Determination and Estimation of Generalized Entropy Rates for Markov Chains

1. *Shannon entropy rate*
   Entropy rate
   Asymptotic Equipartition Property
2. *Generalized entropy rates*
   Generalized entropy functionals
   Determination of the entropy rate
   Explicite expression fo Markov chains
3. *Estimation of entropy rates for Markov chains*
   Shannon entropy and finite chains
   Generalized entropy and denumerable chains

Valerie Girardin
Université de Caen, France

Joint work with:
Gabriela Ciuperca, U. Lyon,
Loïck Lhote, ENSICAEN,
André Sesboüé, U. Caen.

Shannon entropy rate of a stochastic process

- The **entropy up to time** n of a random sequence $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ with denumerable state space E is

$$-\sum_{i_1,\dots,i_n\in E} p_n(i_1^n)\log p_n(i_1^n),$$

where $p_n(i_1^n) = \mathbb{P}[(X_1,\dots,X_n) = (i_1,\dots,i_n)]$ is the likelihood of the sequence.

- The **entropy rate** of $\mathbf{X}$ is defined by

$$-\frac{1}{n}\sum_{i_1,\dots,i_n\in E} p_n(i_1^n)\log p_n(i_1^n) \longrightarrow \mathbb{H}(\mathbf{X}), \quad n\to+\infty,$$

when this quantity is finite.

- **Asymptotic Equirepartition Property** :

$$-\frac{1}{n}\log p_n(X_1^n) \longrightarrow \mathbb{H}(\mathbf{X}), \quad n\to+\infty,$$

weak if the convergence is in probability,
strong if it holds almost surely.

Generalized entropy functionals

The (h, ϕ)-entropy of any measure ν on E is defined by

$$\mathbb{S}_{h(y),\phi(x)}(\nu) = h\left[\sum_{i\in E} \phi(\nu(i))\right]$$

if $\sum_{i\in E} \phi(\nu(i))$ is finite, and as $+\infty$ either.

The functions $h : \mathbb{R} \to \mathbb{R}$ and $\phi : [0,1] \to \mathbb{R}_+$ are twice continuously differentiable functions, with either ϕ concave and h increasing or ϕ convex and h decreasing.

Some (h, ϕ)-entropies :

$h(y)$	$\phi(x)$	(h,ϕ) – entropies
y	$-x \log x$	Shannon (1948)
$(1-s)^{-1} \log y$	x^s	Renyi (1961)
$[t(t-r)]^{-1} \log y$	$x^{r/t}$	Varma (1966)
y	$(1-2^{1-s})^{-1}(x-x^s)$	Havrda and Charvat (1967)
$(t-1)^{-1}(y^t-1)$	$x^{1/t}$	Arimoto (1971)
$(r-1)^{-1}[y^{(r-1)/(s-1)}-1]$	x^s	Sharma and Mittal 1 (1975)
$(r-1)^{-1}[\exp(r-1)y-1]$	$-x \log x$	Sharma and Mittal 2 (1975)
y	$-x^s \log x$	Taneja (1975)
y	$(t-r)^{-1}(x^r-x^t)$	Sharma and Taneja (1975)
$(r-1)^{-1}(1-y)$	x^r	Tsallis (1988)

- The (h, ϕ)-entropy rate of a random sequence $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ with state space $E \subset \mathbb{N}$ is defined by

$$\frac{1}{n}\mathbb{S}_{h(y),\phi(x)}(p_n) \longrightarrow \mathbb{H}_{h,\phi}(\mathbf{X}), \quad n \to +\infty.$$

where $p_n(i_0^n) = \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1})$ is the distribution of $(X_0, \ldots, X_{n-1})$.

Quasi-power property The process $\mathbf{X}$ satisfies the quasi-power property with parameters $[\sigma_0, \lambda, c, \rho]$ if:

1. $\sup_{i_0^n \in E^{n+1}} p_n(i_0^n) \longrightarrow 0$ when $n \to \infty$.

2. $\exists \sigma_0 \in]-\infty, 1]$, such that $\forall s > \sigma_0$ and $\forall n \in \mathbb{N}$, the series

$$\Lambda_n(s) = \textstyle\sum_{i_0^n \in E^{n+1}} p_n(i_0^n)^s$$

is convergent and satisfies

$$\Lambda_n(s) = c(s) \cdot \lambda(s)^n + R_n(s),$$

with $|R_n(s)| = O\left(\rho(s)^n \lambda(s)^n\right)$, where: c and λ are strictly positive analytic functions for $s > \sigma_0$; λ is strictly decreasing with $\lambda(1) = c(1) = 1$, R_n is also analytic, $\rho(s) < 1$.

Remarks:

The quasi-power property says that $\Lambda_n(s)$ behaves like the n-th power of some analytic function.

In dynamical systems theory, $\Lambda_n(s)$ is called the Dirichlet series of fundamental measures of depth $n + 1$.

Classical entropy rates of a random sequence satisfying the quasi-power property.

Entropy	Parameters	Entropy rate
Shannon		$-\lambda'(1)$
Rényi	$s = 1$	$-\lambda'(1)$
	$s \neq 1$	$\frac{1}{1-s} \log \lambda(s)$
Varma	$r = t$	$-\frac{1}{m^2}\lambda'(1)$
	$r \neq t$	$\frac{1}{t(t-r)} \log \lambda(r/t)$
Havrda-Charvat	$s > 1$	0
	$s = 1$	$\frac{-1}{\log 2}\lambda'(1)$
	$s < 1$	$+\infty$
Arimoto	$t > 1$	$+\infty$
	$t = 1$	$-\lambda'(1)$
	$t < 1$	0
Sharma-Mittal 1	$r < 1$	$+\infty$
	$r > 1$	0
	$s = r = 1$	$-\lambda'(1)$
	$r = 1 \neq s$	$\frac{1}{1-s} \log \lambda(s)$
Sharma-Mittal 2		$(1-s)^{-1}[\exp(-(s-1)\lambda'(1)) - 1]$
Taneja	$r < 1$	$+\infty$
	$r = 1$	$-\lambda'(1)$
	$r > 1$	0
Sharma-Taneja	$r < 1$ or $s < 1$	$+\infty$
	$r > 1$ and $s > 1$	0
	$r = 1$ and $s > 1$	0
	$r = 1$ and $s = 1$	$-\lambda'(1)$
	$r > 1$ and $s = 1$	0
Tsallis	$r < 1$	$+\infty$
	$r = 1$	$-\lambda'(1)$
	$r > 1$	0

For an i.i.d. sequence with common distribution ν

Since $p_n(i_0, i_1, \ldots, i_n) = \nu(i_0)\nu(i_1)\ldots\nu(i_n)$, the Dirichlet series $\Lambda_n(s)$ can simply be written

$$\Lambda_n(s) = \left[\sum_{i \in E} \nu(i)^s\right]^{n+1}.$$

Hence, $\mathbf{X}$ satisfies the quasi-power property for $s > 0$ with functions λ, c and ρ defined by

$$\lambda(s) = \sum_{i \in E} \nu(i)^s, \quad c(s) = 1 \quad \text{and} \quad \rho(s) = 0.$$

For a finite chain

$\Lambda_n(s) = \mathbf{1} \cdot P_s^n \cdot \nu_s$, where $P_s = (p(i,j)^s)_{i,j \in E}$, with ν the initial distribution of the chain, and $\nu_s = (\nu(i)^s)_{i \in E}$.

The following relation defines the functions λ, c and ρ of the quasi-power property:

$$P_s^n \cdot \mathbf{v} = \lambda(s)^n \cdot < \mathbf{v}, \mathbf{r}_s > \mathbf{l}_s + R^n(s) \cdot \mathbf{v},$$

where $\lambda(s)$ is the unique dominant eigenvalue of P_s with maximum modulus, with associated left and right eigenvectors $\mathbf{l}_s$ and $\mathbf{r}_s$.

For a denumerable chain

Theorem *Ciuperca, Girardin, Lhote (2010)*

Let $\mathbf{X} = (X_n)$ be an ergodic Markov chain with transition matrix P and initial distribution ν. Suppose that:

A. $\sup_{(i,j)\in E^2} P(i,j) < 1$

B. $\exists \sigma_0 < 1$ such that $\forall s > \sigma_0$,

$$\sup_{i\in E} \sum_{j\in E} P(i,j)^s < +\infty \quad \text{and} \quad \sum_{i\in E} \nu(i)^s < +\infty,$$

C. $\forall \epsilon > 0$ and $\forall s > \sigma_0$, $\exists A \subset E$ with $|A| < +\infty$ such that

$$\sup_{i\in E} \sum_{j\in E\setminus A} P(i,j)^s < \varepsilon.$$

Then $\mathbf{X}$ satisfies the quasi-power property.

Proof of the theorem

Lemma If Assumptions A, B, C hold true,

then $P_s : (\ell^1, ||\ .\ ||_1) \to (\ell^1, ||\ .\ ||_1)$ is a compact operator, $\forall s > \sigma_0$,

where $\ell^1 = \{u = (u_i)_{i\in E} : ||u||_1 = \sum_{i\in E} |u_i| < \infty\}$.

We deduce from the lemma that the spectrum of P_s is a sequence that converges to zero. Hence, P_s has a finite number of eigenvalues with maximum modulus and there exists a spectral gap separating these dominant eigenvalues from the remainder of the spectrum.

Further, since $\mathbf{X}$ is ergodic, P_s has a unique dominant eigenvalue $\lambda(s)$ which, moreover, is positive. Hence,

$$P_s^n u = \lambda(s)^n Q_s u + R_s^n u, \qquad u \in \ell^1,$$

where Q_s is the projector over the dominant eigenspace and R_s is the projector over the remainder of the spectrum. The spectral radius of R_s can be written $\rho(s) \cdot \lambda(s)$ with $\rho(s) < 1$.

Finally,

$$\Lambda_n(s) = \lambda(s)^n ||Q_s \nu_s||_1 (1 + O(\rho(s)^n \lambda(s)^n)),$$

which means that $\mathbf{X}$ satisfies the quasi-power property.

The analyticity of the involved functions is due jointly to the analyticity of $s \to P_s$ and to perturbation arguments. $\square$

Theorem Let $\mathbf{X}$ be a random sequence satifying the quasi-power property with parameters $[\sigma_0, \lambda, c, \rho]$. Suppose that

$$\phi(x) \underset{x \to 0}{\sim} c_1 \cdot x^s \cdot (\log x)^k \qquad (P)$$

with $s > \sigma_0$, $c_1 \in \mathbb{R}_+^*$ and $k \in \mathbb{N}^*$. Then the entropy rate $\mathbb{H}_{h,\phi}(\mathbf{X})$ is given by the following table.

Value of s	Condition on h	Entropy rate
$s = 1$	$h(x) \underset{x \to +\infty}{\sim} c_2 \cdot x^{1/k}$	$c_2 \cdot c_1^{1/k} \cdot \lambda'(1)$
	$h(x) \underset{x \to +\infty}{=} o(x^{1/k})$	0
	$x^{1/k} \underset{x \to +\infty}{=} o(h(x))$	$+\infty$
$s > 1$	$h(x) \underset{x \to 0^+}{\sim} c_2 \cdot \log x$	$c_2 \cdot \log \lambda(s)$
	$h(x) \underset{x \to 0^+}{=} o(\log x)$	0
	$\log x \underset{x \to 0^+}{=} o(h(x))$	$+\infty$
$\sigma_0 < s < 1$	$h(x) \underset{x \to +\infty}{\sim} c_2 \cdot \log x$	$c_2 \cdot \log \lambda(s)$
	$h(x) \underset{x \to +\infty}{=} o(\log x)$	0
	$\log x \underset{x \to +\infty}{=} o(h(x))$	$+\infty$

Proof $\sup_{i_0^n \in E^{n+1}} \nu_n(i_0^n) \to 0$ and (P) together induce that $\forall \epsilon > 0$, $\exists n_0 \in \mathbb{N} /\ n \geq n_0$ and $i_0^n \in E^{n+1}$,

$$(1-\epsilon) c_1 \nu_n(i_0^n)^s \log^k \nu_n(i_0^n) \leq \phi(\nu_n(i_0^n))$$
$$\leq (1+\epsilon) c_1 \nu_n(i_0^n)^s \log^k \nu_n(i_0^n),$$

from which it follows that

$$(1-\epsilon) c_1 \Lambda_n^{(k)}(s) \leq \sum_{i_0^n \in E^{n+1}} \phi(\nu_n(i_0^n)) \leq (1+\epsilon) c_1 \Lambda_n^{(k)}(s).$$

Due to the analyticity of all involved functions,

$$\Lambda_n^{(k)}(s) = c(s) \cdot \lambda'(s)^k \cdot n^k \cdot \lambda(s)^{n-k} \cdot [1 + O(1/n)].$$

which yields

$$\sum_{i_0^n \in E^{n+1}} \phi(\nu_n(i_0^n)) \sim c_1 \cdot c(s) \cdot \lambda'(s)^k \cdot n^k \cdot \lambda(s)^{n-k}.$$

Since ϕ is nonnegative, this sum converges polynomially to infinity. This leads to the next equivalences:

$$\begin{array}{lll} h(\Sigma_n) \sim c_2 \cdot |c_1|^{1/k} \cdot |\lambda'(1)| \cdot n & \text{if} & h(x) \sim c_2 \cdot x^{1/k}, \\ h(\Sigma_n) \sim o(n) & \text{if} & h(x) = o(x^{1/k}), \\ h(\Sigma_n) \sim s_n \cdot n \text{ with } s_n \to \infty & \text{if} & x^{1/k} = o(h(x)). \end{array}$$

Since by definition, the (h, ϕ)-entropy rate is the limit of $h(\Sigma_n)/n$ when n tends to infinity, the results follow immediately for $s = 1$.

The other cases can be studied similarly. □

Entropy	Parameters	Entropy rate
Shannon		$-\lambda'(1)$
Rényi	$s = 1$	$-\lambda'(1)$
	$s \neq 1$	$\frac{1}{1-s} \log \lambda(s)$
Varma	$r = t$	$-\frac{1}{m^2}\lambda'(1)$
	$r \neq t$	$\frac{1}{t(t-r)} \log \lambda(r/t)$
Havrda-Charvat	$s > 1$	0
	$s = 1$	$\frac{-1}{\log 2}\lambda'(1)$
	$s < 1$	$+\infty$
Arimoto	$t > 1$	$+\infty$
	$t = 1$	$-\lambda'(1)$
	$t < 1$	0
Sharma-Mittal 1	$r < 1$	$+\infty$
	$r > 1$	0
	$s = r = 1$	$-\lambda'(1)$
	$r = 1 \neq s$	$\frac{1}{1-s} \log \lambda(s)$
Sharma-Mittal 2		$(1-s)^{-1}[\exp(-(s-1)\lambda'(1)) - 1]$
Taneja	$r < 1$	$+\infty$
	$r = 1$	$-\lambda'(1)$
	$r > 1$	0
Sharma-Taneja	$r < 1$ or $s < 1$	$+\infty$
	$r > 1$ and $s > 1$	0
	$r = 1$ and $s > 1$	0
	$r = 1$ and $s = 1$	$-\lambda'(1)$
	$r > 1$ and $s = 1$	0
Tsallis	$r < 1$	$+\infty$
	$r = 1$	$-\lambda'(1)$
	$r > 1$	0

Values of classical entropy rates of a random sequence satisfying the quasi-power property with parameters $[\lambda, c, \rho, \sigma_0]$.

For an ergodic **Markov chain** $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ with state space E with s states, transition matrix $P = (P(i,j))$, where $P(i,j) = \mathbb{P}(X_{n+1} = j/X_n = i)$, and stationary distribution π such that $\pi P = \pi$, and entropy

$$\mathbb{H}(\mathbf{X}) = -\sum_{i\in E} \pi(i) \sum_{j\in E} P(i,j) \log P(i,j) = h(P)$$

$$(= -\lambda'(1)) .$$

Proposition *Anderson and Goodman (1957)*
The empirical estimators

$$\widehat{P}_n(i,j) = \frac{\sum_{m=1}^{n} \mathbb{1}_{\{X_{m-1}=i, X_m=j\}}}{\sum_{j\in E} \sum_{m=1}^{n} \mathbb{1}_{\{X_{m-1}=i, X_m=j\}}}$$

are strongly convergent and asymptotically normal:

$$\sqrt{n}\left(\widehat{P}_n(i,j) - P(i,j)\right) \xrightarrow{\mathcal{L}} \mathcal{N}_{s^2}(0, \Gamma^2)$$

where $\Gamma^2_{ij} = \delta_{ik}[\delta_{jl} P(i,j) - P(i,j)P(i,l)]/\pi(i)$.

- We define the plug-in estimator

$$\widehat{\mathbb{H}}_n = h(\widehat{P}_n)$$

of the entropy rate.

Theorem *Ciuperca and Girardin (2007)*

If the transition probabilities are not uniform, the plug-in estimator $\widehat{\mathbb{H}}_n = h(\widehat{P}_n)$ of $\mathbb{H}(\mathbf{X})$ is **strongly convergent** and **asymptotically normal**.

Precisely,

$$\sqrt{n}[\widehat{h}_n - \mathbb{H}(\mathbf{X})] \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\partial_i^j h)\,\Gamma(\partial_i^j h)'),$$

where $\partial_u^v h$ is the differential with order v with respect to variable u of h.

Proof

Continuous mapping theorem and delta method

□

For a two-state chain

The transition matrix of the chain is

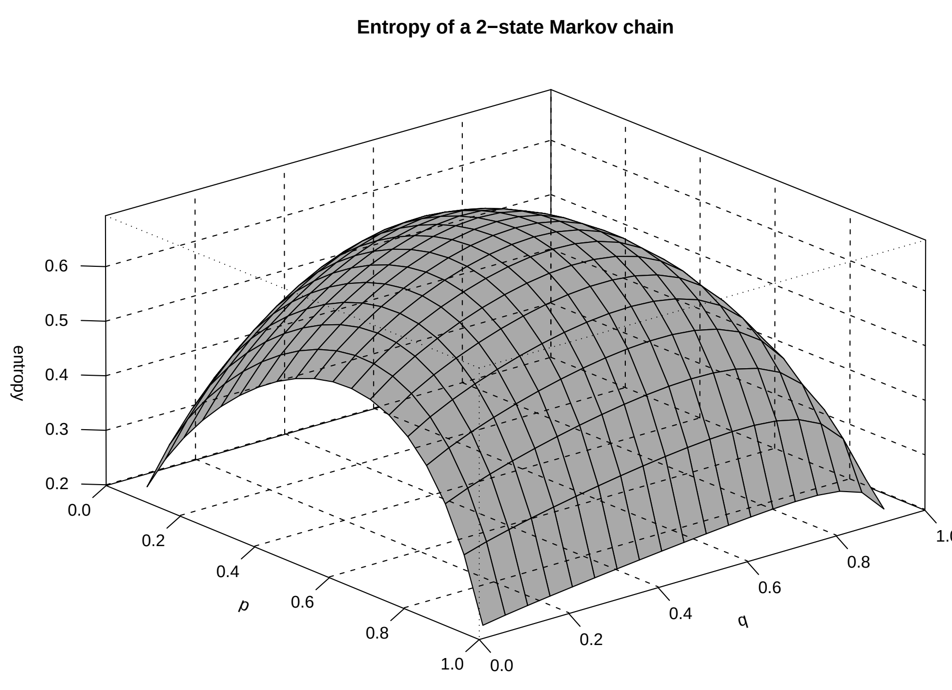
$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

The stationary distribution satisfies $\pi P = \pi$, so

$$\pi(0) = \frac{q}{p+q} \quad \text{and} \quad \pi(1) = \frac{p}{p+q}.$$

The entropy rate is

$$\begin{aligned} \mathbb{H}(\mathbf{X}) &= h(p,q) = \pi(0) S_p + \pi(1) S_q \\ &= \frac{q}{p+q}[-p\log p - (1-p)\log(1-p)] \\ &\quad + \frac{p}{p+q}[-q\log q - (1-q)\log(1-q)]. \end{aligned}$$



Theorem *Girardin and Sesboue (2009)*

$$\widehat{h}_n = h(\widehat{p}_n, \widehat{q}_n) \xrightarrow{a.s.} \mathbb{H}(\mathbf{X}).$$

If the chain is not uniform,

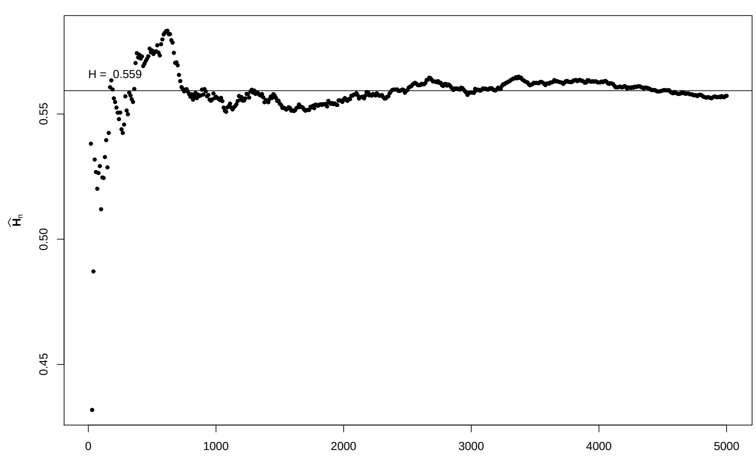
$$\sqrt{n}[\widehat{h}_n - \mathbb{H}(\mathbf{X})] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

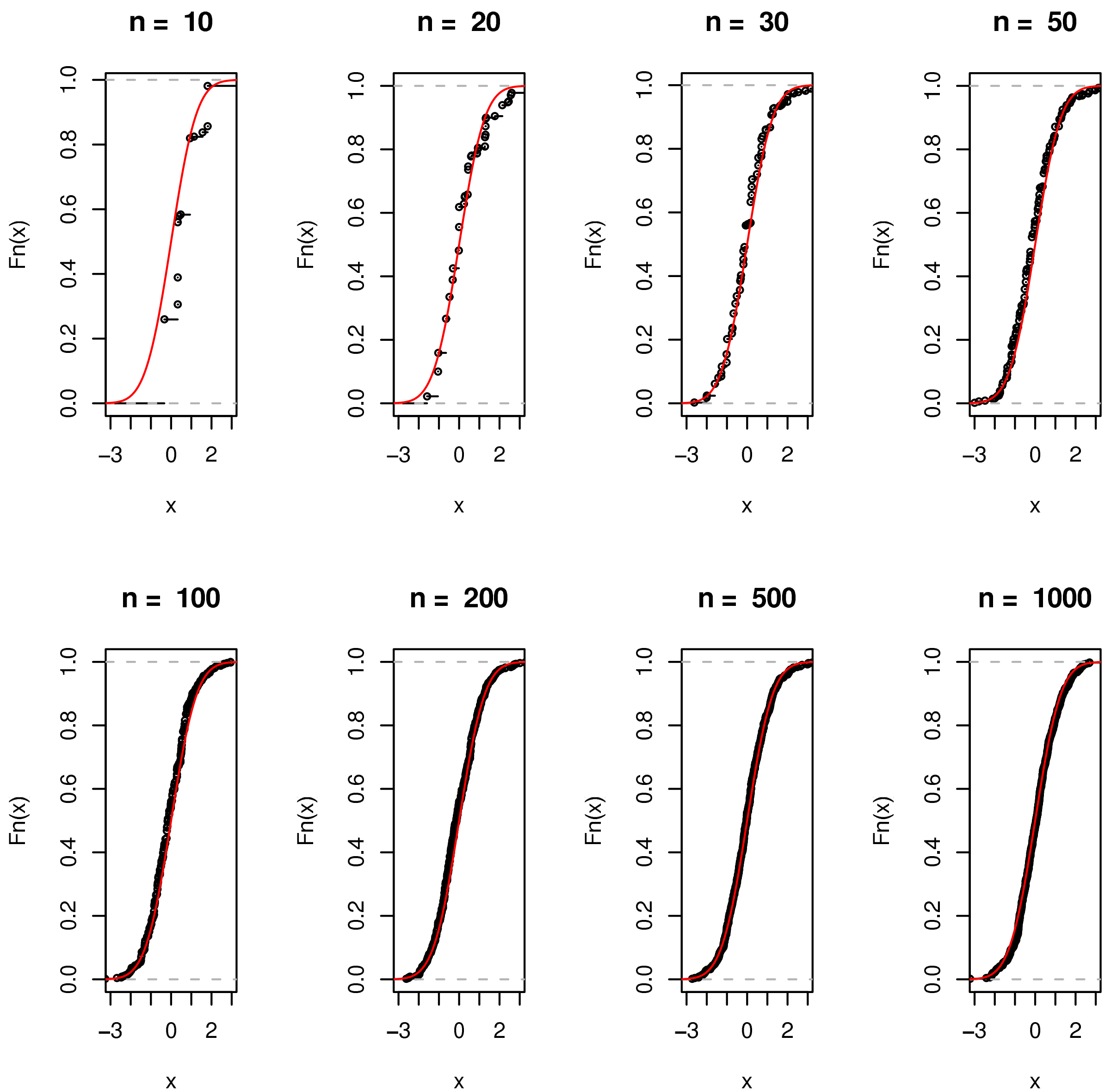
$$\begin{aligned} \text{where } \sigma^2 &= \Gamma(0,0)^2[\partial_1^1 h(p,q)]^2 + \Gamma(1,1)^2[\partial_2^1 h(p,q)]^2 \\ &= \quad pq(1-p)\left[\frac{S_q - S_p}{p+q} - \log\frac{p}{1-p}\right] \\ &\quad +pq(1-q)\left[\frac{S_p - S_q}{p+q} - \log\frac{q}{1-q}\right] \end{aligned}$$

For illustration, we have simulated a chain for $p = 0.2$ and $q = 0.3$, for which $\mathbb{H}(\mathbf{X}) = 0,559$.

The first figure shows the punctual convergence of $\widehat{h}_n$ to $\mathbb{H}(\mathbf{X})$ for $n = 10$ to 5000 by steps of 10.

(computation of $\widehat{h}_n$ for $10 \leq n \leq 5000$ after simulation of one trajectory with length 5000)





This figure compares the empirical distribution function of $\sqrt{n}[\widehat{h}_n - \mathbb{H}(\mathbf{X})]/\widehat{\sigma}_n$ to that of the standard normal distribution for different values of $10 \leq n \leq 1000$.

(for $T = 500$ trajectories simulated for each n)

Theorem *Girardin and Sesboue (2009)*

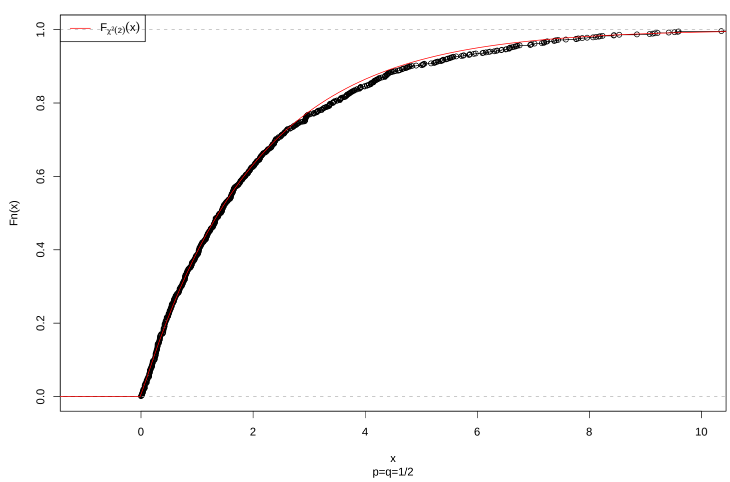
For a uniform chain, $p = q = 1/2$, $\widehat{h}_n$ is strongly convergent and $2n[\mathbb{H}(\mathbf{X}) - \widehat{h}_n] \xrightarrow{\mathcal{L}} \chi^2(2)$.

Proof. $\widehat{h}_n - \mathbb{H}(\mathbf{X}) =$

$$\begin{aligned}
&= [\partial_1^1 h(p,q)][\widehat{P}(0,1) - p] + [\partial_2^1 h(p,q)][\widehat{P}(1,0) - q] \\
&\quad + \frac{1}{2}[\partial_1^2 h(p,q)][\widehat{P}(0,1) - p]^2 + \frac{1}{2}[\partial_2^2 h(p,q)][\widehat{P}(1,0) - q]^2 \\
&\quad + o([\widehat{P}(0,1) - p]^2) + o([\widehat{P}(1,0) - q]^2) \\
&= \tfrac{1}{2\Gamma(0,0)^2}[\widehat{P}(0,1) - p]^2 + \quad \tfrac{1}{2\Gamma(1,1)^2}[\widehat{P}(1,0) - q]^2 \\
&\quad + o([\widehat{P}(0,1) - p]^2) + o([\widehat{P}(1,0) - q]^2).
\end{aligned}$$

and the result follows, since $\frac{\sqrt{n}[\widehat{P}(0,1)-p]}{\Gamma(0,0)}$ and $\frac{\sqrt{n}[\widehat{P}(1,0)-q]}{\Gamma(1,1)}$ are asymptotically standard normal. □

The last figure compares the distribution function of $2n[\widehat{h}_n - \mathbb{H}(\mathbf{X})]$ to that of the $\chi^2(2)$-distribution for $n = 1000$. ($T = 1000$ simulated trajectories for n)



Estimation of generalized entropy rates

All the entropy rates are finite and non-zero only at a threshold where they are equal to the Rényi entropy rate up to a multiplicative factor. Therefore, we only estimate Shannon and Rényi entropy rates, that is

$$\begin{aligned} h(\theta) &= -\lambda'(1;\theta_0), \\ \text{and } h_s(\theta) &= (1-s)^{-1}\log\lambda(s;\theta_0). \end{aligned}$$

The transition probabilities of the ergodic chain $\mathbf{X}$ with denumerable state space are supposed to depend on $\theta \in \Theta^r$, with true value θ^0.

Proposition *Billingsley (1962)* Suppose that:

A. $\forall x$, $\{y : P(x,y;\theta) > 0\}$ does not depend on θ.

B. $\forall (x,y)$, $P_u(x,y;\theta)$, $P_{uv}(x,y;\theta)$ and $P_{uvw}(x,y;\theta)$ are in $\mathcal{C}^1(\Theta)$.

C. $\forall \theta \in \Theta$, $\exists N$, neighborhood such that $\forall u, v$, $P_u(x,y;\theta)$ and $P_{uv}(x,y;\theta)$ are uniformly bounded in $L^1(\mu(dy))$ on N and

$$\mathbb{E}_\theta[\sup_{\theta' \in N} \mid P_u(x,y;\theta') \mid^2] < +\infty.$$

D. $\exists \delta > 0$ such that $\mathbb{E}_\theta[\mid P_u(x,y;\theta) \mid^{2+\delta}]$ is finite $\forall u = 1, \ldots, r$.

E. The Fisher information matrix $\sigma(\theta) = (\mathbb{E}_\theta[P_u(x,y;\theta)P_v(x,y;\theta)])$ is non singular.

Then a strongly consistent maximum likelihood estimator $\widehat{\theta}_n$ of θ exists. Moreover, $\sqrt{n}(\widehat{\theta}_u - \theta_u)$ is asymptotically normal, with covariance matrix $\sigma^{-1}(\theta^0)$.

It is natural to consider the plug-in estimators:

$$h(\hat{\theta}_n) = -\lambda'(1; \theta_n)$$
$$\text{and } h_s(\widehat{\theta_n}) = (1-s)^{-1} \log \lambda(s; \widehat{\theta}_n)$$

of Shannon entropy rate and of Rényi entropy rate.

Theorem If Billingsley's assumptions are satisfied and if $\mathbf{X}$ satisfies the quasi-power property, then $h(\hat{\theta}_n)$ and $h_s(\hat{\theta}_n)$ are strongly consistent and asymptotically normal: $\sqrt{n}[h(\hat{\theta}_n) - h(\theta)] \to \mathcal{N}(0, \Sigma_1)$, where

$$\Sigma_1 = \left\{ \frac{\partial}{\partial \theta}[-\lambda'(1; \theta)] \right\}^t \sigma^{-1}(\theta) \frac{\partial}{\partial \theta}[-\lambda'(1; (\theta)]$$

and $\sqrt{n}[h_s(\hat{\theta}_n) - \mathbf{H}_s(\theta^0)] \to \mathcal{N}(0, \Sigma_s)$, where

$$\Sigma_s = \frac{1}{(1-s)^2} \left\{ \frac{\partial}{\partial \theta} \lambda(s; \theta) \right\}^t \sigma^{-1}(\theta) \frac{\partial}{\partial \theta} \lambda(s; (\theta).$$

Proof Due to operators properties, the eigenvalue $\lambda(s)$ and its derivative $\lambda'(1)$ are continuous with respect to the perturbated operator P_s. For a parametric chain depending on θ, Assumption B induces that P_s is a continuously differentiable function of θ. Therefore both $\lambda(s; \theta)$ and $\lambda'(s; \theta)$ are continuous with respect to θ. The results follow from the continuous mapping theorem and the delta method. □

Estimation of the Entropy Rate of a Countable Markov Chain

Communication in Statistics : Theory and Methods, V36, pp2543–2557, G. Ciuperca & V. Girardin (2007)

Asymptotic study of an estimator of the entropy rate of a two-state Markov chain for one long trajectory,

with A. Sesboüé, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Ed. A. Mohammad-Djafari, AIPCP, V872 pp403–410 (2006).

Comparative Construction of Plug-in Estimators of the Entropy Rate of Two-State Markov Chains

Methodology and Computing in Applied Probability, V11, pp. 181–200, V. Girardin & A. Sesboüé (2009)

Computation of Generalized Entropy Rates. Application and Estimation for Countable Markov Chains

Rapport de recherche Université de Caen, 21 pages, G. Ciuperca, V. Girardin et L. Lhote (2010) to appear in *IEEE Transactions on Information Theory*