Computational Statistics and Data Analysis 158 (2021) 107183

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

# Explicit-duration Hidden Markov Models for quantum state estimation

Alessandra Luati, Marco Novelli *

*Department of Statistical Sciences, University of Bologna, Italy*

ARTICLE INFO

*Article history:*
Received 19 December 2019
Received in revised form 15 January 2021
Accepted 15 January 2021
Available online 10 February 2021

*Keywords:*
Hidden Markov Models
Forward–backward algorithm
Quantum statistics
Kernel estimation
Viterbi algorithm

ABSTRACT

An explicit-duration Hidden Markov Model with a nonparametric kernel estimator of the state duration distribution is specified. The motivation comes from the physical problem of extracting the maximum information from an open quantum system subject to an external perturbation, which induces a change in the dynamics of the system. A nonparametric kernel estimator for discrete data is introduced, which is consistent and improves the estimates accuracy in presence of sparse data. To reconstruct the hidden dynamics, a Viterbi algorithm is used, which is robust against the underflow problem. Finite sample properties are investigated through an extensive Monte Carlo study showing that our formulation outperforms the original one both in small and in large samples.



## 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, such as 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 or generalized measurement, which avoids the so-called wave-function collapse by extending the measurement process to an auxiliary meter-system and then performing the measurement 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 new quantum technologies, namely 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 perturbation which changes the oscillation frequency. The relevance of the quantum experiment is connected with the possibility to track and control quantum systems which undergo a complex evolution, with possible applications to quantum feedback control, high-precision measurements and quantum computing (Ramakrishna and Rabitz, 1996; Wiseman and Milburn, 2009; Dong and Petersen, 2010; Barreiro et al., 2011; Dunjko and Briegel, 2018). 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 Hidden Markov Models (HMM), used for example to model quantum systems under no perturbation (Gammelmark et al.,

* Corresponding author.
*E-mail address:* marco.novelli4@unibo.it (M. Novelli).

https://doi.org/10.1016/j.csda.2021.107183
0167-9473/

2014), are based on the Markovian short-memory assumption and cannot capture the complex dynamics induced by the external force.

We propose an EDHMM associated with a nonparametric kernel estimator of the state duration distributions. The motivation for kernel estimation comes from the following considerations. 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, thus making difficult to obtain reliable information unless very long observation sequences are considered. This high dimensional setting generally results in sparse multinomial table where the cell proportion (maximum likelihood) estimator is neither consistent nor close to the true discrete duration distribution. A smoothed estimator is thus derived, based on a discrete kernel function, which 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).

There are two main contributions in this article. First, to the best of our knowledge, this is the first attempt to model the dynamics of an open quantum system by means of the EDHMM. Second, we take into account the presence of sparse data (lack of information) by introducing a nonparametric kernel estimator for discrete duration distributions. In addition, in order to reconstruct the hidden dynamics of the system, a Viterbi algorithm which is robust against the underflow problem is used.

The remainder of the paper is organized as follows. Explicit duration HMM are discussed in Section 2. Section 3 addresses the estimation issues, including the kernel estimator along with its properties and the Viterbi algorithm used for state reconstruction. Section 4 introduces and discusses the experimental setup. Section 5 illustrates the results of the Monte Carlo study and Section 6 concludes the paper. Additional results are presented in Appendix.

## 2. Hidden semi-Markov models

Hidden semi-Markov Models (HSMM) are extensions of HMM in which each state of the hidden process can have a different duration (*sojourn*) time. Based on the assumptions on the dependence structure, the general HSMM reduces to specific models, such as, for instance, the explicit-duration HMM, where the duration depends on the next state but not on the current one (Rabiner, 1989; Mitchell and Jamieson, 1993) and the variable-transition HMM, where the state transition is dependent on the state duration (Ramesh and Wilpon, 1992; Krishnamurthy et al., 1991); for a complete review of HSMM see Barbu and Limnios (2008) and Yu (2010).

Due to their flexibility, since the earliest formulation of Ferguson (1980), HSMM have been largely studied and applied in a wide variety of fields. In particular, Barbu and Limnios (2006, 2008) 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 (2015) introduced a spectral algorithm for inference in HSMM and Bietti et al. (2015) presented an incremental EM algorithm for online parameters estimation. 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), audio segmentation and clustering (Bietti et al., 2015), accurate heart beat detection (Pimentel et al., 2015), tool wear monitoring (Zhu and Liu, 2017) and for determining duration and timing of up–down state in neocortical neurons (McFarland et al., 2011).

Prior to the introduction of the HSMM, some basic notions concerning semi-Markov chains and Markov renewal chains are provided, based on Barbu and Limnios (2008). In what follows the term chain is used to denote a discrete-time stochastic process. Let us consider a random system with finite state space $\mathcal{S} = \{1, 2, \ldots, M\}$, then the evolution in time of the system is described by the following chains:

- $R = (R_n)_{n\in\mathbb{N}}$, with state space $\mathcal{S}$, represents the state of the system at the $n$th jump time.
- $J = (J_n)_{n\in\mathbb{N}}$, with state space $\mathbb{N}$, is the $n$th jump time, with $J_0 = 0$ and $0 < J_1 < J_2 < \cdots < J_n < J_{n+1} < \cdots$
- $D = (D_n)_{n\in\mathbb{N}}$ with state space $\mathbb{N}$, where $D_n = J_n - J_{n-1}$ for all $n > 0$ and $D_0 = 0$, namely, $D_n$ represents the sojourn time in state $R_{n-1}$, before the $n$th jump.

Note that the chain $(R, J) = (R_n, J_n)_{n\in\mathbb{N}}$ constitutes a Markov renewal chain (MRC) if, for all $n \in \mathbb{N}$, $i, j \in \mathcal{S}$ and $d \in \mathbb{N}$, the condition

$$P(R_{n+1} = j, J_{n+1} - J_n = d \mid R_0, \ldots R_n; J_0, \ldots J_n) = P(R_{n+1} = j, J_{n+1} - J_n = d \mid R_n), \tag{1}$$

is satisfied almost surely. Moreover, if condition (1) is independent of $n$, then $(R, J)$ is said to be a homogeneous MRC. In what follows we shall consider only homogeneous MRC.

Let us define the matrix-valued function $q = q_{ij}(d)$ as the discrete-time semi-Markov kernel where $q_{ij}(d) = P(R_{n+1} = j, D_{n+1} = d \mid R_n = i)$ and the transition matrix of $(R_n)$ defined by

$$a_{ij} = P(R_{n+1} = j \,|R_n = i), i, j \in \mathcal{S}, n \in \mathbb{N},$$

where $a_{ii} = 0$, i.e. the transition to the same state is not allowed. Note that the semi-Markov kernel $q$ verifies the relation

$$q_{ij}(k) = a_{ij}p_{ij}(d),$$

where $p_{ij}(d) = P(D_{n+1} = d \mid R_n = i, R_{n+1} = j)$, $i, j \in \mathcal{S}$ represents the conditional distribution of $D_{n+1}$. Given a MRC $(R, J)$, the chain $X = (X_t)_{t\in\mathbb{N}}$ is the semi-Markov chain (SMC) associated with it if

$$Z_t = R_{N(t)}, t \in \mathbb{N},$$

where $N(t) = \max\{n \in \mathbb{N} \mid J_n \leq t\}$ is the discrete-time counting process of the number of jumps in $[1, t]$. Finally, let $\Pi = \{\pi_1, \ldots, \pi_M\}$ denote the vector collecting the initial distribution of the SMC $X$,

$$\pi_i = P(X_0 = i), i \in \mathcal{S}.$$

In the HSMM formulation, the unobserved process is modeled through a SMC and the observed process is conditionally independent on the value of the semi-Markov chain. Consider now the double chain $(X_t, Y_t)_{t\in\mathbb{N}}$, where $(X_t)$ is $\mathcal{S}$-valued semi-Markov chain and $(Y_t)$ is the observed process taking values on $\mathcal{V} = \{v_1, \ldots, v_K\}$, where the conditional distribution of chain $Y$ is given by

$$b_j(v_k) = P(Y_t = v_k \mid X_t = j), (v_k, j) \in \mathcal{V} \times \mathcal{S},$$

called emission probabilities. The pair $(X_t, Y_t)_{t\in\mathbb{N}}$ is a HSMM with finite state space $\mathcal{S} \times \mathcal{V}$. If, in addition, the conditional distribution of sojourn times $p_{ij}(d)$ depends only on the next visited state, i.e. $p_{ij}(d) = p_j(d)$, so that the kernel reads $q_{ij}(k) = a_{ij}p_j(d)$, we have the so-called EDHMM formulation, which is the focus of our work.

Finally, we denote the set of model parameters as

$$\lambda = \{q_{ij}(d), b_j(v_k), \pi_j\},$$

for all $i, j \in \mathcal{S}$, $d \in \mathbb{N}$ and $v_k \in \mathcal{V}$.

## 3. Estimation

The parameter estimation of HMM is based on the well-known Baum–Welch or Forward–Backward (FB) algorithm, developed by Baum et al. (1970) and primarily based on the forward and backward variables, $\alpha$ and $\beta$, respectively. One of the most severe issues in practical implementations is the numerical underflow, caused by the exponential decay of the joint probability of the observations, as the sample size increases. To overcome this problem, 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 from the underflow problem.

### 3.1. Modified forward–backward recursion

This section briefly sketches the modified FB algorithm in Yu and Kobayashi (2006). Given a sample path of observations, namely $y_{0:T} = (y_0, \ldots, y_T)$, the goal is to estimate the characteristics of the underlying semi-Markov chain along with the conditional distribution of $Y$. Note that, in practical applications, the conditional sojourn time distributions have finite support $\mathcal{D}$; we shall denote as $\bar{D}$ the maximum time duration in any given state. Let $\alpha_{t|k}(i, d)$ be the *forward variable* defined by

$$\alpha_{t|k}(i, d) = P(X_t = i, \tau_t = d \mid y_{0:k}, \lambda),$$

where $y_{0:k}$ denotes the observation sequence from time 0 to $k$, and $k = t - 1, t$ or $T$, the "predicted", "filtered" and "smoothed" probability, respectively. Furthermore, $\tau_t = J_{N(t)} - t$ denotes the forward recurrence time, also called the residual or excess lifetime. Let $b_i^\star(Y_t)$ be the modified emission probabilities,

$$b_i^\star(y_t) = \frac{\alpha_{t|t}(i, d)}{\alpha_{t|t-1}(i, d)} = \frac{b_i(y_t)}{P(y_t \mid y_{0:t-1})}, \tag{2}$$

where $P(y_t \mid y_{1:t-1})$ is the one-step ahead prediction of the observation that can be obtained by

$$P(y_t \mid y_{0:t-1}) = \sum_{i=1}^{M}\sum_{d=1}^{\bar{D}} \alpha_{t|t-1}(i, d)b_i(y_t) = \sum_{i=1}^{M} \gamma_{t|t-1}(i)b_i(y_t),$$

and $\gamma_{t|k}(i) = \sum_d \alpha_{t|k}(i, d)$ is the marginal probability distribution of $X_t$. Then, the likelihood function for the entire sequence of observations is obtained as follows:

$$\mathcal{L}(\lambda) = P(y_{0:T}|\lambda) = \prod_{t=1}^{T} P(y_t \mid y_{0:t-1}).$$

For convenience, in the forward recursion, we introduce two additional variables:

$$\mathcal{E}_t(i) = P(X_t = i, \tau_t = 1 \mid y_{0:t}, \lambda) = \alpha_{t|t-1}(i, 1) b_i^\star(y_t),$$

and

$$\mathcal{S}_t(i) = P(\tau_t = 1, X_{t+1} = i \mid y_{0:t}, \lambda) = \sum_j \mathcal{E}_t(j) a_{ji}.$$

To obtain the smoothed estimates, let $\beta_t(i, d)$ be the *backward variable* defined as

$$\beta_t(i, d) = \frac{P(X_t = i, \tau_t = d \mid y_{0:T}, \lambda)}{P(X_t = i, \tau_t = d \mid y_{0:t-1}, \lambda)} = \frac{P(y_{t:T} \mid X_t = i, \tau_t = d \mid y_{0:T}, \lambda)}{P(y_{t:T} \mid y_{0:t-1}, \lambda)}.$$

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

$$\mathcal{E}_t^\star(i) = \frac{P(y_{t:T} \mid X_t = i, \tau_{t-1} = 1 \mid y_{0:T}, \lambda)}{P(y_{t:T} \mid y_{0:t-1}, \lambda)} = \sum_d p_i(d) \beta_t(i, d),$$

and

$$\mathcal{S}_t^\star(i) = \frac{P(y_{t:T} \mid X_{t-1} = i, \tau_{t-1} = 1 \mid y_{0:T}, \lambda)}{P(y_{t:T} \mid y_{0:t-1}, \lambda)} = \sum_j a_{ij} \mathcal{E}_t^\star(j).$$

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

(i) the forward recursion becomes

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

with initial condition $\alpha_0(i, d) = \pi_i p_i(d)$;

(ii) the backward recursion is

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

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:

$$\mathcal{D}_{t|T}(i, d) = P(\tau_{t-1} = 1, X_t = i, \tau_t = d \mid y_{0:T}, \lambda) = \mathcal{S}_{t-1}(i) p_i(d) \beta_t(i, d), \tag{3}$$

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

$$\mathcal{T}_{t|T}(i, j) = P(X_{t-1} = i, \tau_{t-1} = 1, X_t = j \mid y_{0:T}, \lambda) = \mathcal{E}_{t-1}(i) a_{ij} \mathcal{E}_{t-1}^\star(j). \tag{4}$$

*3.1.1. Parameter re-estimation*

Generally, in practical application the parameters are unknown, hence an iterative procedure which maximizes the probability of the given observation sequence, $P(y_{0:T}|\lambda)$, 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 Eqs. (3) and (4) are used to re-estimate the parameters as follows:

$$\hat{a}_{ij} = \sum_{t=1}^{T} \frac{\mathcal{T}_{t|T}(i, j)}{K_a},$$

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

$$\hat{p}_i(d) = \sum_{t=1}^{T} \frac{\mathcal{D}_{t|T}(i, d)}{K_p}, \tag{5}$$

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

$$\hat{\pi}_i = \frac{\gamma_{0|T}(i)}{K_\pi},$$

where $K_\pi = \sum_i \gamma_{0|T}(i)$ and $\sum_i \hat{\pi}_i = 1$. The term $\gamma_{0|T}(i)$ can be obtained with the following backward recursion:

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

The emission probabilities can be obtained by

$$\hat{b}_i(v_k) = \sum_{t=0}^{T} \frac{\gamma_{t|T}(i)\delta(Y_t, v_k)}{K_b},$$

with $K_b = \sum_{v_k} \sum_{t=0}^{T} \gamma_{t|T}(i)\delta(Y_t, v_k)$ 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.

### 3.2. Estimation of the duration distribution

One of the novelties of this paper is the introduction of a nonparametric 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

$$\tilde{p}_i(d) = \sum_{j=1}^{\bar{D}} K(t_j)\hat{p}_i(j),$$

where $\hat{p}_i(j)$ is the sample relative frequency estimator defined in Eq. (5), $t_j = (d-j)/h$ with $h$ and $d$ positive integers, and $K(t_j)$ is a discrete kernel, i.e. a non negative, symmetric function of $t_j$ 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 occurs in many practical applications. Indeed, the cell proportion estimator (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, 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 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 of the Epanechnikov kernel, which belongs to the class of the kernels generated from Beta distributions and has optimal asymptotic properties for continuous smoothing, in the sense that it minimizes the asymptotic integrated mean square error (AMISE), Wand and Jones (1994).

In practice, we have a quadratic kernel $K(t_j) = at_j^2 + b$, where $t_j$ takes values 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 normalizing constants $a$ and $b$ both in the interior and at the boundaries. These values must be selected in order for the kernel to add up to one, $\sum_j K(t_j) = 1$, and satisfy the properties of symmetry $K(-t_j) = K(t_j)$, positivity, $K(t_j) > 0$ within the bandwidth, and unbiasedness $\sum_j K(t_j)t_j = 0$. In the interior, i.e. for $h+1 < d < \bar{D}-h-1$, the parameters $a$ and $b$ are

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

while, at the boundaries, namely for $1 < d < h+1$ and $d > \bar{D}-h-1$, $a$ and $b$ become

$$a = \frac{-6h^2}{(d-1+h)(d-2+h)(d-3+h)}, \quad b = \frac{3(2-h+h^2-3d+d^2)}{(d-1+h)(d-2+h)(d-3+h)}.$$

While the kernel generally has a limited impact on the estimation, the bandwidth plays a crucial role. 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 $h$ can be found by minimizing the LSCV function

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

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

### 3.3. State reconstruction

This section concerns the Viterbi algorithm for explicit-duration HMM. Viterbi algorithm is a dynamic programming algorithm which finds the optimal hidden state sequence. Some modified versions for HSMM have been proposed, see for example Yu (2010) and Pertsinidou and Limnios (2015). Chen et al. (1995) consider a modified algorithm for continuous density variable-duration HMM, Ramesh and Wilpon (1992) 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. Recently, Pertsinidou and Limnios (2015) develop a new Viterbi algorithm for HSMM that achieves the same complexity of the one for HMM.

Within the modified FB algorithm of Yu and Kobayashi (2006) described in Section 3.1, the Viterbi algorithm for EDHMM can be sketched as follows. Let $\delta_t(j,d)$ be the forward variable for the extended Viterbi algorithm defined by

$$\delta_t(j,d) = \max_{X_{0:t-d}} P(X_{0:t-d}, X_{t-d+1:t} = j, y_{0:t} \mid \lambda), \tag{6}$$

for $0 \le t \le T,\ j \in \mathcal{S},\ d \in \mathcal{D}$. 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, Eq. (6) reads

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

where $b_j^\star(y_{t-d+1:t})$ represents the emission probabilities related to the partial state sequence from $t-d$ to $t$. The algorithm is initialized as follows:

$$\delta_0(j,1) = \pi_j p_j(1) b_j^\star(y_0) \quad \forall j \in \mathcal{S},\ d = 1,$$

$$\delta_0(j,d) = 0 \qquad\qquad \forall j \in \mathcal{S},\ d > 1,$$

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

$$\delta_t(j,1) = \max_{d_1 \in \mathcal{D}, i \in \mathcal{S}\setminus\{j\}} \{\delta_{t-1}(i,d_1) a_{ij} p_j(1)\} b_j^\star(y_t), \qquad d = 1, \tag{7}$$

$$\delta_t(j,d) = \max_{d_1 \in \mathcal{D}, i \in \mathcal{S}\setminus\{j\}} \{\delta_{t-d}(i,d_1) a_{ij} p_j(d)\} b_j^\star(y_{t-d+1:t}), \quad d > 1. \tag{8}$$

Moreover, for $t \le \bar{D}$ and $d = t$ Eq. (8) 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 Eqs. (7) and (8) 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:

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

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 $\tilde{x}_0$ is obtained.

## 4. Quantum state estimation

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 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 system), initialized in a superposition state, are sent through the system $A$ and weakly interact with it. Then, a projection measurement on each system $B$ 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, since the measurement results on system $B$ 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 from one state to another depend on the state duration and the emission probabilities are obtained by the propagation of the wave function at different frequency levels.

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 ancilla $B$, which reveals only partial information about the system $A$.

The model can be sketched as follows. Let us consider a semi-Markov chain with $M$ states $\mathcal{S} = \{\omega_1, \ldots, \omega_M\}$ representing the Rabi frequencies of the system $A$ and two output symbols '0' and '1', representing the measurement results. The quantum system $A$ under study is initialized in a known 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 $|\Psi\rangle$ represents a unit vector in the two dimensional complex Hilbert space $\mathcal{H}_2$, $\{|0_A\rangle, |1_A\rangle\}$ denotes an orthonormal basis for $\mathcal{H}_2$ and $a, b$ are complex numbers satisfying $|a(\omega_i)|^2 + |b(\omega_i)|^2 = 1$.

We are adopting the Dirac bra-ket notation for column vectors $|\Psi\rangle$ (ket) and their conjugate transpose row vectors $\langle\Psi|$ (bra) commonly used for representing physical entities. In the rest of the paper, we will suppress the dependence of $a$ and $b$ on $\omega_i$ 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 outcomes. The initial state of the composite system is thus $|\Psi_A(\omega_i)_{t_0}\rangle \otimes |\Psi_B\rangle$ 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 becomes

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

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

$$P_{t_1}(B = 0 \mid |\Psi_{AB}(\omega_i)\rangle_{t_1}) = |\alpha_{t_1}\sqrt{p}|^2 + |\beta_{t_1}\sqrt{q}|^2, \tag{9}$$

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 $\omega_i$ value which governs the evolution of the system $A$. For more details on quantum probability and quantum statistical inference, the reader is referred to Helstrom (1976), Holevo (1982) and Barndorff-Nielsen et al. (2003).

In the HMM framework, Eq. (9) can be viewed as the emission probability, at time $t_1$, of symbol '0' when the system is in state $\omega_i$, that is $P_{t_1}(B = 0 \mid \omega_i)$. The measurement resulting in the output '0' leaves the system $A$ in the (unnormalized) state

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

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

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

and the (unnormalized) $A$ state is

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

## 5. Monte Carlo study

To validate the finite samples properties of the EDHMM formulation introduced so far, a Monte Carlo study is carried out. Let us remind that the aim is to characterize the behavior of an open quantum system subject to an external force by using only the partial information obtained from a set of ancillary systems. The observed sequence $y_{0:T}$ is a binary sequence obtained by measuring the systems $B$.

We present 3 simulation scenarios in which the hidden chain is composed of $M = 3$, 4 and 5 states, respectively. These correspond to different frequency values in the range $(0, \frac{\pi}{2})$. 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 note that the value $M = 5$ is far beyond the one found in commonly applied models, which usually consider only two or three frequency values.

The truncation value $\bar{D}$, representing the maximum number of consecutive time steps in the same state, is set to 120. The choice of $\bar{D}$ is of crucial relevance: a small value could prevent the model to fully capture the dependence structure of the chain, while a large value can cause a curse of dimensionality that makes the estimation process impractical. From a practical point of view, the choice of the maximum $\bar{D}$ can be done by a grid search over several values of $\bar{D}$. We have fixed a large value of $\bar{D}$ since the simulations have been carried out using an optimized R/C++ code and is not time consuming.

The initial values for the transition probabilities and the initial state distribution of the semi-Markov chain 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 $\omega_i$: this represents a point of strength in the estimation/re-estimation procedure since they do not enter in the estimation algorithm (Rabiner, 1989).

Four observations sequence' lengths are considered: $T = 300, 500, 1000, 5000$. The number of Monte Carlo replications is set to 5000 for each sample size. In the first step of the simulation, the semi-Markov chain representing the time evolution of the Rabi frequencies is generated. The duration of each state is modeled using a zero-truncated Poisson distribution with probability mass function given by

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

where $d \in \mathcal{D}$ and for each $\omega_i$ a different value of $\theta_i$ is chosen. In the case of $M = 5$ states, we set $\theta_i = 33, 41, 25, 53$ and 38 respectively.

A comparison between the EDHMM model with and without the nonparametric kernel estimator (henceforth NKE) is provided. Although the kernel choice has generally a limited impact on the estimation, other results – not shown here for the sake of brevity – highlight that adopting the parabolic kernel described in Section 3.2 leads to an improvement in the estimation accuracy with respect to other functions, like triangular, rectangular and binomial kernels, among others. Moreover, the optimal bandwidth $h \in \{1, 2, \ldots, 20\}$ is selected by minimizing the LSCV function. The transition matrices used to simulate the hidden chains are reported in Appendix. All numerical simulations are performed using an R/C++ code and a Dell Laptop with Processor Intel CoreTM i7-8550U (1.8 GHz) and RAM 16 GB DDR4 2133 MHz.

### 5.1. Simulation results

Tables 1 and 2 summarize the results for the chains with 3 and 4 hidden states, respectively. Specifically, the tables show the Monte Carlo averages and standard errors for: the percentage of correctly reconstructed states using Viterbi algorithm, the sum of the absolute difference between the true and the estimated transition matrix and the percentage of bias in the estimated expected values of the duration distribution.

**Table 1**
Simulation results for a chain with 3 hidden states: Viterbi estimates, transition matrix and duration distributions.

| Parameter | T | w NKE | | w/o NKE | |
|---|---|---|---|---|---|
| | | Mean | MC SE | Mean | MC SE |
| Viterbi | 300 | .928 | .038 | .909 | .046 |
| | 500 | .930 | .028 | .914 | .033 |
| | 1000 | .937 | .028 | .907 | .038 |
| | 5000 | .965 | .008 | .931 | .009 |
| Trans. matrix | 300 | 1.490 | .596 | 1.536 | .623 |
| | 500 | 1.044 | .459 | 1.082 | .478 |
| | 1000 | .750 | .400 | .801 | .384 |
| | 5000 | .297 | .133 | .319 | .144 |
| Bias state 1 | 300 | .013 | .213 | .009 | .229 |
| | 500 | .009 | .123 | −.019 | .144 |
| | 1000 | .000 | .077 | .003 | .097 |
| | 5000 | −.014 | .032 | −.003 | .036 |
| Bias state 2 | 300 | .009 | .208 | −.052 | .237 |
| | 500 | −.025 | .118 | −.015 | .149 |
| | 1000 | −.030 | .061 | −.047 | .071 |
| | 5000 | −.003 | .028 | −.007 | .032 |
| Bias state 3 | 300 | .034 | .199 | −.047 | .232 |
| | 500 | −.015 | .130 | −.031 | .154 |
| | 1000 | −.020 | .083 | −.026 | .101 |
| | 5000 | .007 | .036 | −.017 | .037 |

**Table 2**
Simulation results for a chain with 4 hidden states: Viterbi estimates, transition matrix and duration distributions.

| Parameter | T | w NKE | | w/o NKE | |
|---|---|---|---|---|---|
| | | Mean | MC SE | Mean | MC SE |
| Viterbi | 300 | .840 | .097 | .796 | .106 |
| | 500 | .846 | .078 | .801 | .085 |
| | 1000 | .859 | .059 | .837 | .064 |
| | 5000 | .890 | .024 | .861 | .028 |
| Trans. matrix | 300 | 4.031 | .676 | 4.281 | .702 |
| | 500 | 3.377 | .724 | 3.632 | .739 |
| | 1000 | 2.603 | .686 | 2.792 | .699 |
| | 5000 | 1.083 | .303 | 1.333 | .401 |
| Bias state 1 | 300 | .091 | .467 | .085 | .491 |
| | 500 | .031 | .298 | .046 | .332 |
| | 1000 | .021 | .175 | .056 | .194 |
| | 5000 | −.026 | .071 | .093 | .099 |
| Bias state 2 | 300 | −.013 | .488 | .081 | .543 |
| | 500 | .176 | .439 | .213 | .486 |
| | 1000 | .120 | .300 | .240 | .335 |
| | 5000 | −.019 | .102 | .146 | .142 |
| Bias state 3 | 300 | −.085 | .497 | −.173 | .545 |
| | 500 | .358 | .617 | .473 | .745 |
| | 1000 | .284 | .474 | .434 | .489 |
| | 5000 | −.056 | .103 | .097 | .142 |
| Bias state 4 | 300 | −.008 | .195 | −.061 | .237 |
| | 500 | −.030 | .115 | −.077 | .144 |
| | 1000 | −.018 | .068 | −.041 | .091 |
| | 5000 | −.007 | .026 | −.037 | .055 |

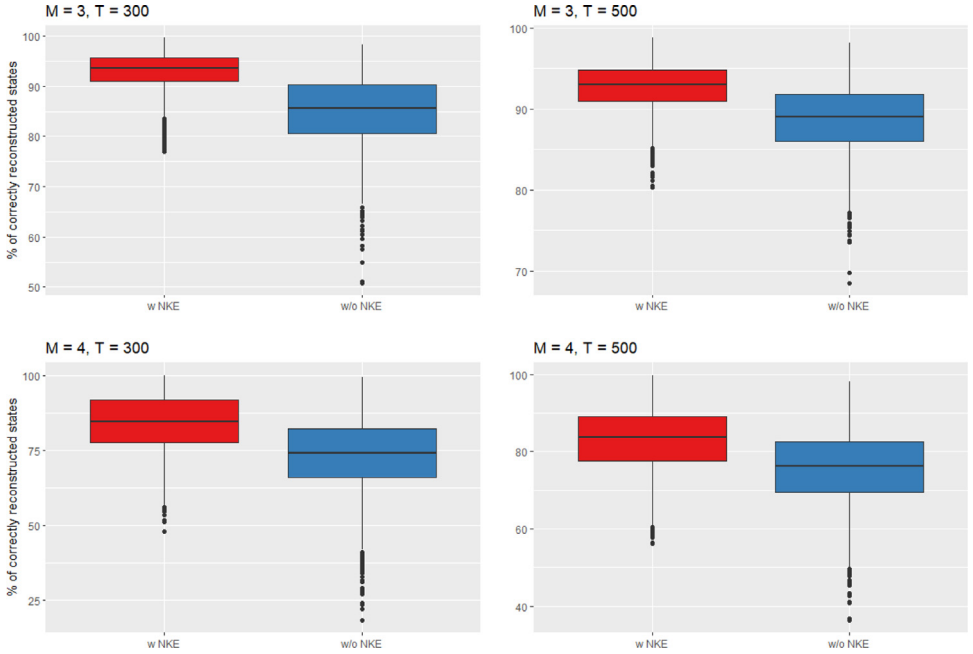


**Fig. 1.** Distribution of the proportion of correctly reconstructed states using Viterbi algorithm, with and without NKE, for $T = 300$ and 500 and $M = 3$ and 4.

The percentage of states correctly reconstructed is uniformly higher using the NKE, which provides an increase of about 2−3% with respect to the standard formulation in the case with 3 states. This is even more evident with 4 states in which the improvement rises up to 4−5%. Besides, adopting the NKE the Monte Carlo standard errors are lower, indicating a higher precision and stability of the estimates. As an example, in Fig. 1 for $T = 300$ and 500 and $M = 3$ and 4 states, the distribution of the proportion of correctly reconstructed states with and without NKE is provided. For what concerns the

**Table 3**
Simulation results for a chain with 5 hidden states: Viterbi estimates, transition matrix and duration distributions.

| Parameter | T | w NKE | | w/o NKE | |
|---|---|---|---|---|---|
| | | Mean | MC SE | Mean | MC SE |
| Viterbi | 300 | .806 | .073 | .743 | .148 |
| | 500 | .818 | .091 | .780 | .101 |
| | 1000 | .827 | .047 | .806 | .054 |
| | 5000 | .857 | .018 | .837 | .021 |
| Trans. matrix | 300 | 5.779 | .801 | 5.990 | .796 |
| | 500 | 5.019 | .870 | 5.274 | .865 |
| | 1000 | 3.540 | .711 | 3.840 | .738 |
| | 5000 | 1.833 | .431 | 1.910 | .448 |
| Bias state 1 | 300 | −.177 | .494 | −.175 | .505 |
| | 500 | .005 | .490 | −.026 | .319 |
| | 1000 | −.008 | .148 | −.013 | .195 |
| | 5000 | −.0301 | .051 | −.001 | .071 |
| Bias state 2 | 300 | −.204 | .484 | −.257 | .465 |
| | 500 | −.030 | .406 | −.064 | .484 |
| | 1000 | −.020 | .252 | −.029 | .284 |
| | 5000 | −.072 | .090 | −.057 | .153 |
| Bias state 3 | 300 | .031 | .609 | −.053 | .649 |
| | 500 | .158 | .470 | .165 | .577 |
| | 1000 | .091 | .244 | .117 | .287 |
| | 5000 | .090 | .071 | .207 | .222 |
| Bias state 4 | 300 | .024 | .316 | −.030 | .332 |
| | 500 | .090 | .313 | .128 | .347 |
| | 1000 | .034 | .142 | .073 | .157 |
| | 5000 | −.023 | .045 | .048 | .088 |
| Bias state 5 | 300 | −.038 | .303 | −.026 | .326 |
| | 500 | −.009 | .204 | −.096 | .254 |
| | 1000 | −.011 | .099 | −.021 | .111 |
| | 5000 | −.027 | .044 | .010 | .050 |

transition matrix, the sum of the absolute difference between the true value and the estimated one is used as a measure of goodness of fit. The latter is always lower with the NKE, confirming also the increase in the estimation precision. As regards 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. Clearly, moving from 3 to 4 states the model complexity increases considerably and the overall performance of the algorithm decreases. This is particularly evident for the case with 5 hidden states in Table 3 in which the ability to recover the true state sequence, with $T = 300$, reduces to about 80% and 74.3% with and without NKE, respectively. Indeed, one of the drawbacks of HSMM is that they require long sequences of observations to obtain reliable estimates, especially for complex models with high-dimensional parameter spaces. However, the results show that adopting the NKE is particularly useful with short observation sequences and complex models, since the presence of sparsity may affect the estimation accuracy of the standard formulation based on the cell proportion estimator. Furthermore, using the NKE increases the estimation precision for all the parameters, indeed it seems to improve both the estimation of the transition matrix and the initial state distributions (see Appendix).

Finally, Table 4 shows the estimated initial distributions with 5 states and Fig. 2 provides a graphical representation of the case with $T = 300$. As it can be seen, for $T$ up to 1000, the improvement in the estimated initial distribution is evident, especially for small sample sizes.

## 6. Concluding remarks

The paper provides the model based and computational solution to a complex problem arising in modern quantum physics. The problem is the extraction of information from an open quantum system, subject to an external perturbation. The solution consists in the specification of an Explicit Duration Hidden Markov Model which takes into account the features of the experimental problem, i.e. presence of sparsity and the characterization of a complex (not directly observable) evolution using only partial information. By means of an extensive Monte Carlo study, we showed that our formulation outperforms the original one in the accuracy of the chain reconstruction along with a consistent reduction in the variability of the estimates, especially for short observation sequences.

Several additional aspects may be considered. Firstly, the ability to reliably track the evolution of the hidden system decreases considerably as the complexity of the model increases. This is mostly due to both the dichotomous nature of the observed signal and to the precision of the quantum measurement that is applied. Secondly, the choice of the maximum number of consecutive time steps in the same state $\bar{D}$ can be possibly carried out through a grid search over several

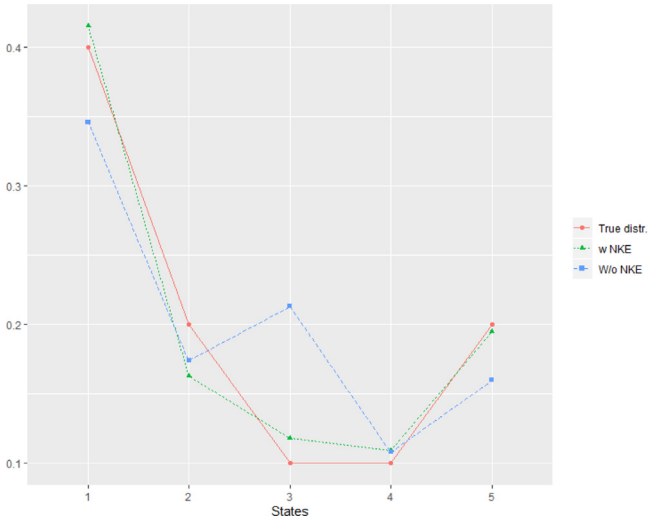


**Fig. 2.** Initial distribution, T = 300.

**Table 4**
Simulation results for 5 hidden states: initial distribution.

| | T | State 1 | State 2 | State 3 | State 4 | State 5 |
|---|---|---|---|---|---|---|
| True values | | .4 | .2 | .1 | .1 | .2 |
| w NKE | 300 | .415 | .163 | .118 | .109 | .195 |
| | 500 | .403 | .202 | .084 | .118 | .193 |
| | 1000 | .395 | .203 | .038 | .156 | .207 |
| | 5000 | .390 | .216 | .102 | .096 | .201 |
| w/o NKE | 300 | .346 | .174 | .213 | .108 | .160 |
| | 500 | .336 | .233 | .132 | .191 | .183 |
| | 1000 | .309 | .277 | .059 | .153 | .202 |
| | 5000 | .359 | .245 | .051 | .149 | .197 |

candidate values. Finally, assuming a known and pre-specified number of hidden states may limit the applicability of the methodology whenever the a priori information about the hidden system is scarce. Future research may focus on a Bayesian model selection procedure for estimating the number of hidden states.

## Acknowledgments

We would like to thank Ole Barndorff-Nielsen and Klaus Mølmer for the nice and useful discussions. We also would like to thank the Editor and two Referees for their insightful comments and suggestions.

## Appendix. Additional results

See Tables 5–9.

**Table 5**
Transition matrix used to simulate the semi-Markov chain with 3 hidden states.

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.0 | 0.75 | 0.25 |
| 2 | 0.4 | 0.0 | 0.6 |
| 3 | 0.5 | 0.5 | 0.0 |

**Table 6**
Transition matrix used to simulate the semi-Markov chain with 4 hidden states.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0.0 | 0.33 | 0.11 | 0.56 |
| 2 | 0.33 | 0.0 | 0.44 | 0.22 |
| 3 | 0.25 | 0.25 | 0.0 | 0.5 |
| 4 | 0.33 | 0.5 | 0.17 | 0.0 |

**Table 7**
Transition matrix used to simulate the semi-Markov chain with 5 hidden states.

| | 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 8**
Simulation results: initial distribution with 4 states.

| | T | State 1 | State 2 | State 3 | State 4 |
|---|---|---|---|---|---|
| True values | | .3 | .1 | .5 | .1 |
| w NKE | 300 | .295 | .129 | .474 | .101 |
| | 500 | .300 | .089 | .449 | .162 |
| | 1000 | .300 | .172 | .430 | .098 |
| | 5000 | .310 | .097 | .491 | .102 |
| w/o NKE | 300 | .262 | .163 | .479 | .096 |
| | 500 | .284 | .121 | .446 | .149 |
| | 1000 | .280 | .200 | .417 | .102 |
| | 5000 | .306 | .104 | .487 | .104 |

**Table 9**
Simulation results: initial distribution with 3 states.

| | T | State 1 | State 2 | State 3 |
|---|---|---|---|---|
| True values | | .3 | .2 | .5 |
| w NKE | 300 | .302 | .193 | .505 |
| | 500 | .308 | .198 | .494 |
| | 1000 | .301 | .205 | .492 |
| | 5000 | .290 | .208 | .502 |
| w/o NKE | 300 | .297 | .202 | .501 |
| | 500 | .320 | .193 | .487 |
| | 1000 | .301 | .209 | .492 |
| | 5000 | .286 | .213 | .500 |

## References


Aitchison, J., Aitken, C.G., 1976. Multivariate binary discrimination by the kernel method. Biometrika 63 (3), 413–420.

Barbu, V.S., Limnios, N., 2006. Maximum likelihood estimation for hidden semi-Markov models. C. R. Math. 342 (3), 201–205.

Barbu, V.S., Limnios, N., 2008. 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., Gill, R.D., Jupp, P.E., 2003. On quantum statistical inference. J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (4), 775–804.

Barreiro, J.T., Müller, M., Schindler, P., Nigg, D., Monz, T., Chwalla, M., Hennrich, M., Roos, C.F., Zoller, P., Blatt, R., 2011. An open-system quantum simulator with trapped ions. Nature 470 (7335), 486.

Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Stat. 41 (1), 164–171.

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

Bishop, Y., Fienberg, S., Holland, P., 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA.

Bulla, J., Bulla, I., 2006. Stylized facts of financial time series and hidden semi-Markov models. Comput. Statist. Data Anal. 51 (4), 2192–2209.

Chen, M.-Y., Kundu, A., Srihari, S.N., 1995. Variable duration hidden Markov model and morphological segmentation for handwritten word recognition. IEEE Trans. Image Process. 4 (12), 1675–1688.

Dong, D., Petersen, I.R., 2010. Quantum control theory and applications: a survey. Control Theory Appl. IET 4 (12), 2651–2671.

Dunjko, V., Briegel, H.J., 2018. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Rep. Progr. Phys. 81 (7), 074001.

Ferguson, J.D., 1980. Variable duration models for speech. In: Proceedings of the Symposium on the Application of HMMs to Text and Speech. Institute for Defense Analyses, Princeton, NJ, pp. 143–179.

Fienberg, S.E., Holland, P.W., 1973. Simultaneous estimation of multinomial cell probabilities. J. Amer. Statist. Assoc. 68 (343), 683–691.
Gammelmark, S., Mølmer, K., Alt, W., Kampschulte, T., Meschede, D., 2014. Hidden Markov model of atomic quantum jump dynamics in an optically probed cavity. Phys. Rev. A 89 (4), 043839.
Gleyzes, S., Kuhr, S., Guerlin, C., Bernu, J., Deleglise, S., Hoff, U.B., Brune, M., Raimond, J.-M., Haroche, S., 2007. Quantum jumps of light recording the birth and death of a photon in a cavity. Nature 446 (7133), 297–300.
Goggin, M., Almeida, M., Barbieri, M., Lanyon, B., O'Brien, J., White, A., Pryde, G., 2011. Violation of the leggett–garg inequality with weak measurements of photons. Proc. Natl. Acad. Sci. 108 (4), 1256–1261.
Guerlin, C., Bernu, J., Deleglise, S., Sayrin, C., Gleyzes, S., Kuhr, S., Brune, M., Raimond, J.-M., Haroche, S., 2007. Progressive field-state collapse and quantum non-demolition photon counting. Nature 448 (7156), 889–893.
Hall, P., Titterington, D., 1987. On smoothing sparse multinomial data. Aust. J. Statist. 29 (1), 19–37.
Hatridge, M., Shankar, S., Mirrahimi, M., Schackert, F., Geerlings, K., Brecht, T., Sliwa, K., Abdo, B., Frunzio, L., Girvin, S.M., et al., 2013. Quantum back-action of an individual variable-strength measurement. Science 339 (6116), 178–181.
Haussler, D.K.D., Eeckman, M.G.R.F.H., 1996. A generalized hidden Markov model for the recognition of human genes in DNA. In: Proc. Int. Conf. on Intelligent Systems for Molecular Biology, St. Louis. pp. 134–142.
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 Theory. North-Holland, Amsterdam.
Johnson, M.J., Willsky, A.S., 2013. Bayesian nonparametric hidden semi-Markov models. J. Mach. Learn. Res. 14 (1), 673–701.
Kirchmair, G., Zähringer, F., Gerritsma, R., Kleinmann, M., Gühne, O., Cabello, A., Blatt, R., Roos, C., 2009. State-independent experimental test of quantum contextuality. Nature 460 (7254), 494–497.
Krishnamurthy, V., Moore, J.B., Chung, S.-H., 1991. On hidden fractal model signal processing. Signal Process. 24 (2), 177–192.
Kubanek, A., Koch, M., Sames, C., Ourjoumtsev, A., Pinkse, P.W., Murr, K., Rempe, G., 2009. Photon-by-photon feedback control of a single-atom trajectory. Nature 462 (7275), 898–901.
Kundu, A., He, Y., Chen, M.-Y., 1997. Efficient utilization of variable duration information in HMM based HWR systems. In: Image Processing, 1997. Proceedings, International Conference on, Vol. 3. IEEE, pp. 304–307.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Ann. Statist. 20 (2), 712–736.
McFarland, J.M., Hahn, T.T., Mehta, M.R., 2011. Explicit-duration hidden Markov model inference of UP-DOWN states from continuous signals. PLoS One 6 (6), e21606.
Melnyk, I., Banerjee, A., 2015. A spectral algorithm for inference in hidden semi-Markov models. In: Artificial Intelligence and Statistics, Vol. 38. pp. 690–698.
Mitchell, C., Harper, M., Jamieson, L., et al., 1995. On the complexity of explicit duration HMM's. IEEE Trans. Speech Audio Process. 3 (3), 213–217.
Mitchell, C.D., Jamieson, L.H., 1993. Modeling duration in a hidden Markov model with the exponential family. In: Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on, Vol. 2. IEEE, pp. 331–334.
Nielsen, M.A., Chuang, I.L., 2010. Quantum Computation and Quantum Information. Cambridge university press.
Palacios-Laloy, A., Mallet, F., Nguyen, F., Bertet, P., Vion, D., Esteve, D., Korotkov, A.N., 2010. Experimental violation of a bell/'s inequality in time with weak measurement. Nat. Phys. 6 (6), 442–447.
Pertsinidou, C.-E., Limnios, N., 2015. Viterbi algorithms for hidden semi-Markov models with application to DNA analysis. RAIRO-Oper. Res. 49 (3), 511–526.
Pimentel, M.A., Santos, M.D., Springer, D.B., Clifford, G.D., 2015. Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. Physiol. Meas. 36 (8), 1717.
Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77 (2), 257–286.
Rajagopalan, B., Lall, U., 1995. A kernel estimator for discrete distributions. J. Nonparametr. Stat. 4 (4), 409–426.
Ramakrishna, V., Rabitz, H., 1996. Relation between quantum computing and quantum controllability. Phys. Rev. A 54 (2), 1715.
Ramesh, P., Wilpon, J.G., 1992. Modeling state durations in hidden Markov models for automatic speech recognition. In: Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, Vol. 1. IEEE, pp. 381–384.
Schmidler, S.C., Liu, J.S., Brutlag, D.L., 2000. Bayesian segmentation of protein secondary structure. J. Comput. Biol. 7 (1–2), 233–248.
Simonoff, J.S., 1983. A penalty function approach to smoothing large sparse contingency tables. Ann. Statist. 11 (1), 208–218.
Simonoff, J.S., 2012. Smoothing Methods in Statistics. Springer Science & Business Media.
Squire, K., Levinson, S.E., 2005. Recursive maximum likelihood estimation for hidden semi-Markov models. In: Machine Learning for Signal Processing, 2005 IEEE Workshop on. IEEE, pp. 329–334.
Titterington, D., 1980. A comparative study of kernel-based density estimates for categorical data. Technometrics 22 (2), 259–268.
Vijay, R., Slichter, D., Siddiqi, I., 2011. Observation of quantum jumps in a superconducting artificial atom. Phys. Rev. Lett. 106 (11), 110502.
Wand, M.P., Jones, M.C., 1994. Kernel Smoothing. Crc Press.
Wang, M.-C., Van Ryzin, J., 1981. A class of smooth estimators for discrete distributions. Biometrika 68 (1), 301–309.
Wiseman, H.M., Milburn, G.J., 2009. Quantum Measurement and Control. Cambridge University Press.
Yu, S.-Z., 2010. Hidden semi-Markov models. Artificial Intelligence 174 (2), 215–243.
Yu, S.-Z., Kobayashi, H., 2006. Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model. IEEE Trans. Signal Process. 54 (5), 1947–1951.
Zhu, K., Liu, T., 2017. Online tool wear monitoring via hidden semi-Markov model with dependent durations. IEEE Trans. Ind. Inf. 14 (1), 69–78.