



**HAL**  
open science

# Different Closed-Form Expressions for Generalized Entropy Rates of Markov Chains

Valerie Girardin, Loïck Lhote, Philippe Regnault

► **To cite this version:**

Valerie Girardin, Loïck Lhote, Philippe Regnault. Different Closed-Form Expressions for Generalized Entropy Rates of Markov Chains. *Methodology and Computing in Applied Probability*, 2018, 10.1007/s11009-018-9679-3 . hal-01966810

**HAL Id: hal-01966810**

**<https://hal.science/hal-01966810>**

Submitted on 29 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Different closed-form expressions for generalized entropy rates of Markov chains.

Valérie Girardin · Loick Lhote ·  
Philippe Regnault

Received: date / Accepted: date

**Abstract** Closed-form expressions for generalized entropy rates of Markov chains are obtained through pertinent averaging. First, the rates are expressed in terms of Perron-Frobenius eigenvalues of perturbations of the transition matrices. This leads to a classification of generalized entropy functionals into five exclusive types. Then, a weighted expression is obtained in which the associated Perron-Frobenius eigenvectors play the same role as the stationary distribution in the well-known weighted expression of Shannon entropy rate. Finally, all terms are shown to bear a meaning in terms of dynamics of an auxiliary absorbing Markov chain through the notion of quasi-limit distribution.

**Keywords** Entropy rate · Ergodic Markov chain · Entropy distribution · Quasi-limit distribution

## 1 Introduction

The concept of entropy was introduced in probability theory in [29] by defining

$$\mathbb{S}(m) = - \sum_{i \in E} m_i \log m_i, \quad (1)$$

with the convention  $0 \log 0 = 0$ , for any distribution  $m$  taking values in a set  $E$ , either finite or denumerable up to convergence restrictions. Apart from information theory to which it gave birth, entropy is an essential tool in all scientific fields involving probability measures; see e.g., [6]. Shannon's original entropy naturally

---

V. Girardin  
Laboratoire de mathématiques N. Oresme, Université de Caen Normandie, BP5186, 14032 Caen, France. E-mail: valerie.girardin@unicaen.fr

L. Lhote  
ENSICAEN, GREYC, UMR 6072, Université de Caen Normandie, BP5186, 14032 Caen, France E-mail: loick.lhote@unicaen.fr

P. Regnault  
Laboratoire de Mathématiques de Reims, EA 4535, Université de Reims Champagne-Ardenne, BP 1039, 51687 Reims cedex 2, France.

generalizes to various entropy functionals, such as Rényi's, Tsallis', etc., introduced in the literature for a better fit to complex systems. Most of them belong to the class of  $(h, \varphi)$ -entropy functionals  $\mathbb{S}_{h(z), \varphi(x)}$ , defined by

$$\mathbb{S}_{h(z), \varphi(x)}(m) = h \left( \sum_{i \in E} \varphi(m(i)) \right); \quad (2)$$

see [26] and [2]. In particular, Rényi entropy functionals – defined in [23] and depending on a parameter  $q > 0$  – are given by

$$\mathbb{S}_{\frac{1}{1-q} \log z, x^q}(m) = \frac{1}{1-q} \log \left( \sum_{i \in E} m_i^q \right), \quad q \neq 1,$$

and for  $q = 1$  by  $\mathbb{S}(m)$  defined in (1).

The entropy of a random variable is the entropy of its distribution. Thus, for any random vector  $\mathbf{X}_0^n = (X_0, \dots, X_{n-1})$  defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  with distribution  $\mathbf{m}_0^n$ , we get  $\mathbb{S}_{h(z), \varphi(x)}(\mathbf{X}_0^n) = \mathbb{S}_{h(z), \varphi(x)}(\mathbf{m}_0^n)$ . For a random sequence  $\mathbf{X} = (X_n)_{n \geq 0}$ , averaging per time unit leads to

$$\mathbb{H}_{h(z), \varphi(x)}(\mathbf{X}) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{S}_{h(z), \varphi(x)}(\mathbf{X}_0^n);$$

this limit – when well defined – is called the entropy rate of the process.

Markov chains and entropy are linked since [29]. The Shannon entropy rate  $\mathbb{H}(\mathbf{X}) = \mathbb{H}_{z, -x \log x}(\mathbf{X})$  of any ergodic (homogeneous aperiodic irreducible) Markov chain is the sum of the entropic values of the transition probabilities  $p_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i)$  weighted by the probability of occurrence of each state according to the stationary distribution  $\pi$  of the chain; in mathematical words,

$$\mathbb{H}(\mathbf{X}) = - \sum_{i \in E} \pi_i \sum_{j \in E} p_{ij} \log p_{ij}. \quad (3)$$

The convergence of  $\frac{1}{n} \mathbb{S}(\mathbf{X}_0^n)$  to (3) is proven in [29] for ergodic finite chains, by the chain rule specific to Shannon entropy. It has since been extended to the denumerable case, and then proven to hold under hypotheses weakened in many directions; see [6], and [11].

Another closed-form expression for the Shannon entropy rate of any ergodic Markov chain is

$$\mathbb{H}(\mathbf{X}) = \lambda'(1), \quad (4)$$

where  $\lambda(s)$  denotes the dominant eigenvalue of the perturbed matrix

$$P(s) = (p_{ij}^s)_{(i,j) \in E^2}, \quad s > 0; \quad (5)$$

see [31] for links to dynamical systems theory. Note that  $\lambda(s)$  is a short-hand notation for  $\lambda(P, s)$  – a function of both  $P$  and  $s$ , that will be used thereafter when no ambiguity may arise on  $P$ .

The Rényi entropy rates of finite state space ergodic Markov chains have been shown in [21] and [22] to be well-defined and given by

$$\mathbb{H}_{\frac{1}{1-q} \log z, x^q}(\mathbf{X}) = (1-q)^{-1} \log \lambda(q), \quad q \neq 1. \quad (6)$$

Extension to denumerable Markov chains is proven to hold in [12]. Further, a closed form weighted expression similar to (3) is established in [24], involving the normalized left eigenvector  $\mathbf{u}(s)$  of  $P(s)$  associated to  $\lambda(s)$ , namely

$$\mathbb{H}_{\frac{1}{1-\mathfrak{q}} \log z, x^{\mathfrak{q}}}(\mathbf{X}) = \frac{1}{1-\mathfrak{q}} \log \left[ \sum_{i \in E} u_i(\mathfrak{q}) \sum_{j \in E} p_{ij}^{\mathfrak{q}} \right]. \quad (7)$$

The proof relies on a chain rule specific to Rényi entropy, that is based on the concept of escort distributions of the marginal distributions of Markov chains. As pointed out in [24], no chain rule can hold for other entropy functionals, hindering to obtain by following similar steps such weighted expressions for other  $(h, \varphi)$ -entropy functionals.

In [5], generalized entropy rates per time unit of a wide class of random sequences including Markov chains are shown to be trivial – either null or infinite – out of a threshold where they are equal to either Shannon or Rényi entropy rates, up to a multiplicative factor. In [12], non trivial divergence rates are obtained through considering suitable averaging sequences  $\mathbf{r} = (r_n)_{n \in \mathbb{N}}$  instead of  $n$ ; entropy appears only as a by-product of divergence, mainly through examples, with rates defined by

$$\mathbb{H}_{h, \varphi, \mathbf{r}}(\mathbf{X}) = \lim_{n \rightarrow \infty} \frac{1}{r_n} \mathbb{S}_{h(z), \varphi(x)}(\mathbf{X}_0^n) = \lim_{n \rightarrow \infty} \frac{1}{r_n} \mathbb{S}_{h(z), \varphi(x)}(\mathbf{m}_0^n). \quad (8)$$

The present paper focuses on closed-form expressions of such generalized entropy rates for Markov chains, closely linked to the dynamics of the chains. Highlighting, particularizing and completing the results of [12] for the special case of entropy will lead us to obtain formulas similar to (3) and (4), and then to (6) and (7), for a large class of  $(h, \varphi)$ -entropy functionals.

First, formulas extending (4) and (6) will be obtained, based on the functions that map  $s > 0$  to the dominant eigenvalue  $\lambda(s)$  of (5) and to its derivative. A classification in five classes of generalized entropy functionals, that holds true for a large class of Markov chains, is derived, depending on the asymptotic behavior of the marginal entropy  $\mathbb{S}_{h(z), \varphi(x)}(\mathbf{X}_0^n)$ . The transition matrices of the considered ergodic Markov chains are only required to satisfy easy to check assumptions, inducing the marginal entropy to behave like the power of some analytic function. Then a relation between  $\lambda(s)$  and the left and right eigenvectors  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$  of  $P(s)$  associated to  $\lambda(s)$  leads to weighted formulas extending (3) and (7), in which  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$  replace the stationary distribution  $\pi$ .

Finally, all involved quantities are interpreted in terms of dynamics of Markov chains. In particular, the probability distributions defined by  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$  are respectively the quasi limit distribution (QLD) and the asymptotic odds of an auxiliary absorbing Markov chain, that is naturally linked to the original chain and to the parameter  $s$ . Furthermore, the probability distributions defined by  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$  and hence all involved quantities are also interpreted in terms of marginals of escort distributions.

The availability of these closed-form expressions opens the way to extend to generalized entropy functionals the sophisticated classical toolbox linked to (3) and (4), including maximum entropy methods and other statistical tools, that require entropy rates to be written as smooth functions of the transition matrix of the chain.

For example, [9], [15] and [10] address the problem of maximization of the entropy rate of Markov chains, subject to a large variety of constraints. Most results rely on the method of Lagrange multipliers and hence make intensive use of the expression (3) of the entropy rate as a function of the transition matrix  $P$ . Application of the maximum entropy principle to various frameworks follows; see, e.g., [5] in queueing theory, or [25] in control in networks, with the so-called randomized shortest-path problem. Many applications in information theory are detailed in [6], especially for computing the capacity of communication channels. From a statistical point of view, the closed form expression (3) allows plug-in estimation of  $\mathbb{H}(\mathbf{X})$  from the estimation of  $P$ ; see [4] for the asymptotic properties of these estimators. Plug-in estimation is extensively used in applications. See for instance [7] and the references therein for computation of the entropy rate of paths connecting two states. See also [17] for plug-in estimation of the entropy rate of Markov trajectories between two states, conditional to given intermediate states, a quantity used as a discriminant measure in classification procedures.

Both (4) and (6) are central in coding theory. (4) is an important tool in dynamical systems theory, associated to Dirichlet series; see [31]. (6) is developed in [21] with application to search algorithms. Finally, statistical estimators of Rényi entropy rates for parametric Markov chains are determined in [5] by making use of (6).

The paper is organized as follows. The considered class of  $(h, \varphi)$ -entropy functionals is detailed in Section 2. In Section 3, the  $(h, \varphi)$ -entropy rates are derived as explicit functions of the dominant eigenvalue  $\lambda(s)$  of  $P(s)$ . In Section 4, the classification of  $(h, \varphi)$ -entropy functionals into five exclusive classes is derived from the asymptotic behavior of the marginal entropy of the considered Markov chain. In Section 5, the  $(h, \varphi)$ -entropy rates are derived as explicit weighted expressions of the same kind as (3) and (7). All involved quantities are interpreted in terms of marginals of escort distributions in Section 6.1 and in terms of dynamics of an auxiliary absorbing Markov chain in Section 6.2.

## 2 Quasi power log entropy functionals

In the literature, most  $(h, \varphi)$ -entropy functionals  $\mathbb{S}_{h(z), \varphi(x)}$  defined by (2), with  $\varphi : [0, 1] \rightarrow \mathbb{R}_+$  and  $h : \mathbb{R}^+ \rightarrow \mathbb{R}$ , are such that either  $\varphi$  is concave and  $h$  increasing or  $\varphi$  is convex and  $h$  decreasing. Moreover  $h$  and  $\varphi$  behave such that  $\mathbb{S}_{h, \varphi}$  takes only non-negative values. For most functionals of interest,  $h$  and  $\varphi$  are locally equivalent to products of power and logarithmic functions, which leads us to define the class of QPL entropy functionals.

**Definition 1** *Any  $(h, \varphi)$ -entropy functional is called a quasi power log (QPL) entropy if  $a \in \mathbb{R}^*$ ,  $s \in \mathbb{R}$ ,  $\delta \in \{0, 1\}$ ,  $b \in \mathbb{R}_+^*$ ,  $\alpha \in \mathbb{N}$  and  $\varepsilon \in \{0, 1\}$  exist such that  $\varphi$  is QPL at 0, in the sense that*

$$\varphi(x) \underset{0}{\sim} ax^s (\log x)^\delta, \quad (9)$$

and  $h$  is QPL at  $\Phi$ , i.e.,

$$h(z) \underset{\Phi}{\sim} bz^\alpha (\log z)^\varepsilon, \quad (10)$$

where

$$\Phi = \begin{cases} 0 & \text{if } s > 1, \\ a & \text{if } s = 1, \delta = 0, \\ \mathcal{S}(a)\infty & \text{if } s < 1, \delta = 0, \\ -\mathcal{S}(a)\infty & \text{if } s \leq 1, \delta = 1, \end{cases} \quad (11)$$

with  $\mathcal{S}(a)$  denoting the sign of  $a$ .

Of course, (10) makes sense only if  $z^\alpha(\log z)^\epsilon$  is well defined in a neighborhood of  $\Phi$ , inducing the following restrictions on the parameters:

$$\begin{aligned} (\delta = 0) &\Rightarrow (a > 0) \text{ or } (\alpha \in \mathbb{Z}, \epsilon = 0), \\ (\delta = 1) &\Rightarrow (a < 0) \text{ or } (\alpha \in \mathbb{Z}, \epsilon = 0), \end{aligned} \quad (12)$$

that will be supposed to hold thereafter.

Table 2 shows various QPL entropy functionals that are of theoretical or practical interest in the literature. The most classical ones are Shannon entropy  $\mathbb{S}(m) = \mathbb{S}_{h(z), \varphi(x)}(m)$ , with  $h(z) = z$  and  $\varphi(x) = -x \log x$ , and Rényi entropy with  $h(z) = (1 - \mathfrak{q})^{-1} \log z$  and  $\varphi(x) = x^\mathfrak{q}$ , where  $\mathfrak{q} > 0$ ,  $\mathfrak{q} \neq 1$ . The subclass for which  $h$  is the identity includes the well-known  $\varphi$ -entropy functionals – also called Csiszár's. Note that for  $\varphi(x) = ax$ , the  $\varphi$ -entropy collapses to a constant  $\mathbb{S}_{z, \varphi(x)}(m) = a$ , for all  $m$ ; this degenerate case is excluded thereafter.

### 3 Rates as functions of Perron-Frobenius eigenvalues

This section is devoted to extending the closed form expression (4) from Shannon entropy to other QLD entropy functionals. This presentation will allow us to obtain a classification of QPL entropy functionals in Section 4 and will lead to weighted expressions extending (3) in Section 5.

First of all, although (4) is standard in analytic combinatorics for finite Markov sources, no original reference seems to be available for its proof. Since this formula is of paramount importance in the following, we will now detail a proof, relying on the asymptotic behavior of the Dirichlet series associated to  $\mathbf{X}$ ,

$$A_n(s) = \sum_{i_0^n \in E^n} [\mathbf{m}(i_0^n)]^s, \quad \text{for } s \in \mathbb{R}. \quad (13)$$

Obviously,  $\mathbb{S}(X_0^n) = -A'_n(1)$ , so that  $\mathbb{H}(\mathbf{X}) = \lim_{n \rightarrow \infty} -A'_n(1)/n$ . We can write

$$A_n(s) = {}^t \mu^s \cdot P(s)^{n-1} \cdot \mathbf{1}, \quad (14)$$

where  $\mu = (\mu_i)_{i \in E}$  denotes the initial distribution of the chain,  $\mu^s = (\mu_i^s)_{i \in E}$  its perturbation,  $P(s)$  is given by (5) and  $\mathbf{1}$  is the unit vector – with all components equal to 1. Note that all vectors are column vectors, and that the dot stands for matrix multiplication.

If  $\mathbf{X}$  is ergodic and finite, then  $P(s)$  is positive, irreducible and aperiodic, and hence the Perron-Frobenius theorem applies; see e.g., [20]. A unique positive eigenvalue  $\lambda(s)$  exists – called Perron-Frobenius or dominant eigenvalue, with a spectral gap, i.e., a positive difference in modulus between the two largest eigenvalues; precisely

$$\lambda(s) - \sup\{|\lambda| : \lambda \text{ eigenvalue of } P(s), \lambda \neq \lambda(s)\} > 0. \quad (15)$$

Moreover, left and right positive eigenvectors  $u(s)$  and  $v(s)$  for  $P(s)$  associated to  $\lambda(s)$  exist – the so-called Perron-Frobenius eigenvectors of  $P(s)$ , such that  ${}^t u(s).P(s) = \lambda(s){}^t u(s)$  and  $P(s).v(s) = \lambda(s)v(s)$ .

Hence, the  $n$ -th power of  $P(s)$  can be written

$$P(s)^n = \lambda(s)^n C(s) + R_n(s), \quad (16)$$

where  $C(s) = v(s).{}^t u(s)$  is the matrix (in the canonical basis) of the Perron-Frobenius projector – the projector on the linear subspace spanned by  $v(s)$ .

The spectral gap induces that

$$\rho(s) = \frac{1}{\lambda(s)} \sup\{|\lambda| : \lambda \text{ eigenvalue of } P(s), \lambda \neq \lambda(s)\} \in (0, 1).$$

It also induces that the subordinated (with respect to the  $\ell_1$ -norm) matrix norm of the matrix remainder  $R_n(s)$  is of order  $O([\rho\lambda(s)]^n)$  as  $n$  tends to infinity, for all  $\rho$  satisfying  $\rho(s) < \rho < 1$ , where  $\ell_1 = \{u = (u_i)_{i \in E} : \sum_{i \in E} |u_i| < \infty\}$ . In other words,

$$A_n(s) = \lambda(s)^{n-1} c(s) + O([\rho\lambda(s)]^{n-1}), \quad (17)$$

where

$$c(s) = {}^t \mu^s . v(s) . {}^t u(s) . \mathbf{1} = {}^t \mu^s . C(s) . \mathbf{1}. \quad (18)$$

Thus, differentiating (14) with respect to  $s$  at  $s = 1$  is justified, and yields  $A'_n(1) \sim (n-1)\lambda'(1)c(1)$ . Finally, observing that  $P(1) = P$  is a stochastic matrix, so that  $v(1) = \mathbf{1}$  and  $c(1) = 1$ , leads to (3).

A corner stone for the results to follow is (17), known as the quasi-power property (QPP); see [31], [5] and [12]. For denumerable Markov chains, a Perron-Frobenius eigenvalue  $\lambda(s)$  for  $P(s)$  again generally exists. Still, a sequence of other eigenvalues converging to  $\lambda(s)$  may hinder the existence of a spectral gap. The following sufficient conditions for the QPP to hold are easy to check. Determining optimal conditions is still an open question.

**Assumptions 1** *Let  $\mathbf{X}$  be an ergodic Markov chain with finite or denumerable state space  $E$ , transition matrix  $P = (p_{ij})_{(i,j) \in E^2}$ , and initial distribution  $\mu = (\mu_i)_{i \in E}$ .*

*C1.  $\sup_{(i,j) \in E^2} p_{ij} < 1$ ;*

*C2. some  $\sigma_0 < 1$  exists such that for all  $s > \sigma_0$ ,  $\sup_{i \in E} \sum_{j \in E} p_{ij}^s < \infty$  and  $\sum_{i \in E} \mu_i^s < \infty$ ;*

*C3. for all  $\varepsilon > 0$  and all  $s > \sigma_0$ , some  $A \subset E$  exists, with a finite number of elements, such that  $\sup_{i \in E} \sum_{j \in E \setminus A} p_{ij}^s < \varepsilon$ .*

If  $E$  is finite, then both Conditions C2 and C3 are satisfied with  $\sigma_0 = -\infty$ , thus leading to (17). In the denumerable case, necessarily  $\sigma_0 > 0$ . Condition C1 ensures mixing between the states of the chain, hindering that a few very probable paths may asymptotically capture the chain. Condition C3 forbids the probability mass to escape to infinity, and hence the asymptotic distribution of the chain to be a degenerate Dirac distribution at  $+\infty$ .

Thanks to C2,  $P(s)$  induces an operator on  $\ell_1$ : if  $u \in \ell_1$ , then  ${}^t u.P(s) \in \ell_1$  too. All together, Assumptions 1 ensure that it is a compact operator, and hence are sufficient to induce both a spectral gap and the QPP. This is precisely stated in the next proposition, whose proof is a straightforward consequence of [5, Lemma 1 and Theorem 2].

**Proposition 1** *Let  $\mathbf{X}$  be an ergodic Markov chain satisfying Assumptions 1. Then the perturbed matrix  $P(s)$  defined by (5) has a unique positive dominant eigenvalue  $\lambda(s)$ , with a spectral gap (15). Moreover, positive left and right eigenvectors  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$  exist satisfying*

$${}^t\mathbf{u}(s).P(s) = \lambda(s){}^t\mathbf{u}(s) \quad \text{and} \quad P(s).\mathbf{v}(s) = \lambda(s)\mathbf{v}(s).$$

These Perron-Frobenius eigenvectors can be chosen such that

$$\sum_{i \in E} u_i(s) = 1 \quad \text{and} \quad \sum_{i \in E} u_i(s)v_i(s) = 1. \quad (19)$$

Finally, the QPP (17) holds true for all real number  $s > \sigma_0$  and all integer  $n \geq 1$ .

The next lemma – see [31, Proposition 8] – will be of use first for the proof of Theorem 1 below, then for making explicit entropy rates associated to classical entropy functionals, and finally for their classification in Section 4. For clarity of statement, the eigenvalue  $\lambda(s)$  is here fully denoted by  $\lambda(P, s)$ .

**Lemma 1** *For any transition matrix  $P$  satisfying Assumptions 1, the dominant eigenvalue function  $s \mapsto \lambda(P, s)$  is differentiable, with  $\lambda'(P, s) < 0$  for all  $s > \sigma_0$ . In particular, it is a decreasing function and  $\lambda(P, 1) = 1$ .*

The next theorem makes explicit both the averaging sequences and the ensuing expressions of the related entropy rates as functions of the dominant eigenvalue of the perturbed transition matrix and initial distribution of the chain. Even though it is formally a special case of the divergence rates determined in [12], no general expressions are developed therein for Markov chains, and only a few examples of classical entropy functionals are presented for generic sequences of random variables satisfying the QPP. Note that the averaging sequence is defined up to the asymptotic equivalence  $r'_n \sim_{\infty} r_n$ ; also,  $\mathbb{H}_{h(z), \varphi(x), k\mathbf{r}}(\mathbf{X}) = \frac{1}{k} \mathbb{H}_{h(z), \varphi(x), \mathbf{r}}(\mathbf{X})$  for all constant  $k \neq 0$ .

**Theorem 1** *Let  $\mathbb{S}_{h(z), \varphi(x)}$  be a QPL entropy. Let  $\mathbf{X}$  be an ergodic Markov chain satisfying Assumptions 1. The entropy rate  $\mathbb{H}_{h(z), \varphi(x), \mathbf{r}}(\mathbf{X})$  given by (8) is well defined and non degenerate (neither zero nor infinite) for the averaging sequence*

$$r_n = \begin{cases} (n-1)^{\delta\alpha + \epsilon} \lambda(s)^{\alpha(n-1)} & \text{if } s \neq 1, \\ (n-1)^{\delta\alpha} (\log(n-1))^{\delta\epsilon} & \text{if } s = 1. \end{cases} \quad (20)$$

Precisely

$$\mathbb{H}_{h(z), \varphi(x), \mathbf{r}}(\mathbf{X}) = \begin{cases} b a^\alpha c(s)^\alpha [\log \lambda(s)]^\epsilon & \text{if } s \neq 1, \delta = 0, \\ b c(s)^\alpha [a \lambda'(s) \lambda(s)^{-1}]^\alpha [\log \lambda(s)]^\epsilon & \text{if } s \neq 1, \delta = 1, \\ b a^\alpha [\log a]^\epsilon & \text{if } s = 1, \delta = 0, \\ b [a \lambda'(1)]^\alpha & \text{if } s = 1, \delta = 1. \end{cases} \quad (21)$$

	$s < 1$	$s = 1$	$s > 1$	
$\delta = 0$	$\mathcal{S}(a)\infty$ $\lambda(s)^{n-1}$ $ac(s)$	$a$ $1$ $a$	$0$ $\lambda(s)^{n-1}$ $ac(s)$	$\Phi$ $r_n$ $\mathbb{H}_{z,\varphi(x),\mathbf{r}}(\mathbf{X})$
$\delta = 1$	$-\mathcal{S}(a)\infty$ $(n-1)\lambda(s)^{n-1}$ $ac(s)\frac{\lambda'(s)}{\lambda(s)}$	$-\mathcal{S}(a)\infty$ $n-1$ $a\lambda'(1)$	$0$ $(n-1)\lambda(s)^{n-1}$ $ac(s)\frac{\lambda'(s)}{\lambda(s)}$	$\Phi$ $r_n$ $\mathbb{H}_{z,\varphi(x),\mathbf{r}}(\mathbf{X})$

**Table 1** Asymptotic value  $\Phi = \lim_n \mathbb{S}_{z,\varphi(x)}(X_0^n)$  of the  $\varphi$ -entropy, averaging sequence  $\mathbf{r} = (r_n)_{n \in \mathbb{N}}$ , and  $\varphi$ -entropy rate  $\mathbb{H}_{z,\varphi(x),\mathbf{r}}(\mathbf{X})$ , according to the values of  $a$ ,  $s$ ,  $\delta$ , and sign  $\mathcal{S}(a)$  of  $a$ .

*Proof* Thanks to Assumptions 1, for any sequence  $i_0^n$ , the probability  $\mathbf{m}(i_0^n)$  converges to 0 as  $n$  tends to infinity. Since, by (9),  $\varphi$  is QPL at 0, and thanks to the QPP (17),

$$\begin{aligned} \mathbb{S}_{z,\varphi(x)}(\mathbf{X}_0^n) &= \sum_{i_0^n \in E^n} \varphi(\mathbf{m}(i_0^n)) \\ &\sim_{\infty} \begin{cases} aA_n(s) \sim_{\infty} \lambda(s)^{n-1} ac(s) & \text{if } \delta = 0, \\ aA'_n(s) \sim_{\infty} (n-1)\lambda(s)^{n-1} a \frac{\lambda'(s)}{\lambda(s)} c(s) & \text{if } \delta = 1. \end{cases} \end{aligned}$$

Together with Lemma 1, the above equivalence implies that  $\mathbb{S}_{z,\varphi(x)}(X_0^n)$  converges, as  $n$  goes to infinity, to  $\Phi$  given by (11). Note that the limit does not depend on the dynamics of  $\mathbf{X}$ , but only on the parameters  $a$ ,  $s$  and  $\delta$ . For instance, if  $\delta = 0$  and  $s < 1$ , then  $\lambda(s) > 1$ , so that  $\mathbb{S}_{z,\varphi(x)}(X_0^n) = ac(s)\lambda(s)^{n-1}$  converges to  $\pm\infty$  according to the sign of  $a$ .

For QPL entropy functionals,  $h$  satisfies (10) and hence, for  $s \neq 1$ ,

$$\mathbb{S}_{h(z),\varphi(x)}(\mathbf{X}_0^n) \sim_{\infty} \begin{cases} \lambda(s)^{(n-1)\alpha} b a^\alpha c(s)^\alpha & \text{if } \delta = 0, \varepsilon = 0, \\ (n-1)\lambda(s)^{(n-1)\alpha} b a^\alpha c(s)^\alpha \log \lambda(s) & \text{if } \delta = 0, \varepsilon = 1, \\ (n-1)^\alpha \lambda(s)^{(n-1)\alpha} b c(s)^\alpha \left[ a \frac{\lambda'(s)}{\lambda(s)} \right]^\alpha & \text{if } \delta = 1, \varepsilon = 0, \\ (n-1)^{\alpha+1} \lambda(s)^{(n-1)\alpha} \\ \quad \times b c(s)^\alpha \left[ a \frac{\lambda'(s)}{\lambda(s)} \right]^\alpha \log \lambda(s) & \text{if } \delta = 1, \varepsilon = 1, \end{cases}$$

and for  $s = 1$ ,

$$\mathbb{S}_{h(z),\varphi(x)}(\mathbf{X}_0^n) \sim_{\infty} \begin{cases} b a^\alpha & \text{if } \delta = 0, \varepsilon = 0, \\ b a^\alpha \log a & \text{if } \delta = 0, \varepsilon = 1, \\ (n-1)^\alpha b [a\lambda'(1)]^\alpha & \text{if } \delta = 1, \varepsilon = 0, \\ (n-1)^\alpha \log(n-1) b [a\lambda'(1)]^\alpha & \text{if } \delta = 1, \varepsilon = 1. \end{cases}$$

The expression of  $\mathbb{H}_{h(z),\varphi(x),\mathbf{r}}(\mathbf{X})$  immediately follows by taking the limit, for  $\mathbf{r}$  given by (20).

For the special case of  $\varphi$ -entropy, (20) reduces to

$$r_n = \lambda(s)^{n-1} (n-1)^\alpha, \quad (22)$$

and hence the  $\varphi$ -entropy rate

$$\mathbb{H}_{z,\varphi(x),\mathbf{r}}(\mathbf{X}) = \lim_{n \rightarrow \infty} \frac{1}{r_n} \mathbb{S}_{z,\varphi(x)}(\mathbf{X}_0^n) = a c(s) \left[ \frac{\lambda'(s)}{\lambda(s)} \right]^\delta$$

is well-defined and non degenerate. See Table 1 for closed-form expressions and details.

Further, Table 2 shows with full details the expressions of the entropy rates associated to the most classical entropy functionals.

#### 4 Classification of QPL entropy functionals

A classification of all QPL  $(h, \varphi)$ -entropy functionals into five exclusive classes derives from Lemma 1 and Theorem 1, according to the asymptotic behavior of the marginal entropy  $\mathbb{S}_{h(z),\varphi(x)}(\mathbf{X}_0^n)$ , or, equivalently, to the type of averaging sequence  $\mathbf{r}$ . The important point is that this classification depends only on the functions  $\varphi$  and  $h$  – through the parameters  $[s, \delta, \alpha, \varepsilon]$  involved in the QPL properties (9) and (10), and not on the specific dynamics of the chain – as soon as the latter satisfies the QPP. In particular, the classification is valid for all finite Markov chains.

The classification of  $\varphi$ -entropy functionals into four classes easily derives from either Table 1 or Equation (22). Indeed, four types of behavior for the sequence of marginal entropy (or, equivalently, rescaling sequence  $\mathbf{r}$ ), are possible, according to the values of  $s$  and  $\delta$  and independently of the transition matrix of the chain. First, for  $s < 1, \delta = 0$  and  $s < 1, \delta = 1$ , the rescaling sequences are respectively  $\lambda(s)^{n-1}$  and  $(n-1)\lambda(s)^{n-1}$ . Lemma 1 states that  $\lambda(s) > 1$  as soon as  $s < 1$ ; the marginal entropy explodes exponentially fast up to  $\Phi = \lim_n \mathbb{S}_{z,\varphi(x)}(X_0^n)$ . Second, for  $s > 1$  (and  $\delta = 0$  or  $1$ ),  $r_n$  and the marginal entropy decrease (in absolute value) exponentially fast to 0. Third, for  $s = 1$  and  $\delta = 0$ , the marginal entropy converges to some finite value, and the rescaling sequence is constant. Fourth, for  $s = 1$  and  $\delta = 1$  (Shannon entropy), the marginal entropy  $\mathbb{S}_{h(z),\varphi(x)}(\mathbf{X}_0^n)$  increases linearly with  $n$  and the rescaling sequence is the time averaging  $r_n = n$ .

The above discussion easily extends to  $(h, \varphi)$ -entropy functionals through the asymptotic behavior of marginal entropies or of the rescaling sequence  $r_n$  given by (20), as made explicit in Table 3. A classification of  $(h, \varphi)$ -entropy functionals ensues, including the four previous classes plus a fifth one. Let us begin by the so-called linear case, associated to classical averaging per time unit.

**Linear case:**  $r_n = kn$ , with  $k \in \mathbb{R}_+$ .

The marginal entropy increases linearly with the number of variables of  $\mathbf{X}_0^n$ . Intuitively, all variables contribute equally to the global information of the chain. This is obtained for  $[\delta = 1, s = 1, \alpha = 1, \varepsilon = 0]$ ,  $[\delta = 0, s \neq 1, \alpha = 0, \varepsilon = 1]$  and also  $[\delta = 1, s \neq 1, \alpha = 0, \varepsilon = 1]$ . The two first sets of parameters correspond respectively to Shannon, Rényi and their two-parameter extensions of [32]:

$$\mathbb{V}_{\mathbf{p},\mathbf{q}}^1(m) = \frac{1}{\mathbf{p} - \mathbf{q}} \log \left( \sum_{i \in E} m_i^{\mathbf{q} - \mathbf{p} + 1} \right), \quad \mathbb{V}_{\mathbf{p},\mathbf{q}}^2(m) = \frac{1}{\mathbf{p}(\mathbf{p} - \mathbf{q})} \log \left( \sum_{i \in E} m_i^{\mathbf{q}/\mathbf{p}} \right),$$

for  $\mathbf{p}, \mathbf{q} > 0$ , with  $\mathbf{q} \neq \mathbf{p}$ .

**Over-linear (or expanding) case:**  $r_n \in \omega(n)$ .

The marginal entropy increases faster than  $n$ , typically exponentially fast. This is obtained for  $[\delta = 0, s > 1, \alpha < 0]$ ,  $[\delta = 0, s < 1, \alpha > 0]$ ,  $[\delta = 1, s < 1, \alpha > 0]$ ,  $[\delta = 1, s > 1, \alpha < 0]$ ,  $[\delta = 1, s = 1, \alpha = 1, \varepsilon = 1]$ , and  $[\delta = 1, s = 1, \alpha > 1]$ . Main examples are Sharma-Mittal and Sharma-Taneja for  $p < 1$ , Arimoto for  $q > 1$ , and Taneja for  $q < 1$ . The marginal entropy explodes with the number of variables. This explosive behavior is counter-balanced by a suitable averaging. The resulting entropy rate focuses essentially on the initial behavior of the chain, and hence depends mainly on the initial distribution.

**Sub-linear case:**  $r_n = o(n)$  and  $r_n \in \omega(1)$ .

The marginal entropy increases slower than  $n$ , typically as  $n^\alpha$  or  $n^\alpha \log n$ , with  $0 < \alpha < 1$ . This is obtained for  $[\delta = 1, s = 1, \alpha = 0, \varepsilon = 1]$  and  $[\delta = 1, s = 1, 0 < \alpha < 1]$ . The resulting rescaled entropy rate is either constant – does not depend on the chain dynamics, or equals Shannon up to the power  $\alpha$  and a multiplicative constant. Interestingly, none of the classical entropy functionals behaves in this way.

**Contracting case:**  $r_n = o(1)$ .

The marginal entropy decreases to 0, typically exponentially fast. This behavior is obtained for  $[\delta = 0, s < 1, \alpha < 0]$ ,  $[\delta = 0, s > 1, \alpha > 0]$ ,  $[\delta = 1, s \leq 1, \alpha < 0]$ , and  $[\delta = 1, s > 1, \alpha > 0]$ . Main examples are Sharma-Taneja for  $p > 1$  and Taneja for  $q > 1$ . The marginal entropy decreases with the considered number of variables. Again, the resulting rescaled entropy rate depends mainly on the initial distribution of the chain.

**Constant case:**  $|r_n| = k \in \mathbb{R}_+$ .

The marginal entropy converges to a non degenerate value. This is obtained when either  $\alpha = 0$  and  $\varepsilon = 0 - h$  is constant in a neighborhood of  $\Phi$ , or  $\delta = 0$  and  $s = 1$  – so that  $\varphi(x) \sim_0 a$ . Classical examples are Arimoto for  $q < 1$ , Sharma-Mittal for  $p > 1$ , Ferreri and Tsallis for  $q > 1$ . The resulting entropy rates are degenerate, in the sense that they take the same value for all Markov chains satisfying the QPP.

## 5 Rates as functions of Perron-Frobenius eigenvectors

Some relevant properties and relations between the quantities involved in the expression of  $\mathbb{H}_{h(z), \varphi(x), r}(\mathbf{X})$  in (21) will allow us to obtain weighted expressions extending (3) and (4) to all QPL entropy rates of Markov chains satisfying the QPP. This will be mostly based on the following computational results.

**Lemma 2** *Let  $\mathbf{X}$  be an ergodic Markov chain satisfying Assumptions 1. Let  $s > \sigma_0$ . The following relations hold true:*

$$\lambda(s) = \sum_{i \in E} u_i(s) \sum_{j \in E} p_{ij}^s, \quad (23)$$

$$\lambda'(s) = \sum_{i \in E} u_i(s) \sum_{j \in E} p_{ij}^s \log p_{ij} + \sum_{i \in E} u'_i(s) \sum_{j \in E} p_{ij}^s. \quad (24)$$

$$c(s) = \sum_{i \in E} \mu_i^s v_i(s). \quad (25)$$

*Proof* We compute  ${}^t\mathbf{u}(s).P(s).\mathbf{1} = \lambda(s){}^t\mathbf{u}(s).\mathbf{1} = \lambda(s)$ , and (23) follows straightforwardly. Differentiating with respect to  $s$  gives (24).

Thanks to the QPP,  $c(s)$  is given by (18). The normalizing condition in (19) yields  ${}^t\mathbf{u}(s).\mathbf{1} = 1$ , and hence (25).

All entropy rates obtained in Theorem 1 can thus be written as explicit functions of the perturbed transition matrix  $P(s)$  and Perron-Frobenius eigenvectors  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$ .

**Theorem 2** *Let  $\mathbf{X}$  be an ergodic Markov chain satisfying Assumptions 1. For the averaging sequence  $\mathbf{r}$  defined by (20), the  $(h, \varphi)$ -entropy rate is*

$$\mathbb{H}_{h(z), \varphi(x), \mathbf{r}}(\mathbf{X}) = \begin{cases} b a^\alpha \left[ \sum_{i \in E} \mu_i^s \mathbf{v}_i(s) \right]^\alpha \left[ \frac{\sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s \log p_{ij} + \sum_{i \in E} \mathbf{u}'_i(s) \sum_{j \in E} p_{ij}^s}{\sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s} \right]^\alpha \\ \quad \times \left[ \log \sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s \right]^\varepsilon, & \text{if } s \neq 1, \delta = 1, \\ b a^\alpha \left[ \sum_{i \in E} \mu_i^s \mathbf{v}_i(s) \right]^\alpha \left[ \log \sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s \right]^\varepsilon, & \text{if } s \neq 1, \delta = 0, \\ b a^\alpha \left[ \sum_{i \in E} \mathbf{u}_i(1) \sum_{j \in E} p_{ij} \log p_{ij} \right]^\alpha, & \text{if } s = 1, \delta = 1, \end{cases} \quad (26)$$

*Proof* Plugging the expressions of  $c(s)$ ,  $\lambda(s)$  and  $\lambda'(s)$  of Lemma 2 into (21) readily yields the result.

In particular, the  $\varphi$ -entropy rates (for  $b = \alpha = 1$  and  $\varepsilon = 0$ ) are given by

$$\mathbb{H}_{h(z), \varphi(x), \mathbf{r}}(\mathbf{X}) = a \left[ \frac{\sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s \log p_{ij} + \sum_{i \in E} \mathbf{u}'_i(s) \sum_{j \in E} p_{ij}^s}{\sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s} \right] \sum_{i \in E} \mu_i^s \mathbf{v}_i(s),$$

for  $s \neq 1, \delta = 1$ . Note that the above ratio is the logarithmic derivative of the weighted sum  $\sum_{i \in E} \mathbf{u}_i(s) \sum_{j \in E} p_{ij}^s$ .

For Shannon entropy,  $b = \alpha = s = \delta = 1$ , and (23) and (25) for  $s = 1$  amount to  $\lambda(1) = 1$  and  $c(1) = 1$ , yielding (3). For Rényi entropy, setting  $a = 1$ ,  $s = q$ ,  $\delta = 0$ ,  $b = (1 - q)^{-1}$ ,  $\alpha = 0$ ,  $\varepsilon = 1$ , yields (7). Clearly, both (3) and (7) involve the Perron-Frobenius eigenvectors too, and are simple particular cases of (26). Therefore, the above extension to the generalized entropy functionals of the well-known weighted expression for Shannon entropy rate opens the way to extension of the associated toolbox, including maximum entropy methods; see the introduction above for details. The interpretation in terms of dynamics of an auxiliary absorbing Markov chain in the next section is also to take its part in connection to this toolbox.

## 6 Dynamical interpretation of all spectral elements and entropy rates

The closed form expression (3) for Shannon entropy has a meaningful dynamical interpretation as the average of entropy of transition probabilities weighted by the stationary distribution of the chain. The expressions of entropy rates as weighted sums obtained in Section 5, similar to (3) in their form, also have a dynamical interpretation: the eigenvectors  $\mathbf{u}(s)$  and  $\mathbf{v}(s)$  will be proven in Section 6.1 to be the limit distributions of sequences of escort distributions of the chain. In

Section 6.2, these escorts are linked with the asymptotic behavior of some auxiliary absorbing Markov chain associated to  $\mathbf{X}$ , conditionally to non absorption. Relevant interpretations of the spectral elements  $u(s)$ ,  $v(s)$  and  $\lambda(s)$ , as well as  $c(s)$ , ensue in terms of dynamics of this auxiliary chain.

### 6.1 Eigenvectors as limits of marginals of escorts distributions

The escort distribution – in short escort – with parameter  $s$  of any probability distribution  $m = (m_i)_{i \in E}$  is the distribution  $m^{*s}$  given by

$$m_i^{*s} = \frac{m_i^s}{\sum_{j \in E} m_j^s}, \quad i \in E, \quad (27)$$

when the sum  $\sum_{j \in E} m_j^s$  is finite. Escort distributions, first introduced in [3], provide a tool for zooming at different parts of a distribution, or for adapting it to constraints, through their ability to scan its structure. They also constitute the geodesics of information geometry with respect to an affine connection naturally induced by the Kullback-Leibler divergence; see [1], [28], and also [13] for an application to hypothesis testing. In Section 6.2, they are shown to describe the behavior, conditionally to non-absorption, of some absorbing Markov chain.

Little attention in the literature has been focused at the properties of escorts of random vectors and sequences, but in [24]. We are interested here in the asymptotic behavior of the escorts  $(m_0^n)^{*s}$  associated to the  $n$  first coordinates  $X_0^n$  of an ergodic discrete Markov chain  $\mathbf{X}$ , in terms of the left and right eigenvectors  $u(s)$  and  $v(s)$ . Interesting by itself, this behavior is also to be linked to the asymptotic behavior of some auxiliary absorbing Markov in Section 6.2.

**Proposition 2** *Let  $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$  be an ergodic Markov chain with transition matrix  $P$  and initial distribution  $\mu$  satisfying Assumptions 1.*

*For any  $n \in \mathbb{N}^*$  and  $s > \sigma_0$ , let  $m_0^n$  denote the distribution of  $\mathbf{X}_0^n$  and  $(m_0^n)^{*s}$  its escort distribution defined in (27). Then, as  $n$  tends to infinity:*

1. *the first marginal of  $(m_0^n)^{*s}$  converges to  $\mu^s v(s)/c(s)$ , where  $\mu^s v(s)$  denotes the component by component product of the vectors  $\mu^s$  and  $v(s)$ .*
2. *the  $n$ -th marginal of  $(m_0^n)^{*s}$  converges to  $u(s)$ ;*

*Proof 1.* Thanks to Point 2 of Assumptions 1, the escort  $(m_0^n)^{*s}$  of  $m_0^n$  is well-defined for  $s > \sigma_0$  and is given by

$$(m_0^n)^{*s}(i_0^n) = \frac{m_0^n(i_0^n)^s}{\Lambda_n(s)},$$

where  $\Lambda_n(s)$  is the Dirichlet series defined in (13).

Let  $\text{mar}_1((m_0^n)^{*s})$  and  $\text{mar}_n((m_0^n)^{*s})$  denote respectively the first and  $n$ -th marginal distribution of  $(m_0^n)^{*s}$ , given by

$$\begin{aligned} \text{mar}_1((m_0^n)^{*s})(i) &= \frac{1}{\Lambda_n(s)} \sum_{i_1^n \in E^{n-1}} m_0^n(i, i_1^n), \\ \text{mar}_n((m_0^n)^{*s})(i) &= \frac{1}{\Lambda_n(s)} \sum_{i_0^{n-1} \in E^{n-1}} m_0^n(i_0^{n-1}, i), \end{aligned}$$

where  $(i, i_1^n) = (i, i_1, \dots, i_{n-1})$  and  $(i_0^{n-1}, i) = (i_0, \dots, i_{n-2}, i)$ . Straightforward linear algebra gives

$$\sum_{i_1^n \in E^{n-1}} m_0^n(i, i_1^n) = \mu_i^s (P(s)^{n-1} \cdot \mathbf{1})_i \quad \text{and} \quad \sum_{i_0^{n-1} \in E^{n-1}} m_0^n(i_0^{n-1}, i) = \left( {}^t \mu^s \cdot P(s)^{n-1} \right)_i,$$

where  $(w)_i$  denotes here the  $i$ -th component of the vector  $w$ . The above equalities together with (14) yield

$$\text{mar}_1((m_0^n)^{*s})(i) = \frac{\mu_i^s (P(s)^{n-1} \cdot \mathbf{1})}{{}^t \mu^s \cdot P(s)^{n-1} \cdot \mathbf{1}} \quad \text{and} \quad {}^t \text{mar}_n((m_0^n)^{*s}) = \frac{{}^t \mu^s \cdot P(s)^{n-1}}{{}^t \mu^s \cdot P(s)^{n-1} \cdot \mathbf{1}},$$

where  $\mu^s (P(s)^{n-1} \cdot \mathbf{1})$  denotes the component by component product of the vectors  $\mu^s$  and  $P(s)^{n-1} \cdot \mathbf{1}$ . From (16), we derive

$$P(s)^{n-1} \cdot \mathbf{1} = \lambda(s)^{n-1} C(s) \cdot \mathbf{1} + R_{n-1}(s) \cdot \mathbf{1} = \lambda(s)^{n-1} \mathbf{v}(s) + R_{n-1}(s) \cdot \mathbf{1},$$

and

$$\begin{aligned} {}^t \mu^s \cdot P(s)^{n-1} &= \lambda(s)^{n-1} {}^t \mu^s \cdot \mathbf{v}(s) \cdot {}^t \mathbf{u}(s) + {}^t \mu^s \cdot R_{n-1}(s) \\ &= \lambda(s)^{n-1} c(s) \cdot {}^t \mathbf{u}(s) + {}^t \mu^s \cdot R_{n-1}(s), \end{aligned}$$

where, thanks to the spectral gap (15),  $\|R_{n-1}(s) \cdot \mathbf{1}\|_1$  and  $\|{}^t \mu^s \cdot R_{n-1}(s)\|_1$  are  $O((\rho(s)\lambda(s))^{n-1})$ . The above relations together with the QPP imply that

$$\text{mar}_1((m_0^n)^{*s}) = \frac{\lambda(s)^{n-1} \mu^s \mathbf{v}(s) + R_{n-1}(s) \cdot \mathbf{1}}{\lambda(s)^{n-1} c(s) + O((\rho(s)\lambda(s))^{n-1})} = \frac{\mu^s \mathbf{v}(s)}{c(s)} + o_{\|\cdot\|_1}(1),$$

and

$${}^t \text{mar}_n((m_0^n)^{*s}) = \frac{\lambda(s)^{n-1} c(s) \cdot {}^t \mathbf{u}(s) + {}^t \mu^s \cdot R_{n-1}(s)}{\lambda(s)^{n-1} c(s) + O((\rho(s)\lambda(s))^{n-1})} = {}^t \mathbf{u}(s) + o_{\|\cdot\|_1}(1),$$

from which both points 1. and 2. immediately follow.

## 6.2 Rates through quasi-limit distributions and asymptotic odds

An interesting interpretation of the generalized entropy rates derives from the notion of quasi-limit distribution (QLD) linked to the dynamics of an auxiliary absorbing Markov chain. Specific concepts and results have first to be recalled for full understanding.

As previously,  $\mathbf{X}$  denotes an ergodic Markov chain on the finite or denumerable set  $E$ , with transition matrix  $P$  and initial distribution  $\mu$  satisfying Assumptions 1. The transition matrix of the auxiliary chain is to be built from a sub-stochastic matrix. If  $s > 1$ , the matrix  $P(s)$  associated is indeed sub-stochastic: all its coefficients are positive or null with all row sums less or equal to 1. If  $s < 1$ , consider the matrix  $\widetilde{P}(s) = \frac{1}{\|P(s)\|} P(s)$ , for some matrix-norm  $\|P(s)\|$  of  $P(s)$ ; possible choices are  $\|P(s)\| = |E| \max_{i,j \in E} p_{ij}^s$ ,  $\|P(s)\| = \sum_{i,j \in E} p_{ij}^s$ , or  $\|P(s)\| = \max_{i \in E} (P(s) \cdot \mathbf{1})_i$ . The matrix  $\widetilde{P}(s)$  is sub-stochastic and has the same Perron-Frobenius eigenvectors as  $P(s)$ ; see [16]. Then  $\widetilde{P}(s)$  can replace  $P(s)$  in the following.

Let us add to  $E$  an absorbing state  $\diamond$  – usually called coffin state – in order to build a stochastic matrix  $T(s)$  from  $P(s)$  – or  $\overline{P}(s)$ , specifically

$$T(s) = \left( \begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline q(s) & P(s) & & \end{array} \right), \text{ where } q(s) = \begin{pmatrix} 1 - \sum_{j \in E} p_{1j}^s \\ 1 - \sum_{j \in E} p_{2j}^s \\ \vdots \end{pmatrix}.$$

Then, let  $\mathbf{Z} = (Z_n)_{n \in \mathbb{N}}$  be a Markov chain taking values into  $\overline{E} = E \cup \{\diamond\}$ , with transition matrix  $T(s)$ . This auxiliary chain has one absorbing state  $\diamond$  and a transient class  $E$ . Time until absorption in  $\diamond$  may be quite long; see [14] or [18] among many others for examples in biology modeling, population processes, etc.. This phenomenon gave birth to the notion of QLD, also called Yaglom limit; see [27] or [19] for details. A probability distribution on  $E$ , say  $\mathbf{w} = (w_i)_{i \in E}$ , is a QLD of  $\mathbf{Z}$  if for any  $i \in E$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}_i(Z_n = j | Z_n \neq \diamond) = w_j$ , for all  $j \in E$ , where  $\mathbb{P}_i$  denotes as usual conditional probability with respect to  $(Z_0 = i)$ ; note that the QLD does not depend on the initial state  $i$ .

For a finite state space  $E$ , [8] establishes that the QLD is the left Perron-Frobenius eigenvector of  $P(s)$ ; in other words, with the notation of Theorem 1,

$$u_j(s) = \lim_{n \rightarrow \infty} \mathbb{P}_i(Z_n = j | Z_n \neq \diamond), \quad j \in E. \quad (28)$$

For a denumerable  $E$ , absorption in  $\diamond$  may not be certain, in which case (28) would not be true. However, [30] shows that absorption is certain provided that  $P$  satisfies some weak conditions covered by Assumptions 1, and then  $\mathbf{u}(s)$  indeed satisfies (28).

Moreover, the right Perron eigenvector  $\mathbf{v}(s)$  can be written in terms of asymptotic conditional odds; see [30] for details. Indeed, the ratio  $v_j(s)/v_i(s)$  quantifies how much more asymptotically probable it is for  $\mathbf{Z}$  not to be absorbed when starting from  $j$  than  $i$ . In mathematical words,

$$\frac{v_j(s)}{v_i(s)} = \lim_{n \rightarrow \infty} \frac{\mathbb{P}(Z_n \neq \diamond | Z_0 = j)}{\mathbb{P}(Z_n \neq \diamond | Z_0 = i)}, \quad i, j \in E. \quad (29)$$

Interestingly, the behavior of  $\mathbf{Z}$  conditional to non-absorption can be linked to escort distributions as follows. For any  $n \in \mathbb{N}^*$  and any  $i_0^n \in E^n$ ,

$$\mathbb{P}(Z_0^n = i_0^n | Z_{n-1} \neq \diamond) = \frac{\nu_{i_0} \prod_{l=0}^{n-2} p_{i_l i_{l+1}}^s}{\sum_{j_0^n \in E^n} \nu_0(j_0) \prod_{l=0}^{n-2} p_{j_l j_{l+1}}^s},$$

where  $\nu$  is the initial distribution of  $\mathbf{Z}$ . Assume that  $s > 1$  and set  $\nu_i = \mu_i^s$  for all  $i \in E$  and  $\nu_\diamond = 1 - \sum_{j \in E} \mu_j^s$ . Then,

$$\mathbb{P}(Z_0^n = i_0^n | Z_{n-1} \neq \diamond) = (m_{X_0^n})^{*s}(i_0^n).$$

In other words, the distribution of  $Z_0^n$  conditional to non absorption at time  $n - 1$  is the  $s$ -escort distribution of  $X_0^n$ . As a consequence, the distribution of  $Z_{n-1}$  conditional to non absorption is the  $n$ th-marginal of  $(m_{X_0^n})^{*s}$ . Hence, (28) appears as a straightforward consequence of Point 2 of Proposition 2.

Similarly, the distribution of the initial state conditionally to non absorption at time  $n - 1$  is the first marginal of  $(m_{X_0^n})^{*s}$ . Hence, Point 1 of Proposition 2 yields

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_0 = i | Z_{n-1} \neq \diamond) = \lim_{n \rightarrow \infty} \text{mar}_1(m_{X_0^n})^{*s}(i) = \frac{\mu_i^s \nu_i(s)}{c(s)}, \quad i \in E,$$

so that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}(Z_0 = i | Z_{n-1} \neq \diamond)}{\mathbb{P}(Z_0 = j | Z_{n-1} \neq \diamond)} = \frac{\mu_i^s \nu_i(s)}{\mu_j^s \nu_j(s)}, \quad i, j \in E. \quad (30)$$

The odds  $\mu_i^s \nu_i(s) / \mu_j^s \nu_j(s)$ ,  $i, j \in E$ , then quantify how much more asymptotically probable it is for  $\mathbf{Z}$  to have started from one state than from another one, conditionally to non-absorption. Note that a simple use of Bayes' theorem in (30) yields (29).

The assumptions  $s > 1$  and  $\nu_i = \mu_i^s$ ,  $i \in E$  ensure in above results that the distributions of  $Z_0$  and  $Z_{n-1}$  conditional to  $(Z_{n-1} \neq \diamond)$  are respectively given by the first and  $n$ -th marginals of  $m_{X_0^n}^{*s}$ . Further, (28) and (30) can be proven for arbitrary  $s > \sigma_0$  and  $\nu$  by a straightforward extension of Proposition 2, that would lead to the alternative limit  $\nu_i^s \nu_i(s) / \nu_j^s \nu_j(s)$  in (30).

In conclusion, all the factors involved in the weighted expression (21) of the generalized entropy rates get a natural interpretation in terms of dynamics of Markov chains through the auxiliary chain  $\mathbf{Z}$ :

$\lambda(s) = \sum_i u_i(s) \sum_j p_{ij}^s = \lim_{n \rightarrow \infty} \mathbb{P}(Z_{n+1} \neq \diamond | Z_n \neq \diamond)$  is the asymptotic probability for  $\mathbf{Z}$  to transit from  $E$  to  $E$ ;

$\lambda'(s) = \sum_i u_i(s) \sum_j p_{ij}^s \log p_{ij} + \sum_i u_i'(s) \sum_j p_{ij}^s$  is the growth rate of this asymptotic probability;

$c(s) = \sum_i \mu_i^s \nu_i(s)$  is a residual information factor focusing on the initial behavior of  $\mathbf{Z}$ .

## References

1. AMARI, S. I. AND NAGAOKA, H. (2007). *Methods of information geometry* (Vol. 191), American Mathematical Soc.
2. BASSEVILLE, M. (2013). Divergence measures for statistical data processing. *J. Signal Proc.* **93**, 621–633.
3. BECK, C. AND SCHGL, F. (1995). *Thermodynamics of chaotic systems: an introduction* (No. 4), Cambridge University Press.
4. CIUPERCA, G. AND GIRARDIN, V. (2007). Estimation of the entropy rate of a countable Markov chain, *Comm. Stat. Th. Meth.*, **36**, 2543–2557.
5. CIUPERCA, G. AND GIRARDIN, V. AND LHOUE, L. (2011). Computation of generalized entropy rates. Application and estimation for countable Markov chains, *IEEE Trans. Info. Th.* **57**, 4026–4034.
6. COVER, L. AND THOMAS, J. (1991). *Elements of information theory*, Wiley series in telecommunications, New-York.
7. EKROOT, L. AND COVER, T. M. (1993). The entropy of Markov trajectories. *IEEE Trans. Info. Th.* **39**, 1418–1421.
8. DARROCH, J. N. AND E. SENETA, E. (1965). On Quasi-Stationary Distributions in Absorbing Discrete-Time Finite Markov Chains, *J. App. Probab.* **2**, 88–100.
9. GERCHAK, Y. (1981). Maximal entropy of Markov chains with common steady-states probabilities, *J. Oper. Res. Soc.* **32**, 233–234.
10. GIRARDIN, V. (2004). Entropy maximization for Markov and semi-Markov processes, *Meth. Comp. App. Prob.* **6**, 109–127.

11. GIRARDIN, V. (2005). On the Different Extensions of the Ergodic Theorem of Information Theory. In *Recent Advances in Applied Probability*, eds. R. Baeza-Yates, J. Glaz, H. Gzyl, J. Hüsler & J. L. Palacios. Springer-Verlag, San Francisco, pp. 163–179.
12. GIRARDIN, V. AND LHOTE, L. (2015). Rescaling Entropy and Divergence Rates. *IEEE Trans. Info. Th.* **61**, 5868–5882.
13. GIRARDIN, V. AND REGNAULT, P. (2016). Escort distributions minimizing the Kullback-Leibler divergence for a large deviations principle and tests of entropy level. *Ann. Inst. Stat. Math.* **68**, 439–468.
14. GOSSELIN, F. (2001). Asymptotic Behavior of Absorbing Markov Chains Conditional on Nonabsorption for Applications in Conservation Biology. *Adv. App. Prob.* **11**, 261–284.
15. HOHOLDT, T. AND JUSTESEN, J. (1984). Maxentropic Markov chains, *IEEE Trans. Info. Th.* **30**, 665–667.
16. HUILLET, T. (2009). Random walks pertaining to a class of deterministic weighted graphs. *J. Physics A* **42**, article id. 275001.
17. Kafsi, M. and Grossglauser, M. and Thiran, P. (2015). Traveling Salesman in Reverse: Conditional Markov Entropy for Trajectory Segmentation. *IEEE International Conference on Data Mining 2015*, 201–210.
18. LAMBERT, A. (2008). Population dynamics and random genealogies. *Stoch. Models* **24**, 45–163.
19. LEDOUX, J. AND RUBINO, G. AND SERICOLA, B. (1994). Exact aggregation of absorbing Markov processes using quasi-stationary distribution. *J. App. Prob.* **31**, 626–634.
20. MEYER, C. D. (2000). *Matrix analysis and applied linear algebra*, SIAM Philadelphia.
21. PRONZATO, L. AND WYNN, H. P. AND ZHIGLJAVSKY, A. A. (1997). Using Renyi entropies to measure uncertainty in search problems. In *Mathematics of Stochastic Manufacturing Systems: AMS-SIAM Summer Seminar in Applied Mathematics* **33**. Williamsburg, USA, pp. 253–268.
22. RACHED, Z. AND ALAJAJI, F AND CAMPBELL, L. L. (2001). Rényi’s divergence and entropy rates for finite alphabet Markov sources. *IEEE Trans. Info. Th.* **47**, 1553–1561.
23. RÉNYI, A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, The Regents of the University of California.
24. REGNAULT, P. AND GIRARDIN, V. AND LHOTE, L. (2017). Escort distributions and the Rényi entropy rates of Markov chains, *Geometric Science of Information 2017*, Paris.
25. SAERENS, M. AND ACHBANY, Y. AND FOUSS, F. AND YEN, L. (2009). Randomized Shortest-Path Problems: Two Related Models. *Neural Comp.* **21**, 2363–2404.
26. MENÉNDEZ, M.L. AND MORALES, D. AND PARDO, L. AND SALICRÚ, M. (1997).  $(h, \varphi)$ -entropy differential metric, *Appl. Math.* **42**, 81–98.
27. SENETA, E. (2006). *Non-negative matrices and Markov chains*, Springer Series in Statistics No. 21. New York.
28. SGARRO, A. (1978). An informational divergence geometry for stochastic matrices. *Calcolo* **15**, 41–49.
29. SHANNON, C. (1948). A mathematical theory of communication. *Bell Syst. Techn. J.* **27**, 379–423, 623–656.
30. SENETA, E. AND VERE-JONES, D. (1966). On Quasi-Stationary Distributions in Discrete-Time Markov Chains with a DenumerableInfinity of States. *J. App. Prob.* **3**, 403–434.
31. VALLÉE, B. (2001). Dynamical sources in information theory: Fundamental intervals and word prefixes. *Algorithmica* **29**, 262–306.
32. VARMA, R. S. (1966). Generalizations of Rényi’s entropy of order  $\alpha$ . *J. Math. Sc.* **1**, 34–48.

Entropy $\varphi(x), h(z)$	$[a, s, \delta]$	$\Phi$	$[b, \alpha, \varepsilon]$	$h\{\Phi\}$	$r_n$	$\mathbb{H}_{h(z), \varphi(x), \mathbf{r}}$	Type
Shannon $-x \log x, z$	$[-1, 1, 1]$	$+\infty$	$[1, 1, 0]$	$+\infty$	$n - 1$	$-\lambda'(1)$	<b>2</b>
Rényi $x^q, \frac{1}{1-q} \log z$	$[1, q, 0]$	$q > 1: 0$ $q < 1: +\infty$	$[\frac{1}{1-q}, 0, 1]$	$+\infty$	$n - 1$	$\frac{1}{1-q} \log \lambda(q)$	<b>2</b>
Tsallis $x^q, \frac{1}{q-1}(1-z)$	$[1, q, 0]$	$q > 1: 0$ $q < 1: +\infty$	$[\frac{1}{q-1}, 0, 0]$ $[\frac{1}{1-q}, 1, 0]$	$\frac{1}{q-1}$ $+\infty$	$1$ $\lambda(q)^{n-1}$	$\frac{1}{q-1}$ $\frac{1}{1-q} c(q)$	<b>0</b> <b>3</b>
Taneja $-x^q \log x, z$	$[-1, q, 1]$	$q > 1: 0$ $q < 1: +\infty$	$[1, 1, 0]$	$\frac{0}{+\infty}$	$(n-1)\lambda(q)^{n-1}$	$-c(q) \frac{\lambda'(q)}{\lambda(q)}$	<b>-1</b> <b>3</b>
Arimoto $x^{1/q}, \frac{1}{q-1}(z^q - 1)$	$[1, \frac{1}{q}, 0]$	$q > 1: +\infty$ $q < 1: 0$	$[\frac{1}{q-1}, q, 0]$ $[\frac{1}{1-q}, 0, 0]$	$+\infty$ $\frac{1}{1-q}$	$\lambda(\frac{1}{q})^{(n-1)q}$ $1$	$\frac{1}{q-1} c(\frac{1}{q})^q$ $\frac{1}{1-q}$	<b>3</b> <b>0</b>
Sharma-Taneja $x^q - x^p, \frac{z}{p-q}$	$q > p: [-1, p, 0]$ $q < p: [1, q, 0]$	$p > 1: 0$ $p < 1: -\infty$ $p > 1: 0$ $p < 1: +\infty$	$[\frac{1}{p-q}, 1, 0]$	$\frac{0}{+\infty}$ $\frac{0}{+\infty}$	$\lambda(p)^{n-1}$ $\lambda(q)^{n-1}$	$\frac{1}{q-p} c(p)$ $\frac{1}{p-q} c(q)$	<b>-1</b> <b>3</b> <b>-1</b> <b>3</b>
Sharma-Mittal $x^q, \frac{1}{1-p} \left( z^{\frac{p-1}{q-1}} - 1 \right)$	$[1, q, 0]$	$q > 1: 0$ $q < 1: +\infty$	$p > 1: [\frac{1}{p-1}, 0, 0]$ $p < 1: [\frac{1}{1-p}, \frac{p-1}{q-1}, 0]$ $p > 1: [\frac{1}{p-1}, 0, 0]$ $p < 1: [\frac{1}{1-p}, \frac{p-1}{q-1}, 0]$	$\frac{1}{p-1}$ $+\infty$ $\frac{1}{p-1}$ $+\infty$	$1$ $\lambda(q)^{\frac{p-1}{q-1}(n-1)}$ $1$ $\lambda(q)^{\frac{p-1}{q-1}(n-1)}$	$\frac{1}{p-1}$ $\frac{1}{1-p} c(q)^{\frac{p-1}{q-1}}$ $\frac{1}{p-1}$ $\frac{1}{1-p} c(q)^{\frac{p-1}{q-1}}$	<b>0</b> <b>3</b> <b>0</b> <b>3</b>
Ferreri $(1+qx) \log(1+qx),$ $(1+\frac{1}{q}) \log(1+q) - \frac{1}{q}z$	$[q, 1, 0]$	$q$	$[b, 0, 0]$ $b = -1+$ $(1+\frac{1}{q}) \log(1+q)$	$b$	$1$	$b$	<b>0</b>

**Table 2** For classical entropy functionals with  $p, q > 0$ , from left to right: Parameters of (9), Asymptotic  $\varphi$ -Entropy  $\Phi = \lim \mathbb{S}_{z, \varphi(x)}(X_0^n)$ , Parameters of (10), Limit Marginal Entropy  $h\{\Phi\} = \lim_n \mathbb{S}_{h(z), \varphi(x)}(X_0^n)$ , Rescaling Sequence  $\mathbf{r} = (r_n)$ , Entropy Rate  $\mathbb{H}_{h(z), \varphi(x), \mathbf{r}}(\mathbf{X})$ , and Type (-1: contracting, 0: constant, 1: sub-linear, 2: linear, 3: over-linear); see Section 4.

		$\alpha < 0$	$\alpha = 0, \varepsilon = 0$	$\alpha = 0, \varepsilon = 1$	$\alpha > 0$					
$\delta = 0,$	$S(a)\infty$							$\Phi$		
$s < 1$		0	$b$	$S(ba^\alpha)\infty$	$S(b)\infty$		$h\{\Phi\}$			
		$\lambda(s)^{\alpha(n-1)}(n-1)^\varepsilon$	$-1$	$1$	$n-1$	$2$	$\lambda(s)^{\alpha(n-1)}(n-1)^\varepsilon$	$3$	$r_n$	
		$ba^\alpha c(s)^\alpha (\log \lambda(s))^\varepsilon$		$b$	$b \log \lambda(s)$		$ba^\alpha c(s)^\alpha (\log \lambda(s))^\varepsilon$		$\mathbb{H}$	
$\delta = 0,$	$a$							$\Phi$		
$s = 1$		$ba^\alpha (\log a)^\varepsilon$	$0$	$b$	$b \log a$	$0$	$ba^\alpha (\log a)^\varepsilon$	$0$	$h\{\Phi\}$	
		$1$	$0$	$1$	$1$	$0$	$1$	$0$	$r_n$	
		$ba^\alpha (\log a)^\varepsilon$		$b$	$b \log a$		$ba^\alpha (\log a)^\varepsilon$		$\mathbb{H}$	
$\delta = 0,$	$0$							$\Phi$		
$s > 1$		$(-1)^\varepsilon S(ab)\infty$	$3$	$b$	$-S(b)\infty$	$2$	$0$	$-1$	$h\{\Phi\}$	
		$\lambda(s)^{\alpha(n-1)}(n-1)^\varepsilon$		$1$	$n-1$		$\lambda(s)^{\alpha(n-1)}(n-1)^\varepsilon$		$r_n$	
		$ba^\alpha c(s)^\alpha (\log \lambda(s))^\varepsilon$		$b$	$b \log \lambda(s)$		$ba^\alpha c(s)^\alpha (\log \lambda(s))^\varepsilon$		$\mathbb{H}$	
$\delta = 1,$	$-S(a)\infty$							$\Phi$		
$s < 1$		0	$b$	$S(b)\infty$	$(-1)^\varepsilon S(b)S(-a)^\alpha \infty$	$2$	$(-1)^\varepsilon S(b)S(-a)^\alpha \infty$	$3$	$h\{\Phi\}$	
		$\lambda(s)^{\alpha(n-1)}(n-1)^{\alpha+\varepsilon}$	$-1$	$1$	$n-1$		$\lambda(s)^{\alpha(n-1)}(n-1)^{\alpha+\varepsilon}$		$r_n$	
		$ba^\alpha c(s)^\alpha \left(\frac{\lambda'(s)}{\lambda(s)}\right)^\alpha (\log \lambda(s))^\varepsilon$		$b$	$b \log \lambda(s)$		$ba^\alpha c(s)^\alpha \left(\frac{\lambda'(s)}{\lambda(s)}\right)^\alpha (\log \lambda(s))^\varepsilon$		$\mathbb{H}$	
$\delta = 1,$	$0$							$\Phi$		
$s > 1$		$(-1)^\varepsilon S(b)S(-a)^\alpha \infty$	$3$	$b$	$-S(b)\infty$	$2$	$0$	$-1$	$h\{\Phi\}$	
		$\lambda(s)^{\alpha(n-1)}(n-1)^{\alpha+\varepsilon}$		$1$	$n-1$		$\lambda(s)^{\alpha(n-1)}(n-1)^{\alpha+\varepsilon}$		$r_n$	
		$ba^\alpha c(s)^\alpha \left(\frac{\lambda'(s)}{\lambda(s)}\right)^\alpha (\log \lambda(s))^\varepsilon$		$b$	$b \log \lambda(s)$		$ba^\alpha c(s)^\alpha \left(\frac{\lambda'(s)}{\lambda(s)}\right)^\alpha (\log \lambda(s))^\varepsilon$		$\mathbb{H}$	
		$\alpha < 0$	$\alpha = 0, \varepsilon = 0$	$\alpha = 0, \varepsilon = 1$			$0 < \alpha < 1$			
$\delta = 1,$	$-S(a)\infty$							$\Phi$		
$s = 1$		0	$b$	$S(b)\infty$	$S(b)S(-a)^\alpha \infty$		$(n-1)^\alpha (\log(n-1))^\varepsilon$	$1$	$h\{\Phi\}$	
		$(n-1)^\alpha (\log(n-1))^\varepsilon$	$-1$	$1$	$\log(n-1)$	$1$	$(n-1)^\alpha (\log(n-1))^\varepsilon$		$r_n$	
		$ba^\alpha \lambda'(1)^\alpha$		$b$	$b$		$ba^\alpha \lambda'(1)^\alpha$		$\mathbb{H}$	
				$\alpha = 1, \varepsilon = 0$	$\alpha = 1, \varepsilon = 1$			$\alpha > 1$		
$\delta = 1,$	$-S(a)\infty$							$\Phi$		
$s = 1$		$-S(ab)\infty$	$2$	$S(ab)\infty$	$S(b)S(-a)^\alpha \infty$	$3$	$(n-1)^\alpha (\log(n-1))^\varepsilon$	$3$	$h\{\Phi\}$	
		$n-1$		$(n-1) \log(n-1)$	$(n-1)^\alpha (\log(n-1))^\varepsilon$		$(n-1)^\alpha (\log(n-1))^\varepsilon$		$r_n$	
		$ba\lambda'(1)$		$ba\lambda'(1)$	$ba^\alpha \lambda'(1)^\alpha$		$ba^\alpha \lambda'(1)^\alpha$		$\mathbb{H}$	

**Table 3** From left to right: QLP Parameters, Asymptotic  $\varphi$ -Entropy  $\Phi = \lim_{S_{z,\varphi(x)} X_0^n$ , Asymptotic Marginal Entropy  $h\{\Phi\} = \lim_{h(z),\varphi(x)} S_{h(z),\varphi(x)}(X_0^n)$ , Averaging Sequence  $r_n$ , Entropy Rate  $\mathbb{H} = \mathbb{H}_{h(z),\varphi(x),r}(\mathbf{X})$ , and Type ( $-1$ : contracting,  $0$ : constant,  $1$ : sub-linear,  $2$ : linear,  $3$ : over-linear) according to  $a, s, \delta, b, \alpha$  and  $\varepsilon$  satisfying (12).