Asymptotic and non-asymptotic analysis for a hidden Markovian process with a quantum hidden system

Masahito Hayashi$^{1,2,3}$ and Yuuya Yoshida$^4$

1 Graduate School of Mathematics, Nagoya University, Nagoya 464-8602, Japan
2 Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Nanshan District, Shenzhen 518055, People’s Republic of China
3 Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117542, Singapore

E-mail: masahito@math.nagoya-u.ac.jp and m17043e@math.nagoya-u.ac.jp

Received 13 February 2018, revised 19 June 2018
Accepted for publication 20 June 2018
Published 10 July 2018

Abstract

We focus on a data sequence produced by repetitive quantum measurement on an internal hidden quantum system, and call it a hidden Markovian process. Using a quantum version of the Perron–Frobenius theorem, we derive novel upper and lower bounds for the cumulant generating function of the sample mean of the data. Using these bounds, we derive the central limit theorem and large and moderate deviations for the tail probability. Then, we give the asymptotic variance by using the second derivative of the cumulant generating function. We also derive another expression for the asymptotic variance by considering the quantum version of the fundamental matrix. Further, we explain how to extend our results to a general probabilistic system.

Keywords: quantum system, hidden Markov, central limit theorem, large deviation, moderate deviation, asymptotic variance

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

1. Introduction

Consider a physical system with an internal quantum system. Usually, it is not so easy to observe the internal system directly. When the physical system has a classical output, we can observe the classical output but other parts cannot be observed. For example, a quantum random number generator has such an internal system and a classical output [1]. As another
example, such a system appears in a quantum memory of a channel [2–4]. Such a correlated system also appears in quantum spin chains [15, 16]. This kind of system is formulated to be a hidden Markovian process with quantum hidden system as shown in figure 1. Initially, a classical Markovian process is formulated as a probability transition matrix. Then, a classical hidden Markovian process is formulated as two probability transition matrices, one of which describes the Markovian process in the hidden system, and the other the relation between the hidden system and the observed system.

While there are several formulations of quantum analogue of Markovian process, one natural formulation is a trace-preserving completely positive map (TP-CP map). However, in this formulation, nobody observes the system, i.e. no observation of the quantum system is discussed. To introduce a measurability on this system, we need to introduce a hidden Markovian process with quantum hidden system. When the quantum system can be measured, the resultant state depends on the classical output. That is, the state evolution depends on the classical output \( \omega \in \Omega \), and is described by a set of CP maps \( C_\omega \), which is often called an instrument [5]. In this case, the sum \( \sum_{\omega \in \Omega} C_\omega \) needs to be trace-preserving. When the initial state is \( \rho \), the classical output \( \omega \) is observed with the probability \( \text{Tr} C_\omega (\rho) \) so that the resultant state is \( C_\omega (\rho) / \text{Tr} C_\omega (\rho) \). Such a system is initially formulated as a quantum measuring process [5], and the set \( \{ C_\omega \}_{\omega \in \Omega} \) is called an instrument.

For the classical case, there are many studies of Markovian processes. These studies focus on the random variables \( X_i \) generated subject to this process and consider the sample mean \( \bar{X}^n := (1/n) \sum_{i=1}^{n} X_i \). Similarly to the independently and identically distributed case, the sample mean \( X^n \) converges to the expectation in probability. Also, the central limit theorem holds for the sample mean [6–9]. Further, the large and moderate deviations also hold [10] [11, theorem 3.1.2] [13, corollaries 8.3 and 8.4]. However, the non-asymptotic analysis has not been discussed sufficiently. While, in the non-asymptotic analysis, we derive upper and lower bounds for the tail probability, we need to consider requirements for a good bound because we need to distinguish good bounds from trivial bounds. Similarly to [12], we impose the following requirements on good bounds.

1.1. Computational complexity
In order that the bound works efficiently, we need to calculate the bound efficiently. With this aim, we need to clarify the computational complexity to calculate the bound, and the complexity needs to be polynomial with respect to the number \( n \) of observations, at least.
1.2. Asymptotic tightness

The bound needs to achieve optimality in one of the following regimes:

- **C1** Large deviation.
- **C2** Moderate deviation.
- **C3** Central limit theorem.

The paper [13] derived upper and lower bounds for the tail probability, which have computational complexity $O(1)$ and achieve asymptotic tightness in the sense of C1 and C2. These upper and lower bounds are derived from the evaluation of the cumulant generating function. Large deviation applies to events for which the difference between the sample mean and the expectation is greater than a certain threshold, this threshold being a constant. That is, in the large deviation case, the event of interest is strongly deviated. Thus, an event of large deviation has exponentially small probability. Moderate deviation applies to events in which the threshold constant is larger than that in the central limit regime but goes to zero. That is, in the moderate deviation case, the event of interest is moderately deviated. Since an event subject to the central limit theorem converges to a constant, the event of the moderate deviation can be regarded as the intermediate situation.

In the quantum setting, the paper [14] derived large deviation in a similar setting. The papers [15, 16] discussed large deviation in quantum spin chains. However, no study has derived upper and lower bounds to satisfy the above requirements for moderate deviation. In addition, the paper [17] addressed local asymptotic normality in the context of system identification for quantum Markov chains. However, it did not discuss the central limit theorem of the sample mean when the hidden system is given as a quantum system.

In this paper, we consider the extension of the upper and lower bounds derived in [13] to a hidden Markovian process with quantum hidden system. That is, when a real-valued variable $X_i$ is generated subject to the process, we focus on the sample mean $X^n := (1/n) \sum_{i=1}^n X_i$, and discuss the asymptotic behavior. More precisely, we show that the random variable $\sqrt{n}(X^n - E)$ converges to the Gaussian distribution, where $E$ is the expectation value. Next, we focus on the tail probability of this process. We derive large and moderate deviations for the tail probability. In addition, we derive a finite-length evaluation for the tail probability, i.e. upper and lower bounds of the tail probability that derive the large and moderate deviations for the tail probability. In these derivations, we first focus on the quantum version of the Perron–Frobenius theorem [18, 19], which characterizes the Perron–Frobenius eigenvalue. Using the Perron–Frobenius eigenvalue, we derive upper and lower bounds of the cumulant generating function of the sample mean, whose computational complexity is $O(1)$. Then, employing the same method as [13], we show the asymptotic tightness in the sense of C1 and C2. Further, we derive the central limit theorem and calculate the asymptotic variance.

The remaining part of this paper is organized as follows. Sections 2 and 3 are devoted to mathematical preparations. Section 4 explains Bregman divergence and its variants, which are powerful tools for our purpose. Section 5 gives the central limit theorem. Section 6 derives upper and lower bounds for the tail probability. These bounds achieve tightness in the sense of the large and moderate deviations. Section 7 gives the concrete form of the variance, which appears in the moderate deviation and central limit theorem. Section 8 explains that our model is equivalent to the model of finitely correlated states discussed in [15].

2. Perron–Frobenius theorem for quantum systems

In preparation, we summarize basic knowledge for a quantum version of the Perron–Frobenius theorem on a finite-dimensional quantum system $\mathcal{H}$ of interest. With this aim, we explain the
results of [20] in the case of completely positive maps. First, let us define a few terms used in linear algebra. The spectral radius \( r(\Lambda) \) of a linear map \( \Lambda \) is defined as the maximum of the absolute values of all eigenvalues of \( \Lambda \) on the linear space \( \mathcal{T}(\mathcal{H}) \) of all Hermitian matrices on \( \mathcal{H} \) [19]. The relations \( r(\Lambda_1 \otimes \Lambda_2) = r(\Lambda_1)r(\Lambda_2) \) and \( r(\Lambda^*) = r(\Lambda) \) hold, where \( \Lambda^* \) denotes the adjoint map of a linear map \( \Lambda \). For a linear map \( \Lambda \) and its eigenvalue, the multiplicity as a root of the characteristic polynomial is called the algebraic multiplicity, and the dimension of the eigenspace is called the geometric multiplicity [20]. The spectral radius plays a special role as follows. Thus, we have the following proposition.

**Proposition 1 ([20, theorem 6]).** For any completely positive map \( \Lambda \), the following conditions are equivalent.

(I-i) The inequality \( r(\Lambda) > 0 \) holds, and there exist strictly positive definite matrices \( \rho_0 \) and \( A_0 \) such that \( \text{Tr} \rho_0 A_0 = 1 \), and any Hermitian matrix \( H \) satisfies

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (r(\Lambda)^{-1} \Lambda^k)(H) = (\text{Tr} A_0 H) \rho_0.
\]

(I-ii) Both \( \Lambda \) and \( \Lambda^* \) have strictly positive definite eigenvectors associated with \( r(\Lambda) > 0 \), and the eigenvalue \( r(\Lambda) \) of \( \Lambda \) has the geometric multiplicity 1.

(I-iii) For any state \( \rho > 0 \), there exists a positive number \( t \) such that \( e^{t \Lambda}(\rho) > 0 \).

(I-iv) Any state \( \rho \) satisfies \( (\Lambda + \Lambda^*)^{(\text{dim } \mathcal{H})} - 1(\rho) > 0 \).

(I-v) If a state \( \rho \) and a nonnegative matrix \( \alpha \) satisfy \( \Lambda(\rho) \leq \alpha 0 \), then \( \rho > 0 \).

(I-vi) \( \Lambda \) has no eigenvectors on the boundary of the set of all positive semi-definite matrices.

(I-iii\( ' \)) For any states \( \rho \) and \( \sigma \) whose ranks equal one, there exists a natural number \( n \) such that \( \text{Tr} \sigma \Lambda^n(\rho) > 0 \).

Here, for two Hermitian matrices \( H \) and \( H' \) on \( \mathcal{H} \), the relations \( H \leq H' \) and \( H < H' \) mean that \( H' - H \) is respectively positive semi-definite or strictly positive definite. \( \iota \) denotes the identity map on \( \mathcal{T}(\mathcal{H}) \).

A \( d \times d \) nonnegative matrix \( W \) is called irreducible when for any \( i,j \), there exists a natural number \( n \) such that the \((i,j)\) component \((W^n)_{ij}\) of \( W^n \) is positive. It can easily be shown that this condition is equivalent to the condition that \((e^{t \Lambda})_{ij} > 0 \) for any \( i,j \). Hence, condition (I-iii) can be regarded as the quantum extension of the irreducibility. Hence, a completely positive map \( \Lambda \) on \( \mathcal{T}(\mathcal{H}) \) is called irreducible when at least one condition in proposition 1 holds. The value \( r(\Lambda) \) and the matrix \( \rho_0 \) in condition (I-i) are called the Perron–Frobenius eigenvalue and Perron–Frobenius eigenvector of \( \Lambda \), respectively. It can be shown that the irreducibility of \( \Lambda \) implies that of \( \Lambda^* \), and that the two matrices \( \rho_0 \) and \( A_0 \) in condition (I-i) are eigenvectors of \( \Lambda \) and \( \Lambda^* \), respectively [20]. Condition (I-iii) is called ergodicity in [18].

An irreducible completely positive map \( \Lambda \) satisfies condition (I-i), but \((r(\Lambda)^{-1} \Lambda)^n(\rho) \) might behave periodically as \( n \to \infty \). Combining corollary 3 and theorem 6 of [20], we obtain the following conditions that exclude such periodicity.

**Proposition 2.** For any completely positive map \( \Lambda \), the following conditions are equivalent.

(P-i) The inequality \( r(\Lambda) > 0 \) holds and there exist strictly positive definite matrices \( \rho_0 \) and \( A_0 \) such that \( \text{Tr} \rho_0 A_0 = 1 \), and any Hermitian matrix \( H \) satisfies
\[
\lim_{n \to \infty} (r(\Lambda)^{-1} \Lambda)^n(H) = (\text{Tr} A_0 H) \rho_0.
\]

(P-ii) Both \(\Lambda^{\otimes 2}\) and \((\Lambda^{\otimes 2})^*\) have strictly positive definite eigenvectors associated with \(r(\Lambda)^2 > 0\), and the eigenvalue \(r(\Lambda)^2\) of \(\Lambda^{\otimes 2}\) has the geometric multiplicity 1.

(P-iii) For any state \(\rho > 0\), there exists a positive number \(t\) such that \(e^{t \Lambda^{\otimes 2}}(\rho) > 0\).

(P-iv) Any state \(\rho\) satisfies \((\rho^{\otimes 2} + \Lambda^{\otimes 2})^{(\dim \mathcal{H})^2 - 1}(\rho) > 0\).

(P-v) If a state \(\rho\) and a nonnegative number \(\alpha\) satisfy \(\Lambda^{\otimes 2}(\rho) \leq \alpha \rho\), then \(\rho > 0\).

(P-vi) \(\Lambda^{\otimes 2}\) has no eigenvector on the boundary of the set of all positive semi-definite matrices.

(I-iii) For any states \(\rho\) and \(\sigma\) whose ranks equal one, there exists a natural number \(n\) such that \(\text{Tr} \sigma \Lambda^{\otimes 2^n}(\rho) > 0\).

A completely positive map \(\Lambda\) on \(\mathcal{T}(\mathcal{H})\) is called primitive when at least one condition in proposition 2 holds. It can be shown that the primitivity implies the irreducibility and that the primitivity of \(\Lambda\) implies that of \(\Lambda^*\) [20]. When a completely positive map maps all diagonal matrices to themselves, the above irreducibility and primitivity are respectively equivalent to the irreducibility and primitivity in the classical case. As for other equivalent conditions for the primitivity, see [19, theorem 6.7].

Proposition 2 has been derived from corollary 3 and theorem 6 of [20]. For a completely positive map, corollary 3 of [20] is simple, as follows.

**Proposition 3 ([20, corollary 3]).** Let \(\Lambda\) be a completely positive map. Then, \(\Lambda\) is primitive if and only if \(\Lambda^{\otimes 2}\) is irreducible.

The spectral radius of a trace-preserving completely positive map equals one, and its adjoint map has the eigenvector \(I\) [20, section 3.2], where \(I\) denotes the identity matrix on \(\mathcal{H}\). Thus, conditions (I-ii) and (P-ii) lead immediately to the following corollary.

**Corollary 1.** Let \(\Lambda\) be a trace-preserving completely positive map. Then, \(\Lambda\) is irreducible if and only if \(\Lambda\) has a fixed state with full rank and the equation \(\dim \text{Ker}(\Lambda - \iota) = 1\) holds. \(\Lambda\) is primitive if and only if \(\Lambda\) has a fixed state with full rank and the equation \(\dim \text{Ker}(\Lambda^{\otimes 2} - \iota^{\otimes 2}) = 1\) holds.

Further, we have the following proposition.

**Proposition 4 ([20, corollary 6]).** Let \(\Omega\) be a nonempty finite set, \(\Lambda_{\omega}\) be a completely positive map and \(a_{\omega}\) be a positive number for each \(\omega \in \Omega\). If \(\Lambda := \sum_{\omega} \Lambda_{\omega}\) is irreducible, then so is \(\Lambda_{\omega} := \sum_{\omega} a_{\omega} \Lambda_{\omega}\). If \(\Lambda := \sum_{\omega} \Lambda_{\omega}\) is primitive, then so is \(\Lambda_{\omega} := \sum_{\omega} a_{\omega} \Lambda_{\omega}\).

All the statements in this section hold under a more general setting [20]. Sections 3–6 derive several properties only with the irreducibility. Only section 7 requires the primitivity.

As a foundation for understanding of the statements in this section, we give a trace-preserving completely positive map that is irreducible but not primitive, as follows.

**Example 1.** Let \([\langle i \rangle]\) be an orthonormal basis of \(\mathbb{C}^d\) and \(\Lambda\) be the trace-preserving completely positive map on \(\mathcal{T}(\mathbb{C}^d)\) which maps any Hermitian matrix \(H\) to

\[
\Lambda(H) := \sum_{i=0}^{d-1} [i - 1] H [i - 1] [i] [i],
\]

where \(i \in \mathbb{Z}/d\mathbb{Z}\). To show that \(\Lambda\) is irreducible, we adopt condition (P-v). Let \(\rho\) and \(\alpha\) be a state and a nonnegative number, respectively, satisfying \(\Lambda(\rho) \leq \alpha \rho\). Suppose \(\langle i | \rho | i \rangle = 0\) for
an $i$. Then,

$$0 = \alpha \langle i | \rho | i \rangle \geq \langle i | \Lambda(\rho) | i \rangle = \langle i - 1 | \rho | i - 1 \rangle,$$

whence $\langle i - 1 | \rho | i - 1 \rangle = 0$. Iterating this, we have $\operatorname{Tr} \rho = 0$, which contradicts that $\rho$ is a state. Therefore, $\langle i | \rho | i \rangle > 0$ for any $i$. Since the inequality $0 < \Lambda(\rho) \leq \alpha \rho$ means $\rho > 0$, $\Lambda$ is irreducible.

To show that $\Lambda$ is not primitive, we adopt condition (P-ii). The separable state $(1/d) \sum_{i=0}^{d-1} | i \rangle \langle i | \otimes | i \rangle \langle i |$ is mapped to itself by $\Lambda^{\otimes 2}$. The maximally mixed state $(1/d) | i \rangle \langle i | \otimes | i \rangle \langle i |$ and $(1/d^2) I^{\otimes 2}$ are eigenvectors associated with $r(\Lambda) = 1$ of $\Lambda$. Thus, the two separable states $(1/d) | i \rangle \langle i | \otimes | i \rangle \langle i |$ and $(1/d^2) I^{\otimes 2}$ are eigenvectors associated with $r(\Lambda)^2 = 1$ of $\Lambda^{\otimes 2}$. Therefore, condition (P-ii) does not hold. Moreover, we show that $\Lambda$ does not satisfy condition (P-i). The sequence $\{\Lambda^n(|0\rangle \langle 0|)\}_{n}$ does not converge because $\Lambda(|i \rangle \langle i|) = |i + 1\rangle \langle i + 1|$ for any $i$. Hence, condition (P-i) does not hold.}

\section{3. Evaluations for cumulant generating function}

We discuss the data sequence of hidden Markovian process with quantum hidden system generated from quantum measurement process described by a set of CP maps (an instrument) $C = \{C_\omega\}_{\omega \in \Omega}$ when $\Lambda := \sum_{\omega \in \Omega} C_\omega$ is irreducible. For a symbol $\omega \in \Omega$, we assign it to a real number $x_\omega$. Since proposition 4 guarantees that $\Lambda_\theta := \sum_{\omega \in \Omega} e^{\theta x_\omega} C_\omega$ is also an irreducible completely positive map, the completely positive map $\Lambda_\theta$ has the Perron--Frobenius eigenvalue $\lambda_\theta$ and the Perron–Frobenius eigenvector $\rho_\theta$ whose trace equals one. We define the function $\phi(\theta)$ to be $\log \lambda_\theta$. Here, the adjoint map $\Lambda_\theta^*$ has the same Perron–Frobenius eigenvalue $\lambda_\theta$. We choose the Perron–Frobenius eigenvector $\rho_\theta$ of $\Lambda^*$ such that $\Lambda_\theta - I$ is positive semi-definite but not strictly positive definite. This is possible because once we take a Perron–Frobenius eigenvector $A$ of $\Lambda_\theta$, $A$ is strictly positive definite and an appropriate positive number $\alpha$ satisfies the condition that $\alpha A - I$ is positive semi-definite but not strictly positive definite.

We denote the sequence of observed real numbers by $X_1, \ldots, X_n$ as figure 2. When the initial state is $\rho$, the variables $X_1, \ldots, X_n$ take the values $x_1, \ldots, x_n$, respectively, with the probability $\operatorname{Tr} C_{x_1} \circ \cdots \circ C_{x_n} (\rho)$, where $x_i = x_{x_i}$. In this case, we denote the cumulant generating function of the variable $nX^n = \sum_{i=1}^n X_i$ by $\phi_{n, \rho}(\theta)$. When the initial state is the eigenvector $\rho_\theta$ of $\Lambda_\theta$, the cumulant generating function $\phi_{n, \rho_\theta}(\theta)$ is calculated as

$$e^{\phi_{n, \rho_\theta}(\theta)} = \sum_{\omega_1 \cdots \omega_n} \operatorname{Tr} e^{\theta x_{\omega_1}} C_{\omega_1} \circ \cdots \circ e^{\theta x_{\omega_n}} C_{\omega_n}(\rho_\theta)$$

$$= \operatorname{Tr} \Lambda_\theta^n(\rho_\theta) = \lambda_\theta^n \operatorname{Tr} \rho_\theta = \lambda_\theta^n,$$

which implies

$$\phi_{n, \rho_\theta}(\theta) = n \phi(\theta). \quad (2)$$

Now, we evaluate the cumulant generating function $\phi_{n, \rho}(\theta)$ for a general initial state $\rho$ by defining

$$\tilde{\delta}(\theta) := \log \operatorname{Tr} A_{\theta} \rho, \quad \tilde{\delta}_n(\theta) := \log \operatorname{Tr} A_{\theta} \rho - \log \|A_{\theta}\|.$$  \quad (3)

Thus, we have the following main theorem.

\textbf{Theorem 1.} The cumulant generating function $\phi_{n, \rho}(\theta)$ of observed variables $X_1, \ldots, X_n$ is evaluated as
While this evaluation is very simple, this evaluation leads to many fruitful derivations for the asymptotic behavior of a hidden Markovian process with quantum hidden system.

**Proof.** Since $A_\theta \geq I$, we have
\[
e^{\alpha_{n,\theta}} = \sum_{\omega_1, \ldots, \omega_n} \text{Tr} e^{\theta \omega_n} C_{\omega_n} \circ \cdots \circ e^{\theta \omega_1} C_{\omega_1}(\rho)
= \text{Tr} \Lambda_{\theta}^n(\rho) \leq \text{Tr} A_\theta \Lambda_{\theta}^n(\rho) = \text{Tr} (\Lambda_{\theta}^n)^n(A_\theta) \rho
= \lambda^n_0 \text{Tr} A_\theta \rho = \lambda^n_0 e^{\delta_\rho(\theta)},
\]
which implies the second inequality in (4). Conversely, since $\frac{1}{\|A_\theta\|} A_\theta \leq I$, we have
\[
e^{\alpha_{n,\theta}} = \text{Tr} \Lambda_{\theta}^n(\rho) \geq \text{Tr} \frac{1}{\|A_\theta\|} A_\theta \Lambda_{\theta}^n(\rho)
= \frac{1}{\|A_\theta\|} \lambda^n_0 \text{Tr} A_\theta \rho = \lambda^n_0 e^{\delta_\rho(\theta)},
\]
which implies the first inequality in (4).

**Lemma 1.**
\[
\lim_{\theta \to 0} \delta_\rho(\theta) = 0, \quad \lim_{\theta \to 0} \delta_\rho(\theta) = 0.
\]

**Proof.** From the construction of $A_\theta$, $A_\theta$ is continuous for $\theta$. Hence,
\[
\lim_{\theta \to 0} \text{Tr} A_\theta \rho = \text{Tr} A_0 \rho = \text{Tr} I \rho = 1,
\]
which implies the first equation of (7). Similarly,
\[
\lim_{\theta \to 0} \|A_\theta\| = \|I\| = 1.
\]
Combining (8) and (9), we obtain the second equation of (7).

By taking the limit in (4) of lemma 1, we have the following.

**Corollary 2.** For $\theta \in \mathbb{R}$, we have
\[
\lim_{n \to \infty} \frac{1}{n} \phi_{n,\theta}(\theta) = \phi(\theta).
\]
In fact, $\lambda_\theta$ is defined as the minimum solution of the eigenvalue of $\Lambda_{\theta}$. Hence, the implicit theorem guarantees that $\lambda_\theta$ is $C^1$-continuous. Hence, we find that $\phi$ is $C^1$-continuous.
Since $\phi_{n,\rho}$ is convex, the limit $\lim_{n \to \infty} \frac{1}{\rho} \phi_{n,\rho}(\theta)$ is also convex, i.e. $\phi$ is convex. Therefore, $\frac{d\phi}{df}$ is a continuously increasing function.

4. Bregman divergence and its variants

Since the cumulant generating function $\phi_n(\theta)$ is convex, this corollary implies that $\phi(\theta)$ is convex. Since the eigenvalue is given as the solution of the eigenequation of the linear map $\Lambda_\theta$, and $\Lambda_\theta$ is differentiability with respect to $\theta$, the maximum eigenvalue $\lambda_\theta$ and the eigenvector $\rho_\theta$ are differentiable with respect to $\theta$. Therefore, $\phi'(\theta) := \frac{d\phi}{df}(\theta)$ is monotonically increasing with respect to $\theta$. So, we can define the inverse function $\phi'^{-1}$. For the latter discussion, we introduce Bregman divergence $D(\theta||\theta)$ [21] and the Rényi type of Bregman divergence $D_{1+s}(\theta||\theta)$ [22, 4.16] for the convex function $\phi$ as

$$D(\theta||\theta) := (\theta - \bar{\theta})\phi'(\theta) - \phi(\theta) + \phi(\bar{\theta}),$$  \hspace{1cm} (11)$$

$$D_{1+s}(\theta||\theta) := \frac{\phi((1+s)\theta - s\bar{\theta}) - (1+s)\phi(\theta) + s\phi(\bar{\theta})}{s}.$$  \hspace{1cm} (12)$$

Since the relation

$$\frac{dD_{1+s}(\theta||\theta)}{ds}$$

$$= \frac{\phi(\theta) - \phi(\theta + s(\theta - \bar{\theta})) - s(\theta - \bar{\theta})\phi'(\theta + s(\theta - \bar{\theta}))}{s^2}$$

$$= \frac{1}{s^2}\phi''(\theta + s(\theta - \bar{\theta}))(\xi s(\theta - \bar{\theta}))^2 > 0$$

holds with some parameter $\xi \in (0, 1)$, the Rényi type of Bregman divergence $D_{1+s}(\theta||\theta)$ is monotonically increasing with respect to $s$. Also, the Rényi type of Bregman divergence recovers Bregman divergence with the limit $s \to 0$ as

$$\lim_{s \to 0} D_{1+s}(\theta||\theta) = D(\theta||\theta).$$  \hspace{1cm} (13)$$

Thus, the properties of a differentiable convex function lead to the following lemma.

**Lemma 2.** When $a > \phi'(0)$,

$$\inf_{\theta > \phi'^{-1}(a)} \frac{\phi((1+s)\theta) - (1+s)\phi(\theta)}{s}$$

$$= \phi'^{-1}(a) - \phi(\phi'^{-1}(a)) = \sup_{\theta \geq 0} [\theta a - \phi(\theta)]$$

$$= D(\phi'^{-1}(a)||0).$$  \hspace{1cm} (14)$$

Similarly, when $a < \phi'(0)$,

$$\inf_{\theta < \phi'^{-1}(a)} \frac{\phi((1+s)\theta) - (1+s)\phi(\theta)}{s}$$

$$= \phi'^{-1}(a) - \phi(\phi'^{-1}(a)) = \sup_{\theta \leq 0} [\theta a - \phi(\theta)]$$

$$= D(\phi'^{-1}(a)||0).$$  \hspace{1cm}
5. Central limit theorem

Next, we discuss the central limit theorem for the sample mean $X^n$. Since $\text{Tr} \Lambda_{0}^{n}(\frac{d}{d\theta}|_{\theta=0}) = \text{Tr} \Lambda_{0}^{n} |_{\theta=0} = 0$, taking the derivative with respect to $\theta$ at $\theta = 0$ in (1), we have

$$n \sum_{\omega} x_{\omega} \text{Tr} C_{\omega}(\rho_{0}) = \text{Tr} \frac{d\Lambda_{0}^{n}}{d\theta}|_{\theta=0}(\rho_{0}) + \text{Tr} \Lambda_{0}^{n}(\frac{d\rho_{\theta}}{d\theta}|_{\theta=0})$$

$$= \frac{d\text{Tr} \Lambda_{0}^{n}(\rho_{0})}{d\theta}|_{\theta=0} = \frac{d\lambda_{0}^{n}(\theta)}{d\theta}|_{\theta=0} = n \phi'(0).$$ (15)

That is, the derivative $\phi'(0)$ expresses the expectation of $X$ when the initial state is the stationary state $\rho_{0}$. That is, to calculate the derivative $\phi'(0)$, it is enough to find the eigenvector $\rho_{0}$ of $\Lambda_{0}$ and calculate the expectation under the state $\rho_{0}$. Even when the initial state is not the stationary state $\rho_{0}$, we can see that the sample mean $X^n$ converges to the derivative $\phi'(0)$ in probability as follows.

Using theorem 1, we can characterize the cumulant generating function of the random variable $\sqrt{n}(X^n - \phi'(0))$ as follows.

**Theorem 2.** The cumulant generating function of the random variable $\sqrt{n}(X^n - \phi'(0))$ converges as follows.

$$\log E \left[ \exp \left[ \delta \sqrt{n}(X^n - \phi'(0)) \right] \right]$$

$$= \phi_{n}(\frac{\delta}{\sqrt{n}}) - \delta \sqrt{n}\phi'(0) + \frac{\delta^2}{2} \phi''(0).$$ (16)

**Proof.** Using (4) and (7), we have

$$\lim_{n \to \infty} \phi_{n}(\frac{\delta}{\sqrt{n}}) - \delta \sqrt{n}\phi'(0)$$

$$\leq \lim_{n \to \infty} n\phi(\frac{\delta}{\sqrt{n}}) - \delta \sqrt{n}\phi'(0) + \delta \phi(\frac{\delta}{\sqrt{n}})$$

$$= \lim_{n \to \infty} \delta^2 \phi\left(\frac{\delta}{\sqrt{n}}\right) - \left(\frac{\delta}{\sqrt{n}}\right)\phi'(0) = \frac{\delta^2}{2} \phi''(0).$$

Similarly, the opposite inequality can be shown by (4) and (7). Hence, we obtain the desired relation.

The right-hand side of (16) is the cumulant generating function of Gaussian distribution with the variance $\phi''(0)$ and average zero. Since the limit of the cumulant generating functions uniquely decides the limit of distribution functions [23], lemma 2 reproduces the central limit theorem as a corollary.

**Corollary 3.** The sample mean $X^n$ converges to $\phi'(0)$ in probability. The limiting distribution of $\sqrt{n}(X^n - \phi'(0))$ is characterized as

$$\lim_{n \to \infty} P \left\{ \sqrt{n}(X^n - \phi'(0)) \leq \delta \right\} = \Phi\left(\frac{\delta}{\sqrt{\phi''(0)}}\right),$$ (17)

where $\Phi(y) := \int_{-\infty}^{y} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$. 

M Hayashi and Y Yoshida

J. Phys. A: Math. Theor. 51 (2018) 335304
6. Tail probability

Next, we proceed to the evaluations for the tail probability. Using the notations \( \phi'^{-1}(a) \), \( \delta_{\rho}(\theta) \), \( \overline{\delta}_{\rho}(\theta) \), and \( \phi(\theta) \) defined in section 3 and (4) of theorem 1, we can derive the following lower bound on the exponent.

**Theorem 3.** For any \( a > \phi'(0) \), we have

\[
- \log \mathbb{P}\{X^n \geq a\} \geq \sup_{\theta \geq 0} [n \theta a - n \phi(\theta) - \overline{\delta}_{\rho}(\theta)] - n \phi'(0)(a) - \overline{\delta}_{\rho}(\phi'(0)(a)).
\]

Similarly, for \( a < \phi'(0) \), we have

\[
- \log \mathbb{P}\{X^n \leq a\} \geq \sup_{\theta \leq 0} [n \theta a - n \phi(\theta) - \overline{\delta}_{\rho}(\theta)] - n \phi'(0)(a) - \overline{\delta}_{\rho}(\phi'(0)(a)).
\]

**Proof.** These evaluations follow from proposition A.1 and lemma 4.1 of [13] and (4) of theorem 1. The proof is the same as theorem 8.1 of [13].

Further, using theorem 1, we can derive bounds of the opposite direction as follows.

**Theorem 4.** For any \( a > \phi'(0) \), we have

\[
- \log \mathbb{P}\{X^n \leq a\} \geq \inf_{\theta > \phi'(0)} \left\{ \frac{1}{s} \left[ n \phi((1 + s)\theta) - n(1 + s) \phi(\theta) + \overline{\delta}_{\rho}((1 + s)\theta) - \overline{\delta}_{\rho}(\theta) - (1 + s) \log \left( 1 - e^{-n \theta a - n \phi(\theta) + \overline{\delta}_{\rho}(\theta)} \right) \right] - (1 + s) \log \left( 1 - e^{-n \theta a - n \phi(\theta) + \overline{\delta}_{\rho}(\theta)} \right) \right\}
\]

\[
\leq \inf_{\theta > \phi'(0)} \left\{ \frac{1}{s} \left[ n \phi((1 + s)\theta) - n(1 + s) \phi(\theta) + \overline{\delta}_{\rho}((1 + s)\theta) - \overline{\delta}_{\rho}(\theta) - (1 + s) \log \left( 1 - e^{n \phi'(0)(a) - n \phi(\theta) + \overline{\delta}_{\rho}(\theta)} \right) \right] - (1 + s) \log \left( 1 - e^{n \phi'(0)(a) - n \phi(\theta) + \overline{\delta}_{\rho}(\theta)} \right) \right\}
\]

\[
= \inf_{\theta > \phi'(0)} \left\{ \frac{nD_{1+s}(\theta)}{s} + \frac{1}{s} \left[ \overline{\delta}_{\rho}((1 + s)\theta) - \overline{\delta}_{\rho}(\theta) \right] - \frac{1 + s}{s} \log \left( 1 - e^{-nD_{1+s}(\theta)} + \overline{\delta}_{\rho}((1 + s)\theta) - \overline{\delta}_{\rho}(\theta) \right) \right\}.
\]

Similarly, for any \( a < \phi'(0) \), we have
\[ \begin{align*}
&- \log P \{ X^n \leq a \} \\
&\leq \inf_{\theta > 0} \frac{1}{s} \left[ n \phi((1+s)\theta) - n(1+s)\phi(\theta) + \delta(1+s)\theta - \tilde{\delta}(\theta) \right] \\
&\quad - (1+s) \log \left( 1 - e^{-n[\delta(1+s)\theta - \phi(\theta) + \tilde{\delta}(\theta)]} \right) \\
&= \inf_{\theta < \phi^{-1}(a)} nD_{1+s}(\theta\|0) + \frac{1}{s} [\tilde{\delta}(1+s)\theta - \tilde{\delta}(\theta)] \\
&\quad - \frac{1 + s}{s} \log \left( 1 - e^{-nD(\phi^{-1}(a)\theta + \tilde{\delta}(\phi^{-1}(a)) - \tilde{\delta}(\theta))} \right).
\end{align*} \]

**Proof.** The proof can be shown from theorem A.2 of [13] in the same way as theorem 8.2 of [13].

The computational complexity does not depend on the number \( n \) of observations in the above upper and lower bounds in theorems 3 and 4. Hence, the above upper and lower bounds are \( O(1) \)-computable. These also attain asymptotic tightness in the large and moderate deviation regimes as shown in the following corollaries, although the large and moderate deviation regimes follow immediately from the combination of Gärtner–Ellis theorem [24] and corollary 2.

From lemma 2 and theorems 3 and 4, we can derive the evaluation in the large deviation regime.

**Corollary 4.** For arbitrary \( \delta > 0 \), we have

\[ \begin{align*}
\lim_{n \to \infty} - \frac{1}{n} \log P \{ X^n - \phi(0) \geq \delta \} &= \sup_{\theta > 0} \left[ \theta \phi(0) + \theta \right], \\
\lim_{n \to \infty} - \frac{1}{n} \log P \{ X^n - \phi(0) \leq -\delta \} &= \sup_{\theta > 0} \left[ \theta \phi(0) - \theta \right].
\end{align*} \]  

**Proof.** As mentioned at the end of section 3, \( \frac{\delta}{\phi} \) is a continuously increasing function. Hence, the RHSs of (20) and (21) are continuous.

Lemma 2 guarantees that (RHS of (18))\( n \) goes to (RHS of (20)). In the RHS of (19), for given \( s > 0 \) and \( \theta > \phi^{-1}(a) \), the value \( e^{-nD(\phi^{-1}(a)\theta + \tilde{\delta}(\phi^{-1}(a)) - \tilde{\delta}(\theta))} \) and \( \frac{1}{n}[\tilde{\delta}(1+s)\theta - \tilde{\delta}(\theta)] \) go to zero as \( n \to \infty \). So, (RHS of (19))\( n \) goes to (RHS of (20)). We can show (21) in the same way.

From theorems 3 and 4, we can derive the evaluation in the moderate deviation regime.

**Corollary 5.** For arbitrary \( t \in (0, 1/2) \) and \( \delta > 0 \), we have

\[ \begin{align*}
\lim_{n \to \infty} - \frac{1}{n^{1-2t}} \log P \{ X^n - \phi(0) \geq n^{-t} \delta \} &= \frac{\delta^2}{2\phi''(0)}.
\end{align*} \]
\[ \lim_{n \to \infty} \frac{1}{n^{1-2t}} \log \mathbb{P} \{ X^n - \phi'(0) \leq -n^{-t} \delta \} = \frac{\delta^2}{2\phi''(0)} . \]  

**Proof.** Corollary 5 can be shown by using lemma 2 and theorems 3 and 4 in the same way as the proof of corollary 8.4 of [13]. \( \square \)

7. Calculation formula of \( \phi''(0) \)

When \( \Lambda = \sum \omega C_\omega \) is primitive, as a quantum version of theorem 7.7 of [13], this section gives a useful formula for the calculation of \( \phi''(0) \), which is used for the central limit theorem and the evaluation in the moderate deviation regime. In this derivation, condition (P-i) for the primitivity plays an essential role.

**Proposition 5.** Let \( \tilde{\Lambda} \) be the trace-preserving completely positive map on \( \mathcal{T}(\mathcal{H}) \) which maps any Hermitian matrix \( H \) to \( (\text{Tr} H)\rho_0 \). Then, the map \( Z := (\iota - (\Lambda - \tilde{\Lambda}))^{-1} \) exists and the equations

\[ \tilde{\Lambda} = \lim_{n \to \infty} \Lambda^n , \]

\[ Z = \sum_{n=0}^\infty (\Lambda - \tilde{\Lambda})^n = \iota + \sum_{n=1}^\infty (\Lambda^n - \tilde{\Lambda}) \]  

hold, where \( \iota \) is the identity map on \( \mathcal{T}(\mathcal{H}) \).

The map \( Z \) is called the fundamental matrix in the classical case [25], and so \( Z \) can be regarded as a quantum version of the fundamental matrix.

**Proof.** The equation (24) follows from condition (P-i). From the definition, the equation \( \tilde{\Lambda} \circ \Lambda = \Lambda \circ \tilde{\Lambda} = \tilde{\Lambda}^2 = \tilde{\Lambda} \) holds. We show that the magnitude of any eigenvalue of \( \Lambda - \tilde{\Lambda} \) is smaller than one. Let \( \lambda \in \mathbb{C} \) be an eigenvalue of \( \Lambda - \tilde{\Lambda} \) and \( M \) be an eigenvector associated with \( \lambda \) of \( \Lambda - \tilde{\Lambda} \). Then any natural number \( n \) satisfies \( \lambda^n M = (\Lambda - \tilde{\Lambda})^n M = (\Lambda^n - \tilde{\Lambda})(M) \). The equation (24) implies that \( \lambda^n M = (\Lambda^n - \tilde{\Lambda})(M) \to 0 \). Hence, we obtain \( \lambda^n \to 0 \), which means that the magnitude of \( \lambda \) is smaller than one.

Since the magnitude of any eigenvalue of \( \Lambda - \tilde{\Lambda} \) is smaller than one, the map \( Z \) exists. Let \( Z' := \sum_{n=0}^\infty (\Lambda - \tilde{\Lambda})^n \). Then the equation \( Z' \circ (\iota - (\Lambda - \tilde{\Lambda})) = \iota \) means \( Z = Z' \). \( \square \)

**Theorem 5.** Define the random variable \( X \) as \( X(\omega) = x_\omega \) and the map \( C_X \) on \( \mathcal{T}(\mathcal{H}) \) as \( C_X := \sum \omega X(\omega)C_\omega \). Then,

\[ \phi''(0) = \mathcal{V}_{\rho_0}[X] + 2\text{Tr} C_X \circ (Z - \tilde{\Lambda}) \circ C_X(\rho_0) , \]

where \( \mathcal{V}_{\rho}[X'] \) is the variance of a random variable \( X' \) under an initial state \( \rho \).

The symbols \( \rho_0, \tilde{\Lambda}, C_X \) and \( Z \) can be calculated from the instrument \( \{ C_\omega \}_{\omega \in \Omega} \) and the random variable \( X \). Hence, we can calculate \( \phi''(0) \) using the above formula.

**Remark 1.** Here, we explain the relationship with the results presented in [14, 17]. The paper [14] considered the case when the correlation is generated by an isometry, and derived the central limit theorem for the sample mean of the measurement outcomes. The combina-
tion of the isometry and the measurement can be written as an instrument \( \{ C_\omega \}_{\omega \in \Omega} \). However, the paper [14] addressed multiple variables. Hence, when we focus on a single variable, our model contains the model in [14] as a special case. In addition, the paper [14] did not give the relation between the asymptotic variance and the second derivative of \( \phi \).

In contrast, the paper [17] addressed local asymptotic normality in the context of system identification for quantum Markov chains. To discuss local asymptotic normality, the paper [17] calculated the Fisher information matrix. Hence, it did not discuss the random variable \( X^n \).

**Proof.** Using the equation \( \phi''_{n,\rho}(0) = V_\rho[nX^n] \), lemma 4 in appendix, and theorem 2, we have

\[
\lim_{n \to \infty} (1/n) \phi''_{n,\rho}(0) = \lim_{n \to \infty} V_\rho[\sqrt{n}X^n] = \lim_{n \to \infty} V_\rho[\sqrt{n}(X^n - \phi'(0))] = \lim_{n \to \infty} \left( E_\rho[(\sqrt{n}(X^n - \phi'(0)))^2] - E_\rho[\sqrt{n}(X^n - \phi'(0))]^2 \right) = \phi''(0),
\]

where \( E_\rho[X^n] \) denotes the expectation of a random variable \( X^n \) under an initial state \( \rho \). Put \( \rho = \rho_0 \). Noting \( \Lambda(\rho) = \rho \), we have

\[
E_\rho[X_i] = \sum_{\omega_1, \ldots, \omega_j} X(\omega_i) Tr C_{\omega_j} \cdots C_{\omega_j}(\rho) = \sum_{\omega_j} X(\omega_i) Tr \Lambda^{n-i} C_{\omega_j} \Lambda^{i} = \sum_{\omega_j} X(\omega_i) Tr C_{\omega_j}(\rho) = E_\rho[X].
\]

Similarly, \( E_\rho[X_i^2] = E_\rho[X^2] \) holds. Combining these equations, we have

\[
\phi''_{n,\rho}(0) = V_\rho[nX^n] = E_\rho\left[ \left( \sum_{i=1}^{n} X_i \right)^2 - n E_\rho \left[ \sum_{i=1}^{n} X_i \right]^2 \right]
\]

which gives

\[
= E_\rho[X_i] - n(n-1) E_\rho[X]^2 + 2 \sum_{i< j} E_\rho[X_i X_j] \quad \text{(a)}
\]

where \( \rho = \rho_0 \) and \( \Lambda \) is the covariance matrix of \( X \).
where the equation $\Lambda(\rho) = \rho$ has been used to obtain the equality (a) and the last equality follows from $\text{Tr} C_X \circ \Lambda \circ C_X(\rho) = (\text{Tr} C_X(\rho))^2 = \rho[2]^2$.

Taking the Cesáro mean for $Z - \iota = \sum_{n=1}^{\infty} (\Lambda^n - \tilde{\Lambda})$, we have

$$Z - \iota = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (n - k + 1)(\Lambda^k - \tilde{\Lambda}).$$

Hence, we obtain (25) as

$$\phi^n(0) = \lim_{n \to \infty} \frac{1}{n} \phi^n_{a\rho}(0)$$

$$= \mathcal{V}_\rho [X] + \lim_{n \to \infty} \frac{2}{n} \sum_{k=0}^{n-2} (n - k - 1) \text{Tr} C_X \circ (\Lambda^k - \tilde{\Lambda}) \circ C_X(\rho)$$

$$= \mathcal{V}_\rho [X] + 2 \text{Tr} C_X \circ (Z - \iota) \circ C_X(\rho) + 2 \text{Tr} C_X \circ (\iota - \tilde{\Lambda}) \circ C_X(\rho).$$

8. Relation to finitely correlated states

The paper [15] considered finitely correlated states. Finitely correlated states play a key role in analysis on quantum spin chains. The meaning of large deviation type evaluation in such a physical system is explained in [15, 16].

Here, we use a notation for a Hermitian matrix $B$ and a real number $a$: $\{B > a\}$ is defined to be the projection $\sum_{j, b_j > a} E_j$ when the Hermitian matrix $B$ has the spectral decomposition $\sum b_j E_j$. In the one-dimensional case of this model, we consider a TP-CP map $\Gamma$ from $\mathcal{T}(\mathcal{H})$ to $\mathcal{T}(\mathbb{C}^d) \otimes \mathcal{T}(\mathcal{H})$. Then, from an initial state $\rho$ on $\mathcal{H}$, we generate the state on $(\mathbb{C}^d)^\otimes n$ as

$$\rho_n := (\text{id}_{\mathcal{T}(\mathbb{C}^d) \otimes \Gamma}) \circ \cdots \circ (\text{id}_{\mathcal{T}(\mathbb{C}^d) \otimes \Gamma}) \circ \Gamma(\rho).$$

(26)

Generally, in this model, we consider the behavior of the average of the observables across several systems. To discuss the relation with our paper, we focus only on the case when the observables are given as a Hermitian matrix $A$ on the single system $\mathbb{C}^d$.

In this special case, the paper [15] discussed the probability $\text{Tr} \rho_n \{1/n \sum_{i=1}^{n} A_i > a\}$, where $A_i$ is defined as $I \otimes \cdots \otimes I \otimes A \otimes I \otimes \cdots \otimes I$. That is, it derived the exponential decreasing rate of the probability $\text{Tr} \rho_n \{1/n \sum_{i=1}^{n} A_i > a\}$ when $n$ goes to infinity. To convert this model to our model, we make the spectral decomposition of $A$ as $A = \sum_{\omega} x_{\omega} E_{\omega}$. Then, we define the CP map $C_\omega$ on $\mathcal{H}$ as $C_\omega(\rho) := \text{Tr}_{\mathbb{C}^d} (E_{\omega} \otimes I) \Gamma(\rho)$. We define the $i$th variable $X_i$ to be $x_{\omega_i}$ when $\omega_i$ is the $i$th observation. Thus, we have the relation

$$P \{X^n > a\} = \text{Tr} \rho_n \left( \left\{ \frac{1}{n} \sum_{i=1}^{n} A_i > a \right\} \otimes I \right),$$

(27)

which shows that our model includes finitely correlated states.

Conversely, when given a set of CP maps (an instrument) $C = \{C_\omega\}_{\omega \in \Omega}$, the map $\sum_{\omega \in \Omega} C_\omega$ is trace-preserving, we define the TP-CP map $\Gamma(\rho) := \sum_{\omega \in \Omega} |\omega\rangle\langle\omega| \otimes C_\omega(\rho)$ from $\mathcal{T}(\mathcal{H})$ to $\mathcal{T}(\mathbb{C}^d) \otimes \mathcal{T}(\mathcal{H})$ and define the Hermitian matrix $A := \sum_{\omega} x_{\omega} |\omega\rangle\langle\omega|$, where $d$ denotes the
number of elements of $\Omega$. Then, the relation (27) holds. Therefore, we can conclude that our model is equivalent to the model of finitely correlated states when the observable is given as a Hermitian matrix $A$ on the single system $C^d$.

When we consider the behavior of the average of the observables across more than one system, we need to care about the non-commutativity between several observables in general. This type analysis is more difficult than that in this paper, but if we obtain analysis similar to theorem 1, we can derive analysis similar to theorems 3–5 by using the same discussion. Therefore, corollary 4 can be regarded as a special case of [15, theorem 1.2 & 27]. However, in this case, our analysis has the following two advantages over the analysis of the paper [15]. Indeed, the paper [15] did not give any non-asymptotic evaluation. On the other hand, we have given the non-asymptotic evaluation of the probability (27) as theorems 3 and 4, and have shown their asymptotic tightness. In this sense, the analysis of this paper is more advanced in this special case of finitely correlated states than that in [15]. The contribution of this paper is the derivation of the non-asymptotic evaluation to achieve asymptotic tightness.

9. Discussion

This paper has addressed a hidden Markovian process with quantum hidden system, and has derived several asymptotic behaviors of the sample mean—namely, the central limit theorem and the large and moderate deviations for the tail probability—while no existing paper has treated the non-asymptotic behaviors of such a hidden Markovian process with quantum hidden system. Using the quantum version of the Perron–Frobenius theorem, we have derived upper and lower bounds of the cumulant generating function. Based on these bounds, we have obtained the above result by using the same method as [13].

In particular, corollary 3 can be regarded as the central limit theorem for a hidden Markov process with quantum hidden system. In the classical case, as the refinement of corollary 3, the paper [26, theorem 2] showed the Markov version of the Berry–Esseen theorem. So, it is an interesting problem to derive the Berry–Esseen theorem for the hidden Markov process with quantum hidden system.

Further, our method relies only on the quantum version of the Perron–Frobenius theorem, and the quantum version of the Perron–Frobenius theorem can be extended to a finite-dimensional vector space with a general closed convex cone, in which the dual convex cone need not be equal to the original convex cone [20, 27–29]. Therefore, our result can be extended to a more general setting, e.g. general probabilistic theory [30–35].

When the system is given as a hidden system $X_1, \ldots, X_n$, the memory effect does not vanish, i.e. $P_{X_i|X_{i-1}, \ldots, X_{i-k}} (x_i|x_{i-1}, \ldots, x_{i-k})$ cannot be simplified to $P_{X_i|X_{i-1}, \ldots, X_{i-k'}} (x_i|x_{i-1}, \ldots, x_{i-k'})$ with $k' < k$ [37]. Hence, the outcome has long-period memory. When discussing some information theoretic problems, we need to discuss information theoretical quantities, e.g. entropy and conditional entropy, instead of the sample mean [12]. In such cases, we need to be careful with respect to such a memory. Hence, it is not sufficient to discuss the sample mean of a random variable for this purpose. For such applications, we need a more complicated calculation.

Acknowledgments

MH is very grateful to Professor Shun Watanabe for helpful discussions and comments. He was supported in part by a JSPS Grant-in-Aids for Scientific Research (B) No.16KT0017
and for Scientific Research (A) No.17H01280, the Okawa Research Grant and Kayamori Foundation of Information Science Advancement.

Appendix

For readers’ convenience, we prove a few well-known lemmas used in section 7. The textbook [36, corollary in p 338] has a more general statement than lemma 4.

**Lemma 3.** Let \( \mu_n \) and \( \mu \) be probability distributions on \( \mathbb{R} \), and \( C \) be the set of all continuous points of the cumulative distribution function of \( \mu \). Assume \( \mu_n \to \mu \). Then, any continuous function \( f : \mathbb{R} \to \mathbb{R} \) and any real numbers \( a, b \in C \) with \( a < b \) satisfy

\[
\lim_{n \to \infty} \int_{(a,b]} f(x) \mu_n(dx) = \int_{(a,b]} f(x) \mu(dx).
\]

Further, any continuous function \( f : \mathbb{R} \to \mathbb{R} \) with \( \sup_{x>0} |f(x)| < \infty \) and any real number \( a \in C \) satisfy

\[
\lim_{n \to \infty} \int_{(a,\infty)} f(x) \mu_n(dx) = \int_{(a,\infty)} f(x) \mu(dx).
\]

Here, \( \mu_n \to \mu \) means that \( \lim_{n \to \infty} \int_{\mathbb{R}} f(x) \mu_n(dx) = \int_{\mathbb{R}} f(x) \mu(dx) \) for any bounded continuous function \( f : \mathbb{R} \to \mathbb{R} \).

**Proof.** Assume \( \mu_n \to \mu \). Define the cumulative distribution functions \( F_n \) and \( F \) of \( \mu_n \) and \( \mu \) as \( F_n(x) := \mu_n((-\infty, x]) \) and \( F(x) := \mu((-\infty, x]) \), respectively. First, we prove that any point \( a \in C \) satisfies \( \lim_{n \to \infty} F_n(a) = F(a) \). Take an arbitrary positive number \( \epsilon \) and define the two bounded continuous functions \( h_\pm \) as

\[
h_+(x) := \begin{cases} 
1 & x \leq a, \\
(a+\epsilon-x)/\epsilon & a < x < a+\epsilon, \\
0 & x \geq a+\epsilon,
\end{cases}
\]

\[
h_-(x) := \begin{cases} 
1 & x \leq a-\epsilon, \\
(a-\epsilon-x)/\epsilon & a-\epsilon < x < a, \\
0 & x \geq a.
\end{cases}
\]

Then, since the inequalities \( F_n(a) \leq \int_{\mathbb{R}} h_+(x) \mu_n(dx) \) and \( F_n(a) \geq \int_{\mathbb{R}} h_-(x) \mu_n(dx) \) hold, by using the assumption \( \mu_n \to \mu \), we obtain

\[
\limsup_{n \to \infty} F_n(a) \leq \limsup_{n \to \infty} \int_{\mathbb{R}} h_+(x) \mu_n(dx) = \int_{\mathbb{R}} h_+(x) \mu(dx) \leq F(a+\epsilon),
\]

\[
\liminf_{n \to \infty} F_n(a) \geq \liminf_{n \to \infty} \int_{\mathbb{R}} h_-(x) \mu_n(dx) = \int_{\mathbb{R}} h_-(x) \mu(dx) \geq F(a-\epsilon).
\]

By taking the limit \( \epsilon \downarrow 0 \), these inequalities turn out \( \limsup_{n \to \infty} F_n(a) = F(a) \) and \( \liminf_{n \to \infty} F_n(a) = F(a) \) because of \( a \in C \). That is, the desired equation \( \lim_{n \to \infty} F_n(a) = F(a) \) holds.

Next, we prove the first equation in this lemma. Let \( f : \mathbb{R} \to \mathbb{R} \) be a bounded continuous function and \( a, b \in C \) be two real numbers satisfying \( a < b \). Then, we define the function \( h : \mathbb{R} \to \mathbb{R} \) as
\( h(x) := \begin{cases} f(a) & x \leq a, \\ f(x) & a < x < b, \\ f(b) & x \geq b. \end{cases} \)

Since the assumption \( \mu_n \to \mu \) and the above proof imply that \( \lim_{n \to \infty} \int_{\mathbb{R}} h(x) \mu_n(dx) = \int_{\mathbb{R}} h(x) \mu(dx) \) and \( \lim_{n \to \infty} F_n(x) = F(x) \) with \( x = a, b \), respectively, we obtain

\[
\lim_{n \to \infty} \int_{[a,b]} f(x) \mu_n(dx) + f(a)F(a) + f(b)(1 - F(b)) = \lim_{n \to \infty} \left[ \int_{[a,b]} f(x) \mu_n(dx) + f(a)F_n(a) + f(b)(1 - F_n(b)) \right] = \lim_{n \to \infty} \int_{[a,b]} h(x) \mu_n(dx) = \int_{[a,b]} h(x) \mu(dx) = \int_{[a,b]} f(x) \mu(dx) + f(a)F(a) + f(b)(1 - F(b)).
\]

Therefore, \( \lim_{n \to \infty} \int_{[a,b]} f(x) \mu_n(dx) = \int_{[a,b]} f(x) \mu(dx) \) holds. Similarly, the second equation in this lemma can also be shown. \( \square \)

**Lemma 4 (Convergence of moments).** Let \( \mu_n \) and \( \mu \) be probability distributions on \( \mathbb{R} \). If the cumulant generating functions of \( \mu_n \) and \( \mu \) exist and the cumulant generating functions of \( \mu_n \) converge pointwise to that of \( \mu \), then any natural number \( k \) satisfies

\[
\lim_{n \to \infty} \int_{\mathbb{R}} x^k \mu_n(dx) = \int_{\mathbb{R}} x^k \mu(dx).
\]

**Proof.** Assume the cumulant generating functions of \( \mu_n \) converge to that of \( \mu \), which implies \( \mu_n \to \mu \). Let \( \theta \) be an arbitrary positive number and take a continuous point \( a \in \mathcal{C} \). It is sufficient to prove \( \lim_{n \to \infty} \int_{\mathbb{R}} (x-a)^k \mu_n(dx) = \int_{\mathbb{R}} (x-a)^k \mu(dx) \) for any natural number \( k \). Since \( e^{\theta(x-a)} \) is a continuous function, lemma 3 implies

\[
\lim_{n \to \infty} \int_{(-\infty,a]} e^{\theta(x-a)} \mu_n(dx) = \int_{(-\infty,a]} e^{\theta(x-a)} \mu(dx).
\]

This equation and the convergence of the cumulant generating functions imply

\[
\lim_{n \to \infty} \int_{(a,\infty)} e^{\theta(x-a)} \mu_n(dx) = \int_{(a,\infty)} e^{\theta(x-a)} \mu(dx).
\]

Combining the above equation and the inequality \( e^{\theta(x-a)} \geq \theta(x-a) + 1 \), we have

\[
\lim_{n \to \infty} \int_{(a,\infty)} (x-a)^k \mu_n(dx) \leq \lim_{n \to \infty} \int_{(a,\infty)} \left( \frac{e^{\theta(x-a)} - 1}{\theta} \right)^k \mu_n(dx) = \int_{(a,\infty)} \left( \frac{e^{\theta(x-a)} - 1}{\theta} \right)^k \mu(dx).
\]

17
Since \((e^{\theta(x-a)} - 1)/\theta\) monotonically increases with respect to \(\theta\), the monotone convergence theorem yields
\[
\int_{(a,\infty)} \left(\frac{e^{\theta(x-a)} - 1}{\theta}\right)^k \mu(dx) \xrightarrow{\theta \downarrow 0} \int_{(a,\infty)} (x-a)^k \mu(dx).
\]

Hence,
\[
\limsup_{n \to \infty} \int_{(a,\infty)} (x-a)^k \mu_n(dx) \leq \int_{(a,\infty)} (x-a)^k \mu(dx).
\]

Take an arbitrary real number \(b \in \mathcal{C}\) satisfying \(a < b\). Since \((x-a)^k\) is a continuous function, lemma 3 implies
\[
\liminf_{n \to \infty} \int_{(a,\infty)} (x-a)^k \mu_n(dx) \geq \liminf_{n \to \infty} \int_{[a,b]} (x-a)^k \mu_n(dx) = \int_{[a,b]} (x-a)^k \mu(dx).
\]

Note that \(\mathbb{R} \setminus \mathcal{C}\) is a countable set because the cumulative distribution function \(F\) increases monotonically. Thus, we can take the limit \(b \to \infty\) while satisfying \(b \in \mathcal{C}\). Taking this limit, we have
\[
\liminf_{n \to \infty} \int_{(a,\infty)} (x-a)^k \mu_n(dx) \geq \int_{(a,\infty)} (x-a)^k \mu(dx).
\]

Thus, the equation \(\lim_{n \to \infty} \int_{(a,\infty)} (x-a)^k \mu_n(dx) = \int_{(a,\infty)} (x-a)^k \mu(dx)\) holds. The other equation \(\lim_{n \to \infty} \int_{(-\infty,a]} (x-a)^k \mu_n(dx) = \int_{(-\infty,a]} (x-a)^k \mu(dx)\) can be also shown in a similar way. From these equations, we obtain the desired equation. \(\square\)

**ORCID iDs**

Masahito Hayashi @ https://orcid.org/0000-0003-3104-1000
Yuuya Yoshida @ https://orcid.org/0000-0003-1071-2098

**References**

[1] Herrero-Collantes M and Garcia-Escartin J C 2017 Quantum random number generators Rev. Mod. Phys. 89 015004
[2] Caruso E, Giovannetti V, Lupo C and Mancini S 2014 Quantum channels and memory effects Rev. Mod. Phys. 86 1203
[3] Cao M X and Vontobel P O 2017 Estimating the information rate of a channel with classical input and output and a quantum state (arXiv:1705.01041)
[4] Kretschmann D and Werner R F 2005 Quantum channels with memory Phys. Rev. A 72 062323
[5] Ozawa M 1984 Quantum measuring processes of continuous observables J. Math. Phys. 25 79
[6] Kontoyiannis I and Meyn S P 2003 Spectral theory and limit theorems for geometrically ergodic Markov processes Ann. Appl. Probab. 13 304–62
[7] Meyn S P and Tweedie R L 1993 Markov Chains and Stochastic Stability (London: Springer)
[8] Jones G L 2004 On the Markov chain central limit theorem Probab. Surv. 1 299–320
[9] Ben-Ari I and Neumann M 2012 Probabilistic approach to perron root, the group inverse, and applications Linear Multilinear Algebra. 60 39–63
[10] Donsker M D and Varadhan S R S 1975 Asymptotic evaluation of certain Markov process expectations for large time, I, II Commun. Pure Appl. Math. 28 1–47
Donsker M D and Varadhan S R S 1975 Asymptotic evaluation of certain Markov process expectations for large time, I, II Commun. Pure Appl. Math. 2 279–301

[11] Dembo A and Zeitouni O 1998 Large Deviations Techniques and Applications 2nd edn (New York: Springer)

[12] Hayashi M and Watanabe S 2013 Finite-length analyses for source and channel coding on Markov chains (arXiv:1309.7528)

[13] Watanabe S and Hayashi M 2017 Finite-length analysis on tail probability for Markov chain and application to simple hypothesis testing Ann. Appl. Probab. 27 811–45

[14] Horssen M and Guta M 2015 Sanov and central limit theorems for output statistics of quantum Markov chains J. Math. Phys. 56 022109

[15] Ogata Y 2010 Large deviations in quantum spin chains Commun. Math. Phys. 296 35–68

[16] Budini A A, Turner R M and Garrahan J P 2014 Fluctuating observation time ensembles in the thermodynamics of trajectories J. Stat. Mech. P03012

[17] Guta M and Kiukas J 2015 Equivalence classes and local asymptotic normality in system identification for quantum Markov chains Commun. Math. Phys. 335 1397–428

[18] Schrader R 2001 Perron–Frobenius theory for positive maps on trace ideals (Mathematical physics in mathematics and physics) (Siena, 2000) Fields Inst. Commun. 30 361–78

[19] Wolf M M 2012 Quantum channels & operations: guided tour (www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf)

[20] Yoshida Y and Hayashi M 2018 Mixing and asymptotically decoupling properties in general probabilistic theory (arXiv:1801.03988)

[21] Bregman L 1967 The relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming USSR Comput. Math. Math. Phys. 7 200–17

[22] Hayashi M and Watanabe S 2016 Information geometry approach to parameter estimation in Markov chains Ann. Stat. 44 1495–535

[23] Rao C R 1973 Linear Statistical Inference and its Applications 2nd edn (New York: Wiley)

[24] Gurtman J 1977 On large deviations from the invariant measure Theory Probab. Appl. 22 24–39

[25] Kemeny J G and Snell J L 1960 Finite Markov Chains (New York: Springer)

[26] Hervé L, Ledoux J and Patilea V 2012 A uniform Berry–Esseen theorem on M-estimators for geometrically ergodic Markov chains Bernoulli 18 703–34

[27] Vandergraft J S 1968 Spectral properties of matrices which have invariant cones SIAM J. Appl. Math. 16 1208–22

[28] Barker G P 1972 On matrices having an invariant cone Czech. Math. J. 22 49–68

[29] Barker G P and Schneider H 1975 Algebraic Perron–Frobenius theory Linear Algebr. Appl. 11 219–33

[30] Barnum H, Barrett J, Leifer M and Wilce A 2007 Generalized no-broadcasting theorem Phys. Rev. Lett. 99 240501

[31] Barnum H, Barrett J, Leifer M and Wilce A 2012 Teleportation in general probabilistic theories foundations of information flow Proc. Sympos. Appl. Math. 71 25–47

[32] Barrett J, Hardy L and Kent A 2005 No signaling and quantum key distribution Phys. Rev. Lett. 95 010503

[33] Gudder S P 1979 Stochastic Method in Quantum Mechanics (Amsterdam: North-Holland)

[34] Gudder S P 1988 Quantum Probability (New York: Academic)

[35] Kimura G, Miyadera T and Imai H 2009 Optimal state discrimination in general probabilistic theories Phys. Rev. A 79 062306

[36] Billingsley P 1995 Probability and Measure 3rd edn (New York: Wiley)

[37] Cappé O, Moulines E and Ryden T 2005 Inference in Hidden Markov Models (Springer Series in Statistics) (New York: Springer)