



# The continuous-time hidden Markov model based on discretization. Properties of estimators and applications

María Luz Gámiz<sup>1</sup> · Nikolaos Limnios<sup>2</sup> · Mari Carmen Segovia-García<sup>1</sup>

Received: 28 October 2022 / Accepted: 16 May 2023  
© The Author(s) 2023

## Abstract

In this paper we consider continuous-time hidden Markov processes (CTHMM). The model considered is a two-dimensional stochastic process  $(X_t, Y_t)$ , with  $X_t$  an unobserved (*hidden*) Markov chain defined by its generating matrix and  $Y_t$  an observed process whose distribution law depends on  $X_t$  and is called the emission function. In general, we allow the process  $Y_t$  to take values in a subset of the  $q$ -dimensional real space, for some  $q$ . The coupled process  $(X_t, Y_t)$  is a continuous-time Markov chain whose generator is constructed from the generating matrix of  $X$  and the emission distribution. We study the theoretical properties of this two-dimensional process using a formulation based on semi-Markov processes. Observations of the CTHMM are obtained by discretization considering two different scenarii. In the first case we consider that observations of the process  $Y$  are registered regularly in time, while in the second one, observations arrive at random. Maximum-likelihood estimators of the characteristics of the coupled process are obtained in both scenarii and the asymptotic properties of these estimators are shown, such as consistency and normality. To illustrate the model a real-data example and a simulation study are considered.

**Keywords** Hidden Markov model · Poisson process · Maximun-likelihood estimation · Asymptotic properties

---

María Luz Gámiz, Nikolaos Limnios and Mari Carmen Segovia-García have contributed equally to this work.

---

✉ María Luz Gámiz  
mgamiz@ugr.es

Nikolaos Limnios  
nikolaos.limnios@utc.fr

Mari Carmen Segovia-García  
msegovia@ugr.es

<sup>1</sup> Departamento de Estadística e I.O., Universidad de Granada, Granada, Spain

<sup>2</sup> Laboratoire de Mathématiques Appliquées de Compiègne, Université de Technologie de Compiègne, Compiègne, France

# 1 Introduction

Hidden Markov Models (HMM) appear in a large number of real-world estimation problems where a process with unobservable states produce observable outputs normally referred as signals. Mor et al. (2021) conduct a review of the published work on HMM during the last few decades, and the different application areas of these models, e.g., Pattern recognition, Bioinformatics, Economy and Finance, Network security, Meteorology, Reliability engineering, etc.

The problem is presented as follows. Consider a coupled process,  $(X, Y)$ , in discrete or continuous time, where  $X$  is an unobserved process, a hidden process, and  $Y$  is the observed process. Based on the observations, the law of the coupled process has to be estimated. In general, the hidden process  $X$  is a Markov chain, and  $Y$  is a process whose distribution depends on the value of  $X_t$ , i.e.,  $\mathbb{P}(Y_t \in B | Y_s, X_s; s \leq t) = \mathbb{P}(Y_t \in B | X_t)$ , or  $\mathbb{P}(Y_t \in B | Y_s, X_s; s \leq t) = \mathbb{P}(Y_t \in B | Y_t, X_t)$ . In the first case we have an M1-M0 model and in the second an M1-M1 model, where M means Markov order 1 or 0. These are the most used orders, but in general we can consider more general orders.

In the literature, the considered Markov chain is a finite-state space process with transition probability  $P_{ij}$  being a function of a parameter  $\theta$  (in general a vector). It is written as  $P_{ij}(\theta)$ . Therefore, the estimation of the parameter vector  $\theta$ , i.e.,  $\hat{\theta}$ , gives us a plug-in estimator of the transition probability, i.e.,  $\hat{P}_{ij} := P_{ij}(\hat{\theta})$ .

Basic theoretical results concerning HMM in discrete time are given in Baum and Petrie (1966), and Leroux (1992) where consistency of estimators is proven, as well as in Bickel et al. (1998) that prove asymptotic normality of the estimator for a stationary process. These results concerning discrete-time hidden Markov models (DTHMM) case are used in our CTHMM in order to provide asymptotic results for our estimators.

Here, a continuous-time HMM (CTHMM) is considered, where  $X$  is defined by its generating matrix  $\mathbf{A}$  and  $Y$  by its probability law, with conditional distribution  $G(i, B)$  given  $\{X_t = i\}$ , and  $B$  a measurable set in  $\mathbb{R}^q$ , i.e.,  $\mathbb{P}(Y_t \in B | X_t = i) = G(i, B)$ , so the considered model is an M1-M0.

Our results concern the consistency and asymptotic normality of the estimator of the parameter  $\theta$  and applications to a real data set and simulated data.

It is assumed that the generator  $\mathbf{A}$  of  $X$  is a matrix depending on the parameter  $\theta$ , i.e.,  $\mathbf{A} = A(\theta)$  as well as the probability distribution of  $Y$ , i.e.,  $G = G_\theta$ . Let us denote  $g_\theta$  the density of  $G_\theta$  with respect to some dominating measure  $\mu$ .

The methodology to tackle the problem is based on results in DTHMM, as Bickel et al. (1998), Baum and Petrie (1966), Leroux (1992) and also on our results in Gamiz et al. (2023). In the present paper the CTHMM is approached by discretization, so that previous results can be applied here, and the parameter  $\theta$  can be estimated by the EM-algorithm. From the estimator  $\hat{\theta}$  we obtain an estimator of the generating matrix  $\hat{\mathbf{A}}$  and prove the asymptotic properties of this estimator such as normality and consistency.

We consider two different scenarii.

## 1.1 Scenario 1: Regular inspections in time

The true state of the system at time  $t$ ,  $X_t$  is non-observable. However, at regular intervals of length  $h$ ,  $0 < h < T$ , certain information related to the system is observed.

Let us show an example of the type of problems that can be treated with this approach. Let us consider a continuous-time Markov chain (CTMC)  $\{X_t, t > 0\}$  with generating matrix given by

$$\mathbf{A} = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\lambda + \mu) & \lambda \\ 0 & 2\mu & -2\mu \end{pmatrix}.$$

This model is a typical description of the functioning of a system with two units identical and independent. The units fail at a constant rate equal to  $\lambda$ . Once a unit fails is sent for repair. The repair system has capacity for the two units and the repair rate is also constant and equal to  $\mu$ . The state of the system is expressed as the number of units functioning, then the state space is  $E = \{0, 1, 2\}$ . For assessing the system performance, we focus on certain functionals  $\phi(\mathbf{A}, t)$ , such as the reliability of the system (the probability that the system has not suffered from failure before a given time), the availability function (the probability that the system is operative at a given time) or the expected hitting times (the expected time a particular class of states is reached by the system), among others. The usual procedure is to take a sample of identical systems that work under similar conditions and to estimate the unknown parameters  $\lambda$  and  $\mu$  from the data. The main problem here is that we consider that the system is not directly observable and then we are not able to register information regarding the number of units that are functioning in the system at any given time. In contrast, at some regular times,  $t_0 = 0, t_1 = h, t_2 = 2h, \dots, t_n = nh, \dots$ , we have access to certain indicators that provide useful, though partial, information about the state of the system. Let us denote by  $Y_n$  the accessible information about the system that we can observe at time  $t_n$ , for  $n > 0$ . We assume that  $\{Y_0, Y_1, \dots, Y_n\}$  are (conditionally) independent and that for each  $n$  the distribution of the random variable  $Y_n$  is determined by a parameter that depends on the current state of the system, that is, at time  $t_n = nh$ .

Let us define  $Q_h(i, j) = \mathbb{P}(X_h = j | X_0 = i)$ , for  $i, j \in E$ , and denote  $\mathbf{Q}_h$  the corresponding matrix. Using basic properties of CTMC we have that  $\mathbf{Q}_h = \mathbf{I} + \mathbf{A}h + o(h)$ , with  $\mathbf{I}$  the identity matrix and  $o(h)/h \rightarrow 0$  for  $h \rightarrow 0$ , then we can approximate  $\mathbf{Q}_h$ , for a fixed small enough  $h$ , by the following

$$\mathbf{Q}_h = \begin{pmatrix} 1 - 2\lambda h & 2\lambda h & 0 \\ \mu h & 1 - (\lambda + \mu)h & \lambda h \\ 0 & 2\mu h & 1 - 2\mu h \end{pmatrix}.$$

In this situation,  $\{Y_0, Y_1, \dots, Y_n\}$  can be seen as a sample from a DTHMM whose parameters can be written as a vector  $\theta$  that includes  $\lambda, \mu$  and some other parameters that determine the distribution of  $Y_n$  given  $\hat{X}_n = X_{nh} = i$ , for  $i \in E$ .

In general, we define  $\mathbf{A}_h = h^{-1}(\mathbf{Q}_h - \mathbf{I})$ . As  $h \rightarrow 0$ ,  $\mathbf{A}_h \rightarrow \mathbf{A}$ , uniformly. As  $\mathbf{Q} = Q(\theta)$ , then  $\mathbf{A} = A(\theta)$ . We estimate  $\hat{\theta}_n$  from observations of the hidden Markov model at times  $\{t_n = nh, n \geq 0\}$ . Leroux (1992) showed the strong consistency of this estimator, that is  $\hat{\theta}_n \rightarrow \theta$ , and in Bickel et al. (1998) the weak convergence of the estimator to a Normal law is proven, that is,  $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow N(0, \Sigma_0)$ , in the case of HMM.

Subsequently, for the plug-in estimator of the transition matrix  $\mathbf{Q}_h$ , that is  $\hat{\mathbf{Q}}_{h,n} = Q_h(\hat{\theta}_n)$ , the strong consistency and asymptotic normality are proven. At the same time the strong consistency and the asymptotic normality of the plug-in estimator of  $\mathbf{A}_h$ , that is  $\hat{\mathbf{A}}_{h,n} = A(\hat{\mathbf{Q}}_{h,n})$ , is also demonstrated in this paper.

And as a consequence, for any functional within the following class  $H(t) = \Phi(\mathbf{A}, t)$ ,  $H_h(t) = \Phi(\mathbf{A}_h, t)$  is defined. Then the plug-in estimator for  $H_h(t)$ ,  $\hat{H}_{h,n}$ , is shown to be strongly consistent and asymptotically normal. As an application of this result, the transition matrix function,  $H(t) = \mathbf{P}(t) = e^{\mathbf{A}t}$ , can be considered, as well as other functions such as reliability, availability, etc.

## 1.2 Scenario 2: Random inspections in time

For a given interval length  $T$ , the true state of the system at time  $t$ ,  $X_t$  is non-observable. Observations related to the underlying state are received at random times following an Homogenous Poisson Process (HPP),  $N(t)$ , with intensity  $\lambda$ , unknown and estimated from the data. We consider the embedded Markov chain obtained from the visited states of  $X$  at the successive arrival times of the HPP. We denote  $\mathbf{Q}$  its transition probability matrix and  $\mathbf{A}$  the generating matrix of the CTMC. We assume that  $\lambda \geq \max\{a_i, i \in E\}$ . Using the uniformization method, see, e.g., Kulkarni (2011), we define  $\mathbf{Q} = (1/\lambda)\mathbf{A} + \mathbf{I}$ . Using the HMM, an estimator  $\hat{\mathbf{Q}}$  of  $\mathbf{Q}$  is obtained, and subsequently the corresponding estimator of the generating matrix  $\mathbf{A}$  is obtained as  $\hat{\mathbf{A}} = \hat{\lambda}(\hat{\mathbf{Q}} - \mathbf{I})$ . It is demonstrated that these estimators  $\hat{\mathbf{Q}}$  and  $\hat{\mathbf{A}}$  are strongly consistent and asymptotically normal.

The model discussed in this scenario should not be confused with the so called Markov-modulated Poisson process, which consists of a Poisson process whose rate is controlled by a non-observable continuous-time Markov chain. In Freed and Shepp (1982) it is considered a switched Poisson process, i.e., with only two states for the hidden chain. The authors assume that the rate of one of the states is zero and derive a simple formula for the asymptotic likelihood ratio that allows to estimate the state at any time from a stream of past events. In our case, a Poisson process is assumed to deal with the number of signals registered by an external observer but it is not responsible for the number of signals emitted by the hidden source nor for the nature of such signals, as it is the case of the Poisson process involved in a Markov-modulated Poisson process.

The results summarized here can be utilized in different areas, such as system reliability, biology, etc., e.g., in Wei et al. (2002) a CTHMM is proposed for evaluating network performance considering discrete time observations. Zhou et al. (2020) proposed as well a CTHMM where the observations can be collected regularly, irregularly or continuously. The number of states is unknown. They apply this model to bladder cancer data. In Verma et al. (2018) the authors develop a CTHMM under a generalized linear modeling framework to model the evolution of chronic obstructive pulmonary disease (COPD), again, the model considers discrete-time observations although they are irregularly-spaced observations. Similarly Hulme et al. (2021) proposed a CTHMM to estimate health conditions of patients that are monitored by wearable and mobile technology, the observations are as well irregularly-spaced. Finally, Liu et al. (2016) considers the use of CTHMM for modeling disease progression as well.

This paper is organized as follows. In Sect. 2 the model and its main characteristics are defined. The generator of the two-dimensional process is defined and obtained in terms of the generating matrix  $\mathbf{A}$  and the emission distribution  $G$ . The basic properties of the model are obtained using a formulation based on semi-Markov processes. In Sect. 3, maximum-likelihood estimators for the parameters of the model are obtained based on data observed under two different time discretization schemes. That is, scenario 1, where observations arrive at times regularly pre-specified; and, scenario 2, where observations arrive randomly according to a homogeneous Poisson process. Some applications and particular cases of the model are described in Sect. 3.6.1. Some numerical examples are presented in Sect. 5 and the conclusions are given in Sect. 6.

## 2 The model

We follow the formulation of Bickel et al. (1998) and consider the “usual parametrization” of the model. Specifically, let  $\{X_t; t \geq 0\}$  be a continuous-time Markov chain with finite state space  $E = \{1, \dots, d\}$  and generating matrix  $\mathbf{A} = (a_{ij}; i, j \in E)$ . Let  $\{Y_t; t \geq 0\}$  an  $\mathcal{Y}$ -valued sequence such that given  $X_t = i$ ,  $Y_t$  is conditionally distributed with density  $g(i, y)$  with respect to some  $\sigma$ -finite measure  $\mu$  on  $\mathcal{Y}$ , and fixed  $i \in E$ .

All processes are defined on a complete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , are right continuous and then progressively measurable. This will imply also a family of probabilities  $\{\mathbb{P}_i; i \in E\}$ , where  $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot | X_0 = i)$ , and expectations  $\{\mathbb{E}_{(i,y)}; i \in E, y \in \mathcal{Y}\}$ , where  $\mathbb{E}_{(i,y)}[\cdot] = \mathbb{E}[\cdot | X_0 = i, Y_0 = y]$ , which are defined with respect to a family of probabilities  $\{\mathbb{P}_{(i,y)}; i \in E, y \in \mathcal{Y}\}$ , where  $\mathbb{P}_{(i,y)}(\cdot) = \mathbb{P}(\cdot | X_0 = i, Y_0 = y)$ .

Both, the generating matrix  $\mathbf{A}$  as well as the family of densities  $\{g(i, \cdot); i \in E\}$ , depend on a vector of parameters  $\theta$ , that is  $A(i, j) = a_{ij}(\theta)$  and  $g(i, \cdot) = g_\theta(i, \cdot)$ , for all  $i, j \in E$ . The set of possible values of the vector  $\theta$  is denoted  $\Theta \subset \mathbb{R}^k$ , and  $\theta$  has to be estimated from a set of observations of the process  $\{Y_t\}$ .

The vector  $\theta$  usually includes the transition rates and also some parameters characterizing the densities  $g$ . As a particular case, it can be assumed that  $g(i, y)$  denotes a concrete parametric family of distributions with parameters  $\beta = (\beta_1, \beta_2, \dots, \beta_s)$ , then we can write  $\theta = (\mathbf{A}^*, \beta)$ , with  $\mathbf{A}^*$  the matrix  $\mathbf{A}$  without its principal diagonal, since the diagonal entries are functions of the off diagonal entries. In our previous example of a 3-state MC, if we have that  $g(i, \cdot)$  is the Normal distribution  $N(\kappa_i, \sigma_i^2)$ ,  $i \in E$ ; the parameter vector  $\theta$  will be:  $\theta = (\lambda, \mu, \kappa_1, \sigma_1^2, \kappa_2, \sigma_2^2, \kappa_3, \sigma_3^2)$ .

In general  $\mathcal{Y}$  can be seen as a subset in  $\mathbb{R}^q$  for some  $q$ . If  $Y_t$  has density function  $g(i, \cdot)$ , given that  $X_t = i$ , then  $\mathbb{P}(Y_t \in B | X_t = i) = \int_B g(i, y)\mu(dy)$ , for  $B \subset \mathcal{Y}$ , is the probability that the process  $Y_t$  takes values in a subset  $B$  given  $X_t = i$ , for  $i \in E$ . We also denote  $G(i, B) = \mathbb{P}(Y_t \in B | X_t = i)$ .

At some points of the paper we will discuss the simpler case that  $Y_t$  takes values in a finite set, that is  $\mathcal{Y} = \{y_1, \dots, y_s\}$  and then the emission function will be a  $d \times s$ -dimensional matrix  $\mathbf{G}$ , with elements  $G(i, y) = G(i, \{y\}) = \mathbb{P}(Y_t = y | X_t = i)$ , for all  $i \in E$  and  $y \in \mathcal{Y}$ , and all  $t > 0$  (see Gamiz et al. 2023).

For the rest of the paper we will consider the process  $(X, Y)$  defined as follows.

**Definition 1** Let  $X = \{X_t; t \geq 0\}$  be an irreducible homogeneous Markov process in a finite set  $E$  and with generating matrix  $\mathbf{A} = (a_{ij})_{i,j \in E}$ , and  $Y = \{Y_t; t \geq 0\}$  is a homogeneous process in a general set  $\mathcal{Y} \subset \mathbb{R}^q$ , with  $q \geq 1$  such that, for  $t > 0$ , the distribution of  $Y_t$  is determined  $G(i, \cdot)$ , over the event  $X_t = i$ , for  $i \in E$ . Then we say that  $(X, Y)$  is a two-dimensional process with dependence structure M1-M0. That is, if  $\mathcal{B}_\mathcal{Y}$  denotes the set of Borel subsets of  $\mathcal{Y}$ , then for any  $i, j \in E, y \in \mathcal{Y}$  and  $B \in \mathcal{B}_\mathcal{Y}$ , and for all  $s, t > 0$ ,

$$\begin{aligned} \mathbb{P}(X_{t+s} = j, Y_{t+s} \in B | X_u, Y_u, 0 \leq u \leq s) \\ = \mathbb{P}(X_{t+s} = j, Y_{t+s} \in B | X_s) \\ = \mathbb{P}(X_t = j, Y_t \in B | X_0), \end{aligned} \tag{1}$$

where we also use homogeneity of the processes  $X$  and  $Y$ . Moreover, for  $t > 0$ ,

$$\mathbb{P}(X_t = j, Y_t \in B | X_0 = i, Y_0 = y) = P_{ij}(t)G(j, B),$$

and, for  $t = 0$ ,  $\mathbb{P}(X_t = j, Y_t \in B | X_0 = i, Y_0 = y) = 1$  if  $i = j$  and  $y \in B$ ; and, 0 otherwise.

### 2.1 Generators

Let us focus on the two-dimensional process  $\{(X_t, Y_t); t \geq 0\}$ , where, as above,  $X$  is a continuous-time Markov chain taking values in the set  $E = \{1, 2, \dots, d\}$ , with generating matrix  $\mathbf{A}$  and initial law  $\alpha$ ; and  $Y$  is a stochastic process taking values in a set  $\mathcal{Y} \subset \mathbb{R}^q$  whose distribution depends on  $X_t$ .

First we recall the concept of *generator* operator for a two-dimensional process with a general state space.

**Definition 2** (Generator of two-dimensional process with general state space) Let  $(X, Y) = \{(X_t, Y_t); t > 0\}$  a two-dimensional process, with  $X$  taking values in a set  $E \subset \mathbb{R}^d$  and  $Y$  taking values in a set  $\mathcal{Y} \subset \mathbb{R}^q$  and let  $f \in \mathcal{C}(E \times \mathcal{Y})$ , where  $\mathcal{C}(E \times \mathcal{Y})$  denotes the set of all continuous bounded functions defined on  $E \times \mathcal{Y}$ . We define the generator  $\tilde{\mathbf{A}}$  of  $(X, Y)$  as follows

$$\tilde{\mathbf{A}}f(x, y) = \lim_{t \rightarrow 0} \frac{1}{t} E_{x,y} [f(X_t, Y_t) - f(x, y)],$$

for  $(x, y) \in E \times \mathcal{Y}$ .

We consider that  $X$  is a MC with finite state space  $E$ , while  $Y$  is a process taking values in a subset  $\mathcal{Y} \subset \mathbb{R}^q$ , for  $q \geq 1$ . Since  $X$  is a CTMC then its generator is a matrix  $\mathbf{A} = (a_{ij}; i, j \in E)$ , where  $a_{ij} \geq 0, i \neq j$  and  $a_{ii} = -\sum_{j \neq i} a_{ij}$  for all  $i \in E$ . On the other hand,  $\{Y_t\}$  is a sequence conditionally independent, where the law of  $Y_t$  depends on the value of  $X_t$ .

**Proposition 1** Let  $(X, Y)$  be a stochastic process where  $X$  is a CTMC with finite state space  $E$  and generating matrix  $\mathbf{A} = (a_{ij}; i, j \in E)$ , and let the transition probabilities of the process  $(X, Y)$  be given by (1). The generator of the two-dimensional MC  $(X_t, Y_t)$  can be written as

$$\tilde{\mathbf{A}}f(i, y) = \sum_{j \in E \setminus \{i\}} a_{ij} \int_{\mathcal{Y}} G(j, du) [f(j, u) - f(i, y)]$$

for all  $i \in E$  and  $y \in \mathcal{Y}$ .

**Proof** From definition 2 we have, for  $i \in E$  and  $y \in \mathcal{Y}$ ,

$$\tilde{\mathbf{A}}f(i, y) = \lim_{t \rightarrow 0} \frac{1}{t} \sum_{j \in E} \int_{\mathcal{Y}} [f(j, u) - f(i, y)] \mathbb{P}(X_t = j, Y_t \in du | X_0 = i, Y_0 = y)$$

From the expression in (1) we get

$$\begin{aligned} \tilde{\mathbf{A}}f(i, y) &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_{j \in E} \int_{\mathcal{Y}} (f(j, u) - f(i, y)) P_{ij}(t) G(j, du) \\ &= \sum_{j \in E} \int_{\mathcal{Y}} f(j, u) G(j, du) \left( \lim_{t \rightarrow 0} \frac{P_{ij}(t) - \delta_{ij} \delta_{yu}}{t} \right) \\ &= a_{ii} f(i, y) + \sum_{j \neq i} a_{ij} \int_{\mathcal{Y}} f(j, u) G(j, du) \end{aligned}$$

□

In particular, when  $\mathcal{Y}$  is a finite set, the generator is given by a matrix  $\tilde{\mathbf{A}}$  as we show next. The general expression in definition 2 becomes

$$\tilde{\mathbf{A}}f = \sum_{i \in E} \sum_{y_2 \in \mathcal{Y}} f(j, y_2) \left( \lim_{t \rightarrow 0} \frac{\mathbb{P}(X_t = j, Y_t = y_2 | X_0 = i, Y_0 = y_1) - \delta_{ij} \delta_{y_1 y_2}}{t} \right)$$

We have that  $\mathbb{P}(X_t = j, Y_t = y_2 | X_0 = i, Y_0 = y_1) = P_{ij}(t)G(j, y_2)$ , and we assume that  $\mathbb{P}(X_t = j, Y_t = y_2 | X_0 = i, Y_0 = y_1) \rightarrow 0$  as  $t \rightarrow 0$  when  $y_1 \neq y_2$ . We consider the following cases:

- If  $i = j, y_1 = y_2, (1/t) [P_{ij}(t)G(j, y_2) - 1] \rightarrow a_{ii}$ ;
- If  $i = j, y_1 \neq y_2, (1/t) [P_{ij}(t)G(j, y_2)] \rightarrow 0$ ; and,
- If  $i \neq j, (1/t) [P_{ij}(t)G(j, y_2)] \rightarrow a_{ij}G(j, y_2)$

Then the generator can be written as a matrix whose elements are

$$\tilde{a}((i, y_1), (j, y_2)) = \begin{cases} a_{ii}, & i = j, y_1 = y_2 \\ 0, & i = j, y_1 \neq y_2 \\ a_{ij}G(j, y_2), & i \neq j \end{cases} \tag{2}$$

Moreover, we have

$$\begin{aligned} \sum_{j \in E} \sum_{y_2 \in \mathcal{Y}} \tilde{a}((i, y_1), (j, y_2)) &= a_{ii} + \sum_{j \neq i} \sum_{y_2 \in \mathcal{Y}} a_{ij}G(j, y_2) \\ &= a_{ii} + \sum_{j \neq i} a_{ij} = 0 \end{aligned}$$

**Example 1** For illustrative purposes, let us construct the 2-dimensional generator for a simple case where  $E = \{1, 2\}$  and  $\mathcal{Y} = \{y_1, y_2, y_3\}$ . Then, the state space of the coupled process is  $\tilde{E} = \{(1, y_1), (1, y_2), (1, y_3), (2, y_1), (2, y_2), (2, y_3)\}$ , and the corresponding generating matrix is

$$\tilde{\mathbf{A}} = \begin{pmatrix} a_{11} & 0 & 0 & a_{12}G(2, y_1) & a_{12}G(2, y_2) & a_{12}G(2, y_3) \\ 0 & a_{11} & 0 & a_{12}G(2, y_1) & a_{12}G(2, y_2) & a_{12}G(2, y_3) \\ 0 & 0 & a_{11} & a_{12}G(2, y_1) & a_{12}G(2, y_2) & a_{12}G(2, y_3) \\ a_{21}G(1, y_1) & a_{21}G(1, y_2) & a_{21}G(1, y_3) & a_{22} & 0 & 0 \\ a_{21}G(1, y_1) & a_{21}G(1, y_2) & a_{21}G(1, y_3) & 0 & a_{22} & 0 \\ a_{21}G(1, y_1) & a_{21}G(1, y_2) & a_{21}G(1, y_3) & 0 & 0 & a_{22} \end{pmatrix}$$

where  $a_{..}$  denote the elements of the generating matrix  $\mathbf{A}$  of  $X$ , and  $G(\cdot, \cdot)$  denote the emission probabilities.

From the generator in Proposition 1 we obtain the semigroup

$$P(t) = \exp(\tilde{\mathbf{A}}t). \tag{3}$$

### 2.2 Markov renewal equation for Markov processes

The above semigroup is difficult to handle numerically. For this reason we will consider another equivalent formulation via semi-Markov processes (see for example Limnios and Oprisan 2001). In fact, Markov processes are particular cases of semi-Markov processes and we may use the Markov renewal equation in order to obtain the above formulation (3) in a much easier numerical scheme.

Let us define the semi-Markov kernel of the Markov process  $X_t$  in the following way.

Let the process  $(X, Y)$  defined as in Definition 1.

For  $i, j \in E$  and  $B \in \mathcal{B}_Y$ , we define the following functions

- $M(i, j, B, t) = G(i, B)e^{-a_i t} \delta_{ij}$ ,
- $L(i, j, t) = (a_{ij}/a_i)(1 - e^{-a_i t})$ , and,
- $\tilde{P}(i, j, B, t) = \mathbb{P}_i(X_t = j, Y_t \in B)$ ,

where  $\mathbb{P}_i(X_t = j, Y_t \in B) = \mathbb{P}(X_t = j, Y_t \in B | X_0 = i)$ , and  $\delta_{ij}$  denotes the Kronecker delta and  $a_i = -a_{ii}$ .

Note that the function  $L(i, j, t)$  is the *semi-Markov kernel of the Markov process*  $\{X_t\}$ .

In particular, when  $Y_t$  takes values in a finite set of size  $s$ , we can define the matrix of transition functions of the process  $(X, Y)$ ,  $\tilde{\mathbf{P}}_t$  with dimension  $(d \cdot s) \times (d \cdot s)$  and elements

$$\tilde{P}_t((i, y'), (j, y)) = P_{ij}(t)G(j, y), \quad i, j \in E, \quad y', y \in \mathcal{Y}, \quad \text{and } t \geq 0. \tag{4}$$

In the same way we can construct matrices  $\mathbf{M}_t$  and  $\mathbf{L}_t$  whose elements are given above.

Given  $\phi$  a measurable function defined on  $E \times R_+$  we define its convolution by  $L$  as follows

$$(L * \phi)(i, t) = \sum_{j \in E} \int_0^t L(i, j, ds)\phi(j, t - s), \quad i \in E, t \geq 0. \tag{5}$$

**Proposition 2** (Markov renewal equation) *The function  $\tilde{P}(i, j, B, t)$  verifies the following Markov renewal equation*

$$\tilde{P}(i, j, B, t) = M(i, j, B, t) + \sum_{k \in E} \int_0^t L(i, k, ds)\tilde{P}(k, j, B, t - s), \tag{6}$$

where  $i, j \in E, B \in \mathcal{B}_Y$ , and  $t \geq 0$ .

**Proof** We have that

$$\begin{aligned} \mathbb{P}_i(X_t = j, Y_t \in B) &= G(i, B)e^{-a_i t} \delta_{ij} \\ &+ \sum_{k \in E} \int_0^t \mathbb{P}_i(X_t = j, Y_t \in B, J_1 = k, T_1 \in ds) \end{aligned}$$

where  $(J_1, T_1)$  describes the first jump of the Markov renewal process  $\{(J_n, T_n), n > 0\}$  associated to the process  $X$ , that is  $J_1 = X_{T_1}$ .

Then

$$\begin{aligned} \mathbb{P}_i(X_t = j, Y_t \in B) &= G(i, B)e^{-a_i t} \delta_{ij} \\ &+ \sum_{k \in E} \int_0^t \mathbb{P}_i(X_t = j, Y_t \in B | J_1 = k, T_1 = s)\mathbb{P}_i(J_1 = k, T_1 \in ds) \\ &= G(i, B)e^{-a_i t} \delta_{ij} \\ &+ \sum_{k \in E} \int_0^t \mathbb{P}_i(X_{t-s} = j, Y_{t-s} \in B)a_{ik}e^{-a_i s} ds \\ &= M(i, j, B, t) + \sum_{k \in E} \int_0^t L(i, k, ds)\tilde{P}(k, j, B, t - s) \\ &= M(i, j, B, t) + (L * \tilde{P})(i, j, B, t) \end{aligned}$$

□

It has been proven that  $\tilde{P}(i, j, B, t) = M(i, j, B, t) + (L * \tilde{P})(i, j, B, t)$ , where  $*$  denotes the convolution operation defined in (5).

For the case that  $Y_t$  takes values in a finite set we can write Eq. (6) in matrix notation as follows

$$\tilde{\mathbf{P}}_t = \mathbf{M}_t + \mathbf{L}_t * \tilde{\mathbf{P}}_t$$

Let us define  $\Psi$  the Markov renewal function of the process  $X$ , that is  $\Psi(t) = \sum_{n \geq 0} \mathbf{L}_t^{(n)}$ , where  $\mathbf{L}_t^{(n)}$  denotes the  $n$ th fold-convolution of  $\mathbf{L}_t$ , that is, the  $(i, j)$ -element given by

$$L^{(n)}(i, j, t) = \sum_{k \in E} \int_0^t L(i, k, ds) L^{(n-1)}(k, j, t - s), \quad i, j \in E, \quad t > 0,$$

being  $L^{(0)}(i, j, t) = \delta_{ij} 1_{\mathbb{R}_+}(t)$ , and  $L^{(1)}(i, j, t) = L(i, j, t)$ .

Then the solution of Eq. (6) is given by  $\tilde{P}(i, j, B, t) = (\Psi * M)(i, j, B, t)$ , see for example Limnios (2012).

Using the Markov renewal theorem (MRE), see Shurenkov (1984), the stationary distribution of the process  $(X, Y)$  can be obtained in the following proposition.

**Proposition 3** (Stationarity) *Under the conditions above and if  $X$  is irreducible, we have*

$$\lim_{t \rightarrow \infty} \tilde{P}(i, j, B, t) = \pi_j G(j, B), \quad j \in E, \quad B \in \mathcal{B}_Y.$$

*So we have that the probability distribution  $\tilde{\pi}$  whose elements are  $\tilde{\pi}(j, B) = \pi_j G(j, B)$ , for  $j \in E$ , and  $B \in \mathcal{B}_Y$  is the limit distribution of the process  $(X, Y)$ .*

*Moreover,  $\tilde{\pi}(i, B)$  is the stationary distribution of the process  $(X, Y)$ , that is, it verifies that*

$$\tilde{\pi} \tilde{\mathbf{A}} f = 0, \quad f \in \mathbb{C}(E \times \mathcal{Y}).$$

**Proof** For the first part of the proposition, let  $\rho$  denote the stationary distribution of the embedded Markov chain  $J_n$  associated to the MC  $X_t$ , with transition probabilities  $p_{ij} = a_{ij}(1 - \delta_{ij})/a_i$ .

Let  $\pi$  be the row vector of stationary probabilities of the CTMC  $X_t$ , that is, verifying that  $\pi \mathbf{A} = \mathbf{0}$ . It is satisfied that  $\pi_i = m_i \rho_i / m$ , with  $m_i = 1/a_i$ , the mean sojourn time in state  $i$  and  $m = \sum_{j \in E} m_j \rho_j$ .

Then, by the irreducibility of  $X$  and, as  $\tilde{P}(i, j, B, t) = (\Psi * M)(i, j, B, t)$  by MRT, we have that

$$\begin{aligned} \lim_{t \rightarrow \infty} (\Psi * M)(i, j, B, t) &= \frac{1}{m} \sum_{k \in E} \rho_k \int_0^\infty M(k, j, B, t) dt \\ &= \frac{1}{m} \sum_{k \in E} \rho_k \int_0^\infty G(k, B) e^{-a_k t} \delta_{kj} dt \\ &= \frac{1}{m} \rho_j G(j, B) \frac{1}{a_j} = \pi_j G(j, B) \end{aligned}$$

Then we define  $\tilde{\pi}(j, B) = \pi_j G(j, B)$ , for  $j \in E$  and  $B \in \mathcal{B}_Y$  the stationary distribution of the coupled process  $(X, Y)$ .

For the second part, we can write

$$\tilde{\pi} \tilde{\mathbf{A}} f = \sum_{j \in E} \int_{\mathcal{Y}} \tilde{\pi}(j, dy) \tilde{\mathbf{A}} f(j, y),$$

and, from the definition of the generator  $\tilde{\mathbf{A}}$  and the definition of  $\tilde{\pi}$ , we get

$$\begin{aligned}\tilde{\pi}\tilde{\mathbf{A}}f &= \sum_{j \in E} \int_{\mathcal{Y}} \pi_j G(j, dy) \sum_{i \in E} a_{ji} \int_{\mathcal{Y}} G(i, du) [f(i, u) - f(j, y)] \\ &= \sum_{i \in E} \sum_{j \in E} \pi_j a_{ji} \int_{\mathcal{Y}} \int_{\mathcal{Y}} G(j, dy) G(i, du) [f(i, u) - f(j, y)] \\ &= 0\end{aligned}$$

where the last equality is deduced from the fact that  $\pi \mathbf{A} = 0$ .  $\square$

**Remark 1** As the component  $G(j, B)$  does not depend on the time, we can also use relation (4) to derive directly the limit in the first part of Proposition 3.

### 3 Estimation

As described in Sect. 2, the two-dimensional process  $(X, Y)$  is fully specified by the generating matrix of the continuous-time MC  $X$ ,  $\mathbf{A}$  and the emission distribution  $G$ . Let us assume that  $\mathbf{A} = A(\theta)$ , and that  $G = G_\theta$ , that is, they both depend on a vector of unknown parameters  $\theta$ . The problem here is to obtain estimators of the parameter from data. Let us consider that the MC  $X$  is unobservable while we can register information about the state of the process  $Y$ . In other words, consider that our data come from a continuous-time HMM  $(X_t, Y_t)$ , where  $X_t$  is the hidden process and we aim at estimating  $\theta$  from a sample of observations of the process  $Y_t$  in an interval of time  $[0, T]$ .

#### 3.1 Time discretization

##### 3.2 Scenario 1. Regular inspections in time

In this section, observations emitted by a CTHMM are recorded at regular time points.

Let us consider a stochastic system according to a continuous-time Markov chain  $\{X_t; t \geq 0\}$  with state space  $E = \{1, 2, \dots, d\}$  as above; generating matrix  $\mathbf{A}$ , that is  $\mathbf{P}(t) = e^{\mathbf{A}t}$ , for all  $t > 0$ .

The true state of the system  $X_t$  is non observable at any time  $t \in (0, T]$ . However, at  $n$  regular pre-specified times denoted by  $0 < h < 2h < \dots < nh = T$ , observations of a random process  $\{Y_t; t \geq 0\}$  are available providing certain information about the true state of the system. Let us denote  $\hat{Y}_k = Y_{kh}$  the  $k$ -th observed signal, which is recorded at time  $kh$ , for  $k \in \{0, 1, \dots, n\}$ .

Associated to (or embedded in) the continuous-time process we can consider the following discrete time Markov chain  $\hat{X}_k = X_{kh}$ , with transition matrix  $\mathbf{Q}_h = e^{\mathbf{A}h}$ , whose  $(i, j)$ -element is given by  $Q_h(i, j) = \mathbb{P}(X_h = j | X_0 = i)$ . The discrete Markov chain  $\hat{X}_k$  is called the *discrete skeleton* (at scale  $h$ ) of  $X_t$ , (see Kingman 1963) for ergodicity properties of Markov chains based on discrete skeletons).

Consider the sequence of times  $0 < h < 2h < \dots < nh$ , then  $\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n$  can be seen as a realization of the *embedded* discrete hidden Markov model that we denote by  $(\hat{X}, \hat{Y})$ . We can use the results in Bickel et al. (1998) to estimate the transition matrix of this *embedded* hidden Markov chain  $\{\hat{X}_n\}$ , that is we obtain  $\hat{\mathbf{Q}}_h$ , as well as the emission functions  $G(i, dy) = \mathbb{P}(\hat{Y}_k \in dy | \hat{X}_k = i)$ , for  $i \in E$ .

For an interval of length  $h$ , we have  $\mathbf{Q}_h = \mathbf{P}(h) = e^{\mathbf{A}h}$ ; where  $\mathbf{A} = (a_{ij}; i, j \in E)$  is the generating matrix of the hidden continuous-time Markov chain;  $\mathbf{P}(h)$  is the corresponding transition function matrix; and,  $\mathbf{Q}_h$  is the transition matrix of the Markov chain  $\hat{X}_k$ . Then, using Taylor expansion, we approximate

$$a_{ij} = \frac{1}{h} (Q_h(i, j) - \delta_{ij}) + o(1)$$

where  $o(h)/h \rightarrow 0$  as  $h \rightarrow 0$ , and  $\delta_{ij}$  equal to 1 if  $i = j$  and 0 otherwise. Then, we define the estimator of the generating matrix as

$$\hat{\mathbf{A}}_h := \frac{1}{h} (\hat{\mathbf{Q}}_h - \mathbf{I}), \tag{7}$$

where  $\hat{\mathbf{Q}}_h$  is the maximum likelihood estimator of the transition matrix corresponding to the discrete-time chain  $\hat{X}$ , obtained as in Gamiz et al. (2023), and  $\mathbf{I}$  the identity matrix.

Note that since we write  $\mathbf{A} = A(\theta)$  we also have that  $\mathbf{Q}$  is a function of the parameter vector  $\theta$ , that is,  $\mathbf{Q} = Q(\theta)$ .

### 3.3 Scenario 2. Random inspections in time

In this case, we consider that the observations of the process  $\{Y_t\}$  are received at random times following a Poisson process  $N(t)$ , independent of  $X_t$  and  $Y_t$ , with intensity  $\lambda$ , which is unknown and can be estimated from the data.

The number of observations into the interval  $(0, T)$  is finite, random and equal to  $N(T)$ , where  $N(t)$ ,  $t \geq 0$  is a HPP with intensity  $\lambda \geq \max\{a_i\}$ , and  $a_i = -a_{ii} > 0$ , for all  $i = 1, \dots, d$ .

Let us denote  $\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n$  the observations registered in an interval  $(0, T]$ . The states visited by the Markov process  $X$  at the successive arrival times of the Poisson process can be seen as an embedded discrete-time hidden Markov chain  $Z_n$  with transition probability matrix denoted by  $\mathbf{Q}$ . The generating matrix of the hidden process is denoted by  $\mathbf{A}$  already given before, then, using the *uniformization* method (see, e.g., Kulkarni 2011, or Girardin and Limnios 2018), we have that

$$\mathbf{Q} = \frac{1}{\lambda} \mathbf{A} + \mathbf{I}$$

where  $\mathbf{I}$  is the identity matrix.

Using the HMM model we estimate again the transition matrix of the embedded Markov chain,  $\hat{\mathbf{Q}}$ . So we can define the following estimator

$$\hat{\mathbf{A}} = \hat{\lambda} (\hat{\mathbf{Q}} - \mathbf{I}), \tag{8}$$

where  $\lambda$  can be estimated by  $\hat{\lambda} = N(T)/T$ .

### 3.4 Assumptions

In order to establish consistency and asymptotic normality of the estimators, we need the following assumptions.

Let us define  $G_\theta(i, dy) := \mathbb{P}_\theta(Y_t \in dy | X_t = i) = g_\theta(i, y)\mu(dy)$ ,  $i \in E$  and  $\theta \in \Theta \subset \mathbb{R}^k$ , with  $\Theta$  an open set, and where  $\mu$  is some reference measure dominating all  $G_\theta(i, \cdot)$ .

Denote by  $\|\cdot\|$  the euclidean norm in  $\mathbb{R}^k$ , and  $\theta = (\theta_1, \dots, \theta_k)$ . The true value of the vector of parameters is denoted as  $\theta_0 = (\theta_{01}, \dots, \theta_{0k})$ .

- A1: The MP  $X$  is irreducible, i.e., ergodic. We consider moreover that  $X$  is stationary.  
 A2: The mixtures of  $G_\theta(i, \cdot)$  are identifiable.  
 A3: The functions  $\theta \mapsto P_\theta(i, j)$  and  $\theta \mapsto G_\theta(i, \cdot)$  belongs to  $C^2(\Theta)$ .  
 A4: For some  $\delta > 0$ , and all  $i \in E$ ,

$$\begin{aligned} \mathbb{E}_{\theta_0} \|\ln g_\theta(i, Y_0)\| &< +\infty, \\ \mathbb{E}_{\theta_0} \left[ \sup_{\|\theta - \theta_0\| < \delta} (\ln g_\theta(i, Y_0))^+ \right] &< +\infty. \end{aligned}$$

- A5: For some  $\delta > 0$

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[ \sup_{\|\theta - \theta_0\| < \delta} \left\| \frac{\partial}{\partial \theta_i} \ln g_\theta(i, Y_0) \right\|^2 \right] &< +\infty \\ \mathbb{E}_{\theta_0} \left[ \sup_{\|\theta - \theta_0\| < \delta} \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln g_\theta(i, Y_0) \right\|^2 \right] &< +\infty \\ \int \sup_{\|\theta - \theta_0\| < \delta} \left\| \frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} \ln g_\theta(i, y) \right\| \mu(dy) &< +\infty. \end{aligned}$$

- A6:  $k(y) := \sup_{\|\theta - \theta_0\| < \delta} \max_{i, j \in E} \frac{g_\theta(i, y)}{g_\theta(j, y)}$  for some  $\delta > 0$ , and  $\mathbb{P}_{\theta_0}(k(Y_0) = +\infty | X_0 = i) < 1$ , for any  $i \in E$ .

### 3.5 Asymptotic properties

Let  $(X, Y)$  be the two-dimensional process as defined in Definition 1 in Sect. 2. Let  $\mathbf{A} = (a_{ij})$  denote the generating matrix of  $X_t$ ,  $t \geq 0$ , with state space  $E = \{1, 2, \dots, d\}$ .

Consider the related Markov chains:

**Scenario 1:**  $\hat{X}_n = X_{nh}$ , with transition probability matrix  $Q_h = e^{\mathbf{A}h}$ .

**Scenario 2:**  $Z_n = X_{T_n}$  where  $T_n$ ,  $n \geq 1$ , are the arrival times of a  $HPP(\lambda)$ , with  $\lambda \geq \max\{a_i, i = 1, \dots, d\}$ , and  $T_0 = 0$ , where  $a_i := -a_{ii}$ . Its transition probability is  $\mathbf{Q} = \mathbf{I} + \lambda^{-1}\mathbf{A}$ .

- Lemma 1** (i) *If the Markov process  $X$  is irreducible then the above Markov chains,  $\hat{X}$  and  $Z$ , are irreducible and aperiodic, i.e., ergodic.*  
 (ii) *The Markov process  $X$ , and the Markov chains  $\hat{X}$  and  $Z$  have the same stationary probability, say  $\pi$ .*  
 (iii) *If the Markov process  $X$  is stationary, then the Markov chains  $\hat{X}$  and  $Z$  are also stationary, and the sequence  $\hat{Y}$  is stationary with the stationary distribution  $\pi G(\cdot) = \sum_{j \in E} \pi_j G(j, \cdot)$  in both scenarii.*  
 (iv) *In our case M1 – M0, if  $\{X_n\}$  is ergodic, the Markov chain  $(X_n, Y_n)$  is ergodic with stationary probability  $\tilde{\pi}(j, B) = \pi_j G(j, B)$ , with  $j \in E$  and  $B \in \mathcal{B}_Y$ , Borel sets in  $\mathcal{Y}$ .*

**Proof** (i) By construction the  $\hat{X}$  and  $Z$  are irreducible too as  $X$ . Moreover they are aperiodic since  $Q_h(i, i) > 0$  and  $Q(i, i) > 0$ .

- (ii) The stationary probability  $\pi$  of the MP  $X$  is given by the solution of the equation  $\pi e^{\mathbf{A}h} = \pi$  which is the same as for  $\hat{X}$ . Moreover this is the same for  $Z$  too (see, e.g., Ross 1996).  
 (iii) If the MP  $X$  is stationary, it means that  $\mathbb{P}(X_t = j) = \pi_j$ , for all  $t \geq 0$  and all  $j \in E$ , and we have  $\mathbb{P}(X_{kh} = j) = \mathbb{P}(\hat{X}_k = j) = \pi_j$ , so that  $\hat{X}$  is stationary too. The same applies for  $Z$ .

(iv) The transition probability for the Markov process  $(X_t, Y_t)$  is obtained by the limit of the transition probability. We showed in Sect. 2.2 that the transition probabilities for the process  $(X_t, Y_t)$  are given by  $\mathbb{P}(X_t = j, Y_t \in B | X_0 = i, Y_0 = y) = P_{ij}(t)G(j, B)$ , for  $i, j \in E, B \in \mathcal{B}_Y$  and  $y \in \mathcal{Y}$ . Also we proved in Sect. 2.2 that  $\tilde{\pi}(j, B) = \pi_j G(j, B)$ . □

### 3.6 Scenario 1. Regular inspections in time

In this section we establish asymptotic properties of the estimators. Let  $h > 0$ , and let  $\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n$  observations of a HMM as described in Sect. 3.2 for Scenario 1.

We extend the sequence  $\{Y_n, n \geq 0\}$  to the sequence  $\{Y_n, n \in \mathbb{Z}\}$ .

#### Notation

1. Likelihood:

$$p_\theta(\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n) = \sum_{(i_0, \dots, i_n) \in E^{n+1}} \pi_\theta(i_0) \prod_{k=1}^n P_\theta(i_{k-1}, i_k) \prod_{l=0}^n G_\theta(i_l, \hat{Y}_l).$$

2. Fisher information matrix:  $I_n(\theta_0) = -E_{\theta_0} \left[ \frac{\partial^2 \ln p_\theta(\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta_0} \right]_{ij}$ .

3. Asymptotic Fisher information matrix;  $I(\theta_0) = -E_{\theta_0} \left[ \frac{\partial^2 \ln \mathbb{P}_\theta(\hat{Y}_0 | \hat{Y}_{-1}, \hat{Y}_{-2}, \dots)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\theta_0} \right]_{ij}$ . This matrix is nonsingular provided that there exists an integer  $n \in \mathbb{N}$  such that  $I_n(\theta_0)$  is nonsingular.

Under assumptions A1, A2, A3, A4, Leroux (1992) proves that the maximum-likelihood estimator is strongly consistent in euclidean norm. Given consistency and under assumptions A1, A3, A5, A6, Bickel et al. (1998) prove asymptotic normality. Specifically they prove the following theorem.

**Theorem 1** (Bickel et al. (1998). Theorem 1.) *Assume that A1, A3, A5, A6 hold, that the maximum-likelihood estimator  $\hat{\theta}_n$  is consistent and that  $I(\theta_0)$  is nonsingular. Then*

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma_0),$$

as  $n \rightarrow +\infty$ , with  $\Sigma_0 = I(\theta_0)^{-1}$ .

Now for the processes  $X_t, Y_t$  in continuous time, we get the following results.

**Proposition 4** (Consistency) *Let  $h > 0$  be fixed. Under assumptions A1–A4, given a sample of observations  $\{\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n\}$ , the estimator given in (7) is uniformly strongly consistent, that is*

$$\max_{i, j \in E} |\hat{\mathbf{A}}_{h,n}(i, j) - \mathbf{A}_h(i, j)| \xrightarrow{a.s.} 0, \text{ as } n \rightarrow \infty.$$

**Proof** Uniform consistency of  $\hat{\theta}_n$  implies uniform consistency of  $\hat{\mathbf{Q}}_{h,n}$ , since  $\hat{\mathbf{Q}}_{h,n} = Q(\hat{\theta}_n)$  is a continuous transformation of  $\hat{\theta}_n$ , and so is  $\hat{\mathbf{A}}_{h,n}$ , since we define  $\hat{\mathbf{A}}_{h,n} = A(\hat{\theta}_n) = h^{-1} (Q(\hat{\theta}_n) - \mathbf{I})$ , then  $\hat{\mathbf{A}}_{h,n}$  is a uniform consistent estimator of  $\mathbf{A}_h$ . □

**Proposition 5** (Asymptotic normality) *Under assumptions A1, A3, A5, A6, given a sample of observations  $\{\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n\}$ , where for each  $k \geq 0$ ,  $\hat{Y}_k = Y_{kh}$ , for  $h > 0$  fixed, the random matrix  $\sqrt{n}(\hat{\mathbf{A}}_{h,n} - \mathbf{A}_h)$  is asymptotically Normal, as  $n \rightarrow \infty$ , with mean  $\mathbf{0}$  and variance-covariance matrix  $\Sigma_{\mathbf{A}} = \nabla A^\top \Sigma_0 \nabla A$ , with  $\nabla A$  the gradient of function  $A(\theta) = \mathbf{A}$  evaluated at  $\theta_0$ .*

**Proof** For a fixed  $h > 0$ , and  $T = nh$ , we have, as  $T \rightarrow \infty$ ,  $n \rightarrow \infty$ , and we can use Theorem 1 of Bickel et al. (1998).

Let  $\mathbf{A}$  be the generating matrix of  $\{X_t\}$ , with  $\mathbf{A} = A(\theta)$ ; then for  $h > 0$  fixed, define  $\mathbf{Q}_h = \mathbf{A}h + \mathbf{I}$ , then  $\mathbf{Q}_h = Q(\theta) = A(\theta)h + \mathbf{I}$ , and  $\nabla Q = h \nabla A$ . Then

$$\sqrt{n}(\hat{\mathbf{Q}}_{h,n} - \mathbf{Q}_h) \rightarrow N(0, \Sigma_{\mathbf{Q}_h}),$$

with  $\Sigma_{\mathbf{Q}_h} = h^2 \Sigma_{\mathbf{A}} = h^2 \nabla A^\top \Sigma_0 \nabla A$ .

Now, we define  $\mathbf{A}_h = \frac{\mathbf{Q}_h - \mathbf{I}}{h}$ , for  $h > 0$  fixed. The estimator of the matrix  $\mathbf{A}_h$  defined in (7) leads to

$$\sqrt{n}(\hat{\mathbf{A}}_{h,n} - \mathbf{A}_h) = \frac{\sqrt{n}}{h} (\hat{\mathbf{Q}}_{h,n} - \mathbf{Q}_h)$$

and then, again using the Delta method, we obtain  $\Sigma_{\mathbf{A}} = h^{-2} \Sigma_{\mathbf{Q}_h} = \nabla A^\top \Sigma_0 \nabla A$ , which does not depend on  $h$ . □

In this scenario, it is important to check the accuracy, in terms of the value of  $h$ , of the approximation of  $\mathbf{A}$  by  $\mathbf{A}_h$ . The following proposition proves that  $\hat{\mathbf{A}}_{h,n}$  converges uniformly to  $\mathbf{A}$ , when  $h$  goes to 0 and  $n$  goes to  $\infty$ , for fixed  $T$ .

**Proposition 6** *Under assumptions A1–A4, given a sample of observations  $\{\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_n\}$ , it is true that:*

(a)  $\hat{\mathbf{A}}_{h,n}$  is an estimator of  $\mathbf{A}$  uniformly strongly consistent. That is:

$$\max_{i,j \in E} |\hat{\mathbf{A}}_{h,n}(i, j) - \mathbf{A}(i, j)| \xrightarrow{a.s.} 0, \text{ as } n \rightarrow \infty, h \rightarrow 0,$$

(b)  $\sqrt{n}(\hat{\mathbf{A}}_{h,n} - \mathbf{A}) \xrightarrow{d} N(0, \Sigma_{\mathbf{A}})$

**Proof** For any  $h > 0$ , as  $\mathbf{A}_h = \frac{1}{h}(\mathbf{Q}_h - \mathbf{I})$ , strong consistency is deduced because  $\hat{\mathbf{A}}_{h,n}$  is a continuous transformation of  $\hat{\theta}$ , which gives us (a).

For (b) we have

$$\sqrt{n}(\hat{\mathbf{A}}_{h,n} - \mathbf{A}) = \sqrt{n}(\hat{\mathbf{A}}_{h,n} - \mathbf{A}_h) + \sqrt{n}(\mathbf{A}_h - \mathbf{A}). \tag{9}$$

From Proposition 5,  $\sqrt{n}(\hat{\mathbf{A}}_{h,n} - \mathbf{A}_h) \xrightarrow{d} N(0, \Sigma_{\mathbf{A}})$ , for any  $h$  as  $n \rightarrow +\infty$  and  $T \rightarrow +\infty$ . We can take  $h_n = (n \ln n)^{-1/2}$  in order that  $T = nh_n \rightarrow +\infty$  as  $n \rightarrow +\infty$  so we can apply Proposition 5 in the first term of the right hand side of (9). On the other hand,  $\mathbf{Q}_h = e^{\mathbf{A}h}$ , and, by the definition of  $\mathbf{A}_h$ , we have that  $\mathbf{A}_{h_n} - \mathbf{A} = \frac{\mathbf{A}^2}{2}h_n + o(h_n)$ , then, for the second term in the right hand side of (9) we have that  $\sqrt{n}(\mathbf{A}_{h_n} - \mathbf{A}) = \frac{\mathbf{A}^2}{2} \frac{1}{\sqrt{\ln n}} + o(1/\sqrt{\ln n})$ , and the conclusion follows from the uniqueness of the limit. □

### 3.6.1 Applications

Proposition 6 allows us to prove the almost sure convergence of the plug-in estimators of functionals of type  $H(t) = \Phi(\mathbf{A}, t)$ . That is, let  $h > 0$ , then we can define the  $H_h(t) = \Phi(\mathbf{A}_h, t)$  as  $H$  but based on  $\mathbf{A}_h$ . If a sample of size  $n$  of the HMM is available, we can define the plug-in estimator of  $H_h(t)$  and deduce the properties of this estimator as above. Moreover, we have that

$$\widehat{H}_{h,n}(t) \xrightarrow{a.s.} H(t), \text{ as } n \rightarrow \infty, h \rightarrow 0$$

and also

$$\sqrt{n}(\widehat{H}_n - H) \xrightarrow{d} N(0, \Sigma_H), \text{ as } n \rightarrow \infty, h \rightarrow 0.$$

As an example let us consider the reliability or survival function. Let  $\mathbf{A}_0$  and  $\alpha_0$  the restrictions of  $\mathbf{A}$  and  $\alpha$  on the up states. Then we can write the reliability formula

$$R(t) = \alpha_0 e^{\mathbf{A}_0 t} \mathbf{1},$$

and in discrete time, for  $t = nh$ ,

$$R_h(t) = \alpha_0 (\mathbf{Q}_{h,0})^n \mathbf{1},$$

where  $\mathbf{Q}_{h,0}$  is the restriction of  $\mathbf{Q}_h$  on the up states, and  $\mathbf{1}$  is a column-vector of 1 s with the appropriate dimension.

**Proposition 7** For any arbitrary but fixed  $t > 0$ , such that  $t = nh$ , we have

$$R_h(t) \longrightarrow R(t), \text{ as } n \longrightarrow \infty (h \rightarrow 0).$$

**Proof** This comes from the well known result

$$(\mathbf{Q}_{h,0})^n = (\mathbf{I} + n^{-1}\mathbf{A}_0 + o(h))^n \xrightarrow[n \rightarrow \infty]{} e^{\mathbf{A}_0 t}.$$

□

### 3.7 Scenario 2. Random inspections in time

In this section we establish asymptotic properties for  $T \rightarrow +\infty$ .

**Proposition 8** (Consistency) Under assumptions A1-A4, given a sample of observations  $\{N(T), \hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_{N(T)}\}$ , the estimator given in (8) is strongly consistent, that is  $\hat{A}_T \rightarrow A$  (a.s.), as  $T \rightarrow +\infty$ .

**Proof** We have that  $\hat{\lambda} = \frac{N(T)}{T} \xrightarrow[T \rightarrow +\infty]{} \lambda$ , a.s. (see for example Ross 1996). Also from the results in Gamiz et al. (2023) we have that  $\widehat{\mathbf{Q}}_T \rightarrow \mathbf{Q}$ . Then using Slutsky theorem (Gut 2013), we get

$$\widehat{\mathbf{A}}_T = \hat{\lambda}_T (\widehat{\mathbf{Q}}_T - \mathbf{I}) \xrightarrow[T \rightarrow +\infty]{} \lambda(\mathbf{Q} - \mathbf{I}) = \mathbf{A}; \quad (a.s.).$$

□

**Proposition 9** (Asymptotic normality) *Under assumptions A1, A3, A5, A6, given a sample of observations  $\{N(T), \hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_{N(T)}\}$ , the random matrix  $\sqrt{T}(\hat{\mathbf{A}} - \mathbf{A})$  is asymptotically Normal as  $T \rightarrow +\infty$ , with mean  $\mathbf{0}$  and variance-covariance matrix  $\Sigma_{\mathbf{A}} = \Sigma_1 + \Sigma_2$ , where  $\Sigma_1 = \lambda \nabla \mathbf{Q}^\top \Sigma_0 \nabla \mathbf{Q}$ , and  $\Sigma_2 = \bar{\mathbf{B}} \bar{\mathbf{B}}^\top$ , where  $\mathbf{B} = \sqrt{\lambda}(\mathbf{Q} - \mathbf{I})$  and  $\bar{\mathbf{B}}$  is a vector representation of the matrix  $\mathbf{B}$ .*

**Proof** We can write

$$\sqrt{T}(\hat{\mathbf{A}} - \mathbf{A}) = \sqrt{T} \hat{\lambda}(\hat{\mathbf{Q}}_T - \mathbf{Q}) + \sqrt{T}(\hat{\lambda} - \lambda)(\mathbf{Q} - \mathbf{I}). \quad (10)$$

To check the convergence to a Normal distribution we consider the two terms of expression (10) separately. First, we have

$$\sqrt{T} \hat{\lambda}(\hat{\mathbf{Q}}_T - \mathbf{Q}) = \sqrt{\frac{N(T)}{T}} \sqrt{N(T)}(\hat{\mathbf{Q}}_T - \mathbf{Q}), \quad (11)$$

where we use that  $\hat{\lambda} = N(T)/T$ .

From the results in Bickel et al. (1998) and the fact that  $N(T) \rightarrow \infty$  (a.s.) as  $T \rightarrow \infty$ , we have that

$$\sqrt{N(T)}(\hat{\mathbf{Q}}_T - \mathbf{Q}) \xrightarrow{d} N(0, \Sigma_{\mathbf{Q}}), \quad \text{as } T \rightarrow \infty,$$

with variance-covariance matrix  $\Sigma_{\mathbf{Q}}$  obtained by the Delta method applied to the function  $Q(\theta) = \mathbf{Q}$ , similar to proposition 5.

Using that  $N(T)/T \rightarrow \lambda$ , almost surely, as  $T \rightarrow +\infty$ , we get that the first term of the sum converges to a Normal distribution, that is

$$\sqrt{T} \hat{\lambda}(\hat{\mathbf{Q}}_T - \mathbf{Q}) \xrightarrow{d} N(0, \lambda \Sigma_1),$$

with  $\Sigma_1 = \lambda \Sigma_{\mathbf{Q}}$ , a matrix of dimension  $d^2 \times d^2$ .

On the other hand,  $N(T)$  has Poisson distribution with mean  $\lambda T$ , which can be approximated by a Normal distribution with both mean and variance equal to  $\lambda T$ , that is

$$\sqrt{T}(\hat{\lambda}_T - \lambda) \xrightarrow{d} N(0, \lambda).$$

Then, the second term in the sum is also asymptotically Normal with variance-covariance matrix  $\Sigma_2 = \bar{\mathbf{B}} \bar{\mathbf{B}}^\top$ , where  $\mathbf{B} = \sqrt{\lambda}(\mathbf{Q} - \mathbf{I})$ , that is

$$\sqrt{T}(\hat{\lambda}_T - \lambda)(\mathbf{Q} - \mathbf{I}) \xrightarrow{d} N(0, \Sigma_2).$$

Specifically  $\bar{\mathbf{B}}$  is a  $d^2$ -dimensional vector obtained by stacking the columns of  $\mathbf{B}$  into a column vector, then  $\Sigma_2$  is also a matrix of dimension  $d^2 \times d^2$ , and then  $\Sigma = \Sigma_1 + \Sigma_2$  is well defined.

□

## 4 Applications

### 4.1 Reliability for CTHMM with discrete state space

In this section we consider that  $Y$  takes values in a finite set, that is  $\mathcal{Y} = \{y_1, \dots, y_s\}$ , then the corresponding two-dimensional process  $(X_t, Y_t)$  has a finite number of states.

We can write the set of states for the two-dimensional process as  $\tilde{E} = E \times \mathcal{Y}$ , with  $\tilde{E} = \{(1, y_1), \dots, (d, y_1), (1, y_2), \dots, (d, y_2), \dots, (1, y_s), (2, y_s), \dots, (d, y_s)\}$ .

Then  $\{(X_t, Y_t); t > 0\}$  is a two-dimensional continuous-time Markov chain with state-space  $\tilde{E} = E \times \mathcal{Y}$  and transition matrix  $\tilde{\mathbf{P}}$  with elements

$$\tilde{P}_t((i, y'), (j, y)) = P_{ij}(t)G(j, y),$$

for  $i, j \in E$  and  $y', y \in \mathcal{Y}$ . In matrix form, we have

$$\tilde{\mathbf{P}}(t) = \left( \begin{array}{c|c|c|c} \mathbf{B}_1(t) & \mathbf{B}_2(t) & \cdots & \mathbf{B}_s(t) \\ \hline \mathbf{B}_1(t) & \mathbf{B}_2(t) & \cdots & \mathbf{B}_s(t) \\ \hline \cdots & \cdots & \cdots & \cdots \\ \hline \mathbf{B}_1(t) & \mathbf{B}_2(t) & \cdots & \mathbf{B}_s(t) \end{array} \right)$$

where for each  $k = 1, \dots, s$ , the corresponding block is a  $d \times d$  sub-matrix  $\mathbf{B}_k(t) = \mathbf{P}(t) \cdot \text{diag}(G_k)$ , where  $\text{diag}(G_k)$  is a  $d$ -dimensional diagonal matrix with the  $k$ th column of the matrix  $\mathbf{G}$ . The generator  $\mathbf{A}$  is a matrix of dimension  $(d \cdot s) \times (d \cdot s)$  with elements  $\tilde{a}((i, y')(j, y))$  as detailed in (2).

Following similar arguments as in Gamiz et al. (2023), we consider that the state-space of the process  $X$  is split into two subsets  $U := \{1, \dots, r\}$ , the working states, and  $D := \{r + 1, \dots, d\}$ , the down states. Additionally, the system up states can be defined not only by  $U \subset E$  but also by some subset of  $\mathcal{Y}$ . Then we consider also a partition in the set of possible observations, that is  $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$ , where in  $\mathcal{Y}_1$  we consider indicators of good performance of the system, and in  $\mathcal{Y}_2$  we consider the indicators warning of some serious problem in the system.

Let us denote  $\tau$  the first time the system visits the set of down states  $D$ , i.e., the hitting time of set  $D$ . Let us consider  $\tilde{U} = U \times \mathcal{Y}_1$  and  $\tilde{D} = \tilde{E} \setminus \tilde{U}$ , being  $\tilde{E} = E \times \mathcal{Y}$ . Then  $\tau = \inf\{t > 0 : \tilde{X} = (X_t, Y_t) \in \tilde{D}\}$ . Therefore the reliability of the system can be defined as  $\tilde{R}(t) = \mathbb{P}(\tau > t)$ , for  $t \geq 0$ . Conditioning on the initial state  $(i, y) \in \tilde{U} = U \times \mathcal{Y}_1$ , we write

$$\begin{aligned} \tilde{R}_{(i,y)}(t) &= \mathbb{P}_i(\tau > t) = \mathbb{P}_i(X_s \in U, Y_s \in \mathcal{Y}_1, 0 < s \leq t) \\ &= \mathbb{P}_i((X_t, Y_t) \in \tilde{U}, 0 < s \leq t), \end{aligned} \tag{12}$$

and then

$$\tilde{R}(t) = \sum_{(i,y) \in \tilde{U}} \tilde{R}_{(i,y)}(t),$$

for  $t > 0$ . Using matrix notation, we can write

$$\tilde{R}(t) = \tilde{\alpha} e^{\tilde{\mathbf{A}}t} \mathbf{1}_{\tilde{U}}$$

where  $\tilde{\mathbf{A}}_{\tilde{U}}$  denotes the sub-matrix of  $\tilde{\mathbf{A}}$  with all transition rates among states of subset  $\tilde{U}$ .

### 4.2 Reliability for CTHMM with a general state space

When the  $Y$  process takes values in a finite set, the formula in (12) can be applied because the generator  $\mathbf{A}_{\tilde{U}}$  is a matrix, but when  $Y$  takes values in a subset of  $\mathbb{R}^q$ ,  $\mathbf{A}_{\tilde{U}}$  is an operator and we can not make use of the formula in (12). In that case we propose to work by Markov renewal equations (in the semi-Markov way) in this section. Let us define  $H(i, t) = e^{-a_i t} G(i, \mathcal{Y}_1)$ , for  $i \in U$ , and let  $\Psi_U$  be the Markov renewal function corresponding to the sub semi-Markov kernel  $L_U(i, j, t) = \frac{a_{ij}}{a_i} (1 - e^{-a_i t})$ , for  $i, j \in U$ . The following result can be proved similarly to Proposition 2.

**Proposition 10** (Reliability) *The conditional reliability function  $\tilde{R}_{(i,y)}(t)$  satisfies the Markov renewal equation (MRE)*

$$\tilde{R}(i, t) = H(i, t) + \sum_{j \in U} \int_0^t \tilde{R}(k, t - s) L(i, k, ds)$$

Therefore, the conditional reliability is given by the only solution of the above equation, i.e.,

$$\tilde{R}(i, t) = (\Psi_U * H)(i, t)$$

## 5 Numerical examples

In this section we illustrate the two scenarios discussed in the rest of the paper. In the first example, a real dataset is considered whereas in the second example a simulation study is carried out. In both cases, we discretize the continuous-time problem, and then estimate using our algorithms in Gamiz et al. (2023), and finally we establish estimator properties in continuous-time, our initial problem.

### 5.1 Scenario 1. Regular inspections in time

As an illustration we consider a comparison study of the suicide-rate in the US and Japan during the period 1985–2015. The data have been taken from the data platform <https://www.kaggle.com/>. We focus on the following variables: the *Suicide Rate*, which is measured as the number of suicides/100k population, and, the *Gross Domestic Product per capita* (*GDP\_per\_capita*).

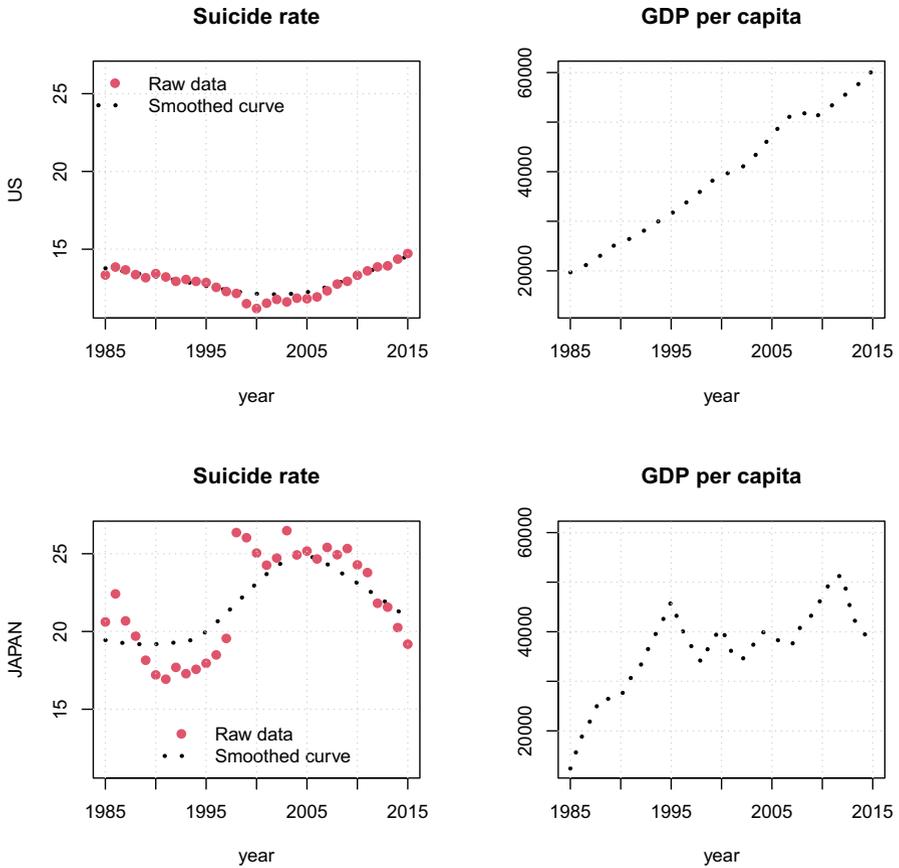
#### 5.1.1 Preliminary

Although the data registry provides information from a total of 94 countries we limit ourselves to the US and Japan. We have chosen two of the most developed countries in the world with the aim of comparing them with respect to the number of suicides per year.

In Fig. 1 we give a graphical description of the situation. On the top panel we describe the case of US while on the bottom panel we represent the information from Japan.

Looking at the figure from left to right we have the following. The plots on the left display the suicide-rate per year for US (top) and Japan (bottom) from 1995 to 2015, respectively. The suicide-rate has been calculated as the number of suicides registered in the country every year per 100,000 inhabitants. In each graph it is shown the value calculated every year as well as a smoothed curve obtained from these values and that helps in visualizing the tendency of the suicide-rate along the period of observation. As we can see from the curves, not only suicides occur in Japan at a significantly higher rate than in US, also it is remarkable the existence of a high variability in the rate of suicides in Japan over the years, which is not appreciated in the case of US where the tendency is more steady.

On the right-side panels, a scatterplot is given representing the GDP per capita from 1995 until 2015 in US (top) and Japan (bottom). What catches our attention is that again the situation in US seems to be more stable and suggests an increasing trend, with an almost constant slope, of the population standard of living in the US. We can see just one abrupt decay coinciding with the period of the economical crisis around 2008. However the GDP curve in



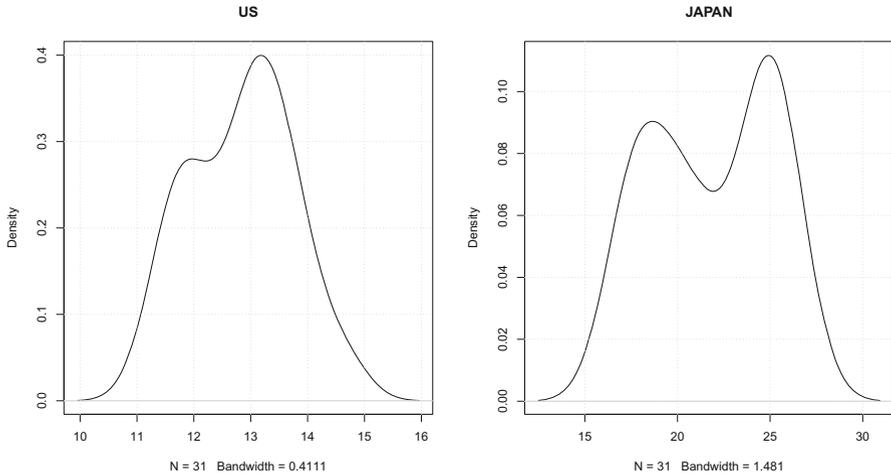
**Fig. 1** Overview of the situation in the US (top panel) and Japan (bottom panel) during the period 1995–2015

Japan, although suggests a slightly increasing tendency from 1995 onwards, it shows high variability at the same period of time when the highest suicide rate is seen in the corresponding plot (on the left panel).

Figure 2 shows smooth density estimations of the rate of suicides all along the observation period for both countries, US on the left panel, and Japan on the right one. As we can see, the curves in both cases seem to suggest two different “regimes” (hidden or latent) in the country. These regimes could be the result of the interaction of the country’s wealth (measured in terms of the GDP) and possibly many other intrinsic factors which in conjunction can explain, to a certain extent, the observed rate of suicides at any particular time.

With this in mind, we propose to study this phenomenon using a CTHMM with the following specifications.

- $\{X(t); t > 0\}$  is the CTMC that represents the internal regime that is not directly observable.  
Let us assume that  $X$  takes values in  $E = \{1, 2\}$ ;
- $\{Y(t), t > 0\}$  is the observable process. We define  $Y(t)$  as the rate of suicides at time  $t$ . Considering the plots in Fig. 2, we assume that  $Y(t)$  conditioned to the event  $\{X(t) = i\}$  follows a Normal distribution with mean  $\mu_i$  and standard deviation  $\sigma_i$ , for  $i \in \{1, 2\}$ ;



**Fig. 2** Density curve of suicide rate in the US (left) and in Japan (right)

We do not have a continuous-time follow-up of the process. We rather have annual observations, so we use the methodology presented in this paper and estimate the model considering Scenario 1 where observations arrive regularly in time, specifically with  $h = 1$ . Regarding the DTHMM for each case study (US and Japan), the following parameters have to be estimated

$$\theta = (Q_h(1, 1), Q_h(2, 1), \mu_1, \mu_2, \sigma_1, \sigma_2),$$

where  $Q_h(i, 1) = \mathbb{P}(X_h = 1|X_0 = i)$ , for  $i \in \{1, 2\}$ . Then an estimator  $\hat{\mathbf{A}}_h$  is obtained as explained in Sect. 3.1.

### 5.1.2 Results

We use the EM-Algorithm. The function  $Q(\theta|\theta^{(m)}) = \mathbb{E}_{\theta^{(m)}}[\ln f(\hat{X}, \hat{Y}|\theta)]$  will give us by successive iterations an approximation of the estimate of  $\theta$ , where  $\ln f(\hat{X}, \hat{Y}|\theta)$  is the log-likelihood corresponding to the complete dataset, which would be obtained from the bivariate process  $(\hat{X}, \hat{Y})$ .

$$\begin{aligned}
 Q(\theta|\theta^{(m)}) &= \sum_{n=1}^{N-1} \sum_{i,j \in E} \mathbb{P}_{\theta^{(m)}}(\hat{X}_n = i, \hat{X}_{n+1} = j|\hat{Y}) \ln P_{\theta}(i, j) \\
 &\quad + \sum_{n=1}^N \sum_{i \in E} \mathbb{P}_{\theta^{(m)}}(\hat{X}_n = i|\hat{Y}) \sum_{k=1}^K \ln g(\hat{Y}_{kn}; \phi_i). \tag{13}
 \end{aligned}$$

• **E-Step**

For given  $\theta^{(m)}$ , compute the probabilities:

$$\mathbb{P}_{\theta^{(m)}}(\hat{X}_n = i, \hat{X}_{n+1} = j|\hat{Y}), \quad n \in \{1, \dots, N - 1\}; \quad i, j \in E$$

and

$$\mathbb{P}_{\theta^{(m)}}(\hat{X}_n = i|\hat{Y}), \quad n \in \{1, \dots, N\}; \quad i \in E.$$

**Table 1** Estimated parameters of the hidden model

United States (US)	Japan
$\widehat{\mathbf{Q}}_h = \begin{pmatrix} 0.915 & 0.085 \\ 0.052 & 0.948 \end{pmatrix}$	$\widehat{\mathbf{Q}}_h = \begin{pmatrix} 0.941 & 0.059 \\ 0.072 & 0.928 \end{pmatrix}$
$G(1, \cdot) \rightarrow N(\widehat{\mu}_1 = 11.90, \widehat{\sigma}_1 = 0.437)$	$G(1, \cdot) \rightarrow N(\widehat{\mu}_1 = 19.53, \widehat{\sigma}_1 = 2.340)$
$G(2, \cdot) \rightarrow N(\widehat{\mu}_2 = 13.41, \widehat{\sigma}_2 = 0.556)$	$G(2, \cdot) \rightarrow N(\widehat{\mu}_2 = 24.79, \widehat{\sigma}_2 = 1.047)$
$\widehat{\alpha} = (1, 0)$	$\widehat{\alpha} = (0, 1)$
$\widehat{\mathbf{A}} = \begin{pmatrix} -0.0854 & 0.0854 \\ 0.052 & -0.052 \end{pmatrix}$	$\widehat{\mathbf{A}} = \begin{pmatrix} -0.059 & 0.059 \\ 0.072 & -0.072 \end{pmatrix}$
$\widehat{\pi} = (0.3789, 0.6211)$	$\widehat{\pi} = (0.5514, 0.4486)$

To solve this step we use of the forward-backward probabilities defined as in Gamiz et al. (2023).

• **M-Step**

Update  $\theta^{(m)}$  to  $\theta^{(m+1)}$ . The maximization step M is realized directly by

$$\widehat{P}^{(m+1)}(i, j) = \frac{\sum_{n=1}^N \mathbb{P}_{\theta^{(m)}}(\widehat{X}_n = i, \widehat{X}_{n+1} = j | \widehat{\mathbf{Y}})}{\sum_{n=1}^N \mathbb{P}_{\theta^{(m)}}(\widehat{X}_n = i | \widehat{\mathbf{Y}})}. \tag{14}$$

Considering that the emission function follows a Normal law,  $g_\theta(i, y) = (\sigma_i \sqrt{2\pi})^{-1} \exp(-(y - \mu_i)^2 / (2\sigma_i^2))$ , for  $i = 1, 2$ . The optimal value of the second term  $Q_2((\phi_i; i \in E) | \theta^{(m)})$  is obtained for

$$\widehat{\mu}_i = \frac{\sum_{n=1}^N \mathbb{P}_{\theta^{(m)}}(\widehat{X}_n = i | \widehat{\mathbf{Y}}) \widehat{Y}_n}{\sum_{n=1}^N \mathbb{P}_{\theta^{(m)}}(\widehat{X}_n = i | \widehat{\mathbf{Y}})};$$

and,

$$\widehat{\sigma}_i^2 = \frac{\sum_{n=1}^N \mathbb{P}_{\theta^{(m)}}(\widehat{X}_n = i | \widehat{\mathbf{Y}}) (\widehat{Y}_n - \widehat{\mu}_i)^2}{\sum_{n=1}^N \mathbb{P}_{\theta^{(m)}}(\widehat{X}_n = i | \widehat{\mathbf{Y}})},$$

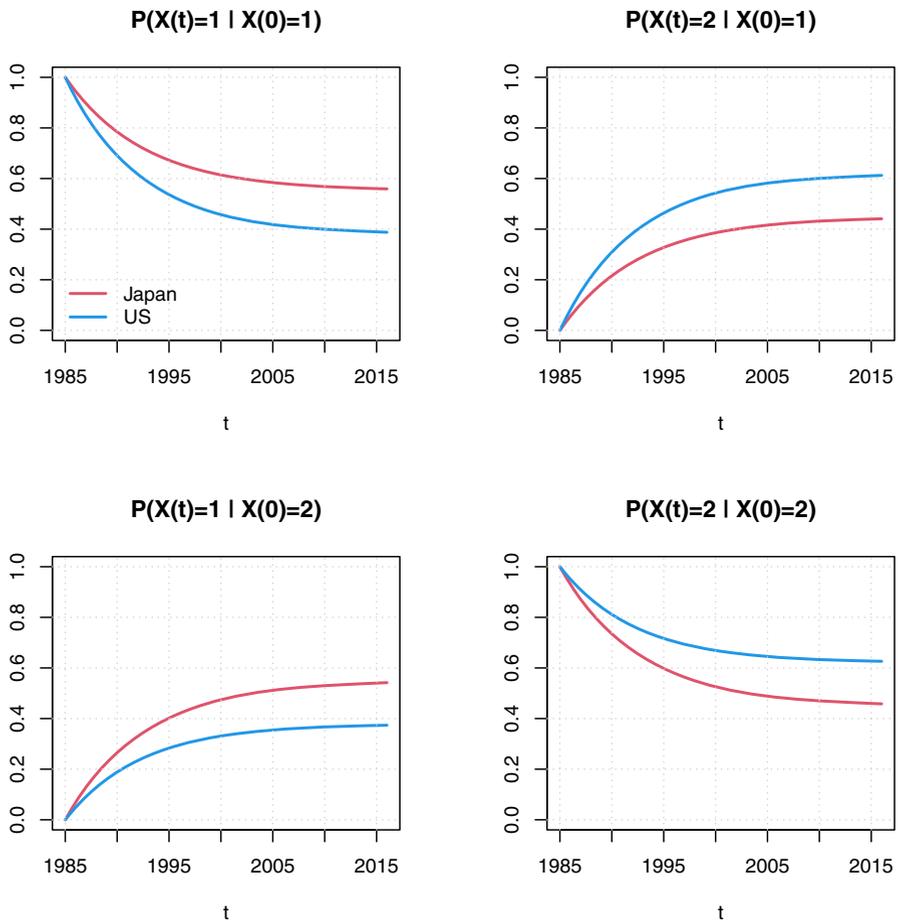
for  $i \in \{1, 2\}$ .

We have estimated the model in both cases, i.e., with data from US as well as Japan. The estimated values reported in Table 1 are the following: the transition matrix  $\mathbf{Q}_h$ ; the parameters of the emission law,  $\{G(i, \cdot) \rightarrow N(\mu_i, \sigma_i); i = 1, 2\}$ ; the initial distribution,  $\alpha$ ; the generating matrix  $\mathbf{A}$  and the stationary distribution of the continuous MC  $\pi$ .

Figure 3 represents the transition probability functions between hidden states.

**5.2 Scenario 2. Random inspections in time**

To illustrate this approach we present a simulation study. In particular, we consider a system with four possible states, that is  $E = \{1, 2, 3, 4\}$ , with  $\{1, 2\}$  the functioning states whereas  $\{3, 4\}$  are the down states. The observed output may vary in the set  $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$ . The



**Fig. 3** Transition-function matrix of the hidden MC governing the suicide rate observed in Japan (solid red line) and in US (solid blue line)

true generating and the emission matrices are given by

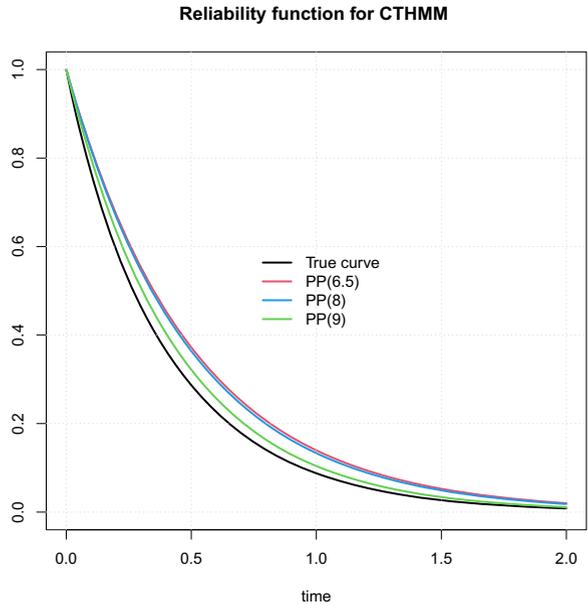
$$\mathbb{A} = \begin{pmatrix} -6 & 4 & 1 & 1 \\ 3 & -5 & 1 & 1 \\ 1 & 2 & -5 & 2 \\ 0 & 2 & 2 & -4 \end{pmatrix}, \quad \text{and}, \quad \mathbb{G} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.5 & 0.2 & 0 \\ 0.1 & 0.2 & 0.5 & 0.2 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}.$$

Besides, we assume that the system is inspected at times that follow a Poisson process with intensity  $\lambda$ . Let us assume that  $X_0 = 1$  and  $Y_0 = y_1$ .

We have simulated a total of 500 samples of size  $N = 100$  each according to the following algorithm.

1. Generate a sample trajectory  $(X_1, T_1), \dots, (X_N, T_N)$  of the MC  $X$  from the generating matrix  $\mathbb{A}$ , where  $X_n$  is the  $n$ th state visited by the process and  $T_n$  the sojourn time of the system in the  $n - 1$  for  $n \in \{1, \dots, N\}$ . Denote  $Z_n = \sum_{k=1}^n T_k$ , the successive jump times of the MC.

**Fig. 4** Scenario 2: Observations follow a Poisson process



**Table 2** Mean values of sample sizes for different values of  $\lambda$

$\lambda$	6.5	7	8	9
$\bar{N}_0$	131.1	141.2	161.0	180.8

$\bar{N}_0 = \sum_{k=1}^{500} N_0^k$   
 $N_0^k$ : total number of observations obtained at the  $k$ th repetition of the experiment

2. Generate a sample  $Y_1, \dots, Y_N$  of the process  $Y$ , following the rule  $\mathbb{P}(Y = y | X = i) = G_{i,y}$ , with  $i \in \{1, 2, 3, 4\}$  and  $y \in \mathcal{Y}$ .
3. Generate a sample trajectory of the PP with rate  $\lambda$ , that is,  $S_1 < S_2 < \dots < S_{N_0}$ . Simulate arrival times until  $S_{N_0} \geq Z_N$ .
4. Put  $\hat{Y}_0 = Y_0$ . For each  $r \in \{1, \dots, N_0\}$ , find  $n > 0$  such that  $Z_n \leq S_r < Z_{n+1}$  and define  $\hat{Y}_r = Y_n$ .

We have considered three cases,  $\lambda \in \{6.5, 8, 9\}$

Again we use the EM-algorithm, as in Gamiz et al. (2023), to fit the discrete HMM  $(\hat{X}, \hat{Y})$  and estimate the corresponding parameters  $\hat{Q}$  and  $\hat{G}$ . Also, for each repetition of the experiment, estimate  $\hat{\lambda} = \frac{n_0}{s_{n_0}}$ . Then, obtain an estimation of the generating matrix  $\hat{A}$  as explained in Sect. 3.1. Using the estimation of the emission matrix  $\hat{G}$  we construct the generating matrix of the 2-dimensional process  $(X, Y)$  and following 4.1 we can obtain the estimation of the reliability function, which is shown in Fig. 4.

Notice that  $N_0 \geq N$ , since we choose  $\lambda > \max\{-a_{ii}, i \in E\}$ . In our case we have obtained the statistics displayed in Table 2.

## 6 Conclusion

Continuous-time hidden Markov processes can be seen as a two-dimensional continuous-time Markov process  $(X_t, Y_t)$  where the first component  $X_t$  is an unobservable continuous-time Markov chain and the second one  $Y_t$  is an observable process whose distribution law depends on  $X_t$  through a function called the emission function. In this paper, we have defined the generator function corresponding to the coupled process in terms of the generating matrix of  $X_t$  and the emission function. The theoretical properties of this type of processes have been obtained using a semi-Markov formulation of the model. To estimate the characteristics of the process we have considered two different discretization schemes under which observations can arrive. On the one hand, we assume that observations arrive regularly in time, and on the other hand, we assume that observations arrive at random. In both cases, maximum-likelihood estimator of the parameters of the CTHMM have been obtained and their asymptotic properties have been proven.

To our knowledge, the approach to HMM models considered in this paper, basing on Markov renewal theory, is completely new and provides a powerful tool to get other insights in this context of HMMs.

With respect to the estimation problem, as pointed out in Liu et al. (2017), when discretization is too coarse, many transitions of the hidden states can occur between two consecutive observations and the dynamics of the hidden process might be poorly caught by the model. This situation may happen in our first scenario when the time-span between observations ( $h$ ) is not fine enough. Then, more flexible models are required and it is suitable a continuous follow-up of the process that will be considered in a future work.

In the applications discussed in this paper we have assumed that the set of hidden states is finite and the number of states is pre-specified. Generalizations of our work can be to consider selecting the optimal number of hidden states, as in Lin and Song (2022) where the number of states is unknown and to be determined by the data. Another approach that will be considered in future works, consists of defining the state space of the hidden chain, in general as a measurable set. See for example, Dorea and Zhao (2002) where kernel density estimation is discussed for the observed process in the context of DTHMM.

**Acknowledgements** This work was jointly supported by the Spanish Ministry of Science and Innovation-State Research Agency through grants numbered PID2020-120217RB-I00 and PID2021-123737NB-I00, and the IMAG-Maria de Maeztu grant CEX2020-001105-/AEI/10.13039/501100011033; and by the Spanish Junta de Andalucía through grant number B-FQM-284-UGR20.

**Funding** Funding for open access publishing: Universidad de Granada/CBUA

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was written by MLG and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Consent for publication** All the authors approve the version to be published.

**Ethical approval** All authors declare to comply with Journal Ethics.

**Consent to participate** All the authors consent to participate in this research.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Stat* 37(6):1554–1563
- Bickel PJ, Ritov Y, Ryden T (1998) Asymptotic normality of the maximum-likelihood estimator for a general hidden Markov model. *Ann Stat* 26(4):1614–1635
- Dorea CC, Zhao LC (2002) Nonparametric density estimation in hidden Markov models. *Stat Inference Stoch Process* 5:55–64
- Freed DS, Shepp LA (1982) A Poisson process whose rate is a hidden Markov process. *Adv Appl Probab* 14(1):21–36
- Gamiz ML, Limnios N, Segovia-Garcia MC (2023) Hidden Markov models in reliability and maintenance. *Eur J Oper Res* 304:1242–1255
- Girardin V, Limnios N (2018) Applied probability. From random sequences to stochastic processes. Springer, Cham
- Gut A (2013) Probability: a graduate course. Springer, Berlin
- Hulme WJ, Martin GP, Sperrin M, Casson AJ, Bucci S, Lewis S, Peek N (2021) Adaptive symptom monitoring using hidden Markov models—an application in ecological momentary assessment. *IEEE J Biomed Health Inform* 25(5):1770–1780
- Kingman JFC (1963) Ergodic properties of continuous-time Markov processes and their discrete skeletons. *Proc Lond Math Soc* 13(1):593–604
- Kulkarni VG (2011) Introduction to modeling and analysis of stochastic systems. Springer, Berlin
- Leroux BG (1992) Maximum-likelihood estimation for hidden Markov models. *Stoch Process Appl* 40:127–143
- Limnios N (2012) Reliability measures of semi-Markov systems with general state space. *Methodol Comput Appl Probab* 14:895–917
- Limnios N, Oprisan G (2001) Semi-Markov processes and reliability. Springer Science & Business Media, Berlin
- Lin Y, Song X (2022) Order selection for regression-based hidden Markov model. *J Multivar Anal* 192:105061
- Liu YY, Li S, Li F, Song L, Rehg JM (2016) Efficient learning of continuous-time hidden Markov models for disease progression. *Adv Neural Inf Process Syst* 28:3599–3607
- Liu YY, Moreno A, Li S, Li F, Song L, Rehg JM (2017) Learning continuous-time hidden Markov models for event data. In: Rehg J, Murphy S, Kumar S (eds) Mobile health. Springer, Cham. [https://doi.org/10.1007/978-3-319-51394-2\\_19](https://doi.org/10.1007/978-3-319-51394-2_19)
- Mor B, Garhwal S, Kumar A (2021) A systematic review of hidden Markov models and their applications. *Arch Comput Methods Eng* 28:1429–1448
- Ross SM (1996) Stochastic processes, 2nd edn. Wiley, Hoboken
- Sadek A, Limnios N (2002) Asymptotic properties for maximum likelihood estimators for reliability and failure rates of Markov chains. *Commun Stat Theory Methods* 31(10):1837–1861
- Sadek A, Limnios N (2005) Nonparametric estimation of reliability and survival function for continuous-time finite Markov processes. *J Stat Plan Inference* 133:1–21
- Shurenkov VM (1984) On the theory of Markov renewal. *Theory Probab Appl* 19(2):247–265
- Verma A, Powell G, Luo Y, Stephens D, Buckeridge DL (2018) Modeling disease progression in longitudinal EHR data using continuous-time hidden Markov models. *Machine Learning for Health Workshop*. [arxiv.org/1812.00528](https://arxiv.org/abs/1812.00528)
- Wei W, Wang B, Towsley D (2002) Continuous-time hidden Markov models for network performance evaluation. *Perform Eval* 49:129–146
- Zhou J, Song X, Sun L (2020) Continuous time hidden Markov model for longitudinal data. *J Multivar Anal* 179:1–16

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.