Mathematics

AIMS Mathematics, 8(11): 26328–26342.
DOI: 10.3934/math.20231343
Received: 30 March 2023
Revised: 21 August 2023
Accepted: 28 August 2023
Published: 14 September 2023

http://www.aimspress.com/journal/Math

Research article

New remarks on the Kolmogorov entropy of certain coarse-grained deterministic systems

Michel Moreau[1,*] and Bernard Gaveau[2]

[1] Laboratory of Theoretical Physics of Condensed Matter, Faculty of Sciences, Sorbonne Université, Paris, France
[2] Former Professor at the Faculty of Mathematics, Sorbonne Université, Paris, France

* **Correspondence:** Email: michel.moreau@sorbonne-universite.fr; Tel: +33144277237.

Abstract: Unless an appropriate dissipation mechanism is introduced in its evolution, a deterministic system generally does not tend to equilibrium. However, coarse-graining such a system implies a mesoscopic representation which is no longer deterministic. The mesoscopic system should be addressed by stochastic methods, but they lead to practically infeasible calculations. However, following the pioneering work of Kolmogorov, one finds that such mesoscopic systems can be approximated by Markov processes in relevant conditions, mainly, if the microscopic system is ergodic. So, the mesoscopic system tends to stationarity in specific situations, as expected from thermodynamics. Kolmogorov proved that in the stationary case, the instantaneous entropy of the mesoscopic process, conditioned by its past trajectory, tends to a finite limit at infinite times. Thus, one can define the Kolmogorov entropy. It can be shown that in certain situations, this property remains true even in the nonstationary case. We anticipated this important conclusion in a previous article, giving some elements of a justification, whereas it is precisely derived below in relevant conditions and in the case of a discrete system. It demonstrates that the Kolmogorov entropy is linked to basic aspects of time, such as its irreversibility. This extends the well-known conclusions of Boltzmann and of more recent researchers and gives a general insight to the fascinating relation between time and entropy.

Keywords: stochastic processes; coarse-graining; Kolmogorov entropy; Markov processes; martingales
Mathematics Subject Classification: 60G99

1. Introduction

Following the work of Kolmogorov (see [1] and references herein), the stochastic theory of coarse-grained deterministic, ergodic systems has been studied in [2]. In particular, it was shown that such mesoscopic systems can be approximated by Markov processes, which may explain why Markov models are so widely used in the literature. Here we present new remarks on such systems, in particular concerning their Kolmogorov entropy, which has been introduced for stationary processes in the work of this author [1]. In particular, we will show that the Kolmogorov entropy can be defined for a nonstationary process obeying simple properties which should be satisfied for generic, realistic systems. Here, we will focus on finite systems, which significantly simplifies the reasoning.

Our definitions and notations are identical with those of [2]. Nevertheless, for the sake of clarity, we summarize them below in Section 2, as well as the known results. New outcomes are presented in Section 3, with simplified demonstrations. Conclusion and discussion are given in Section 4. Detailed derivations are postponed to Appendix A. A more complete and rigorous theory will be presented elsewhere (see Section 4).

2. Material and methods

2.1. Microscopic and mesoscopic descriptions of a deterministic system

It is known [1] that a coarse-grained deterministic system $\mathcal{S}$ can be represented by a non-Markovian stochastic process. One has to define this stochastic process on the space $\mathcal{M}$ of the observable mesoscopic states and during all the period of observation we will assume that this period begins at time $t = 0$, without assigning it a finite end.

2.1.1. Microscopic, deterministic dynamical system

For the sake of simplicity, we consider a finite microscopic dynamical system: The space X of microscopic states is finite and contains $\mathcal{N}$ microscopic states x, each of them corresponding to the ultimate possible description of the system, according to the usual conventions of statistical mechanics. Furthermore, time will be discretized: $t = 0, 1, \ldots k, \ldots$, the elementary time step τ being taken as time unit.

A probability measure $\mu(t)$ is defined on the finite microscopic space X at time $t \geq 0$, including $\mathcal{N}$ microscopic states. The probability of a set A of microstates states at time t is $\mu(A,t)$. The system obeys a deterministic stationary process which transfers an initial microscopic state x into the microscopic state $\varphi_t(x)$ after time t, where the evolution function φ_t satisfies the standard property of such functions (see for instance the book by Arnold and Avez [1] and references therein). So, we assume that φ_t is measure-preserving, i.e., for any measurable subspace A of X, $\mu(A,t) = \mu(\varphi_{-t}A,0)$. From now on, we also assume that μ is stationary: $\mu(A,t) = \mu(A,0)$.

We adopt the current hypothesis that the microscopic dynamical system considered here is ergodic [1]. There is no measure-invariant subspace Y of the microscopic space X, except X itself and the empty space $\varnothing$. It is well-known [1,3] that if the microscopic system is ergodic, the stationary measure is unique.

In the absence of any microscopic information before time 0, it can be assumed that the initial microscopic probability distribution μ is uniform in the whole space X: $\mu(x) \equiv \mu(x, 0) = 1/\mathcal{N}$, the uniform law being obviously stationary. So, in this article we suppose that the stationary law μ is uniform, although the following reasoning can in many cases be extended to more general stationary measures.

2.1.2. Mesoscopic, coarse-grained system

Because the possible observations are limited and because the measure accuracies are finite, these microscopic states are not directly observable. On the other hand, the limited accuracy of actual, available experiments allows one to define M observable mesostates i, constituting the mesospace $\mathcal{M}$, in such a way that each microscopic state x belongs to one and only one mesostate i. On the other hand, a mesostate i corresponds to n_i different microstates, with $n_i \geq 1$. Clearly, these are the usual conventions of classical statistical mechanics discussed in all textbooks.

The initial mesoscopic stationary distribution of a mesostate i is proportional to the number n_i of microstates included in the mesostate i:

$$p^0(i,0) = \mu(i) = n_i / M$$

where the upper index 0 denotes the stationary case, and the probability of i at time k is

$$p^0(i,k) = \mu(\varphi_{-k} i) = p^0(i,0) \tag{1}$$

for all $k > 0$.

2.1.3. Evolution of the coarse-grained system: stationary case

The stochastic process representing the coarse-grained states i_0, i_1, …i_k, … at the respective times 0, 1, …k,…. is defined by the probability $p_k^0(i_0,0;i_1,1;\dots;i_{k-1},k-1)$ of any k-times trajectory, for all $k > 0$. The complete stationary probability law, for all k, is denoted p^0.

It is easily seen [2] that the probability of a (k+1)-times trajectory from time 0 is

$$p_{k+1}^0(i_0,0;\ i_1,1;\dots;i_k,k) = \mu(\varphi_{-k}\, i_{t_k} \cap \dots \varphi_{-1} i_1 \cap i_0) \ . \tag{2}$$

To simplify the notations, we will now omit the lower index k+1 in the probability p_{k+1}^0 when it is possible without confusion, for instance when the variables are explicitly mentioned.

With this convention, conditional probabilities can be defined and written in the usual elementary way. For instance, if $p^0(i_0,0;i_1,1;\dots;i_{k-1},k-1) > 0$

$$p^0(i_k,k|i_{k-1},k-1;\dots;i_0,0) = \frac{p^0(i_0,0;i_1,1;\dots;i_k,k)}{p^0(i_0,0;i_1,1;\dots;i_{k-1},k-1)} \tag{3}$$

If $p^0(i_0,0;i_1,1;\dots;i_{k-1},k-1) = p^0(i_0,0;i_1,1;\dots;i_k,k) = 0$, it is well known that the conditional probability is not defined by (3), but this indetermination has no influence on the following calculations.

2.1.4. Coarse-grained system: Nonstationary case

The system can be prepared in order that the initial probability of any mesostate i obeys some arbitrary distribution $p(i)$. Then, the microscopic dynamics of the system and the initial probability law

$p(i)$ determine the law of the stochastic coarse-grained process over the space of mesostates $\mathcal{M} \equiv (i^m)$, m= 1, …M. Since no available observation can distinguish two microscopic states inside the same mesostate i, it can be logically assumed that the microscopic initial distribution is piecewise uniform, being uniform in each mesostate i. Then, if the initial probability of i is $p(i) \neq \mu(i)$, it can be shown [2] that the mesoscopic k-times law is

$$p_k(i_0, 0; i_1, 1; \ldots; i_{k-1}, k-1) = \frac{p(i_0)}{\mu(i_0)} \mu(\varphi_{-k+1} i_{k-1} \cap \ldots \varphi_{-1} i_1 \cap i_0) \tag{4}$$

where μ is again the stationary measure on X. This formula allows one to obtain all probabilities concerning finite mesoscopic trajectories, such as the probability that the system is in some mesostate i_k at time k, as well as all relevant conditional probabilities.

If, for instance, the microscopic system is prepared to be initially localized inside some initial mesostate i_0, it will no stay concentrated in i_0 at the next step (this is forbidden by ergodicity) but it will generally be distributed between several mesostates states i'_1, i''_1; …. Only in very specific cases, all the microscopic cases included in i_0 at time 0, are transferred to the same mesostate i_1 at time 1 and to a mesostate state i_2 at time 2, etc. Since φ is measure- preserving, $\mu(i_0) = \mu(i_1)$ and the microscopic states are uniformly distributed inside i_1 at time 1, then transferred to i_2 at time 2, etc. In this special case, the mesoscopic trajectory i_0, i_1, i_2, … is periodic, as well as the microscopic trajectories, because the system is supposed to be finite. We will discard such an exceptional situation, which presents no interest and is not realized in current phenomena. On the contrary, we will assume that the coarse-graining is such that, after a relatively small number of steps, the microstates initially concentrated in some i_0, are essentially distributed between all the mesostates of $\mathcal{M}$. This is a usual hypothesis, adopted in most textbooks of statistical thermodynamics. It provides an intuitive justification of the memory erasing precisely defined in Section 3.4, Eq (12) and derived previously [2].

In these conditions, following Kolmogorov [1] and using intuitive extensions of his methods, we will present some remarks on the non-stationary case in Section 3, mainly concerning the entropy of these processes.

2.2. Entropy of a process. Kolmogorov entropy of the mesoscopic stationary process

2.2.1. The n-times entropy of the system

We call n-times entropy $S_n(p)$ of the process the Shannon entropy [2,5–8] $S(p_n)$ of the n-times trajectory $(i)_n = (i_0, \ldots, i_{n-1})$ in the phase space

$$S_n(p) = -\sum_{i_0,\ldots i_{n-1}} p_n(i_0, 0; \ldots; i_{n-1}, n-1) \ln p_n(i_0, 0; \ldots; i_{n-1}, n-1) \equiv S(p_n). \tag{5}$$

So, among other interpretations [9], this quantity measures the uncertainty, or disorder contained in the n-times probability p_n. Equivalently, according to Shannon [5], this is the information recovered after an experiment where an actual trajectory is observed, whereas before the experiment one only knew the probability of this trajectory. Clearly, this entropy vanishes if the trajectory is deterministic.

It is well known that suppressing correlations between the different states increases the disorder in the stochastic system and increases its entropy. So, the maximum n-times entropy S_n occurs when all the states are statistically independent. In this very special case, the entropy of the n-times process is

$$\overline{S}_n = \sum_{k=0}^{n-1} \left[-\sum_{i_k=1}^{N} p_1(i_k,k) \ln p_1(i_k,k) \right] \equiv \sum_{k=0}^{n-1} s_k(p_1(t=k)) \tag{6}$$

where $s_k(p_1(t=k))$ is the one-time entropy at time *k* (see below).

In case the process is stationary, the one-time probability p_1 is time invariant and the one-time entropy s_1 as well, so that the maximum *n*-times entropy is $\overline{S}_n = n\, s_1$.

2.2.2. The instantaneous entropy at time *n*

The new information obtained by observing the system in the mesoscopic state i_n at time *n*, knowing that it was in the respective states i_0, …i_{n-1} at the prior times 0, …$n-1$, will be called the (average) instantaneous entropy s_n at time *n* [1]

$$\begin{aligned} s_n(p) &= S_{n+1}(p) - S_n(p) = -\sum_{i_0,\ldots i_n} p(i_0,0;\ldots.;i_n,n) \ln p(i_n,n|i_{n-1},n-1;\ldots;i_0,0) \ \geq 0 \\ &= \sum_{i_0,\ldots i_{n-1}} p(i_0,0;\ldots.;i_{n-1},n-1)\, S(p(\bullet,n|i_{n-1},n-1;\ldots;i_0,0)). \end{aligned} \tag{7}$$

Here,

$$S(p(\bullet,n|i_{n-1},n-1;\ldots;i_0,0)) = -\sum_{i_n} p(i_n,n|i_{n-1},n-1;\ldots;i_0,0) \ln p(i_n,n|i_{n-1},n-1;\ldots;i_0,0) \tag{8}$$

is the entropy of the conditional probability at time *n*, conditioned by the past trajectory. It generally differs from the usual 1-time entropy $S(p(.,n))$, which is often used in physics [10,11] when one does not know the previous states of the system. This 1-time entropy is

$$S(p(.,n)) = -\sum_{i_n} p(i_n,n) \ln p(i_n,n). \tag{9}$$

$S(p(.,n))$ is a state function, as defined in thermodynamics. It is seen that

$$S(p(.,n)) - s_n(p) = \sum_{i_0,\cdots n} p(i_0,0;\ldots;.i_n,n) \ln \frac{p(i_n,n|i_{n-1},n-1;\ldots;i_0,0)}{p(i_n,n)} \geq 0 \tag{10}$$

the equality holding only if the state of $\mathcal{S}$ at time *n* is independent of its prior trajectory.

2.2.3. The stationary situation and Kolmogorov entropy

The properties of $S(p_n)$ and $s_n(p)$ have been extensively studied by Kolmogorov and other authors in the case of the stationary process [1,2]. In particular, Kolmogorov [1] showed that if *p* is the stationary process p^0, $s_n(p^0) \equiv s_n^0$ decreases with time *n*

$$s_{n+1}^0 - s_n^0 \leq 0.$$

As a result, s_n^0 tends to a non-negative limit $\overline{s}$ when $n \to \infty$ and

$$\frac{S_n(p^{\grave{a}})}{n} \text{ and } s_n(p^0) \quad \to \overline{s}(p^0) \in [0, s_0(p^0)] \quad \text{if } n \to \infty. \tag{11}$$

With some simplification [1,2], $\overline{s}(p^0)$ is the Kolmogorov entropy of the stationary process p^0.

It may be noticed that since p^0 is stationary from time 0, the state entropy $S(p^0(.,n))$ is clearly a constant s_0, whereas $s_n(p^0)$ decreases from s_0 to $\overline{s}$ when n increases from 0 to infinity.

2.2.4. Memory erasing in the stationary situation

It has been shown recently [2] that the memory of the stationary mesoscopic distribution p^0 can be approximately limited to the n last past events, n depending of the accuracy required for the approximation. More precisely, for any positive number ε, it is possible to find a positive integer $n(\varepsilon)$ such that for any integer $k > n(\varepsilon)$

$$0 < \left\langle \Sigma_{i_k} \ p^0(i_k,k|\, i_{k-1},k-1;...;i_0,0) \ln \frac{p^0(i_k,k|\, i_{k-1},k-1;...;i_0,0)}{p^0(i_k,k|\, i_{k-1},k-1;...;i_{k-n},k-n} \right\rangle_{p^0(0,\ ...k-1)} < \ \varepsilon\,. \tag{12}$$

Here $\langle A \rangle_{p^0(0,\ ...k-1)}$ denotes the average of A with respect to the k-times stationary probability $p^0(i_0,0;....;i_{k-1},k-1)$ .This property implies [2] that if n is large enough, with overwhelming probability (in the space $\mathcal{M}^k \equiv \{i_0, \ldots, i_{k-1}\}$) the absolute distance[(1)] between the complete conditional probability $p^0(i_k,k|\ ...;i_0,0)$ and the truncated conditional probability $p^0(i_k,k|\ ...;i_{k-n},k-n)$ is less than ε.

Thus,

$$\begin{aligned} p^0(i_k,k;\, i_{k-1},k-1;...;i_0,0) &= p^0(i_k,k|\, i_{k-1},k-1;...;i_0,0)\ p(i_{k-1},k-1;...;i_0,0) \\ &\approx p^0(i_k,k|\, i_{k-1},k-1;..;i_{k-n},k-n)\ p(i_{k-1},k-1;...;i_0,0) \\ &\equiv w(i_k|\, i_k,...,i_{k-n})\, p(i_{k-1},k-1;...;i_0,0) \end{aligned} \tag{13}$$

where, because of the stationarity of p^0, w is defined by

$$\begin{aligned} p^0(i_k,k|\, i_{k-1},k-1;...;i_{k-n},k-n) &= p^0(i_k,n|\, i_{k-1},n-1;...;i_{k-n},0) \\ &\equiv w(i_k|\, i_{k-1};...,i_{k-n}) \end{aligned}. \tag{14}$$

Equation (14) defines the transition probability w.

[(1)]***Note.*** The absolute distance [2] between two probabilities p and q on the same discrete space (j) is

$$d(p,q) = \frac{1}{2} \sum_j \left| p_j - q_j \right|.$$

It can be shown [2] that the following Pinsker inequality holds

$$2\big(d(p,q)\big)^2 < S(p|\,q) \equiv \sum_j \ p_j \ln \frac{p_j}{q_j}$$

which implies (13).

2.2.5. Truncated memory, stationary approximation

2.2.5.1. Definitions

Because the memory of the process can be approximately limited to the first n past times, we can define an approximation $\bar{p}^0$ of the stationary process p^0 which is a n-Markov process, as defined below.

(1) n-Markov process

We say that a process q is a n-Markov process if it satisfies the following property:
For any integer $N > n$, one can define a function $w(i_N, N|i_{N-1}, n-1;...;i_{N-n}, N-n)$ of the partial trajectory $i(_{N-n}, i_{N-n-1},...i_N)$ between times N-n and N, such that

$$q(i_N, N;...;i_0, 0) = w(i_N; N|i_{N-1}, N-1;...; i_{N-n}, N-n)\, q(i_{N-1}, N-1;...; i_0, 0) \tag{15}$$

or equivalently

$$q(i_N, N|\ i_{N-1}, N-1;...; i_0, 0) = w(i_N, N|i_{N-1}, N-1;...; i_{N-n}, N-n)\ . \tag{15'}$$

Summing (15) on i_0, … i_{N-n+1}, one sees that, for $N > n$

$$q(i_N, N;...; i_{N-n}, N-n) = w(i_N, N|i_{N-1}, N-1;...; i_{N-n}N-n)\ q(i_{N-1}, N-1;...; i_{N-n}, N-n)$$

and consequently

$$\begin{aligned} q(i_N, N|\ i_{N-1}, N-1;...; i_{N-n}, N-n) &= w(i_N, N|i_{N-1}, N-1; ...; i_{N-n}, N-n) \\ &= q(i_N, N|\ i_{N-1}, N-1;...; i_0, 0) \end{aligned} \tag{16}$$

This is the characteristic property of a n-Markov process: the conditional probability of any state at time N>n, conditioned by its complete past trajectory from time 0, is identical to the conditional probability of this state, conditioned by its past trajectory during the n previous times only.

Note that if q is stationary, formula (15) shows that w is invariant if time N is replaced by N+ h, where h is any positive integer, so w is independent of N

$$w(i_N, N|i_{N-1}, N-1; ...; i_{N-n}, N-n) = w(i_N, n|i_{N-1}, n-1; ...; i_0, 0) \equiv w(i_N|i_{N-1}, ..., i_{N-n}) \tag{17}$$

It is seen that, by the approximate equality (13), the coarse-grained stationary distribution p^0 of Section 2 is almost an n-Markov process. We now state this property more rigorously.

(2) Truncated stationary process

The truncated process $\bar{p}^0$ is defined from p^0 by

$$\bar{p}^0(i_k, k;...;i_0, 0) = p^0(i_k, k;...;i_0, 0) \quad \text{for} \quad k \le n \tag{18}$$

and for $k > n$, it is obtained by repeated applications of the following iterative formula simulating Eq (13)

$$\begin{aligned} \bar{p}^0(i_k, k; i_{k-1}, k-1;...i_0, 0) &= p^0(i_k, k|\, i_{k-1}, k-1;...;i_{k-n}, k-n)\ \bar{p}^0(i_{k-1}, k-1;...;i_0, 0) \\ &= w(i_k|\, i_k, ..., i_{k-n})\, \bar{p}^0(i_{k-1}, k-1;...;i_0, 0) \end{aligned} \tag{19}$$

All probabilities for fragmentary trajectories between time N-n and N are obtained by summing (19) on irrelevant states. The conditions of Kolmogorov theorem [3] are then satisfied and the n-Markov process $\bar{p}^0$ is defined by its probability law.

The fact that $\bar{p}^0$ is an approximation of p^0, as described in 3.4, was derived in [2].

2.2.5.2. Evolution of the truncated process

Equation (19) can easily be generalized to (writing now times in *decreasing* order)

$$\bar{p}^0(i_{k+n-1}, k+n-1; \dots; i_k, k; \dots; i_0, 0) = w(i_{k+n-1}, \dots i_k \mid i_{k-1}, \dots, i_{k-n}) \times \bar{p}^0(i_{k-1}, \dots; i_{k-n}, k-n; \dots; i_0, 0) \tag{20}$$

with

$$w(i_{k+n-1}, \dots, i_k \mid i_{k-1}, \dots i_{k-n}) = w(_{k+n-1} \mid i_{k+n-2}, \dots i, _{k-1}) \dots w(, i_k \mid i_{k-1}, \dots, i_{k-n}) \ . \tag{21}$$

Summing (20) on i_0, $\dots i_{k-1}$, we obtain

$$\bar{p}^0(i_{k+n-1}, k+n-1; \dots; i_k, k) = \sum_{i_{k-1, \dots k-n}} w(i_{k+n-1}, \dots i_k \mid i_{k-1}, k-1; \dots i_{k-n}) \times \bar{p}^0(i_{k-1}, k-1; \dots; i_{k-n}, k-n) \tag{22}$$

which is clearly a kind of master equation. It is written more easily in the following formalism of partial trajectories.

2.2.5.3. Partial, *n*-steps trajectories and master equation

We consider integers $K \geq 0$ and we define the partial *n*-steps trajectories (again writing times in decreasing order) $I_K = (i_{(K+1)n-1}), \dots, i_{Kn+1}, i_{Kn})$ at the n corresponding, decreasing times $T_K = ((K+1)n-1, \dots, Kn+1, Kn)$.

Here, n is the integer determined by the accuracy needed for the approximation, according to formula (12).

So, in abbreviated notations, we can write

$$(I_K, T_K) \equiv (i_{(K+1)n-1}, (K+1)n-1; \dots; I_{Kn+1}, Kn+1; i_{Kn}, Kn)$$

and

$$\bar{P}^0(I_K, T_K) \equiv \bar{p}^0(i_{K+n-1}, K+n-1; \dots; i_K, K) \ . \tag{23}$$

Choosing $k = Kn$ in (22) for some positive integer K, this equation takes the condensed form

$$\bar{P}^0(I_K, T_K) = \Sigma_{I_{K-1}} W(I_K \mid I_{K-1}) \, \bar{P}^0(I_{K-1}, T_{K-1}) \ . \tag{24}$$

Because of the stationarity of p^0, $W(I_K \mid I_{K-1}) \equiv P^0(I_K, T_K \mid I_{K-1}, T_{K-1})$ is independent of T_K, T_{K-1}. So, for $K \geq 1$, (24) is a generalized master equation [3,12] for the *n*-steps partial trajectories and the generalized transition rate $W(I_K \mid I_{K-1})$ can be explicitly computed from (21).

At this step, Eq (24) essentially has a formal interest, since the solution $\bar{P}^0(I_K, T_K)$ is known from (19). However, it will prove to be very useful in Section 3.1.2.

3. Results

3.1. Kolmogorov entropy in the non-stationary situation

3.1.1. Non stationary mesoscopic process

Assuming that the microscopic probability at the initial time is piecewise uniform, we saw that the mesoscopic *k*-times law is given by (3) or

$$p(i_{k-1},k-1;\ ...;i_1,1;i_0,0) \;=\; \frac{p(i_0)}{\mu(i_0)}\,\mu(\varphi_{-k+1}\,i_{t_{k-1}} \cap ...\varphi_{-1}\,i_1 \cap i_0) \qquad . \tag{25}$$

Thus,

$$p(i_k \mid i_{k-1},k-1;\ ...;\ i_1,1;i_0,0) \;=\; p^0(i_k \mid i_{k-1},k-1;\ ...;\ i_1,1;i_0,0)\ . \tag{26}$$

Many similar equalities can be found between conditional probabilities of the non-stationary law *p* and the similar conditional probabilities of the stationary law p^0, provided that *the conditioning trajectory includes the initial mesostate* i_0 *at time* 0. In fact, although the following notations may be heavy, one can easily deduce from (26) the following property:

If $A = \{i_h\}, h, = 1, 2, \ldots k$, is any set of $k \geq 1$ mesostates, and if $T = \{t_h\}$, $h = 1, 2, \ldots k$, *i*s any set of *k* positive integer times t_h, then the conditional probability of the partial trajectory $(A, T) = \{i_h, t_h\}$, conditioned by some partial trajectory at times not included in *T*, but *including state* i_0 *at time* 0, is identical to the corresponding conditional probability calculated with the stationary law p_k^0.

These very simple properties allow one to show [2] that the nonstationary process obeys an approximate generalized master equation which holds on the partial trajectories (I_K, T_K) of length *n* defined in 3.5.2.

They have been evoked in previous papers [2,13]. New results will now be presented and justified. Detailed derivations are postponed to Appendix A.

3.1.2. Nonstationary finite memory approximation

Thanks to (12), we can write for $k > n$

$$\begin{aligned} p(i_k,k;\ ...;i_1,1;\ i_0,0) &= p(i_k,k \mid i_{k-1},k-1;\ ...;\ i_1,1;\ i_0,0)\ \ p(i_{k-1},k-1;\ ...;\ i_1,1;\ i_0,0) \\ &= p^0(i_k,k \mid i_{k-1},k-1;\ ...;\ i_1,1;\ i_0,0)\ \ p(i_{k-1},k-1;\ ...;\ i_1,1;i_0,0)\ . \\ &\approx p^0(i_k,k \mid i_{k-1},k-1;\ ...;i_{k-n},k-n)\ \ p(i_{k-1},k-1;\ ...;\ i_1,1;i_0,0) \end{aligned} \tag{27}$$

Thus, summing on $i_0, \ldots i_{n-k-1}$ we obtain

$$p(i_k,k;\ ...;i_{k-n},k-n) \;\approx\; p^0(i_k,k \mid i_{k-1},k-1;\ ...;\ i_{k-n},k-n)\ \ p(i_{k-1},k-1;\ ...;\ i_{k-n},k-n)\,. \tag{28}$$

Comparing (28) and (29), it is seen that the memory of the non-stationary process *p* at time *k* is approximately limited to the first *n* past times $k-1,\ldots\ k-n$, as for the memory of the stationary process

$$p(i_k,k|i_{k-1},k-1;...;i_{k-n},k-n) \approx p(i_k,k|i_{k-1},k-1;...;i_0,0)$$
$$= p^0(i_k,k|\, i_{k-1},k-1;\, ...;\, i_{k-n},k-n) \; = w(i_k|\, i_{k-1},\, ...,i_{k-n}) \quad (29)$$

Similar to the reasoning of paragraph 2.2.5.2, Eq (29) implies

$$p(i_{k+n-1},k+n-1;...;i_k,k) \approx \sum_{i_{k-1,...k-n}} w(i_{k+n-1},\, ...,i_k|\, i_{k-1},....,i_{k-n})$$
$$\times\; p(i_{k-1},k-1;...;i_{k-n},k-n) \quad (30)$$

Taking $k = Kn$ for $K \geq 1$ and using the condensed notations defined in 3.5.2, it can be concluded that, with a high probability

$$P(I_K,T_K) \approx \sum_{I_{K-1}} W(I_K\,|\,I_{K-1})\; P(I_{K-1},T_{K-1})\;. \quad (31)$$

So, the non-stationary probability $P(I_K,T_K)$ approximately satisfies the master Eq (24). Assume that the (nM)x(nM) stochastic matrix $W = \left(W(I_K\,|\,I_{K-1})\right)$is regular. Then, the exact stationary solution P^0 of Eq (31) is unique.

From the theory of stochastic matrices [14] it is expected that, for any n-steps partial trajectory $J = j_0,\, j_1....,\, j_{n-1}$ at the successive times $T \equiv Kn,\, Kn+1;....,(K+1)n-1$, we have for any I_0

$$P(J,T|\,I_0,T_0) \rightarrow P^0(J) \text{ when } K\rightarrow \infty. \quad (32)$$

This can be proved with relevant assumptions (see [2] and remarks below). Consequently, if $K\rightarrow\infty$ (again writing times in decreasing order)

$$p(j_{n-1},(K+1)n-1;...;j_0,Kn|\,i_{n-1},n-1:...:i_0,0) \rightarrow \mu(j_{n-1},...j(_0) \quad . \quad (33)$$

Renumbering the times and summing on appropriate indexes, (33) implies that, for any positive integers h, k, m and for $m < k$

$$p(i_{k+h},k+h;...;i_k,k|i_m,m;...;i_0,0) \rightarrow \mu(i_{k+h},k+h;...;i_k,k) \text{ when } k \rightarrow\infty. \quad (34)$$

So, the non-stationary process is mixing [1,3], as well as the stationary process.
The meaning of approximation 3.1.2 is further discussed in Appendix A.

Remarks. The reasoning from Eq (31) to Eq (34) applies to the truncated approximation $\overline{p}$ that can be defined from p as $\overline{p}^0$ is defined from the stationary distribution p^0 in Section 2.2.5.1. It is shown in [2] that in the notation of partial trajectories, $\overline{P}(I_K,T_K)$ satisfies (31). So, it results [2] from the matrix theory that if the matrix $W(I_K|I_{K-1})$ is regular, $\overline{P}(I_K,T_K)$ tends to the stationary solution of (31) for any initial partial trajectory I_0 when $K\rightarrow\infty$

$$\overline{P}(I_K,T_K|\,I_0,T_0) \rightarrow P^0(I_K) \text{ for any } I_0 \text{ if } K\rightarrow\infty. \quad (35)$$

3.1.3. Instantaneous entropy of the non-stationary process

We now present our main, new result: Under certain, reasonable conditions, the instantaneous entropy $s_n(p)$ tends to a finite limit $s(p)$ when $n\rightarrow\infty$, thus defining the Kolmogorov entropy $s(p)$ of the mesoscopic non-stationary process p. So, with relevant assumptions the basic result established by Kolmogorov for the stationary situation p^0 is extended to non-stationarity. The discussion of this assertion needs some detailed calculations which are discussed in Appendix **A**. The proof can be summarized as follows.

With a given accuracy ε, the memory of the process can be neglected at times larger than $n(\varepsilon)$, in the sense of Section 2.2.4. Then, it is probable that the instantaneous entropy (6) at time $N= k + n(\varepsilon)$ is, for any $k> 0$,

$$\begin{aligned} s_{k+n}(p) &= -\sum_{i_0} p(i_0)\sum_{i_1,\dots i_{k+n}} \frac{\mu(i_{k+n},k+n;\dots;i_0,0)}{\mu(i_0)} \ln \mu(i_{k+n},k+n|i_{k+n-1},k+n-1;\dots;i_0,0) \\ &\approx -\sum_{i_0} p(i_0)\sum_{i_1,\dots i_{k+n}} \frac{\mu(i_{k+n},k+n;\ \dots;i_1,1;i_0,0)}{\mu(i_0)} \ln \mu(i_{k+n},k+n|i_{k+n-1},k+n-1;\dots;i_k,k) \\ &= -\sum_{i_0} p(i_0)\sum_{i_k,\dots i_{k+n}} \frac{\mu(i_{k+n},k+n;\dots;i_k,k;i_0,0)}{\mu(i_0)} \ln \mu(i_{k+n},k+n|i_{k+n-1},k+n-1;\dots;i_k,k) \end{aligned} \tag{36}$$

Here, $p(i_0)$ is the initial, non-stationary probability of i_0 and $\mu(.i_k,k;\dots;i_0,0)$ is the stationary probability of a trajectory at times $k. \dots, 0$.
Considering that by (34), the stationary process is mixing [1,3], we have

$$\begin{aligned} \mu(i_{k+n},k+n;\ \dots;i_k,k|i_0,0) &\equiv \frac{\mu(i_{k+n},k+n;\dots;\ i_k,k;\ i_0,0)}{\mu(i_0)} \\ &\to \mu(i_{k+n},k+n;\dots;i_k,k) \qquad if \quad k\to\infty \end{aligned} \tag{37}$$

Thus, if k and n are large enough,

$$s_{k+n}(p) \approx -\sum_{i_k,\dots i_{k+n}} \mu(i_k,k;\dots;\ .i_{k+n},k+n) \ln \mu(i_{k+n},k+n|i_{k+n-1},k+n-1;\dots;i_k,k) = s_n^0 \tag{38}$$

where we used the stationarity of the measure μ. It follows from (38) and (11) that $s_N(p)$ and $s_N(p^0) \equiv s_N^0$ have a common limit $\bar{s}$ when $N\to\infty$

$$s_N(p) \to s(p^0) = \bar{s} \quad \text{if} \quad N\to\infty. \tag{39}$$

So, in appropriate conditions the Kolmogorov entropy $\bar{s}$ is defined even for a non-stationary mesoscopic process. This is our main result: Its detailed justification is commented in Appendix A.

4. Discussion and conclusions

We have proved that with relevant hypothesis, even in the non-stationary situation, the instantaneous entropy of the mesoscopic coarse-grained process tends to a finite limit, depending only on its stationary measure μ. Thus, we have completed the analysis presented in [2], which essentially proved that the partial trajectories traveled during a mesoscopic time interval n can be approximated by a n-times Markov process if n is large enough. This was done with relatively simple methods, although the notations and the calculations may be somewhat heavy. A more general theory, using the formalism of martingales, will be presented in a further publication (B. Gaveau, M. Moreau, Coarse-graining a deterministic system: Martingale theory, unpublished work). A complete discussion of our approach, including a comparison with other points of view (see for instance [24] and references herein) should profitably be performed at the light of the forthcoming article.

From the present results, the entropy introduced by Kolmogorov can suggest new remarks concerning time and its relations with physics and probabilities. This subject has been addressed in a vast literature and over the course of centuries innumerable philosophers tried to analyze time [14]. Of course, we do not intend to discuss or even to evoke all these works. We would only like to point out that the Kolmogorov entropy presents an innovative point of view on two basic aspects of time: Its irreversibility and its (apparently) regular progress. Since Boltzmann [16,17], time irreversibility (the arrow of time) is linked to the growth of entropy for isolated systems. This principle stems from classical thermodynamics [10,11,18], but it received a first analytical basis thanks to Boltzmann [16,17], who not only gave a theoretical definition of entropy but also proved, by his celebrated H theorem, that the one-time entropy of a non-equilibrium isolated system increases with time, within the collision model of low density gases. However, this is a very specific model. In spite of some possible extensions, it gives no assurance on the generality of this conclusion. More recently, the relation between time and irreversibility again attracted the attention of many scientists (see for instance [19–25]). In particular it was given an original form by I. Prigogine [20–23]. However, despite their interest, it seems that many of these works are restricted to special examples and can hardly represent a general approach.

Things are clearer if, with Kolmogorov, one considers the entropy of a stochastic process. On the one hand, this entropy is directly associated with its time evolution, which is included in the very definition of stochastic processes [3]. Furthermore, it is based on the memory of the events. Not only does this point agree with current observation, but it also meets the opinion of philosophers who pointed out that subjective time is related to human memory (see for instance [16] and references therein). On the other hand, the trajectory entropy clearly increases with time in all circumstances, whereas the average, instantaneous entropy (conditioned by the past) does not necessarily increase with time: It is even non-increasing for a stationary trajectory, as proved by Kolmogorov [1].

Eventually, the instantaneous entropy tends to a finite limit for a stationary process and in certain conditions for a non-stationary process as well, as shown in this article. This fact appears to be linked with the regular flow of time commonly experienced. In the Kolmogorov approach, the time scale of the process is related to the rate of information creation due to the possible observations, according to Shannon [5], or it arises from the rate of disorder production due to the stochastic evolution, according to Boltzmann.

We think that these simple remarks, concerning a very old and complex problem, could deserve to be developed in the framework of Kolmogorov entropy.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

Appendix A. On the derivation of the main result, formula (39)

A.1. Memory erasing and instantaneous entropy of the non-stationary mesoscopic distribution

Although the memory erasing expressed by formula (12) only holds in the stationary situation, it has important consequences for the instantaneous entropy s_n of a nonstationary mesoscopic process, defined by (6). This is due to the fact that the nonstationary conditional probabilities are equal to the stationary conditional probabilities in the specific case considered here and described in Section 3.1.1. So, the instantaneous entropy s_N at time N can be written by (6)

$$\begin{aligned} s_N &= -\sum_{i_0} p(i_0)\sum_{i_1,\dots i_N} \frac{\mu(i_N,N;\dots;i_0,0)}{\mu(i_0)} \ln \mu(i_N,N|i_{N-1},N-1;\dots;i_N,N-n) \\ &\quad -\sum_{i_0} p(i_0)\sum_{i_1,\dots i_N} \frac{\mu(i_N,N;\dots;i_0,0)}{\mu(i_0)} \ln \frac{\mu(i_N,N|i_{N-1},N-1;\dots;i_0,0)}{\mu(i_N,N|i_{N-1},N-1;\dots;i_{N-n},N-n)} \\ &\equiv s_N^{(1)} + s_N^{(2)} \quad \text{(the respective sums appearing in the 1st line and in the 2nd line abose).} \end{aligned} \tag{A.1}$$

Let $N = k + n$, n and k be positive integers. Assuming that the stationary process μ is mixing in the sense defined by (32), it is found that if k is large enough,

$$s_{n+k}^{(1)} \approx s_n^0$$

where s_n^0 is the instantaneous entropy of the stationary process at time n (see Eq (37)). So, if k and $n\to\infty$

$$s_{n+k}^{(1)} \to \bar{s} \tag{A.2}$$

and s is the Kolmogorovb entropy (39) of the stationary process.
On the other hand, it results from the memory erasing property (12) that, in $s_N^{(2)}$, the ratio

$$\lambda \equiv \frac{\mu(i_N,N|i_{N-1},N-1;\dots;i_0,0)}{\mu(i_N,N|i_{N-1},N-1;\dots;i_{N-n},N-n)} \tag{A.3}$$

is (probably) very close to 1 if n is large enough. As a consequence, it can be proved, with relevant assumptions (see below) that

$$s_N^{(2)} \to 0 \text{ if } N \to \infty. \tag{A.4}$$

From (A.1-3), we can conclude with some additional hypotheses that

$$s_N \to \bar{s} \quad \text{if } N \to \infty. \tag{A.5}$$

Then, the instantaneous entropy of a nonstationary process tends to the same limit $\bar{s}$ as the stationary process corresponding to the stationary measure μ.

It should be pointed out that supplementary hypotheses are necessary to prove (A.5). In fact, the memory erasing relation (12), which is essential for the derivation, is most probably satisfied but with a very low probability it can fail to be verified. A sufficient condition for obtaining (A.5) is that the ratio λ of (A.3) has finite upper and lower bounds, independent of n and N.

With this assumption, it is not difficult to derive the previous results, but the calculations are lengthy. Rather than detailing them we prefer to summarize and discuss the main hypotheses used in the reasoning.

A.2. Complementary discussion

It has been shown that the mesoscopic process approximately satisfies a generalized Markov process, whose probability tends to the stationary probability at infinite times. This fact does not necessarily imply that the mesoscopic probability has the same property. We gave arguments in this sense elsewhere [2,13].

This is a general problem in modeling, when the evolution of some actual system is shown to obey approximate, theoretical equations: It is difficult to know whether their formal asymptotics still represent the natural system correctly. This point is overlooked in some physical publications. Even if mathematical conditions are found to ensure the relevance of the model at large times, it is difficult to check whether these conditions are verified in practice. The present study does not avoid this difficulty completely.

Another basic point is that the stationary, mesoscopic probability μ is supposed to be mixing (see (32)). This property seems to be a natural extension of the memory erasing, proved for the mesoscopic process. In fact, this assertion is not obvious, but it is true for a process with a finite n-steps memory, or generalized Markov process. According to the previous discussion, this should be true also for the stationary probability μ, since we have shown that, with any accuracy ε, μ can be approximated by a process $\overline{p}^0$ with a finite memory of $n(\varepsilon)$, steps.

Eventually, the assumption that the ratio λ, defined by (A.3), has finite, nonzero upper and lower bounds seems reasonable because it just completes and reinforces the fact that λ is almost everywhere close to 1. One can notice that the ratio λ should only be considered if $\mu(i_N, N|i_{N-1}, N-1;...;i_{N-n}, N-n) > 0$. The case $\mu(i_N, N|i_{N-1}, N-1;...;i_{N-n}, N-n) = 0$ implies $\mu(i_N, N; i_{N-1}, N-1;...;i_{N-n}, N-n) = 0$ and $\mu(i_N, N; i_{N-1}, N-1;...;i_0, 0) = 0$. In this situation, the partial trajectory $(i_0, 0;...;i_N, N)$ does not contribute to the entropy s_N and it can be ignored for calculating λ. Because the space $\mathcal{M}$ of mesostates i is finite and consists in M elements, the number of partial trajectories $i_{N-n}, \dots i_N$ is M^n and the upper and lower bounds of λ are finite and independent of N. Their values can only be estimated in specific cases or for some academic models.

References

1. V. I. Arnold, A. Avez, *Ergodic problems of classical mechanics*, Mathematical Physics Monographs, Benjamin, 1968.
2. B. Gaveau, M. Moreau, On the stochastic representation and Markov approximation of Hamiltonian systems, *Chaos*, **30** (2020), 083104. https://doi.org/10.1063/5.0001435
3. J. Doob, *Stochastic processes*, Wiley, New York, 1953.
4. P. Levy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937.
5. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27** (1948), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x
6. A. I. Khinchin, *Mathematical foundations of information theory*, Dover, New York, 1957.
7. R. MacMillan, The basic theorems of information theory, *Ann. Math. Stat.*, **24** (1953), 193. https://doi.org/10.1214/aoms/1177729028
8. T. Cover, J. Thomas, *Elements of information theory*, Wiley, New York, 1991.
9. L. Brillouin, *Science an information theory*, Academic Press, New York, 1956.
10. L. D. Landau, E. M. Lifshitz, *Statistical physics*, 3 Eds., Pergamon Press, Oxford, 1980.

11. F. Reif, *Fundamentals of statistical and thermal physics*, Mc Graw Hill, New York, 1965.
12. W. Feller, *An introduction to probability theory and its applications*, Wiley, New York, 1971.
13. M. Moreau, B. Gaveau, *Stochastic theory of coarse-grained deterministic systems: Martingales and Markov applications*, In: Advances in Dynamical Systems Theory, Models, Algorithms and Applications, 2021. https://doi.org/10.5772/intechopen.95903
14. G. Akemann, J. Baik, P. Di Francesco, *The Oxford handbook of random matrix theory*, Oxford University Press, 2015. https://doi.org/10.1093/oxfordhb/9780198744191.001.0001
15. C. Hoerl, T. McCormack, *Time and memory*: *Issues in philosophy and psychology*, Clarendon, Oxforf, 2001.
16. L. Boltzmann, *Vorlesungenüber gastheorie, Part 2,* J. A. Barth, Leipzig, 1898.
17. L. Boltzmann, *Reply to Zermelo's remarks on the theory of heat*, In: History of modern physical sciences: The kinetic theory of gases, Imperial College Press, **57** (1896), 567.
18. H. B. Callen, *Thermodynamics and an introduction to thermostatistics*, John Wiley and Sons, New York, 1985.
19. L. S. Schulman, *Time's arrows and quantum measurement*, Cambridge University Press, Cambridge, 1997. https://doi.org/10.1017/CBO9780511622878
20. L. S. Schulman, arXiv.cond-mat/991101, 1999.
21. I. Prigogine, *From being to becoming*, Freeman, 1980.
22. I. Prigogine, *Time and change*, 2003.
23. I. Prigogine, *Time, dynamics and chaos: Integrating Poincare's non-integrable systems*, Center for Studies in Statistical Mechanics and Complex Systems at the University of Texas-Austin, United States Department of Energy-Office of Energy Research, Commission of the European Communities, 1990.
24. J. B. Gao, Y. H. Cao, W. W. Tung, J. Hu, *Multiscale analysis of complex time series-integration of chaos and random fractal theory, and beyond*, Wiley, New York, 2007.
25. J. B. Gao, F. Y. Liu, J. F. Zhang, J. Hu, Y. H. Cao, Information entropy as a basic building block of complexity theory, *Entropy*, **15** (2013), 3396. https://doi.org/10.3390/e15093396