Bernoulli **12**(3), 2006, 515–533

Entropy for semi-Markov processes with Borel state spaces: asymptotic equirepartition properties and invariance principles

VALERIE GIRARDIN[1] and NIKOLAOS LIMNIOS[2]

[1]*LMNO, UMR6139, Campus II, Université de Caen, BP 5186, 14032 Caen, France.*
E-mail: girardin@math.unicaen.fr
[2]*LMAC, Université de Technologie de Compiègne, BP 20529, 60205 Compiègne Cedex, France.*
E-mail: nikolaos.limnios@utc.fr

The aim of this paper is to define the entropy rate of a semi-Markov process with a Borel state space by extending the strong asymptotic equirepartition property (also called the ergodic theorem of information theory or Shannon–McMillan–Breiman theorem) to this class of non-stationary processes. The mean asymptotic equirepartition property (also called the Shannon–McMillan theorem) is also proven to hold. The relative entropy rate between two semi-Markov processes is defined. All earlier results concerning entropy for semi-Markov processes, jump Markov processes and Markov chains thus appear as special cases. Two invariance principles are established for entropy, one for the central limit theorem and the other for the law of the iterated logarithm.

Keywords: asymptotic equirepartition property; entropy rate; functional central limit theorem; functional law of the iterated logarithm; invariance principle; relative entropy; semi-Markov processes; Shannon–McMillan theorem; Shannon–McMillan–Breiman theorem

1. Introduction

Semi-Markov processes constitute an extension of jump Markov processes and renewal processes. They allow the use of any distributions for the sojourn times instead of the exponential (geometric) distributions of Markov processes (chains); therefore, numerous real phenomena can be modelled by semi-Markov processes (see, for example, Limnios and Oprişan 2001a).

The concept of entropy was introduced in statistical mechanics by Boltzmann in the nineteenth century and subsequently in probability by Shannon (1948) in order to study communication systems. Its different forms (entropy, relative entropy, cross-entropy, mutual entropy, conditional entropy, entropy rate, etc.) and different names (not only entropy but also information, divergence, distance, distortion measure, etc.) apply to discrete or absolutely continuous probability distributions, to random variables or vectors, to discrete or continuous time stochastic processes of any dimension, defined on any probability space.

1350–7265

Entropy is used in communication theory, statistical mechanics, finance, signal analysis, image reconstruction, and also in psychology, linguistics, etc.

The Shannon entropy of a discrete probability P with countable support E was originally defined by Shannon (1948) as $\mathbb{S}(P) = -\sum_{x\in E} P(x)\log P(x)$. Kullback and Leibler (1951) defined the relative entropy of a probability with respect to another, also called the Kullback–Leibler information or cross-entropy. They apply to random variables and processes as follows.

Definition 1. *Let P be a probability absolutely continuous with respect to a σ-finite reference measure μ ($P \ll \mu$). The entropy of P with respect to μ is defined by*

$$\mathbb{S}(P|\mu) = \mathbb{E}_P\left(-\log \frac{\mathrm{d}P}{\mathrm{d}\mu}\right) = -\int \log \frac{\mathrm{d}P(x)}{\mathrm{d}\mu(x)}\,\mathrm{d}P(x).$$

If X is a random variable with distribution $P_X \ll \mu$, then

$$\mathbb{S}(P_X|\mu) = \mathbb{E}\left[-\log \frac{\mathrm{d}P_X}{\mathrm{d}\mu}(X)\right].$$

If Y is another random variable with distribution P_Y, then the relative entropy of X with respect to Y is $\mathbb{S}(X|Y) = \mathbb{S}(P_X|P_Y)$.

Let $\mathbf{Z} = (Z_t;\ t \in \mathbb{T})$, where $\mathbb{T} = \mathbb{N}$ or $\mathbb{R}_+$, be a real stochastic process. Let f_T denote the likelihood function of $Z_{[0,T]} = (Z_t)_{0\leqslant t\leqslant T}$ with respect to a σ-finite reference measure μ_T. The entropy of $\mathbf{Z}$ up to time T is $\mathbb{H}_T(\mathbf{Z}) = -\mathbb{E}\,[\log f_T(Z_{[0,T]})]$. If $\mathbb{H}_T(\mathbf{Z})/T$ converges to a finite limit when T tends to infinity, the limit is called the entropy rate $\mathbb{H}(\mathbf{Z})$ of the process.

The relative entropy rate of a process with respect to another is defined similarly. Note that if $\mathbf{Z}$ is defined on its canonical space $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}), \mathbb{P})$, with natural filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$, then μ_T appears as the restriction on $\mathcal{F}_T$ of a reference measure μ on $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}))$.

Theorem A (Asymptotic equirepartition property). *Under suitable conditions on the process $\mathbf{Z}$, its index space $\mathbb{T}$, its state space E and the reference measure μ, the sequence $(-\log f_T(Z_{[0,T]})/T)$ converges in mean or almost surely to the entropy rate of the process.*

The mean convergence result is known in the literature as the Shannon–McMillan theorem and the almost sure convergence as the Shannon–McMillan–Breiman theorem or ergodic theorem of information theory. Both are also called mean and strong asymptotic equirepartition properties, due to their use for characterizing the asymptotic behaviour of the marginal vectors of the process; see Barron (1985) for details and references. In the following, for the sake of brevity, we will call them the mean and strong AEP.

Shannon (1948) stated and McMillan (1953) proved the mean AEP and Breiman (1958) proved the strong AEP, for an ergodic stationary process with finite state space. Perez (1964) proved the mean AEP for an ergodic stationary discrete or continuous time process with a measurable state space. The strong AEP for an ergodic stationary discrete time process was extended to any Borel state space independently by Barron (1985) and Orey (1985). For the strong AEP to hold for an ergodic stationary process, the weakest possible assumption for the reference measure seems to be homogeneous 'nearly' Markovian (see Orey 1985). The strong AEP was proven to hold for a null-recurrent stationary homogeneous Markov chain by Krengel (1967) using ergodic theory arguments, and for a non-homogeneous Markov chain by Wen and Weiguo (1996) using specific arguments. It can be proven to hold for other kinds of discrete or continuous time ergodic stationary processes by using ergodic theory arguments (see Pinsker 1964; Kifer 1986). Bad Dumitrescu (1988) proved the mean AEP for an ergodic stationary pure jump Markov process with a finite state space, by using Perez (1964). See Girardin (2005) for details on extensions to other stationary processes.

Semi-Markov processes are not stationary, and hence none of the above AEPs apply to their case. The strong and mean AEPs for a countable semi-Markov process were obtained in Girardin and Limnios (2003) for the continuous time case and in Girardin and Limnios (2004) for the discrete time case (by proving convergence results on the number of transitions from one state to another), with an explicit analytic expression for the entropy rate. To our knowledge, no other strong AEPs exist for non-stationary processes. The main aim of this paper is to prove the strong and mean AEPs for any semi-Markov process with a Borel state space, giving an explicit analytic expression for the entropy rate.

Considering entropy with respect to a given reference probability measure amounts to considering relative entropy with respect to a given distribution. Different AEPs have been proven in this context. Bad Dumitrescu (1988) computed the relative entropy rate of a pure jump Markov process with respect to another. The AEP for the relative entropy between two finite semi-Markov processes proven in Girardin and Limnios (2003) is here generalized to Borel state space processes.

Maximization of entropy and minimization of relative entropy are widely used to study distributions or random variables. The entropy methods are justified by probabilistic or statistical arguments: see, for example, Garret (2001), Csizár (1996) for Bayesian statistics, Grendar and Grendar (2001) for links to maximum likelihood, and Johnson (2004) for limit theorems. An explicit analytic expression of the entropy rate, as computed here, is necessary for extending these methods to random processes. See Girardin (2004) and Girardin and Limnios (2004) for application to discrete or continuous time finite ergodic homogeneous Markov or semi-Markov processes. See also Shannon (1948) for application to information theory via the well-known first Shannon coding theorem.

The entropy rate was first defined by Shannon (1948) for an ergodic Markov chain $\mathbf{Z}$ with finite state space E as the sum of the entropies of the transition probabilities $(P(x, y))_{y\in E}$ weighted by the probability of occurrence of each state according to the stationary distribution ν of the chain, namely

$$\mathbb{H}(\mathbf{Z}) = -\sum_{x\in E} \nu(x) \sum_{y\in E} P(x, y) \log P(x, y). \tag{1}$$

Shannon (1948) proved the convergence in probability of the sequence $(-\log \mathbb{P}(Z_1 = x_1, \ldots, Z_n = x_n)/n)$ to this quantity. Bad Dumitrescu (1988) obtained the entropy rate of an ergodic stationary pure jump Markov process with finite state space E in the form

$$\mathbb{H}(\mathbf{Z}) = -\sum_{x\in E} \pi(x) \sum_{y\neq x} a_{xy}(\log a_{xy} - 1), \tag{2}$$

where (a_{xy}) is the infinitesimal generator and $(\pi(x))$ is the stationary distribution of the process. Both mean and strong AEPs have been extended in Girardin and Limnios (2003, 2004) to an irreducible positive recurrent semi-Markov process with countable state space E. The entropy rate has an explicit expression as the sum of the entropies of the derivatives $(q_{xy})_{y\in E}$ of the semi-Markov kernel weighted by the probability of occurrence of each state according to the stationary distribution ν of the embedded Markov chain, balanced by the mean $\hat{m}$ of the mean sojourn times with respect to ν:

$$\mathbb{H}(\mathbf{Z}) = -\frac{1}{\hat{m}} \sum_{x\in E} \nu(x) \sum_{y\in E} \int_{\mathbb{R}_+} q_{xy}(s) \log q_{xy}(s) \mathrm{d}s \tag{3}$$

for a continuous time semi-Markov process, and

$$\mathbb{H}(\mathbf{Z}) = -\frac{1}{\hat{m}} \sum_{x\in E} \nu(x) \sum_{y\in E} \sum_{k\in\mathbb{N}} q_{xy}(k) \log q_{xy}(k) \tag{4}$$

for a semi-Markov chain. We generalize these results to a Borel state space, obtaining

$$\mathbb{H}(\mathbf{Z}) = -\frac{1}{\hat{m}} \iiint_{E\times E\times\mathbb{R}_+} \nu(\mathrm{d}x) q(x, y, s) \log q(x, y, s) \mathrm{d}y\, \mathrm{d}s, \tag{5}$$

where $q(x, y, s)$ denotes the derivative of the semi-Markov kernel.

Both AEPs also have many applications linked to central limit theorems or large-deviations results. We prove here (see Theorems 4 and 5) a weak invariance principle for entropy, also called the functional central limit theorem, and a strong invariance principle, also called the functional law of the iterated logarithm, two tools necessary for statistical purposes; see Billingsley (1968) or Gut (1988) for details. These also induce such results as the extension to semi-Markov processes of the central limit theorem, proven in O'Neil (1990), or of the law of the iterated logarithm, proven in Dym (1966), for the entropy of Markov chains.

This paper is organized as follows. General notation, definitions of semi-Markov processes and necessary assumptions are given in Section 2. The strong and mean AEPs are extended to semi-Markov processes with Borel state spaces in Section 3, with an explicit expression for the entropy rate. The particular cases of countable semi-Markov processes and chains are derived in Section 4. The important special case of pure jump Markov processes is studied in Section 5. The AEP for the relative entropy between two semi-Markov processes is obtained in Section 6, with an explicit relative entropy rate. The weak and strong invariance principles are given in Sections 7 and 8.

2. Semi-Markov setting

To define semi-Markov processes, it is natural first to define semi-Markov kernels and Markov renewal processes (see, for example, Limnios and Oprişan 2001a). All the following random processes will be supposed to be defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2. *Let $(E, \mathcal{E})$ be a Borel measurable space. Let $\mathcal{B}_+$ denote the Borel σ-algebra of $\mathbb{R}_+$. Let $(P((x, s); A \times \Gamma), (x, s) \in E \times \mathbb{R}_+, A \in \mathcal{E}, \Gamma \in \mathcal{B}_+)$ be a Markov transition function on $(E \times \mathbb{R}_+, \mathcal{E} \times \mathcal{B}_+)$. The function Q defined by*

$$Q(x, A \times (\Gamma - s)) = P((x, s); A \times \Gamma), \qquad (x, s) \in E \times \mathbb{R}_+, A \in \mathcal{E}, \Gamma \in \mathcal{B}_+,$$

where $\Gamma - s = \{t \in \mathbb{R}_+ : t + s \in \Gamma\}$, is a semi-Markov kernel.

For any $(x, s) \in E \times \mathbb{R}_+$, there exist a probability measure $\mathbb{P}_{(x,s)}$ on $(\Omega, \mathcal{F})$ and a sequence of random variables $(J_n, S_n;\ n \in \mathbb{N})$, such that $\mathbb{P}_{(x,s)}(J_0 = x, S_0 = s) = 1$ and, for $n \geqslant 0$, $A \in \mathcal{E}$, and $\Gamma \in \mathcal{B}$,

$$\begin{aligned}\mathbb{P}_{(x,s)}((J_{n+1}, S_{n+1}) \in A \times \Gamma \sigma(J_k, S_k;\ k \leqslant n)) &= \mathbb{P}_{(x,s)}((J_{n+1}, S_{n+1}) \in A \times \Gamma J_n, S_n) \\ &= Q(J_n, A \times (\Gamma - S_n)).\end{aligned}$$

Let $(N(t);\ t \in \mathbb{R}_+)$ be defined by

$$N(t) = \sup\{n \geqslant 0 : S_n \leqslant t\}. \tag{6}$$

Definition 3. *The stochastic process $(J_n, S_n;\ n \in \mathbb{N})$ is called a Markov renewal process. The stochastic process $\mathbf{Z} = (Z(t);\ t \in \mathbb{R}_+)$, defined by $Z(t) = J_{N(t)}$ for $t \geqslant 0$ (or $J_n = Z(S_n)$ for $n \geqslant 0$), is the semi-Markov process associated with (J_n, S_n).*

To be exact, $\mathbf{Z}$ is an $(E, \mathcal{E})$-valued cadlag homogeneous semi-Markov process. The process $\mathbf{J} = (J_n;\ n \in \mathbb{N})$ (called the embedded Markov chain of $\mathbf{Z}$) is a Markov chain with state space $(E, \mathcal{E})$ and transition probability kernel $P(x, \mathrm{d}y) = Q(x, \mathrm{d}y \times \mathbb{R}_+)$. The process (S_n) is the sequence of jump times of $\mathbf{Z}$, with $S_0 \leqslant S_1 \leqslant \ldots \leqslant S_n \leqslant S_{n+1} \leqslant \ldots$, and inter-jump times $X_n = S_n - S_{n-1}$, for $n \geqslant 1$. The process $(J_n, X_n;\ n \in \mathbb{N}^*)$ is a Markov chain with state space $(E \times \mathbb{R}_+, \mathcal{E} \times \mathcal{B}_+)$ and transition probability kernel $Q(x, \mathrm{d}y \times \mathrm{d}t)$. The point process $(N(t);\ t \in \mathbb{R}_+)$ counts the jumps in the time interval $(0, t]$.

Let H denote the distribution function of the sojourn times, that is, $H(x, t) = Q(x, E \times [0, t])$ for $(x, t) \in E \times \mathbb{R}_+$, and set $\overline{H} = 1 - H$. Let r be the hazard function of H, defined by

$$r(x, t) = \frac{-\dfrac{\partial H}{\partial t}(x, t)}{\overline{H}(x, t)},$$

so that

$$\log[\overline{H}(x,\ t)] = -\int_0^t r(x,\ u)\mathrm{d}u, \qquad x \in E,\ t \in \mathbb{R}_+. \tag{7}$$

Also let $m_k(x)$ denote the kth moment of the sojourn time in state $x \in E$, that is,

$$m_k(x) = \int_{\mathbb{R}_+} Q(x,\ E \times \mathrm{d}s)s^k = \int_{\mathbb{R}_+} H(x,\ \mathrm{d}s)s^k, \qquad k \in \mathbb{N}^*,\ x \in E; \tag{8}$$

we set $m(x) = m_1(x)$ for the mean sojourn time in state $x \in E$.

We will also need the following definitions. Let $\mathcal{P}(E)$ be the set of all probability distributions α on $(E,\ \mathcal{E})$; note that α will denote both the distribution and its density when it exists.

Definition 4. *Let α be a probability measure in $\mathcal{P}(E)$. The probability measure $\mathbb{P}_\alpha$ is defined on $(\Omega,\ \mathcal{F})$ by*

$$\mathbb{P}_\alpha(C) = \int_E \alpha(\mathrm{d}x)\mathbb{P}_x(C), \qquad C \in \mathcal{F},$$

where $\mathbb{P}_x(C) = \mathbb{P}(C \mid J_0 = x)$.

Let β be a probability measure on $(E \times \mathbb{R}_+,\ \mathcal{E} \times \mathcal{B}_+)$. The probability measure $\mathbb{P}_\beta$ is defined on $(\Omega,\ \mathcal{F})$ by

$$\mathbb{P}_\beta(C) = \int\int_{E\times\mathbb{R}_+} \beta(\mathrm{d}x \times \mathrm{d}s)\mathbb{P}_{(x,s)}(C), \qquad C \in \mathcal{F},$$

where $\mathbb{P}_{(x,s)}(C) = \mathbb{P}(C|J_0 = x,\ S_0 = s)$.

Let $L^1(\beta Q)$ be the space of all real βQ-measurable integrable functions g defined on $E \times E \times \mathbb{R}_+$. The functional μQ is defined on $L^1(\mu Q)$ by

$$\mu Q g = \int\int\int_{E\times E\times\mathbb{R}_+} \mu(\mathrm{d}x)Q(x,\ \mathrm{d}y \times \mathrm{d}s)g(x,\ y,\ s), \qquad g \in L^1(\beta Q).$$

We set $\hat{g} = \nu Q g$, and, if g is a function of only one variable $x \in E$, we put $\mu g = \mu Q g = \int_E \mu(\mathrm{d}x)g(x)$.

We will denote by $\mathbb{E}_\beta$ the expectation, by var_β the variance, and by cov_β the covariance with respect to $\mathbb{P}_\beta$.

The following assumptions are stated here for future reference. Note that μ is some reference measure on $(E,\ \mathcal{E})$.

Assumption 1. *The Markov chain* **J** *is Harris positive with a stationary distribution ν.*

Assumption 2. *The mean sojourn times are integrable with respect to the stationary distribution of* **J***, that is, $\hat{m} = \int_E \nu(\mathrm{d}x)m(x) < +\infty$.*

Assumption 3. *The semi-Markov process* **Z** *is regular, that is, $\mathbb{P}_{(x,0)}(N(t) < +\infty) = 1$, for $(x,\ t) \in E \times \mathbb{R}_+$.*

Assumption 4. *For every $x \in E$, the semi-Markov kernel is absolutely continuous with respect to $\mu \times \lambda$, with Radon–Nikodym derivative q, that is $Q(x, A \times \Gamma) = \int\int_{A\times\Gamma} q(x, y, s)\mu(\mathrm{d}y)\mathrm{d}s$, for $x \in E$, $A \in \mathcal{E}$, and $\Gamma \in \mathcal{B}_+$.*

Assumption 5. *The logarithm of q is νQ-integrable, that is, $\int\int\int_{E\times E\times\mathbb{R}_+} |\log q(x, y, s)|\nu(\mathrm{d}x)Q(x, \mathrm{d}y \times \mathrm{d}s) < +\infty$.*

Assumption 6. *The initial distribution of the process is absolutely continuous with respect to $\mu \times \lambda$, with density α, that is, $\mathbb{P}((J_0, S_0) \in B \times \Gamma) = \int\int_{B\times\Gamma} \alpha(x, s)\mu(\mathrm{d}x)\mathrm{d}s$ for $B \in \mathcal{E}$ and $\Gamma \in \mathcal{B}_+$, with α bounded away from 0 and infinity.*

Assumption 7. *The hazard function r is uniformly bounded on E by an increasing function: $\sup_{x\in E} r(x, t) \leqslant \sqrt{t}/\sqrt{2\log\log t}$, for $t \geqslant t_0$, for some $t_0 > 0$.*

Assumption 8. *We have $C_1 = \mathbb{E}\sup_{k\geqslant 1}|\log q(J_{k-1}, J_k, X_k)| < +\infty$.*

Assumptions 1–3 amount to the ergodicity of the semi-Markov process $\mathbf{Z}$. Assumptions 4 and 6 are necessary for the log-likelihood of $\mathbf{Z}$ to be defined; Assumption 6 is especially satisfied for $\alpha = \delta_{(x,s)}$, that is, if the process is known to start from state x at time s. Assumption 5 is necessary for the entropy rate of $\mathbf{Z}$ to be defined (see relation (5)). Assumption 7 is sufficient for the backward recurrence time of $\mathbf{Z}$ to be asymptotically negligible; it is trivially fulfilled by Markov processes with bounded infinitesimal generators, and more generally by semi-Markov processes with decreasing hazard functions — for example, with Weibull transition functions with shape parameters less than 1. Assumption 8 is classical for inducing mean convergence from almost sure convergence.

3. Entropy of a semi-Markov process

Suppose we are given one observation of the semi-Markov process $\mathbf{Z}$ on the time interval $[0, T]$ for some $T > 0$, that is, $Z_{[0,T]} = (Z(t); 0 \leqslant t \leqslant T)$, or equivalently $(J_0, S_0, J_1, X_1, \ldots, J_{N(T)}, X_{N(T)}, U_T)$, with $U_T = t - S_{N(T)}$. Clearly, the likelihood of $Z_{[0,T]}$ with respect to $\mu \times \mu \times \lambda$ is

$$f_T(Z_{[0,T]}) = \alpha(J_0, S_0)\left[\prod_{k=1}^{N(T)} q(J_{k-1}, J_k, X_k)\right]\overline{H}(J_{N(T)}, U_T). \tag{9}$$

The proof of the strong AEP will be based on the following four lemmas. First, it is easy to see that $(J_{n-1}, J_n, S_n; n \in \mathbb{N}^*)$ is a Markov renewal process with state space $E \times E \times \mathbb{R}_+$ and semi-Markov kernel defined by

$$\tilde{Q}((x, y), A \times B \times \Gamma) = \mathbb{1}_A(y)Q(y, B \times \Gamma), \qquad x, y \in E,\ A, B \in \mathcal{E}, \Gamma \in \mathcal{B}_+, \tag{10}$$

where $\mathbb{1}_A$ is the indicator function of the set A.

Lemma 1. *The distribution probability* $\nu^\sharp = \nu Q$ *is a stationary distribution of the Markov chain* $(J_{n-1}, J_n, X_n;\ n \in \mathbb{N})$.

Proof. We have

$$\begin{aligned}
\nu^\sharp \tilde{Q}(A \times B \times \Gamma) &= \iiint_{E\times E\times\mathbb{R}_+} \nu^\sharp(\mathrm{d}x \times \mathrm{d}y \times \mathrm{d}s)\tilde{Q}((x, y), A \times B \times \Gamma) \\
&= \int\int\int_{E\times E\times\mathbb{R}_+} \nu(\mathrm{d}x)Q(x, \mathrm{d}y \times \mathrm{d}s)\mathbb{1}_A(y)Q(y, B \times \Gamma) \\
&= \int\int_{E\times A} \nu(\mathrm{d}x)P(x, \mathrm{d}y)Q(y, B \times \Gamma) \\
&= \int_A \nu(\mathrm{d}y)Q(y, B \times \Gamma) \\
&= \nu^\sharp(A \times B \times \Gamma).
\end{aligned}$$

Moreover, $\nu^\sharp(E \times E \times \mathbb{R}_+) = \int_E \nu(\mathrm{d}x)Q(x, E \times \mathbb{R}_+) = 1$, which completes the proof. □

Note that under Assumption 1, the Markov chain (J_{n-1}, J_n, X_n) has a unique stationary distribution, given by Lemma 1.

Lemma 2. *Let* $(N(t);\ t \in \mathbb{R}_+)$ *be as defined in* (6). *Under Assumptions 1–3, if* $g \in L^1(\nu Q)$, *the following convergence holds for any* $\alpha \in \mathcal{P}(E)$:

$$\frac{1}{N(t)} \sum_{k=1}^{N(t)} g(J_{k-1}, J_k, X_k) \to \hat{g}, \qquad t \to +\infty,\ \mathbb{P}_\alpha\text{-a.s.}$$

Proof. From the ergodic theorem for Markov chains (see, for example, Meyn and Tweedie 1996, Theorem 17.0.1) we know that $n^{-1}\sum_{k=1}^{n} g(J_{k-1}, J_k, X_k)$ converges to $\mathbb{E}_{\nu^\sharp}[g(J_0, J_1, X_1)]$. From Lemma 1 applied to the Markov chain (J_{n-1}, J_n, X_n), we obtain that $\mathbb{E}_{\nu^\sharp}[g(J_0, J_1, X_1)] = \nu^\sharp g = \hat{g}$, and hence $n^{-1}\sum_{k=1}^{n} g(J_{k-1}, J_k, X_k)$ tends $\mathbb{P}_\alpha$-a.s. to $\hat{g}$ when n tends to infinity. From Assumption 3, we know that $N(t)$ tends $\mathbb{P}_\alpha$-a.s. to infinity as t tends to infinity (see, for example, Limnios and Oprişan 2001a). Then the law of large numbers for a sum of a random number of terms (see, for example, Gut 1988, Theorem I.2.1) yields the result. □

The following lemma is proven for example in Limnios and Oprişan (2001a) for a countable state space. For the sake of completeness, we will here give a proof for a Borel state space.

Lemma 3. *Under Assumptions 1–3, the following convergence holds for any* $\alpha \in \mathcal{P}(E)$:

$$\frac{N(t)}{t} \to \frac{1}{\hat{m}}, \qquad t \to +\infty, \ \mathbb{P}_\alpha\text{-a.s.}$$

Proof. From Grigorescu and Oprişan (1976) (see also Limnios and Oprişan 2001a, Theorem 3.13, p. 79), we know that $n^{-1}\sum_{k=0}^{n} X_k$ tends $\mathbb{P}_\alpha$-a.s. to $\hat{m}$ as n tends to infinity. Since $N(t)$ tends $\mathbb{P}_\alpha$-a.s. to infinity as t tends to infinity, we obtain that $N(t)^{-1}\sum_{k=0}^{N(t)} X_k$ tends $\mathbb{P}_\alpha$-a.s. to $\hat{m}$ as t tends to infinity (see the proof of Lemma 2). We have $\sum_{k=0}^{N(t)} X_k = S_{N(t)}$ and

$$\frac{S_{N(t)}}{N(t)} \leqslant \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)}.$$

For t tending to infinity, this yields the result. □

Lemma 4. *Let* $\overline{H}$ *be given as in* (7). *Under Assumptions 1–3 and 7, the following convergence holds for any* $\alpha \in \mathcal{P}(E)$:

$$\frac{\log[\overline{H}(J_{N(T)}, U_T)]}{T} \to 0, \qquad T \to +\infty, \ \mathbb{P}_\alpha\text{-a.s.}$$

Proof. We have, for $B_t = \sqrt{2t \log\log t}$,

$$\left| \frac{1}{T} \log[\overline{H}(J_{N(T)}, U_T)] \right| = \frac{1}{T} \int_0^{U_T} r(J_{N(T)}, u) \mathrm{d}u \leqslant U_T \frac{U_T}{T B_{U_T}} \leqslant \frac{U_T}{B_T} \leqslant \frac{X_{N(T)+1}}{B_T}.$$

Let us prove that

$$\frac{X_{N(T)+1}}{B_T} = \frac{X_{N(T)+1}}{B_{N(T)+1}} \frac{B_{N(T)+1}}{B_T} \to 0, \qquad T \to +\infty, \ \mathbb{P}_\alpha\text{-a.s.} \tag{11}$$

Using Lemma 3, it is easy to prove that

$$\frac{B_{N(T)+1}}{B_T} = \sqrt{\log\left[\frac{\log N(T)}{\log T}\right]} \sqrt{\frac{N(T)}{T}} \to \frac{1}{\sqrt{\hat{m}}}, \qquad T \to +\infty, \ \mathbb{P}_\alpha\text{-a.s.}$$

Since the inter-jump times process (X_n) satisfies the law of the iterated logarithm (see Limnios and Oprişan 2001a, Theorem 3.14, p. 79), we know that X_n/B_n tends $\mathbb{P}_\alpha$-a.s. to zero as n tends to infinity. Finally, since $\log\log(n+1)/\log\log n$ tends to 1 as n tends to infinity and $N(t)$ tends $\mathbb{P}_\alpha$-a.s. to infinity as t tends to infinity, (11) holds true (see the proof of Lemma 2), which proves the lemma. □

The AEPs for semi-Markov processes with Borel state spaces follow.

Theorem 1 (Strong AEP for a semi-Markov process). *Let* $\mathbf{Z}$ *be a semi-Markov process with Borel state space* $(E, \mathcal{E})$. *If Assumptions 1–7 are fulfilled, then the following convergence holds for any* $\alpha \in \mathcal{P}(E)$:

$$-\frac{1}{T} \log f_T(Z_{[0,T]}) \to \mathbb{H}(\mathbf{Z}) = -\frac{\nu Q \log q}{\hat{m}} = -\frac{\widehat{\log q}}{\hat{m}}, \qquad T \to +\infty, \ \mathbb{P}_\alpha\text{-a.s.}, \tag{12}$$

where $\log q$ *stands for the composite function* $\log \circ q$.

This defines the entropy rate of the process $\mathbf{Z}$, which can also be written as in (5).

Proof. Under Assumption 4, the entropy $\mathbb{H}_T(\mathbf{Z})$ of $\mathbf{Z}$ up to time T is defined for any $T \in \mathbb{R}_+$. By (9), the log-likelihood of $Z_{[0,T]}$ is

$$\log f_T(Z_{[0,T]}) = \log \alpha(J_0, S_0) + \sum_{k=1}^{N(T)} \log q(J_{k-1}, J_k, X_k) + \log[\overline{H}(J_{N(T)}, U_T)]. \tag{13}$$

From Assumption 6, we deduce that $[\log \alpha(J_0, S_0)]/T$ tends $\mathbb{P}_\alpha$-a.s. to 0 as T tends to infinity. From Lemma 4, $[\log \overline{H}(J_{N(T)}, U_T)]/T$ tends $\mathbb{P}_\alpha$-a.s. to 0. From both Lemmas 2 and 3 and Assumption 5, we obtain that $T^{-1}\sum_{k=1}^{N(T)} \log q(J_{k-1}, J_k, X_k)$ tends $\mathbb{P}_\alpha$-a.s. to $\widehat{\log q}/\hat{m}$, which concludes the proof. □

Theorem 2 (Mean AEP for a semi-Markov process). *If Assumptions 1–8 are fulfilled, then the convergence in* (12) *holds also in mean.*

Proof. From Assumption 6, $\log \alpha(J_0, X_0)$ is finite so $\log \alpha(J_0, X_0)/T$ is bounded for T large enough, say by C_2. From Lemma 4, we deduce that $\log[\overline{H}(J_{N(T)}, U_T)]/T$ is also bounded for T large enough, say by C_3. From Lemma 3, $N(T)/T$ tends $\mathbb{P}_\alpha$-a.s. to a constant, and, thanks to Assumption 3, $N(T)/T$ is $\mathbb{P}_\alpha$-a.s.finite, so $N(T)/T$ is bounded for T large enough, say by K. Therefore, we deduce from (13) that for T large enough,

$$\left| \frac{\log f_T(Z_{[0,T]})}{T} \right| \leqslant K \sup_{k \geqslant 1} |\log q(J_k, J_{k-1} X_k)| + C_2 + C_3, \qquad \mathbb{P}_\alpha\text{-a.s.}$$

Hence, thanks to Assumption 8, the dominated convergence theorem applies to prove that the convergence in (12), which holds $\mathbb{P}_\alpha$-a.s., also holds in mean. □

It is worth noticing that convergence in distribution instead of almost sure convergence is sufficient to induce convergence in mean in Theorem 2 under Assumption 8.

4. Some particular semi-Markov processes

For some subclasses of semi-Markov processes, that are especially used in applications, the entropy rate and the assumptions for the AEP to hold deserve to be specifically stated. First, we consider the AEP for countable semi-Markov processes already obtained in Girardin and Limnios (2003) by proving convergence results on the number of transitions from one state to another. Then we consider semi-Markov chains with Borel state spaces for which no AEP has ever been proven before. Finally, we recall the conditions for the AEP to hold for countable semi-Markov chains, a result already obtained in Girardin and Limnios (2004) through specific means too. Note that the setting here is even more general than the particular cases generally encountered in the literature, since time 0 need not be a renewal time.

4.1. Countable semi-Markov processes

We deduce from Theorem 1 the explicit expression for the entropy rate of a countable semi-Markov process $\mathbf{Z}$. The semi-Markov kernel of the process $\mathbf{Z}$ can be written $Q = (Q_{xy}(t);\, x,\, y \in E,\, t \geqslant 0)$, where $Q_{xy}(t) = Q(x, \{y\} \times [0, t]) = \mathbb{P}(J_{n+1} = y, S_{n+1} - S_n \leqslant t | J_n = x)$. The reference measure μ is the counting measure on E.

Assumption 6 is naturally fulfilled since α denotes here the initial distribution itself, and Assumptions 1–5 and 8 can respectively be restated as follows:

Assumption 1′. $\mathbf{J}$ *is recurrent positive with a stationary distribution* ν.

Assumption 2′. $\sum_{x \in E} \nu(x) m(x) < +\infty$.

Assumption 3′. $\mathbf{Z}$ *is regular.*

Assumption 4′. $Q_{xy} \ll \lambda$ *with derivative* q_{xy}.

Assumption 5′. $\log q \in L^1(\nu Q)$.

Assumption 8′. $\mathbb{E} \sup_{k \geqslant 1} |\log q_{J_{k-1}, J_k}(X_k)| < +\infty$.

The expression for the entropy rate follows straightforwardly.

Corollary 1 (AEP for a countable semi-Markov process). *Let* $\mathbf{Z} = (Z(t);\, t \in \mathbb{R}_+)$ *be a semi-Markov process with countable state space E. If Assumptions 1′–5′ and 7 are fulfilled, then* $-\log f_T(Z_{[0,T]})/T$ *tends* $\mathbb{P}_\alpha$*-a.s. to* $\mathbb{H}(\mathbf{Z})$ *for any* $\alpha \in \mathcal{P}(E)$*, where* $\mathbb{H}(\mathbf{Z})$ *is given by* (3)*, with* $\hat{m} = \sum_{x \in E} \nu(x) m(x)$*. If Assumption 8′ is fulfilled, the above convergence also holds in mean.*

4.2. Semi-Markov chains

The strong AEP specializes to discrete time semi-Markov processes, also called semi-Markov chains, in the following way. The counting measure on $\mathbb{N}$ replaces the Lebesgue measure λ and we set $q(x, A, k) = Q(x, A \times \{k\}) = \mathbb{P}(J_{n+1} \in A, S_{n+1} - S_n = kJ_n = x)$, for $k \in \mathbb{N}$ (see, for example, Barbu *et al.* 2004).

Assumption 3 is always fulfilled and Assumption 4 is irrelevant. Since the hazard rate is a probability here, Assumption 7 is also always fulfilled.

Corollary 2 (AEP for a semi-Markov chain). *Let* $\mathbf{Z} = (Z_n;\, n \in \mathbb{N})$ *be a semi-Markov chain* *with Borel state space $(E, \mathcal{E})$. If Assumptions 1, 2, 5′ and 6 are fulfilled, then the following convergence holds $\mathbb{P}_\alpha$-a.s. for any $\alpha \in \mathcal{P}(E)$:*

$$-\frac{1}{n}\log f_n(Z_{[0,n]}) \to \mathbb{H}(\mathbf{Z}) = -\frac{1}{\hat{m}}\sum_{k\in\mathbb{N}}\int\int_{E\times E}\nu(\mathrm{d}x)q(x,\,\mathrm{d}y,\,k)\log q(x,\,y,\,k), \qquad n \to +\infty.$$

If Assumption 8 is fulfilled, the above convergence also holds in mean.

The AEP for a countable semi-Markov chain also appears as a direct consequence of Theorem 1. The semi-Markov kernel of the process is $(q_{xy}(k);\, x,\, y \in E,\, k \in \mathbb{N}^*)$ with $q_{xy}(k) = \mathbb{P}(J_{n+1} = y,\, S_{n+1} - S_n = k | J_n = x)$. The mean sojourn time in $x \in E$ is $m(x) = \sum_{k\geqslant 0}\sum_{y\in E} q_{xy}(k)$.

Assumptions 3 and 7 are always fulfilled and Assumptions 4 and 6 are irrelevant. Under Assumptions 1′, 2′ and 5′, we easily obtain (4), with $\hat{m} = \sum_{x\in E}\nu(x)m(x)$.

The expression (1) for the entropy rate of a countable Markov chain follows straightforwardly.

5. AEP for pure jump Markov processes

Jump Markov processes and Markov chains also constitute particular cases of semi-Markov processes. Due to their paramount importance, we present them separately.

If the process $\mathbf{Z}$ is an ergodic homogeneous Markov process, it is stationary with respect to $\mathbb{P}_\pi$, where π is its stationary distribution. Thus, for a Markov chain with a Borel state space, the strong AEP with respect to $\mathbb{P}_\pi$ and the mean AEP can be obtained by applying Barron (1985) or Orey (1985). The process may be asymptotically mean stationary or Markovian with order more than 1 (see Barron 1985) or nearly Markovian (see Orey 1985). If the process is not ergodic, convergence to an invariant function can still be proven to hold (see Barron 1985). For a continuous time Markov process, the mean AEP can be obtained by applying Perez (1964). None of these AEPs include an explicit expression for the limit entropy rate. Note that Bad Dumitrescu (1988) computes the explicit entropy rate (2) of a pure jump Markov process with finite state space as the limit of its entropy up to time ar T divided by T.

The ergodic homogeneous Markov process $\mathbf{Z}$ is no longer stationary with respect to $\mathbb{P}_\alpha$ for any initial distribution α on E and the above results do not apply. The strong AEP with respect to $\mathbb{P}_\alpha$, and the mean AEP for any pure jump Markov process or Markov chain with a Borel state space, with an explicit expression for the entropy rate, are obtained here as special cases of Theorem 1.

5.1. Borel state space jump Markov processes

Let us consider a pure jump Markov process $\mathbf{Z}$ with generating operator A, defined on the set of continuous functions φ on E by $A\varphi(x) = a(x)\int_E P(x,\,\mathrm{d}y)[\varphi(y) - \varphi(x)]$, where $(P(x,\,A);\, x \in E,\, A \in \mathcal{E})$ is the transition function of the embedded Markov chain $\mathbf{J}$, and

$(a(x);\ x \in E)$ is the intensity of the jump function (see Limnios and Oprișan 2001a). The semi-Markov kernel of $\mathbf{Z}$ is given by $Q(x, A \times [0, t]) = P(x, A)[1 - \mathrm{e}^{-a(x)t}]$ for $x \in E$, $A \in \mathcal{E}$, and $t \geqslant 0$, where P is the transition function of $\mathbf{J}$. Note that $r(x, t) = a(x)$ and $H(x, t) = 1 - \exp[-a(x)t]$ for $x \in E$ and $t \in \mathbb{R}_+$.

A jump Markov process $\mathbf{Z}$ fulfilling Assumptions 1–3 is said to be ergodic. These three assumptions are equivalent to the two following conditions: 0 is a simple eigenvalue of the generator of $\mathbf{Z}$ and $\|P_t(x, A) - \pi(A)\|$ tends to zero as t tends to infinity, for $x \in E$ and $A \in \mathcal{E}$ (for the total variation norm), where $P_t(x, B) = \mathbb{P}(Z_t \in B | Z_0 = x)$ is the transition function and π the stationary distribution of $\mathbf{Z}$. Assumptions 4–8 can here respectively be restated as follows:

Assumption 4″. $P(x, A) = \int_A p(x, y)\mu(\mathrm{d}y)$, *for* $x \in E$ *and* $A \in \mathcal{E}$.

Assumption 5″. $\int\int_{E\times E} p(x, y)\log[p(x, y)a(x)]\nu(\mathrm{d}x)\mu(\mathrm{d}y) < +\infty$.

Assumption 6″. $\mathbb{P}(J_0 \in A) = \int_A \alpha(x)\mathrm{d}x$ *with* α *bounded away from 0 and infinity.*

Assumption 7″. $\mathbf{Z}$ *has a bounded generating operator, that is,* $\sup_{x\in E} a(x) < +\infty$.

Assumption 8″. $\mathbb{E}\sup_{k\geqslant 1} p(J_{k-1}, J_k) < +\infty$.

Corollary 3 (AEP for a jump Markov process). *Let* $\mathbf{Z} = (Z(t);\ t \in \mathbb{R}_+)$ *be a pure jump Markov process with Borel state space* $(E, \mathcal{E})$*. If Assumptions 1–3 and 4″–7″ are fulfilled, the following convergence holds* $\mathbb{P}_\alpha$*-a.s. for any* $\alpha \in \mathcal{P}(E)$*:*

$$-\frac{1}{T}\log f_T(Z_{[0,T]}) \to \mathbb{H}(\mathbf{Z}) = -\int\int_{E\times E} \pi(\mathrm{d}x)p(x, y)a(x)(\log[p(x, y)a(x)] - 1)\mu(\mathrm{d}y), \qquad T \to +\infty.$$

If, moreover, Assumption 8″ is fulfilled, then the convergence also holds in mean.

Proof. Thanks to Assumption 4″, Q is absolutely continuous with respect to $\mu \times \lambda$. Its Radon–Nikodym derivative is $q(x, y, s) = p(x, y)a(x)\mathrm{e}^{-a(x)s}$. Theorem 1 gives

$$\begin{aligned}\mathbb{H}(\mathbf{Z}) &= -\frac{1}{\hat{m}}\int\int\int_{E\times E\times\mathbb{R}_+} \nu(\mathrm{d}x)p(x, y)a(x)\mathrm{e}^{-a(x)s}\log[p(x, y)a(x)\mathrm{e}^{-a(x)s}]\mu(\mathrm{d}y)\mathrm{d}s \\ &= -\frac{1}{\hat{m}}\int\int_{E\times E} \nu(\mathrm{d}x)p(x, y)(\log[p(x, y)a(x)] - 1)\mu(\mathrm{d}y).\end{aligned}$$

The stationary distributions ν of $\mathbf{J}$ and π of $\mathbf{Z}$ are linked through $\pi(\mathrm{d}x)a(x) = \nu(\mathrm{d}x)\int_E a(y)\pi(\mathrm{d}y)$. We compute

$$\hat{m} = \int_E \frac{\nu(\mathrm{d}x)}{a(x)} = \int_E \frac{\pi(x)}{\int_E a(z)\pi(\mathrm{d}z)} = \left[\int_E a(z)\pi(\mathrm{d}z)\right]^{-1},$$

and the result follows. □

For a countable state space, the semi-Markov kernel takes the form $Q_{xy}(t) = P(x, y)(1 - \exp[-a(x)t])$, and the stationary distributions of the process and of its embedded chain are linked through $\pi(x)a(x) = \nu(x)\sum_{x \in E}\pi(x)a(x)$. The expression (2) follows straightforwardly.

5.2. Borel state space Markov chains

The semi-Markov kernel Q of a Markov chain $\mathbf{Z}$ with Borel state space $(E, \mathcal{E})$ and transition function P is as follows. Set $\rho(x) = P(x, \{x\})$ for $x \in E$. If $\rho(x) = 0$ then $Q(x, A \times \{k\}) = P(x, A)\mathbb{1}_{(k=1)}$. If $\rho(x) \neq 0$ (i.e. if x is an atom of P), then $Q(x, A \times \{k\}) = P_0(x, A)\rho(x)^{k-1}[1 - \rho(x)]$, where $P_0(x, A) = P(x, A)/\rho(x)$ if $A \neq \{x\}$, and $P_0(x, \{x\}) = 0$, are the transition probabilities of the embedded chain $\mathbf{J}$.

Assumption 5 amounts to $\int\int_{E \times E}\pi(\mathrm{d}x)P(x, \mathrm{d}y)|\log p(x, y)| < +\infty$, where π is the stationary distribution of $\mathbf{Z}$. Assumptions 1–3 amount to the ergodicity of $\mathbf{Z}$. Assumptions 4 and 6 take the form of Assumptions 4″ and 6″. Assumption 7 is always fulfilled. Under these assumptions, we compute

$$\mathbb{H}(\mathbf{Z}) = \int\int_{E \times E}\pi(\mathrm{d}x)P(x, y)\log P(x, y)\mathrm{d}x\,\mathrm{d}y.$$

6. Relative entropy

The relative entropy up to time T of a process $\mathbf{Z}$ with respect to another $\mathbf{W}$, say $\mathbb{H}_T(\mathbf{Z}|\mathbf{W})$, is defined as the relative entropy between their marginal distributions on $[0, T]$. If $\mathbb{H}_T(\mathbf{Z}|\mathbf{W})/T$ converges, the limit $\mathbb{H}(\mathbf{Z}|\mathbf{W})$ is called the relative entropy rate of $\mathbf{Z}$ with respect to $\mathbf{W}$.

Note that the entropy $\mathbb{H}_T(\mathbf{Z})$ with respect to a probability reference measure can also be seen as the relative entropy $\mathbb{H}_T(\mathbf{Z}|\mathbf{W})$ of the marginal distribution of $\mathbf{Z}$ with respect to the marginal distribution μ_T of some process $\mathbf{W}$; for this, see especially Pinsker (1964) and Perez (1964).

We consider here two semi-Markov processes $\mathbf{Z}$ and $\mathbf{W}$ with the same Borel state space. We will denote with a $\mathbf{W}$ or $\mathbf{Z}$ superscript the functions linked to $\mathbf{W}$ or $\mathbf{Z}$, respectively. The following assumptions are necessary for stating the strong and mean AEPs for the relative entropy:

Assumption 9. *The initial distribution of* $\mathbf{Z}$ *is absolutely continuous with respect to the initial distribution of* $\mathbf{W}$, *that is,* $\alpha^{\mathbf{Z}} \ll \alpha^{\mathbf{W}}$, *and* $q^{\mathbf{Z}}$ *is supported within the support of* $q^{\mathbf{W}}$.

Assumption 10. *The logarithm of* $q^{\mathbf{W}}$ *is* $\nu^{\mathbf{Z}}Q^{\mathbf{Z}}$ *integrable.*

Assumption 11. $\mathbb{E}\sup_{k \geqslant 1}|\log q^{\mathbf{W}}(J_k^{\mathbf{Z}}, J_{k-1}^{\mathbf{Z}}, X_k^{\mathbf{Z}})| < +\infty$.

Note that Assumption 9 is necessary for the relative entropy $\mathbb{H}_T(\mathbf{Z}|\mathbf{W})$ to be defined for any $t \in \mathbb{R}_+$, and implies Assumption 6 for $\mathbf{Z}$.

Theorem 3 (Relative entropy rate). *Let* $\mathbf{Z}$ *and* $\mathbf{W}$ *be two semi-Markov processes with the same Borel state space* $(E, \mathcal{E})$*. If Assumptions 1–5 and 7 for* $\mathbf{Z}$*, Assumptions 1–4 and 6–7 for* $\mathbf{W}$ *and Assumption 9 and 10 are fulfilled, then the following convergence holds* $\mathbb{P}_\alpha^{\mathbf{Z}}$*-a.s. for any* $\alpha \in \mathcal{P}(E)$*:*

$$\frac{1}{T}\log\left[\frac{f_T^{\mathbf{Z}}}{f_T^{\mathbf{W}}}(Z_{[0,T]})\right] \to \frac{1}{\hat{m}^{\mathbf{Z}}}\nu^{\mathbf{Z}}Q^{\mathbf{Z}}\log\left(\frac{q^{\mathbf{Z}}}{q^{\mathbf{W}}}\right), \qquad T \to +\infty.$$

If, moreover, Assumption 8 for $\mathbf{Z}$ *and Assumption 11 are fulfilled, then the above convergence also holds in mean.*

This defines the relative entropy rate of $\mathbf{Z}$ with respect to $\mathbf{W}$: explicitly,

$$\mathbb{H}(\mathbf{Z}\mathbf{W}) = \frac{1}{\hat{m}^{\mathbf{Z}}}\int\int\int_{E\times E\times\mathbb{R}_+} \nu^{\mathbf{Z}}(\mathrm{d}x)q^{\mathbf{Z}}(x, y, s)\log\left[\frac{q^{\mathbf{Z}}(x, y, s)}{q^{\mathbf{W}}(x, y, s)}\right]\mu(\mathrm{d}y)\mathrm{d}s.$$

Proof. On the one hand, by Theorem 1, $\mathbb{P}_\alpha^{\mathbf{Z}}$-a.s., as T tends to infinity,

$$\frac{1}{T}\log f_T^{\mathbf{Z}}(Z_{[0,T]}) \to -\mathbb{H}(\mathbf{Z}) = \frac{1}{\hat{m}^{\mathbf{Z}}}\int\int\int_{E\times E\times\mathbb{R}_+} \nu^{\mathbf{Z}}(\mathrm{d}x)q^{\mathbf{Z}}(x, y, s)\log q^{\mathbf{Z}}(x, y, s)\mathrm{d}y\,\mathrm{d}s.$$

On the other hand, by Assumption 6 for $\mathbf{W}$, by Lemma 4 applied to $\overline{H}^{\mathbf{W}}(J^{\mathbf{Z}}_{N^{\mathbf{Z}}(T)}, U^{\mathbf{Z}}_T)$ and by Lemma 2 applied to the function $g = \log q^{\mathbf{W}}$, we obtain that $-[N^{\mathbf{Z}}(T)]^{-1}\log f_T^{\mathbf{W}}(Z_{[0,T]})$ tends $\mathbb{P}_\alpha^{\mathbf{Z}}$-a.s. to $-\nu^{\mathbf{Z}}Q^{\mathbf{Z}}(\log q^{\mathbf{W}})$ as T tends to infinity. Therefore, by Lemma 3, $-[\log f_T^{\mathbf{W}}(Z_{[0,T]})]/T$ tends $\mathbb{P}_\alpha^{\mathbf{Z}}$-a.s. to $-\nu^{\mathbf{Z}}Q^{\mathbf{Z}}(\log q^{\mathbf{W}})/\hat{m}^{\mathbf{Z}}$ and the proof is complete. □

Explicit expressions for the relative entropy rate between Markov processes, semi-Markov chains or countable semi-Markov processes can be derived straightforwardly.

For application in information theory, mutual information is widely used; see Pinsker (1964), Perez (1964) or Barron (1985). The mutual information between two random variables X and Y with respective marginal distributions $\mathbb{P}_X$ and $\mathbb{P}_Y$ and joint distribution $\mathbb{P}_{(X,Y)}$ can be written as

$$\mathbb{I}(X, Y) = \mathbb{S}(\mathbb{P}_X) + \mathbb{S}(\mathbb{P}_Y) - \mathbb{S}(\mathbb{P}_{(X,Y)}) = \mathbb{S}(\mathbb{P}_Y) - \mathbb{S}(\mathbb{P}_Y\mathbb{P}_X) = \mathbb{S}(\mathbb{P}_X) - \mathbb{S}(\mathbb{P}_X\mathbb{P}_Y).$$

The definition of the mutual information rate of two processes follows. Mutual information thus appears just as a special case of relative entropy; therefore, all results concerning mutual entropy derive straightforwardly from the above results on entropy and relative entropy rate.

7. The weak invariance principle

The following additional assumptions are necessary for stating the functional central limit theorem for entropy for a semi-Markov process:

Assumption 12. *The means, with respect to the stationary distribution of the embedded chain, of the second moments of the sojourn times defined in* (8) *are finite, that is,* $\int_E \nu(\mathrm{d}x) m_2(x) < +\infty$.

Assumption 13. *The logarithm of the Radon–Nikodym derivative* q *of* Q *is* νQ*-square integrable.*

Assumptions 12 and 13 are necessary to define the asymptotic variance (14) below. Note that Assumption 13 obviously implies Assumption 4.

The weak invariance principle will be deduced from the invariance principle for an additive functional of a two-dimensional Markov renewal process. Let us consider the Markov renewal process $(J_{n-1}, J_n, X_n;\ n \in \mathbb{N})$, with the notation of Section 2. Its semi-Markov kernel is given by (10). As shown in Lemma 1, its stationary distribution is $\nu^\sharp = \nu Q$. Its embedded Markov chain $\tilde{\mathbf{J}} = (J_n, J_{n+1};\ n \in \mathbb{N})$ has the transition kernel $\tilde{P}((x, y); \mathrm{d}x' \times \mathrm{d}y') = \mathbb{1}_y(\mathrm{d}x') P(y, \mathrm{d}y')$ and the stationary distribution $\tilde{\nu}(\mathrm{d}x \times \mathrm{d}y) = \nu(\mathrm{d}x) P(x, \mathrm{d}y)$.

Let $V : E \times E \to [1, +\infty)$ be an $\mathcal{E} \times \mathcal{E}$-measurable function, and let $L_V^\infty(E \times E)$ denote the space of all Borel functions $\varphi : E \times E \to \mathbb{R}$ such that $|\varphi|_V = \sup_{(x,y)\in E\times E}[|\varphi(x, y)|/V(x, y)] < +\infty$.

Let γ be a signed measure on $(E \times E, \mathcal{E} \times \mathcal{E})$, bounded for the V-norm defined by $\|\gamma\|_V = \sup\{|\gamma\varphi|_V : \varphi \in L_V^\infty(E \times E), |\varphi|_V \leqslant 1\} = \sup_{\varphi \in L_V^\infty(E\times E)}[|\gamma\varphi|_V / |\varphi|_V]$. Note that for V constant and equal to 1, the V-norm $\|\cdot\|_1$ is the total variation norm.

Assumption 14. *There exists a function* $\phi : \mathbb{N} \to \mathbb{R}_+$*, with* $\sum_{n\geqslant 1} \sqrt{\phi(n)} < +\infty$*, such that* $\|\tilde{P}^n((x, y), \cdot \times \cdot) - \tilde{\nu}(\cdot \times \cdot)\|_V \leqslant \phi(n)$*, for* $x, y \in E$ *and* $n \in \mathbb{N}^*$.

Assumption 14 is a mixing condition. It is fulfilled if $\tilde{\mathbf{J}}$ is V-uniformly ergodic (with $\phi(n) = Rr^{-n}$ for non-negative constants $R \in \mathbb{R}_+$ and $r \in [0, 1[$), or equivalently if there exists a small set $\Delta \in \mathcal{E} \times \mathcal{E}$ such that the drift condition $\int_E P(y, \mathrm{d}y') V(y', y') \leqslant cV(x, y) + b\mathbb{1}_\Delta(x, y)$, for $x, y \in E$, holds for some constants $c \in]0, 1[$ and $b \in \mathbb{R}_+{}^*$; see Meyn and Tweedie (1996) for details.

Let us set $q_k = q(J_{k-1}, J_k, X_k)$ for $k \in \mathbb{N}^*$ and $G_n = \sum_{k=1}^n [-\log q_k - \mathbb{H}(\mathbf{Z})]$ for $n \geqslant 1$, and define the asymptotic variance

$$\sigma^2 = \mathrm{var}_{\tilde{\nu}}[\log q(J_0, J_1, X_1)] + 2 \sum_{k\geqslant 2} \mathrm{cov}_{\tilde{\nu}}[\log q(J_0, J_1, X_1), \log q(J_{k-1}, J_k, X_k)]. \quad (14)$$

Let $\Rightarrow$ denote weak convergence in the Skorohod space $D[0, +\infty)$, with respect to the probability $\mathbb{P}_{\tilde{\mu}}$, where $\tilde{\mu}(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x) P(x, \mathrm{d}y)$ for a probability measure μ on $(E, \mathcal{E})$.

Theorem 4 (Weak invariance principle for entropy). *If Assumptions 1–7 and 12–14 hold, then* $0 \leqslant \sigma^2 < +\infty$. *If, moreover,* $\sigma^2 > 0$, *then*

$$\frac{1}{\sqrt{n}}[-\log f_{nt}(Z_{[0,nt]}) - N(nt)\mathbb{H}(\mathbf{Z})] \Rightarrow bW, \qquad n \to +\infty,$$

where $b = \sigma/\sqrt{\hat{m}}$ *and* W *is the standard Brownian motion.*

Proof. On the one hand, thanks to Assumption 14, we can prove as in Limnios and Oprişan (2001b, Lemma 2) that $|\mathbb{P}_{\tilde{\nu}}(A \cap B) - \mathbb{P}_{\tilde{\nu}}(A)\mathbb{P}_{\tilde{\nu}}(B)| \leqslant \varphi_{\mathbb{P}_{\tilde{\nu}}}(n)\mathbb{P}_{\tilde{\nu}}(A)$, for all $A \in \sigma(J_{m-1}, J_m, X_m;\ m \leqslant k)$, $B \in \sigma(J_{m-1}, J_m, X_m;\ m \geqslant n+k)$, and $n, k \in \mathbb{N}^*$, where $\varphi_{\mathbb{P}_{\tilde{\nu}}}(n) = \phi(n-1)$ with ϕ given in Assumption 14. Since $\phi(n)$ tends to zero as n tends to infinity, the Markov chain (J_{n-1}, J_n, X_n) is ϕ-mixing under $\mathbb{P}_{\tilde{\nu}}$.

On the other hand, we obtain in the same way that $|\mathbb{P}_{\tilde{\nu}}(B) - \mathbb{P}_{\mu}(B)| \leqslant \phi(n)$, for all $B \in \sigma(J_{m-1}, J_m, X_m;\ m \geqslant n + k)$. Thanks to Assumptions 12 and 13, σ^2 is well defined. Therefore we deduce from Limnios and Oprişan (2001b, Theorem 2) that $(\sigma\sqrt{n})^{-1}G_{[nt]} \Rightarrow W$ as n tends to infinity. Finally, Lemma 3 and Gut (1988, Theorem V 2.1) together imply that $(\sqrt{n})^{-1}G_{N(nt)} \Rightarrow bW$ as n tends to infinity. □

8. The strong invariance principle

The following technical assumptions are necessary for stating the invariance principle for the law of the iterated logarithm:

Assumption 15. *The quantity* $\log q_n/B_n$ *tends* $\mathbb{P}_\alpha$*-a.s. to zero as* n *tends to infinity.*

Assumption 16. *The quantity* $\sup_{(x,t)\in E\times\mathbb{R}_+} r(x, t)$ *is finite.*

Theorem 5 (Strong invariance principle for entropy). *Under Assumptions 1–6 and 13–16, for any* $t \in \mathbb{R}_+$, *the sequence*

$$\left(\frac{\sqrt{\hat{m}}[-\log f_{nt} - N(nt)\mathbb{H}(\mathbf{Z})]}{\sigma B_n};\ n \geqslant 3\right),$$

where $B_n = \sqrt{2n\log\log n}$, *is* $\mathbb{P}_\alpha$*-a.s.a relatively compact subset of* $D[0, +\infty)$ *(in the uniform topology). The set of its limit points coincides with*

$$K = \left\{h \in \mathcal{AC}[0, \infty) : h(0) = 0,\ \int_0^{+\infty}\left[\frac{\mathrm{d}h}{\mathrm{d}t}(t)\right]^2 \mathrm{d}t \leqslant 1\right\},$$

in which $\mathcal{AC}[0, \infty)$ *stands for the space of absolutely continuous functions defined on* $\mathbb{R}_+$.

Proof. Let us set $\xi_n(t) = \sigma^{-1}[G_{[nt]} - (nt - [nt])\log q_{[nt]+1}]$. According to Herkenrath *et al.* (2003) (see also Heyde and Scott 1973), the sequence $(\xi_n/B_n;\ n \in \mathbb{N})$ is $\mathbb{P}_\alpha$-a.s. a relatively compact subset of $D[0, +\infty)$ and the set of its limit points coincides with K.

Since $B_{[nt]+1}/B_n$ tends to $\sqrt{t}$ for any $t > 0$, thanks to Assumption 15, we obtain that $[\log q_{[nt]+1}]/B_n$ tends $\mathbb{P}_\alpha$-a.s. to zero as n tends to infinity. Thanks to Assumptions 12 and 13, σ^2 is well defined. Hence, the sequence $(G_{[nt]}/\sigma B_n;\ n \geqslant 3)$ is also relatively compact, with limit set K, $\mathbb{P}_\alpha$-a.s.

Using Lemma 3, we obtain from Gut (1988, Theorem V 7.1) that $(\sqrt{\hat{m}} G_{N(nt)}/\sigma B_n;\ n \geqslant 3)$ is also $\mathbb{P}_\alpha$-a.s. relatively compact, with limit set K. We have $-\log f_{S_{N(nt)}} = -\log \alpha(J_0, S_0) - G_{N(nt)} + N(nt)\mathbb{H}(\mathbf{Z})$. Thus, due to Assumption 6, $(\sqrt{\hat{m}}[-\log f_{S_{N(nt)}} - N(nt)\mathbb{H}(\mathbf{Z})]/\sigma B_n;\ n \geqslant 3)$ is also relatively compact, with limit set K. Finally, $-\log f_{nt} = -\log f_{S_{N(nt)}} - \log[\overline{H}(J_{N(nt)}, U_{nt})]$, so

$$M_{N(nt)} = \sup_{S_{N(nt)} \leqslant s < S_{N(nt)+1}} |\log f_s - \log f_{S_{N(nt)}}| = \sup_{S_{N(nt)} \leqslant s < S_{N(nt)+1}} \log[\overline{H}(J_{N(nt)}, U_s)].$$

Thanks to Assumption 16, $(\sqrt{n})^{-1} M_{N(nt)}$ tends $\mathbb{P}_\alpha$-a.s. to zero, and the conclusion follows. □

Note that if Assumption 15 is not fulfilled, the result still holds true at the jump times, that is, for all $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$ such that $nt = S_n$.

Acknowledgement

The authors are grateful to an anonymous referee for his valuable comments which have greatly contributed to improving the presentation of this paper.

References

Bad Dumitrescu, M. (1988) Some informational properties of Markov pure-jump processes. *Časopis Pěstovani Mat.*, **113**, 429–434.

Barbu, V., Boussemart, M. and Limnios, N. (2004) Discrete time semi-Markov model for reliability and survival analysis. *Comm. Statist. Theory Methods*, **33**, 2833–2868.

Barron, A. (1985) The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman theorem. *Ann. Probab.*, **13**, 1292–1303.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Breiman, L. (1958) The individual ergodic theorem of information theory. *Ann. Math. Statist.*, **28**, 809–811. Correction (1960): **31**, 809–810.

Csizár, I. (1996) Maxent, mathematics, and information theory. In K.M. Hanson and R.N. Silver (eds), *Maximum Entropy and Bayesian Methods*, pp. 35–50. Dordrecht: Kluwer Academic.

Dym, H. (1966) A note on limit theorems for the entropy of Markov chains. *Ann. Math. Statist.*, **37**, 522–524.

Garret, A. (2001) Maximum entropy from the laws of probability. In M. Mohammad-Djafari (ed.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, pp. 3–22. Melville, NY: American Institute of Physics.

Girardin, V. (2004) Entropy maximization for Markov and semi-Markov processes. *Methodol. Comput. Appl. Probab.*, **6**, 109–127.

Girardin, V. (2005) On the different extensions of the ergodic theorem of information theory. In R. Baeza-Yates, J. Glaz, H. Gzyl, J. Hüsler and J.L. Palacios (eds), *Recent Advances in Applied Probability*, pp. 163–179. New York: Springer-Verlag.

Girardin, V. and Limnios, N. (2003) On the entropy of semi-Markov processes. *J. Appl. Probab.*, **40**, 1060–1068.

Girardin, V. and Limnios, N. (2004) Entropy rate and maximum entropy methods for countable semi-Markov chains. *Comm. Statist. Theory and Methods*, **33**, 609–622.

Grendar, M. and Grendar, M. (2001) What is the question MaxEnt answers? A probabilistic interpretation. In M. Mohammad-Djafari (ed.), *Bayesian inference and Maximum Entropy Methods in Science and Engineering*, pp. 83–93. Melville, NY: American Institute of Physics.

Grigorescu, S. and Oprişan, G. (1976) Limit theorems for J-X processes with a general state space. *Z. Wahrscheinlichkeitsheorie Verw. Geb.*, **35**, 65–73.

Gut, A. (1988) *Stopped Random Walks, Limit Theorems and Applications*. New York: Springer-Verlag.

Herkenrath, U., Iosifescu, M. and Rudolph, A. (2003) Letter to the editor. A note on invariance principles for iterated random functions. *J. Appl. Probab.*, **40**, 834–837.

Heyde, C.C. and Scott, D.J. (1973) Invariance principles for the law of iterated logarithm for martingales and processes with stationary increments. *Ann. Probab.*, **1**, 428–436.

Johnson, O. (2004) *Information Theory and the Central Limit Theorem*. London: Imperial College Press.

Kifer, Y. (1986) *Ergodic Theory of Random Transformations*. Boston: Birkhäuser.

Krengel, U. (1967) Entropy of conservative transformations. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **7**, 161–181.

Kullback, S. and Leibler, R.A. (1951) On information and sufficiency. *Ann. Math. Statist.*, **29**, 79–86.

Limnios, N. and Oprişan, G. (2001a) *Semi-Markov Processes and Reliability.* Boston: Birkhäuser.

Limnios, N. and Oprişan, G. (2001b) The invariance principle of an additive functional of a semi-Markov process. *Rev. Roumaine Math. Pures Appl.*, **44**, 75–83.

McMillan, M. (1953) The basic theorems of information theory. *Ann. Math. Statist.*, **24**, 196–219.

Meyn, S.P. and Tweedie, R.L. (1996) *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

O'Neil, R. (1990) The Shannon information on a Markov chain approximately normally distributed. *Colloq. Math.*, **60/61**, 569–577.

Orey, S. (1985) On the Shannon–Perez–Moy theorem. *Contemp. Math.*, **41**, 319–327.

Perez, A. (1964) Extensions of Shannon–McMillan's limit theorem to more general stochastic processes. In *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pp. 545–574. Prague: Czechoslovak Academy of Sciences.

Pinsker, M.S. (1964) *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day.

Shannon, C. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, 623–656.

Wen, L. and Weiguo, Y. (1996) An extension of Shannon–McMillan theorem and some limit properties for nonhomogeneous Markov chains. *Stochastic Process. Appl.*, **61**, 129–145.

Received December 2004 and revised October 2005