Electronic Journal of Probability

Vol. 11 (2006), Paper no. 40, pages 1049–1067.

Journal URL
http://www.math.washington.edu/~ejpecp/

# Relative entropy and waiting times for continuous-time Markov processes

J.-R. Chazottes*    C. Giardina†    F. Redig‡

**Abstract**

For discrete-time stochastic processes, there is a close connection between return (resp. waiting) times and entropy (resp. relative entropy). Such a connection cannot be straightforwardly extended to the continuous-time setting. Contrarily to the discrete-time case one needs a reference measure on path space and so the natural object is relative entropy rather than entropy. In this paper we elaborate on this in the case of continuous-time Markov processes with finite state space. A reference measure of special interest is the one associated to the time-reversed process. In that case relative entropy is interpreted as the entropy production rate. The main results of this paper are: almost-sure convergence to relative entropy of the logarithm of waiting-times ratios suitably normalized, and their fluctuation properties (central limit theorem and large deviation principle)

**Key words:** continuous-time Markov chain, law of large numbers, central limit theorem, large deviations, entropy production, time-reversed process

**AMS 2000 Subject Classification:** Primary 60J27;60F10.

Submitted to EJP on January 10 2006, final version accepted October 4 2006.

*CPhT, CNRS-Ecole Polytechnique, 91128 Palaiseau Cedex, France, and CMM, UMI CNRS 2807, Universidad de Chile, Av. Blanco Encalada 2120, Santiago, Chile, jeanrene@cpht.polytechnique.fr

†Eurandom, Postbus 513, 5600 MB Eindhoven, The Netherlands, giardina@eurandom.tue.nl

‡Mathematisch Instituut Universiteit Leiden, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands, redig@math.leidenuniv.nl

# 1 Introduction

Many limit theorems in the theory of stochastic processes have a version for discrete-time as well as for continuous-time processes. The ergodic theory of Markov chains e.g. is more or less identical in discrete and in continuous time. The same holds for the ergodic theorem, martingale convergence theorems, central limit theorems and large deviations for additive functionals, etc. Usually, one obtains the same results with some additional effort in the continuous-time setting, where e.g. extra measurability issues can pop up.

For discrete-time ergodic processes, there is a famous theorem by Ornstein and Weiss connecting return times and entropy (see (9), (14), (11)). In words, it states that the logarithm of the first time the process repeats its first $n$ symbols typically behaves like $n$ times the entropy of the process. This provides a way to estimate the entropy of a process by observing a single trajectory. This result seems a natural candidate to transport to a continuous-time setting. The relation between entropy and return times is sufficiently intuitive so that one would not expect major obstacles on the road toward such a result for continuous-time ergodic processes. There is however one serious problem. On the path space of continuous-time processes (on a finite state space, say), there is no natural flat measure. In the discrete-time and finite state space setting, one cannot distinguish between entropy of a process and relative entropy between the process and the uniform measure on trajectories. These only differ by a constant (i.e., a quantity not depending on the process but only on the cardinality of the state space of the process). As we shall see below, the difference between relative entropy and entropy starts to play an important role in turning to continuous-time processes. In fact, it turns out that there is no naive continuous-time analogue of the relation between return times and entropy; the logarithm of return times turns out to have no suitable way of being normalized, even for very simple processes in continuous time such as Markov chains. To circumvent this drawback, we propose here to consider *the logarithm of waiting time ratios* and relate them to *relative entropy*.

Of course, *defining* a return time is already a problem if we are working in continuous time: a process never exactly reproduces a piece of trajectory. Our approach with respect to this problem is to first discretize time, and show that the relation between waiting times and relative entropy persists in the limit of vanishing discrete time-step. From the physical point of view, this is a natural procedure, and the time-step of the discretization can be associated to the acquisition frequency of the device one uses to sample the process. We can also think of numerical simulations for which the discretization of time is unavoidable. Of course, the natural issue is to verify if the results obtained with the discretized process give the correct ones for the original process, after a suitable rescaling and by letting the time-step go to zero. This will be done in the present context.

In this paper, we will restrict ourselves to continuous-time Markov chains with finite state space, for the sake of simplicity and also because the aforementioned problem already appears in this special, yet fundamental, setting. The main body of this paper is: a law of large numbers for the logarithm of waiting times ratios and its connection to relative entropy, a large deviation result and a central limit theorem for the same quantities. One possible application is the estimation of the relative entropy density between the forward and the backward process which is physically interpreted as the *mean entropy production*, and which is strictly positive if and only if the process is reversible (i.e., in "detailed balance", or in "equilibrium").

Our paper is organized as follows. In Section 2, we show why the naive generalization of the

Ornstein-Weiss theorem fails. Section 3 contains the main results about law of large numbers, large deviations and central limit theorem for the logarithm of ratios of waiting times. In the final section we consider the problem of "shadowing" a given continuous-time trajectory drawn from an ergodic distribution on path space.

## 2 Naive approach

In this section we introduce some basic notation and start with an informal discussion motivating the quantities which we will consider in the sequel. Let $\{X_t, t \geq 0\}$ be a continuous-time Markov chain on a finite state space $A$, with stationary measure $\mu$, and with generator

$$Lf(x) = \sum_{y\in A} c(x)p(x,y)(f(y) - f(x))$$

where $p(x,y)$ are the transition probabilities of a *discrete-time* irreducible Markov chain on $A$, with $p(x,x) = 0$, and where the escape rates $c(x)$ are strictly positive.

Given a time-step $\delta$, we introduce the "$\delta$-discretization" $\{X_{i\delta}, i = 0,1,2,\ldots\}$, which is then a discrete-time irreducible Markov chain. Next we define the first time the $\delta$-discretized process repeats its first $n$ symbols via the random variable

$$R_n^\delta(X) := \inf\{k \geq 1 : (X_0,\ldots,X_{(n-1)\delta}) = (X_{k\delta},\ldots,X_{(k+n-1)\delta})\}\,. \tag{2.1}$$

For $x_1,\ldots,x_n \in A$ we denote by $x_1^n$ the set of those discrete-time trajectories $\omega \in A^{\mathbb{N}}$ such that $\omega_1 = x_1,\ldots,\omega_n = x_n$. We denote by $\mathbb{P}^\delta$ the probability measure on $A^{\mathbb{N}}$ given by the joint distribution of $\{X_{n\delta} : n \in \mathbb{N}\}$, starting from the stationary measure $\mu$, i.e., with $X_0$ distributed according to $\mu$. By $\mathbb{P}^\delta(X_1^n)$ we denote the probability of the *random* $n$-cylinder given by the first $n$ symbols of the discretized Markov chain, i.e.,

$$\mathbb{P}^\delta(X_1^n) = \mu(X_1)\prod_{i=1}^{n-1} p_\delta^X(X_i,X_{i+1}) \tag{2.2}$$

The analogue of the Ornstein-Weiss theorem (see (9), (11)) for this continuous-time process would be a limit theorem for a suitably normalized version of $\log R_n^\delta$ for $n\to\infty, \delta\to 0$. However, for $\delta > 0$ fixed, the $\delta$-discretization $\{X_0, X_\delta, X_{2\delta},\ldots,X_{n\delta},\ldots\}$ is a *discrete-time* ergodic Markov chain for which we can *apply* the Ornstein-Weiss theorem. Moreover, in combination with Shannon-McMillan-Breiman theorem we can write

$$\lim_{n\to\infty}\frac{1}{n}\log\left[R_n^\delta(X)\mathbb{P}^\delta(X_1^n)\right] = 0 \text{ a.s.}$$

Using the fact that $X$ is an ergodic Markov chain we obtain

$$\begin{aligned}
-\frac{1}{n}\log R_n^\delta(X) &= \frac{1}{n}\log\mu(X_0) + \frac{1}{n}\left(\sum_{i=1}^{n-1}\log p_\delta^X(X_i,X_{i+1})\right) + o(1)\\
&= \mathbb{E}\left[\log p_\delta^X(X_0,X_1)\right] + o(1)
\end{aligned} \tag{2.3}$$

where by $o(1)$ we mean a random variable converging to zero almost surely as $n \to \infty$, where $\mathbb{E}$ denotes expectation in the Markov chain started from its stationary distribution and where $p_\delta^X$ denotes the transition probability of the $\delta$-discretized Markov chain $\{X_{i\delta}, i = 0, 1, 2, \ldots\}$, i.e.,

$$p_\delta^X(x,y) = \mathbb{P}(X_\delta = y | X_0 = x) = (e^{\delta L})_{xy} = 1\!\!1_{xy}(1 - \delta c(x)) + \delta c(x) p(x,y) + \mathcal{O}(\delta^2)\,.$$

Therefore,

$$-\lim_{n\to\infty} \frac{1}{n} \log R_n^\delta(X) =$$
$$\sum_{x\in A} \mu(x)(1-\delta c(x))\log(1-\delta c(x)) + \sum_{x,y\in A} \mu(x)\delta c(x) p(x,y) \log(\delta c(x) p(x,y)) + \mathcal{O}(\delta^2)\,.$$

In the rhs, the first sum is of order $\delta$ whereas the second one is of order $\delta \log \delta$. Therefore, this expression does not seem to have a natural way to be normalized. This is a typical phenomenon for continuous-time processes: we will need a suitable reference process in order to define "entropy" as "relative entropy" with respect to this reference process. In the discrete-time context this reference measure is obviously the uniform measure on trajectories. Instead of considering return times, as we will see in the next sections, by considering *differences of waiting times* one is able to cancel the $\delta \log \delta$ term and obtain expressions that do converge in the limit $\delta \downarrow 0$ to relative entropy.

# 3 Main results: waiting times and relative entropy

We consider continuous-time Markov chains with a finite state-space $A$. We will always work with irreducible Markov chains with a unique stationary distribution. The process is denoted by $\{X_t : t \geq 0\}$. The associated measure on path space starting from $X_0 = x$ is denoted by $\mathbb{P}_x$ and by $\mathbb{P}$ we denote the path space measure of the process started from its unique stationary distribution. For $t \geq 0$, $\mathcal{F}_t$ denotes the sigma-field generated by $X_s, s \leq t$, and $\mathbb{P}^{[0,t]}$ denotes the measure $\mathbb{P}$ restricted to $\mathcal{F}_t$.

## 3.1 Relative entropy: comparing two Markov chains

Consider two continuous-time Markov chains, one denoted by $\{X_t : t \geq 0\}$ with generator

$$Lf(x) = \sum_{y\in A} c(x) p(x,y)(f(y) - f(x)) \tag{3.1}$$

and the other denoted by $\{Y_t : t \geq 0\}$ with generator

$$\tilde{L}f(x) = \sum_{y\in A} \tilde{c}(x) p(x,y)(f(y) - f(x))$$

where $p(x,y)$ is the Markov transition function of an irreducible discrete-time Markov chain. We further assume that $p(x,x) = 0$, and $c(x) > 0$ for all $x \in A$. We suppose that $X_0$, resp. $Y_0$, is distributed according to the unique stationary measure $\mu$, resp. $\tilde{\mu}$ so that both processes are stationary and ergodic.

**Remark 1.** *The fact that the Markov transition function $p(x, y)$ is the same for both processes is only for the sake of simplicity. All our results can be reformulated in the case that the Markov transition functions would be different.*

We recall Girsanov's formula (see Proposition 2.6, in Appendix 1 of (6) for a straightforward approach, or section 19 of (7) for more background on exponential martingales in the context of point processes).

$$\frac{d\mathbb{P}^{[0,t]}}{d\tilde{\mathbb{P}}^{[0,t]}}(\omega) = \frac{\mu(\omega_0)}{\tilde{\mu}_(\omega_0)} \exp\left( \int_0^t \log \frac{c(\omega_s)}{\tilde{c}(\omega_s)}\ dN_s(\omega) - \int_0^t (c(\omega_s) - \tilde{c}(\omega_s))\ ds \right) \tag{3.3}$$

where $N_s(\omega)$ is the number of jumps of the path $\omega$ up to time $s$. The relative entropy of $\mathbb{P}$ w.r.t. $\tilde{\mathbb{P}}$ up to time $t$ is defined as

$$s_t(\mathbb{P}|\tilde{\mathbb{P}}) = \int d\mathbb{P}(\omega) \log \left( \frac{d\mathbb{P}^{[0,t]}}{d\tilde{\mathbb{P}}^{[0,t]}}(\omega) \right) \cdot \tag{3.4}$$

Using (3.3), stationarity, and the fact that

$$M_t := N_t - \int_0^t c(\omega_s)\ ds$$

is a (mean zero) martingale, we have that for every function $\varphi : A \to \mathbb{R}$

$$\mathbb{E}\left( \int_0^t \varphi(\omega_s) dN_s \right) = \mathbb{E}\left( \int_0^t \varphi(\omega_s) c(\omega_s) ds \right) = t \sum_x \mu(x)\varphi(x)c(x)$$

Therefore, returning to (3.3) we obtain

$$\begin{aligned} s_t(\mathbb{P}|\tilde{\mathbb{P}}) &= t \left( \sum_{x\in A} \mu(x)c(x) \log \frac{c(x)}{\tilde{c}(x)} - \sum_{x\in A} \mu(x)(c(x) - \tilde{c}(x)) \right) + C(\mu, \tilde{\mu}) \\ &=: t\mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) + C(\mu, \tilde{\mu}) \end{aligned} \tag{3.5}$$

where

$$C(\mu, \tilde{\mu}) = \sum_{x\in A} \mu(x) \log \frac{\mu(x)}{\tilde{\mu}(x)}$$

and where

$$\mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) = \left( \sum_{x\in A} \mu(x)c(x) \log \frac{c(x)}{\tilde{c}(x)} - \sum_{x\in A} \mu(x)(c(x) - \tilde{c}(x)) \right) \tag{3.6}$$

is called the *relative entropy per unit time* (or relative entropy density) of $\mathbb{P}$ with respect to $\tilde{\mathbb{P}}$. Notice also that, by the assumed ergodicity of the Markov process $X$,

$$\lim_{t\to\infty} \frac{1}{t} \log \frac{d\mathbb{P}^{[0,t]}}{d\tilde{\mathbb{P}}^{[0,t]}}(\omega) = \mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) \quad \mathbb{P} - \text{a.s.}.$$

In the case $\{Y_t : t \geq 0\}$ is Markov chain with generator

$$\tilde{L}f(x) = \sum_{y\in A} \tilde{c}(x)\tilde{p}(x, y)(f(y) - f(x))$$

(3.5) would give

$$\mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) = \sum_{x,y\in A} \mu(x)c(x)p(x,y)\log\frac{c(x)p(x,y)}{\tilde{c}(x)\tilde{p}(x,y)} - \sum_{x\in A}\mu(x)(c(x)-\tilde{c}(x)) \tag{3.7}$$

**Remark 2.** *An important particular case is met when $\{Y_t : t \geq 0\}$ is the time-reversed process of $\{X_t : t \geq 0\}$, i.e.,*

$$(Y_t)_{0\leq t\leq T} = (X_{T-t})_{0\leq t\leq T} \text{ in distribution.}$$

*This is a Markov chain with transition rates*

$$\tilde{c}(x,y) = c(x)\tilde{p}(x,y) \tag{3.9}$$

*where*

$$\tilde{p}(x,y) = \frac{c(y)\mu(y)p(y,x)}{c(x)\mu(x)} \tag{3.10}$$

*which coincides with $p(x,y)$ in the case where $\mu$ is a reversible measure for the Markov process $X$, i.e., when*

$$c(y)\mu(y)p(y,x) = c(x)\mu(x)$$

*For the choice* (3.9)*, the random variable*

$$S_T(\omega) = \log\frac{d\tilde{\mathbb{P}}^{[0,T]}}{d\mathbb{P}^{[0,T]}} \tag{3.11}$$

*has the interpretation of "entropy production", and the relative entropy density $s(\mathbb{P}|\tilde{\mathbb{P}})$ has the interpretation of "mean entropy production per unit time"; see e.g. (5; 8).*

## 3.2 Law of large numbers

For $\delta > 0$, we define the discrete-time Markov chain $X^\delta := \{X_0, X_\delta, X_{2\delta}, \ldots\}$. This Markov chain has transition probabilities

$$\begin{aligned} p^X_\delta(x,y) &= (e^{\delta L})_{xy} \\ &= \mathbb{1}_{xy}(1-\delta c(x)) + \delta c(x)p(x,y) + \mathcal{O}(\delta^2) \end{aligned} \tag{3.12}$$

where $\mathbb{1}$ is the identity matrix. Similarly, we define the Markov chain $Y^\delta$ with transition probabilities

$$\begin{aligned} p^Y_\delta(x,y) &= (e^{\delta\tilde{L}})_{xy} \\ &= \mathbb{1}_{xy}(1-\delta\tilde{c}(x)) + \delta\tilde{c}(x)p(x,y) + \mathcal{O}(\delta^2)\,. \end{aligned} \tag{3.13}$$

The path-space measure (on $A^{\mathbb{N}}$) of $X^\delta$, resp. $Y^\delta$, starting from the stationary measure, is denoted by $\mathbb{P}^\delta$, resp. $\tilde{\mathbb{P}}^\delta$. As before, (see (2.2)) we will use the notation $\mathbb{P}^\delta(X_1^n)$ for the probability of the random $n$-cylinder $X_1^n$ in the discretized process.

We define waiting times, which are random variables defined on $A^{\mathbb{N}}\times A^{\mathbb{N}}$, by setting

$$W^\delta_n(X|Y) = \inf\{k\geq 1 : (X^\delta_1,\ldots,X^\delta_n) = (Y^\delta_{k+1},\ldots,Y^\delta_{k+n})\} \tag{3.14}$$

where we make the convention $\inf \varnothing = \infty$. In words, this is the first time that in the process $Y^\delta$, the first $n$ symbols of the process $X^\delta$ are observed. Similarly, if $X'^\delta$ is an independent copy of the process $X^\delta$, we define

$$W_n^\delta(X|X') = \inf\{k \geq 1 : (X_1^\delta, \dots, X_n^\delta) = (X'^\delta_{k+1}, \dots, X'^\delta_{k+n})\}\,. \tag{3.15}$$

In what follows, we will always choose the processes $\{Y_t : t \geq 0\}$ and $\{X_t : t \geq 0\}$, as well as $\{X_t : t \geq 0\}$ and $\{X'_t : t \geq 0\}$ to be independent. However, no independence of $\{Y_t : t \geq 0\}$ and $\{X'_t : t \geq 0\}$ will be required. The joint distribution on path space of these three processes $(X, Y, X')$ will be denoted by $\mathbb{P}_{123}$. Correspondingly the joint law of the $\delta$-discretization of $(X, Y, X')$ will be denoted by $\mathbb{P}^\delta_{123}$.

Our first result is a law of large numbers for the logarithm of the ratio of waiting times, in the limit $n \to \infty$, $\delta \downarrow 0$.

**Theorem 1.** $\mathbb{P}_{123}$-*almost surely:*

$$\lim_{\delta\to 0}\lim_{n\to\infty} \frac{1}{n\delta} \log \frac{W_n^\delta(X|Y)}{W_n^\delta(X|X')} = \mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}})\,. \tag{3.17}$$

Before proving this theorem, we state a theorem from (1) about the exponential approximation for the hitting-time law, which will be the crucial ingredient in the whole of this paper. For an $n$-block $x_1^n := x_1, \dots, x_n \in A^n$ and a continuous-time trajectory $\omega \in A^{[0,\infty)}$, we define the hitting time of the $\delta$-discretization by

$$T^\delta_{x_1^n}(\omega) = \inf\{k \geq 1 : \omega_{k\delta} = x_1, \dots, \omega_{(k+n-1)\delta} = x_n\}\,. \tag{3.18}$$

We then have the following result, see (1).

**Theorem 2.** *For all $\delta > 0$, there exist $\eta_1, \eta_2, C, c, \beta, \kappa \in ]0, \infty[$ such that for all $n \in \mathbb{N}$ and for all $x_1^n \in A^n$, there exists $\eta = \eta(x_1^n)$, with $0 < \eta_1 \leq \eta \leq \eta_2 < \infty$ such that for all $t > 0$*

$$\begin{aligned}\left|\mathbb{P}\left(T^\delta_{x_1^n}(\omega) > \frac{t}{\mathbb{P}^\delta(X_1^n = x_1^n)}\right) - e^{-\eta t}\right| &\leq Ce^{-ct}\left(\mathbb{P}^\delta(X_1^n = x_1^n)\right)^\kappa \\ &\leq Ce^{-ct}e^{-\beta n}\,. \end{aligned}\tag{3.20}$$

*The same theorem holds with $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$.*

The constants appearing in Theorem 2 (except $C$) depend on $\delta$, and more precisely we have $\beta = \beta(\delta) \to 0$, $\eta_1 = \eta_1(\delta) \to 0$ as $\delta \to 0$. We will come back to this later on.

From Theorem 2 we derive (see (2), theorem 2.4):

**Proposition 1.** *For all $\delta > 0$, there exist $\kappa_1, \kappa_2 > 0$ such that*

$$-\kappa_1 \log n \leq \log\left(W_n^\delta(X|Y)\tilde{\mathbb{P}}^\delta(X_1^n)\right) \leq \log(\log n^{\kappa_2}) \quad \mathbb{P}\otimes\tilde{\mathbb{P}} \text{ eventually a.s.}$$

*and*

$$-\kappa_1 \log n \leq \log\left(W_n^\delta(X|X')\mathbb{P}^\delta(X_1^n)\right) \leq \log(\log n^{\kappa_2}) \quad \mathbb{P}\otimes\mathbb{P} \text{ eventually a.s.}$$

With these ingredients we can now give the proof of Theorem 1.

*Proof of Theorem 1.* From Proposition 1 it follows that, for all $\delta > 0$, $\mathbb{P}_{123}$ almost surely

$$\lim_{n\to\infty}\frac{1}{n}\left(\log W_n^\delta(X|Y) - \log W_n^\delta(X|X') + \sum_{i=0}^{n-1}\log p_\delta^Y(X_i,X_{i+1}) - \sum_{i=0}^{n-1}\log p_\delta^X(X_i,X_{i+1})\right) = 0\,. \tag{3.22}$$

By ergodicity of the continuous-time Markov chain $\{X_t : t \geq 0\}$, the discrete Markov chains $X^\delta, Y^\delta$ are also ergodic and therefore we obtain

$$\lim_{n\to\infty}\frac{1}{n}\left(\log W_n^\delta(X|Y) - \log W_n^\delta(X|X') + \sum_{x,y\in A}\mu(x)p_\delta^X(x,y)\log\left(\frac{p_\delta^Y(x,y)}{p_\delta^X(x,y)}\right)\right) = 0\,. \tag{3.23}$$

Using (3.12), (3.13) and $p(x,x) = 0$, this gives

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\Big(\log W_n^\delta(X|Y) - \log W_n^\delta(X|X')\Big)\\
= \;& -\sum_{x\in A}\mu(x)(1-\delta c(x))\log(1-\delta\tilde{c}(x)) - \sum_{x,y\in A}\mu(x)\delta c(x)p(x,y)\log(\delta\tilde{c}(x)p(x,y))\\
+ \;& \sum_{x\in A}\mu(x)(1-\delta c(x))\log(1-\delta c(x)) + \sum_{x,y\in A}\mu(x)\delta c(x)p(x,y)\log(\delta c(x)p(x,y)) + \mathcal{O}(\delta^2)\\
= \;& \delta\left(\sum_{x,y\in A}\mu(x)c(x)p(x,y)\log\frac{c(x)}{\tilde{c}(x)} + \sum_{x\in A}\mu(x)(\tilde{c}(x)-c(x))\right) + \mathcal{O}(\delta^2)\\
= \;& \delta\,\mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) + \mathcal{O}(\delta^2)\,.
\end{aligned} \tag{3.24}$$

where in the last line we used the expression (3.6) for the relative entropy. □

Let us now specify the dependence on $\delta$ of the various constants appearing in Theorem 2. For the lower bound on the parameter we have (see (1), section 5)

$$\eta_1(\delta) \geq \frac{1}{C' + K} \tag{3.25}$$

where $C'$ is a positive number independent of $\delta$ and

$$K = 2\sum_{l=1}^{\infty}\alpha(l) + \sum_{k=1}^{n/2}\sup_{\{x_1^{(n-k)}\}}\mathbb{P}^\delta(X_1^{(n-k)} = x_1^{(n-k)})\,.$$

Here $\alpha(l)$ denotes the classical $\alpha$-mixing coefficient:

$$\alpha(l) = \sup_{j\geq 1}\sup_{S_1\in\mathcal{F}_0^{j-1}, S_2\in\mathcal{F}_{j+l}^\infty}\Big(\mathbb{P}^\delta(S_1\cap S_2) - \mathbb{P}^\delta(S_1)\mathbb{P}^\delta(S_2)\Big)$$

where, for $0 \leq m \leq n < \infty$, $\mathcal{F}_m^n$ denotes sigma-field of subsets of $A^{\mathbb{N}}$ generated by $X_i$ ($m \leq i \leq n$). By the assumption of ergodicity of the continuous Markov chain, the generator $L$ (resp. $\tilde{L}$)

has an eigenvalue 0, the maximum of the real part of the other eigenvalues is strictly negative and denoted by $-\lambda_1 < 0$. One then has

$$\alpha(l) \leq \exp(-\lambda_1 \delta l)\,. \tag{3.26}$$

Using (3.12) there exists $\lambda_2 > 0$ such that

$$\mathbb{P}^\delta(X_1^{(n-k)} = x_1^{(n-k)}) \leq \exp(-\lambda_2 \delta n/2)$$

for $k = 1, \dots, n/2$. Therefore, there exists $\hat{c} > 0$ such that

$$\eta_1(\delta) > \hat{c}\delta\,. \tag{3.27}$$

Similarly, from the proof of Theorem 2.1 in (2) one obtains easily the dependence on $\delta$ of the constants appearing in the error term of (3.20).

$$c = c(\delta) > \gamma_1 \delta, \beta = \beta(\delta) > \gamma_2 \delta \tag{3.28}$$

for some $\gamma_1, \gamma_2 > 0$.

In applications, e.g., in the estimation of the relative entropy from two given sample paths, one would like to choose the word-length $n$ and the discretization $\delta = \delta_n$ *together*. This possibility is precisely provided by the estimates (3.27) and (3.28), as the following analogue of Proposition 1 shows.

**Proposition 2.** *Let $\delta_n \to 0$ as $n \to \infty$, then there exists $\kappa_1, \kappa_2 > 0$*

$$-\kappa_1 \frac{\log n}{\delta_n} \leq \log\left(W_n^{\delta_n}(X|Y)\tilde{\mathbb{P}}^{\delta_n}(X_1^n)\right) \leq \log\left(\frac{\kappa_2 \log n}{\delta_n}\right) \quad \mathbb{P}\otimes\tilde{\mathbb{P}} \text{ eventually a.s.} \tag{3.30}$$

*and*

$$-\kappa_1 \frac{\log n}{\delta_n} \leq \log\left(W_n^{\delta_n}(X|X')\mathbb{P}^{\delta_n}(X_1^n)\right) \leq \log\left(\frac{\kappa_2 \log n}{\delta_n}\right) \quad \mathbb{P}\otimes\mathbb{P} \text{ eventually a.s.}\,.$$

*Proof.* The proof is analogous to the proof of Theorem 2.4 in (2). For the sake of completeness, we prove the upper bound of (3.30). We can assume that $\delta_n \leq 1$. By the exponential approximation (3.20) we have, for all $t > 0$, $n \geq 1$, the estimates

$$\begin{aligned} \mathbb{P}^\delta \otimes \tilde{\mathbb{P}}^\delta\left(\log\left(W_n^\delta(X|Y)\tilde{\mathbb{P}}^\delta(X_1^n)\right) \geq \log t\right) \quad &\leq \quad e^{-\eta(\delta_n)t} + Ce^{-\beta(\delta_n)n}e^{-c(\delta_n)t} \\ &\leq e^{-\eta_1\delta_n t} + Ce^{-\gamma_1\delta_n n}e^{-\gamma_2\delta_n t}\,. \end{aligned} \tag{3.31}$$

Choosing $t = t_n = \frac{\kappa_2 \log n}{\delta_n}$, with $\kappa_2 > 0$ large enough makes the rhs of (3.31) summable and hence a Borel-Cantelli argument gives the upper bound. □

The use of propositions 1, 2 lies in the fact that the logarithm of the waiting time can be well approximated by the logarithm of the probability of an $n$-block under the measure $\mathbb{P}^\delta$, or $\tilde{\mathbb{P}}^\delta$. By the Markov property, the logarithm of this probability is an ergodic sum which is much easier to deal with from the point of view of obtaining law of large numbers, large deviation theorem and central limit behavior.

Of course, whether proposition (2) is still useful, i.e., whether it still gives the law of large numbers with $\delta = \delta_n$ depends on the behavior of ergodic sums such as

$$\sum_{i=1}^{n} f(X_i)$$

under the measure $\mathbb{P}^{\delta_n}$, i.e., the behavior of

$$\sum_{i=1}^{n} f(X_{i\delta_n})$$

under $\mathbb{P}$. This is made precise in the following theorem:

**Theorem 3.** *Suppose that $\delta_n \to 0$ as $n \to \infty$ such that $\frac{\log n}{n\delta_n^2} \to 0$ then in $\mathbb{P}_{123}$ probability:*

$$\lim_{n\to\infty} \frac{1}{n\delta_n} \log \frac{W_n^{\delta_n}(X|Y)}{W_n^{\delta_n}(X|X')} = \mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) \,. \tag{3.33}$$

*Proof.* By Proposition 2 we can write

$$\begin{aligned} &\log W_n^{\delta_n}(X|Y) - \log W_n^{\delta_n}(X|X') \\ = \;& \sum_{i=1}^{n} 1\!\!I_{X_i = X_{i+1}} \log \frac{1-\delta_n c(X_i)}{1-\delta_n \tilde{c}(X_i)} + \sum_{i=1}^{n} 1\!\!I_{X_i \neq X_{i+1}} \log \frac{c(X_i)}{\tilde{c}(X_i)} + \mathcal{O}(\log n/\delta_n) \,. \end{aligned} \tag{3.34}$$

Both sums on the right hand side of (3.34) are of the form

$$\sum_{i=1}^{n} F_{\delta_n}(X_{i\delta_n}, X_{(i+1)\delta_n}) \tag{3.35}$$

with

$$\mathbb{E}(F_{\delta_n} - \mathbb{E}(F_{\delta_n}))^2 \le C\delta_n \tag{3.36}$$

where $C > 0$ is some constant. Now, using ergodicity of the continuous-time Markov chain $\{X_t, t \ge 0\}$, we have the estimate

$$\mathbb{E}\left(\left(F_{\delta_n}(X_{i\delta_n}, X_{(i+1)\delta_n}) - \mathbb{E}(F_{\delta_n})\right)\left(F_{\delta_n}(X_{j\delta_n}, X_{(j+1)\delta_n}) - \mathbb{E}(F_{\delta_n})\right)\right) \le \|F_{\delta_n}\|_2^2 \; e^{-\delta_n \lambda_1 |i-j|} \tag{3.37}$$

with $\lambda_1 > 0$ independent of $n$.
Combining these estimates gives

$$\mathrm{Var}\left(\sum_{i=1}^{n} F_{\delta_n}(X_{i\delta_n}, X_{(i+1)\delta_n})\right) \le Cn\delta_n + \sum_{i=1}^{n} \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} \delta_n e^{-\delta_n\lambda_1|i-j|} \le Cn\delta_n + C'\delta_n \frac{n}{\delta_n} \tag{3.38}$$

where $C' > 0$ is some constant. Therefore,

$$\frac{1}{n^2\delta_n^2} \mathrm{Var}\left(\sum_{i=1}^{n} F_{\delta_n}(X_{i\delta_n}, X_{(i+1)\delta_n})\right) = \mathcal{O}(1/n\delta_n^2) \,. \tag{3.39}$$

Combining (3.34) and (3.39) with the assumption $\frac{\log n}{n\delta_n^2} \to 0$ concludes the proof. $\square$

### 3.3 Large deviations

In this subsection, we study the large deviations of

$$\frac{1}{n}\log\left(\frac{W_n^\delta(X|Y)}{W_n^\delta(X|X')}\right).$$

under the measure $\mathbb{P}_{123}$. More precisely, we compute the scaled-cumulant generating function $\mathcal{F}^\delta(p)$ in the limit $\delta\to 0$ and show that it coincides with the scaled-cumulant generating function for the loarithm of the Radon-Nikodym derivative $d\mathbb{P}^{[0,t]}/d\tilde{\mathbb{P}}^{[0,t]}$ in the range $p\in(-1,1)$. As in the case of waiting times for discrete-time processes, see e.g. (3), the scaled-cumulant generating function is only finite in the interval $(-1,1)$.

We introduce the function

$$\mathcal{E}(p)=\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{P}^{[0,t]}}{d\tilde{\mathbb{P}}^{[0,t]}}\right)^p=$$

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}_{\mathbb{P}}\left(\exp\left(p\left(\int_0^t\log\frac{c(\omega_s)}{\tilde{c}(\omega_s)}\,dN_s(\omega)-\int_0^t(c(\omega_s)-\tilde{c}(\omega_s))\,ds\right)\right)\right). \qquad (3.40)$$

By standard large deviation theory for continuous-time Markov chains (see e.g. (12) p. 107-) this function exists and is the scaled cumulant generating function for the large deviations of

$$\int_0^t\log\frac{c(\omega_s)}{\tilde{c}(\omega_s)}\,dN_s(\omega)-\int_0^t(c(\omega_s)-\tilde{c}(\omega_s))\,ds$$

as $t\to\infty$, i.e., the large deviations of this quantity are governed by the entropy function

$$I(x)=\sup_p(px-\mathcal{E}(p))$$

We can now formulate our large deviation theorem.

**Theorem 4.** *a) For all $p\in\mathbb{R}$ and $\delta>0$ the function*

$$\mathcal{F}^\delta(p):=\lim_{n\to\infty}\frac{1}{n\delta}\log\mathbb{E}_{\mathbb{P}^\delta_{123}}\left(\frac{W_n^\delta(X|Y)}{W_n^\delta(X|X')}\right)^p \qquad (3.42)$$

*exists, is finite in $p\in(-1,1)$ whereas*

$$\mathcal{F}^\delta(p)=\infty\quad\text{for } |p|\geq 1\,.$$

*b) Moreover, as $\delta\to 0$, we have, for all $p\in(-1,1)$:*

$$\mathcal{F}(p):=\lim_{\delta\to 0}\mathcal{F}^\delta(p)=\mathcal{E}(p)\,.$$

The following notion of logarithmic equivalence will be convenient later on.

**Definition 1.** *Two non-negative sequences $a_n$, $b_n$ are called* logarithmically equivalent *(notation $a_n\simeq b_n$) if*

$$\lim_{n\to\infty}\frac{1}{n}(\log a_n-\log b_n)=0\,.$$

*Proof.* To prove Theorem 4, we start with the following lemma.

**Lemma 1.** *1. For all $\delta > 0$ and for $|p| < 1$,*

$$\mathbb{E}_{\mathbb{P}^{\delta}_{123}} \left( \frac{W_n^{\delta}(X|Y)}{W_n^{\delta}(X|X')} \right)^p \simeq \mathbb{E}_{\mathbb{P}^{\delta}} \exp \left( p \sum_{i=0}^{n-1} \log \left( \frac{p_{\delta}^X(X_i, X_{i+1})}{p_{\delta}^Y(X_i, X_{i+1})} \right) \right). \tag{3.45}$$

*2. For $|p| > 1$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\mathbb{P}^{\delta}_{123}} \left( \frac{W_n^{\delta}(X|Y)}{W_n^{\delta}(X|X')} \right)^p = \infty \,. \tag{3.46}$$

*Proof.* The proof is similar to that of Theorem 3 in (3).

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}^{\delta}_{123}} \left( \frac{W_n^{\delta}(X|Y)}{W_n^{\delta}(X|X')} \right)^p \\ = & \sum_{x_1,\ldots,x_n} \mathbb{P}^{\delta}(X_1^n = x_1^n) \left( \frac{\mathbb{P}^{\delta}(X'^n_1 = x_1^n)}{\tilde{\mathbb{P}}^{\delta}(Y_1^n = x_1^n)} \right)^p \mathbb{E}_{\tilde{\mathbb{P}}^{\delta}\otimes\mathbb{P}^{\delta}} \left( \frac{T_{x_1^n}(Y^{\delta})\tilde{\mathbb{P}}^{\delta}(Y_1^n = x_1^n)}{T_{x_1^n}(X'^{\delta})\mathbb{P}^{\delta}(X'^n_1 = x_1^n)} \right)^p \\ = & \sum_{x_1,\ldots,x_n} \mathbb{P}^{\delta}(X_1^n = x_1^n)^{1+p}\, \tilde{\mathbb{P}}^{\delta}(Y_1^n = x_1^n)^{-p}\, \mathbb{E}_{\tilde{\mathbb{P}}^{\delta}\otimes\mathbb{P}^{\delta}} \left( \frac{\xi_n}{\zeta_n} \right)^p \end{aligned}$$

where

$$\xi_n = T_{x_1^n}(Y^{\delta})\, \tilde{\mathbb{P}}^{\delta}(Y_1^n = x_1^n)$$

and

$$\zeta_n = T_{x_1^n}(X'^{\delta})\, \mathbb{P}^{\delta}(X'^n_1 = x_1^n)\,.$$

The random variables $\xi_n, \zeta_n$ have approximately an exponential distribution (in the sense of Theorem 2). Using this fact, we can repeat the arguments of the proof of Theorem 3 in (3) -which uses the exponential law with the error-bound given by Theorem 2- to prove that for $p \in (-1, 1)$

$$0 < C_1 \le \mathbb{E}_{\tilde{\mathbb{P}}^{\delta}\otimes\mathbb{P}^{\delta}} \left( \frac{\xi_n}{\zeta_n} \right)^p \le C_2 < \infty$$

where $C_1, C_2$ do not depend on $n$, whereas for $|p| > 1$,

$$\mathbb{E}_{\tilde{\mathbb{P}}^{\delta}\otimes\mathbb{P}^{\delta}} \left( \frac{\xi_n}{\zeta_n} \right)^p = \infty \,. \tag{3.47}$$

(Remark that up to corrections to the exponential law (details spelled out in proof of Theorem 3 in (3)) this simply comes from the fact that for $\gamma > 0$, the integral

$$\int_0^{\infty} \frac{1}{x^{\gamma}} e^{-x}$$

is convergent if and only if $\gamma < 1$.) Therefore, with the notation of Definition 1, for $|p| < 1$

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}^{\delta}_{123}} \left( \frac{W_n^{\delta}(X|Y)}{W_n^{\delta}(X|X')} \right)^p \\ \simeq & \sum_{x_1,\ldots,x_n} \mathbb{P}^{\delta}(X_1 = x_1, \ldots, X_n = x_n)^{1+p}\, \tilde{\mathbb{P}}^{\delta}(Y_1 = x_1, \ldots, Y_n = x_n)^{-p} \\ = & \mathbb{E}_{\mathbb{P}^{\delta}} \exp \left( p \sum_{i=1}^{n} \log \left( \frac{p_{\delta}^X(X_i, X_{i+1})}{p_{\delta}^Y(X_i, X_{i+1})} \right) \right). \end{aligned} \tag{3.48}$$

and for $|p| > 1$ we obtain (3.46) from (3.47). $\square$

This proves the existence of $\mathcal{F}^\delta(p)$. Indeed, the limit

$$\mathcal{F}^\delta(p) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\mathbb{P}^\delta} \exp\left(p \sum_{i=1}^n \log\left(\frac{p_\delta^X(X_i, X_{i+1})}{p_\delta^Y(X_i, X_{i+1})}\right)\right) \tag{3.49}$$

exists by standard large deviation theory of (discrete-time, finite state space) Markov chains (since $\delta > 0$ is *fixed*), see e.g. (4), section 3.1.1.
In order to deal with the limit $\delta \to 0$ of $\mathcal{F}^\delta(p)$, we expand the expression in the rhs of (3.45), up to order $\delta^2$. This gives

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}^\delta} \exp\left(p \sum_{i=1}^n \log\left(\frac{p_\delta^X(X_i, X_{i+1})}{p_\delta^Y(X_i, X_{i+1})}\right)\right) \\
= \quad & e^{\mathcal{O}(n\delta^2)} \mathbb{E}_{\mathbb{P}^\delta}\left(\exp\left(p \sum_{i=1}^n \log \frac{1\!\!1_{X_i, X_{i+1}} + \delta c(X_i) p(X_i, X_{i+1}) - \delta c(X_i)}{1\!\!1_{X_i, X_{i+1}} + \delta \tilde{c}(X_i) p(X_i, X_{i+1}) - \delta \tilde{c}(X_i)}\right)\right)
\end{aligned}$$

$$\begin{aligned}
= e^{\mathcal{O}(n\delta^2)} \mathbb{E}_{\mathbb{P}^\delta}\Big[ & \exp\Big(p \sum_{i=1}^n \delta 1\!\!1(X_i = X_{i+1}) \ (\tilde{c}(X_i) - c(X_i))\Big) \\
& + \ p \sum_{i=1}^n 1\!\!1(X_i \neq X_{i+1}) \log \frac{c(X_i)}{\tilde{c}(X_i)} \Big]
\end{aligned}$$

$$\begin{aligned}
= e^{\mathcal{O}(n\delta^2)} \mathbb{E}_{\mathbb{P}}\Big[ & \exp\Big(p \sum_{i=1}^n \delta 1\!\!1(X_{i\delta} = X_{(i+1)\delta}) \ (\tilde{c}(X_{i\delta}) - c(X_{i\delta}))\Big) \\
& + \ p \sum_{i=1}^n 1\!\!1(X_{i\delta} \neq X_{(i+1)\delta}) \log \frac{c(X_{i\delta})}{\tilde{c}(X_{i\delta})} \Big].
\end{aligned}$$

Next we prove that for all $K \in \mathbb{R}$

$$\log \mathbb{E}_{\mathbb{P}^\delta}\left[\frac{\exp\left(K \sum_{i=1}^n \left(\delta 1\!\!1(X_{i\delta} = X_{(i+1)\delta})(\tilde{c}(X_{i\delta}) - c(X_{i\delta})) + 1\!\!1(X_{i\delta} \neq X_{(i+1)\delta}) \log \frac{c(X_{i\delta})}{\tilde{c}(X_{i\delta})}\right)\right)}{\exp\left(K \int_0^{n\delta} (\tilde{c}(X_s) - c(X_s)) ds + K \int_0^{n\delta} \log \frac{c(X_s)}{\tilde{c}(X_s)} \ dN_s\right)}\right] \tag{3.50}$$

$= \mathcal{O}(n\delta^2)$.

This will give the result of the theorem by an application of Hölder's inequality, as is shown in lemma 2 below. We first consider the difference

$$A(n, \delta) := \Big| \sum_{i=1}^n 1\!\!1(X_{i\delta} \neq X_{(i+1)\delta}) \log \frac{c(X_{i\delta})}{\tilde{c}(X_{i\delta})} - \int_0^{n\delta} \log \frac{c(X_s)}{\tilde{c}(X_s)} \ dN_s \Big|.$$

If there does not exist an interval $[i\delta, (i+1)\delta[, i \in \{0, \ldots, n-1\}$ where at least two jumps of the process $\{N_t, t \geq 0\}$ occur, then $A(n, \delta) = 0$. Indeed, if there is no jump in $[i\delta, (i+1)\delta[$, both $\mathbb{1}(X_{i\delta} \neq X_{(i+1)\delta}) \log \frac{c(X_{i\delta})}{\tilde{c}(X_{i\delta})}$ and $\int_{i\delta}^{(i+1)\delta} \log \frac{c(X_s)}{\tilde{c}(X_s)} dN_s$ are zero and if there is precisely one jump, then they are equal. Therefore, using the independent increment property of the Poisson process, and the strict positivity of the rates, we have the bound

$$A(n,\delta) \leq C \sum_{i=1}^{n} \mathbb{1}(\chi_i \geq 2)$$

where the $\chi_i$'s, $i = 1, \ldots, n$, form a collection of independent Poisson random variables with parameter $\delta$, and $C$ is some positive constant. This gives

$$\mathbb{E}_{\mathbb{P}^\delta} e^{2KA(n,\delta)} = (\mathcal{O}(\delta^2)e^{2K} + \mathcal{O}(1))^n = \mathcal{O}(e^{n\delta^2}). \tag{3.51}$$

Next, we tackle

$$B(n,\delta) := \Big| \sum_{i=1}^{n} \delta \mathbb{1}(X_{i\delta} = X_{(i+1)\delta})(\tilde{c}(X_{i\delta}) - c(X_{i\delta})) - \int_0^{n\delta} (\tilde{c}(X_s) - c(X_s))ds \Big|. \tag{3.52}$$

If there is no jump in any of the intervals $[i\delta, (i+1)\delta[$, this term is zero. Otherwhise, the contibution of the interval $[i\delta, (i+1)\delta[$ is bounded by $C'\delta$. Therefore it is bounded by

$$B(n,\delta) \leq C'\delta \sum_{i=1}^{n} \mathbb{1}(\chi_i \geq 1)$$

where the $\chi_i$'s, $i = 1, \ldots, n$, form once more a collection of independent Poisson random variables with parameter $\delta$, and $C'$ is some positive constant. This gives

$$\mathbb{E}_{\mathbb{P}} e^{2KB(n,\delta)} \leq (\mathcal{O}(\delta e^{C''\delta}) + 1 - \delta)^n = \mathcal{O}(e^{n\delta^2})$$

where $C''$ is some positive constant. Hence, (3.50) follows by combining (3.51) and (3.52) and using Cauchy-Schwarz inequality. ☐

The following lemma referred to before is easy and standard, but we state and prove it here for the sake of completeness.

**Lemma 2.** *Let $X_n$ and $Y_n$ be two sequences of (real-valued) random variables such that*

$$\lim_{n\to\infty} \frac{1}{a_n} \log \mathbb{E} e^{t(X_n - Y_n)} = 0 \tag{3.54}$$

*for all $t \in \mathbb{R}$, and for some sequence of positive numbers $a_n \uparrow \infty$. Suppose that for all $t \in \mathbb{R}$,*

$$F_X(t) = \limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{E} e^{tX_n} \tag{3.55}$$

*and*

$$F_Y(t) = \limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{E} e^{tY_n} \tag{3.56}$$

*are finite. Then for all $t \in \mathbb{R}$*

$$F_X(t) = F_Y(t) \tag{3.57}$$

*Proof.* Put

$$\epsilon_n = X_n - Y_n$$

then by Hölder's inequality and the assumption (3.54) we have for all $p > 1$,

$$\begin{aligned} F_X(t) &= \limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{E}(e^{tY_n} e^{t(\epsilon_n)}) \\ &\leq \limsup_{n\to\infty} \frac{1}{pa_n} \log \mathbb{E}(e^{ptY_n}) = \frac{1}{p} F_Y(tp) \end{aligned}$$

So we obtain the inequalities

$$F_X(t) \leq \frac{1}{p} F_Y(tp) \leq \frac{1}{pp'} F_X(tpp')$$

for all $p, p' > 1$. Both functions $F_X$ and $F_Y$ are convex and hence continuous. Therefore the result of the lemma follows by taking the limits $p, p' \to 1$. □

The following large deviation result for *fixed* $\delta > 0$ is an application of Theorem 4 and (10).

**Proposition 3.** *For all $\delta > 0$, $\mathcal{F}^\delta$ is real-analytic and convex, and the sequence $\{\log W_n^\delta(X|Y) - \log W_n^\delta(X|X') : n \in \mathbb{N}\}$ satisfies the following large deviation principle: Define the open interval $(c_-, c_+)$, with*

$$c_\pm := \lim_{p\to\pm 1} \frac{d\mathcal{E}^\delta}{dp} < 0$$

*Then, for every interval $J$ such that $J \cap (c_-, c_+) \neq \varnothing$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{123}^\delta \left\{ \frac{1}{n} \log \left( \frac{W_n^\delta(X|Y)}{W_n^\delta(X|X')} \right) \in J \right\} = - \inf_{q \in J\cap(c_-,c_+)} \mathcal{I}^\delta(q)$$

*where $\mathcal{I}^\delta$ is the Legendre transform of $\mathcal{F}^\delta$.*

**Remark 3.** *In the case $\{Y_t : t \geq 0\}$ is the time reversed process of $\{X_t : t \geq 0\}$, the scaled-cumulant generating function function $\mathcal{E}(p)$ satisfies the so-called fluctuation theorem symmetry*

$$\mathcal{E}(p) = \mathcal{E}(-1-p)\,.$$

*The large deviation result of Theorem 4 then gives that the entropy production estimated via waiting times of a discretized version of the process has the same symmetry in its scaled-cumulant generating function for $p \in [0, 1]$.*

### 3.4 Central limit theorem

**Theorem 5.** *For all $\delta > 0$,*

$$\frac{1}{\sqrt{n}} \left( \log \left( \frac{W_n^\delta(X|Y)}{W_n^\delta(X|X')} \right) - n\mathbf{s}(\mathbb{P}|\tilde{\mathbb{P}}) \right)$$

*converges in distribution to a normal law $\mathcal{N}(0, \sigma_\delta^2)$, where*

$$\sigma_\delta^2 = \lim_{n\to\infty} \frac{1}{n} \mathrm{Var} \left( \log \left( \frac{\mathbb{P}^\delta(X_1^n)}{\tilde{\mathbb{P}}^\delta(X_1^n)} \right) \right) .$$

*Moreover*

$$\lim_{\delta\to 0}\frac{1}{\delta^2}\sigma_\delta^2=\theta^2$$

*where*

$$\theta^2=\lim_{t\to\infty}\frac{1}{t}\mathrm{Var}\left(\log\left(\frac{d\mathbb{P}^{[0,t]}}{d\tilde{\mathbb{P}}^{[0,t]}}\right)\right).$$

*Proof.* First we claim that for all $\delta>0$

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\mathbb{P}^\delta_{123}}\left(\log\left(\frac{W_n^\delta(X|Y)}{W_n^\delta(X|X')}\right)-\sum_{i=1}^n\log\frac{p_\delta^X(X_i,X_{i+1})}{p_\delta^Y(X_i,X_{i+1})}\right)^2=0\,. \tag{3.61}$$

This follows from the exponential law, as is shown in (3), proof of Theorem 2.
Equation (3.61) implies that a CLT for $\left(\log W_n^\delta(X|Y)-\log W_n^\delta(X|X')\right)$ is equivalent to a CLT for $\sum_{i=1}^n\log\frac{p_\delta^X(X_i,X_{i+1})}{p_\delta^Y(X_i,X_{i+1})}$ and the variances of the asymptotic normals are equal. For $\delta$ fixed, $\sum_{i=1}^n\log\frac{p_\delta^X(X_i,X_{i+1})}{p_\delta^Y(X_i,X_{i+1})}$ satisfies the CLT (for $\delta>0$ fixed, $X_i$ is a discrete-time ergodic Markov chain), so the only thing left is the claimed limiting behavior for the variance, as $\delta\to 0$.
As in the proof of the large deviation theorem, we first expand up to order $\delta$:

$$\begin{aligned}&\sum_{i=1}^n\log\frac{p_\delta^X(X_i,X_{i+1})}{p_\delta^Y(X_i,X_{i+1})}\\ =\;&\sum_{i=1}^n 1\!\!1(X_i=X_{i+1})\delta(\tilde{c}(X_i)-c(X_i))+\sum_{i=1}^n 1\!\!1(X_i\neq X_{i+1})\log\frac{c(X_i)}{\tilde{c}(X_i)}\\ =:\;&\sum_{i=1}^n(\xi_i^\delta+\zeta_i^\delta)\,.\end{aligned}$$

It is then sufficient to verify that

$$\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n\delta}\sum_{i=1}^n\mathbb{E}\left(\left(\xi_i^\delta-\int_{i\delta}^{(i+1)\delta}(\tilde{c}(X_s)-c(X_s))ds\right)^2+\left(\zeta_i^\delta-\int_{i\delta}^{(i+1)\delta}\log\frac{c(X_s)}{\tilde{c}(X_s)}dN_s\right)^2\right)=0$$

which is an analogous computation with Poisson random variables as the one used in the proof of Theorem 4. □

# 4 Shadowing a given trajectory

In the context of discrete-time processes, besides the study of return and waiting times, one is also interested in the hitting time of a *given pattern*. It turns out that if the process satisfies certain mixing conditions, then the exponential law for the hitting time of a pattern holds for *all* patterns, with a parameter that depends on self-repetitive properties of the pattern, see e.g., (1). This exponential law for *all* patterns can then be used in the context of waiting times, where one chooses a pattern "typical" for measure $\mathbb{P}$ and looks for the hitting time of it in a random sequence with distribution $\mathbb{P}$, or for return times, where the pattern consists of the first $n$ symbols of the process.

In the same spirit, one can consider *a fixed trajectory* and ask for the limiting behavior of the logarithm of the time that one has to wait before seeing this trajectory in a continuous time process. As before, this question can only be asked properly by first discretizing, and then taking the limit of zero discretization $\delta$. Moreover, in the spirit of what preceded, one has to consider differences of logarithms of hitting times in order to find a proper normalizable quantity in the limit of discretization step $\delta \downarrow 0$.

In the previous sections we required the processes $X, Y, X'$ to be Markov. To compare with the other sections, in this section we require $Y, X'$ to be Markov, while $X$ will be replaced by a *fixed* trajectory $\gamma$. We recover in particular the analogous (to Theorem 1) law of large numbers (Theorem 6 below) if we require $\gamma$ to be distributed according to a stationary ergodic (but now not necessarily Markovian) process.

Let $\gamma \in D([0,\infty), X)$ be a given trajectory. The jump process associated to $\gamma$ is defined by

$$N_t(\gamma) = \sum_{0\le s\le t} \mathbb{1}(\gamma_{s^-} \neq \gamma_{s^+})\,.$$

For a given $\delta > 0$, define the "jump times" of the $\delta$-discretization of $\gamma$:

$$\Sigma_n^\delta(\gamma) = \{i \in \{1, \dots, n\} : \gamma_{(i-1)\delta} \neq \gamma_{i\delta}\}\,.$$

For the Markov process $\{X_t, t \ge 0\}$ with generator

$$Lf(x) = \sum_{y\in A} c(x)p(x,y)(f(y) - f(x))$$

define the hitting time

$$T_n^\delta(\gamma|X) = \inf\{k \ge 0 : (X_{k\delta}, \dots, X_{(k+n)\delta}) = (\gamma_0, \gamma_\delta, \dots, \gamma_{n\delta})\}\,.$$

In words this is the first time after which the $\delta$-discretization of the process imitates the $\delta$-discretization of the given trajectory $\gamma$ during $n+1$ time-steps. For fixed $\delta > 0$, the process $\{X_{n\delta}; n \in \mathbb{N}\}$ is an ergodic discrete-time Markov chain for which we can apply the results of (1) for hitting times. More precisely there exist $0 < \Lambda_1 < \Lambda_2 < \infty$ and $C, c, \alpha > 0$ such that for all $\gamma$, $n \in \mathbb{N}$, there exists $\Lambda_1 < \lambda_n^\gamma < \Lambda_2$ such that

$$\left|\mathbb{P}\Big(T_n^\delta(\gamma|X)\mathbb{P}(X_{i\delta} = \gamma_{i\delta}\ , \forall i = 1, \dots, n+1) > t\Big) - e^{-\lambda_n^\gamma t}\right| \le Ce^{-ct}e^{-\alpha n}\,. \tag{4.1}$$

As a consequence of (4.1) we have

**Proposition 4.** *For all $\delta > 0$, there exist $\kappa_1, \kappa_2 > 0$ such that for all $\gamma \in D([0,\infty), X)$, $\mathbb{P}\otimes\tilde{\mathbb{P}}$ eventually almost surely*

$$-\kappa_1 \log n \le \log\Big(T_n^\delta(\gamma|Y)\tilde{\mathbb{P}}(Y_\delta = \gamma_\delta, \dots, Y_{n\delta} = \gamma_{n\delta})\Big) \le \log(\log n^{\kappa_2})$$

*and*

$$-\kappa_1 \log n \le \log\Big(T_n^\delta(\gamma|X)\mathbb{P}(X_\delta = \gamma_\delta, \dots, X_{n\delta} = \gamma_{n\delta})\Big) \le \log(\log n^{\kappa_2})\,.$$

Therefore, for $\delta > 0$ fixed, we arrive at

$$\log T_n^\delta(\gamma|X) = \sum_{i\in\Sigma_n^\delta(\gamma)} \log(\delta c(\gamma_{(i-1)\delta})p(\gamma_{(i-1)\delta},\gamma_{i\delta})) + \sum_{i\in\{1,\ldots,n\}\setminus\Sigma_n^\delta(\gamma)} \log(1-\delta c(\gamma_{(i-1)\delta})) + o(n)\,. \tag{4.3}$$

The presence of the $\log(\delta)$ term in the rhs of (4.3) causes the same problem as we have encountered in Section 2. Therefore, we have to subtract another quantity such that the $\log(\delta)$ term is canceled. In the spirit of what we did with the waiting times, we subtract $\log T_n^\delta(\gamma|Y)$, where $\{Y_t : t \geq 0\}$ is another independent Markov process with generator

$$Lf(x) = \sum_{y\in A} \tilde{c}(x)\tilde{p}(x,y)(f(y)-f(x))\,.$$

We then arrive at

$$\log\frac{T_n^\delta(\gamma|X)}{T_n^\delta(\gamma|X)} = \sum_{i\in\Sigma_n^\delta(\gamma)} \log\frac{c(\gamma_{(i-1)\delta})p(\gamma_{(i-1)\delta},\gamma_{i\delta})}{\tilde{c}(\gamma_{(i-1)\delta})\tilde{p}(\gamma_{(i-1)\delta},\gamma_{i\delta})} + \sum_{i\in\{1,\ldots,n\}\setminus\Sigma_n^\delta(\gamma)} \log\frac{(1-\delta c(\gamma_{(i-1)\delta}))}{(1-\delta\tilde{c}(\gamma_{(i-1)\delta}))} + o(n) \tag{4.4}$$

We then have the following law of large numbers

**Theorem 6.** *Let $\mathbb{P}$ (resp. $\tilde{\mathbb{P}}$) denote the stationary path space measure of $\{X_t : t \geq 0\}$ (resp. $\{Y_t : t \geq 0\}$ and let $\gamma \in D([0,\infty), X)$ be a fixed trajectory. We then have $\mathbb{P}\otimes\tilde{\mathbb{P}}$-almost surely:*

$$\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n\delta}\left(\log\frac{T_n^\delta(\gamma|Y)}{T_n^\delta(\gamma|X)} - \int_0^{n\delta}\log\frac{c(\gamma_s)p(\gamma_{s^-},\gamma_{s^+})}{\tilde{c}(\gamma_s)\tilde{p}(\gamma_{s^-},\gamma_{s^+})}dN_s(\gamma) - \int_0^{n\delta}(\tilde{c}(\gamma_s)-c(\gamma_s))ds\right) = 0\,. \tag{4.6}$$

*Moreover, if $\gamma$ is chosen according to a stationary ergodic measure $\mathbb{Q}$ on path-space, then $\mathbb{Q}$-almost surely*

$$\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n\delta}\left(\log T_n^\delta(\gamma|Y) - \log T_n^\delta(\gamma|X)\right) = \sum_{x,y\in A} q(x,y)\log\frac{c(x)p(x,y)}{\tilde{c}(x)\tilde{p}(x,y)} + \sum_{x\in A} q(x)(\tilde{c}(x)-c(x)) \tag{4.7}$$

*where*

$$\begin{aligned} q(x,y) &= \lim_{t\to\infty}\mathbb{E}_{\mathbb{Q}}\left(\frac{N_t^{xy}}{t}\right) \\ q(x) &= \mathbb{Q}(\gamma_0 = x) \end{aligned}$$

*and where $N_t^{xy}(\gamma)$ denotes the number of jumps from $x$ to $y$ of the trajectory $\gamma$ in the time-interval $[0,t]$.*

*Proof.* Using proposition 4, we use the same proof as that of Theorem 1, and use that the sums in the rhs of (4.4) is up to order $\delta^2$ equal to the integrals appearing in the lhs of (4.6). The other assertions of the theorem follow from the ergodic theorem. □

**Remark 4.** *If we choose $\gamma$ according to the path space measure $\mathbb{P}$, i.e., $\gamma$ is a "typical" trajectory of the process $\{X_t : t \geq 0\}$, and choose $p(x,y) = \tilde{p}(x,y)$, then we recover the limit of the law of large numbers for waiting times (Theorem 1):*

$$\begin{aligned}\lim_{\delta\to 0}\lim_{n\to\infty}\frac{1}{n\delta}\Big(\log T_n^\delta(\gamma|Y) - \log T_n^\delta(\gamma|X)\Big) &= \sum_x \mu(x)c(x)\log\frac{c(x)}{\tilde{c}(x)} + \sum_x \mu(x)(\tilde{c}(x) - c(x)) \\ &= s(\mathbb{P}|\tilde{\mathbb{P}}).\end{aligned}$$

# References


[1] M. Abadi, *Exponential approximation for hitting times in mixing processes*, Math. Phys. Electron. J. **7** (2001).

[2] M. Abadi, J.-R. Chazottes F. Redig and E. Verbitskiy, *Exponential distribution for the occurrence of rare patterns in Gibbsian random fields*, Commun. Math. Phys. **246** no. 2, 269–294 (2004).

[3] J.-R. Chazottes and F. Redig, *Testing the irreversibility of a Gibbsian process via hitting and return times*, Nonlinearity, **18**, 2477–2489 (2005).

[4] A. Dembo and O. Zeitouni, Large deviation techniques and applications, Springer, (1998).

[5] D.-Q. Jiang, M. Qian, M.-P. Qian, *Mathematical theory of nonequilibrium steady states. On the frontier of probability and dynamical systems.* Lecture Notes in Mathematics **1833**, Springer, (2004).

[6] C. Kipnis and C. Landim, *Scaling limits of interacting particle systems*, Springer, Berlin, (1999).

[7] R.S. Lipster and A.N. Shiryayev, *Statistics of random processes*, Volume II, Springer, New York (1977).

[8] C. Maes, *The fluctuation theorem as a Gibbs property*, J. Stat. Phys. **95**, 367-392, (1999).

[9] D. Ornstein, B. Weiss, *Entropy and recurrence rates for stationary random fields*, Special issue on Shannon theory: perspective, trends, and applications. IEEE Trans. Inform. Theory **48**, No. 6, 1694–1697 (2002).

[10] D. Plachky, J. Steinebach, *A theorem about probabilities of large deviations with an application to queuing theory*, Period. Math. Hungar. **6**, no. 4, 343–345 (1975).

[11] P.C. Shields, The ergodic theory of discrete sample paths. Graduate Studies in Mathematics **13**, American Mathematical Society, Providence, RI, (1996).

[12] D. Stroock, An Introduction to the theory of large deviations Springer, (1984).

[13] S.R.S. Varadhan, Large deviations and applications. Philadelphia: Society for Industrial and Applied Mathematics, (1984).

[14] A.J. Wyner, *More on recurrence and waiting times*, Ann. Appl. Probab. **9**, No. 3, 780–796 (1999).