Alma Mater Studiorum · Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE STATISTICHE

Ciclo XXVIII

Settore Concorsuale di afferenza: 13/D1

Settore Scientifico disciplinare: SECS-S01

STATISTICAL INFERENCE IN OPEN QUANTUM SYSTEMS

Presentata da: Marco Novelli

Coordinatore dottorato:
Prof.ssa Alessandra Luati

Relatore:
Prof.ssa Alessandra Luati

Esame finale anno 2016

To my beloved parents

English Abstract

This thesis concerns the statistical analysis of open quantum systems, that is, systems that are in interaction with an external quantum system, often called the environment. Specifically, the aim is to develop inferential techniques to extract the maximum information from an open quantum system subject to an external and non-stationary perturbation. The thesis is composed of three paper.

In the first paper, a generalization of the explicit-duration hidden Markov models (EDHMM) which takes into account the presence of sparse data is presented. Introducing a kernel estimator in the estimation procedure increases the accuracy of the estimates, and thus allows one to obtain a more reliable information about the evolution of the unobservable system. A generalization of the Viterbi algorithm to EDHMM, used to reconstruct the hidden chain, is developed. The finite sample properties of our formulation are assessed through an extensive simulation study.

In the second paper, we develop a Markov Chain Monte Carlo (MCMC) procedure for estimating the EDHMM. We improve the flexibility of our formulation by adopting a Bayesian model selection procedure which allows one to avoid a direct specification of the number of states of the hidden chain. Motivated by the presence of sparsity, we make use of a non-parametric estimator to obtain more accurate estimates of the model parameters. The formulation presented turns out to be straightforward to implement, robust against the underflow problem and provides accurate estimates of the parameters. A comparative analysis with the EM-based method is also carried out.

In the third paper, an extension of the Cramér-Rao inequality for quantum discrete parameter models is derived. The latter are models in which the parameter space is restricted to a finite set of points. The extension presented sets the ultimate accuracy of an estimator, and determines a discrete counterpart of the quantum Fisher information. This is particularly useful in many experiments in which the parameters can assume only few different values: for example, the direction which the magnetic field points to. We also provide an illustration related to a quantum optics problem.

Italian Abstract

Lo scopo della tesi è sviluppare metodi inferenziali in sistemi quantistici aperti, ovvero, quei sistemi che interagiscono con un sistema quantistico esterno chiamato ambiente. In particolare, lo scopo è quello di estrarre la massima informazione da un sistema quantistico aperto soggetto ad una perturbazione non stazionaria. La tesi si compone di tre articoli.

Nel primo articolo viene presentata una generalizzazione dei modelli explicit-duration HMM che tiene conto della presenza di dati sparsi (ovvero, mancanza di infomazione nei piccoli campioni). L'introduzione di uno stimatore kernel migliora considerevolmente l'accuratezza delle stime e permette di ottenere un'informazione affidabile sull'evoluzione del sistema non osservabile. Inoltre, al fine di rocostruire l'evoluzione del sistema nascosto, viene presentata una generalizzazione dell'algoritmo di Viterbi per i modelli EDHMM.

Nel secondo articolo viene sviluppata una procedura Monte Carlo Markov Chain (MCMC) per la stima dei modelli EDHMM. La flessibilità della procedura di stima viene aumentata adottando un metodo di selezione del modello in ottica Bayesiana, il quale permette di non specificare a priori il numero degli stati della catena nascosta. L'introduzione di uno stimatore non parametrico, motivata dalla presenza di dati sparsi, permette di ottenere stime dei parametri del modello più precise. Il modello presentato risulta essere facile ed intuitivo da implementare, robusto rispetto al problema dell'underflow e fornisce stime dei parametri accurate.

Nel terzo articolo viene sviluppata una estensione della disuguaglianza di Cramér-Rao per modelli quantistici a parametri discreti, ovvero quei modelli il cui spazio parametrico è ristretto ad un insieme finito di punti. L'estensione presentata determina la minima varianza ottenibile da uno stimatore e fornisce una controparte discreta dell'informazione quantistica di Fisher. Essa trova applicazione negli esperimenti in cui i parametri possono assumere soltanto pochi valori: ad esempio, la direzione del campo magnetico. Inoltre, viene presentata un'illustrazione relativa ad un modello di ottica quantistica.

Contents

1 Introduction

According to the von Neumann projection postulate (Von Neumann (1955)), describing the effects of a measurement on a quantum system, a quantum measurement induces a discontinuous stochastic change in the state of the measured system, the so-called wave function collapse. The latter implies that, when a particular measurement is performed, the future results of the same measurement on the same system are fixed, preventing to obtain further information from the same system.

In recent years, the first quantum technologies are emerging due to the experimental advances that allow the researchers to perform repeated measurements on the same quantum system. The novelty lies in the possibility to control and measure a quantum system without causing the collapse of the wave function. This has been possible thanks to the development of a new measurement technique, the indirect or generalized measurement (Davies (1976); Kraus et al. (1983)).

The latter is realized by coupling a quantum system to a "meter", i.e. an auxiliary quantum system, and then measuring the meter. This implies that no direct perturbation affects the system. Here, the difference between classical and quantum mechanics emerges: in fact, in classical systems, the effect of the measurement on the system can be neglected; on the other hand, in quantum systems, measurement always causes an effect on the system, even if carried out "indirectly".

In the context of generalized measurement, the central role is played by open quantum systems, that is, systems that interacts with an external quantum system, often called the environment. Indeed, since the interaction with the environment is typically continuous, it is possible to monitor the environment so as to realize a continuous generalized measurement on the system. (Wiseman and Milburn (2009), Nielsen and Chuang (2010)). Open quantum systems find wide applications in many fields such as quantum measurement theory, quantum optics, quantum statistical mechanics and quantum computing (Breuer and Petruccione (2002); Attal et al. (2006); Carmichael (2009)).

This thesis deals with the problem of extracting the maximum information form an open quantum system. Specifically, the aim is to develop inferential techniques to provide a statistical analysis of an open quantum system subject to an external and non-stationary perturbation.

The work can be divided into two main parts.

1. In the first part we develop a physical model (experimental set-up) which could be a useful and versatile tool to address a broad variety of physical and statistical problems. (The model was developed in collaboration with Prof. Klaus Mølmer at Aarhus University, Denmark)

2. In the second part, inferential methods based on discrete time non stationary HMM are developed with the aim of analyzing and extracting the maximum information from the experimental design. Theoretical results on the asymptotic efficiency in discrete parameters models are also derived.

The experimental setting designed in the first part of the thesis can be thought of as a modification of the model presented by Nobel Laureate Serge Haroche (Guerlin et al. (2007)), where the gradual collapse of the wave function induced by repeated measurements has been observed experimentally for the first time.

The model can be briefly sketched as follows. A two-level atom (system A) is driven by an external and non-homogeneous force causing a change in its oscillation frequency. Two-level atoms (system B, also called ancilla systems), are sent through the system A and let weakly interact with it. The objective is to reconstruct the system A evolution through the measurement of the systems B, which reveal only partial information about the system A. Basically, the external force leads to a time-dependent evolution of the transition rate among the states of system A.

The experimental setting just presented can be nested in the framework of Hidden Semi-Markov Models (HSMM), since the measurement results on the B are governed by the current state of the (hidden) system A, which, in turn, is driven by a semi-Markov evolution process.

The second part of the thesis deals with the modeling and the estimation of the non-stationary evolution of the quantum system through an extension of a particular class of HSMM, the explicit-duration HMM.

We present a more general class of models in which the state of the hidden system is driven by an external force modeled as a non-stationary Markov chain. This implies that the transition probabilities from one state to the others are time-dependent. Moreover, due to the propagation of the wave functions, the so-called emission probabilities are time-dependent as well.

We introduce a non-parametric kernel estimator which allows one to considerably improve the accuracy of the estimates and then to obtain a more reliable information about the unobservable process in the presence of sparsity (especially for short observation sequences).

From a different perspective, one could be interested in estimating the value of the oscillation frequency of the system. As a matter of fact, since the external force acting on the system is represented by a finite state semi-Markov chain, each state of the chain can be thought of as a different value for the oscillation frequency of the atom. Basically, the physical theory provides us with additional information which leads to a restriction of the parameter space. We derive an extension of the quantum Cramér-Rao bound for discrete parameter models which provides the lower bound on the variance of an estimator and determines a discrete counterpart of the quantum Fisher information.

1.1 Summary and main contributions of the thesis

The thesis is organized in three papers.

Paper A is concerned with the problem of extracting the maximum information from open quantum systems. This finds applications in quantum computing and high-precision measurements. Specifically, we present an extension of the explicit-duration HMM formulation in which the hidden process is a non-stationary Markov chain with time-dependent emission probabilities and non-parametric state duration distributions.

Despite being a very flexible solution, the non-parametric specification of the duration, based on the sample relative frequencies, may require very long series of observations. To overcome this problem, we introduce a non-parametric kernel estimator for discrete distributions in the iterative step of the (estimation) Forward-Backward algorithm. Smoothing methods for discrete data, indeed, are proven to be optimal in the mean summed squared error (MSSE) sense in the presence of sparsity (Hall and Titterington (1987)), as it is the case in the experimental setup considered in this paper.

Furthermore, in order to obtain the optimal state sequence of the hidden chain associated with the given observations, we develop a generalization of the Viterbi algorithm for the explicit-duration HMM which is also robust against the so-called underflow problem.

The finite sample properties of our formulation are assessed through an extensive Monte Carlo study. All the simulations were carried out using an

optimized C++ code that takes advantage of parallel computing. The results of the simulation show that our formulation outperform the standard one both in the ability to reconstruct the hidden chain and in the accuracy of the estimates. It is remarkable that the introduction of the kernel estimator provides an improvement in the accuracy of the estimates for all the parameters, even for long observation sequences.

We also found that the performance of the model strongly depends on the type of measurement carried out. Since in our experiment the quantum measurement has only two outputs, the model is able to obtain a good fit up to five chain states, beyond which the accuracy of the estimates decreases. It is worth to remark that a chain with five states is far beyond the number of states in commonly applied models, which usually consider only one state.

One of the main limitations of HSMM consists in the prior specification of the number of states for the hidden chain, which has to be known. In order to overcome this problem, we adopt a Bayesian perspective and employ a model selection procedure which allows one to compare models with different number of chain states.

In Paper B, we develop a Markov Chain Monte Carlo (MCMC) sampling scheme for estimating the explicit-duration hidden Markov models and we make use of a Bayesian model selection procedure (Congdon (2006)) which allows one to avoid a prior specification of the number of states. Specifically, the procedure uses a Monte Carlo approximation based on independent MCMC sampling to produce posterior model probabilities and compares different models using Bayes factor estimates.

We provide both parametric and non-parametric specification for the state duration distributions. In the non-parametric specification, motivated by the presence of sparse data (lack of information), we make use of the kernel estimator for discrete distribution introduced in Paper A. This choice provides a significant improvement in the precision of the estimates for all the parameters, in particular for the transition probabilities.

As an application, we discuss an explicit-duration HMM characterization of a two-level open quantum system subject to an external and non-homogeneous perturbation causing a change in the oscillation frequency.

By means of an extensive Monte Carlo study, we show that our formulation is straightforward to implement, robust against the underflow problem and provides an accurate reconstruction of the hidden evolution with precise estimates of the parameters. The comparison of the two procedures, presented in these two papers, shows that for short observation sequences the

EM-based method offers better performance, while for medium and long sequences the Bayesian setup provides estimates less variable and more precise.

The third paper contains some theoretical results in the field of quantum Fisher information. Specifically, in Paper C we develop an extension of the Cramér-Rao inequality for quantum discrete parameter models. Basically, in our setup the parameter space is an enumerable set of values, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$, representing different oscillation frequencies for the system. We prove that the extension presented sets the ultimate accuracy of an estimator, and determines a discrete counterpart of the quantum Fisher information. This is particularly useful in many experiments in which the parameters can assume only few different values: for example, the direction which the magnetic field points to. We also provide an illustration related to a quantum optics problem.

Main contributions of the thesis

In summary, the main contributions of the thesis, besides the specification of the experimental setup, are:

- the generalization of the Viterbi algorithm for an extension of explicit-duration HMM
- an efficient and optimal (MMSE) non-parametric kernel estimator for discrete distributions in the case of sparse data
- the development of a new MCMC-based method for estimating the explicit-duration HMM
- A Bayesian model selection procedure based on posterior probabilities
- the extension of Cramér-Rao inequality for quantum discrete parameter models

1.2 Basic concepts of Quantum Mechanics

This section introduces the basic concepts of quantum statistics that will be used in the thesis. In particular, we shall consider only finite dimensional quantum systems. For more detail on quantum statistical inference and quantum probability, the reader is referred to Helstrom (1976), Holevo (1982) and Barndorff-Nielsen et al. (2003), while Wang (2012) and Gill (2001) provide a statistics-oriented introduction to quantum computation and quantum information.

1.2.1 States and evolution

Let us consider the Hilbert space of k-dimensional complex vectors, $\mathcal{H} = \mathbb{C}^k$, endowed with the inner product $\langle \cdot, \cdot \rangle$. The state of a finite dimensional quantum system can be characterized by an Hermitian, nonnegative and trace-one matrix ρ, called density matrix. If ρ has rank 1, the state is called a pure state and it can be alternatively represented as a unit vector in $\mathcal{H}$. Using the Dirac's bra-ket notation, a pure state is denoted by $|\psi\rangle$ or $\rho = |\psi\rangle \langle\psi|$, where $|\cdot\rangle$ and $\langle\cdot|$ represent column and row vectors, respectively.

One of the purely quantum phenomena is the quantum superposition. The latter can be explained referring to the property of linearity of the Schröedinger equation. Basically, like waves in classical physics, any (complex) linear combination of states results in another well-defined quantum state but, when the a measurement is performed, only one of the "composing" states can be observed, randomly. Probably the most known thought experiment linked to superposition is the so-called Schröedinger's cat, in which a cat may be simultaneously both alive and dead, as a result of a random subatomic event that may or may not occur.

A quantum system is completely described by its state and the evolution. The latter follows two principles: a deterministic, unitary evolution given by the Schröedinger equation, and a stochastic, discontinuous evolution represented by the measurement of an observable.

1.2.2 Parametric quantum model

We assume that the state of the system depends on an unknown parameter θ which is the object of our inferential problem. Then, a parametric quantum model $\{\rho(\theta), \mathbf{M};\ \theta \in \Theta \subseteq \mathbf{R}^k\}$ is defined by i) the state of the quan-

tum system, $\rho(\theta)$, which depends on θ and ii) the measurement $\mathbf{M}$ to be performed on the system.

A measurement, $\mathbf{M}$, is mathematically described by a collection of Hermitian non-negative matrices $m(x)$ indexed by the possible outcomes x of the measure. They also must add up to the identity matrix $\mathbf{I}$.

Let us consider pure states, that is $|\psi(\theta)\rangle$. The probability distribution of a random variable X, describing the outcome of the measurement $\mathbf{M}$ on the system in state $|\psi(\theta)\rangle$, is given by the trace rule,

$$P_X(\cdot, \theta) = tr\{\mathbf{M}(\cdot)\,|\psi(\theta)\rangle\,\langle\psi(\theta)|\}.$$

Measuring the system will alter its state, causing the so-called wave function collapse. The latter implies that when a particular measurement is performed, the future results of the same measurement on the same system, are fixed, preventing to obtain further information from the same system. Specifically, the state of the system after applying the measurement $\mathbf{M}$ and conditional on observing the outcome x_1 is given by

$$\frac{m(x_1)\,|\psi(\theta)\rangle}{tr\{m(x_1)\,|\psi(\theta)\rangle\,\langle\psi(\theta)|\}}$$

This procedure is known as the von Neumann projection postulate and describes the effect of a measurement on a system. In our work, we exploit the advantages of the indirect measurement which allows repeated measures on the same quantum system.

1.2.3 Generalized measurements

The generalized measurement in an open quantum framework consists of two steps: the system is coupled to a meter-system (the probe) and then a measurement on the probe is performed, avoiding a direct perturbation of the system. More formally, let us consider an auxiliary meter-system in a state $|0\rangle$, with orthonormal basis $\{|x\rangle \,|\, x \in X\}$, where X is the set of possible measurement results, and $|\psi(\theta)\rangle$, the state of the system studied.

The initial state of the two systems is $|\psi(\theta)\rangle \otimes |0\rangle$. In order to measure some properties of $|\psi(\theta)\rangle$ the two systems must interact and the state of the "composite" system is

$$U\,|\psi(\theta)\rangle \otimes |0\rangle$$

where $U = \sum_{x,x' \in X} U_{x,x'} \otimes |x\rangle\,\langle x'|$ is the unitary operator induced by the interaction. After the interaction, the two systems became entangled, that is,

"correlated". Entanglement is a quantum phenomenon which creates a correlation between two systems and allows to obtain information about $|\psi(\theta)\rangle$ measuring only the auxiliary system.

In order to avoid a direct perturbation of the system, a measurement operator that acts only on the auxiliary system is applied to the composite system. The new state of the system will change accordingly to the measurement result obtained. For example, let us assume that the measurement on the auxiliary system is resulted to be x_1, leaving it in state $|x_1\rangle$. This means that the measurement operator $\mathbf{1} \otimes |x_1\rangle \langle x_1|$ is applied to the composite system,

$$(\mathbf{1} \otimes |x_1\rangle \langle x_1|) U |\psi(\theta)\rangle \otimes |0\rangle = U_{x_1,0} |\psi(\theta)\rangle \otimes |x_1\rangle$$

leaving the system studied in the unnormalized state $U_{x_1,0} |\psi(\theta)\rangle$. The probability to obtain the result x_1 is $P(x_1) = tr\{U_{x_1,0} |\psi(\theta)\rangle \langle\psi(\theta)| U^*_{x_1,0}\}$. The two systems are no longer entangled and the interaction with a new auxiliary system can be implemented in order to acquire more information on the system evolution.

References

Attal, S., A. Joye, et al., 2006: *Open Quantum Systems I: The Hamiltonian Approach*, Springer.

Barndorff-Nielsen, O. E., R. D. Gill, and P. E. Jupp, 2003: On quantum statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 775–804.

Breuer, H.-P. and F. Petruccione, 2002: *The theory of open quantum systems*, Oxford university press.

Carmichael, H., 2009: *An open systems approach to quantum optics: lectures presented at the Université Libre de Bruxelles, October 28 to November 4, 1991*, vol. 18, Springer Science & Business Media.

Congdon, P., 2006: Bayesian model choice based on monte carlo estimates of posterior model probabilities. *Computational statistics & data analysis*, **50**, 346–357.

Davies, E. B., 1976: Quantum theory of open systems.

Gill, R. D., 2001: *Teleportation into quantum statistics*, EURANDOM.

Guerlin, C., J. Bernu, S. Deleglise, C. Sayrin, S. Gleyzes, S. Kuhr, M. Brune, J.-M. Raimond, and S. Haroche, 2007: Progressive field-state collapse and quantum non-demolition photon counting. *Nature*, **448**, 889–893.

Hall, P. and D. Titterington, 1987: On smoothing sparse multinomial data. *Australian Journal of Statistics*, **29**, 19–37.

Helstrom, C. W., 1976: *Quantum detection and estimation theory, Mathematics in Science and Engineering*, Academic press.

Holevo, A., 1982: *Probabilistic and Statistical Aspects of Quantum Mechanics*, North-Holland, Amsterdam.

Kraus, K., A. Böhm, J. D. Dollard, and W. Wootters, 1983: States, effects, and operations fundamental notions of quantum theory, *States, Effects, and Operations Fundamental Notions of Quantum Theory*, vol. 190.

Nielsen, M. A. and I. L. Chuang, 2010: *Quantum computation and quantum information*, Cambridge university press.

Von Neumann, J., 1955: *Mathematical foundations of quantum mechanics*, no. 2, Princeton university press.

Wang, Y., 2012: Quantum computation and quantum information. *Statistical Science*, **27**, 373–394.

Wiseman, H. M. and G. J. Milburn, 2009: *Quantum measurement and control*, Cambridge University Press.

Explicit-Duration Hidden Markov Models for Quantum State Estimation

Contents

Abstract

This paper is concerned with the problem of extracting the maximum information from an open quantum system subject to an external and non-stationary perturbation. An extension of the explicit-duration HMM formulation in which the hidden process is a non-stationary Markov chain with time-dependent emission probabilities and nonparametric state duration distributions. A non-parametric kernel estimator for discrete data is introduced in the estimation procedure, in order to improve the estimates accuracy in presence of sparse data. A generalization of the Viterbi algorithm for the explicit-duration HMM which is robust against the underflow problem is developed. We investigate the finite sample properties of our formulation in a Monte Carlo study and show that our model outperform the standard one even for long observation sequences. We also show that our formulation can reliably reconstruct the hidden process with few information.

1 Introduction

In recent years, the experimental advances in the field of quantum mechanics have allowed physicists to perform repeated measurements on the same quantum system, namely trapped atoms and molecules (Kirchmair et al. (2009); Kubanek et al. (2009)), optical cavities (Gleyzes et al. (2007); Goggin et al. (2011)) and superconducting systems (Palacios-Laloy et al. (2010); Vijay et al. (2011); Hatridge et al. (2013)). The novelty lies in the development of the indirect (generalized) measurement which avoids the so-called wave-function collapse by extending the measurement process to an auxiliary meter-system and then performing the measure only on the latter (Wiseman and Milburn (2009), Nielsen and Chuang (2010)). Such procedure involves (at least) two systems which are called open quantum systems since they can interact and exchange information.

Open quantum systems are the key tool in developing the new quantum technologies, for example quantum computers, quantum sensing and quantum secure communication, since any real quantum system behaves as an open system due to the extreme difficulty to isolate it from its environment.

This paper is motivated by a quantum physical problem, that is, extracting the maximum information from an open quantum system subject to an external and non-stationary perturbation causing a change in the oscillation frequency. The relevance of the quantum experiment is connected with the possibility to track and control quantum systems which undergo complex non-homogeneous evolution, with possible application to quantum feedback control, high-precision measurement and quantum computing (Ramakrishna and Rabitz (1996); Wiseman and Milburn (2009); Dong and Petersen (2010)).

The theoretical background which lies behind the experimental setting considered in the paper can be envisaged in the class of explicit-duration hidden Markov models (EDHMM). As a matter of fact, the conventional HMM, based on the Markovian short-memory assumption, cannot capture the non-stationary dynamics of the system.

Here, we present an extension of the explicit-duration HMM formulation in which the hidden process is a non-stationary Markov chain with time-dependent emission probabilities and nonparametric state duration distributions. The standard estimation procedure in the non parametric setup is based on the cell proportion estimator, i.e. the sample relative frequency. This choice, although improving flexibility, inevitably increases the dimension of the parameter set making difficult to obtain reliable information unless

very long observation sequences are considered. This high dimensional setting results in sparse multinomial table where the cell proportion (maximum likelihood) estimator is neither consistent nor a "good" estimator of the true discrete duration distribution. A smoothed estimator is thus derived, based on a discrete kernel function, which highly improves the estimation and is shown to be consistent and computationally efficient, according to the theory of sparse asymptotics (Fienberg and Holland (1973); Bishop et al. (1975)), for multinomial data (Aitchison and Aitken (1976); Titterington (1980); Wang and Van Ryzin (1981); Simonoff (1983), Hall and Titterington (1987)).

A generalization of the Viterbi algorithm, used to reconstruct the hidden dynamics is developed. The latter turns out to be robust against the underflow problem.

There are three main contributions in this article. First, to the best of our knowledge, this is the first attempt to provide an accurate statistical description of an open quantum system subject to a complex dynamics using the explicit-duration HMM formulation. Second, we take into account the presence of sparse data (lack of information) by introducing a non-parametric kernel estimator for discrete distribution in the estimation procedure. Third, we propose an extension of the Viterbi algorithm for the explicit-duration HMM which is also robust against the so-called underflow problem.

The remainder of the paper is organized as follows. Section 1.1 reviews the basics of quantum mechanics and Section 2 presents the experimental setup. The explicit duration HMM are discussed in section 3. Sections 4 address the estimation issues, including the kernel estimator and the generalized Viterbi algorithm. Section 5 presents the results the Monte Carlo study and concludes the paper in Section 6.

1.1 Basic concepts of Quantum Mechanics

This section introduces the basic concepts of quantum statistics that will be used in the paper. In particular, we shall consider only finite dimensional quantum systems. For more detail on quantum statistical inference and quantum probability, the reader is referred to Helstrom (1976), Holevo (1982) and Barndorff-Nielsen et al. (2003), while Wang (2012) and Gill (2001) provide a statistics-oriented introduction to quantum computation and quantum information.

1.1.1 States and evolution

Let us consider the Hilbert space of k-dimensional complex vectors, $\mathcal{H} = \mathbb{C}^k$, endowed with the inner product $\langle \cdot, \cdot \rangle$. The state of a finite dimensional quantum system can be characterized by an Hermitian, nonnegative and trace-one matrix ρ, called density matrix. If ρ has rank 1, the state is called a pure state and it can be alternatively represented as a unit vector in $\mathcal{H}$. Using the Dirac's bra-ket notation, a pure state is denoted by $|\psi\rangle$ or $\rho = |\psi\rangle \langle\psi|$, where $|\cdot\rangle$ and $\langle\cdot|$ represent column and row vectors, respectively.

One of the purely quantum phenomena is the quantum superposition. The latter can be explained referring to the property of linearity of the Schröedinger equation. Basically, like waves in classical physics, any (complex) linear combination of states results in another well-defined quantum state but, when the a measurement is performed, only one of the "composing" states can be observed, randomly. Probably the most known thought experiment linked to superposition is the so-called Schröedinger's cat, in which a cat may be simultaneously both alive and dead, as a result of a random subatomic event that may or may not occur.

A quantum system is completely described by its state and the evolution. The latter follows two principles: a deterministic, unitary evolution given by the Schröedinger equation, and a stochastic, discontinuous evolution represented by the measurement of an observable.

1.1.2 Parametric quantum model

We assume that the state of the system depends on an unknown parameter θ which is the object of our inferential problem. Then, a parametric quantum model $\{\rho(\theta), \mathbf{M};\ \theta \in \Theta \subseteq \mathbf{R}^k\}$ is defined by i) the state of the quantum system, $\rho(\theta)$, which depends on θ and ii) the measurement $\mathbf{M}$ to be performed on the system.

A measurement, $\mathbf{M}$, is mathematically described by a collection of Hermitian non-negative matrices $m(x)$ indexed by the possible outcomes x of the measure. They also must add up to the identity matrix $\mathbf{I}$.

Let us consider pure states, that is $|\psi(\theta)\rangle$. The probability distribution of a random variable X, describing the outcome of the measurement $\mathbf{M}$ on the system in state $|\psi(\theta)\rangle$, is given by the trace rule,

$$P_X(\cdot, \theta) = tr\{\mathbf{M}(\cdot)\, |\psi(\theta)\rangle \langle\psi(\theta)|\}.$$

Measuring the system will alter its state, causing the so-called wave function collapse. The latter implies that when a particular measurement is performed, the future results of the same measurement on the same system, are fixed, preventing to obtain further information from the same system. Specifically, the state of the system after applying the measurement $\mathbf{M}$ and conditional on observing the outcome x_1 is given by

$$\frac{m(x_1)\,|\psi(\theta)\rangle}{tr\{m(x_1)\,|\psi(\theta)\rangle\,\langle\psi(\theta)|\}}$$

This procedure is known as the von Neumann projection postulate and describes the effect of a measurement on a system. In our work, we exploit the advantages of the indirect measurement which allows repeated measures on the same quantum system.

1.1.3 Generalized measurements

The generalized measurement in an open quantum framework consists of two steps: the system is coupled to a meter-system (the probe) and then a measurement on the probe is performed, avoiding a direct perturbation of the system. More formally, let us consider an auxiliary meter-system in a state $|0\rangle$, with orthonormal basis $\{|x\rangle \mid x \in X\}$, where X is the set of possible measurement results, and $|\psi(\theta)\rangle$, the state of the system studied.

The initial state of the two systems is $|\psi(\theta)\rangle \otimes |0\rangle$. In order to measure some properties of $|\psi(\theta)\rangle$ the two systems must interact and the state of the "composite" system is

$$U\,|\psi(\theta)\rangle \otimes |0\rangle$$

where $U = \sum_{x,x' \in X} U_{x,x'} \otimes |x\rangle\,\langle x'|$ is the unitary operator induced by the interaction. After the interaction, the two systems became entangled, that is, "correlated". Entanglement is a quantum phenomenon which creates a correlation between two systems and allows to obtain information about $|\psi(\theta)\rangle$ measuring only the auxiliary system.

In order to avoid a direct perturbation of the system, a measurement operator that acts only on the auxiliary system is applied to the composite system. The new state of the system will change accordingly to the measurement result obtained. For example, let us assume that the measurement on the auxiliary system is resulted to be x_1, leaving it in state $|x_1\rangle$. This means

that the measurement operator $\mathbf{1} \otimes |x_1\rangle \langle x_1|$ is applied to the composite system,

$$(\mathbf{1} \otimes |x_1\rangle \langle x_1|) U |\psi(\theta)\rangle \otimes |0\rangle = U_{x_1,0} |\psi(\theta)\rangle \otimes |x_1\rangle$$

leaving the system studied in the unnormalized state $U_{x_1,0} |\psi(\theta)\rangle$. The probability to obtain the result x_1 is $P(x_1) = tr\{U_{x_1,0} |\psi(\theta)\rangle \langle\psi(\theta)| U^*_{x_1,0}\}$. The two systems are no longer entangled and the interaction with a new auxiliary system can be implemented in order to acquire more information on the system evolution. In the next section the experimental setup developed is presented.

2 Quantum experiment

We design an experimental setup that can be thought of as a modification of the one presented by Nobel Laureate Serge Haroche (Guerlin et al. (2007)), in which the gradual step-by-step 'state collapse' caused by repeated measurements is experimentally observed for the first time.

Here, a two-level atom (system A), initialized in a superposition state, is driven by an external and non-homogeneous force causing a change in the Rabi frequency. The latter represents the characteristic angular frequency of the atom corresponding to the oscillation between the two levels, caused by the time evolution. Two-level atoms (system B, also called ancilla, meter or probe systems), initialized in a superposition state, are sent through the system A and weakly interact with it. Then, a projection measurement on each B system is performed, thus avoiding a direct perturbation of the system A. The projection measurement of the ancilla systems causes a non-projective 'weak' disturbance of the system A, influencing its quantum state.

The experimental set up introduced so far can be nested in the framework of hidden semi-Markov models (HSMM), since the measurement results on the B are governed by the current (hidden) state of the system A, which, in turn, is driven by a semi-Markov evolution process. Specifically, the transition probabilities form one state to another depend on the state duration, and thus they are time varying. The emission probabilities are time-dependent as well, due to the propagation of the wave function at different frequency values. Basically, the external force, characterized as a semi-Markov chain, affects the evolution of oscillation frequency of the system, making it time dependent.

Our purpose is to keep track of the evolution of system A and to estimate the state of the chain (the value of the Rabi frequency) at each time-step through measurements on the ancillas B, which reveal only partial information about the system A.

The model can be sketched as follows. Let us consider a semi-Markov chain with n states $\mathcal{S} = \{\omega_1, \ldots, \omega_n\}$ representing the Rabi frequencies of the system and two output symbols '0' and '1', representing the measurement results. The quantum system A under study is initialized in a (known) superposition state

$$|\Psi_A(\omega_i)\rangle_{t_0} = a_{t_0}(\omega_i)\,|0_A\rangle + b_{t_0}(\omega_i)\,|1_A\rangle$$

where we stress the dependency of the A state on the value of $\omega_i \in \mathcal{S}$. Furthermore, a, b are complex numbers satisfying $|a(\omega_i)|^2 + |b(\omega_i)|^2 = 1$, $|\Psi\rangle \in \mathcal{H}_2$ represents a unit vector in the two dimensional complex Hilbert space and $\{|0_A\rangle, |1_A\rangle\}$ denotes an orthonormal basis for the two dimensional vector space. In the rest of the paper, we will suppress the dependence of α and β on ω_i and we will maintain it only in the quantum state(in order to avoid a too cumbersome notation). The ancilla system is initially in the state $|\Psi_B\rangle \in \mathcal{H}_2$, with orthonormal basis $\{|v_B\rangle; v \in \mathcal{V}\}$ where $\mathcal{V} = \{0, 1\}$ is the set of all possible measurement outcome. The initial state of the composite system is $|\Psi_A(\omega_i)_{t_0}\rangle \otimes |\Psi_B\rangle$ and belongs to $\mathcal{H} = \mathcal{H}_2 \otimes \mathcal{H}_2$. When the interaction between the systems takes place, the system A and the meter become correlated, and the subsequent entangled state reads

$$\begin{aligned} |\Psi_{AB}(\omega_i)\rangle_{t_1} &= \alpha_{t_1}\,|0_A\rangle \otimes \left|\Psi_B^0\right\rangle + \beta_{t_1}\,|1_A\rangle \otimes \left|\Psi_B^1\right\rangle \\ &= \alpha_{t_1}\,|0_A\rangle \otimes \left(\sqrt{p}\,|0_B\rangle + \sqrt{q}\,|1_B\rangle\right) + \beta_{t_1}\,|1_A\rangle \otimes \left(\sqrt{q}\,|0_B\rangle + \sqrt{p}\,|1_B\rangle\right) \end{aligned}$$

where $p+q = 1$, representing the ability to distinguish between the two states. Then, the ancilla system is measured and the result '0' is obtained with probability

$$P_{t_1}(B = 0\,|\,|\Psi_{AB}(\omega_i)\rangle_{t_1}) = |\alpha_{t_1}\sqrt{p}|^2 + |\beta_{t_1}\sqrt{q}|^2 \tag{1}$$

This quantity represents the probability to obtain the output value '0' conditional on the system dynamics up to time t_1 and then on the ω_i value which governs the evolution of the system A. In the HMM framework, such quantity can be viewed as the emission probability, at time t_1, of symbol '0' when the system is in state ω_i, that is $P_{t_1}(B = 0\,|\,\omega_i)$. Furthermore, it

is important to note that this probability is time-varying, since, depending on the ω_i driving the system dynamics at time t_1, the probability to observe each output can change. The measurement resulting in the output '0' leaves the system A in the (unnormalized) state

$$|\Psi_A(\omega_i \,|\, B = 0)\rangle = \alpha_{t_1}\sqrt{p}\,|0_A\rangle + \beta_{t_1}\sqrt{q}\,|1_A\rangle\,,$$

where we explicitly indicate that the state is conditional on the previous result on the B system. Instead, the measurement result '1' is obtained with probability

$$P_{t_1}(B = 1 \,|\, |\Psi_{AB}(\omega_i)\rangle_{t_1}) = |\alpha_{t_1}\sqrt{q}|^2 + |\beta_{t_1}\sqrt{p}|^2 \tag{2}$$

and the (unnormalized) A state is

$$|\Psi_A(\omega_i \,|\, B = 1)\rangle = \alpha_{t_1}\sqrt{q}\,|0_A\rangle + \beta_{t_1}\sqrt{p}\,|1_A\rangle\,.$$

From equations (1) and (2) it can be seen that the quantities p and q represent the ability to distinguish between the states and then they can be thought of as a measurement effectiveness. In the following section, we shall introduce the general HSMM, and the Explicit-duration HMM.

3 EDHMM with time varying emission probabilities

HSMM are extensions of the HMM in which each state has a different duration time (*sojourn*) modeled by a random variable that takes integer values in the set $\mathcal{D} = \{1, 2, \ldots, D\}$. Based on hypotheses on the dependence structure of the chain and the duration of each state, the general HSMM reduces to specific models, such as, for instance, explicit-duration HMM, where the duration is assumed to be dependent on the current state and independent to the previous state (Rabiner (1989), Mitchell and Jamieson (1993)) and variable-transition HMM, where the state transition is dependent on the state duration (Ramesh and Wilpon (1992a), Krishnamurthy et al. (1991), for a complete review of HSMM see Yu (2010))

The earliest formulation of HSMM dates back to Ferguson (1980), who was the first to formalize the HSMM, which he called "variable duration HMM". Since then, HSMM have been largely studied due to their applications in a wide variety of fields. In particular, Barbu and Limnios (2006)

and Barbu and Limnios (2009) proved consistency and asymptotic normality for nonparametric maximum likelihood estimators, Johnson and Willsky (2013) introduced the explicit-duration Hierarchical Dirichlet Process HSMM in a nonparametric Bayesian setting and Squire and Levinson (2005) propose a recursive maximum-likelihood algorithm for online estimation. Recently, Melnyk and Banerjee (2014) introduced a spectral algorithm for inference in HSMM and Bietti et al. (2015) presented an incremental EM algorithm for online parameters estimation. Besides, HSMM have been successfully applied in many areas among which, financial time series modeling (Bulla and Bulla (2006)), recognition of human genes in DNA (Haussler and Eeckman (1996)), handwritten word recognition (Kundu et al. (1997)), protein structure prediction (Schmidler et al. (2000)) and for determining duration and timing of up-down state in neocortical neurons (McFarland et al. (2011)).

The improvement of HSMM with respect to conventional HMM lies in the explicit specification of the duration of each state, in contrast to the implicit duration (geometric or exponential) in the HMM. Hence, a HSMM is a discrete-time, finite-state non homogeneous Markov chain observed through a finite set of transition densities indexed by the states of the semi-Markov chain.

This paper is concerned with the class of explicit-duration hidden Markov models (EDHMM). More formally, an EDHMM is characterized by the following quantities:

- $\{X_t\}_{t\in\mathbb{N}}$ the non stationary Markov chain on $\mathcal{S} = \{1, 2, \ldots, M\}$ the set of M states, where $X_{1:T} = \{X_1, \ldots, X_T\}$ denotes the state sequence up to time T. Here, $\mathcal{S}$ corresponds to the set of the Rabi frequencies for the system, that is, $\mathcal{S} = \{\omega_1, \ldots, \omega_5\}$;

- $\mathcal{A} = \{a_{i(j,d)}\}_{i,j=1,\ldots,M,\ d=1,\ldots,D}$ the set of transition probabilities from state i to state j with duration d,

$$a_{i(j,d_2)} = a_{ij} p_j(d_2), \tag{3}$$

where

$$a_{ij} = P(X_{t+1} = j \,|\, X_t = i), \qquad 1 \le i, j \le M; \tag{4}$$

the transition probability from state i to state j, with null self-transition probability, $a_{ij} = 0$, and

$$p_j(d) = P(X_{t+1:t+d]} = j \,|\, X_{[t+1} = j), \quad 1 \le j \le M,\ 1 \le d \le D$$

the probability that state j has a duration d, starting from time $t+1$ until $t+d$;

- $\{Y_t\}_{t\in\mathbb{N}}$ the sequence of conditionally independent random variables on $\mathcal{V} = \{v_1, \ldots, v_K\}$, representing the observed signal. In particular, Y_t has a conditional distribution which depends on the chain only through its current value X_t, and $\mathcal{V}$ is the set of K measurement outputs per state.Here, for each state there are two output symbols '0' and '1', representing the results of the quantum measurement;

- $\mathcal{B} = \{b_j(Y_t = v_k)\}_{j=1,\ldots,M,\ k=1,\ldots,K}$ the set of emission probabilities,

$$b_j(Y_t = v_k) = P(Y_t = v_k \,|\, X_t = j), \qquad 1 \le j \le M,\ 1 \le k \le K \tag{5}$$

  representing the probability to observe the value $Y_t = v_k$ when $X_t = j$. In our experiment the emission probabilities (1) and (2) are time-dependent and are obtained by propagating the wave-function;

- $\Pi = \{\pi_i\}_{i=1,\ldots,M}$ the initial state distribution, where

$$\pi_i = P(X_1 = i), \qquad 1 \le i \le M.$$

The set of model parameters is defined by

$$\lambda = (\mathcal{A}, \mathcal{B}, \Pi).$$

In the following section, we shall address the estimation issues, starting from the Forward-Backward algorithm to the re-estimation of the model parameters.

4 Estimation

State and parameter estimation of HMM is based on the well-known Baum-Welch or Forward-Backward (FB) algorithm, developed in the seminal work of Baum, Petrie, Soules and Weiss (Baum et al. (1970)) and primarily based on the so called forward and backward variables, α and β, respectively.

One of the most severe issues in practical implementations of the FB algorithm is the numerical underflow, caused by the exponential decay of the joint probability of the observations, as the sample size increases. For

the conventional HMM, this problem can be circumvented by rescaling the Forward-Backward variables, α and β. However, there exists no such procedure for HSMM. As an alternative, we make use of a modified version of the FB algorithm, due to Yu and Kobayashi (2006), where new forward and backward variables are defined conditionally on the observations. The resulting algorithm is computationally efficient and does not suffer of the underflow problem.

4.1 Modified Forward-Backward Recursion

In this section, the modified FB algorithm used in this paper is presented. The latter defines an estimation procedure for HSMM which is robust against the underflow problem, and does not increase the computational complexity of the standard procedure.

Let $\alpha_{t|k}(i,d)$ be the *forward variable* defined by

$$\alpha_{t|k}(i,d) = P(X_t = i, \tau_t = d \,|\, Y_{1:k}, \lambda)$$

where $Y_{1:k}$ denotes the observation sequence from time 1 to k, and $k = t-1, t$ or T, indicates the "predicted", "filtered" and "smoothed" probability respectively. Furthermore, τ_t denotes the remaining *sojourn* time of the current state X_t. Let us also define the modified emission probabilities $b_i^\star(Y_t)$, as the ratio of the filtered probability $\alpha_{t|t}(i,d)$ over the predicted one $\alpha_{t|t-1}(i,d)$,

$$b_i^\star(Y_t) = \frac{\alpha_{t|t}(i,d)}{\alpha_{t|t-1}(i,d)} = \frac{b_i(Y_t = v_k)}{P(Y_t \,|\, Y_{1:t-1})} \tag{6}$$

where $P(Y_t \,|\, Y_{1:t-1})$ is the one-step ahead prediction of the observation. This can be obtained by

$$P(Y_t \,|\, Y_{1:t-1}) = \sum_{i=1}^{M} \sum_{d=1}^{D} \alpha_{t|t-1}(i,d) b_i(Y_t) = \sum_{i=1}^{M} \gamma_{t|t-1}(i) b_i(Y_t)$$

where $\gamma_{t|k}(i) = \sum_d \alpha_{t|k}(i,d)$ is the marginal probability distribution of X_t which also represents the predicted, filtered or smoothed conditional probability of state X_t given the observed sequence $Y_{1:k}$. The likelihood function for the entire sequence of observations is obtained using the the one-step ahead predictions $P(Y_t \,|\, Y_{1:t-1})$ as follows:

$$\mathcal{L}(\lambda) = P(y_{1:T}|\lambda) = \prod_{t=1}^{T} P(Y_t \,|\, Y_{1:t-1}).$$

For convenience in the forward recursion, we introduce two additional variables: the conditional probability of state i ending at time t given the whole observation sequence $Y_{1:t}$,

$$\mathcal{E}_t(i) = P(X_t = i, \tau_t = 1 \,|\, Y_{1:t}, \lambda) = \alpha_{t|t-1}(i,1) b_i^\star(Y_t,)$$

and the conditional probability of state i starting at time $t+1$ given $Y_{1:t}$,

$$\mathcal{S}_t(i) = P(\tau_t = 1, X_{t+1} = i \,|\, Y_{1:t}, \lambda) = \sum_j \mathcal{E}_t(j) a_{ji}.$$

To obtain the smoothed estimates, let $\beta_t(i,d)$ be the *backward variable* defined as the ratio between the smoothed probability $\alpha_{t|T}(i,d)$ and the predicted one $\alpha_{t|t-1}(i,d)$, that is

$$\beta_t(i,d) = \frac{P(X_t = i, \tau_t = d \,|\, Y_{1:T}, \lambda)}{P(X_t = i, \tau_t = d \,|\, Y_{1:t-1}, \lambda)} = \frac{P(Y_{t:T} \,|\, X_t = i, \tau_t = d \,|\, Y_{1:T}, \lambda)}{P(Y_{t:T} \,|\, Y_{1:t-1}, \lambda)}.$$

As for the forward recursion, we define two more variables that are symmetric to $\mathcal{S}_t(i)$ and $\mathcal{E}_t(j)$ and will be used in the backward recursion

$$\mathcal{E}_t^\star(i) = \frac{P(Y_{t:T} \,|\, X_t = i, \tau_{t-1} = 1 \,|\, Y_{1:T}, \lambda)}{P(Y_{t:T} \,|\, Y_{1:t-1}, \lambda)} = \sum_d p_i(d) \beta_t(i,d),$$

and

$$\mathcal{S}_t^\star(i) = \frac{P(Y_{t:T} \,|\, X_{t-1} = i, \tau_{t-1} = 1 \,|\, Y_{1:T}, \lambda)}{P(Y_{t:T} \,|\, Y_{1:t-1}, \lambda)} = \sum_j a_{ij} \mathcal{E}_t^\star(j).$$

The modified forward-backward algorithm for explicit-duration HMM can be implemented as follows:

i) the forward recursion becomes

$$\alpha_{t|t-a}(i,d) = \mathcal{S}_{t-1}(i) p_i(d) + b_i^\star(Y_{t-1}) \alpha_{t-1|t-2}(i, d+1) \tag{7}$$

with initial condition $\alpha_{1|0}(i,d) = \pi_i p_i(d)$;

ii) the backward recursion is

$$\beta_t(i,d) = \begin{cases} \mathcal{S}_{t+1}^\star(i) b_i^\star(Y_t) & d = 1 \\ \beta_{t+1}(i, d-1) b_i^\star(Y_t) & d > 1 \end{cases} \tag{8}$$

with initial condition $\beta_T(i,d) = b_i^\star(Y_T)\ \forall d \in \mathcal{D}$.

We conclude this section by defining two more variables which will enter in the re-estimation step of the algorithm: the smoothed probability that state i starts at time t and lasts for d time units:

$$\mathcal{D}_{t|T}(i,d) = P(\tau_{t-1} = 1, X_t = i, \tau_t = d \,|\, Y_{1:T}, \lambda) = \mathcal{S}_{t-1}(i) p_i(d) \beta_t(i,d) \qquad (9)$$

and the smoothed probability that a transition from state i to j occurs at time t:

$$\mathcal{T}_{t|T}(i,j) = P(X_{t-1} = i, \tau_{t-1} = 1, X_t = j \,|\, Y_{1:T}, \lambda) = \mathcal{E}_{t-1}(i) a_{ij} \mathcal{E}^{\star}_{t-1}(j). \qquad (10)$$

4.1.1 Parameter re-estimation

So far, we have assumed that the set of model parameters λ is known. When λ is unknown, an iterative procedure which maximises the probability of the given observation sequence, $P(Y_{1:T}|\lambda)$, is commonly adopted, with initial values that are either randomly selected or uniformly distributed. The model parameters are estimated and then re-estimated until the likelihood is locally maximized.

Specifically, the smoothed probabilities in equations (9) and (10) are used to re-estimate the parameters as follows:

$$\hat{a}_{ij} = \sum_{t=2}^{T} \frac{\mathcal{T}_{t|T}(i,j)}{K_a}$$

where $K_a = \sum_{j \neq i} \sum_{t=2}^{T} \mathcal{T}_{t|T}(i,j)$ is the normalizing constant such that $\sum_j \hat{a}_{ij} = 1 \; \forall i$,

$$\hat{p}_i(d) = \sum_{t=2}^{T} \frac{\mathcal{D}_{t|T}(i,d)}{K_p}$$

with $K_p = \sum_{d=1}^{D} \sum_{t=2}^{T} \mathcal{D}_{t|T}(i,d)$ and $\sum_d \hat{p}_i(d) = 1 \; \forall i$,

$$\hat{\pi}_i = \frac{\gamma_{1|T}(i)}{K_\pi}$$

where $K_\pi = \sum_i \gamma_{1|T}(i)$ and $\sum_i \hat{\pi}_i = 1$. The term $\gamma_{1|T}(i)$ can be obtained with the following backward recursion:

$$\gamma_{t-1|T}(i) = \gamma_{t|T}(i) + \mathcal{E}_{t-1}(i)\mathcal{S}^{\star}_t(i) - \mathcal{S}_{t-1}(i)\mathcal{E}^{\star}_t(i).$$

The static emission probabilities may be obtained by

$$\hat{b}_i(Y_t = v_k) = \sum_{t=1}^{T} \frac{\gamma_{t|T}(i)\delta(Y_t, v_k)}{K_b}$$

with $K_b = \sum_{v_k} \sum_{t=1}^{T} \gamma_{t|T}(i)\delta(Y_t, Y_t = v_k)$ such that $\sum_k \hat{b}_i(v_k) = 1$, and $\delta(Y_t, v_k) = 1$ if $Y_t = v_k$, and zero otherwise.

In the next section, we shall deal with the estimation of the duration distribution.

4.2 Estimation of the duration distribution

One of the novelties of this paper is the introduction of a non-parametric kernel estimator for the discrete distribution of the duration in the iterative step of the estimation algorithm. Specifically, we shall consider an estimator of the form

$$\hat{p}_i(d) = \sum_{j-d=-h}^{h} K((j-d)/h)p_i(j) \tag{11}$$

where $p_i(j) = \frac{n_{ij}}{n_i}$ is the sample relative frequency, h and d are positive integer, $h + 1 < d < D - (h - 1)$, and $K(t)$ is a discrete kernel, that is a non negative, symmetric function of t which adds up to one.

The relevance of the contribution lies in the fact that smoothing methods for discrete data are optimal in the mean summed squared error (MSSE) sense in the case of sparsity (Hall and Titterington (1987)), as it is the case of the experiment considered in this paper. In fact, the cell proportion estimator $p_i = \frac{n_i}{n}$ (the MLE for a multinomial distribution) traditionally used in the probability mass function estimation is a consistent estimator of $p_i(d)$ only when the sample size becomes large compared with the number of cells (Fienberg and Holland (1973); Simonoff (1983); Simonoff (2012)). In the case when the number of cells is close to or greater than the number of observations, which results in a sparse table with many small or zero cell counts, the cell proportion estimator p_i is inconsistent and provides estimates characterized by roughness and multimodality. Smoothing methods based on kernel estimators have been proved to be effective for sparse multinomial data (Aitchison and Aitken (1976); Titterington (1980); Wang and Van Ryzin (1981); Simonoff (1983)). Besides, Hall and Titterington (1987) proved the optimality of kernel estimators for sparse multinomial data in the framework

of sparse asymptotic earlier introduced by Fienberg and Holland (1973) and Bishop et al. (1975).

Using a kernel smoother requires that the shape and the bandwidth of the kernel function are selected. As far as the kernel function is concerned, we have chosen the discrete version the Epanechnikov kernel, which belongs to the class of the kernels generated as Beta distributions and has optimal asymptotic properties for continuous smoothing, in the sense that it minimises the asymptotic integrated mean square error (AMISE), (Wand and Jones (1994)).

In practice, we have a discrete quadratic kernel $K(t) = at^2 + b$, where t takes value in a discrete set, which is a common choice in kernel smoothing as it represents a good compromise between fitting ad smoothing. (Marron and Wand (1992); Simonoff (2012)). Actually, we use the quadratic kernel formulation of Rajagopalan and Lall (1995), with minor corrections, who derive the value of the normalising constants a and b both in the interior and in the boundaries. These value must be selected in order for the kernel to add up to one, $\sum_t K(t) = 1$, and satisfy the properties of symmetry $K(t) = K(t)$, positivity, $K(t) > 0$ within the bandwidth, and unbiasedness $\sum_t K(t)t = 1$. For the interior region, i.e. for $h + 1 < d < D - h - 1$, the parameters a and b are

$$a = \frac{3h}{(1 - 4h^2)}, \; b = -a.$$

At the boundaries, i.e. for $1 < d < h + 1$ and $d > D - h - 1$, a and b become

$$a = \frac{-D}{2h(h+d)}\left(\frac{E}{4h^3} - \frac{CD}{12h^3(h+d)}\right)^{-1}, \quad b = \left(1 - \frac{aC}{6h^2}\right)(h+d)^{-1}$$

where

$$C = h(h+1)(2h+1) + (d-2)(d-1)(2h-3) \tag{12}$$
$$D = -h(h+1) + (d-2)(d-1) \tag{13}$$
$$E = (-h(h+1))^2 + ((d-2)(d-1))^2. \tag{14}$$

If the kernel has a limited impact on the estimation, the bandwidth is a crucial parameter. In the context of sparse asymptotics, Hall and Titterington (1987) have shown the optimality of the least cross validation criterion (LSCV) in terms of rate of convergence.

Hence, the optimal h is found by minimizing the LSCV function defined as

$$LSCV(h) = \sum_{d=1}^{D} (\hat{p}_i(d))^2 - \frac{2}{D} \sum_{d=1}^{D} \hat{p}_{\setminus i}(d) d_i \tag{15}$$

where $\hat{p}_{\setminus i}(d)$ represents the estimate of $p_i(d)$ without the i-th's contribute.

In the next section we shall discuss the reconstruction of the optimal state sequence of th ehidden chain, namely the Viterbi algorithm.

4.3 State reconstruction

One of the main results of this paper consists in the extension of the Viterbi algorithm for explicit-duration HMM. The latter turns out to be robust against the underflow problem. Viterbi algorithm is a dynamic programming algorithm which finds the single best state sequence maximizing $P(X_{1:T} \,|\, Y_{1:T}, \lambda)$, taking also into account the probabilities of occurrence of states. Some modified versions of this algorithm for HSMM have been proposed (Yu (2010)). Specifically, Chen et al. (1995) propose a modified algorithm for continuous density variable-duration HMM, Ramesh and Wilpon (1992b) use the Viterbi algorithm for modeling state duration in inhomogeneous HMM, Mitchell et al. (1995) introduce a new recursion which reduces the complexity of the estimation procedure.

The algorithm can be sketched as follows. Let $\delta_t(j, d)$ be the forward variable for the extended Viterbi algorithm defined by

$$\delta_t(j, d) = \max_{X_{1:t-d}} P(X_{1:t-d}, X_{[t-d+1:t]} = j, Y_{1:t} \,|\, \lambda) \tag{16}$$

for $1 \leq t \leq T$, $j \in \mathcal{S}$, $d \in \mathcal{D}$. Here $\delta_t(j, d)$ denotes the probability of the best partial state sequence which ends at time t in state j with duration d. In the explicit-duration HMM, equation (16) reads

$$\delta_t(j, d) = \max_{i \in \mathcal{S} \setminus \{j\}, d_1 \in \mathcal{D}} \{\delta_{t-d}(i, d_1) a_{ij} p_j(d) b_j^\star(Y_{t-d+1:t})\} \tag{17}$$

where $b_j^\star(Y_{t-d+1:t}) = \prod_{\tau=t-d+1}^{t} b_j^\star(Y_\tau)$ due to the assumption of conditional independence of outputs given the state of the system. The algorithm is initialized as follows:

$$\delta_1(j, 1) = \pi_j p_j(1) b_j^\star(Y_1) \qquad \forall j \in \mathcal{S},\ d = 1$$

$$\delta_1(j, d) = 0 \qquad \forall j \in \mathcal{S},\ d > 1$$

Hence, $\delta_t(j,d)$ can be recursively obtained by

$$\delta_t(j,1) = \max_{d_1 \in \mathcal{D}, i \in \mathcal{S}\setminus\{j\}} \{\delta_{t-1}(i,d_1) a_{ij} p_j(1)\} b_j^\star(Y_t), \qquad d = 1 \tag{18}$$

$$\delta_t(j,d) = \max_{d_1 \in \mathcal{D}, i \in \mathcal{S}\setminus\{j\}} \{\delta_{t-d}(i,d_1) a_{ij} p_j(d)\} b_j^\star(Y_{t-d+1:t}), \quad d > 1. \tag{19}$$

Moreover, for $t \leq D$ and $d = t$ (19) reduces to

$$\delta_t(j,d) = \pi_j p_j(d) b_j^\star(Y_{t-d+1:t}).$$

For backtracking the optimal state sequence, we keep track of the arguments which maximize equations (18) and (19) defining two variables as follows:

$$\Psi(t,j) = \arg\max_{i \in \mathcal{S}\setminus\{j\}} \{\delta_{t-d}(i, \omega = \Delta_t(j,i)) a_{ij}\},$$

which records the state selected by $\delta_t(j,d)$ that ends at time $t-d$, and

$$\Delta(t,j,i) = \arg\max_{\omega \in \mathcal{D}} \{\delta_{t-d}(i,\omega) a_{ij}\}$$

which records its duration. The probability of the best state sequence is given by

$$P^* = \max_{i \in \mathcal{S}} \{\delta_T(i, \eta(i))\}$$

where

$$\eta(i) = \arg\max_{d \in \mathcal{D}} \{\delta_T(i,d)\}, \ \forall i \in \mathcal{S}$$

The best path is obtained by finding the last state which maximizes the likelihood, that is

$$\tilde{x}_T = \arg\max_{i \in \mathcal{S}, d \in D} \delta_T(i,d).$$

Hence, by letting $\tilde{d} = \eta(\tilde{x}_T)$, $t = T$ and $z = \tilde{d}$, the sequence can be tracked back as follows:

$$\begin{aligned} \tilde{x}_{[t-\tilde{d}+1:t]} &= \tilde{x}_t \\ \tilde{x}_{t-\tilde{d}} &= \Psi(t-\tilde{d}+1, \tilde{x}_t) \\ z &= \Delta(t-\tilde{d}+1, \tilde{x}_t, \tilde{x}_{t-\tilde{d}}) \\ t &= t-\tilde{d}, \qquad \tilde{d} = z \end{aligned}$$

until the first state $\tilde{x}_1$ is obtained. In the following, to investigate the estimation performance of our formulation, a simulation study is carried out.

5 Monte Carlo study

5.1 Design of the simulation study

To validate the finite samples properties of the ED-HMM model introduced above, a Monte Carlo study is carried out. Let us remind that the aim is to to characterize the behavior of an open quantum system subject to an external force by using only the partial information obtained by measuring a set of ancillary systems.

The observed signal $Y_{1:T}$ is a binary sequence obtained by measuring the systems B. It is characterized as a discrete-time random process driven by an underlying (hidden) semi-Markov chain represented by the evolution of system A.

The hidden chain is composed of $M = 5$ states corresponding to 5 different frequency values in the range $(0, \frac{\pi}{2})$, that is $\mathcal{S} = \{\omega_1 = 0.003,\ \omega_2 = 0.4,\ \omega_3 = 0.85,\ \omega_4 = 1.17,\ \omega_5 = 1.55\}$. The model's ability to correctly characterise the hidden states strongly depends on the type of measurement performed. Since the output is a binary time series, our simulations show that the model is able to obtain a good fit up to 5 chain states, beyond which the accuracy of the estimates decreases. It is worth to remark that the value $M = 5$ is far beyond the one found in commonly applied models, which usually consider only one frequency value.

The truncation value D, representing the maximum number of consecutive time steps in the same state, is set to 120. The choice of the D value can be a crucial aspect. On the one hand, choosing a small value of D does not allow the model to be flexible enough to capture the full dependence structure of the chain. On the other hand, a large value of D can cause a curse of dimensionality that makes the estimation process impractical. From a practical point of view, the choice of the maximum D can be done by a grid search over several values of D. We have have fixed a large value of D, that is $D = 120$, in that the simulations were carried out using an optimized code that takes advantage of parallel computing and is not time consuming.

The initial values for the transition probabilities $\mathcal{A}$ and the initial state distribution Π are assumed to be either uniformly distributed or randomly selected. The final results are found to be robust in terms of the choice of the initial condition. The emission probabilities are obtained by propagating the wave function at each frequency ω_i: this represents a point of strength in the estimation/re-estimation procedure (see Rabiner (1989)), in that they

do not enter in the estimation algorithm. The latter is based on the non parametric kernel estimator described in section 4.2

Five different sample sizes are considered: T = 300, 500, 2000, 5000, 20000. The number of Monte Carlo replications is set to 5000 for each sample size.

In the first step of the simulation, we generate the semi-Markov chain representing the time evolution of the Rabi frequencies. The duration of each state is modeled using a zero-truncated Poisson distribution with probability mass function given by

$$p_i(d) = \frac{\theta_i^d}{(e^{\theta_i} - 1)d!} \qquad i = 1, 2, \ldots, 5$$

where $d \in \mathcal{D}$ and for each state ω_i we set a different θ_i: 33, 41, 25, 53 and 38 respectively. The transition matrix used to simulate the hidden chain is shown in Table 1.

	1	2	3	4	5
1	0.0	0.1	0.3	0.1	0.5
2	0.3	0.0	0.1	0.4	0.2
3	0.1	0.1	0.0	0.6	0.2
4	0.2	0.3	0.1	0.0	0.4
5	0.2	0.1	0.5	0.2	0.0

Table 1: transition matrix used to simulate the semi-Markov chain

In what follows, the results of the experiment are presented.

5.2 Simulation results

Table 2 shows the estimated initial distributions with the non parametric kernel estimator (w NKE) and without it (w/o NKE). Figure 1 is a graphical representation of the case when $T = 300$. As it can be seen, for T up to 2000, the improvement in the estimated initial distribution is evident, especially for small sample sizes.

	T	State 1	State 2	State 3	State 4	State 5
True values		.4	.2	.1	.1	.2
w NKE	300	.3761	.1964	.1158	.1126	.1991
	500	.3773	.1927	.1203	.1111	.1985
	2000	.3832	.1874	.1151	.1189	.1953
	5000	.3787	.1967	.1133	.1171	.1939
	20000	.3702	.1879	.1179	.1197	.2042
w/o NKE	300	.2893	.2481	.1715	.1153	.1755
	500	.3202	.2351	.1485	.1091	.1871
	2000	.3451	.2211	.1263	.1140	.1932
	5000	.3457	.2297	.1192	.1145	.1907
	20000	.3467	.2123	.1232	.1139	.2038

Table 2: Simulation results I: initial distribution

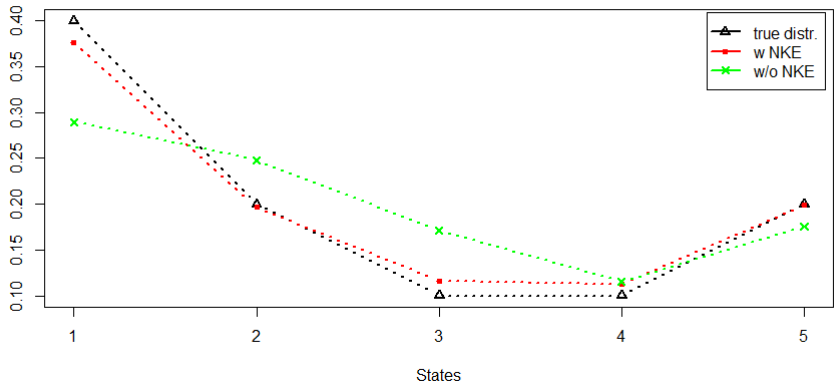


Figure 1: Initial distribution, T = 300

Table 3 presents Monte Carlo averages and standard errors for the parameter estimates. Specifically, the table shows: the percentage of states

correctly reconstructed using Viterbi algorithm, the sum of the absolute difference between the true value and the estimated one for the transition matrix and the percentage of bias in the estimated expected values for the duration distribution.

The percentage of states that are correctly reconstructed is uniformly higher when using the NKE, which gives an increase of about 2% w.r.t the standard formulation. Besides, the Monte Carlo standard errors are lower, which indicates an improvement in the precision of the estimates. For what concerns the transition matrix, we use, as a measure of goodness of fit, the sum of the absolute difference between the true value and the estimated one. The latter is always lower when the NKE is used, but the estimates shows some variability for small sample sizes such as $T = 300$ and $T = 500$. As regards for the state duration distributions, the NKE provides a reduction in the bias of the estimated expected values with only few exceptions, while the estimates variability turns out to be always lower.

One of the drawbacks of HSMM is that they require long sequences of observations to be correctly estimated, especially for complex models where the parameter space is high-dimensional. The result of the simulation show that, in presence of sparsity, our formulation outperforms the standard one, based on the cell proportion estimator, already for $T = 300$ and $T = 500$.

Parameter	T	w NKE		w/o NKE	
		Mean	MC SE	Mean	MC SE
Viterbi	300	.8270	.1005	.7960	.1088
	500	.8376	.0822	.8163	.0851
	2000	.8743	.0401	.8515	.0429
	5000	.8948	.0211	.8709	.0249
	20000	.9082	.0092	.8958	.0107
Trans. matrix	300	5.8249	.8755	6.1230	.8058
	500	5.0385	.8364	5.1691	.8295
	2000	2.6564	.5680	2.8687	.6255
	5000	1.6792	.3609	1.9083	.4427
	20000	.9156	.2228	1.1011	.3584
Bias State 1	300	-.2654	.2769	-.3055	.3006
	500	-.1079	.2697	-.1183	.3106
	2000	-.0434	.0944	-.0424	.1167
	5000	-.0401	.0547	-.0367	.0710
	20000	-.0368	.0254	-.0311	.0317
Bias State 2	300	-.3767	.3250	-.4313	.3908
	500	-.1909	.3147	-.2319	.3451
	2000	-.1252	.1683	-.1378	.1914
	5000	-.1325	.1218	-.1385	.1636
	20000	-.1570	.0813	-.2032	.1508
Bias State 3	300	-.0334	.5184	-.1189	.5819
	500	.1518	.4654	.1353	.5101
	2000	.0473	.1798	.0698	.2251
	5000	.0051	.0903	.0234	.1211
	20000	-.0129	.0396	.0101	.0495
Bias State 4	300	-.2125	.2900	-.2556	.3198
	500	-.0623	.2015	-.0791	.2184
	2000	-.0314	.0892	-.0345	.1005
	5000	-.0311	.0546	-.0312	.0561
	20000	-.0241	.0258	-.0180	.0397
Bias State 5	300	-.1373	.2701	-.1631	.3047
	500	-.0308	.1753	-.0275	.1918
	2000	-.0201	.0738	-.0161	.0891
	5000	-.0224	.1005	-.0201	.1088
	20000	-.0191	.0203	-.0131	.0248

Table 3: Simulation results II: Viterbi estimates, transition matrix and duration distributions

6 Concluding remarks

This paper was concerned with the problem of extracting the maximum information from an open quantum system subject to an external and non-stationary perturbation. We have presented a generalization of the explicit-duration Hidden Markov Models (EDHMM) which takes into account the presence of sparse data. Introducing a kernel estimator in the estimation procedure increases considerably the accuracy of the estimates, allowing to obtain a more reliable information about the evolution of the unobservable system.

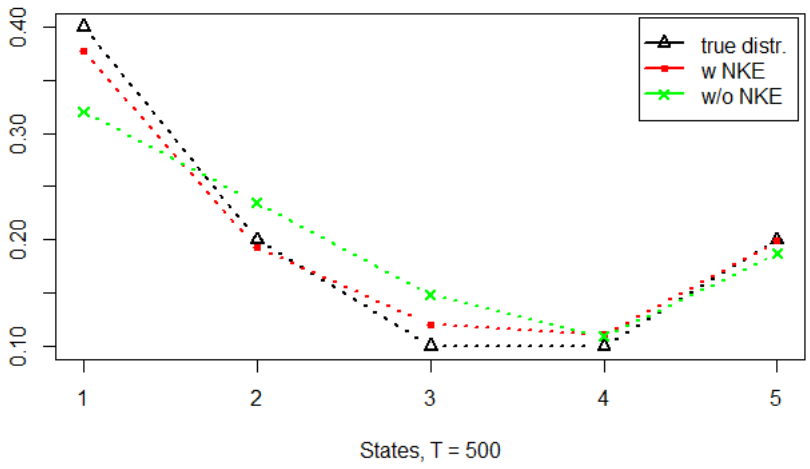
A generalization of the Viterbi algorithm to the extension of EDHMM, used to reconstruct the hidden chain, was developed. By means of an extensive Monte Carlo study, we showed that our formulation outperform the standard one both in the precision and in the variability of the estimates. We also found that our model can accurately tracks the hidden dynamic even for short observation sequences, which is a major improvement w.r.t. the standard formulation.

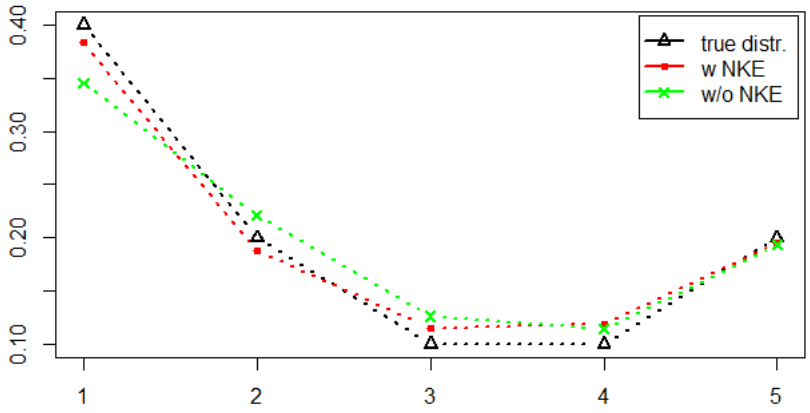
7 Appendix

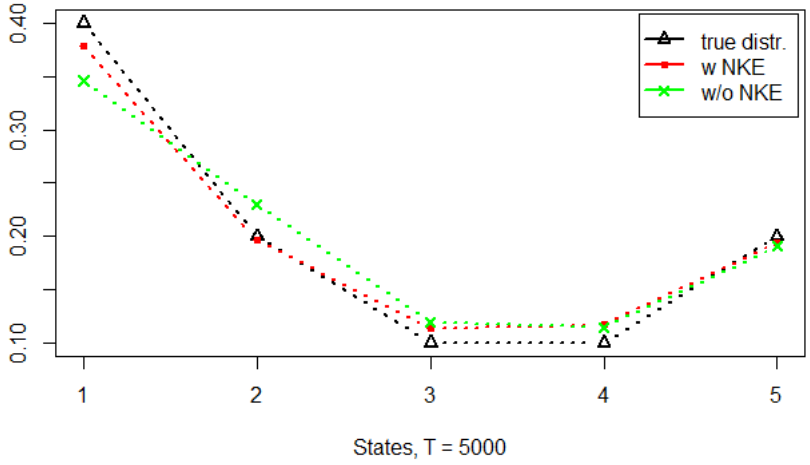
Additional results

In this appendix we present some additional simulation results. Specifically, Figure 2 shows the estimated initial distribution with and without the NKE for all the sample size analyzed. In Figure 3, the density of the percentage of correctly reconstructed states using Viterbi algorithm are reported; Figure 4 shows the Monte Carlo distribution of the goodness of fit measure used for the transition matrix.

Figure 2: Initial distribution for T = 500, 2000, 5000 and 20000




0.40
0.30
0.20
0.10
1
2
3
4
5
true distr.
w NKE
w/o NKE
States, T = 2000


0.40
0.30
0.20
0.10
1
2
3
4
5
true distr.
w NKE
w/o NKE
States, T = 5000

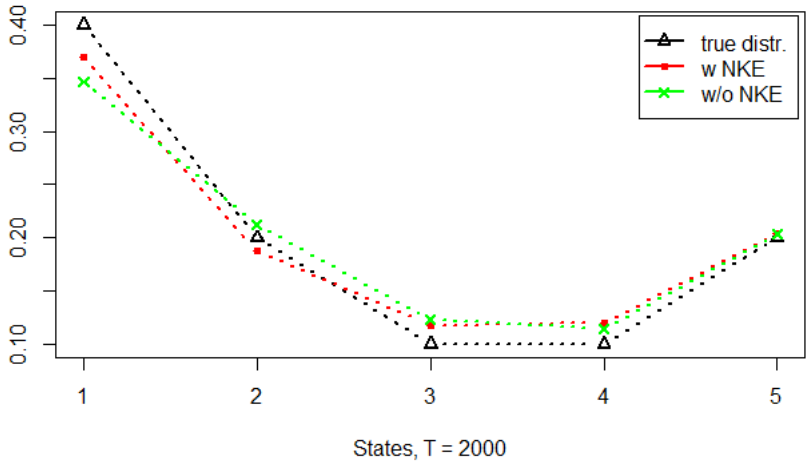

0.40
0.30
0.20
0.10
1
2
3
4
5
States, T = 2000
true distr.
w NKE
w/o NKE

Figure 3: Viterbi estimates, percentage of states correctly reconstructed

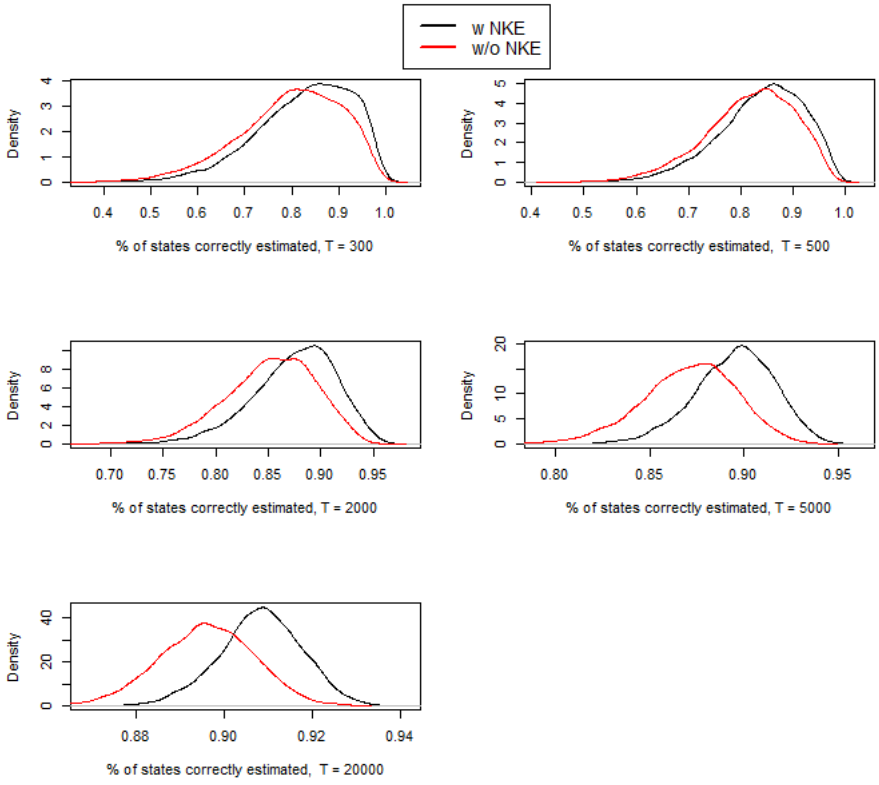
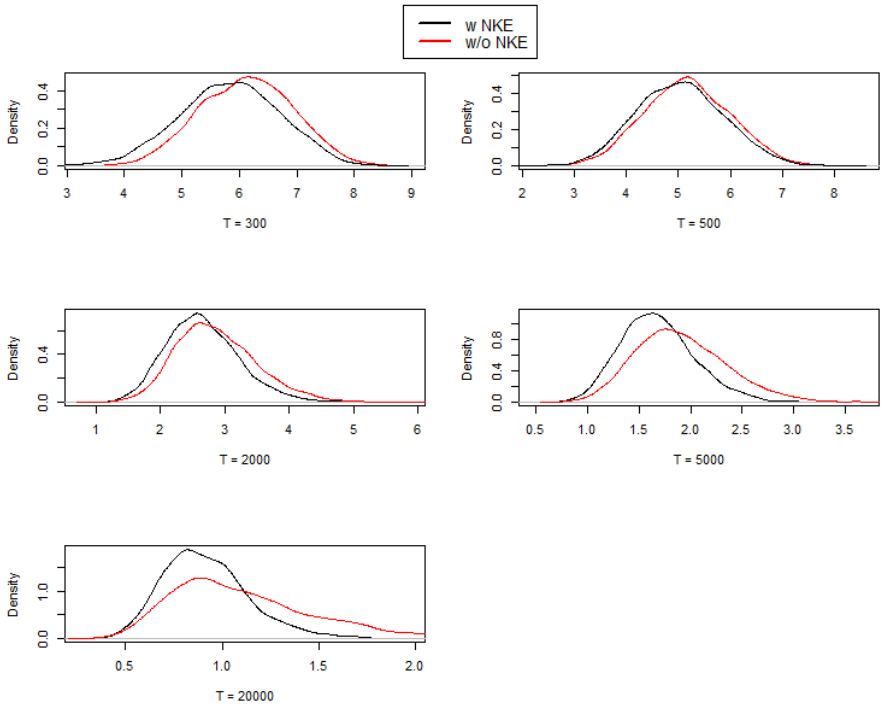


Figure 4: Transition matrix, sum of the absolute difference between the true value and the estimated one



References

Aitchison, J. and C. G. Aitken, 1976: Multivariate binary discrimination by the kernel method. *Biometrika*, **63**, 413–420.

Barbu, V. and N. Limnios, 2006: Maximum likelihood estimation for hidden semi-markov models. *Comptes Rendus Mathematique*, **342**, 201–205.

Barbu, V. S. and N. Limnios, 2009: *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*, vol. 191, Springer Science & Business Media.

Barndorff-Nielsen, O. E., R. D. Gill, and P. E. Jupp, 2003: On quantum statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 775–804.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss, 1970: A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 164–171.

Bietti, A., F. Bach, and A. Cont, 2015: An online em algorithm in hidden (semi-) markov models for audio segmentation and clustering, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Bishop, Y., S. Fienberg, and P. Holland, 1975: Discrete multivariate analysis: Theory and practice. 1975.

Bulla, J. and I. Bulla, 2006: Stylized facts of financial time series and hidden semi-markov models. *Computational Statistics & Data Analysis*, **51**, 2192–2209.

Chen, M.-Y., A. Kundu, and S. N. Srihari, 1995: Variable duration hidden markov model and morphological segmentation for handwritten word recognition. *Image Processing, IEEE Transactions on*, **4**, 1675–1688.

Dong, D. and I. R. Petersen, 2010: Quantum control theory and applications: a survey. *Control Theory & Applications, IET*, **4**, 2651–2671.

Ferguson, J. D., 1980: Variable duration models for speech, *Proceedings of the Symposium on the Application of HMMs to Text and Speech*.

Fienberg, S. E. and P. W. Holland, 1973: Simultaneous estimation of multinomial cell probabilities. *Journal of the American Statistical Association*, **68**, 683–691.

Gill, R. D., 2001: *Teleportation into quantum statistics*, EURANDOM.

Gleyzes, S., S. Kuhr, C. Guerlin, J. Bernu, S. Deleglise, U. B. Hoff, M. Brune, J.-M. Raimond, and S. Haroche, 2007: Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature*, **446**, 297–300.

Goggin, M., M. Almeida, M. Barbieri, B. Lanyon, J. OâĂŹBrien, A. White, and G. Pryde, 2011: Violation of the leggett–garg inequality with weak measurements of photons. *Proceedings of the National Academy of Sciences*, **108**, 1256–1261.

Guerlin, C., J. Bernu, S. Deleglise, C. Sayrin, S. Gleyzes, S. Kuhr, M. Brune, J.-M. Raimond, and S. Haroche, 2007: Progressive field-state collapse and quantum non-demolition photon counting. *Nature*, **448**, 889–893.

Hall, P. and D. Titterington, 1987: On smoothing sparse multinomial data. *Australian Journal of Statistics*, **29**, 19–37.

Hatridge, M., S. Shankar, M. Mirrahimi, F. Schackert, K. Geerlings, T. Brecht, K. Sliwa, B. Abdo, L. Frunzio, S. M. Girvin, et al., 2013: Quantum back-action of an individual variable-strength measurement. *Science*, **339**, 178–181.

Haussler, D. K. D. and M. G. R. F. H. Eeckman, 1996: A generalized hidden markov model for the recognition of human genes in dna, *Proc. Int. Conf. on Intelligent Systems for Molecular Biology, St. Louis.*

Helstrom, C. W., 1976: *Quantum detection and estimation theory, Mathematics in Science and Engineering*, Academic press.

Holevo, A., 1982: *Probabilistic and Statistical Aspects of Quantum Mechanics*, North-Holland, Amsterdam.

Johnson, M. J. and A. S. Willsky, 2013: Bayesian nonparametric hidden semimarkov models. *The Journal of Machine Learning Research*, **14**, 673–701.

Kirchmair, G., F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. Roos, 2009: State-independent experimental test of quantum contextuality. *Nature*, **460**, 494–497.

Krishnamurthy, V., J. B. Moore, and S.-H. Chung, 1991: On hidden fractal model signal processing. *Signal Processing*, **24**, 177–192.

Kubanek, A., M. Koch, C. Sames, A. Ourjoumtsev, P. W. Pinkse, K. Murr, and G. Rempe, 2009: Photon-by-photon feedback control of a single-atom trajectory. *Nature*, **462**, 898–901.

Kundu, A., Y. He, and M.-Y. Chen, 1997: Efficient utilization of variable duration information in hmm based hwr systems, *Image Processing, 1997. Proceedings., International Conference on*, vol. 3, IEEE.

Marron, J. S. and M. P. Wand, 1992: Exact mean integrated squared error. *The Annals of Statistics*, 712–736.

McFarland, J. M., T. T. Hahn, and M. R. Mehta, 2011: Explicit-duration hidden markov model inference of up-down states from continuous signals. *PloS one*, **6**, e21606.

Melnyk, I. and A. Banerjee, 2014: A spectral algorithm for inference in hidden semi-markov models. *arXiv preprint arXiv:1407.3422*.

Mitchell, C., M. Harper, L. Jamieson, et al., 1995: On the complexity of explicit duration hmm's. *IEEE transactions on speech and audio processing*, **3**, 213–217.

Mitchell, C. D. and L. H. Jamieson, 1993: Modeling duration in a hidden markov model with the exponential family, *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 2, IEEE.

Nielsen, M. A. and I. L. Chuang, 2010: *Quantum computation and quantum information*, Cambridge university press.

Palacios-Laloy, A., F. Mallet, F. Nguyen, P. Bertet, D. Vion, D. Esteve, and A. N. Korotkov, 2010: Experimental violation of a bell/'s inequality in time with weak measurement. *Nature Physics*, **6**, 442–447.

Rabiner, L., 1989: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 257–286.

Rajagopalan, B. and U. Lall, 1995: A kernel estimator for discrete distributions. *Journal of Nonparametric Statistics*, **4**, 409–426.

Ramakrishna, V. and H. Rabitz, 1996: Relation between quantum computing and quantum controllability. *Physical Review A*, **54**, 1715.

Ramesh, P. and J. G. Wilpon, 1992a: Modeling state durations in hidden markov models for automatic speech recognition, *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 1, IEEE.

Ramesh, P. and J. G. Wilpon, 1992b: Modeling state durations in hidden markov models for automatic speech recognition, *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 1, IEEE.

Schmidler, S. C., J. S. Liu, and D. L. Brutlag, 2000: Bayesian segmentation of protein secondary structure. *Journal of computational biology*, **7**, 233–248.

Simonoff, J. S., 1983: A penalty function approach to smoothing large sparse contingency tables. *The Annals of Statistics*, 208–218.

Simonoff, J. S., 2012: *Smoothing methods in statistics*, Springer Science & Business Media.

Squire, K. and S. E. Levinson, 2005: Recursive maximum likelihood estimation for hidden semi-markov models, *Machine Learning for Signal Processing, 2005 IEEE Workshop on*, IEEE.

Titterington, D., 1980: A comparative study of kernel-based density estimates for categorical data. *Technometrics*, **22**, 259–268.

Vijay, R., D. Slichter, and I. Siddiqi, 2011: Observation of quantum jumps in a superconducting artificial atom. *Physical review letters*, **106**, 110502.

Wand, M. P. and M. C. Jones, 1994: *Kernel smoothing*, Crc Press.

Wang, M.-C. and J. Van Ryzin, 1981: A class of smooth estimators for discrete distributions. *Biometrika*, **68**, 301–309.

Wang, Y., 2012: Quantum computation and quantum information. *Statistical Science*, **27**, 373–394.

Wiseman, H. M. and G. J. Milburn, 2009: *Quantum measurement and control*, Cambridge University Press.

Yu, S.-Z., 2010: Hidden semi-markov models. *Artificial Intelligence*, **174**, 215–243.

Yu, S.-Z. and H. Kobayashi, 2006: Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model. *Signal Processing, IEEE Transactions on*, **54**, 1947–1951.

MCMC estimation of Explicit-Duration Hidden Markov Models

Contents

Abstract

A Markov chain Monte Carlo (MCMC) sampling scheme for estimating the explicit-duration hidden Markov models is developed. The procedure is straightforward to implement and robust against the underflow problem. Motivated by the presence of sparse data (lack of information), we make use of a kernel estimator for discrete data for estimating the state duration distribution. This choice provides a significant improvement in the precision of the estimates for all the parameters, in particular for the transition probabilities. One of the main limitations of HSMM consists in the prior specification of the number of states for the hidden chain, which has to be known. In order to make our formulation more flexible, we employ a Bayesian model selection procedure (Congdon (2006)), which enables to avoid a prior specification of the number of states. As an application, we discuss an explicit-duration HMM characterization of a two-level open quantum system subject to an external and non-homogeneous perturbation causing a change in the oscillation frequency. We investigate the finite samples properties of our formulation by means of an extensive Monte Carlo study.

1 Introduction

Hidden Markov models (HMM) are statistical models in which the hidden stochastic process is a M-state Markov process $\{X_t\}_{t\in T}$, indexed by a discrete time set T, with transition probabilities $P(X_{t+1} = j \,|\, X_t = i) = a_{ij}$ $1 \leq i, j \leq M$, observed through a stochastic process $\{Y_t\}_{t\in T}$ representing the measured signal. Since $\{X_t\}_{t\in T}$ is not directly observable, parameter estimation and inference have to be based only on the observed signal.

HMM have become a fundamental technique for modeling weakly dependent sequences of random variables, receiving considerable attention due to their great flexibility and ability to model a wide variety of phenomena. For this reasons, in the last decades, HMM have been largely studied by statisticians, engineers and information theorists. In particular, Bickel et al. (1998), Douc et al. (2001), Jensen and Petersen (1999), proved consistency and asymptotic normality of the maximum likelihood estimator for the parameters, Rydén (1994), Leroux (1992) and Rydén (1995) provided the identifiability conditions for of a general HMM and Ito et al. (1992) addressed the identifiability of deterministic functions of non-stationary Markov chains (see also Rabiner (1989); Ephraim and Merhav (2002); Cappe et al. (2007) for a comprehensive treatment)

However, one of the major weaknesses of HMM is the inflexibility in modeling state duration, which is geometric distributed by construction. This drawback can be avoided by extending the conventional HMM to a more general class of models known as hidden semi-Markov models (HSMM), or non-stationary HMM, which enable an explicit specification of the state duration distribution. The HSMM formulation allows to capture the statistical properties of the hidden system dynamics, including short and medium-range dependence structures, which go beyond the Markovian short-memory assumption implied by the standard HMM.

In this paper we present a Markov chain Monte Carlo (MCMC) estimation of a particular class of HSMM called explicit-duration hidden Markov models (Yu (2010), Mitchell and Jamieson (1993), Rabiner (1989)).

As an application, we discuss an HSMM characterization of a two-level open quantum system subject to an external and non-homogeneous perturbation causing a change in the oscillation frequency. Furthermore, we will treat both parametric and non-parametric specifications. A non parametric framework allows the model to be more flexible: it does not require the explicit specification of the state duration distribution, not always available to exper-

imentalists. However, it requires a much longer observation sequence than the one needed by a parametric assumption and involves a greater number of parameters in the estimation algorithm. For this reason, we make use of a non-parametric kernel estimator for discrete distribution which improve the accuracy of the estimates in presence of lack of information (sparsity). We also employ a Bayesian model selection procedure based on posterior models probability (Congdon (2006)) which allows one to avoid a prior specification of the number of states in the hidden chain.

The main contributions of the paper are the extension of the MCMC method for the estimation of the HSMM, and its application for tracking the dynamics of an open quantum systems which undergoes an external time-dependent perturbation. Moreover, by introducing a kernel estimator in the MCMC procedure in order to improve the estimates accuracy in presence of sparse data.

HSMM are generally treated from a non-Bayesian perspective in the literature, estimated using an approximate maximum-likelihood procedure (especially with the Expectation-Maximization algorithm). To the best of our knowledge, few has been done for the Bayesian estimation of non-stationary HMM. Johnson and Willsky (2013) introduced the explicit-duration Hierarchical Dirichlet Process HSMM and develop two Gibbs sampling algorithms, the weak limit and direct assignment samplers. In Djuric and Chun (2002), the authors provide an MCMC-based estimation method for HSMM with a parametric specification of the state duration distribution using a different formulation for the semi-Markov chain. However, such formulation is computationally demanding and it also suffers the under-flow problem. The formulation presented in our work is flexible, straightforward to implement, it does not suffer the underflow problem and provides an excellent accuracy in the estimation of the hidden dynamics.

The remainder of the paper is organized as follows. Section 2 introduces the explicit-duration HMM and the main contributions, namely the MCMC procedure and the kernel estimator used. Section 3 presents the experimental setup analyzed and the results of the Monte Carlo study. Section 4 concludes the paper.

2 Explicit-Duration HMM

HSMM are discrete-time, finite-state non-stationary Markov chains observed through finite set of transition densities indexed by the states of the chains. While in the conventional HMM each state has a geometric (or exponential) duration by construction, in the HSMM the state duration can be explicitly modeled by a random variable that takes integer values in the set $\mathcal{D} = \{1, 2, \dots, D\}$.

Since the pioneering work of Ferguson (Ferguson (1980)) in which the author presented the "variable duration HMM", HSMM have been largely studied and applied to many different fields. In particular, Barbu and Limnios (2006) and Barbu and Limnios (2009) proved consistency and asymptotic normality for nonparametric maximum likelihood estimators, Squire and Levinson (2005) and Bietti et al. (2015) propose recursive maximum-likelihood algorithm and incremental EM algorithm for online estimation and Melnyk and Banerjee (2014) introduced a spectral algorithm for inference in HSMM.

HSMM have been successfully applied in many areas among which, financial time series modeling (Bulla and Bulla (2006)), handwritten word recognition (Kundu et al. (1997)), recognition of human genes in DNA (Haussler and Eeckman (1996)), protein structure prediction (Schmidler et al. (2000)) and for determining duration and timing of up-down state in neocortical neurons (McFarland et al. (2011)).

This paper is concerned with the class of explicit-duration hidden Markov models (EDHMM). More formally, an EDHMM is defined by

- $\{X_t\}_{t\in\mathbb{N}}$ the non stationary Markov chain on $\mathcal{S} = \{1, 2, \dots, M\}$ the set of M states, where $X_{1:T} = \{X_1, \dots, X_T\}$ denotes the state sequence up to time T;
- $\mathcal{A} = \{a_{i(j,d)}\}_{i,j=1,\dots,M,\ d=1,\dots,D}$ the set of transition probability from state i to state j with duration d,

  $$a_{i(j,d_2)} = a_{ij} p_j(d_2),$$

  where

  $$a_{ij} = P(X_{t+1} = j \,|\, X_t = i), \qquad 1 \le i, j \le M;$$

  the transition probability from state i to state j, with null self-transition probability, $a_{ij} = 0$, and

  $$p_j(d) = P(X_{t+1:t+d]} = j \,|\, X_{[t+1} = j), \quad 1 \le j \le M,\ 1 \le d \le D$$

the probability that state j has a duration d, starting from time $t+1$ until $t+d$;

- $\{Y_t\}_{t\in\mathbb{N}}$ the sequence of conditionally independent random variables on $\mathcal{V} = \{v_1, \ldots, v_K\}$, representing the observed signal. Specifically, Y_t has a conditional distribution which depends on the chain only through its current value X_t. Moreover, $\mathcal{V}$ is the set of K measurement results per state, corresponding to the physical output of the system being observed;

- $\mathcal{B} = \{b_j(Y_t = v_k)\}_{j=1,\ldots,M,\ k=1,\ldots,K}$ the set of emission probabilities,

$$b_j(Y_t = v_k) = P(Y_t = v_k \,|\, X_t = j), \qquad 1 \le j \le M,\ 1 \le k \le K$$

representing the probability to observe the value $Y_t = v_k$ when $X_t = j$;

- $\Pi = \{\pi_i\}_{i=1,\ldots,M}$ the initial state distribution, where

$$\pi_i = P(X_1 = i), \qquad 1 \le i \le M.$$

The set of model parameters is defined by

$$\lambda = (\mathcal{A}, \mathcal{B}, \Pi).$$

In the following section, the estimation procedure developed is introduced.

2.1 Estimation

The inference is carried out with Gibbs samplers when possible and Metropolis-Hastings steps within a MCMC sampling scheme. More specifically, when the state duration distribution is assumed to be parametric, the Metropolis-Hastings algorithm is used for sampling from the posterior distribution. On the other hand, if the state duration is treated non-parametrically a Gibbs sampling approach is used.

2.1.1 Specification of the Priors

The first step consists in the prior specification for the unknown parameters. All the chosen priors are non-informative. We will discuss first the non-parametric specification for the state duration and then the parametric one.

If we cannot specify any functional form for the state duration, the approach used is completely non parametric. Since the likelihood functions of the initial state probabilities, the state transition probabilities, and the state duration distributions are modeled by multinomial distributions, the standard priors are the multivariate Dirichlet distributions. This implies that the posterior densities of the parameters will be Dirichlet distributed as well. More specifically:

- the prior for initial state distribution $\Pi = \{\pi_i\},\ 1 \le i \le M$ is the multivariate Dirichlet distribution of dimension $M-1$

$$\Pi \sim \mathcal{D}i(\alpha_1, \ldots, \alpha_M);$$

- for a given state i, the prior for the transition probabilities $A_i = \{a_{i1}, \ldots, a_{i(i-1)}, a_{i(i+1)}, \ldots, a_{iM}\},\ 1 \le j \le M,\ i \ne j$ is the multivariate Dirichlet distribution of dimension $M-2$

$$A_i \sim \mathcal{D}i(\gamma_1, \ldots, \gamma_{M-1});$$

- for a given state i, the prior for the emission probabilities $B_i = \{b_i(v_1), \ldots, b_i(v_K)\},\ 1 \le i \le M$ is the multivariate Dirichlet distribution of dimension $K-1$

$$B_i \sim \mathcal{D}i\Big(\epsilon_1, \ldots, \epsilon_{K-1}\Big);$$

- for each state, the prior for the state duration distribution $p_i(d)$ is the multivariate Dirichlet distribution of dimension $D-1$

$$p_i(d) \sim \mathcal{D}i(\eta_1, \ldots, \eta_D)$$

  where D represents the maximum number of time units in the same system state.

For the parametric specification of the state duration distribution, we choose the zero-truncated Poisson, whose probability mass function reads

$$p_i(d) = \frac{\theta_i^d}{(e^{\theta_i} - 1)d!} \tag{1}$$

where $i = 1, 2, \ldots, M$ and $d \in \mathcal{D} = \{1, \ldots, D\}$. Since (1) does not belong to the conjugate distributions, a Metropolis-Hastings approach must be used. In particular, we choose a uniform distributed prior for the parameter θ.

In the next section the MCMC estimation algorithm will be presented.

2.2 Gibbs Sampling

Once the priors are chosen, the Gibbs sampling procedure can be implemented as follows. At iteration k of the algorithm:

- draw $\Pi^{(k)}$ form the $(M-1)$-dimensional Dirichlet distribution according to
$$\Pi \sim \mathcal{D}i\Big(\alpha_1 + \delta_{X_1,1}^{(k-1)}, \ldots, \alpha_M + \delta_{X_1,M}^{(k-1)}\Big)$$
where $\delta_{X_1,i}^{(k-1)} = 1$ if $X_1^{(k-1)} = i$ and zero otherwise;

- draw $A_i^{(k)}$ for $i = 1, \ldots, M$ from the $(M-2)$-dimensional Dirichlet distribution as follows
$$A_i \sim \mathcal{D}i\Big(\gamma_1 + n_{i1}^{(k-1)}, \ldots, \gamma_{M-1} + n_{i(M-1)}^{(k-1)}\Big)$$
where $n_{ij}^{(k-1)}$ is the number of transition from state i to state j at iteration $k-1$;

- draw $p_i(d)^{(k)}$ for $i = 1, \ldots, M$ from the $(D-1)$-dimensional Dirichlet distribution
$$p_i(d) \sim \mathcal{D}i\Big(\eta_1 + n_{i1}^{(k-1)}, \ldots, \eta_D + n_{iD}^{(k-1)}\Big)$$
where $n_{id}^{(k-1)}$ represent the number of times state i has lasted d time units at iteration $k-1$;

- draw $B_i^{(k)}$ for $i = 1, \ldots, M$ from the $(K-1)$-dimensional Dirichlet distribution
$$B_i \sim \mathcal{D}i\Big(\epsilon_1 + n_{i1}^{(k-1)}, \ldots, \epsilon_{K-1} + n_{i(K-1)}^{(k-1)}\Big)$$
where $n_{i(K-1)}^{(k-1)}$ is the number of measurement output v_k in state i;

- draw $X_t^{(k)}$ according to
$$X_t \sim P\Big(X_t \,\Big|\, X_{t-1}^{(k)}, d_{X_{t-1}^{(k)}}, X_{t+1}^{(k-1)}, d_{X_{t+1}^{(k-1)}}, \lambda^{(k)}, Y_t\Big)$$
where $d_{X_{t-1}^{(k)}}$ and $d_{X_{t+1}^{(k-1)}}$ indicate the duration of state X_{t-1} and X_{t+1} at iteration k and $k-1$, respectively. Furthermore, $\lambda^{(k)}$ represents the set of model parameter obtained at iteration k.

In the next section, we shall deal with the estimation of the duration distribution.

2.3 Estimation of the duration distribution

One of the novelties of this paper is the introduction of a non-parametric kernel estimator for the discrete distribution of the duration in the iterative step of the estimation algorithm. Specifically, we shall consider an estimator of the form

$$\hat{p}_i(d) = \sum_{j-d=-h}^{h} K((j-d)/h)p_i(j)$$

where $p_i(j) = \frac{n_{ij}}{n_i}$ is the sample relative frequency, h and d are positive integer, $h+1 < d < D-(h-1)$, and $K(t)$ is a discrete kernel, that is a non negative, symmetric function of t which adds up to one.

The relevance of the contribution lies in the fact that smoothing methods for discrete data are optimal in the mean summed squared error (MSSE) sense in the case of sparsity (Hall and Titterington (1987)), as it is the case of the experiment considered in this paper. In fact, the cell proportion estimator $p_i = \frac{n_i}{n}$ (the MLE for a multinomial distribution) traditionally used in the probability mass function estimation is a consistent estimator of $p_i(d)$ only when the sample size becomes large compared with the number of cells (Fienberg and Holland (1973); Simonoff (1983); Simonoff (2012)). In the case when the number of cells is close to or greater than the number of observations, which results in a sparse table with many small or zero cell counts, the cell proportion estimator p_i is inconsistent and provides estimates characterized by roughness and multimodality. Smoothing methods based on kernel estimators have been proved to be effective for sparse multinomial data (Aitchison and Aitken (1976); Titterington (1980); Wang and Van Ryzin (1981); Simonoff (1983)). Besides, Hall and Titterington (1987) proved the optimality of kernel estimators for sparse multinomial data in the framework of sparse asymptotic earlier introduced by Fienberg and Holland (1973) and Bishop et al. (1975).

Using a kernel smoother requires that the shape and the bandwidth of the kernel function are selected. As far as the kernel function is concerned, we have chosen the discrete version the Epanechnikov kernel, which belongs to the class of the kernels generated as Beta distributions and has optimal asymptotic properties for continuous smoothing, in the sense that it minimises the asymptotic integrated mean square error (AMISE), (Wand and Jones (1994)).

In practice, we have a discrete quadratic kernel $K(t) = at^2 + b$, where t takes value in a discrete set, which is a common choice in kernel smoothing

as it represents a good compromise between fitting ad smoothing. (Marron and Wand (1992); Simonoff (2012)). Actually, we use the quadratic kernel formulation of Rajagopalan and Lall (1995), with minor corrections, who derive the value of the normalising constants a and b both in the interior and in the boundaries. These value must be selected in order for the kernel to add up to one, $\sum_t K(t) = 1$, and satisfy the properties of symmetry $K(t) = K(t)$, positivity, $K(t) > 0$ within the bandwidth, and unbiasedness $\sum_t K(t)t = 1$. For the interior region, i.e. for $h + 1 < d < D - h - 1$, the parameters a and b are

$$a = \frac{3h}{(1 - 4h^2)}, \; b = -a.$$

At the boundaries, i.e. for $1 < d < h + 1$ and $d > D - h - 1$, a and b become

$$a = \frac{-D}{2h(h+d)}\left(\frac{E}{4h^3} - \frac{CD}{12h^3(h+d)}\right)^{-1}, \quad b = \left(1 - \frac{aC}{6h^2}\right)(h+d)^{-1}$$

where

$$\begin{aligned} C &= h(h+1)(2h+1) + (d-2)(d-1)(2h-3) \\ D &= -h(h+1) + (d-2)(d-1) \\ E &= (-h(h+1))^2 + ((d-2)(d-1))^2. \end{aligned}$$

If the kernel has a limited impact on the estimation, the bandwidth is a crucial parameter. In the context of sparse asymptotics, Hall and Tittering-ton (1987) have shown the optimality of the least cross validation criterion (LSCV) in terms of rate of convergence.

Hence, the optimal h is found by minimizing the LSCV function defined as

$$LSCV(h) = \sum_{d=1}^{D} (\hat{p}_i(d))^2 - \frac{2}{D} \sum_{d=1}^{D} \hat{p}_{\setminus i}(d) d_i$$

where $\hat{p}_{\setminus i}(d)$ represents the estimate of $p_i(d)$ without the i-th's contribute.

In the next section, the finite sample properties of the formulation introduced are investigated through a Monte Carlo study.

3 Simulation study

To validate the finite samples properties of the ED-HMM model introduced above, a Monte Carlo study is carried out. The simulation is concerned

with an HSMM characterization of a two-level open quantum system subject to an external and non-homogeneous perturbation causing a change in the oscillation frequency. The description of the experimental setup is reported in the next section.

3.1 Quantum experiment

We design an experimental setup that can be thought of as a modification of the one presented by Nobel Laureate Serge Haroche (Guerlin et al. (2007)). In the latter, from an initial state exhibiting quantum uncertainty in the measured observable, the system is gradually projected into a state in which this observable becomes precisely known, that is, the step-by-step 'state collapse' caused by repeated soft measurements is experimentally observed for the first time.

Here, a two-level atom (system A) is driven by an external and non-homogeneous force causing a change in the Rabi frequency. The latter represents the characteristic angular frequency of the atom corresponding to the oscillation between the two levels, caused by the time evolution. Two-level atoms (system B, also called ancilla) are sent through the system A and weakly interact with it. Then, a projection measurement on each B system is performed, thus avoiding a direct perturbation of the system A. The projection measurement of the ancilla systems causes a non-projective 'weak' disturbance of the system A, influencing its quantum state.

The experimental set up introduced so far can be nested in the framework of hidden semi-Markov models (HSMM), since the measurement results on the B are governed by the current (hidden) state of the system A, which, in turn, is driven by a semi-Markov evolution process. Specifically, the transition probabilities form one state to another depend on the state duration, and thus they are time varying. The emission probabilities are time-dependent as well, due to the propagation of the wave function at different frequency values.

Our purpose is to keep track of the evolution of system A and to estimate the state of the chain (the value of the Rabi frequency) at each time-step through measurements on the ancillas B, which reveal only partial information about the system A.

The model can be sketched as follows. Let us consider a semi-Markov chain with n states $\mathcal{S} = \{\omega_1, \ldots, \omega_n\}$ representing the Rabi frequencies of the system and two output symbols '0' and '1', representing the measurement

results. The quantum system A is initialized in a (known) superposition state

$$|\Psi_A(\omega_i)\rangle_{t_0} = a_{t_0}(\omega_i)\,|0_A\rangle + b_{t_0}(\omega_i)\,|1_A\rangle$$

where we stress the dependency of the A state on the value of $\omega_i \in \mathcal{S}$. Furthermore, a, b are complex numbers satisfying $|a(\omega_i)|^2 + |b(\omega_i)|^2 = 1$, $|\Psi\rangle \in \mathcal{H}_2$ represents a unit vector in the two dimensional complex Hilbert space and $\{|0_A\rangle, |1_A\rangle\}$ denotes an orthonormal basis for the two dimensional vector space. In the rest of the paper, we will suppress the dependence of α and β on ω_i and we will maintain it only in the quantum state. The ancilla system is initially in the state $|\Psi_B\rangle \in \mathcal{H}_2$, with orthonormal basis $\{|v_B\rangle\,; v \in \mathcal{V}\}$ where $\mathcal{V} = \{0, 1\}$ is the set of all possible measurement outcome. The initial state of the composite system is $|\Psi_A(\omega_i)_{t_0}\rangle \otimes |\Psi_B\rangle$ and belongs to $\mathcal{H} = \mathcal{H}_2 \otimes \mathcal{H}_2$. When the interaction between the systems takes place, the system A and the meter become correlated, and the subsequent entangled state reads

$$\begin{aligned} |\Psi_{AB}(\omega_i)\rangle_{t_1} &= \alpha_{t_1}\,|0_A\rangle \otimes \left|\Psi_B^0\right\rangle + \beta_{t_1}\,|1_A\rangle \otimes \left|\Psi_B^1\right\rangle \\ &= \alpha_{t_1}\,|0_A\rangle \otimes \left(\sqrt{p}\,|0_B\rangle + \sqrt{q}\,|1_B\rangle\right) + \beta_{t_1}\,|1_A\rangle \otimes \left(\sqrt{q}\,|0_B\rangle + \sqrt{p}\,|1_B\rangle\right) \end{aligned}$$

where $p+q = 1$, representing the ability to distinguish between the two states. Then, the ancilla system is measured and the result '0' is obtained with probability

$$P_{t_1}(B = 0\,|\,|\Psi_{AB}(\omega_i)\rangle_{t_1}) = |\alpha_{t_1}\sqrt{p}|^2 + |\beta_{t_1}\sqrt{q}|^2 \tag{2}$$

This quantity represents the probability to obtain the output value '0' conditional on the system dynamics up to time t_1 and then on the ω_i value which governs the evolution of the system A. In the HMM framework, such quantity can be viewed as the emission probability, at time t_1, of symbol '0' when the system is in state ω_i, that is $P_{t_1}(B = 0\,|\,\omega_i)$. Furthermore, it is important to note that this probability is time-varying, since, depending on the ω_i driving the system dynamics at time t_1, the probability to observe each output can change. The measurement resulting in the output '0' leaves the system A in the (unnormalized) state

$$|\Psi_A(\omega_i\,|\,B = 0)\rangle = \alpha_{t_1}\sqrt{p}\,|0_A\rangle + \beta_{t_1}\sqrt{q}\,|1_A\rangle\,,$$

where we explicitly indicate that the state is conditional on the previous result on the B system. Instead, the measurement result '1' is obtained with probability

$$P_{t_1}(B = 1\,|\,|\Psi_{AB}(\omega_i)\rangle_{t_1}) = |\alpha_{t_1}\sqrt{q}|^2 + |\beta_{t_1}\sqrt{p}|^2 \tag{3}$$

and the (unnormalized) A state is

$$|\Psi_A(\omega_i \,|\, B = 1)\rangle = \alpha_{t_1}\sqrt{q}\,|0_A\rangle + \beta_{t_1}\sqrt{p}\,|1_A\rangle\,.$$

From equations (2) and (3) it can be seen that the quantities p and q represent the ability to distinguish between the states and then they can be thought of as a measurement effectiveness.

3.2 Design of the Monte Carlo study

The simulation setting is the same as in Chapter 2.

The observed signal $Y_{1:T}$ is a binary sequence obtained by measuring the systems B. It is characterized as a discrete-time random process driven by an underlying (hidden) semi-Markov chain represented by the evolution of system A.

The hidden chain is composed of $M = 5$ states corresponding to 5 different frequency values in the range $(0, \frac{\pi}{2})$, that is $\mathcal{S} = \{\omega_1 = 0.003,\ \omega_2 = 0.4,\ \omega_3 = 0.85,\ \omega_4 = 1.17,\ \omega_5 = 1.55\}$. The model's ability to correctly characterize the hidden states strongly depends on the type of measurement performed. Since the output is a binary time series, our simulations show that the model is able to obtain a good fit up to 5 chain states, beyond which the accuracy of the estimates decreases. It is worth to remark that the value $M = 5$ is far beyond the one found in commonly applied models, which usually consider only one frequency value.

The truncation value D, representing the maximum number of consecutive time steps in the same state, is set to $D = 150$ for $T = 300$, $D = 250$ for $T = 500$, $D = 1000$ for $T = 2000$ and $D = 2500$ for $T = 5000, 20000$. It is worth mentioning that in the present formulation, the choice of D does not directly affect the computational complexity of the algorithm, as the number of calculations is linear in D. This represents an improvement of our formulation with respect to standard non-Bayesian estimation method based on the Expectation-Maximization algorithm. In fact, in the latter, the computational complexity is a quadratic function of D, that is $O(M^2+MD+MD^2)T$, and then a large value of D can cause a curse of dimensionality that makes the estimation procedure unfeasible.

The emission probabilities are obtained by propagating the wave function at each frequency ω_i: this represents a point of strength in the estimation/re-estimation procedure (see Rabiner (1989)), in that they do not enter in the

estimation algorithm. Moreover, the latter is based on the non parametric kernel estimator described in section 2.3. The initial state sequence is randomly generated.

Five different sample sizes are considered: T = 300, 500, 2000, 5000, 20000. The number of Monte Carlo replications is set to 1000 for each sample size.

In the first step of the simulation, we generate the semi-Markov chain representing the time evolution of the Rabi frequencies. The duration of each state is modeled using a zero-truncated Poisson distribution with probability mass function given by

$$p_i(d) = \frac{\theta_i^d}{(e^{\theta_i} - 1)d!} \qquad i = 1, 2, \ldots, 5$$

where $d \in \mathcal{D}$ and for each state ω_i we set a different θ_i: 33, 41, 25, 53 and 38 respectively. The transition matrix used to simulate the hidden chain is shown in Table 1.

	1	2	3	4	5
1	0.0	0.1	0.3	0.1	0.5
2	0.3	0.0	0.1	0.4	0.2
3	0.1	0.1	0.0	0.6	0.2
4	0.2	0.3	0.1	0.0	0.4
5	0.2	0.1	0.5	0.2	0.0

Table 1: transition matrix used to simulate the semi-Markov chain

In what follows, the results of the experiment are presented.

3.3 Simulation results

Table 2 shows the estimated initial distributions with the non parametric kernel estimator (w NKE) and without it (w/o NKE). As it can be seen, for T up to 2000, the improvement in the estimated initial distribution is evident, especially for small sample sizes.

	T	State 1	State 2	State 3	State 4	State 5
True values		.4	.2	.1	.1	.2
w NKE	300	.2916	.2977	.0805	.1112	.2188
	500	.3713	.2415	.0908	.097	.1991
	2000	.4058	.1914	.1137	.1005	.2083
	5000	.3874	.1925	.1067	.0905	.2226
	20000	.3799	.2106	.0941	.1058	.2213
w/o NKE	300	.6590	.2389	.0706	.0270	.0042
	500	.1608	.2093	.0881	.1857	.3559
	2000	.3369	.1983	.0907	.1429	.2111
	5000	.3599	.1808	.0978	.1471	.2141
	20000	.3822	.1921	.1019	.1199	.2037

Table 2: Simulation results I: initial distribution

In Table 3, Monte Carlo averages and standard errors for the model parameters are presented. In particular, the percentage of the hidden chain correctly reconstructed, the sum of the absolute difference between the true value and the estimated one for the transition matrix and the percentage of bias in the estimated expected values for the duration distribution are reported.

The percentage of correctly reconstructed states is always higher using NKE, which increases up to 94.8%. The differences between the two formulations remains stable for all the sample sizes and is about 4%. The standard errors are lower with NKE, especially for small samples, indicating an improvement in the precision of the estimates in presence of lack of information. Regarding the transition matrix, the goodness of fit measure used (the sum of the absolute difference between the true value and the estimated one) is always lower using NKE especially for $T = 300, 500$. The variability of the estimates for the transition matrix using NKE decreases considerably and it is always about 35% lower than without the kernel estimator. For what concerns the bias in the state duration distributions, using NKE provides an improvement in the accuracy of the estimates with only few exceptions in

particular for $T = 300$, while the estimates variability turns out to be always lower.

The simulation shows that our formulation provides accurate estimates and is able to correctly reconstruct the evolution of the hidden process even for small sample sizes. This represents a point of strength of our formulation, since usually HSMM need long observation sequences in order to obtain reliable estimates, in particular for complex models with high dimensional parameter space.

Parameter	T	w NKE		w/o NKE	
		Mean	MC SE	Mean	MC SE
Recon. states	300	.8003	.1075	.7696	.1781
	500	.8286	.0953	.7839	.1218
	2000	.8958	.0468	.8537	.0665
	5000	.9166	.0252	.8793	.0389
	20000	.9482	.0146	.9099	.0246
Trans. matrix	300	2.6165	.4438	3.7294	.6746
	500	2.6172	.3931	3.4160	.5985
	2000	2.4439	.3605	2.8129	.4646
	5000	2.1666	.2704	2.6670	.3829
	20000	2.0737	.2038	2.4149	.3584
Bias State 1	300	-.3323	.3880	-.3055	.4441
	500	-.1639	.3533	-.2177	.4017
	2000	-.0391	.1030	-.1142	.1887
	5000	-.0431	.0299	-.0863	.0903
	20000	-.0429	.0300	-.0522	.0531
Bias State 2	300	-.2315	.2536	-.3413	.3382
	500	-.1729	.2014	-.2856	.3125
	2000	-.1137	.1080	-.1197	.2112
	5000	-.1346	.0325	-.1171	.0914
	20000	-.1448	.0285	-.1321	.0815
Bias State 3	300	-.1771	.2633	-.1573	.2911
	500	-.1127	.1915	-.1156	.2372
	2000	-.1355	.1017	-.1044	.1513
	5000	-.0392	.0333	-.0871	.0583
	20000	-.0370	.0304	-.0513	.0477
Bias State 4	300	-.2237	.2206	-.2777	.2423
	500	-.1311	.1467	-.1946	.2003
	2000	-.0262	.0929	-.0993	.1447
	5000	-.0289	.0316	-.0711	.0763
	20000	-.0288	.0294	-.0407	.0516
Bias State 5	300	-.1522	.2413	-.1507	.2635
	500	-.0396	.1748	-.0992	.2121
	2000	-.0363	.0783	-.0849	.1237
	5000	-.0328	.0216	-.0622	.0686
	20000	-.0303	.0227	-.0456	.0582

Table 3: Simulation results II: States reconstruction, transition matrix and duration distributions

In order to make our formulation more flexible, we employ a Bayesian model selection procedure proposed by Congdon (2006) which produces posterior model probabilities and allows one to calculate Bayes factor estimates. The procedure uses a Monte Carlo approximation based on independent MCMC sampling of two or more different models. Here, we present the result for the comparison between two models which assume different number of states for the chain:

- M_1 in which we assume 5 states;
- M_2 in which 6 states are assumed.

Table 4 shows the logarithm of the Bayes Factor, $log(BF_{12})$, of M_1 versus M_2 for all the sample sizes analyzed.

T	300	500	2000	5000	20000
$log(BF_{12})$	2.01	2.37	4.29	5.87	7.01

Table 4: Logarithm of the Bayes Factor, M_1 versus M_2

For all the sample sizes, the $log(BF_{12})$ are in favor of M_1, which assumes the correct number of chain states, with strong or very strong evidence. The value of the Bayes factor increases with the sample size, meaning that with more information the procedure's ability to chose the correct model increases.

3.4 Comparison with the EM method

In this section, we compare the formulation presented in this paper with the one introduced in Paper A.

For what concerns the ability to reconstruct the hidden dynamics, the model presented in Paper A performs better for short observation sequences, while for medium and long sequences, the Bayesian setup provides more reliable estimates with the same variability. Regarding the bias in the duration distributions, the two formulation show different patterns. The MCMC method with NKE shows estimates with less variability but higher bias in the estimated expected value. The EM formulation in Paper A, instead, provides estimates more accurate but with higher variability. The estimation of the transition matrix for small samples constitutes one of the strength of

the MCMC formulation. As matter of fact, the value of the goodness of fit measure for $T = 300, 500$ is about one half of the one in the EM formulation, and the SE are uniformly lower for all the sample sizes. For longer sequences, instead, the EM based method offers a more accurate estimates of the transition matrix. Finally, the two formulation performs both well in estimating the initial state distribution, with the MCMC method which increases considerably the precision using the kernel estimator.

4 Conclusions

A MCMC method for estimating explicit duration HMM was presented. The formulation is flexible, provides accurate estimates of the parameters and takes into account the presence of sparsity. Specifically, by introducing a kernel estimator for discrete data in the estimation procedure we obtained a considerable improvement in the precision and a reduction in the variability of the estimates, in particular for the transition probabilities. We improved the flexibility of our formulation by adopting a Bayesian model selection procedure which allows one to avoid a direct specification of the number of states of the hidden chain.

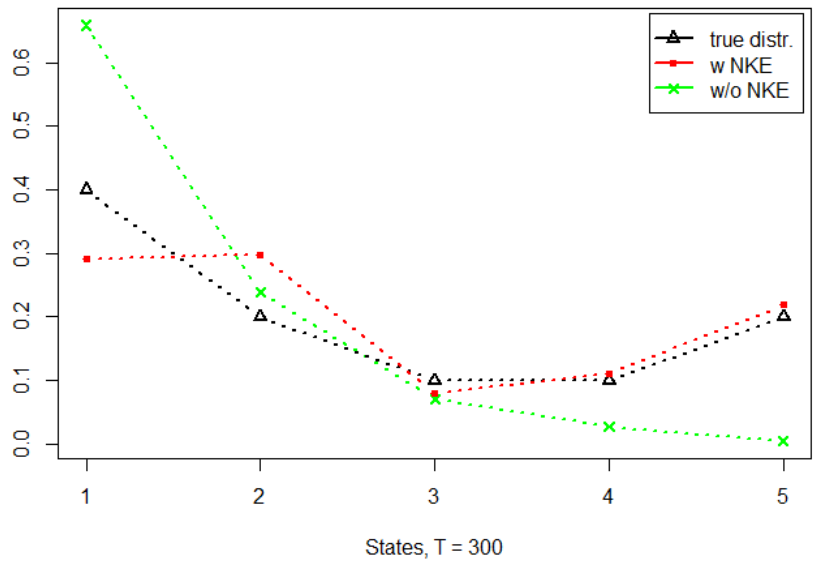
The estimation performances of the formulation proposed were investigated by means of an extensive Monte Carlo study. The results showed that, even for small sample sizes, our model exhibited an excellent ability to reconstruct the hidden dynamics and to estimate accurately all the model parameters.

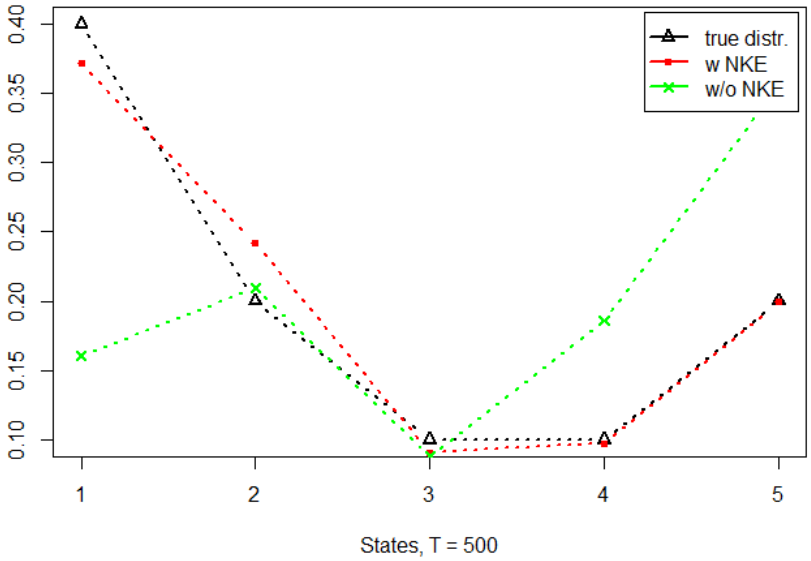
4.1 Appendix

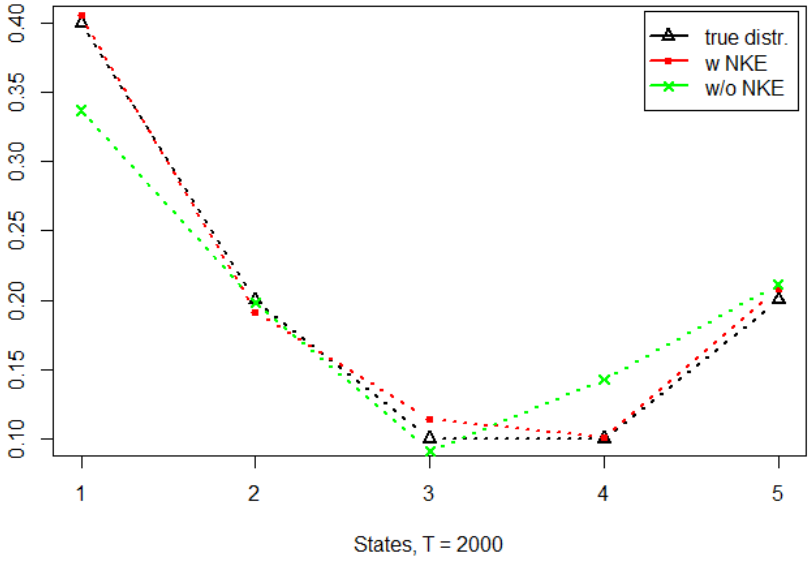
Additional results

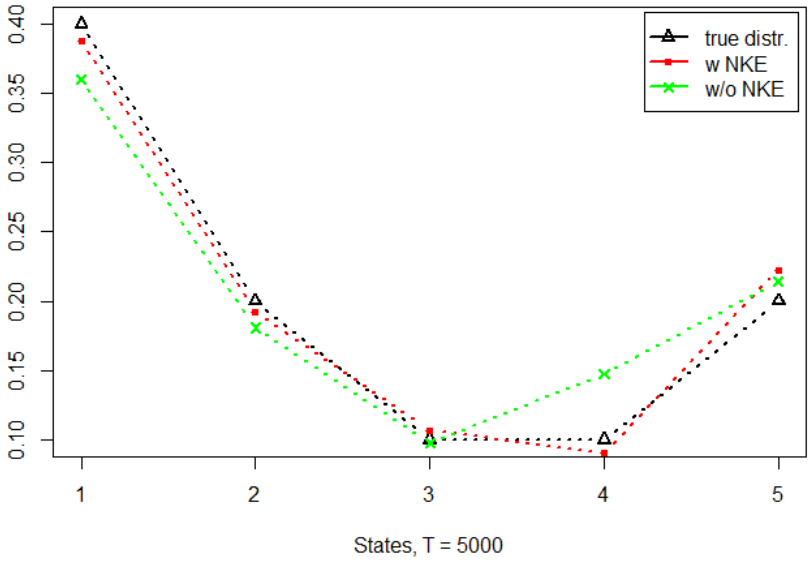
Here, we present some additional simulation results. In particular, Figure 1 shows the estimated initial distribution with and without the NKE for all the sample size simulated. In Figure 2, the density of the percentage of correctly reconstructed states are reported, and in Figure 3 the Monte Carlo distribution of the goodness of fit measure used for the transition matrix are plotted.

Figure 1: Initial distribution




true distr.
w NKE
w/o NKE
0.10
0.15
0.20
0.25
0.30
0.35
0.40
1
2
3
4
5
States, T = 500


true distr.
w NKE
w/o NKE
0.10
0.15
0.20
0.25
0.30
0.35
0.40
1
2
3
4
5
States, T = 2000


true distr.
w NKE
w/o NKE
0.10
0.15
0.20
0.25
0.30
0.35
0.40
1
2
3
4
5
States, T = 5000

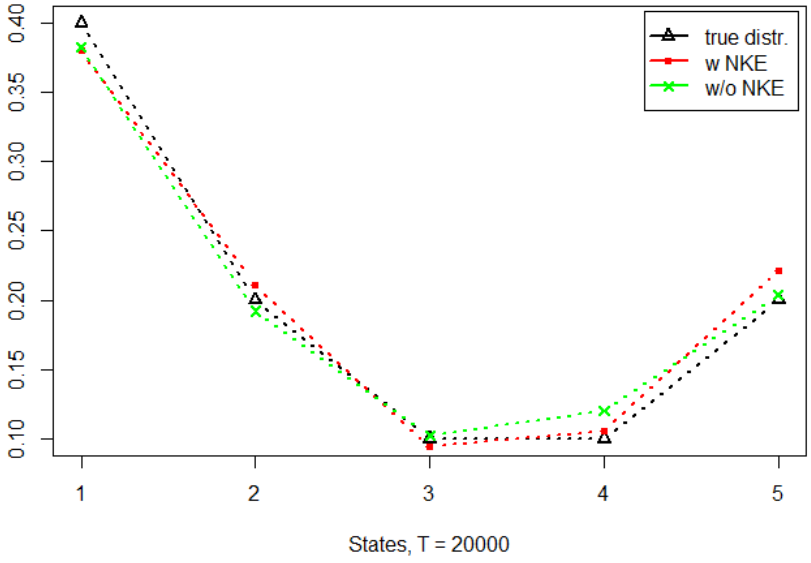

true distr.
w NKE
w/o NKE
0.10
0.15
0.20
0.25
0.30
0.35
0.40
1
2
3
4
5
States, T = 20000

Figure 2: Viterbi estimates, percentage of states correctly reconstructed

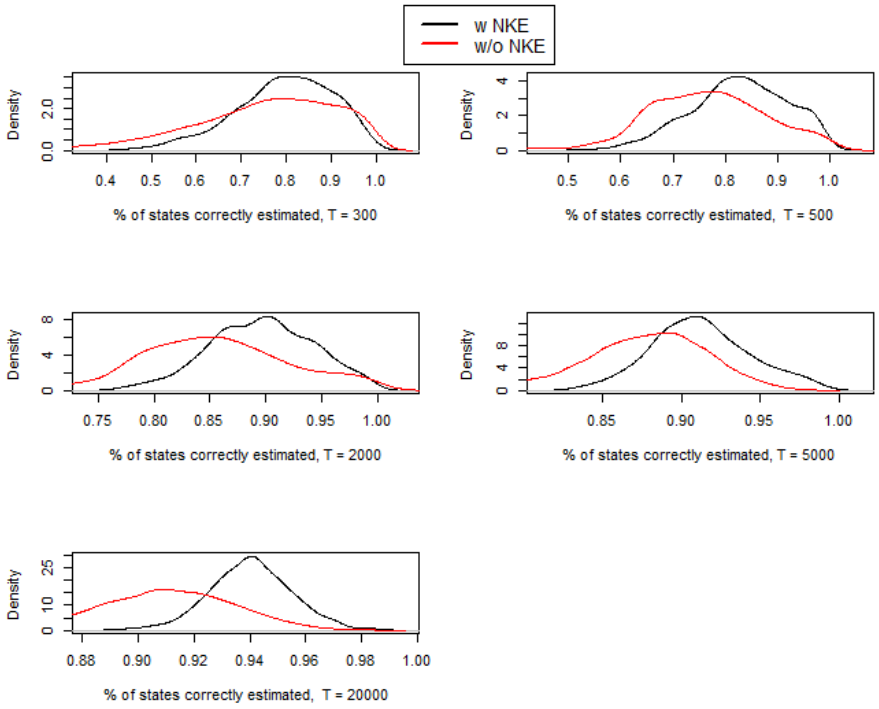
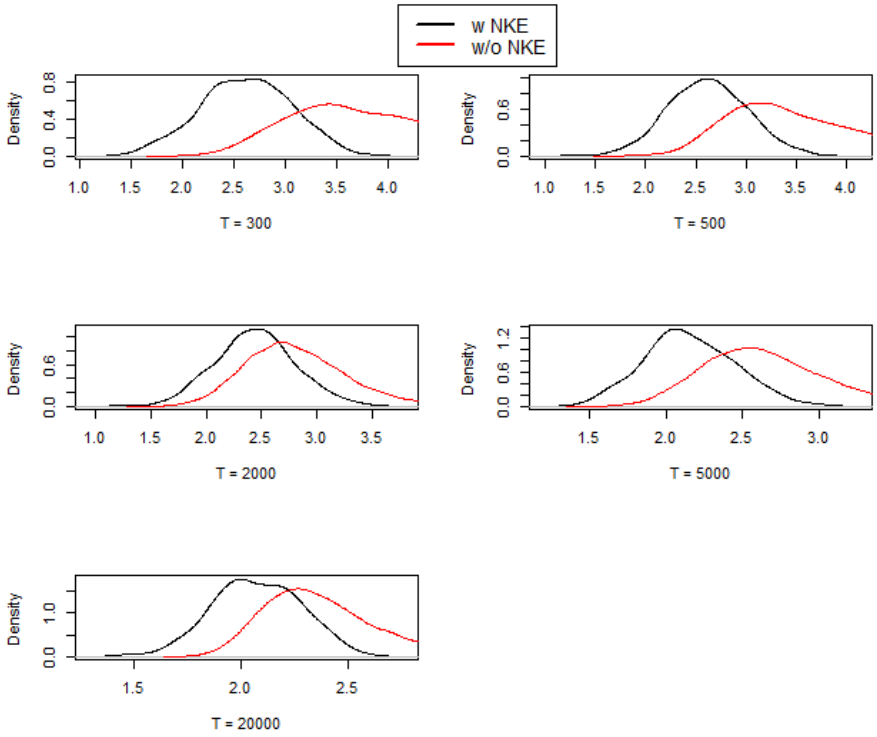


Figure 3: Transition matrix, sum of the absolute difference between the true value and the estimated one



References

Aitchison, J. and C. G. Aitken, 1976: Multivariate binary discrimination by the kernel method. *Biometrika*, **63**, 413–420.

Barbu, V. and N. Limnios, 2006: Maximum likelihood estimation for hidden semi-markov models. *Comptes Rendus Mathematique*, **342**, 201–205.

Barbu, V. S. and N. Limnios, 2009: *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis*, vol. 191, Springer Science & Business Media.

Bickel, P. J., Y. Ritov, T. Ryden, et al., 1998: Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. *The Annals of Statistics*, **26**, 1614–1635.

Bietti, A., F. Bach, and A. Cont, 2015: An online em algorithm in hidden (semi-) markov models for audio segmentation and clustering, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Bishop, Y., S. Fienberg, and P. Holland, 1975: Discrete multivariate analysis: Theory and practice. 1975.

Bulla, J. and I. Bulla, 2006: Stylized facts of financial time series and hidden semi-markov models. *Computational Statistics & Data Analysis*, **51**, 2192–2209.

Cappe, O., E. Moulines, and T. Ryden, 2007: *Inference in hidden Markov models*, New York, Springer.

Congdon, P., 2006: Bayesian model choice based on monte carlo estimates of posterior model probabilities. *Computational statistics & data analysis*, **50**, 346–357.

Djuric, P. M. and J.-H. Chun, 2002: An mcmc sampling approach to estimation of nonstationary hidden markov models. *Signal Processing, IEEE Transactions on*, **50**, 1113–1123.

Douc, R., C. Matias, et al., 2001: Asymptotics of the maximum likelihood estimator for general hidden markov models. *Bernoulli*, **7**, 381–420.

Ephraim, Y. and N. Merhav, 2002: Hidden markov processes. *Information Theory, IEEE Transactions on*, **48**, 1518–1569.

Ferguson, J. D., 1980: Variable duration models for speech, *Proceedings of the Symposium on the Application of HMMs to Text and Speech*.

Fienberg, S. E. and P. W. Holland, 1973: Simultaneous estimation of multinomial cell probabilities. *Journal of the American Statistical Association*, **68**, 683–691.

Guerlin, C., J. Bernu, S. Deleglise, C. Sayrin, S. Gleyzes, S. Kuhr, M. Brune, J.-M. Raimond, and S. Haroche, 2007: Progressive field-state collapse and quantum non-demolition photon counting. *Nature*, **448**, 889–893.

Hall, P. and D. Titterington, 1987: On smoothing sparse multinomial data. *Australian Journal of Statistics*, **29**, 19–37.

Haussler, D. K. D. and M. G. R. F. H. Eeckman, 1996: A generalized hidden markov model for the recognition of human genes in dna, *Proc. Int. Conf. on Intelligent Systems for Molecular Biology, St. Louis*.

Ito, H., S.-I. Amari, and K. Kobayashi, 1992: Identifiability of hidden markov information sources and their minimum degrees of freedom. *Information Theory, IEEE Transactions on*, **38**, 324–333.

Jensen, J. L. and N. V. Petersen, 1999: Asymptotic normality of the maximum likelihood estimator in state space models. *Annals of Statistics*, 514–535.

Johnson, M. J. and A. S. Willsky, 2013: Bayesian nonparametric hidden semi-markov models. *The Journal of Machine Learning Research*, **14**, 673–701.

Kundu, A., Y. He, and M.-Y. Chen, 1997: Efficient utilization of variable duration information in hmm based hwr systems, *Image Processing, 1997. Proceedings., International Conference on*, vol. 3, IEEE.

Leroux, B. G., 1992: Maximum-likelihood estimation for hidden markov models. *Stochastic processes and their applications*, **40**, 127–143.

Marron, J. S. and M. P. Wand, 1992: Exact mean integrated squared error. *The Annals of Statistics*, 712–736.

McFarland, J. M., T. T. Hahn, and M. R. Mehta, 2011: Explicit-duration hidden markov model inference of up-down states from continuous signals. *PloS one*, **6**, e21606.

Melnyk, I. and A. Banerjee, 2014: A spectral algorithm for inference in hidden semi-markov models. *arXiv preprint arXiv:1407.3422*.

Mitchell, C. D. and L. H. Jamieson, 1993: Modeling duration in a hidden markov model with the exponential family, *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 2, IEEE.

Rabiner, L., 1989: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 257–286.

Rajagopalan, B. and U. Lall, 1995: A kernel estimator for discrete distributions. *Journal of Nonparametric Statistics*, **4**, 409–426.

Rydén, T., 1994: Consistent and asymptotically normal parameter estimates for hidden markov models. *The Annals of Statistics*, 1884–1895.

Rydén, T., 1995: Estimating the order of hidden markov models. *Statistics: A Journal of Theoretical and Applied Statistics*, **26**, 345–354.

Schmidler, S. C., J. S. Liu, and D. L. Brutlag, 2000: Bayesian segmentation of protein secondary structure. *Journal of computational biology*, **7**, 233–248.

Simonoff, J. S., 1983: A penalty function approach to smoothing large sparse contingency tables. *The Annals of Statistics*, 208–218.

Simonoff, J. S., 2012: *Smoothing methods in statistics*, Springer Science & Business Media.

Squire, K. and S. E. Levinson, 2005: Recursive maximum likelihood estimation for hidden semi-markov models, *Machine Learning for Signal Processing, 2005 IEEE Workshop on*, IEEE.

Titterington, D., 1980: A comparative study of kernel-based density estimates for categorical data. *Technometrics*, **22**, 259–268.

Wand, M. P. and M. C. Jones, 1994: *Kernel smoothing*, Crc Press.

Wang, M.-C. and J. Van Ryzin, 1981: A class of smooth estimators for discrete distributions. *Biometrika*, **68**, 301–309.

Yu, S.-Z., 2010: Hidden semi-markov models. *Artificial Intelligence*, **174**, 215–243.

Discrete Parameter Models in Quantum Measurement

Contents

Abstract

We derive an extension of the Cramér-Rao inequality for discrete parameter models in quantum measurement. The extension sets the lower bound on the variance of an estimator and determines a discrete counterpart of the quantum Fisher information. This formulation finds application in experiments in which the parameters can assume only few values: for example, the direction which the magnetic field points to, or an atom whose possible oscillation frequencies belong to a finite set of values. We also provide an illustration concerning a quantum optics problem.

1 Introduction

Discrete parameters models are statistical models in which the parameter space is restricted to an enumerable set of points, that is $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. Statistical inference when the parameter space is reduced to a lattice was first considered by Hammersley (1950) in which the author was mainly concerned with a Normal distribution with known variance and unknown integer mean. In Khan (1973, 1978, 2000, 2003) the general admissibility conditions for the mean estimator proposed by Hammersley (1950) are investigated. Cox and Hinkley (1974) discuss the construction of confidence intervals, LaMotte (2008) treats sufficiency and minimal sufficiency for models in which the parameter space and the sample space are both finite. Teunissen (2007) extends the theory of minimum mean squared error prediction by introducing new classes of predictors based on the principle of equivariance. Recently, Choirat et al. (2012) discuss consistency, asymptotic distribution theory, information inequalities and their relations with efficiency and superefficiency for a general class of m-estimators in a discrete parameters setting.

On the side of applications, Baram (1978); Baram and Sandell Jr (1978b,a) highlight the relevance of discrete parameter models in signal processing by deriving the conditions for consistently selecting among a finite set of stationary Gaussian models and obtaining bounds on the performance of the estimators. Moreover, from an information theory perspective, Poor and Verdu (1995) derive a lower bound on the probability of error in multi-hypothesis testing and Kanaya et al. (1995) studies the asymptotic relation between the posterior entropy and the MAP error probability.

This paper is concerned with discrete parameters models in quantum measurement. Specifically, we derive an extension of the Cramér-Rao bound for such models and we provide an illustration concerning a quantum optics problem. The extension sets the ultimate accuracy of an estimator, and determines a discrete counterpart of the quantum Fisher information. This finds application in many experiments in which the parameters can assume only few different values: for example, the direction which the magnetic field points to, or an atom whose possible oscillation frequencies belong to a finite set of values.

The remainder of the paper is organized as follows. Section 1.1 and 1.2 introduce to quantum measurements. Section 2 presents our main contribution, the extended Cramér-Rao bound. In section 3 an illustration is provided and Section 4 concludes the paper.

1.1 Quantum Measurement

Let $(M, \mathscr{A}, P_0)$ be a *quantum* probability space, where M is the set of quantum measurement outcomes, $\mathscr{A}$ a σ-algebra on M and $P_0 : \mathscr{A} \to \mathbb{R}$ the reference measure on M. The reference measure, P_0, has no physical meaning and can be viewed as the measure-theoretic formulation of Wiseman's *ostensible probability* (Wiseman (1996)).

We restrict our attention to finite dimensional quantum systems. For such systems, the state, ρ_0, can be represented by a non-negative complex matrix of trace 1, the density matrix. Let also O be a quantum measurement operator, that is, a map from $\mathscr{A}$ to the set of bounded operators on the system Hilbert space, $\mathcal{H}, O : \mathscr{A} \to \mathcal{B}(\mathcal{H})$, which is normalized as follows,

$$\int_M dP_0(m)\, O(m)^\dagger O(m) = \mathbb{I} \equiv \mathbb{E}_0[O^\dagger O].$$

Since quantum measurement translates a quantum uncertainty in a classical uncertainty (Gammelmark (2013)), once the measurement operators are defined, a classical statistical inference problem is defined (Barndorff-Nielsen et al. (2003)). This can be done defining a classical probability space $(M, \mathscr{A}, P)$, where $P : \mathscr{A} \to \mathbb{R}$ is a new probability measure assigning probability to possible quantum measurement outcomes,

$$P(B) = \int_B dP_0(m)\, \mathrm{Tr}(O(m)^\dagger O(m) \rho_0). \tag{1}$$

Let $\prod(B) = \int_B dP_0(m)\, O^\dagger O$ be the map from M to $\mathcal{B}(\mathcal{H})$ which represents the *positive operator valued measure* POVM associated with the measurement operator O (Holevo (2011)). Recalling the previous definition of P, we can write $P(B) = \mathrm{Tr}[\prod(B)\rho_0]$. The last step consists in defining a random variable on M as the unnormalized quantum state conditional on outcome m:

$$\tilde{\rho}\,|m = O(m)\rho_0 O(m)^\dagger \tag{2}$$

and the normalized quantum state

$$\rho\,|m = \frac{O(m)\rho_0 O(m)^\dagger}{\mathrm{Tr}(O(m)\rho_0 O(m)^\dagger)}$$

From (2) it can be seen that $\tilde{\rho}\,|m$ is a linear function of ρ_0 and allow one to restate (1) as follows:

$$P(B) = \int_B dP_0(m)\,\mathrm{Tr}(\tilde{\rho}|m) \quad \forall B \in \mathscr{A}$$

Therefore, $\mathrm{Tr}(\tilde{\rho}|m)$ is the Radon-Nikodym derivative of P with respect to P_0:

$$\frac{dP}{dP_0}(m) = \mathrm{Tr}(\tilde{\rho}|m)$$

With this formulation it is possible to express expectation w.r.t. P in terms of P_0, i.e. $\mathbb{E}_p[X] = \mathbb{E}_{p_0}[\mathrm{Tr}(\tilde{\rho})X]$ and, most importantly, $\mathrm{Tr}(\tilde{\rho}|m)$ can be thought of as the probability density for measurement result m.

1.2 Measurement Processes

The theory so far delineated can be extended to time-dependent (repeated) measurements. Let $I = \{1, \ldots, N\}$ be the set of times for N sequential measurements and m an N-dimensional vector representing the time-dependent sequence of outputs. Following the same reasoning as before, the measurement result at time $t \in I$, m_t, is a map $m : I \times M \to \mathbb{R}$ which can be thought of as a random variable on M.

In the repeated measurements framework, it is crucial to define the filtration $\mathscr{F}_t$ of a stochastic process. Loosely speaking, $\mathscr{F}_t$ is the σ-algebra generated by m_s for $s \leq t$ and can be viewed as the history of m up to, and including, t.

The stochastic evolution of a quantum state from time t_0 to t_1 is described by a stochastic operator $O(t_1, t_0)$ with the semi-group property $O(t_1, t_0) = O(t_1, s)O(s, t_0)$ for $t_0 \leq s \leq t_1$. The operator $O(t_1, t_0)$ is a $\mathscr{F}_t$-measurable random variable such that $\mathbb{E}_{P_0}[O(t_1, t_0)^\dagger O(t_1, t_0)|\mathscr{F}_{t_0}] = \mathbb{I}$

The normalization condition is a statement about the Markovian nature of quantum measurements and it also implies that $O_t^\dagger O_t$ is a martingale w.r.t. the filtration $\mathscr{F}_t$ and the reference probability measure P_0, since for $s < t$

$$\mathbb{E}_{P_0}[O(s)^\dagger O(s)|\mathscr{F}_t] = \mathbb{E}_{P_0}[O_t^\dagger O(s,t)^\dagger O(s,t) O_t|\mathscr{F}_t] = O_t^\dagger O_t$$

Moreover, as in the case of a single measurement, $\mathrm{Tr}[\tilde{\rho}_T] = \mathrm{Tr}[O_T \rho_0 O_T^\dagger]$ is the Radon-Nikodym derivative of P w.r.t. P_0. The martingale property of O_t implies that the process $L_t = \mathrm{Tr}[\tilde{\rho}_t] = \mathrm{Tr}[O_t \rho_0 O_t^\dagger]$ is also a martingale and represents a well-defined generalization of the likelihood for the observation sequence (Gammelmark and Mølmer (2013)).

Finally, let ρ_0 be the state at time $t_0 = 0$, then,

$$\tilde{\rho}_t|m = O_t \rho_0 O_t^\dagger$$

and

$$\rho_t|m = \frac{O_t \rho_0 O_t^\dagger}{\mathrm{Tr}(O_t \rho_0 O_t^\dagger)}$$

represent the normalized and un-normalized conditional quantum state in repeated measurement. In the next section we shall derive our main result, the extended Cramér-Rao inequality for discrete parameter models in quantum measurement.

2 Discrete parameter models and quantum measurement

In this section we develop an extension of the Cramér-Rao inequality for discrete parameter models in quantum measurement. Let us remind that in our set-up the parameter space is an enumerable set of values $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. Particular attention must be paid in treating the derivatives of the likelihood function with respect to the parameter, since the likelihood is no longer a continuous function w.r.t. the parameter values. Moreover, the restriction imposed on the parameter space can make differentiation inadmissible since $\theta + d\theta$ may not belong to the specified parameter set Θ (Hammersley (1950)). For this reason, instead of derivatives we shall consider differences of the likelihood function for different parameter values. In the following we shall present our main result and provide an application for which an analytical solution is possible.

2.1 Extended Cramér-Rao bound

Our main result concerns the lower bound of the variance of an estimator $\hat{\theta}$ in discrete parameter models in quantum measurement.

Theorem: *Let $(\Omega, \mathscr{A}, P)$ be a classical probability space based on the quantum probability space $(\Omega, \mathscr{A}, P_0)$, and $Tr[\tilde{\rho}_T] = Tr[O_T \rho_0 O_T^\dagger]$ the Radon-Nikodym derivative of P w.r.t. P_0 representing the density of the observations. Then, for $\theta \in \Theta$, $\theta \neq \theta_0$ the lower bound on the variance for the estimator $\hat{\theta}$ is*

$$\mathbb{V}[\hat{\theta}] \geq \max_{\theta \neq \theta_0} \frac{(\theta - \theta_0)^2}{\int_\Omega \frac{Tr(\tilde{\rho}_T \,|\, \theta)^2}{Tr(\tilde{\rho}_T \,|\, \theta_0)} \, dP_0 - 1} \tag{3}$$

where $\mathbb{V}[\hat{\theta}]$ is the of $\hat{\theta}$, and

$$\int_\Omega \frac{Tr(\tilde{\rho}_T \,|\, \theta)^2}{Tr(\tilde{\rho}_T \,|\, \theta_0)} \, dP_0 - 1$$

is the discrete counterpart of the quantum Fisher information in discrete parameter models.

Proof. Since $\mathrm{Tr}[\tilde{\rho}_T] = \mathrm{Tr}[O_T \rho_0 O_T^\dagger]$, the Radon-Nikodym derivative of P w.r.t. P_0, represents the *density* of the observations,

$$\int_M \mathrm{Tr}(\tilde{\rho}_T \,|\, \theta) \, dP_0 = \int_M \frac{dP}{dP_0} \, dP_0 = 1 \tag{4}$$

Let $\hat{\theta}$ be the estimator of the parameter θ which governs the dynamics of the quantum system. Its expected value and variance are,

$$\mathbb{E}_p[\hat{\theta}] = a = \theta + b(\theta) = \int_M \hat{\theta} \, dP = \int_M \hat{\theta} \, \mathrm{Tr}(\tilde{\rho}_T \,|\, \theta) \, dP_0$$

$$\mathbb{V}_p[\hat{\theta}] = \int_M (\hat{\theta} - \theta)^2 \, \mathrm{Tr}(\tilde{\rho}_T \,|\, \theta) \, dP_0$$

where $b(\theta) = \mathbb{E}_p[\hat{\theta}] - \theta$ is the bias of $\hat{\theta}$. Let $\theta_0 \in \Theta$ be the true but unknown parameter value and consider any two values $\theta_1, \theta_2 \in \Theta$. From (4) we get:

$$\int_M [\mathrm{Tr}(\tilde{\rho}_T \,|\, \theta_1) - \mathrm{Tr}(\tilde{\rho}_T \,|\, \theta_2)] \, dP_0 = 0. \tag{5}$$

The difference between the two values in (5) can be thought of as the equivalent of the derivatives in the discrete parameter space. Multiplying (5) by $\hat{\theta}$:

$$\int_M \hat{\theta}\,[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1) - \mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]\,dP_0 = (a_1 - a_2) \tag{6}$$

Multiplying (5) by a_2 and subtracting from (6):

$$\begin{aligned}(a_1 - a_2) &= \int_M (\hat{\theta} - a_2)\,[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1) - \mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]\,dP_0 \qquad (7)\\ &= \int_M (\hat{\theta} - m_2)\,[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]^{1/2}]\frac{[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1) - \mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]}{\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]^{1/2}}\,dP_0\end{aligned}$$

Then, applying the Cauchy–Schwarz inequality we get:

$$\mathbb{V}[\hat{\theta}\,|\theta_0 = \theta_2] \geq \frac{(a_1 - a_2)^2}{\displaystyle\int_M \frac{[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1) - \mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]^2}{\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]}\,dP_0} \tag{8}$$

The denominator of the previous expression can be restated as:

$$\frac{[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1) - \mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]^2}{\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]} = \frac{[\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1)]^2}{\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2)]} - 2\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_1) + \mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_2) \tag{9}$$

Integrating (9) over M gives -2 and $+1$, for the second and third term, respectively. Since the reasoning followed so far holds for all values in Θ, we can rewrite (8) as follows:

$$\mathbb{V}[\hat{\theta}] \geq \max_{\theta \neq \theta_0} \frac{(\theta - \theta_0)^2}{\displaystyle\int_M \frac{\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta)^2}{\mathrm{Tr}(\tilde{\rho}_T\,|\,\theta_0)}\,dP_0 - 1} \tag{10}$$

Where we allow θ to vary over the whole parametric space Θ except for θ_0, and we consider unbiased estimator in the numerator. Equation (10) represents the extension of the Cramér-Rao inequality to the case of discrete parameter models in quantum measurement. The formula for the multidimensional vector of parameters can be derived in the same fashion.

3 Application

In this section we provide an illustration of the result obtained in the previous section. In Nielsen and Mølmer (2008) the authors describe a quantum system with different photon-number states, enumerated by n, coupled to the probe field. The measurement modifies the probability distribution over the n-states. Here, $C_n(t) = C_n(0)\exp(-r_n\hat{y} - r_n^2 t/2)$ is the time dependent probability assigned to n which, in our context, can be interpreted as the probability of parameter θ_i that has to be estimated by the probing. To avoid possible confusion with notation, we replace the index n with i, so the previous formula reads:

$$C_{\theta_i}(t) = C_{\theta_i}(0)\exp(-r_i\hat{y} - r_i^2 t/2) \qquad i = 0, 2, \ldots, n \tag{11}$$

where $\Theta = \{\theta_i\}_{i=0,\ldots,n}$, and $n+1$ is the total number of points in the parameter space.

The probing is Quantum Non Demolition (QND), meaning that the probabilities are updated at each time-step without disturbing the system dynamics. Hence, the probabilities at time t are function of $\hat{y}$, the total integrated noisy current, which has different means for different θ_i.

Assuming that the true value $\theta_0 \in \Theta$, the probability distribution for the integrated noisy current is Gaussian, and leads to a close form for the integrand in (10). Indeed, the probability to observe a quantum state for which the integrated current $\hat{y}$ assumes the value y is described by the following probability density:

$$P_{\rho_t} = \sum_{i=0}^{n} \frac{C_{\theta_i}(0)}{\sqrt{2\pi t}} \exp\left(-\frac{(y + r_i t)^2}{2t}\right) \tag{12}$$

where we consider the sum of the joint probabilities for observing each value of θ_i and the noisy current.

We are interested in the reversal of this problem. In particular, we want to calculate the integral in (10) which, basically, compares two different probabilities:

- the numerator, which corresponds to the square of the probability for observing the quantum state $\tilde{\rho}_T$ given that the driving parameter is θ_i, with $i = 1, \ldots, n$;
- the denominator, which is the probability to observe $\tilde{\rho}_T$ given the true parameter value θ_0.

Thus, in (12) instead of consider all the addends, we pick out only one summand at time. This is due to the fact that we already have the data and we want to know which θ has generated them. Let us restate the integral in (10) in a more tractable form:

$$\int_M \frac{\mathrm{Tr}(\tilde{\rho}_T \,|\, \theta_i)^2}{\mathrm{Tr}(\tilde{\rho}_T \,|\, \theta_0)}\, dP_0 - 1 = \int_{\mathbb{R}} \frac{\left[\frac{C_{\theta_i}(0)}{\sqrt{2\pi t}} \exp\left(-\frac{(y+r_i t)^2}{2t}\right)\right]^2}{\frac{C_{\theta_0}(0)}{\sqrt{2\pi t}} \exp\left(-\frac{(y+r_0 t)^2}{2t}\right)}\, dy - 1 \tag{13}$$

It is easy to see that both the numerator and the denominator in (13) are proportional to a Gaussian distribution with variance t and mean $-r_i t$ or $-r_0 t$, and the latter is exactly the object of our inferential problem. Now, considering the integrand in the right hand side of (13) and taking the square, we get:

$$\frac{\frac{C_{\theta_i}(0)^2}{2\pi t} \exp\left(-\frac{(y+r_i t)^2}{t}\right)}{\frac{C_{\theta_0}(0)}{\sqrt{2\pi t}} \exp\left(-\frac{(y+r_0 t)^2}{2t}\right)} = \frac{C_{\theta_i}(0)^2}{C_{\theta_0}(0)\sqrt{2\pi t}} \exp\left[-\frac{y^2}{2t} - \frac{t^2}{2t}(2r_i^2 - r_0^2) - \frac{2yt}{2t}(2r_i - r_0)\right]$$

completing the square inside the brackets with $\pm t^2(2r_i - r_0)^2$

$$\frac{C_{\theta_i}(0)^2}{C_{\theta_0}(0)} \exp\left[-\frac{t^2}{2t}(2r_i^2 - r_0^2)\right] \exp\left[\frac{t^2}{2t}(2r_i - r_0)^2\right] \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{1}{2t}(y + t(2r_i - r_0))^2\right]$$

integrating the previous equation over the real line,

$$\frac{C_{\theta_i}(0)^2}{C_{\theta_0}(0)} \exp\left[-\frac{t^2}{2t}(2r_i^2 - r_0^2)\right] \exp\left[\frac{t^2}{2t}(2r_i - r_0)^2\right]$$

since $\int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{1}{2t}(y+t(2r_i - r_0))^2}}{\sqrt{2\pi t}}\, dy = 1$. Finally, rearranging terms, the denominator in (10) reads:

$$\frac{C_{\theta_i}(0)^2}{C_{\theta_0}(0)} \exp\left[t(r_i - r_0)^2\right] - 1 \tag{14}$$

We now need to define an (unbiased) estimator for the parameter. In this regard, it is reasonable to assume that the differences in the mean of the integrated signal are of the form $r_i = r_0 + \alpha$, where α is a non zero integer and r_0 is the true but unknown value. Using the previous notation, that is

$\hat{\theta}$ for the estimator and θ_i for the parametr values, we get $\theta_i = \theta_0 + \alpha$. Then (10) reads:

$$\mathbb{V}[\hat{\theta}] \geq \max_{\theta_i \neq \theta_0} \frac{(\theta_i - \theta_0)^2}{\frac{C_{\theta_i}(0)^2}{C_{\theta_0}(0)} \exp\left[t(\theta_i - \theta_0)^2\right] - 1} = \max_{\alpha \neq 0} \frac{C_{\theta_0}(0)\, \alpha^2}{C_{\theta_i}(0)^2 \exp\left[t\alpha^2\right] - 1} \tag{15}$$

Since the previous expression reaches its maximum when α tends to 0:

$$\lim_{\alpha \to 0} \frac{C_{\theta_0}(0)\, \alpha^2}{C_{\theta_i}(0)^2 \exp\left[t\alpha^2\right] - 1} = \frac{C_{\theta_0}(0)\, 2\alpha}{C_{\theta_i}(0)^2\, t\, 2\alpha \exp\left[t\alpha^2\right]} = \frac{C_{\theta_0}(0)}{C_{\theta_i}(0)^2 t} \tag{16}$$

The last term in (16) represents the lower bound for the variance of the mean estimator $\hat{\theta}$ which is linearly dependent on t.

4 Conclusions

We have derived an extension of the Cramér-Rao bound for quantum discrete parameter models, that is, models in which the parameter space is restricted to a finite set of points. We have proved that the extension sets the lower bound on the variance of an estimator, and determines a discrete counterpart of the quantum Fisher information. We have also provided an illustration concerning a quantum optics problem.

References

Baram, Y., 1978: A sufficient condition for consistent discrimination between stationary gaussian models. *Automatic Control, IEEE Transactions on*, **23**, 958–960.

Baram, Y. and N. R. Sandell Jr, 1978a: Consistent estimation on finite parameter sets with application to linear systems identification. *Automatic Control, IEEE Transactions on*, **23**, 451–454.

Baram, Y. and N. R. Sandell Jr, 1978b: An information theoretic approach to dynamical systems modeling and identification. *Automatic Control, IEEE Transactions on*, **23**, 61–66.

Barndorff-Nielsen, O. E., R. D. Gill, and P. E. Jupp, 2003: On quantum statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 775–804.

Choirat, C., R. Seri, et al., 2012: Estimation in discrete parameter models. *Statistical Science*, **27**, 278–293.

Cox, D. and D. Hinkley, 1974: Theoretical statistics chapman and hall, london. *See Also*.

Gammelmark, S., 2013: Efficient parametric inference, estimation and simulation of open quantum systems, Ph.D. thesis, Aarhus UniversitetAarhus University, Science and TechnologyScience and Technology, Institut for Fysik og AstronomiDepartment of Physics and Astronomy.

Gammelmark, S. and K. Mølmer, 2013: Bayesian parameter inference from continuously monitored quantum systems. *Physical Review A*, **87**, 032115.

Hammersley, J., 1950: On estimating restricted parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, **12**, 192–240.

Holevo, A. S., 2011: *Probabilistic and statistical aspects of quantum theory*, vol. 1, Springer Science & Business Media.

Kanaya, F. et al., 1995: The asymptotics of posterior entropy and error probability for bayesian estimation. *Information Theory, IEEE Transactions on*, **41**, 1988–1992.

Khan, R. A., 1973: On some properties of hammersley's estimator of an integer mean. *The Annals of Statistics*, 756–762.

Khan, R. A., 1978: A note on the admissibility of hammersley's estimator of an integer mean. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 113–119.

Khan, R. A., 2000: A note on hammersley's estimator of an integer mean. *Journal of Statistical Planning and Inference*, **88**, 37–45.

Khan, R. A., 2003: A note on hammersley's inequality for estimating the normal integer mean. *International Journal of Mathematics and Mathematical Sciences*, **2003**, 2147–2156.

LaMotte, L. R., 2008: Sufficiency in finite parameter and sample spaces. *The American Statistician*, **62**, 211–215.

Nielsen, A. E. and K. Mølmer, 2008: Stochastic master equation for a probed system in a cavity. *Physical Review A*, **77**, 052111.

Poor, V. H. and S. Verdu, 1995: A lower bound on the probability of error in multihypothesis testing. *Information Theory, IEEE Transactions on*, **41**, 1992–1994.

Teunissen, P., 2007: Best prediction in linear models with mixed integer/real unknowns: theory and application. *Journal of Geodesy*, **81**, 759–780.

Wiseman, H., 1996: Quantum trajectories and quantum measurement theory. *Quantum and Semiclassical Optics: Journal of the European Optical Society Part B*, **8**, 205.