal{M}_T(\mathcal{X}^\alpha)$  have a uniform large deviation property with the same rate function  $I_{\pi,\alpha}$  as  $\{M_{n,\alpha}(\omega, \cdot)\}$ . Using the Kolmogorov extension theorem and the Tychonoff product theorem, one easily shows that the projective limit of the spaces  $\{\mathcal{M}_T(\mathcal{X}^\alpha), \alpha \in \mathbb{N}\}$  with the topology of weak convergence equals the space  $\mathcal{M}_T(\mathcal{X}^\mathbb{Z})$  with the topology of weak convergence. An obvious uniformization of Theorem 3.3 in the paper by Dawson and Gärtner (1987) implies that the  $P_x$ -distributions of  $\{R_n(\omega, \cdot)\}$  on  $\mathcal{M}_T(\mathcal{X}^\mathbb{Z})$  have a uniform large deviation property with the rate function

$$J_{\pi,\infty}(P) = \sup_{\alpha \in \mathbb{N}} I_{\pi,\alpha}(m_\alpha P).$$

We now identify  $J_{\pi,\infty}(P)$  with the function  $I_{\pi,\infty}(P)$  defined in Theorem 1.3. For  $\mathbf{x} = (\dots, x_{-2}, x_{-1}, x_0, x_1, \dots)$  a point in  $\mathcal{X}^\mathbb{Z}$  and  $\alpha \in \mathbb{N}$ , let  $x_{-1}^{-\alpha} = (x_{-\alpha}, \dots, x_{-2}, x_{-1})$ . We denote by  $\hat{X}_{-1}^{-\alpha}$  the mapping taking  $x$  to  $x_{-1}^{-\alpha}$  [ $\hat{X}_{-1}^{-\alpha} = (\hat{X}_{-\alpha}^-, \dots, \hat{X}_{-1}^-)$ ]. For  $\alpha \in \mathbb{N}$  and  $P \in \mathcal{M}_T(\mathcal{X}^\mathbb{Z})$ , we define  $P^*(x_{-1}^{-\alpha}, dx_0)$  to be a regular conditional distribution of  $\hat{X}_0^-$  given  $\hat{X}_{-1}^{-\alpha} = x_{-1}^{-\alpha}$ . We write  $P(dx_{-1}^{-\alpha})$ ,  $P(dx_0^-)$ ,  $P(dx_{-1}^{-\infty})$  and  $P(dx_0^{-\infty})$  for the  $P$ -distributions of  $\hat{X}_{-1}^{-\alpha}$ ,  $\hat{X}_0^- = (\hat{X}_{-\alpha}^-, \dots, \hat{X}_{-1}^-, \hat{X}_0^-)$ ,  $\hat{X}_{-1}^{-\infty} = (\dots, \hat{X}_{-2}^-, \hat{X}_{-1}^-)$  and  $\hat{X}_0^{-\infty} = (\dots, \hat{X}_{-2}^-, \hat{X}_{-1}^-, \hat{X}_0^-)$ , respectively. According to Theorem 1.4 in Ellis (1988),

$$\begin{aligned} I_{\pi,\alpha+1}(m_{\alpha+1}P) &= \int_{x_{-1}^{-\alpha}} I(P^*(x_{-1}^{-\alpha}, dx_0), \pi(x_{-1}, dx_0)) P(dx_{-1}^{-\alpha}) \\ &= I(P(dx_0^-), P(dx_{-1}^{-\alpha}) \otimes \pi(x_{-1}, dx_0)), \end{aligned}$$

where  $\mathcal{X}_{-1}^{-\alpha} = \prod_{j=-1}^{-\alpha} \mathcal{X}_j$ . The contraction principle implies that the sequence  $\{I_{\pi, \alpha}(m_\alpha P), \alpha \in \mathbb{N}\}$  is nondecreasing. Hence in order to identify  $J_{\pi, \infty}(P) = \sup_{\alpha \in \mathbb{N}} I_{\pi, \alpha}(m_\alpha P)$  with  $I_{\pi, \infty}(P)$ , it suffices to prove that

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} I_{\pi, \alpha+1}(m_{\alpha+1} P) &= I(P(dx_0^{-\infty}), P(dx_{-1}^{-\infty}) \otimes \pi(x_{-1}, dx_0)) \\ &= I_{\pi, \infty}(P). \end{aligned}$$

This limit, which is well known, follows straightforwardly from any of the following: Lemma 2.1 in Donsker and Varadhan (1975), Exercise IV-5-4 in Neveu (1964), Barron (1985) or Pinsker (1964). The proof of Theorem 1.3 is complete.  $\square$

**Acknowledgments.** One of us (R. S. E.) would like to thank Francis Comets, Hans Föllmer, Eddy Mayer-Wolf, Steven Orey and Daniel Stroock for useful conversations.

## REFERENCES

- BARRON, A. R. (1985). The strong ergodic theorem for densities: Generalized Shannon–McMillan–Breiman theorem. *Ann. Probab.* **13** 1292–1303.
- DAWSON, D. A. and GÄRTNER, J. (1987). Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. *Stochastics* **20** 247–308.
- DE ACOSTA, A. (1988) Large deviations for empirical measures of Markov chains. Technical Report, Case Western Reserve Univ. To appear.
- DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. *Comm. Pure Appl. Math.* **28** 1–47.
- DONSKER, M. D. and VARADHAN, S. R. S. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** 183–212.
- DONSKER, M. D. and VARADHAN, S. R. S. (1985). Large deviations for stationary Gaussian processes. *Comm. Math. Phys.* **97** 187–210.
- ELLIS, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York.
- ELLIS, R. S. (1988). Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Ann. Probab.* **16** 1496–1508.
- NEVEU, J. (1964). *Calcul des Probabilités*. Masson, Paris.
- OREY, S. (1985). Large deviations in ergodic theory. In *Seminar on Stochastic Processes, 1984* (E. Çinlar, K. L. Chung and R. K. Gettoor, eds.) 195–249. Birkhäuser, Boston.
- OREY, S. and PELIKAN, S. (1988). Large deviation principles for stationary processes. *Ann. Probab.* **16** 1481–1495.
- PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco. (A. Feinstein, transl. and ed.)
- STROOCK, D. W. and DEUSCHEL, J. -D. (1989). *An Introduction to the Theory of Large Deviations*. Academic, New York.

DEPARTMENT OF MATHEMATICS AND STATISTICS  
LEDERLE GRADUATE RESEARCH CENTER TOWERS  
UNIVERSITY OF MASSACHUSETTS  
AMHERST, MASSACHUSETTS 01003

AT & T BELL LABORATORIES  
MURRAY HILL, NEW JERSEY 07974