RESEARCH

Generalized kinetic theory of coarse-grained systems. I. Partial equilibrium and Markov approximations

Bernard Gaveau[1] and Michel Moreau[1][*]

[*]Correspondence:
[1]Sorbonne University, Paris, France

Abstract

The general kinetic theory of coarse-grained systems is presented in the abstract formalism of communication theory developed by Shannon and Weaver, Khinchin and Kolmogorov. The martingale theory shows that, under reasonable, general hypotheses, coarse-grained systems can be approximated by generalized Markov systems. For mixing systems, the Kolmogorov entropy production can be defined for nonstationary processes as Kolmogorov defined it for stationary processes.

Mathematics Subject Classification: 60G99

Keywords: Stochastic processes; Coarse-graining; Kolmogorov entropy; Markov processes; Martingales

1 Introduction

The purpose of this article (and the one to follow) is to define a generalized kinetic theory of classical thermodynamic systems at a coarse-grained level (see Sect. 2 for definitions). The microscopic evolution of the system induces an evolution on the coarse-grained states, which is generally non-Markovian.

In the same context, it has been shown recently [1] that coarse-grained deterministic dynamical systems can be approximated by generalized Markov systems, which may explain why Markov processes are so popular in modeling actual phenomena. These conclusions were obtained by applying and extending some pioneering results of Kolmogorov [2–4]. The formalism used in our previous works was relatively intuitive, even if sometimes lengthy, but it was sufficient for our first aim. However, we had to adopt some hypotheses that could seem reasonable but were difficult to justify precisely.

In the present work, we adopt a more abstract and rigorous formalism and show that the previous results can be generalized to a much broader framework, as mentioned above. This formalism is the one used in communication theory by Shannon-Weaver [5] and by Khinchin [6] to define optimal coding, by Kolmogorov [2, 3] (and also [4] for a pedagogical exposition) to define entropic invariants of dynamical systems. It was also introduced in [7] in the Markovian situation only.

The system evolution is specified by a stationary distribution on the path space X^Z, where X is the finite set of coarse-grained states, and Z represents the discrete time (see Sect. 2). At the coarse-grained level, the stationary evolution is not Markovian, but the advantage is that the evolution takes place on the *finite* state space X and that we avoid all controversial discussions concerning ergodicity and time scales for reaching equilibrium [8, 9].

Sections 2 and 3 fix notations and definitions and give basic examples. Section 4 introduces nonstationary processes: the initial condition is not the stationary state on X, but the evolution is given by the stationary process. It corresponds to the notion of partial equilibrium of Landau-Lifschitz [10]. We define the entropy production of both processes and show that they are equal, assuming a mixing property in Sect. 5. Although this result seems obvious, its proof is quite lengthy.

In Sect. 6, we address the main question of kinetic theory, namely why the evolution can be approximated by a Markovian evolution, as in the theory of Brownian motion, Fokker-Planck equation, etc. Obviously, one also has to use a coarse-grained time scale. We define various Markovian evolutions and prove that they approximate the exact evolution on the coarse-grained time scale using the production of relative entropy.

We want to dedicate this article to the memory of Prof. Mark Kac, who introduced one of us to the problems of justification of the Markov processes in statistical mechanics.

2 Notations and definitions

In this article, X denotes a finite set. Elements of X are denoted as $x \in X$. Z is the set of positive, 0, or negative integers.

2.1 The spaces X^Z, X

X^Z is the space of sequences $(x(n))$, $n \in Z$, $x(n) \in X$.

We define the shift $\tau \colon X^Z \to X^Z$ by

$$\tau\big(x(.)\big)(n) = x(n-1) \tag{2.1}$$

Let $I = \{i_1, \ldots, i_l\}$ be a finite subset of Z ordered by $i_1 < i_2 < \cdots < i_l$. We define for $k \in Z$

$$I + k = \{i_{1+k}, \ldots, i_l + k\} \tag{2.2}$$

X^I denotes the set of maps $x(I) \in X^I$.

In expanded notations, we write

$$x(I) \equiv (x_1, i_1; x_2, i_2; \ldots; x_l, i_l) \tag{2.3}$$

with

$$x_{k=s(I)(i_k)}$$

The shift is also defined as $\tau \colon X^I \to X^{I+1}$ by

$$\tau\big(x(I)\big) \equiv (x_1, i_1 + 1 : x_2, i_2 + 1 :; \ldots; x_l, i_l + 1) \tag{2.4}$$

for $x(I)$ given by (2.3).

If $J \subset I$ is a subset of I, $x(I)|_J$ is the restriction of the map $x(I)$ to the subset J.

Finally, if $m \leq n$, and $I = \{m, m + 1, \dots, n\}$ is the interval of integers between m and n, we denote

$$x(I) \equiv x[m, n] \tag{2.5}$$

2.2 Probabilities on X^Z

A stochastic process on X is the data of a system of probabilities p_I on X^I or all finite set $I \subset Z$ with the compatibility conditions:

if $J \subset I$ and $x(J) \in X^J$, then

$$p_J(x(J)) = \sum_{\{x(I) \in X^I | x(I)|_J = x(J)\}} p_I(x(I)) \tag{2.6}$$

Obviously, the p_I are known as soon as the $p_{[m,n]}$ are known.

It is known that a system of probabilities p_I satisfying the compatibility conditions of Eq. (2.6) defines a probability p sur X^Z, the p_I being the marginal laws of p

$$p_I(x(I)) = p(x(Z) \in X^Z | x(Z)|_I = x(I)) \tag{2.7}$$

This result is the extension Theorem of Kolmogorov [11]. The probability p is defined on the measurable subsets of X^Z. By definition, the subset appearing in p in Eq. (2.7) is measurable.

The stochastic process p is stationary if for any measurable set $A \subset X^Z$

$$p(\tau(A)) = p(A)$$

or equivalently, for any $x(I) \in X^I$, and any I

$$p_{I+1}(\tau(x(I)) = p_I(x(I))) \tag{2.8}$$

In particular, if p is stationary, it defines a unique probability distribution p_0 on X by

$$p_0(x) = p_{[n]}((x, n))$$

which is independent of n.

Remark (Conventions and definition)

(i) In order to simplify the notations, we shall skip the index I of p_I whenever it is clear that we refer to p_I. For instance, we write $p(x(I))$ instead of $p_I(x(I))$.

(ii) When we use conditional probabilities, the condition is always in the past: for instance, if $m < n$

$$p(x(n))|x[m, n - 1]) \equiv p_{[m,n]}(x(n))|x[m, n - 1]) \equiv \frac{p_{[m,n]}(x[m, n])}{p_{[m,n-1]}(x[m, n - 1])}$$

(iii) According to the usual definition [4], the stochastic process p is ergodic if any measurable set $B \subset X^Z$ is invariant by τ having probability $p(B) = 0$ or 1.

2.3 Coarse-graining

Let A be a partition of X: the elements of $a \in A$ are subsets of $a \subset X$ such that

$$X = \bigcup_{a \in A} a \tag{2.9}$$

$$a \cap a' = \phi \quad \text{if } a' \neq a \tag{2.10}$$

A probability q on X generates a probability $q^{(A)}$ on A by

$$q^{(A)}(a) = \sum_{x \in A} q(x) \tag{2.11}$$

The partition A on X induces the partitions A^Z of X^Z and A^I of X^I. The stochastic process p induces a stochastic process $p^{(A)}$ defined by

$$p_I^{(A)}\big(a(I)\big) = \sum_{x(I) \in a(I)} p_I\big(x(I)\big) \tag{2.12}$$

where the notation $x(I) \in a(I)$ means

$$\big(x(I) \in a(I))\big) \quad \Leftrightarrow \quad \big(x(I)(i) \in a(I)(i) \text{ for all } i \in I\big) \tag{2.13}$$

$p^{(A)}$ is a coarse-grained process of p. If p is stationary, $p^{(A)}$ is stationary. If p is ergodic, $p^{(A)}$ is ergodic.

Such coarse-grained processes are extensively used in physics and applied sciences when inaccurate observations cannot allow one to distinguish two different elements x belonging to the same subset a of A [1, 12].

3 Examples

We only cite a few well-known processes that are of interest to us.

(a) *Bernoulli processes*

Let μ be a probability on X. The Bernoulli process defined by μ is

$$p_I\big(x(I)\big) = \prod_{i \in I} \mu\big(x(I)(i)\big) \tag{3.1}$$

It is stationary. It is ergodic if and only if $\mu(x) > 0$ for any $x \in X$.

(b) *Markov processes*

Let $R = (R_{yx})$, $y, x \in X$ be a stochastic matrix, so

$$\sum_y R_{yx} = 1, \quad 0 \leq R_{yx} \leq 1 \tag{3.2}$$

Let μ be a stationary probability for R:

$$\mu(y) = \sum_x R_{yx} \mu(x) \tag{3.3}$$

Then, we define a stochastic process by

$$p_{[m,n]}\big(x[m,n]\big) = R_{x_n x_{n-1}} \cdots R_{x_k x_{k-1}} \cdots R_{x_{m+1} x_m} \mu(x_m) \tag{3.4}$$

where

$$x([m,n] = (x_m, m; x_{m+1}, m+1 : \cdots x_n, n) \tag{3.5}$$

This process is stationary. It is ergodic if and only if R is irreducible. The Bernoulli process (a) is a particular case when $R_{yx} = \mu(x)$.

(c) *Dynamical systems*

These systems are of special interest to physics (see [1] and Remark 2 below). Let $(M, \mathcal{M}, \mu)$ be a probability space so that M is a measurable space with a σ-algebra $\mathcal{M}$ of measurable subsets and μ a probability defined on $\mathcal{M}$.

Let $f : M \to M$ be a measurable bijection, which is measure-preserving, namely

$$\mu(f^{-1}(\mathcal{B}) = \mu(\mathcal{B}), \quad \mathcal{B} \in \mathcal{M} \tag{3.6}$$

Let X be a finite partition of M in measurable subsets. We define a coarse-grained, stochastic process on X by the formula

$$p_{[m,n]}\big(x[m,n]\big) = \mu\big(x_m \cap f^{-1}(x_{m+1}) \cap \cdots f^{-k}(x_{m+k}) \cap \cdots f^{m-n}(x_n)\big) \tag{3.7}$$

where $x[m,n] \in X^{[m,n]}$ is given by Eq. (3.5).

Then $p_{[m,n]}(x[m,n])$ is the measure of the subset of elements $z \in M$ with

$$z \in x_m, \quad f(z) \in x_{m+1}, \dots, f^{n-m}(z) \in x_n.$$

This process is stationary. It is ergodic if f is ergodic (i.e., if the only measurable subset $\mathcal{B}$ of M invariant by f is of measure $\mu(\mathcal{B}) = 0$ or 1).

Remark 1   This definition, due to Kolmogorov, was introduced to define nonspatial invariants of dynamical systems [4].

Remark 2   A particularly interesting example [4] is the case when M is a phase space, and f is a Hamiltonian map (i.e., the map given by the solution of the Hamilton equation at a given time) and μ is the volume on M, which is preserved by f because of the Liouville theorem.

4 Changing initial conditions: definition of the nonstationary process $\overline{p}$. Production of entropy

4.1 Definition of a particular nonstationary process $\overline{p}$

Let A be a partition of X. The elements of A are subsets $a \subset X$ satisfying Eqs. (2.5)–(2.10).

Let p be a stationary process on X, and q be a probability on A. These two data determine a process on X that is a probability $\overline{p}$ on X^N given by the formulas

$$\overline{p}_{[0,n]}(x\big([0,n]\big) = \frac{q(a(x_0))}{p_0(a(x_0))} p\big(x[0,n]\big) \quad (n \geq 0) \tag{4.1}$$

$$\overline{p}_{[m,n]}\big(x([0,n])\big) = \sum_{x[0,m-1]} \frac{q(a(x_0))}{p_0(a(x_0))} p\big(x[0,n]\big) \quad (m < 0 \le n) \tag{4.2}$$

Here

$$x[0,n] \in X^{[0,n]}$$

$$a(x_0) \in A \quad \text{is the unique } a \text{ such that } x_0 \in a$$

$$p_0(a) = \sum_{x \in a} p_0(x), \quad p_0 \text{ being } p_{[0]}.$$

Definitions (4.1)–(4.2) show that the distributions $\overline{p}_{[0,n]}$ and $p_{[m,n]}$ satisfy the compatibility conditions and define a probability on X^N, and a stochastic process (induced by the integers ≥ 0) on X. This stochastic process is nonstationary (indeed, being indexed by the ≥ 0 integers, the stationarity is meaningless).

The initial distribution is

$$\overline{p}_0(x_0) = \frac{q(a(x_0))}{p(a(x_0))} p_0(x_0) \tag{4.3}$$

and the distribution at time n is

$$\overline{p}_n(x_n) = \sum_{x[0,n-1]} \frac{q(a(x_0))}{p(a(x_0))} p\big(x[0,n-1],x_n\big) \tag{4.4}$$

where $(x[0,n-1],x_n)$ denotes the path $(x(0),0;x(1),1;\ldots;x(n-1),n-1;x_n,n)$.

Convention　As previously mentioned, we skip the indices I for $\overline{p}_I$ when there is no possible confusion.

Lemma 4.1　*The conditional probabilities of the process $\overline{p}$ with the condition starting at time* 0 *are identical to the corresponding conditional probabilities of the process p, so, for* $0 < k \le n$

$$\overline{p}\big(x[k,n]|x[0,k-1]\big) = p\big(x[k,n]|x[0,k-1]\big) \tag{4.5}$$

The proof is obvious using (4.1).

4.2　Entropy and relative entropy

If Z is a finite set, if $|Z|$ is the number of its elements, and if p, q are probabilities on Z, we define the entropy of p and the relative entropy of p and q by the usual formulas [7]

$$S(p) = -\sum_{z \in Z} p(z) \ln p(z) \ge 0 \tag{4.6}$$

$$S(p|q) = \sum_{z \in Z} p(z) \ln \frac{p(z)}{q(z)} \ge 0 \tag{4.7}$$

One has

$$0 \le S(p) \le \ln |Z|$$

$$S(p|q) \ge 0 \quad \text{and} \quad S(p|q) = 0 \text{ if and only if } p = q.$$

4.3 Path entropy

For the stationary process, the nonstationary process, and any positive integer n, we define the path entropy S_n

$$S_n(p) = S(p_{[0,n]}) = -\sum_{x[0,n]} p(x[0,n]) \ln p(x[0,n]) \tag{4.8}$$

$$S_n(\overline{p}) = S(\overline{p}_{[0,n]}) = -\sum_{x[0,n]} \overline{p}(x[0,n]) \ln \overline{p}(x[0,n]) \tag{4.9}$$

Lemma 4.2 (a) *One has the following identities*:

$$S_n(p) = S(p_0) + \sum_{k=1}^{n} d_k S_k(p) \tag{4.10}$$

where

$$d_k S_k(p) = S_k(p) - S_{k-1}(p) \tag{4.11}$$

and the same identities with $\overline{p}$ instead of p.

(b) *One has*

$$d_k S_k(p) = \sum_{x[0,k-1]} p(x[0,k-1]) S\big((.|x[0,k-1])\big) \geq 0 \tag{4.12}$$

$$d_k S_k(\overline{p}) = \sum_{x[0,k-1]} \overline{p}(x[0,k-1]) S\big((.|x[0,k-1])\big) \geq 0 \tag{4.13}$$

Proof (a) is trivial. On the other hand, one has

$$d_k S_k(p) = -\sum p(x[0,k]) \ln p(x[0,k] + \sum p(x[0,k-1]) \ln p(x[0,k-1]$$

$$= -\sum p(x[0,k]) \ln \frac{p(x[0,k]}{p(x[0,k-1]}$$

$$= -\sum_{x[0,k-1]} p(x[0,k-1]) \sum_{x(k)} \frac{p(x[0,k])}{p(x[0,k-1])} \ln \frac{p(x[0,k]}{p(x[0,k-1]}$$

which is (4.12). Similarly, we derive (4.13) using Lemma 4.1.　　　　□

Lemma 4.3 (a) *For the stationary process p, one has the identity*

$$d_k S_k(p_{[0,k]}) - d_{k-1} S_{k-1}(p_{[0,k-1]})$$

$$= -\sum_{x[-k,-1]} p(x[-k,-1]) S\big(p(.|x[-k,-1])|p(.|x[-k+1,-1])\big) \leq 0$$

(b) $d_k S_k(p(p_{[0,k]})$ *is a decreasing sequence with a limit $s(p)$.*

Proof Using the definition of $d_k S$ (Eq. (4.12)), one has by stationarity of p

$$d_k S_k(p(p_{[0,k]}) - d_{k-1} S_{k-1}(p(p_{[0,k-1]})$$

$$
\begin{aligned}
&= -\Big[\sum p\big(x[0,k]\big) \ln p\big(x(k)|x[0,k-1]\big) \\
&\quad - \sum p\big(x[0,k-1]\big) \ln p\big(x(k-1)|x[0,k-2]\big)\Big] \\
&= -\Big[\sum p\big(x[-k,0]\big) \ln p\big(x(0)|x[-k,-1]\big) \\
&\quad - \sum p\big(x[-k+1,0]\big) \ln p\big(x(0)|x[-k+1,-1]\big)\Big] \\
&= -\sum p\big(x\big([-k,0]\big)\big) \ln \frac{p(x(0)|x[-k,-1])}{p(x(0)|x[-k+1,-1])} \\
&\equiv -\sum p\big(x[-k,-1]\big) S\big(p(.|x[-k,-1]\big|p(.|x[-k+1,-1]\big)\big) \le 0 \qquad \square
\end{aligned}
$$

4.4 The case of a stationary probability p

In this case, we will use a theorem that was first presented in Ref. [1], using the concept of martingale (see, for instance, [13, 14], or [12] for a simplified definition).

Theorem 4.4 *For $x = x(0) \in X$, the sequence of random variables $p(x|x[-k,-1])$ is a martingale with respect to the sequence $\mathcal{F}_k$ of σ-algebras generated by $x([-k,-1])$. Moreover, these random variables are positively bounded by* 1.

Using this theorem, Lemma 4.3(b), and the identity (4.10), one obtains the result of Kolmogorov-Shannon [4, 6]:

Theorem 4.5 *For the stationary process p, one has*
(a) $d_k S_k(p)$ *has a limit $s(p)$ for $k \to \infty$*
(b) *One has*

$$
\lim_{n\to\infty} \frac{1}{n} S_n(p) = s(p) \tag{4.14}
$$

Definition $s(p)$ is the (asymptotic) production of entropy per unit time of the process p.

For completeness, the proofs of Theorems 4.4 and 4.5 are given in Appendix A.

Remark 3 In the special case where p comes from a dynamical system (Eq. (3.7)), it is proved in [3] that $d_k S_k(p)$ is a decreasing sequence, and there is no need to use martingale theory.

4.5 Non-stationary probability $\overline{p}$

In the nonstationary case, the production of entropy at time k is $d_k S_k(\overline{p})$. The asymptotic entropy production of $\overline{p}$ is well defined if $d_k S_k(\overline{p})$ tends to a limit when $k \to \infty$. As shown later, further hypotheses are necessary for such a limit to exist. In the general case, we can only prove that

$$
C' s(p) \le \liminf d_k S_k(\overline{p}) \le \limsup d_k S_k(\overline{p}) \le C'' s(p) \tag{4.15}
$$

where C' (resp. C'') is the lower (resp. upper) bound of $q(x)/p(x)$ ($0 \le C' \le 1 \le C''$).

Proof These inequalities straightforwardly result from Eqs. (4.1)–(4.2) and Theorem 4.4. $\qquad \square$

5 Production of entropy for a nonstationary distribution $\overline{p}$

5.1 Mixing process

We say that the stationary process p is a mixing process if for $0 < n < n + k$

$$\lim_{n \to \infty} p_{[0,n+k]}\big(x(0), x[n, n+k]\big) = p_0\big(x(0)\big) p_{[0,k]}\big(\tau^{-n}\big(x[n, n+k]\big)\big) \tag{5.1}$$

In expanded notations, this means that for $n \to \infty$

$$p\big(x(0), 0 : x_n, n; x_{n+1}, n + 1; \ldots; x_{n+k}, n + k\big)$$
$$\to p\big(x(0), 0\big) p(x_n, 0; x_{n+1}, 1; \ldots; x_{n+k}, k) \tag{5.2}$$

for any sequence $x(0), 0 : x_n, n; x_{n+1}, n+1; \ldots; x_{n+k}, n+k$ of $k + 1$ elements.

The mixing property implies ergodicity [4].

Theorem 5.1 *If p is mixing, the nonstationary process $\overline{p}$ defined in Sect. 4 has an asymptotic distribution in X, which is $p(x)$.*

Proof The asymptotic distribution of $\overline{p}$ (if it exists) is

$$\begin{aligned}
\overline{p}_\infty(x) &= \lim_{n \to \infty} \sum_{x[0,n-1]} \overline{p}\big(x[0,n]\big) \quad \text{with } x(n) = x \text{ fixed} \\
&= \lim_{n \to \infty} \sum_{x(0)} \overline{p}(x\big(x(0), 0; x(n), n\big) \\
&= \lim_{n \to \infty} \sum_{a \in A} \frac{q(a)}{p(a)} \sum_{x(0) \in a} p\big(x(0), 0; x(n), n\big)
\end{aligned} \tag{5.3}$$

But $p(x(0), 0; x(n), n) \to p(x(0))p(x)$ by the mixing property. As the sum over $a \in A$ is finite, the limit in Eq. (5.3) exists, and it is $p(x)$. □

5.2 Production of entropy for $\overline{p}$ when p is mixing. The main theorem

Theorem 5.2 *Assume that p is a mixing process. Then*

   (a) $d_n S_n(\overline{p})$ *has the limit* $s(p)$
   (b) $\lim_n \frac{1}{n} S_n(\overline{p}) = s(p)$

Then $\overline{p}$ has a well-defined production of entropy, which is the same as the entropy production of p.

The proof of this basic theorem is the consequence of successive partial results, which are postponed to Section 5.4 and completed in Appendix B.

5.3 Production of entropy for a mixing process p

Theorem 5.3 *If p is a mixing process, one has*

$$d_k S_k(p) = \lim_{n \to \infty} \sum p\big(x(0), x[n, n+k-1]\big) S\big(p\big(.|x(0), x[n, n+k-1]\big)\big) \tag{5.4}$$

the sum being taken over $x(0) \in X$ and over $x[n, n+k-1] \in X^{[n,n+k-1]}$.

Proof The mixing property (5.1) implies that the conditional probability $p_{[0,n+k]}(x(n+k)|x(0),x[n,n+k-1])$ has a limit:

$$
\lim_{n\to\infty} p_{[0,n+k]}\big(x(n+k)|x(0),x[n,n+k-1]\big)
$$
$$
= p_{[0,k]}(x(n+k)|\tau^{-n}\big(x[n,n+k-1]\big)1 \tag{5.5}
$$

where the limit is taken with a fixed k and fixed $x(0),x(n),\dots,x(n+k-1),x(n+k)$. Then

$$
-p\big(.|x(0),x[n,n+k-1]\big)\ln\big(p\big(.|x(0),x[n,n+k-1]\big)\big)
$$
$$
\to -p\big(.|\tau^{-n}\big(x[n,n+k-1]\big)\big)\ln p\big(.|\tau^{-n}\big(x[n,n+k-1]\big)\big) \tag{5.6}
$$

and all these quantities are uniformly bounded by $\max_{0\le\alpha\le 1}|\alpha\ln\alpha|$. As X is finite, we can sum (5.6) on $x(n+k)$ and obtain

$$
S\big(p\big(.|x(0),x[n,n+k-1]\big)\big) \to S\big(p\big(.|\tau^{-n}x[n,n+k-1]\big)\big) \tag{5.7}
$$

while staying uniformly bounded. By the Lebesgue theorem of dominated convergence in $L^1(Z,p)$, we have [15]

$$
\mathrm{E}_p\big\{S\big(p\big(.|x(0),x[n,n+k-1]\big)\big)\big\} \to \mathrm{E}_p\big\{S\big(p\big(.|\tau^{-n}x[n,n+k-1]\big)\big)\big\} \tag{5.8}
$$

where $\mathrm{E}_p\{\}$ is the mathematical expectation for the measure p. Now, the first term in Eq. (5.8) is

$$
\sum_{x(0),x[n,n+k-1]} p\big(x(0),x[n,n+k-1]\big)S\big(p\big(.|x(0),x[n,n+k-1]\big)\big)
$$

and the last member in Eq. (5.8) js, using (4.12)

$$
\sum_{x[0,k-1]} p(x[0,k-1]S\big(p\big(.|x[0,k-1]\big)\big) \equiv d_k S_k(p)
$$

which proves Eq. (5.4). $\square$

Theorem 5.4 *If p is a mixing process, for any probability q in X and for the associated process $\overline{p}$ defined in Sect. 4, one has*

$$
d_k S_k(p) = \lim_{n\to\infty} \sum \overline{p}\big(x(0),x[n,n+k-1]\big)S\big(\overline{p}\big(.|x(0),x[n,n+k-1]\big) \tag{5.9}
$$

where k is fixed and the sum is over $x(0)$, $x[n,n+k]$.

Proof Using the definition of $\overline{p}$ and Lemma 4.1, one has

$$
\sum \overline{p}\big(x(0),x[n,n+k-1]\big)S\big(\overline{p}\big(.|\{x(0),x[n,n+k-1]\big)\big)
$$
$$
\equiv \mathrm{E}_p\left\{ \frac{q(a(x(0)))}{p(a(x(0)))} S\big(p\big(.|x(0),x[n,n+k-1]\big)\right\}
$$
$$
= \sum_{a\in A} \frac{q(a)}{p(a)} \mathrm{E}_p\big\{\mathbf{1}_a\big(x(a)\big)S\big(p\big(.|x(0),x[n,n+k-1]\big)\big\} \tag{5.10}
$$

where $\mathbf{1}_a$ is the characteristic function of the subset a of X. By the mixing property Eq. (5.1) and Lebesgue theorem of dominated convergence, we have

$$
\begin{aligned}
&\lim_{n \to \infty} \mathrm{E}_p\big\{\mathbf{1}_a\big(x(a)\big)S(p(.|x(0), x[n, n+k-1])\big\} \\
&= p(a)\mathrm{E}_p\big\{S(p(.|\tau^{-n}x[n, n+k-1])\big\} \\
&\equiv p(a)\sum p(x([0, k-1]))S(p(.|x[0, k-1]) \equiv p(a)d_k S_k(p)
\end{aligned}
\tag{5.11}
$$

where we have used (4.12). So, by Eq. (5.10), we have proved Eq. (5.9). □

5.4 Proof of Theorem 5.2
We first derive several successive lemmas.

Lemma 5.5 *Let*

$q(x, y, z)$ be a probability distribution on three variables x, y, z taking discrete values,
$q(x)$ and $q(x, y)$ the corresponding marginal laws,
$q(z|x)$ and $q(z|x, y)$ the corresponding conditional laws of z.
Denote by S_Z the entropy of the probability distribution of z. Then, we have the identity

$$
\sum_x q(x)S_Z(q(.|x) - \sum_{x,y} q(x, y)S_Z\bigg(q(.|x, y) = \sum_{x,y} q(x, y)S_Z\big(q(.|x, y))|q(.|x)\bigg) \geq 0 \tag{5.12}
$$

Proof We apply the definitions to the first member to obtain identity (5.12). □

Lemma 5.6 *One has the identity*

$$
\begin{aligned}
&\sum \overline{p}(x0), x[n, n+k-1])S\big(\overline{p}(.|x0), x[n, n+k-1]\big) \\
&\quad - \sum \overline{p}\big(x[0, n+k-1]\big)S(\overline{p}(.|x[0, n+k-1]) \\
&= \sum \overline{p}\big(x[0, n+k-1]\big)S\big(\overline{p}(.|x[0, n+k-1])|\overline{p}(.|x(0), x[n; n+k-1])\big)
\end{aligned}
\tag{5.13}
$$

where each summation is over the variables appearing in the concerned probabilities. For instance, the first sum on the left is over $x(0), x(n), \ldots, x(n+k-1)$.

Proof of Lemma 5.6 We apply Lemma 5.5 to $q = \overline{p}$ with the substitutions

$$
x \to \big(x(0), x[n, n+k-1]\big), \qquad y \to x[1, n-1, \qquad z \to x[n+k]
$$

so that $(x, y) \to x[0, n+k-1]$. □

Lemma 5.7 *We have*

$$
\begin{aligned}
&\lim_{k \to \infty} \lim_{n \to \infty} \sum \overline{p}\big(x[0, n+k-1]\big) \\
&\quad \times S\big(\overline{p}(.|x[0, n+k-1])|\overline{p}(.|x(0), x[n, n+k-1]\big) = 0
\end{aligned}
\tag{5.14}
$$

The proof of this lemma implies that the first member of (5.13) tends to 0 when n and k tend to be infinite, which may seem intuitive from the definition of mixing. However, rigorous proof of Lemma 5.7 requires several further steps, as shown in Appendix B. It allows one to complete the proof of the basic Theorem 5.2.

End of the proof of Theorem 5.2   We start with the identity (5.13) of Lemma 5.6. The second term of the right member of this identity is just

$$- \sum \overline{p}(x[0, n + k - 1]) S(\overline{p}(.|x[0, n + k - 1])) \equiv d_{n+k} S_{n+k}(\overline{p}) \tag{5.15}$$

We see that the limit when $n \to \infty$ of the first term is, using Theorem 5.4, Eq. (5.9)

$$\lim_{n \to \infty} \sum \overline{p}(x(0), x[n, n + k - 1]) S(\overline{p}(.|x(x(0), x[n, n + k - 1])) \equiv d_k S_k(p) \tag{5.16}$$

Now, in the identity (5.13), taking the limits when $n \to \infty$ and then $k \to \infty$, the first member tends to 0 by Lemma 5.7. Taking the same limits in Eq. (5.16), its first term tends to $s(p)$. So, the second term of Eq. (5.13) $d_{n+k} S_{n+k}(\overline{p})$, has a limit, which is $s(p)$. Thus, we have proved that the nonstationary process $\overline{p}$ has a production of entropy $s(\overline{p}) = s(p)$.   □

6 Markov approximations
6.1 The process $p^{(T)}$ of memory T associated to p
In general, the process p has an infinite memory. Let T be a positive integer. We define a process $p^{(T)}$ on X of memory T associated to the process p by the formulas

$$p^{(T)}(x[0, k]) = p(x[0, k]) \quad \text{if } 0 \le k \le T - 1$$

$$p^{(T)}(x[0, k]) = p(x[0, T - 1]) \prod_{j=T}^{k} p(x(j)|x[j - T, j - 1]) \quad \text{if } k \ge T \tag{6.1}$$

Distance between p and $p^{(T)}$ An asymmetric "distance" between p and $p^{(T)}$ for n-step trajectories can be evaluated from the relative entropy of these two processes (see Sect. 6.4. below):

$$S(p_{[0,n]}|p_{[0,n]}^{(T)})$$

This quantity is related to the total variation distance between $p_{[0,n]}$ and $p_{[0,n]}^{(T)}$, as shown in Sect. 6.4.

Theorem 6.1 *For every $\varepsilon > 0$, there exists a time T_ε such that for $n \ge T \ge T_\varepsilon$, one has*

$$0 \le \frac{1}{n} S(p_{[0,n]}|p_{[0,n]}^{(T)}) \le \varepsilon \tag{6.2}$$

So, the distance between $p_{[0,n]}$ and $p_{[0,n]}^{(T)}$ tends to 0 when $n \to \infty$.

Proof   Using the definition of the relative entropy, Eq. (4.7), one has

$$S(p_{[0,n]}|p_{[0,n]}^{(T)}) = -S(p_{[0,n]}) - \sum_{x[0,n]} p(x[0, n]) \ln p^{(T)}(x[0, n]) \tag{6.3}$$

On the other hand, it follows from Eq. (4.9) that

$$-S(p_{[0,n]}) = -S(p_0) - \sum_{k=1}^{n} d_k S_k(p) \tag{6.4}$$

and by definition (6.1), if $n \geq T$

$$-\sum_{x[0,n]} p(x[0,n]) \ln p^{(T)}(x[0,n])$$

$$= S(p[0,T-1]) - \sum_{j=T}^{n} \sum_{x[j-T,j]} p(x[j-T,j]) \ln p(x(j)|x[j-T,j-1]) \tag{6.5}$$

By the stationarity of p, this is

$$S(p_{[0,T-1]}) - \sum_{j=T}^{n} \sum_{x[0,T]} p(x[0,T]) \ln p(x(T)|x[0,T-1])$$

$$= S(p_{[0,T-1]}) + (n - T + 1) d_T S_T(p) \tag{6.6}$$

From Eqs. (6.4), (6.6), and (6.3), we obtain

$$S(p_{[0,n]}|p_{[0,n]}^{(T)}) = \sum_{k=T}^{k=n} (d_T S_T(p) - d_k S_k(p)) \tag{6.7}$$

According to Theorem 4.4, $d_k S_k(p)$ decreases when k increases, so each term of the sum in Eq. (6.7) is ≥ 0, and $d_k S_k(p) \to s(p)$. Choose T_ε so that or $k \geq T \geq T_\varepsilon$

$$s(p) \leq d_k S_k(p) \leq s(p) + \varepsilon \tag{6.8}$$

Then, for $n \geq T \geq T_\epsilon$

$$S(p_{[0,n]}|p_{[0,n]}^{T}) \leq (n - T_\varepsilon)\varepsilon \tag{6.9}$$

which completes the proof of Theorem 6.1.   □

6.2 Partial histories of length T

Definitions A partial history of length T is an element of X^T. The nth history of length T is

$$x^{(T)}(n) = (x(nT)), x(nT+1), \ldots, x((n+1)T - 1)) \in X^T \tag{6.10}$$

If M and N are positive integers ($M < N$), a sequence of partial histories is

$$x^{(T)}(M,N] = (x^{(T)}(M), \ldots, x^{(T)}(N)) \tag{6.11}$$

We also denote by $\tau^{(T)}$ the translation of time T on histories of length T.

Theorem 6.2 (a) *The process $p^{(T)}$ induces a Markov process $\tilde{p}^{(T)}$ on partial histories of length T by the formulas*

$$\tilde{p}^{(T)}\big(x^{(T)}(0)\big) = p\big(x[0, T-1]\big)$$
$$\tilde{p}^{(T)}\big(x^{(T)}[0, N]\big)) = p^{(T)}\big(x[0, (N+1)T-1]\big) \tag{6.12}$$

The transition probabilities between histories of length T are

$$R\big(x^{(T)}(1)|x^{(T)}(0)\big) = \prod_{j=T}^{2T-1} p\big(x(j)|x[j-T, j-1]\big) \tag{6.13}$$

where, according to Eq. (6.10)

$$x^{(T)}(0) = \big(x(0),\ldots,x(T-1)\big), \qquad x^{(T)}(1) = \big(x(T),\ldots,x(2T-1)\big) \tag{6.14}$$

(b) *The stationary probability of the Markov process $\tilde{p}^{(T)}$ is $\tilde{p}^{(T)}(x^{(T)}(0))$.*
(c) *The production of entropy of $\tilde{p}^{(T)}$ is*

$$s\big(\tilde{p}^{(T)}\big) = \sum_{x^{(T)}(0)} \tilde{p}^{(T)}\big(x^{(T)}(0)\big)S(R(.|x^{(T)}(0)) \tag{6.15}$$

where $S(R(.|x^{(T)}(0))$ is the entropy of the probability distribution $x^{(T)}(1) \to R(x^{(T)}(1)| x^{(T)}(0))$.

Proof Using the definitions of Eqs. (6.13) and (6.1), we have

$$\tilde{p}^{(T)}\big(x^{(T)}[0, N]\big) = \tilde{p}^{(T)}\big(x^{(T)}(0)\big) \prod_{j=T}^{(N+1)T-1} p\big(x(j)|x[j-T, j-1]\big)$$
$$\equiv \tilde{p}^{(T)}\big(x^{(T)}(0)\big) \prod_{k=1}^{N} R\big(x^{(T)}(k)|x^{(T)}(k-1)]\big) \tag{6.16}$$

We now show that $\tilde{p}^{(T)}(x^{(T)}(0))$ is the stationary probability. It will prove that (6.16) is the usual formula [16] for a Markov process (Eq. (3.4) in Sect. 3). One has

$$\sum_{x^{(T)}(0)} R\big(x^{(T)}(1)|x^{(T)}(0)]\tilde{p}^{(T)}\big(x^{(T)}(0)\big)\big)$$
$$= \sum_{x[0,T-1]} p\big(x[0, T-1]\big)p\big(x(T)|x[0, T-1]\big)p\big(x(T+1)|x[1, T]\big) \cdots$$
$$\cdots p\big(x(2T-1)|x[T-1, 2T-2]\big)$$
$$= \sum_{x[0,T-1]} p\big(x[0, T]\big)\frac{p(x[1, T+1])}{p(x[1, T])}p\big(x(T+2)|x[2, T+]\big) \cdots \tag{6.17}$$
$$= \sum_{x[1,T-1]} p\big(x[1, T+1]\big)\frac{p(x[2, T+2])}{p(x[1, T+1])} \cdots = \cdots$$
$$= p\big(x[T, 2T-1]\big)$$

$$= p\big(\tau^{-T}x[T, 2T-1]\big) = \tilde{p}^{(T)}\big(\tau^{-1}x^{(T)}(1)\big)$$

So, $\tilde{p}^{(T)}(x^{(T)}(0))$ is indeed the stationary probability of the Markov process $\tilde{p}^{(T)}$.

On the other hand, Eq. (6.15) is just the usual formula for the entropy production of a Markov process [16]. □

6.3 Comparison of p and $\tilde{p}^{(T)}$

The process p induces a stationary process on the partial histories of length T (denoted p_T) by

$$p_T\big(x^{(T)}[M, N]\big) = p\big(x[MT, (N+1)T-1]\big) \tag{6.18}$$

The process p_T is exactly the same as p except that it is restricted to an integer number of time T.

The entropy production of p_T is

$$s_T(p_T) = Ts(p) \tag{6.19}$$

So, one can rewrite Theorem 6.1 as follows.

Theorem 6.3 *Denote by $S_T(.\,|.)$ the relative entropy of two processes defined on histories of length T. Then, for any $\epsilon > 0$, there exists T_ϵ, independent of N, such that for $T \geq T_\epsilon$ one has*

$$\frac{1}{NT}S_T\big(p_{T,[0,n]}||\tilde{p}^{(T)}_{[0,N]}\big) \leq \varepsilon \tag{6.20}$$

6.4 Distance between p_T and $\tilde{p}^{(T)}$

We can interpret this relation as follows. If p, q are probabilities on a finite space Z, the following Pinsker inequality [17] relates the relative entropy of p and q to the total variation distance of distribution p and q:

$$\frac{1}{2}\left(\sum_{z\in Z}|p(z) - q(z)|\right)^2 \leq S(p|q)$$

This shows that $S(p|q)$ represents an asymmetrical distance between p and q. Equation (6.20) implies that the absolute distance between the actual process p_T and the Markov process $\tilde{p}^{(T)}$, divided by T, goes to 0 for long times T.

Theorem 6.4 *One has for the production of entropy*

$$\lim_{T\to\infty}\frac{1}{T}s_T\big(\tilde{p}^{(T)}\big) = s(p) \tag{6.21}$$

Proof We use the expression of the entropy production of a Markov process (Eq. (6.15))

$$s\big(\tilde{p}^{(T)}\big) = -\sum_{x[0,T-1]} p(x[0, T-1]) \prod_{j=T}^{2T-1} p\big(x(j)|x[j-T, j-1]\big)$$

$$\times \sum_{k=T}^{2T-1} \ln p\big(x(k)|x[k-T,k-1]\big)$$

$$= -\sum_{k=T}^{2T-1}\sum p\big(x[0,T-1]\big) \prod_{j=T}^{k} p\big(x(j)|x[j-T,j-1]\big) \ln p\big(x(k)|x[k-T,k-1]\big)$$

$$= -\sum_{k=T}^{2T-1}\sum p\big(x[1,T]\big) \prod_{j=T+1}^{k} p\big(x(j)|x[j-T,j-1]\big) \ln p\big(x(k)|x[k-T,k-1]\big)$$

$$= \cdots$$

$$= -\sum_{k=T}^{2T-1}\sum_{x[k-T,k]} p\big(x[k-T,k-1]\big) \ln p\big(x(k)|x[k-T,k-1]\big)$$

and by the stationarity of p and Eq. (4.12)

$$= -T \sum_{x[0,T]} p\big(x[0,T]\big) \ln p\big(x(T)|x[0,T-1]\big) = T d_T S_T(p)$$

However, by Theorem 4.5, $d_T S_T(p) \to s(p)$ if $T \to \infty$, which gives Theorem 6.4.   □

6.5 Attenuation of the memory

We come back to the process p_T on histories of length T. We now prove

Theorem 6.5 *For a fixed integer $N \geq 1$, one has*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{x^{(T)}[0,N-1]} p\big(x^{(T)}[0,N-1]\big) \tag{6.22}$$

$$\times S_T\big(p_T\big(.|x^{(T)}[0,N-1]\big)|p_T\big(.|x^{(T)}(N-1)\big)\big) = 0$$

Proof We use the definition of relative entropy and decompose the sum of Eq. (6.22) into two terms:

$$A_1 \equiv -d_N S_{T,N}(p_T)$$

$$= - \sum_{x^{(T)}[0,N-1]} p\big(x^{(T)}[0,N-1]\big) \times S_T\big(p\big(.|x^{(T)}[0,N-1]\big)\big) \tag{6.23}$$

$$= -(S(p_{[0,NT-1]}) - S(p_{[0,(N-1)T-1]}) = -\sum_{k=(N-1)T}^{NT-1} d_k S_k(p)$$

and

$$A_2 \equiv - \sum_{x^{(T)}[0,N-1]} p\big(x^{(T)}[0,N-1]\big)$$

$$\times \sum_{x^{(T)}(N)} p_T\big(x^{(T)}(N)|x^{(T)}[0,N-1]\big) \ln p_T\big(x^{(T)}(N)|x^{(T)}(N-1)\big)$$

$$= - \sum_{x^{(T)}[0,1]} p\big(x^{(T)}[0,1]\big) \ln p_T\big(x^{(T)}(1)|x^{(T)}(0)\big) \tag{6.24}$$

$$= S(p_{[0,2T-1]}) - S(p_{[0,T-1]}) = \sum_{k=T}^{2T-1} d_k S_k(p)$$

Both sums (6.23) and (6.24) contain T terms $d_k S_k(p)$ which tend to $s(p)$. This gives the result (6.22). □

As a consequence, if T is large enough, within a given accuracy, ε is possible to neglect the distance between the process at time NT, with complete history from time 0, and the process with history limited to the last period of length T, between times NT and $(N-1)T$. In practice, one can neglect the memory after times larger than T.

7 Conclusion

It has been rigorously proved that coarse-graining dynamical systems induce new systems that partially approximate the original systems. This conclusion is often anticipated intuitively in modeling physical or applied phenomena, which most generally needs simplifying and approximating actual observations. Because of its importance, this question has been, for a long time, the matter of many studies (see, for instance [18], and references therein), but it is difficult to obtain both a general approach and exact results on dynamical problems. Recently, it has been shown that innovative concepts introduced by Kolmogorov somewhat sixty years ago can be combined with the martingale theory to yield novel results in this domain. At first, this point of view was applied to classical Hamiltonian systems [1], and a major result was that under appropriate, realistic conditions, coarse-graining systems generate an approximate Markov system. Here, we have seen that the same reasoning applies to much more general, possibly stochastic, processes. Using a purely mathematical formalism, we obtained new, more general conclusions.

In particular, we have proved that the Kolmogorov entropy, introduced by Kolmogorov for ergodic stationary processes [4], also exists for a class of nonstationary processes defined for coarse-grained systems: these processes are obtained by imposing a nonstationary coarse-grained initial probability distribution, whereas the initial conditional distribution remains stationary in each grain. Such nonstationary coarse-grained distributions can be adopted in realistic mesoscopic systems if they are initially constrained to nonequilibrium. In contrast, local equilibrium is almost instantaneously re-established: these approximations are often valid in realistic examples [10, 19], which justifies studying this special class of nonstationary processes. Moreover, it has been proved that the asymptotic entropy production of these nonstationary processes is identical to the entropy production of the microscopic stationary process, provided this one is mixing. This is our main result, which allows one to approximate a large class of coarse-grained dynamical processes by Markov processes.

Alternatively, within the framework of the previous general theory, a forthcoming article [20] will present further exact results concerning the comparison of different coarse-grainings of dynamical systems that are of interest for modeling Markov and non-Markov processes.

Appendix A: Proofs of Theorems 4.4 and 4.5

Theorem 4.4 *For $x = x(0) \in X$, the sequence of random variables $p(x|x[-k, -1])$ is a martingale with respect to the sequence $\mathcal{F}_k$ of σ-algebras generated by $x([-k, -1])$. Moreover, these random variables are positively bounded by* 1.

Proof Consider the random variables

$$\pi_k = p\big(x|x[-k,-1]\big) = p(x|\mathcal{F}_k).$$

We have because $\mathcal{F}_{k-1} \subset \mathcal{F}_k$

$$E(\pi_k|\mathcal{F}_{k-1}) = E\big(p(x|\mathcal{F}_k)|\mathcal{F}_{k-1}\big) = p(x|\mathcal{F}_{k-1}) = \pi_{k-1},$$

where E{} is the mathematical expectation for the measure p. So, π_k is a martingale on the σ-algebra $\mathcal{F}_N$, and by the convergence theorem of martingales [13], it converges almost surely to a limit π when $k \to \infty$. □

From the previous theorem, one can deduce the theorem of Kolmogorov-Shannon [4–6].

Theorem 4.5 *For the stationary process p, one has*
(a) *$d_k S_k(p)$ has a limit $s(p)$ for $k \to \infty$*
(b) *One has*

$$\lim_{n\to\infty} \frac{1}{n} S_n(p) = s(p) \tag{A.1}$$

Proof (a) We use Eq. (2.13) and the stationarity of p so that

$$d_k S_k(p) = \sum_{x[-k,-1]} p\big(x[-k,-1]\big) S\big(p\big((.|x[-k,-1]\big)\big) \geq 0 \tag{A.2}$$

where

$$S\big(p(.|x[-k,-1]\big) = -\sum_{x(0)} p\big(x(0)|x[-k,-1]\big) \ln p\big(x(0)|x[-k,-1]\big) \tag{A.3}$$

By Theorem 4.4, the sequence of random variables $p(x(0)|x[-k,-1])$ is a martingale with respect to the sequence $\mathcal{F}_k$ of σ-algebras generated by $x([-k,-1])$. Moreover, these random variables are positively bounded by 1. So, by the theorem of convergences of martingales [4], this sequence converges p-almost surely to a certain random variable $p(x(0)|x[-\infty,-1])$ (as well as any Lebesgue space $L^r(X^r,p)$ for $1 \leq r \leq +\infty$). Furthermore, $p(x(0)|x[-k,-1]) \ln p(x(0)|x[-k,-1])$ converges also p-almost surely, as well as the finite sum $S(p(.|x[-k,-1))$ over $x(0) \in X$, while staying uniformly bounded. By Lebesgue of dominated convergence [6], the integral over p of these random variables converges, so $d_k S_k(p)$ converges.

(b) $\frac{1}{n} S_n(p)$ is, up to $\frac{1}{n} S_n(p_0)$, the arithmetic sum of the first n differences $d_k S_k(p)$ (Eq. (4.10)). □

Appendix B: Proofs of Lemma 5.7 and Theorem 5.2

Using the definition of $\overline{p}$ and Lemma 4.1 for the conditional entropy, we have

$$\sum \overline{p}\big(x[0, N + k - 1]\big) S\big(\overline{p}\big(.|x[0, n + k - 1]\big)|\overline{p}\big(.|x(0), x[n, n + k - 1]\big)\big)$$

$$= \sum \frac{q(a(x_0))}{p(a(x_0))} p\big(x[0, n + k - 1]\big) S\big(p\big(.|x[0, n + k - 1]\big)|p\big(.|x(0), x[n, n + k - 1]\big)\big)$$

$$\leq C \sum p\big(x[0, n + k - 1]\big) S\big(p\big(.|x[0, n + k - 1]\big)|p\big(.|x(0), x[n, n + k - 1]\big)\big)$$

where $C = \max q(a)/p(a)$.

To prove Lemma 5.7, we prove

Lemma 5.8 *One has*

$$\lim_{k \to \infty} \lim_{n \to \infty} \sum p\big(x[0, n + k - 1]\big)$$

$$\times S\big(p\big(.|x[0, n + k - 1]\big)|p\big(.|x(0), x[n, n + k - 1]\big)\big) = 0 \tag{B.1}$$

Proof of Lemma 5.8 We split the sum in the first term of (5.16) into two terms

$$\sum p\big(x[0, n + k - 1]\big) S\big(p\big(.|x[0, n + k - 1]\big)|p\big(.|x(0), x[n, n + k - 1]\big)\big)$$

$$= E_1(n + k) + E_2(n, k) \tag{B.2}$$

with

$$-E_1(n + k) = E_p\big\{S\big(p\big(.|x[0, n + k - 1]\big)\big)\big\} \tag{B.3}$$

and

$$E_2(n, k) = -E_p\Big\{\sum p\big(x(n + k)|x[0, n + k - 1]\big) \ln p\big(x(n + k)|x(0), x[n, n + k - 1]\big)\Big\} \tag{B.4}$$

Lemma 5.8 is proved from the next two Lemmas. □

Lemma 5.9 *We have*

$$\lim_{n+k \to \infty} E_1(n + k) = -s(p) \tag{B.5}$$

Proof of Lemma 5.9 By the stationarity of p, we have

$$E_1(n + k) = \sum_{x(0)} E_p\big\{-p\big(x(0)|x[-n - k, -1]\big) \ln p\big(x(0)|x[-n - k, -1]\big)\big\}$$

The martingale $p(x(0)|x(-n - k, -1)$ is uniformly bounded and converges p-almost surely, and it also integrable, so that, by the convergence theorem of martingales [13]

$$\lim_{n+k \to \infty} E_1(n + k) = \sum_{x(0)} E_p\big\{-p\big(x(0)|x[-\infty, -1]\big) \ln p\big(x(0)|x - \infty, -1]\big)\big\} = -s(p). \quad □$$

Lemma 5.10 *One has*

$$
\lim_{k \to \infty} \lim_{n \to \infty} E_2(n,k)
$$
$$
= -\sum_{x(0)} \mathrm{E}_p\big\{p\big(x(0)|x]-\infty,-1]\big) \ln p\big(x(0)|x]-\infty,-1]\big)\big\}
\tag{B.6}
$$

Proof of Lemma 5.10   We have

$$
E_2(n,k) = -\sum p\big(x[0,n+k-1]\big)p\big(x(n+k)|x[0,n+k-1]\big)
$$
$$
\times \ln p\big(x(n+k)|x(0),x[n,n+k-1]\big)
$$
$$
= -\sum p\big(x[0,n+k]\big)\ln p\big(x(n+k)|x(0),x[n,n+k-1]\big)
$$
$$
= -\sum_{x(n+k)} \mathrm{E}_p\big\{p\big((x(n+k)\big)\big\}\ln p\big(x(n+k)|x(0),x[n,n+k-1]\big)
$$
$$
= \mathrm{E}_p\big\{S\big(p\big(.|x(0),x[n,n+k-1]\big)\big)\big\}
\tag{B.7}
$$

Now, we have seen in Eq. (5.8) that this converges when $n \to \infty$ to

$$
\lim_{n \to \infty} E_2(n,k) = \mathrm{E}_p\big\{S\big(p\big(.|\tau^{-n}x[n,n+k-1]\big)\big)\big\}
$$
$$
= \mathrm{E}_p\big\{S\big(p\big(.|x[\grave{a},k-1]\big)\big)\big\} = d_k S_k(p) \quad \big(\text{by Eq. (4.12)}\big)
\tag{B.8}
$$

So,

$$
\lim_{k \to \infty} \lim_{n \to \infty} E_2(n,k) = s(p)
\tag{B.9}
$$

Eqs. (B.8) and (B.9) prove Lemma 5.10.   □

This concludes the proof of Theorem 5.2.

Declarations

References
1. Gaveau, B., Moreau, M.: Chaos **30**, 083104 (2020)
2. Kolmogorov, A.N.: Dokl. Akad. Nauk SSSR **119**, 861 (1958)
3. Kolmogorov, A.N.: Dokl. Akad. Nauk SSSR **124**, 754 (1959)
4. Arnold, V.I., Avez, A.: Ergodic problems of Classical Mechanics. Mathematical Physics, Monographs, Benjamin (1968)
5. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1949)

6. Khinchin, A.I.: Mathematical Foundation of Information Theory. Dover, New York (1957)
7. Gaveau, B., Schulman, L.S.: J. Math. Phys. **37**, 3897 (1996)
8. Huang, K.: Statistical Mechanics. CRC Press, Boca Raton (1987)
9. Gaveau, B., Schulman, L.S.: Eur. Phys. J. **224**, 891 (2015)
10. Landau, L., Lifschitz, E.: Statistical Physics, 3rd edn. Pergamon, Elmsford (1969)
11. Bass, R.F.: Stochastic Processes. Cambridge Universiry Press, Cambridge (2011)
12. Moreau, M., Gaveau, B.: Stochastic theory of coarse-grained deterministic systems: martingales and Markov approximations. In: Carpentieri, B. (ed.) Advances in Dynamical Systems Theory, Models, Algorithms (2021). https://doi.org/10.5772/Intechopen.95903
13. Doob, J.: Stochastic Processes. Wiley, New York (1953)
14. Levy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1937)
15. Bartle, R.G.: The Elements of Integration and Lebesgue Measure. Wiley-Interscience, New York (1995)
16. Dynkin, E.B.: Theory of Markov Processes. Pergamon, Elmsford (2015)
17. Cover, T., Thomas, J.: Elements of Information Theory. Wiley, New York (1991)
18. Achieser, N.I.: Theory of Approximation. Dover Books on Mathematics, Dover (2013)
19. Reif, F.: Fundamental of Statistical and Thermal Physics. McGraw-Hill, New York (1965)
20. Gaveau, B., Moreau, M.: Generalized kinetic theory of coarse-grained systems, entropy and Markov approximations. II: comparison between various coarse grainings. To be published

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.