The explicit form of the rate function for semi-Markov processes and its contractions

Yuki Sughiyama$^{1,2}$ and Testuya J Kobayashi$^1$

$^1$ Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo, 153-8505, Japan

E-mail: yuki.sughiyama@gmail.com

Received 15 September 2017, revised 12 January 2018
Accepted for publication 30 January 2018
Published 21 February 2018

Abstract
We derive the explicit form of the rate function for semi-Markov processes. Here, the ‘random time change trick’ plays an essential role. Also, by exploiting the contraction principle of large deviation theory to the explicit form, we show that the fluctuation theorem (Gallavotti–Cohen symmetry) holds for semi-Markov cases. Furthermore, we elucidate that our rate function is an extension of the level 2.5 rate function for Markov processes to semi-Markov cases.

Keywords: large deviation, semi-Markov process, fluctuation theorem

(Some figures may appear in colour only in the online journal)

1. Introduction

The theory for the large deviation property (LDP) is a significant mathematical tool to describe rare events in a sufficiently large system [1–4]. This theory elucidates a behavior of large fluctuations beyond the variance around the convergence value due to the law of large numbers. To be more precise, an exponential decay of rare events caused by expansion of system size is evaluated by the rate function (the large deviation function).

In terms of statistical mechanics [5], the rate function expresses the entropy function, which characterizes thermodynamic properties of many-body systems. In fact, owing to differentiations of the entropy function (the rate function), we can obtain the heat capacity, the equation of state, and so on. Accordingly, one of the main purposes of statistical physics is to calculate an explicit form of the rate function for a given microscopic system. In equilibrium statistical mechanics, by using the LDP, we consider the limit of increase in particle number.
and volume under a constant density; that is, we calculate a ‘spatial’ thermodynamic limit. In particular, we derive the explicit form of the entropy function from a given microscopic Hamiltonian. On the other hand, in nonequilibrium situations, we focus on the LDP for long time average statistics on stochastic processes describing nonequilibrium dynamics [6–8]; that is, we consider a ‘temporal’ thermodynamic limit. Especially, a large deviation for the flow (current) characterizing a nonequilibrium state plays an essential role. A symmetry of the rate function for the flow is known as the fluctuation theorem (FT) [9–12], which leads to many recent developments in nonequilibrium physics [13–16].

Explicit forms of the rate functions for nonequilibrium dynamics were derived by various approaches. As numerical approaches, recently reported, were the biased method using the tilted processes [17, 18] and the cloning technique employing population dynamics [19–21]. At the other extreme, as analytic approaches, the study by Donsker and Varadhan is well-known. In their series of papers [22–25], they revealed an explicit form of the rate function for the pair empirical measures, i.e. empirical flow (jump), on discrete-time Markov processes. Thus, by the contraction [1, 2, 4] of the explicit form, we can evaluate the rate function for statistics with respect to the flow (i.e. the heat flow and the entropy production). Furthermore, an explicit form of the rate function for continuous-time Markov processes was recently derived in several papers [26–31] by using various analytic methods. This rate function is known as the level 2.5 rate function, which describes fluctuations of the empirical occupation and the empirical jump.

Beyond Markov processes, semi-Markov processes have been studied in various fields [32–41]. In contrast to Markov processes, the semi-Markov processes have a memory, because the event (jump) occurrence probability depends on the elapsed time since the last event occurs. As an extension of the LDP for Markov processes, a LDP on semi-Markov processes has been recently studied in mathematics and mathematical physics with rigorous methods [42–46]. In the application side of physics, it has been discussed with a heuristic approach, accompanied by the FT for the semi-Markov cases [39–41]. In contrast to the rigorous studies as represented by the Donsker–Varadhan study and the level 2.5 rate function, however, the explicit form of the rate function for semi-Markov processes has not yet been obtained with a heuristic presentation.

Apart from these theoretical advancements, the semi-Markov processes find more applications outside of physics, because the memory effect is ubiquitous in chemical and biological systems, which typically reflects hidden and unobserved variables. Examples are a single-molecule enzymatic kinetics [35], firing of neurons [36], and cell cycles of various cells [37, 38]; completion of a reaction, firing of a neuron, and replication of a cell strongly depend on the elapsed time since the last reaction, firing, and division, respectively. Among these examples, cell cycle analysis of dividing cells is an important application of semi-Markov models. The kinetics of cell division can be characterized by its stochastic inter-division interval, the probability density of which is dependent on the type of cells we are investigating, e.g. bacteria, cancer cells, immune cells, and iPS cells [47]. Different stochastic models on the inter-division interval without internal state have been applied to statistical analyses of the cell-type specific inter-division interval [48, 49]. However, recent live-imaging technologies have revealed that the types of cells can change within a relatively short time span by the process of development, by stochastic phenotype switching of cells, and by external signals, all of which may accompany modifications of the epigenetic states of the cells [50]. An example is a direct observation of the epigenetic state transitions in budding yeast by using fluorescent protein markers [51]. Such epigenetic changes can accompany changes in the inter-division statistics, which can be best modeled by a semi-Markov process with a discrete Markov variable (see figure 1). While the direct application of semi-Markov models is currently limited to the image analysis of cells [37, 38], time-series data of the epigenetic state and division time [52–54] are becoming more and more accessible by using experimental devices to capture
single cells for a long time. LDP techniques work as the mathematical basis to estimate the state-dependent inter-division statistics from the empirical measures calculated from time-series data [55]. A simple form of the LDP can substantially benefit from such applications by providing a way to estimate various kinds of statistical bias and variance due to the simplicity of models and the limited number of samples.

The purpose of this study is to reveal the explicit form of the rate function for semi-Markov processes by using a heuristic approach, which means that our derivation of the rate function is not rigorous, but familiar and accessible to physicists. Specifically, we derive the rate function for the empirical jump depending on the waiting (sojourn) time. Furthermore, from the explicit form, we rederive the FT for semi-Markov processes by using the contraction principle [1, 2, 4]. Here, the direction time independence (DTI) [20, 35, 41] of the processes plays an important role. Finally, by using a contraction of the explicit form, we show that our rate function can be reduced to the level 2.5 rate function on Markov processes.

This paper is organized as follows. In the next section, we briefly introduce semi-Markov processes and their properties. In section 3, we derive the explicit form of the rate function for semi-Markov processes by employing the ‘random time change trick’ [56, 57]. In section 4, by assuming DTI, we show that the rate function can be decomposed into two parts: the point process and the Markov process parts. Furthermore, we indicate that this decomposition leads to the FT on semi-Markov processes. In section 5, we consider an age representation of the rate function obtained in section 3, which gives an extension of the level 2.5 rate function to semi-Markov cases. Also, we show that, under the Markov assumption, our rate function can be reduced to the ordinary level 2.5 rate function through the contraction principle. Finally, we summarize this study and discuss its potential applications in section 6.

2. Semi-Markov processes (Markov renewal processes)

Before working on the main result, we devote this section to an introduction of semi-Markov processes (Markov renewal processes) [35, 41] and their properties. First, suppose a point
process [58] on a time interval $[0, t]$, $(T):= \{T_i | 1 \leq i \leq n_t\}$; $T_i \in [0, \infty)$ denotes the inter-event interval between $i-1$th and $i$th events; and $T_1$ exceptionally represents the time when the first event occurs. Here, $n_t$ represents the number of events up to the final time $t$, that is, $n_t = \max \{n | \sum_{i=1}^{n} T_i \leq t\}$. Also, we assign a state $X_i \in \Omega$ for each inter-event interval $T_i$, where $\Omega$ is a finite state space. Then, consider a combination of the histories (series) of the inter-event intervals and the states, $(T, X) := \{T_i, X_i | 1 \leq i \leq n_t\}$. In this joint process $(T, X)$, each event represents either jump ($X_{i+1} \neq X_i$) or reset ($X_{i+1} = X_i$) in the state space $\Omega$, and the inter-event interval $T_i$ is interpreted as the waiting (sojourn) time in the state $X_i$ (see figure 1).

If the updating of the joint process $(T, X)$ is conditionally independent of the past history given the current state $X_i$:

$$
\text{Prob} \{X_{i+1} = x; T_i = \tau'|X_1, \ldots, X_i; T_1, \ldots, T_{i-1}\} = \text{Prob}\{X_{i+1} = x; T_i = \tau'|X_i\}, \tag{1}
$$

we call this process a semi-Markov process or a Markov renewal process. The generator of this process

$$
Q(x; \tau'|x') := \text{Prob} \{X_{i+1} = x; T_i = \tau'|X_i = x'\}, \tag{2}
$$

is known as semi-Markov kernel [41]. Note that this kernel describes the transition probability from $x'$ to $x$ after waiting for time $\tau'$ in the state $x'$. Furthermore, we define a waiting time distribution in the state $x'$ as

$$
\pi(\tau'|x') := \sum_{x \in \Omega} Q(x; \tau'|x'). \tag{3}
$$

By using $\pi(\tau'|x')$, the semi-Markov kernel can be decomposed as

$$
Q(x; \tau'|x') = T(x|\tau', x') \pi(\tau'|x'), \tag{4}
$$

where $T(x|\tau', x')$ is given by the definition of the conditional probability: $T(x|\tau', x') := Q(x; \tau'|x')/\pi(\tau'|x')$. Note that $T(x|\tau', x')$ expresses the transition probability from $x'$ to $x$ under the condition that any event occurs at age $\tau'$, where the age means the elapsed time since the last event occurs. Also, $T(x|\tau', x')$ satisfies the property of the transition matrix: $\sum_{x \in \Omega} T(x|\tau', x') = 1$.

Next, to reveal a relationship between $\pi(\tau'|x')$ and the Poisson point process [58], we rewrite it in terms of an event rate. By using $\pi(\tau'|x')$, we can calculate the survival probability $\Pi(a|x)$ up to age $a$ in state $x$, which means the probability that no event occurs up to age $a$:

$$
\Pi(a|x) := 1 - \int_0^a \pi(\tau|x) \, d\tau = \int_a^\infty \pi(\tau|x) \, d\tau. \tag{5}
$$

Owing to this survival probability, we can represent the semi-Markov kernel as

$$
Q(x; \tau'|x') = r(x; \tau', x') \Pi(\tau'|x'), \tag{6}
$$

where $r(x; \tau', x')$ represents the rate that a jump event from $x'$ to $x$ occurs at age $\tau'$, which is known as the hazard function [35]. By taking summation with respect to $x$ in (6) and using (3) and (5), we have a differential equation:

$$
-\pi(\tau'|x') = \frac{d\Pi(\tau'|x')}{d\tau'} = -\left\{ \sum_{x \in \Omega} r(x; \tau', x') \right\} \Pi(\tau'|x'). \tag{7}
$$
the solution of which is represented as

$$\Pi (\tau' | x') = \exp \left[ - \int_0^{\tau'} r (a, x') \, da \right].$$

(8)

Here, we use the initial condition \( \Pi (0, x') = 1 \) and \( r (a, x') \) is defined as \( r (a, x') := \sum_{x \in \Omega} r (x; a, x') \), which represents the rate that a jump event from \( x' \) to an arbitrary state occurs at age \( a \). In this paper, we call \( r (a, x') \) the event rate at age \( a \) in the state \( x' \). Accordingly, by differentiating (8) with respect to \( \tau' \) and using (5), the waiting time distribution \( \pi (\tau' | x') \) can be expressed by the event rate \( r (a, x') \) as

$$\pi (\tau' | x') = r (\tau', x') \exp \left[ - \int_0^{\tau'} r (a, x') \, da \right].$$

(9)

Substituting (9) into (4), we can represent the semi-Markov kernel as

$$Q (x; \tau' | x') = T (x | \tau', x') r (\tau', x') \exp \left[ - \int_0^{\tau'} r (a, x') \, da \right].$$

(10)

By comparing (10) with (6) and (8), we also have \( r (x; \tau', x') = T (x | \tau', x') r (\tau', x') \). In addition, if \( T (x | \tau', x') \) is independent of age \( \tau' \); \( T (x | \tau', x') = T (x | x') \), we say that the semi-Markov process has direction-time independence (DTI) [20, 35, 41].

Finally, we mention the connection to continuous-time Markov processes [59, 60]. If \( r (a; x') \) does not depend on age \( a \): \( r (a; x') = r (x') \) and \( T \) satisfies DTI, then \( Q (x; \tau' | x') \) can be written as

$$Q (x; \tau' | x') = T (x | x') r (x') e^{-r (x') \tau'}.$$  

(11)

This fact represents that the semi-Markov process is reduced to a continuous-time Markov process with transition rate \( \omega (x | x') := T (x | x') r (x') \), because any jump event occurs as Poissonian with the event rate \( \omega (x) := \sum_{x' \in \Omega} \omega (x | x) = r (x) \).

3. The LDP on semi-Markov processes

We consider the LDP on the semi-Markov process \((T, X)\). In this section, we deal with the following empirical measure for a triplet \((x; \tau', x')\):

$$j_{\varepsilon} (x; \tau', x') := \frac{1}{\varepsilon} \sum_{k=1}^{n_{\varepsilon}} \delta_{X_{k-1}} \delta (\tau' - T_k) \delta_{x'; X_k}.$$  

(12)

which represents how many times a jump (reset) event from \( x' \) to \( x \) at age \( \tau' \) occurs in a realization \((T, X)\). Note that this empirical triplet depends on the realization \((T, X)\), but we abbreviate it from the notation of \( j_{\varepsilon} (x; \tau', x') \) for simplicity. Here, we also assume a periodic condition \( X_{n+1} = X_1 \), which means that we count the number of jump events by regarding the state at time \( t \), \( X_{n+1} \), as \( X_1 \), even though it is not \( X_1 \) in an actual realization \((T, X)\). However, this assumption does not restrict generality of the LDP, since the boundary condition is not effective in the calculation of the empirical measure for \( t \to \infty \). Owing to this assumption, the empirical measure satisfies so-called shift-invariant property [1, 2, 4]:

$$\sum_{x \in \Omega} \int_0^\infty d\tau' j_{\varepsilon} (x; \tau', x') = \sum_{x \in \Omega} \int_0^\infty d\tau' j_{\varepsilon} (x'; \tau', x).$$

(13)
Also, we introduce a marginal measure:

\[ g_{e}(\tau', x') := \sum_{x \in \Omega} j_{e}(x; \tau', x'), \]

which quantifies how often any event at age \( \tau' \) in state \( x' \) occurs in the realization \((T, X)\). Furthermore, the following property is useful:

\[ \frac{n_{t}}{t} = \sum_{x, x' \in \Omega} \int_{0}^{\infty} d\tau' j_{e}(x; \tau', x') = \sum_{x' \in \Omega} \int_{0}^{\infty} d\tau' g_{e}(\tau', x'), \tag{15} \]

which represents the number of events per unit time. Finally, we shall note a normalization relation of \( j_{e}(x; \tau', x') \). By using the definition of \( j_{e}(x; \tau', x') \), (12), we obtain

\[ \sum_{x, x' \in \Omega} \int_{0}^{\infty} d\tau' \tau' j_{e}(x; \tau', x') = \frac{1}{t} \sum_{i=1}^{n_{t}} T_{i}. \tag{16} \]

If we assume that \( t \approx \sum_{i=1}^{n_{t}} T_{i} \) for \( t \to \infty \), which means that we can ignore the elapsed time after the final event occurs, we find the normalization relation at \( t \to \infty \) as

\[ 1 = \sum_{x, x' \in \Omega} \int_{0}^{\infty} d\tau' \tau' j_{e}(x; \tau', x') = \sum_{x' \in \Omega} \int_{0}^{\infty} d\tau' \tau' g_{e}(\tau', x'). \tag{17} \]

In this section, we reveal the explicit form of the rate function for the empirical triplet \( j_{e}(x; \tau', x') \), which is defined as

\[ I[j(x; \tau', x')] := \lim_{t \to \infty} -\frac{1}{t} \log \Prob{j_{e}(x; \tau', x') \approx j(x; \tau', x')} \]. \tag{18} \]

To calculate the rate function, we employ the following two steps: (i) We regard the semi-Markov process as a 2-dimensional Markov process. Then, we calculate the rate function for the pair empirical measure for the 2-dimensional Markov process [1, 2, 4] and its contracted one. (ii) By using the ‘random time change trick’ [56, 57], we obtain the rate function for the empirical triplet \( j_{e}(x; \tau', x') \). We describe the above two steps in the next two subsections.

### 3.1. The LDP on 2-dimensional Markov processes

Suppose a 2-dimensional discrete-time Markov process \((T, X) = \{T_{i}, X_{i}\}\) with a transition probability \( M(\tau, x|\tau', x') := \pi(\tau|x) \pi(x|\tau', x') \), which is equivalent to the semi-Markov process introduced in the previous section (see figure 2). Note that, in terms of this Markov process, \( T = \{T_{i}\} \) is a just discrete-time state sequence, i.e. \( T_{i} \) does not represent the waiting time.

Let us consider the 2-dimensional pair empirical measure:
where we again assume periodic conditions $T_{n+1} = T_1$ and $X_{n+1} = X_1$. Here, we note that this pair empirical measure is normalized by the number of events $n$, differently from $J_e(x; \tau', x')$ in (12); thus $\sum_{x, x' \in \Omega} \int_0^\infty d\tau' J_e(\tau, x; \tau', x') = 1$. The shift-invariant property is also satisfied:

$$
\sum_{x \in \Omega} \int_0^\infty d\tau J_e(\tau, x; \tau', x') = \sum_{x \in \Omega} \int_0^\infty d\tau J_e(\tau', x'; \tau, x) =: G_e(\tau'; x').
$$

If the transition probability $M(\tau, x|\tau', x')$ has ergodicity on the 2-dimensional space $(\tau, x)$, by using Sanov’s theorem for Markov processes [1, 2, 4], the rate function of the pair empirical measure is evaluated as

$$
\bar{I}[J(\tau, x; \tau', x')] = \sum_{x, x' \in \Omega} \int_0^\infty \int_0^\infty d\tau' J(\tau, x; \tau', x') \times \log \frac{J(\tau, x; \tau', x')}{M(\tau, x|\tau', x')} G(\tau'; x').
$$

A brief derivation of this rate function is shown in appendix A, and the well-known rigorous proof is in [1, 2]. In this study, we only deal with cases where $M(\tau, x|\tau', x') = \pi(\tau|x) \mathbb{T}(x|x', x')$ satisfies ergodicity (i.e. $\pi$ and $\mathbb{T}$ are restricted). By substituting the definition of $M(\tau, x|\tau', x')$ into (21), we can rewrite the rate function as

$$
\bar{I}[J(\tau, x; \tau', x')] = \sum_{x, x' \in \Omega} \int_0^\infty \int_0^\infty d\tau' J(\tau, x; \tau', x') \times \log \frac{J(\tau, x; \tau', x')}{\pi(\tau|x) \mathbb{T}(x|x', x')} G(\tau'; x')
$$

where we use the representation of the semi-Markov kernel (4). In the second equality in (22), we also used the equality,

$$
\sum_{x, x' \in \Omega} \int_0^\infty d\tau' J(\tau, x; \tau', x') \log \frac{1}{\pi(\tau|x)} = \sum_{x, x' \in \Omega} \int_0^\infty d\tau' J(\tau, x; \tau', x') \log \frac{1}{\pi(\tau'|x')},
$$

which is derived by the shift-invariant property (20).
Next, we consider the following useful decomposition of the joint probability $J (\tau; x; \tau', x')$ for contraction:

$$J (\tau; x; \tau', x') = J (\tau | x; \tau', x') J (x; \tau', x') ,$$

(24)

where $J (x; \tau', x')$ is defined as $J (x; \tau', x') = \int_0^\infty d\tau J (\tau; x; \tau', x')$ and the conditional probability $J (\tau | x; \tau', x')$ is given by the definition of the conditional probability:

$$J (\tau | x; \tau', x') := J (\tau; x; \tau', x') / J (x; \tau', x').$$

Also, we decompose $G (\tau'; x')$ as

$$G (\tau'; x') = G (\tau' | x') G (x') ,$$

(25)

where we again use $G (\tau' | x') := G (\tau' ; x') / G (x')$. By substituting (24) and (25) into (22), we have

$$\tilde{I} [J (\tau; x; \tau', x')] = \sum_{x,x' \in \Omega} \int_0^\infty d\tau' J (x; \tau', x') \log \frac{J (x; \tau', x')}{G (x; \tau' | x') G (x')}$$

$$+ \sum_{x,x' \in \Omega} \int_0^\infty d\tau' J (x; \tau', x') \int_0^\infty d\tau J (\tau | x; \tau', x') \log \frac{J (\tau | x; \tau', x')}{G (\tau | x)},$$

(26)

where, in the second line, we change the argument of $G$ from $G (\tau' | x')$ to $G (\tau | x)$ by using the shift-invariant property of $J (\tau; x; \tau', x')$ as in (23). By employing the above preparation, we calculate the rate function of the empirical measure for the triplet $(x; \tau', x')$,

$$J_r (x; \tau', x') := \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \delta (\tau' - T_i) \delta_{x_i} = \int_0^\infty d\tau J_r (\tau; x; \tau', x'),$$

(27)

which measures how often a triplet $(x; \tau', x')$ appears in a realization $(T, X)$. Owing to the contraction principle of the LDP [1, 2, 4], the rate function is evaluated as

$$\tilde{I} [J (\tau; x; \tau', x')] = \min_{J (\tau; x; \tau', x')} \left\{ I [J (\tau; x; \tau', x')] \mid \int_0^\infty d\tau J (\tau; x; \tau', x') = J (x; \tau', x') \right\} .$$

(28)

Since $J (x; \tau', x')$ is fixed in the minimization, we can only change the conditional probability $J (\tau | x; \tau', x')$ for the minimization, see (24). If we can choose $J (\tau | x; \tau', x')$ as $G (\tau | x)$, the second term of (26) can be eliminated; and this choice gives a minimizer for (28), because the second term is represented by the relative entropy form for probability densities with respect to $\tau$. Thus, we have $\tilde{I} [J (\tau; x; \tau', x')]$ as

$$\tilde{I} [J (\tau; x; \tau', x')] = \sum_{x,x' \in \Omega} \int_0^\infty d\tau' J (x; \tau', x') \log \frac{J (x; \tau', x')}{G (x; \tau' | x') G (x')},$$

(29)

where we note that the following shift-invariant property is satisfied:

$$\sum_{x \in \Omega} \int_0^\infty d\tau' J (x; \tau', x') = \sum_{x' \in \Omega} \int_0^\infty d\tau' J (x'; \tau', x') = G (x'),$$

(30)

where we use (20). By using the remaining part of this subsection, we prove the reason why we can choose $J (\tau | x; \tau', x')$ as $G (\tau | x)$. Due to the shift-invariant property (20), the conditional probability $J (\tau | x; \tau', x')$ must satisfy
\[ \sum_{x \in \Omega} \int_0^\infty d\tau J(\tau|x, \tau', x') J(x, \tau', x) = \sum_{x \in \Omega} \int_0^\infty d\tau J(\tau'|x; x) J(x', \tau, x). \]  

(31)

In order to choose \( J(\tau|x, \tau', x') = G(\tau|x) \), we have to verify that the following equality holds:

\[ \sum_{x \in \Omega} \int_0^\infty d\tau G(\tau|x) J(x, \tau', x') = \sum_{x \in \Omega} \int_0^\infty d\tau G(\tau'|x') J(x'; \tau, x). \]  

(32)

The left hand side is calculated as \( \Sigma_{x \in \Omega} J(x; \tau', x') = G(\tau', x') \) by the definition of \( G(\tau', x'), (20) \); on the other hand, the right hand side is evaluated as \( G(\tau'|x') G(x') = G(\tau', x') \). Accordingly, we can choose \( J(\tau|x, \tau', x') = G(\tau|x) \), and therefore the rate function of the empirical triplet \((x; \tau', x')\) can be represented by (29).

### 3.2. Random time change

As shown in the previous subsection, the rate function of \( J_\epsilon(x; \tau', x') \) is given by (29). However, since \( J_\epsilon(x; \tau', x') \) is normalized by the number of events \( n \), we need to change \( J_\epsilon(x; \tau', x') \) to \( j_\epsilon(x; \tau', x') \), the latter of which is normalized by time \( t \), see (12). In this subsection, we consider a scaling of \( \frac{1}{t} J(x; \tau', x') \).

We begin with the definition of the time-normalized rate function (18). By substituting (12) into (18), we get

\[ I[j(x; \tau', x')] := \lim_{t \to \infty} -\frac{1}{t} \log \text{Prob}\{ j_\epsilon(x; \tau', x') \approx j(x; \tau', x') \} \]

\[ = \lim_{t \to \infty} -\frac{1}{t} \log \text{Prob}\left\{ \frac{1}{t} \sum_{i=1}^n \delta_{X_i, x'} \delta (\tau' - T_i) \delta_{x', x} \approx j(x; \tau', x') \right\}. \]  

(33)

By using (15), we have

\[ I[j(x; \tau', x')] = \lim_{t \to \infty} -\frac{1}{t} \log \text{Prob}\left\{ \frac{1}{t} \sum_{i=1}^n \delta_{X_i, x'} \delta (\tau' - T_i) \delta_{x', x} \approx j(x; \tau', x') \right\}. \]  

(34)

Dividing both sides of the equation in \( \text{Prob}\{ \cdot \} \) by \( \sum_{i'} \int_0^\infty d\tau' j(x; \tau', x') \), we reach

\[ I[j(x; \tau', x')] = \lim_{t \to \infty} -\frac{1}{t} \sum_{i'} \int_0^\infty d\tau' j(x; \tau', x') \times \log \text{Prob}\left\{ \frac{1}{t} \sum_{i=1}^n \delta_{X_i, x'} \delta (\tau' - T_i) \delta_{x', x} \approx \frac{1}{t} \sum_{i'} \int_0^\infty d\tau' j(x; \tau', x') \right\} \]

\[ \times \sum_{i=1}^n \delta_{X_i, x'} \delta (\tau' - T_i) \delta_{x', x} \approx \frac{j(x; \tau', x')}{\sum_{i'} \int_0^\infty d\tau' j(x; \tau', x')}, \]  

(35)
where we insert \( \Sigma_{x,x'\in\Omega} \int_0^\infty \! \mathrm{d}t' j(\tau', x'; x') / \Sigma_{x,x'\in\Omega} \int_0^\infty \! \mathrm{d}t' j(\tau', x') \) = 1 before the symbol ‘log’. By defining \( n := \Sigma_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x') \) and changing the limit \( t \to \infty \) to \( n \to \infty \), we have
\[
I[j(\tau', x') = \left[ \sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x') \right] \lim_{n \to \infty} - \frac{1}{n} \log \left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda x_i} \delta(\tau' - T_i) \delta_{\lambda x_i} \approx \frac{j(\tau', x')}{\sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x')} \right\}. \tag{36}
\]
Furthermore, from (27), we obtain
\[
I[j(\tau', x')] = \left[ \sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x') \right] \lim_{n \to \infty} - \frac{1}{n} \log \left\{ \frac{j(\tau', x')}{\sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x')} \right\}. \tag{37}
\]
Finally, taking the definition of the rate function \( I[j(\tau', x')] \) into account, we find the scaled equation:
\[
I[j(\tau', x')] = \sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x') \times \bar{I} \left[ \frac{j(\tau', x')}{\sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x')} \right]. \tag{38}
\]
Accordingly, by substituting the explicit form of the rate function \( \bar{I} \), (29), into (38), we obtain the rate function of the empirical triplet \( j_{\bar{e}}(\tau; \tau', x') \) as
\[
I[j(\tau', x')] = \sum_{x,x'\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(\tau', x') \log \frac{j(\tau', x')}{Q(\tau; \tau' | x') g(x')}, \tag{39}
\]
where \( g(x') \) is defined by the shift-invariant property (13) as
\[
g(x') := \int_0^{\infty} \! \mathrm{d}t' g(t', x') = \sum_{x\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(x; \tau', x') = \sum_{x\in\Omega} \int_0^{\infty} \! \mathrm{d}t' j(x'; \tau', x). \tag{40}
\]
This explicit form of the rate function (39) constitutes the foundation of our study. In the remaining part of this paper, we will derive various important rate functions by applying a contraction for this explicit form.

The scaling trick used in this subsection is known as ‘random time change’. Although we only give a brief procedure of the random time change in this paper, its rigorous proof is shown in [56, 57].

4. DTI semi-Markov processes and the fluctuation theorem

In this section, we consider the LDP on a semi-Markov process with DTI, \( \mathbb{T}(x | \tau', x') = \mathbb{T}(x | \tau') \).
In this case, by employing the contraction principle, we can obtain an explicit form of the rate function for the following two empirical measures:

\[
J. Phys. A: Math. Theor. 51 (2018) 125001
\]

Y Sugihyama and T J Kobayashi
\[ g_e(\tau', x') = \sum_{x \in \Omega} j_e(x; \tau', x') = \frac{1}{t} \sum_{j=1}^{ \tilde{t}} \delta(\tau' - T_j) \delta_{x', x}, \] 
\[ c_e(x; x') := \int_{0}^{\infty} d\tau' j_e(x; \tau', x') = \frac{1}{t} \sum_{j=1}^{ \tilde{t}} \delta_{x, \pi(x')} \delta_{x', x}, \]

where \( c_e(x; x') \) measures how often a jump (reset) from \( x' \) to \( x \) occurs in the realization \((T, X)\) and satisfies the shift-invariant property:
\[ \sum_{x \in \Omega} c_e(x; x') = \sum_{x \in \Omega} c_e(x'; x) = g_e(x'). \]

From these empirical measures, we can see that the rate function is composed of two parts: rate functions of point processes and Markov processes. Furthermore, by using the explicit form obtained, we show that the fluctuation theorem (Gallavotti–Cohen symmetry) \([9, 10]\) holds even for DTI semi-Markov cases.

### 4.1. Rate function for DTI semi-Markov processes

We start with the rate function \((39)\). Since we now consider the DTI case, we substitute \( Q(\tau; \tau'|x') = \tau(x'|x) \pi(\tau'|x') \) into \((39)\); then we have
\[ I[j(x; \tau', x')] = \sum_{x, x' \in \Omega} \int_{0}^{\infty} d\tau' j(x; \tau', x') \log \frac{j(x; \tau', x')}{\pi(\tau'|x') \pi(\tau'|x') g(x')} \quad \text{for} \quad x, x' \in \Omega. \]

To prepare for the following calculations, we introduce decompositions:
\[ j(x; \tau', x') = j(\tau'|x') c(x; x'), \]
\[ g(\tau', x') = g(\tau'|x') g(x'), \]

where \( j(\tau'|x') \) and \( g(\tau'|x') \) are conditional measures. Substituting \((45)\) and \((46)\) into \((44)\), we get
\[ I[j(x; \tau', x')] = \sum_{x, x' \in \Omega} \int_{0}^{\infty} d\tau' g(\tau', x') \log \frac{g(\tau', x')}{\pi(\tau'|x') g(x')} \\
+ \sum_{x, x' \in \Omega} c(x; x') \log \frac{c(x; x')}{\pi(\tau'|x') g(x')} \\
+ \sum_{x, x' \in \Omega} c(x; x') \int_{0}^{\infty} d\tau' j(\tau'|x; x') \log \frac{j(\tau'|x; x')}{g(\tau'|x')} \quad \text{for} \quad x, x' \in \Omega. \]

Now, the rate function for the empirical measures \( g_e(\tau', x') \) and \( c_e(x; x') \) is given by the contraction principle as
\[ I[g(\tau', x'), c(x; x')] = \min_{j(x; \tau', x')} \left\{ I[j(x; \tau', x')] \right\} \quad \text{for} \quad x, x' \in \Omega \]
\[ \int_{0}^{\infty} d\tau' j(x; \tau', x') = c(x; x'). \]

\[ \text{for} \quad x, x' \in \Omega. \]
Since \( g(\tau, x) \) and \( c(x; x') \) are fixed in the minimization, we can only sweep the conditional measure \( j(\tau' | x'; x') \) (see (45)) under a constraint that the following equation holds:

\[
\sum_{x \in \Omega} j(\tau' | x'; x') c(x; x') = g(\tau', x') .
\]

(49)

Since the third term in (47) is represented by the relative entropy form, \( j(\tau' | x; x') = g(\tau' | x') \) becomes a minimizer of (48), if the constraint (49) holds. Actually, this choice \( j(\tau' | x; x') = g(\tau' | x') \) satisfies (49), because we can have

\[
\sum_{x \in \Omega} g(\tau' | x') c(x; x') = g(\tau' | x') g(x') = g(\tau', x') .
\]

(50)

Therefore, by substituting \( j(\tau' | x; x') = g(\tau' | x') \) into (47), we obtain the rate function for \( g(\tau, x) \) and \( c(x; x') \) as

\[
I[ g(\tau, x), c(x; x') ] = \sum_{x \in \Omega} \int_0^\infty d\tau \ g(\tau, x) \log \frac{g(\tau, x)}{\pi(\tau | x) g(x)}
+ \sum_{x, x' \in \Omega} c(x; x') \log \frac{c(x; x')}{T(x|x')} g(x') ,
\]

(51)

where we change the summation index and the integration variable in the first term from \( (\tau', x') \) to \( (\tau, x) \). The rate function (51) is composed of two terms. The first term describes the rate function on point processes, which determines the inter-event interval. An explanation of point processes and their LDP is shown in appendix B. On the other hand, the second term represents the rate function for Markov jump processes, which is the same form as the rate function for the pair empirical measure on discreet-time Markov processes, (A.17), in appendix A. Owing to this explicit form, we can find the fluctuation theorem for DTI semi-Markov processes as in the next subsection.

Before closing this subsection, we mention a rigorous proof of the LDP on DTI semi-Markov processes. The explicit form of the rate function for DTI cases is rigorously proved by Mariani and Zambotti [46]. Our rate function (51) corresponds to a simplified version of their result, i.e. equation (1.10) in [46]. However, instead of the empirical measure \( g_e(\tau, x) \) we use, they focus on the following empirical measure:

\[
\hat{g}_e(\tau, x) := \tau g_e(\tau, x) = \frac{1}{\tau} \sum_{i=1}^n T_i \delta(\tau - T_i) \delta_{x_i} .
\]

(52)

Therefore, by substituting \( g(\tau, x) = \hat{g}(\tau, x) / \tau \) into (51), we can obtain their rate function (1.10)\(^3\).

4.2. Fluctuation theorem (Gallavotti–Cohen symmetry)

Various significant developments in statistical physics have recently been brought by the FT [9–16], which describes the time reversal symmetry of entropy production. (To be more precise, it originally expresses the symmetry of the entropy flow (current); however, in a non-equilibrium stationary situation, the entropy flow is equivalent to the entropy production.)

\(^3\) More precisely, our rate function (51) differs slightly from theirs. The difference is in a treatment of \( g_e(\tau, x) \) at \( \tau \to \infty \). While we implicitly assume \( \lim_{\tau \to \infty} g_e(\tau, x) = 0 \) in our heuristic approach, they did not. Due to that, some additional term for the limit \( \tau \to \infty \) appears in their rate function (1.10).
Particularly, in terms of the LDP, the FT appears as a symmetry of the rate function for entropy production, which is called Gallavotti–Cohen symmetry (GCS) [9, 10]. Although many studies concern the FT on Markov processes, some recent studies elucidate that FT can be extended to the case of DTI semi-Markov processes [35, 39, 40, 41]. In this subsection, by using the explicit form (51), we show that the GCS holds on DTI semi-Markov processes; which was originally proved by Maes et al [41] by using a different approach from ours.

According to several studies [35, 41] treating the FT on semi-Markov processes, under the DTI assumption, the entropy production (flow) associated with a jump from $x'$ to $x$ is represented as

$$\Sigma (x; x') := \log \frac{\mathbb{T} (x|x') \theta (x')}{\mathbb{T} (x'|x) \theta (x)},$$

(53)

where $\theta (x)$ describes the effective escape rate from state $x$, which is defined by

$$\frac{1}{\theta (x)} := \int_0^\infty d\tau \pi (\tau |x) = \int_0^\infty d\tau \Pi (\tau |x).$$

(54)

Here, we use

$$\int_0^\infty d\tau \Pi (\tau |x) = \int_0^\infty d\tau \int_0^\infty dt \pi (t|x) = \int_0^\infty dt \tau \pi (t|x).$$

(55)

Equations (53) and (54) respectively indicate extensions of the detailed fluctuation theorem (local detailed balance) and the escape rate to semi-Markov cases. If we assume the Markov condition $r (a, x) = r (x)$, the integration in (54) is calculated as

$$\int_0^\infty d\tau \Pi (\tau |x) = \int_0^\infty d\tau e^{-r (x) \tau} = \frac{1}{r (x)}.$$  

(56)

where we use (8). Therefore, $\theta (x)$ can be reduced to the ordinary escape rate of Markov processes, $\theta (x) = r (x)$ ($= \omega (x)$). Furthermore, recalling that $\mathbb{T} (x|x') r (x')$ expresses the transition rate of Markov processes, $\omega (x|x') := \mathbb{T} (x|x') r (x')$, we find that equation (53) is reduced to the well-known detailed fluctuation theorem on Markov processes.

Consider the time-averaged entropy production rate on a sufficient long path $(T, X)$:

$$\sigma_e := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^n \Sigma (X_{i+1}, X_i) = \sum_{x,x' \in \Omega} c_e (x; x') \Sigma (x, x').$$

(57)

Since $c_e (x; x')$ has the shift-invariant property (43), we get

$$\sigma_e = \sum_{x,x' \in \Omega} c_e (x; x') \log \frac{\mathbb{T} (x|x') \theta (x')}{\mathbb{T} (x'|x) \theta (x)} = \sum_{x,x' \in \Omega} c_e (x; x') \log \frac{\mathbb{T} (x|x')}{\mathbb{T} (x'|x)}.$$  

(58)

We now investigate the symmetry of the rate function for $\sigma_e$. To do that, we firstly elucidate a relationship between $I [g (\tau, x), c (x, x')]$ and $I [g (\tau, x), c (x'; x)]$. For notational simplicity, here we write the transpose matrix of $c (x, x')$ as $c (x'; x)$. By using (51), we have
\[
I [g(\tau, x), c(x', x)] = \sum_{x \in \Omega} \int_0^\infty \! d\tau \, g(\tau, x) \log \frac{g(\tau, x)}{\pi(\tau | x) g(x)} \\
+ \sum_{x, x' \in \Omega} c(x', x) \log \frac{c(x', x)}{T(x|x') g(x')}
\]
= \sum_{x \in \Omega} \int_0^\infty \! d\tau \, g(\tau, x) \log \frac{g(\tau, x)}{\pi(\tau | x) g(x)} \\
+ \sum_{x, x' \in \Omega} c(x, x') \log \frac{c(x, x')}{T(x|x') g(x')}
\]
= I [g(\tau, x), c(x, x')] + \sigma, \quad \text{(59)}

where we use the shift-invariant property and change of the summation index to have the second equality. Finally, by employing the contraction principle, we obtain
\[
\min_{g(\tau, x), c(x, x')} \left\{ I [g(\tau, x), c(x', x)] \mid \sum_{x, x' \in \Omega} c(x, x') \log \frac{T(x|x')}{T(x'|x)} = \sigma \right\}
\]
= \min_{g(\tau, x), c(x, x')} \left\{ I [g(\tau, x), c(x, x')] \mid \sum_{x, x' \in \Omega} c(x, x') \log \frac{T(x|x')}{T(x'|x)} = \sigma \right\} + \sigma. \quad \text{(60)}

Accordingly, we find the GCS:
\[
I (-\sigma) = I (\sigma) + \sigma. \quad \text{(61)}
\]

5. Contraction to the level 2.5 rate function

The fluctuation of current (flow) plays an essential role to characterize nonequilibrium states. For continuous-time Markov jump processes, the explicit form of the joint rate function for the empirical occupation and the empirical jump (reset) has been revealed as
\[
I [\mu(x), c(x', x)] = \sum_{x, x' \in \Omega} \{ \omega(x|x') \mu(x') - c(x, x') \}
\]
+ \sum_{x, x' \in \Omega} c(x, x') \log \frac{c(x, x')}{\omega(x|x') \mu(x')} \quad \text{(62)}

where \( \omega(x|x') \) denotes the transition rate of the Markov process from the state \( x' \) to \( x \); \( \mu(x) \) and \( c(x, x') \) represent the occupation of the state \( x \) and the jump from \( x' \) to \( x \), respectively (also see (69) and (42)). This explicit form describes fluctuation of any thermodynamic quantities concerned with the current (e.g. heat flow and entropy production) on Markov processes through the contraction principle. This rate function (62) is known as the level 2.5 rate function and is derived by various methods. A rigorous proof was given by Bertini et al [29, 30], and a more familiar (heuristic) derivation for physicists was made by Barato and Chetrite by using a tilting or spectral technique [31]. In this section, under the DTI assumption, we derive
an extension of the level 2.5 rate function to semi-Markov cases. Furthermore, from its contraction, we rederive the ordinary level 2.5 rate function on Markov processes, (62).

5.1. The level 2.5 rate function for DTI semi-Markov processes

We introduce an age representation of the rate function (51), which is to express an extension of the level 2.5 rate function to DTI semi-Markov processes. Let us change representation of the semi-Markov process. While we described the semi-Markov process by using the inter-event interval (waiting time) as \((T, X)\) up to the previous section, in this section, we represent the same process by employing time series for age \(A_t\), instead of the inter-event interval, as \((A, X) = \{A_t, X_t\}\) (see figure 3). That is, \((A, X)\) share the same dynamics (probability laws) with \((T, X)\); the difference between them is only representation. To be more precise, an age at time \(t\), \(A_t\), is represented by using the inter-event intervals up to time \(t\), \(\{T_i\}_{i=1}^n\), as

\[
A_t = t - \sum_{i=1}^{n} T_i,
\]

because the age means the elapsed time since the last event occurs.

Consider an empirical occupation:

\[
\mu_e(a, x) := \frac{1}{t} \int_0^t \delta(t - A_t) \delta(x - X_t) \, dt,
\]

which represents how often the set \((a, x)\) appears in the realization \((A, X)\). First, we show a relationship between \(\mu_e(a, x)\) and \(g_e(\tau, x)\). Since the decrease in the occupation \(\mu_e(a, x)\) with respect to aging is caused by the occurrence of events at age \(a\) (see figure 3), we obtain

\[
-\frac{\partial \mu_e(a, x)}{\partial a} \approx \frac{\mu_e(a, x) - \mu_e(a + \Delta a, x)}{\Delta a} = g_e(a, x).
\]

Also, since \(\mu_e(0, x)\) expresses the number of jumps (resets) from an arbitrary state to the state \(x\) in \((A, X)\), it can be expressed as

\[
\mu_e(0, x) = \mu_e(0, x).
\]
where \( g_e(x) = \sum_{x' \in \Omega} c_e(x,x') \) represents inflow to the state \( x \). Solving (64) with the boundary condition (65), we have

\[
\mu_e(a,x) = \int_0^\infty \text{d} \tau \, g_e(\tau,x),
\]

(66)

where we use \( \mu_e(0,x) = g_e(x) = \int_0^\infty \text{d} \tau \, g_e(\tau,x) \). Thus, we find that the correspondence between \( \mu_e(a,x) \) and \( g_e(\tau,x) \) is a bijection. Taking this fact into account, we obtain the rate function for \( \mu_e(a,x) \) and \( c_e(x,x') \) as

\[
I[\mu(a,x),c(x;x')] = \sum_{x' \in \Omega} \int_0^\infty \text{d} a \, \left\{ -\frac{\partial \mu(a,x)}{\partial a} \right\} \log \left\{ \frac{\partial \mu(a,x)}{\partial a} \right\} \pi(\tau|x) \mu(0,x) \\
+ \sum_{x,x' \in \Omega} c(x;x') \log \frac{c(x;x')}{\pi(x|x') \mu(0,x)},
\]

(67)

where we substitute (64) and (65) into (51). Note that the following shift-invariant property holds, due to (65):

\[
\mu(0,x) = \sum_{x' \in \Omega} c(x';x) = \sum_{x' \in \Omega} c(x;x').
\]

(68)

As shown in the next subsection, the contraction of the rate function (67) gives the level 2.5 LDP on continuous-time Markov processes under the Markov condition (11), i.e. the event rate does not depend on age: \( r(a,x) = r(x) \). Accordingly, we can say that the rate function (67) is an extension of the level 2.5 rate function to DTI semi-Markov processes.

Finally, we define the occupation distribution of the state \( x' \) as \( \mu_e(x') \), which is a marginal distribution of \( \mu_e(a,x) \). By using \( g_e(x',x') \), we can represent \( \mu_e(x') \) as

\[
\mu_e(x') := \int_0^\infty \text{d} a \, \mu_e(a,x) = \int_0^\infty \text{d} \tau' \, \tau' g_e(\tau',x'),
\]

(69)

where we use (66).

5.2. Contraction to the level 2.5 rate function for Markov processes

Here, we rederive the level 2.5 rate function (62), by using contraction of the rate function (67). Consider the rate function on the DTI semi-Markov process, (67), with an event rate \( r(a,x) = r(x) \). Then, owing to the contraction principle, the level 2.5 rate function for Markov processes is given by

\[
I[\mu(x),c(x;x')] = \min_{\mu(x)} \left\{ I[\mu(a,x),c(x;x')] \right\} \int_0^\infty \text{d} a \, \mu(a,x) = \mu(x),
\]

\[
\mu(0,x) = \sum_{x' \in \Omega} c(x';x) = \sum_{x' \in \Omega} c(x;x'),
\]

(70)

where the second constraint is due to the shift-invariant property (68). Here, we note that the joint occupation \( \mu(a,x) \) is contracted to the state occupation \( \mu(x) \) through (69). Although we can directly calculate the minimization to obtain the level 2.5 rate function, we here employ another approach. First, we note that the following minimization with respect to \( g(\tau,x) \) instead of \( \mu(a,x) \) is equivalent to that in (70):
where the rate function \( I [g (\tau, x), c (x, x')] \) is given by (51). The first constraint comes from (69) and the second one is from (65). Hereafter, we consider the minimization in (71) instead of one in (70). Next, by substituting the Markov condition \( r (a, x) = r (x) \) into the explicit form of \( I [g (\tau, x), c (x, x')] \), (51), we have

\[
I [\mu (x), c (x, x')] = \min_{g (\tau, x)} \left\{ I [g (\tau, x), c (x, x')] \mid \int_{0}^{\infty} d\tau g (\tau, x) = \mu (x), \right. \\
\left. \int_{0}^{\infty} d\tau g (\tau, x) = \sum_{x' \in \Omega} c (x'; x) = \sum_{x' \in \Omega} c (x; x') \right\},
\]

(71)

where \( g (x) \) is given by the shift-invariant property (43). Since all terms except the first one are fixed by the constraints, we can simplify the minimization problem in (71) as

\[
\min_{g (\tau, x)} \left\{ \sum_{x \in \Omega} \int_{0}^{\infty} d\tau g (\tau, x) \log g (\tau, x) \mid \int_{0}^{\infty} d\tau g (\tau, x) = \mu (x), \right. \\
\left. \int_{0}^{\infty} d\tau g (\tau, x) = \sum_{x' \in \Omega} c (x'; x) = \sum_{x' \in \Omega} c (x; x') \right\}.
\]

(73)

By employing the Lagrange multiplier method, we find that the function \( g^* (\tau, x) \) attaining the above minimization satisfies the following equation:

\[
\sum_{x \in \Omega} \int_{0}^{\infty} d\tau g^* (\tau, x) \log g^* (\tau, x) = \sum_{x \in \Omega} g (x) \log \frac{g^2 (x)}{\mu (x)} - \sum_{x \in \Omega} g (x).
\]

(74)

By substituting (74) into (72), we obtain the level 2.5 rate function as

\[
I [\mu (x), c (x, x')] = \sum_{x \in \Omega} \{ r (x) \mu (x) - g (x) \}
+ \sum_{x, x' \in \Omega} c (x; x') \log \frac{c (x; x')}{\mathbb{P} (x'|x) r (x') \mu (x')}.
\]

(75)

where we use the shift-invariant property (43). Finally, recalling that the transition rate of Markov processes can be represented as \( \omega (x|x') = \mathbb{P} (x|x') r (x') \) (i.e. \( \Sigma_{x' \in \Omega} \omega (x'|x) \mu (x) = r (x) \mu (x) \)), and \( g (x) = \Sigma_{x' \in \Omega} c (x'; x) \), we find that the explicit form (75) is equivalent to (62).

6. Summary and discussion

We have derived the explicit form of the rate function for semi-Markov processes with respect to the empirical triplet (12). Also, we have shown that the explicit form can be decomposed into point-process and Markov-process parts under the DTI assumption. In addition, by
exploiting the contraction principle to the decomposed rate function, we have elucidated that the FT (Gallavotti–Cohen symmetry) holds for DTI semi-Markov cases. Furthermore, we have found that the age representation of our rate function for semi-Markov processes gives an extended version of the level 2.5 rate function for Markov processes.

The explicit forms obtained in this paper can contribute to the analysis of cell cycles and neural firing by providing the way to calculate the statistical bias and variance of parameters estimated from empirical histograms of age, inter-event interval, and states. Moreover, the forms can be utilized for analysis of age-structured populations of cells and organisms. In contrast to the case depicted in figure 1 where we ignore the daughter cells generated by division event, we have to consider a whole population of cells including both mother and daughter cells. The situation is more entangled, but the explicit forms of the rate function for the single-cell division can play a pivotal role in such a problem. We will show the application of our rate function to the problem of an age-structured population in our next paper [61].

Acknowledgment

We thank Rosemary J Harris and Massimo Gavallaro for fruitful discussions. This research is supported by JSPS KAKENHI Grant Number JP16K17763, JP16H06155 and JST PRESTO Grant Number JPMJPR15E4, Japan.

Appendix A. Sanov’s theorem for Markov processes

Here, we show a brief derivation of the rate function (21); A rigorous proof is shown in [1, 2, 4]. For simplicity of calculation, we deal with 1-dimensional Markov processes; an extension to multidimensional processes is straightforward. Consider a time-discrete Markov process \((X) = \{X_i\}\) with an ergodic transition probability \(T(x|x')\), and its scaled cumulant generating function \(J_{e}(x,x')\): 

\[
J_{e}(x,x') = \left\langle \exp \left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{x,X_{i}} \delta_{x',X_{i}} \right\} \right\rangle = \left\langle \exp \left\{ \sum_{i=1}^{n} k(x,x') \right\} \right\rangle = \lim_{n \to \infty} \frac{1}{n} \log \sum_{\{x\}_n} e^{k(x,x') T(x|x')} \rho(x),
\]

where \(\langle \cdot \rangle\) represents the average over all paths \(\{X_i\} 1 \leq i \leq n\) with a path probability \(\Pi_{i=1}^{n-1} T(x_{i+1}|x_i) \rho(x_i); \rho(\cdot)\) is an arbitrary initial distribution. By using diagonalization of \(e^{\langle \cdot \cdot \rangle T(\cdot,\cdot)}\) and taking the limit \(n \to \infty\) into account, we find that the scaled cumulant generating function \(\lambda[k]\) is given by the logarithm of the largest eigenvalue of \(e^{\langle \cdot \cdot \rangle T(\cdot,\cdot)}\). That is, by employing the corresponding right eigenvector \(v_k(\cdot)\) (the right eigenvector corresponding to the largest eigenvalue), we have

\[
\sum_{y \in \Omega} e^{k(x,y) T(y|y)} v_k(y) = e^{\lambda k} v_k(x),
\]

(A.1)
where the uniqueness of the largest eigenvalue and the positivity of the corresponding eigenvector are guaranteed by the Perron–Frobenius theorem. From the Gärtner–Ellis theorem [4], the Legendre transform of $\lambda \left[ k \right]$ gives the rate function for the pair empirical measure $J_r \left( x, y \right)$:

$$ I \left[ J \right] = \max_k \left\{ \sum_{x, y \in \Omega} J \left( x, y \right) k \left( x, y \right) - \lambda \left[ k \right] \right\}. $$ \hspace{1cm} (A.3)

To solve the maximization, we calculate

$$ k^* \left( \cdot, \cdot \right) := \arg \max_k \left\{ \sum_{x, y \in \Omega} J \left( x, y \right) k \left( x, y \right) - \lambda \left[ k \right] \right\}. $$ \hspace{1cm} (A.4)

From the variation of (A.3) with respect to $k \left( x, y \right)$, $k^* \left( x, y \right)$ satisfies

$$ \frac{\delta \lambda \left[ k \right]}{\delta k \left( x, y \right) \bigg|_{k=k^*}} = J \left( x, y \right). $$ \hspace{1cm} (A.5)

To compute the variation of the left-hand side, we consider a procedure like perturbation methods. Now, analyze the following perturbed equation:

$$ \sum_{y \in \Omega} e^{\epsilon \left( x, y \right) + \delta k \left( x, y \right)} T \left( x \mid y \right) \left\{ \nu_k \left( y \right) + \delta \nu_k \left( y \right) \right\} \hspace{1cm} (A.6) $$

$$ = e^{\lambda \left[ k \right] + \delta \lambda \left[ k \right]} \left\{ \nu_k \left( x \right) + \delta \nu_k \left( x \right) \right\}. $$

Evaluation within the first order of $\delta$ leads to

$$ \sum_{y \in \Omega} e^{\epsilon \left( x, y \right) T \left( x \mid y \right)} \delta \nu_k \left( y \right) + \sum_{y \in \Omega} e^{\epsilon \left( x, y \right) T \left( x \mid y \right)} \nu_k \left( y \right) \delta k \left( x, y \right) \hspace{1cm} (A.7) $$

$$ = e^{\lambda \left[ k \right]} \delta \nu_k \left( x \right) + e^{\lambda \left[ k \right]} \nu_k \left( x \right) \delta \lambda \left[ k \right], $$

where we use (A.2) to simplify the equation. Furthermore, by applying the corresponding left eigenvector $u_k \left( \cdot \right)$ to both sides of (A.7) from the left side, we have

$$ \sum_{x, y \in \Omega} u_k \left( x \right) e^{\epsilon \left( x, y \right) T \left( x \mid y \right)} \delta \nu_k \left( y \right) + \sum_{x, y \in \Omega} u_k \left( x \right) e^{\epsilon \left( x, y \right) T \left( x \mid y \right)} \nu_k \left( y \right) \delta k \left( x, y \right) \hspace{1cm} (A.8) $$

$$ = e^{\lambda \left[ k \right]} \sum_{x \in \Omega} u_k \left( x \right) \delta \nu_k \left( x \right) + e^{\lambda \left[ k \right]} \sum_{x \in \Omega} u_k \left( x \right) \nu_k \left( x \right) \delta \lambda \left[ k \right]. $$

Taking into account the fact that $u_k \left( \cdot \right)$ represents the left eigenvector of $e^{\epsilon \left( \cdot, \cdot \right) T \left( \cdot, \cdot \right)}$:

$$ \sum_{x \in \Omega} u_k \left( x \right) e^{\epsilon \left( x, y \right) T \left( x \mid y \right)} = e^{\lambda \left[ k \right]} u_k \left( y \right), $$ \hspace{1cm} (A.9)

we can cancel the first terms in both sides of (A.8). Then, after simplifying (A.8), we obtain

$$ \frac{\delta \lambda \left[ k \right]}{\delta k \left( x, y \right)} = \frac{u_k \left( x \right) e^{\epsilon \left( x, y \right) - \lambda \left[ k \right] T \left( x \mid y \right)} \nu_k \left( y \right)}{\sum_{z \in \Omega} u_k \left( z \right) \nu_k \left( z \right)}. $$ \hspace{1cm} (A.10)

Accordingly, from (A.5), $k^* \left( x, y \right)$ satisfies

$$ J \left( x, y \right) = \frac{u_k^* \left( x \right) e^{\epsilon \left( x, y \right) - \lambda \left[ k^* \right] T \left( x \mid y \right)} \nu_k^* \left( y \right)}{\sum_{z \in \Omega} u_k^* \left( z \right) \nu_k^* \left( z \right)}. $$ \hspace{1cm} (A.11)

Also, we define a marginal distribution $G \left( \cdot \right)$ as
\[ G(y) := \sum_{x \in \Omega} J(x, y) = \sum_{x \in \Omega} J(y, x) = \frac{u_k^*(y) v_k^*(y)}{\sum_{x \in \Omega} u_k^*(z) v_k^*(z)}, \]  
(A.12)

where we use (A.9). By solving (A.11) with respect to \( e^{k^*(x, y) - \lambda k^*} \), we have

\[ e^{k^*(x, y) - \lambda k^*} = \frac{J(x, y) \sum_{x \in \Omega} u_k^*(z) v_k^*(z)}{u_k^*(x) \mathbb{P}(x|y) v_k^*(y)}. \]  
(A.13)

Noting (A.3), we can express the rate function by using \( k^*(\cdot, \cdot') \) as

\[ I[f] = \sum_{x, y \in \Omega} J(x, y) \log e^{k^*(x, y) - \lambda k^*}. \]  
(A.14)

By substituting (A.13) into (A.14), we get

\[ I[f] = \sum_{x, y \in \Omega} J(x, y) \log \frac{J(x, y)}{\mathbb{P}(x|y)} + \sum_{x \in \Omega} J(x, y) \log \frac{\sum_{x \in \Omega} u_k^*(z) v_k^*(z)}{u_k^*(x) v_k^*(y)}. \]  
(A.15)

Finally, by using the shift-invariant property (A.12), we rewrite the second term in (A.15) as

\[
\begin{align*}
&= \sum_{x, y \in \Omega} J(x, y) \log \frac{\sum_{x \in \Omega} u_k^*(z) v_k^*(z)}{u_k^*(x) v_k^*(y)} \\
&= \sum_{x, y \in \Omega} J(x, y) \log \frac{\sum_{x \in \Omega} u_k^*(z) v_k^*(z)}{u_k^*(x) v_k^*(y)} = \sum_{x \in \Omega} J(x, y) \log \frac{1}{G(y)},
\end{align*}
\]  
(A.16)

where we use the expression of \( G(\cdot) \), (A.12). Thus, we obtain the explicit form of the rate function as

\[ I[f] = \sum_{x, y \in \Omega} J(x, y) \log \frac{J(x, y)}{\mathbb{P}(x|y) G(y)}. \]  
(A.17)

By extending this calculation to 2-dimensional cases, we can find (21).

**Appendix B. The LDP on point processes**

In this appendix, we introduce point processes [58] and their LDP. Suppose an inter-event time interval sequence \( (T) := \{T_i\} \), where each element \( T_i \) is distributed with a probability density,

\[ \pi(\tau) := r(\tau) e^{-\int_0^\tau r(a) da}. \]  
(B.1)

Here, \( r(a) \) represents the event rate, i.e. the probability that an event occurs at age \( a \). The age means the elapsed time since the last event occurs. Thus, we regard \( r(a) \) as a simple version of \( r(a, x) \) introduced in section 2. The process \( (T) \) generated by (B.1) is a kind of point processes. If \( r(a) \) is constant, the point process is reduced to a homogeneous Poisson point process.

For the above process, we consider the rate function of the following ‘time-normalized’ empirical measure:

\[ g_\tau(\tau) := \frac{1}{n} \sum_{i=1}^n \delta(\tau - T_i), \]  
(B.2)
which measures how often the inter-event interval $\tau$ appears in the sequence $(T)$. Note that this measure is normalized as $\int_0^\infty d\tau \ 	au g_e(\tau) = 1$ at $t \to \infty$. According to the procedure in section 3, to calculate the rate function, we firstly consider the ‘number-normalized’ empirical measure:

$$ G_e(\tau) := \frac{1}{n} \sum_{i=1}^n \delta (\tau - T_i), $$

which is normalized as $\int_0^\infty d\tau \ G_e(\tau) = 1$, differently from (B.2). Now, since each element $T_i$ is independent and identically distributed (IID) from $\pi(\cdot)$, Sanov’s theorem for IID [4] leads the explicit form of the rate function for $G_e(\tau)$ as

$$ \tilde{I}[G(\tau)] = \int_0^\infty d\tau \ G(\tau) \log \frac{G(\tau)}{\pi(\tau)}. $$

By using the random time change trick shown in section 3.2, we obtain

$$ I[g(\tau)] = \left\{ \int_0^\infty d\tau \ g(\tau) \right\} \tilde{I} \left[ \frac{1}{\int_0^\infty d\tau \ g(\tau)} \right], $$

where $I[g(\tau)]$ is the rate function of $g_e(\tau)$. Substituting (B.4) into (B.5), we finally find the explicit form:

$$ I[g(\tau)] = \int_0^\infty d\tau \ g(\tau) \log \frac{g(\tau)}{\pi(\tau)} g, $$

where the constant $g$ is $g := \int_0^\infty d\tau \ g(\tau)$, which represents the number of events per unit time, that is $n/\lambda t$.

**ORCID iDs**

Yuki Sughiyama 🅰️ https://orcid.org/0000-0001-6688-9373

**References**

[1] Dembo A and Zeitouni O 1998 *Large Deviations Techniques and Applications* 2nd edn (Berlin: Springer)
[2] Ellis R S 1985 *Entropy, Large Deviations, and Statistical Mechanics* (Berlin: Springer)
[3] Freidlin M I and Wentzell A D 1984 *Random Perturbations of Dynamical Systems* (Berlin: Springer)
[4] Touchette H 2009 The large deviation approach to statistical mechanics *Phys. Rep.* 478 1–69
[5] Landau L D and Lifshitz E M 1980 *Statistical Physics* 3rd edn (Oxford: Pergamon)
[6] Chetrite R and Touchette H 2013 Nonequilibrium microcanonical and canonical ensembles and their equivalence *Phys. Rev. Lett.* 111 120601
[7] Verley G 2016 Nonequilibrium thermodynamic potentials for continuous-time Markov chains *Phys. Rev. E* 93 012111
[8] Sughiyama Y and Ohzeki M 2013 Variational principle in Langevin processes *Interdiscip. Inf. Sci.* 19 93–9
[9] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in nonequilibrium statistical mechanics *Phys. Rev. Lett.* 74 2694
[10] Lebowitz J L and Spohn H 1998 A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics *J. Stat. Phys.* 95 333–65
[11] Harris R J and Schütz G M 2007 Fluctuation theorems for stochastic dynamics J. Stat. Mech. P07020
[12] Chetrite R and Gawedzki K 2008 Fluctuation relations for diffusion processes Commun. Math. Phys. 282 469–518
[13] Sagawa T 2013 Thermodynamics of Information Processing in Small Systems (Berlin: Springer)
[14] Hatano T and Sasa S-I 2001 Steady-state thermodynamics of Langevin systems Phys. Rev. Lett. 86 3463
[15] Komatsu T S, Nakagawa N, Sasa S-I and Tasaki H 2010 Entropy and nonlinear nonequilibrium thermodynamic relation for heat conducting steady states J. Stat. Phys. 142 127–53
[16] Nemoto T and Sasa S-I 2011 Variational formula for experimental determination of high-order correlations of current fluctuations in driven systems Phys. Rev. E 83 030105
[17] Nemoto T and Sasa S-I 2011 Thermodynamic formula for the cumulant generating function of time-averaged current Phys. Rev. E 84 061113
[18] Giardinà C, Kurchan J and Peliti L 2006 Direct evaluation of large-deviation functions Phys. Rev. Lett. 96 120603
[19] Donsker M D and Varadhan S R S 1975 Asymptotic evaluation of certain Markov process expectation for large time, I Commun. Pure Appl. Math. 28 1–47
[20] Bertini L, Faggionato A and Gabrielli D 2012 Large deviations of the empirical flow for continuous time Markov chains (arXiv:1210.2004)
[21] Zelniker E E, Bradley A P, Castner J E, Chenery H J, Copland D A and Silburn P A 2008 Estimation of neuronal firing rates with the three-state biological point process model J. Neurosci. Methods 174 281–91
[22] El-Labban A, Zisserman A, Toyoda Y, Bird A W and Hyman A 2012 Discriminative semi-Markov models for automated mitotic phase labelling IEEE Int. Symp. on Biomedical Imaging pp 760–3
[23] Liu A A, Li K and Kanade T 2011 A semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations IEEE Trans. Med. Imaging 31 359–69
[39] Esposito M and Lindenberg K 2008 Continuous time random walk for open systems: fluctuation theorems and counting statistics Phys. Rev. E 77 051119
[40] Andrieux D and Gaspard P 2008 The fluctuation theorem for currents in semi-Markov processes J. Stat. Mech. P11007
[41] Maes C, Netočný K and Wynants B 2009 Dynamical fluctuations for semi-Markov processes J. Phys. A: Math. Theor. 42 365002
[42] Duffy K and Metcalfe A P 2005 How to estimate the rate function of a cumulative process J. Appl. Probab. 42 1044–52
[43] Duffy K R, Macì C and Torrisi G L 2011 On the large deviations of a class of modulated additive processes ESAIM: Probab. Stat. 15 83–109
[44] Lefèvre R, Mariani M and Zambotti L 2011 Large deviations for renewal processes Stoch. Process. Appl. 121 2243–71
[45] Mariani M and Zambotti L 2014 A renewal version of the Sanov theorem Electron. Commun. Probab. 19 1–13
[46] Mariani M and Zambotti L 2016 Large deviations for the empirical measure of heavy-tailed Markov renewal processes Adv. Appl. Probab. 48 648–71
[47] Golubev A 2016 Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression J. Theor. Biol. 393 203–17
[48] Smith J A and Martin L 1973 Do cells cycle? Proc. Natl Acad. Sci. 70 1263–7
[49] Cooper S 1982 The continuum model: statistical implications J. Theor. Biol. 94 783–800
[50] Bintu L, Yong J, Antebi Y E, McCue K, Kazuki Y, Uno N, Oshimura M and Elowitz M B 2016 Dynamics of epigenetic regulation at the single-cell level Science 351 720–4
[51] Yasunobu M, Kobayashi T J, Nakayama J, Uchida H and Oki M 2013 Single cell visualization of yeast gene expression shows correlation of epigenetic switching between multiple heterochromatic regions through multiple generations PLOS Biol. 11 e1001601
[52] Wang P, Robert L, Pelletier J, Dang W L, Taddei F, Wright A and Jun S 2010 Robust growth of Escherichia coli Curr. Biol. 20 1099–103
[53] Sachs C C, Grünberger A, Helfrich S, Probst C, Wiechert W, Kohlheyer D and Nöhl K 2016 Image-based single cell profiling: high-throughput processing of mother machine experiments PloS One 11 e0163453
[54] Tanouchi Y, Pai A, Park H, Huang S, Buchler N E and You L 2017 Long-term growth data of Escherichia coli at a single-cell level Sci. Data 4 170036.
[55] Bouzebda S and Limnios N 2013 On general bootstrap of empirical estimator of a semi-Markov kernel with applications J. Multivariate Anal. 116 52–62
[56] Russell R 1997 The large deviations of random time changes PhD Thesis Trinity College Dublin
[57] Rodgers-Lee M 2003 The large deviations of random time-changes in a metric topology PhD Thesis Trinity College Dublin
[58] Bremaud P 1981 Point Processes and Queues: Martingale Dynamics (Berlin: Springer)
[59] Gardiner C W 1985 Handbook of Stochastic Methods 2nd edn (Berlin: Springer)
[60] Risken H 1989 The Fokker–Planck Equation Methods of Solution and Applications 2nd edn (Berlin: Springer)
[61] Sughiyama Y and Kobayashi T J in preparation