Reliability Engineering and System Safety 239 (2023) 109498

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Dynamic reliability and sensitivity analysis based on HMM models with Markovian signal process

M.L. Gámiz [*], F. Navas-Gómez, R. Raya-Miranda, M.C. Segovia-García

Department of Statistics and Operational Research, University of Granada, Spain

ARTICLE INFO

Keywords:
Double chain Hidden Markov Model
EM-algorithm
Maintenance
Sensitivity measures
Sensors

ABSTRACT

The main objective of this paper is to build stochastic models to describe the evolution-in-time of a system and to estimate its characteristics when direct observations of the system state are not available. One important application area arises with the deployment of sensor networks that have become ubiquitous nowadays with the purpose of observing and controlling industrial equipment. The model is based on hidden Markov processes where the observation at a given time depends not only on the current hidden state but also on the previous observations. Some reliability measures are defined in this context and a sensitivity analysis is presented in order to control for false positive (negative) signals that would lead to believe erroneously that the system is in failure (working) when actually it is not. System maintenance aspects based on the model are considered, and the concept of signal-runs is introduced. A simulation study is carried out to evaluate the finite sample performance of the method and a real application related to a water-pump system monitored by a set of sensors is also discussed.

1. Introduction

The advances in sensing and automated data collection from multiple sensors has given rise to the availability of large datasets providing useful information (signals or indicators) about certain aspects of the real health of industrial equipment [1,2]. The nature of this information can be very heterogeneous, producing multiple types of variables, such as, speed or vibration signals, temperature or pressure measurements, for instance. The key question is how to interpret such signals to establish a reliable diagnostic of the real level of performance of the equipment [2–4].

In summary, multi-sensors or control devices are used to provide raw signals which need to be processed to model the physical state and degradation of components. It is a very important challenge for engineers to provide methods to define the state of their equipment (broken or functioning) given the values of these indicators or signals.

In this context, different techniques have been proposed in the literature to predict the health state of a system and prevent potential failures using sensed data, one of these techniques are Hidden Markov Models (HMM), which are a powerful and popular statistical tool for modelling partially observable systems and has been successfully applied in many application areas in Engineering. In [5] a bibliographic review of HMMs and extensions is presented in general and in [6] in reliability, in particular. In Table 1 we present a short review of recent literature.

In this paper a stochastic model based on Hidden Markov processes is build to describe the evolution-in-time of a system and its properties are estimated from observations that do not directly inform the true current state of the system but provide only some indicators.

Hidden Markov models are based on a coupled process (e.g. Markov chain), say (X,Y), where X is an unobserved random sequence describing the state of the system (i.e., engine), and Y is assumed an observable random sequence, giving the values of the parameters of some indicators (i.e., vibration, pressure, temperature, etc.), whose law depends on the value of the corresponding unobserved sequence X. In order to be able to handle the above coupled process, we have to assume some particular probabilistic structure. For example, for X we can suppose that it is an i.i.d sequence or a Markov or semi-Markov chain; while for Y, usually it is thought as conditionally independent with its law depending on the corresponding value of X. In this case, if X is a Markov chain, we denote the process as M1M0-HMM [7,8].

In the present paper, we assume that the observations not only depend on the hidden state but also on the previous observation [9]. Thus, the hidden process is a Markov chain and the observed process (the measurement process) conditionally to the hidden one is also a Markov chain. Such a coupled process (X,Y) is called a double Markov chain in [10]. We prefer notation M1M1-HMM for the double Markov chain to emphasize its relation to the basic model HMM with structure M1M0. Very close to this model are the so called regime-switching

* Corresponding author.
E-mail address: mgamiz@ugr.es (M.L. Gámiz).

https://doi.org/10.1016/j.ress.2023.109498
Received 21 March 2023; Received in revised form 6 July 2023; Accepted 7 July 2023
Available online 16 July 2023

Table 1
Review of bibliography (in alphabetical order) on recent applications of HMM and related models in Reliability Engineering.

Reference	Technique	Contribution/Application
Chen et al. (2019) [9]	HMM-*AO*	Degradation modelling of manufacturing systems and a maintenance policy based on remaining useful life (RUL)
Cheng et al. (2023) [14]	*CT*-HMM	Optimization of inspection and condition based mission abort policy for a partially observed safety-critical system
Coraça et al. (2023) [15]	HMM	Building an unsupervised learning framework for vibration based health monitoring of a structure
Danisman & Kocer (2021) [16]	*BD*-HMM	Capture first-order Markov dependency with application to earthquakes and stock market
Gamiz et al. (2023a) [7]	HMM	New reliability measures and maintenance strategies based on critical probabilities
Gamiz et al. (2023b) [8]	*CT*-HMM	Discretization strategies for estimating the CT model
Ghasvarian Jahromi et al. (2023) [17]	HMM	Short-term wind speed and power forecasting
Gualeni et al. (2023) [18]	HMM	Maintenance in ships engine room, the space available around machinery and systems
Guo & Liang (2022) [19]	HMM & *MDP*	Optimization of inspection and maintenance strategies for multi-state systems
Habayeb et al. (2018) [20]	HMM	Identify at an early stage the bug reports that would require a long time to fix
Khan & Abuhasel (2021) [21]	HMM & *GA*	Detection of threat in Industrial Internet of things
Li et al. (2020) [22]	HMM	RUL prediction and estimation of suitable maintenance interval for the component of a hydro-turbine that cannot be online monitored
Lin et al. (2023) [23]	HMM & *NN*	Fit the local fluctuations in the degradation process of lithium batteries
Martindale et al. (2021) [24]	*H*-HMM	Smart annotation of cyclic data to reduce the cost of labelling data based on sensors
Soleimani et al. (2021) [25]	HMM & *BN*	Fault diagnosis and prognosis for complex, multi-state, transient, and dynamic systems such as automotive propulsion systems
Zhao et al. (2021) [26]	HMM	Description of multi-state systems degradation with data from a nuclear power plant

Notation: *AR*= Auto-regressive; *AO*=Auto-correlated Observations; *CT*=Continuous-time; *BD*=Binary Dependence; *MDP*= Markov Decision Process; *GA*= Genetic Algorithm; *NN*=Neural Network;*H*= Hierarchical; *BN*=Bayesian Network. HMM refers to the model in discrete time with dependence structure M1-M0.

models in which the system dynamic behaviours are allowed to be changed over time according to the states of an underlying Markov chain, which is also called a modulating Markov chain (see [11,12] or [13] for distinct reliability applications of this model).

In practice it is common that the observations are registered through a longitudinal study and therefore we assume that there is some correlation between the successive observed measurements (see [9,10]).

These models have recently proven great potential in applications as can be seen in [20,27], and [28], to cite some recent works. In particular in [9] a hidden Markov model with auto-correlated observations is developed to handle the degradation modelling of manufacturing systems.

Firstly, we focus on a continuous-time model to describe both the hidden-state process and the observed-indicator process. However, we assume that observations are registered regularly in time, which is a quite realistic situation in practice. This means that observations can be treated as realizations of a discrete-time model. So we also present the model in discrete-time.

For the continuous-time model, we assume that $X = \{X_t; t \geq 0\}$ is a Markov chain taking values in a finite set E, where transitions between states are given by an unknown generating matrix $\mathbf{A}$; and, that $Y = \{Y_t; t \geq 0\}$ is also a Markov chain whose distribution at time t depends on the state of X_t, that is, given the event $\{X_t = i\}$, the generating matrix governing the transitions of Y is $\mathbf{B}_i$, for each $i \in E$. We consider the case where Y takes values in a finite set $\mathcal{Y}$.

The problem here is to estimate the generating matrix of the Markov chain X, i.e. $\mathbf{A}$, as well as the set of generating matrices $\{\mathbf{B}_i, i \in E\}$, defined above. To do it, we will use the discretization approach presented in [8]. The estimation of the characteristics of the discrete-time M1M1-HMM process can be done by extending the algorithms used for M1M0-HMM model such as the Baum-Welch algorithm (see [29–31]). Then, a back transformation will allow us to build the continuous-time HMM.

The main contributions of this paper are the following:

- A new dynamic modelling frame for the reliability analysis of a system based on a longitudinal follow-up that only provides information on certain aspects related to the system operation conditions, while the real state of the system is hidden or, in any case, unobserved during the follow-up. The auto-correlation of the observed measurements is considered in the model, which is a realistic situation (as can be seen from the real application presented) and generalizes [7,8]. We denote our model M1M1-HMM.
- Maximum-likelihood estimators of the model are obtained and their theoretical properties are discussed. The reliability concept is adapted to this context. Both versions, discrete-time and continuous-time, are considered.
- A sensitivity analysis is carried out based on the concepts of *false-positive* and *false-negative* that are also introduced in this paper in the context of HMMs.
- Finally, some maintenance related questions are discussed based on the sequence of registered indicators, which will be called signal-runs in this paper.

To illustrate the method we analyse a dataset from a water pump installed in a small area. The information is collected by a set of sensors which monitor parts of the water pump over time. Specifically there are 50 sensors measuring temperature, pressure, vibration, load capacity, volume, flow density, and others, every minute from 01-04-2018 to 31-08-2018. In total there are 220320 data points and 50 variables. Unlike [2,4,22] that propose different data fusion models, we do not combine the information provided by the full multi-sensors platform, rather we use the individual information provided by a unique sensor at a time. Then we can compare the results provided by all sensors to decide which one determines the true state of the machine with the lowest error.

The present paper is organized as follows. Section 2 describes the model M1M1-HMM in continuous time. In Section 3, the M1M1-HMM model is presented in discrete time, and a discretization strategy is presented to estimate the model in continuous time. In Section 4, the reliability function of the system based on this type of models is introduced, both in discrete time and in continuous time. Sensitivity analysis is introduced in Section 5 and maintenance related questions are also discussed. In Section 6 simulations are carried out and a real case based sensor-data from a water pump is studied. Section 7 gives the conclusions. Finally, in the Appendix supplementary material is provided.

2. The continuous-time M1M1-HMM model

2.1. Example. A two-unit system with hot standby redundancy

Consider the following situation. A structure consists of 2 units operating in redundancy as follows. Each unit can be in one of two states. When the unit i is in state ON, it changes to state OFF at a rate λ_i, and it changes from OFF to ON at a rate μ_i, for $i = 1, 2$. The units operate independently and the current state of each unit is not available to the observer. In other words, the states of the units are hidden.

The only information about the system performance is given by an indicator variable Y_t that gives value 0 when the performance is good and 1 otherwise.

The state-space of the structure (i.e., the two-unit system) can be represented by the set $E = \{1, 2, 3, 4\}$ with the following encoding: 1= "both units are ON"; 2= "unit 1 is ON and unit 2 is OFF"; 3= "unit 1 is OFF and unit 2 is ON"; 4= "both units are OFF".

If we denote $X(t)$ the state occupied by the non-observed structure at time t, then $\{X_t, t > 0\}$ is a continuous-time Markov chain (CTMC) taking values in E, with generating matrix given by

$$\mathbf{A} = \begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_2 & \lambda_1 & 0 \\ \mu_2 & -(\mu_2+\lambda_1) & 0 & \lambda_1 \\ \mu_1 & 0 & -(\mu_1+\lambda_2) & \lambda_2 \\ 0 & \mu_1 & \mu_2 & -(\mu_1+\mu_2) \end{pmatrix}. \quad (1)$$

For $t > 0$, $\mathbf{P}(t) = e^{\mathbf{A}t}$ is the transition probability matrix. We assume that $\{Y_t, t > 0\}$ is a continuous-time Markov chain taking values in $\mathcal{Y} = \{0, 1\}$ with generating matrix conditioned on the true internal configuration of the structure as follows. If $X_t = i$, then the transition intensities of the process Y are given by the matrix $\mathbf{B}_i$, for $i = 1, 2, 3, 4$, and $\mathbf{Q}_i(t) = e^{\mathbf{B}_i t}$ is the corresponding transition probability matrix, that is, the matrix whose entries are $Q_{i;01}(t) = \mathbb{P}(Y_t = 1 \mid X_t = i, Y_0 = 0) = 1 - Q_{i,00}(t)$, and $Q_{i;10}(t) = \mathbb{P}(Y_t = 0 \mid X_t = i, Y_0 = 1) = 1 - Q_{i;11}(t)$, for $i \in E$.

A full description of the system behaviour is be given by the two-dimensional process (X_t, Y_t), whose state space is the set $\widetilde{E} = E \times \mathcal{Y}$. Our main purpose is to derive the distribution characteristics of this two-dimensional process.

2.2. The model

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

$$\widetilde{P}_t((i, y'), (j, y)) = P_{ij}(t) Q_{j;y'y}(t),$$

for $i, j \in E$ and $y', y \in \mathcal{Y}$, and $P_{..}$ and $Q_{.;..}$ defined above.

The corresponding generating matrix is $\widetilde{\mathbf{A}}$ with elements $\widetilde{A}((i, y_l), (j, y_h))$ for $(i, y_l), (j, y_h) \in \widetilde{E}$. This matrix is obtained by studying the following limits

$$\begin{aligned}
&\lim_{t\to 0} \frac{1}{t} \left\{ \mathbb{P}(X_t = j, Y_t = y_h \mid X_0 = i, Y_0 = y_l) - \delta_{ij}\delta_{y_l y_h} \right\} \\
&= \lim_{t\to 0} \frac{\mathbb{P}(X_t = j \mid X_0 = i)\mathbb{P}(Y_t = y_h \mid Y_0 = y_l, X_t = j) - \delta_{ij}\delta_{y_l y_h}}{t} \\
&= \lim_{t\to 0} \frac{P_{ij}(t) Q_{j;y_l y_h}(t) - \delta_{ij}\delta_{y_l y_h}}{t},
\end{aligned}$$

for different values of $i, j \in E$ and $y_l, y_h \in \mathcal{Y}$, and where $\delta_{..}$ is the function delta of Kronecker.

We consider 4 cases:

- $i \neq j$ and $l \neq h$, then
  $$\widetilde{A}((i, y_l), (j, y_h)) = \lim_{t\to 0} \frac{P_{ij}(t) Q_{j;y_l y_h}(t)}{t} = 0,$$
  using that for $i \neq j$, $P_{ij}(t) \to 0$, and, for $l \neq k$, $Q_{j;y_l y_h}(t)/t \to B_{j;y_l y_h} < +\infty$, as $t \to 0$.
- $i = j$ and $l \neq h$, then
  $$\widetilde{A}((i, y_l), (i, y_h)) = \lim_{t\to 0} \frac{P_{ii}(t) Q_{j;y_l y_h}(t)}{t} = B_{j;y_l y_h},$$
  using that $P_{ii}(t) \to 1$, and $Q_{j;y_l y_k}(t)/t \to B_{j;y_l y_h}$, as $t \to 0$.
- $i \neq j$ and $y_l = y_h = y$, then
  $$\widetilde{A}((i, y), (j, y)) = \lim_{t\to 0} \frac{P_{ij}(t) Q_{j;yy}(t)}{t} = A_{ij},$$
  using that $Q_{j;yy}(t) \to 1$, and $P_{ij}(t)/t \to A_{ij}$, as $t \to 0$.
- $i = j$ and $y_l = y_h = y$, then
  $$\widetilde{A}((i, y), (i, y)) = \lim_{t\to 0} \frac{P_{ii}(t) Q_{j;yy}(t) - 1}{t} = A_{ii} + B_{j;yy},$$
  where we write
  $$\begin{aligned}
  \lim_{t\to 0} \frac{P_{ii}(t) Q_{j;yy}(t) - 1}{t} &= \lim_{t\to 0} \frac{(P_{ii}(t) - 1) Q_{j;yy}(t) + Q_{j;yy}(t) - 1}{t} \\
  &= \lim_{t\to 0} \frac{P_{ii}(t) - 1}{t} + \lim_{t\to 0} \frac{Q_{j;yy}(t) - 1}{t},
  \end{aligned}$$
  using that $\lim_{t\to 0} Q_{j;yy}(t) = 1$, for any $y \in \mathcal{Y}$.

2.2.1. Example. A two-unit system with hot standby redundancy

In this example, the generating matrix of the internal structure (unobserved) is given in Eq. (1).

Then, for the two-dimensional process (X_t, Y_t), the state space $\widetilde{E} = E \times \mathcal{Y}$, with $E = \{1, 2, 3, 4\}$ and $\mathcal{Y} = \{0, 1\}$, can be written as

$$\widetilde{E} = \{(1,0), (1,1), (2,0), (2,1), (3,0), (3,1), (4,0), (4,1)\},$$

and the corresponding generating matrix $\widetilde{\mathbf{A}}$ is given in Box I.

2.2.2. Matrix notation

Without loss of generality, we can represent the states of the hidden MC by $E = \{1, \dots, d\}$, and for the observable MC we take $\mathcal{Y} = \{y_1, \dots, y_s\}$. Then the elements of the set $\widetilde{E}$ can be ordered as follows

$$\widetilde{E} = \{(1, y_1), \dots, (1, y_s), (2, y_1), \dots, (2, y_s), \dots, (d, y_1), \dots, (d, y_s)\}.$$

Then, the generating matrix of the two-dimensional process can be expressed as

$$\widetilde{\mathbf{A}} = \mathbf{A} \otimes \mathbf{I}_s + \mathrm{Diag}(\mathbf{B}_1, \dots, \mathbf{B}_d), \quad (2)$$

$$\widetilde{\mathbf{A}} = \begin{pmatrix} A_{11}+B_{1;00} & B_{1;01} & A_{12} & 0 & A_{13} & 0 & A_{14} & 0 \\ B_{1;10} & A_{11}+B_{1;11} & 0 & A_{12} & 0 & A_{13} & 0 & A_{14} \\ A_{21} & 0 & A_{22}+B_{2;00} & B_{2;01} & A_{23} & 0 & A_{24} & 0 \\ 0 & A_{21} & B_{2;10} & A_{22}+B_{2;11} & 0 & A_{23} & 0 & A_{24} \\ A_{31} & 0 & A_{32} & 0 & A_{33}+B_{3;00} & B_{3;01} & A_{34} & 0 \\ 0 & A_{31} & 0 & A_{32} & B_{3;10} & A_{33}+B_{3;11} & 0 & A_{34} \\ A_{41} & 0 & A_{42} & 0 & A_{43} & 0 & A_{44}+B_{4;00} & B_{4;01} \\ 0 & A_{41} & 0 & A_{42} & 0 & A_{43} & B_{4;10} & A_{44}+B_{4;11} \end{pmatrix}.$$

Box I.

where $\otimes$ denotes the Kronecker product of matrices, and with Diag we denote a function that transforms a set of d square matrices of dimension $s \times s$ each into a matrix of dimension $d \cdot s \times d \cdot s$ that can be written in blocks as follows

$$\mathrm{Diag}(\mathbf{B}_1, \dots, \mathbf{B}_s) = \begin{pmatrix} \mathbf{B}_1 & \mathbf{0}_s & \dots & \mathbf{0}_s \\ \mathbf{0}_s & \mathbf{B}_2 & \dots & \mathbf{0}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_s & \dots & \mathbf{0}_s & \mathbf{B}_d \end{pmatrix}, \quad (3)$$

where $\mathbf{0}_s$ is a square matrix of zeros with dimension $s \times s$.

Remark

In the particular case of the two-unit system the transition matrix $\widetilde{\mathbf{P}}$ is determined by the following vector of parameters

$$\theta = (\lambda_1, \lambda_2, \mu_1, \mu_2, B_{1;01}, B_{1;10}, B_{2;01}, B_{2;10}, B_{3;01}, B_{3;10}, B_{4;01}, B_{4;10},).$$

All these parameters are interpretable from the point of view of the system behaviour. For example, $B_{2;01}$ ($B_{3;01}$) gives the failure rate of the system when unit 1 (2) is ON and unit 2 (1) is OFF. If for instance, the corresponding estimations show a relation $\hat{B}_{1;01} > \hat{B}_{2;01}$, this can suggest unit 1 has a bigger impact in the system performance, which can lead to detecting weaknesses in the system structure and, among other things, maintenance strategies can be designed based on priority units. All this, provided that the indicators give reliable information as will be discussed later in Section 5.1.

3. The discrete-time M1M1-HMM model

In this section, we are not concerned with time. Rather, we are just interested in the successive states occupied by the system at a (pre-specified) grid of equispaced times and thus we consider a discrete-time stochastic process to represent the system behaviour.

3.1. Model description

Let $\{(X_n, Y_n), n \geq 0\}$ be a two-dimensional stochastic process such that X_n is a discrete-time Markov chain (DTMC) taking values in the set $E = \{1, 2, \dots, d\}$ (i.e., *states*), and Y_n is a DTMC with values in $\mathcal{Y} = \{y_1, y_2, \dots, y_s\}$ (i.e., *signals*).

Let us denote $\mathbf{P} = (P_{ij};\ i, j \in E)$ the transition matrix of the MC X_n. For a fixed $i \in E$ transitions of the MC Y_n occur according to the matrix $\mathbf{Q}_i$, that is,[1]

$$Q_i(y_l, y_h) = \mathbb{P}(Y_n = y_h \mid Y_{n-1} = y_l, X_n = i),$$

for all $y_l,\ y_h \in \mathcal{Y}$. In total we have $\{\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_d\}$ potential transition laws for the DTMC Y_n.

[1] Since confusion is not likely here, we consider this notation, although we have already used $\mathbf{P}$ to denote the matrix of transitions functions in the continuous-time model. Same discussion can be considered for notation $\mathbf{Q}$.

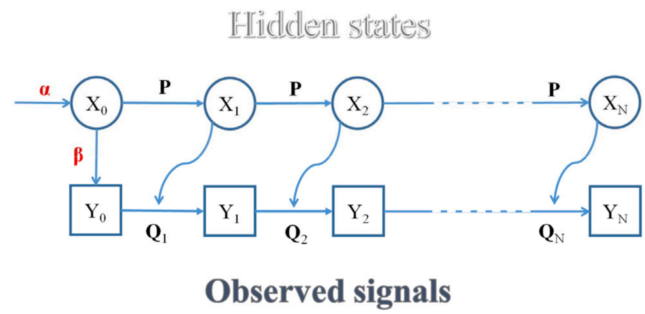

Fig. 1. The discrete-time HMM with M1-M1 structure of dependence.

Then $\{(X_n, Y_n), n \geq 0\}$ is a two-dimensional discrete-time Markov chain with transition matrix that is denoted as $\widetilde{\mathbf{P}}$ and has elements

$$\begin{aligned} \widetilde{P}((i, y_l), (j, y_h)) &= \\ &= \mathbb{P}((X_n, Y_n) = (j, y_h) \mid (X_{n-1}, Y_{n-1}) = (i, y_l)) \\ &= P_{ij} Q_j(y_l, y_h), \end{aligned} \quad (4)$$

for all $(i, y_l), (j, y_h) \in \widetilde{E}$. This matrix, can be written in a compact way as follows. First, for all $i \in E$, let us define the subset of pairs whose first components is i, that is, $\widetilde{i} = \{(i, y_1), \dots, (i, y_s)\}$. Then we can split the state space as follows $\widetilde{E} = \widetilde{1} \cup \widetilde{2} \cup \dots \widetilde{d}$ and finally the transition matrix of the two-dimensional process can be written in blocks as

$$\widetilde{\mathbf{P}} = \left(\begin{array}{c|c|c|c} P_{11}\mathbf{Q}_1 & P_{12}\mathbf{Q}_2 & \dots & P_{1d}\mathbf{Q}_d \\ \hline P_{21}\mathbf{Q}_1 & P_{22}\mathbf{Q}_2 & \dots & P_{2d}\mathbf{Q}_d \\ \hline \vdots & \vdots & \ddots & \dots \\ \hline P_{d1}\mathbf{Q}_1 & P_{d2}\mathbf{Q}_2 & \dots & P_{dd}\mathbf{Q}_d \end{array}\right), \quad (5)$$

where, for $i, j \in E$, the sub-matrix $P_{ij}\mathbf{Q}_j$ gives transitions from the class of states $\widetilde{i}$ to $\widetilde{j}$.

We assume that the first component of the coupled process, i.e. X_n, is not observable while the only information that can be registered over time is about the second component, i.e. Y_n, see Fig. 1. Then $\{(X_n, Y_n), n \geq 0\}$ is a discrete-time hidden Markov process with structure dependency M1-M1. In the following we will denote this process DT-M1M1-HMM.

3.2. Maximum-likelihood estimation

Let us assume that we observe the dataset $Y_1^N = \{Y_1, Y_2, \dots, Y_N\}$, where the following notation is being considered.

Notation:

- $Y_{k_1}^{k_1+k_2} = \{Y_{k_1}, Y_{k_1+1}, \dots, Y_{k_1+k_2}\}$
- $X_{k_1}^{k_1+k_2} = \{X_{k_1}, X_{k_1+1}, \dots, X_{k_1+k_2}\}$
- θ vector of unknown parameters: $\theta = (\theta_1, \theta_2)$; and, with θ_0 denoting the true vector of parameters.

- $\theta_1 = \mathbf{P}^*$ where, $\mathbf{P}^*$ is the matrix $\mathbf{P}$ without the dth column. The number of distinct parameters in θ_1 is: $d \cdot (d-1)$.
- $\theta_2 = (\mathbf{Q}_1^*, \mathbf{Q}_2^*, \dots, \mathbf{Q}_d^*)$ with $\mathbf{Q}_i^*$ matrix $\mathbf{Q}_i$ without the sth column. The total number of parameters to be estimated in θ_2 is: $d\cdot(s-1)\cdot s$. Then, the size of vector θ is $d \cdot (d-1) + d \cdot s \cdot (s-1)$
- Initial conditions: $X_0 = 1$, $Y_0 = y_1$, then $\alpha = (1, 0, \dots, 0)_{(1\times d)}$, $\beta = (1, 0, \dots, 0)_{(1\times s)}$

The likelihood function is given by

$$L(\theta) = p_\theta(Y_1^N) = \sum_{X_1^N \in E^N} \mathbb{P}_\theta(X_1^N, Y_1^N), \tag{6}$$

and then the aim is to find $\hat{\theta}$ maximizing the log-likelihood function, that is

$$\hat{\theta} = \arg\max_\theta \ell(\theta), \tag{7}$$

with $\ell(\theta) = \log L(\theta)$.

3.3. Estimation of the CT-M1M1-HMM based on discretization

To estimate the characteristics of the model described in Section 2 we consider a procedure that uses observations of the state of the system regularly registered in a pre-specified grid of times, i.e. $t_0 < t_1 < \cdots < t_N$, with $t_n - t_{n-1} = h$, for all $n = 1, \dots, N$, and with $h > 0$ and small. We consider the approach presented in [8] where a continuous-time HMM with dependence structure M1-M0 is analysed.

Usually, among others, for economical reasons, a continuous follow-up of the system is not possible, and rather observations are registered regularly in time. In other words, let us assume that the system is observed at times $t_n = n\ h$, with $h > 0$, some constant. Then we can consider a discrete version of the two-dimensional process, that is $\{(\hat{X}_n, \hat{Y}_n), n = 0, 1, \dots\}$, where $\hat{X}_n$ is the internal configuration of the system at time t_n, that is $\hat{X}_n = X_{t_n}$; and $\hat{Y}_n = Y_{t_n}$ is the observed indicator of system performance at time t_n (see [8]).

Given that the generating matrix of the hidden chain is $\mathbf{A} = \lim_{t\to 0}(\mathbf{P}(t) - \mathbf{I})/t$, with $\mathbf{I}$ the identity matrix, we can define

$$\mathbf{P}(h) = \mathbf{A}h + \mathbf{I},$$

and, equally, we have

$$\mathbf{Q}_i(h) = \mathbf{B}_i h + \mathbf{I}.$$

for all $i \in E$.

More specifically, $P_{ij}(h)$ is the (i,j)-element of matrix $\mathbf{P}(h)$ defined above, that is $P_{ij}(h) = A_{ij}h + \delta_{ij}$, for $i, j \in E$; and, similarly, for $y', y \in \mathcal{Y}$, $Q_{i;y'y}(h) = B_{i;y'y}h + \delta_{y'y}$, for $i \in E$.

Based on the observations $\{\hat{Y}_1^N\}$ we obtain the estimators $\hat{\mathbf{P}}(h)$ and $\hat{\mathbf{Q}}_i(h)$, for all $i \in E$, and then define the following estimators of the corresponding parameters of the continuous-time model

$$\hat{\mathbf{A}} = \frac{\hat{\mathbf{P}}(h) - \mathbf{I}}{h},$$

and

$$\hat{\mathbf{B}}_i = \frac{\hat{\mathbf{Q}}_i(h) - \mathbf{I}}{h}, \quad i \in E,$$

for h sufficiently small. The asymptotic properties of these estimators can be deduced using similar arguments as in [8].

4. Reliability based on M1M1-HMM models

In this section we define the reliability function based on both approaches, discrete-time as well as continuous time.

4.1. The continuous-time model

The model $\{(X_t, Y_t); t > 0\}$ is a two-dimensional continuous-time Markov chain with state-space $\widetilde{E} = E \times \mathcal{Y}$ and generating matrix $\widetilde{\mathbf{A}}$ as deduced in Section 2. Following the arguments in [7], we assume that the state-space E is split into two subsets such that $E = U \cup D$, where we can denote for instance $U := \{1, \dots, m\}$, the *working* states, and $D := \{m+1, \dots, d\}$, the *down* states. For simplicity, and without loss of generality, this notation is used for the states of the system. Additionally, the system up states can be defined not only by $U \in E$ but also by some subset of $\mathcal{Y}$. In some situations, the information we get about the system functioning can be categorized into two groups of signals. On the one hand, we have a group of signals indicating a *good* performance, the subset $\mathcal{Y}_1$; and, on the other hand, there is another group of signals $\mathcal{Y}_2$ for *warning* of some serious problem in the system that involves the operation interruption thus causing the system failure, that is, we have also the partition $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$. The set of states of the two-dimensional process (X_n, Y_n) can be written as $\widetilde{E} = E \times \mathcal{Y}$, with $\widetilde{E} = \{(1, y_1), \dots, (d, y_1), (1, y_2), \dots, (d, y_2), \dots, (1, y_s), (2, y_s), \dots, (d, y_s)\}$. Let us define $\widetilde{\mathcal{U}} = U \times \mathcal{Y}_1$ and $\widetilde{\mathcal{D}} = \widetilde{\mathcal{E}} \setminus \widetilde{\mathcal{U}}$, with $\widetilde{E} = E \times \mathcal{Y}$ already defined. Thus, the corresponding generating matrix $\widetilde{\mathbf{A}}$ has elements given in Eq. (2) and can be written in blocks as

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{A}_{\widetilde{U}\widetilde{U}} & \mathbf{A}_{\widetilde{U}\widetilde{D}} \\ \hline \mathbf{A}_{\widetilde{D}\widetilde{U}} & \mathbf{A}_{\widetilde{D}\widetilde{D}} \end{array} \right). \tag{8}$$

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

$$\begin{aligned} \widetilde{R}^c_{(i,y)}(t) &= \mathbb{P}_i(\mathcal{T} > t) = \mathbb{P}_i(X_s \in U, Y_s \in \mathcal{Y}_1, 0 < s \leq t) \\ &= \mathbb{P}_i((X_t, Y_t) \in \widetilde{U}, 0 < s \leq t), \end{aligned} \tag{9}$$

and then

$$\widetilde{R}^c(t) = \sum_{(i,y) \in \widetilde{U}} \widetilde{R}^c_{(i,y)}(t),$$

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

$$\widetilde{R}^c(t) = \widetilde{\alpha} \exp\left(\widetilde{\mathbf{A}}_{\widetilde{U}\widetilde{U}}\ t\right) \mathbf{1}_{\widetilde{U}},$$

where $\widetilde{\mathbf{A}}_{\widetilde{U}\widetilde{U}}$ denotes the sub-matrix of $\widetilde{\mathbf{A}}$ with all transition rates among states of subset $\widetilde{U}$, as given in 8. Based on a sample $\hat{Y}_2, \dots, \hat{Y}_N$, with $\hat{Y} = Y(t_{nh})$ for a fixed h small enough, we define the following estimator of the reliability function in continuous time

$$\widehat{\widetilde{R}}^c_h(t) = \widetilde{\alpha} \cdot \exp(\widehat{\widetilde{\mathbf{A}}}_{h,\widetilde{U}\widetilde{U}} t) \cdot \mathbf{1}_{\widetilde{U}}, \quad t \geq 0.$$

Using similar arguments as in [8], the following result can be obtained.

Proposition 1. *The estimator $\widehat{\widetilde{R}}^c_h(t)$ is asymptotically Normal, as $N \to \infty$, and $h \to 0$, for any $t \geq 0$, i.e.,*

$$\sqrt{n}(\widehat{\widetilde{R}}^c_h(t) - \widetilde{R}^c(t)) \xrightarrow{d} N(0, \Sigma_{R^c}).$$

4.2. The discrete-time model

The model $\{(X_n, Y_n); n = 0, 1, \dots\}$ is a two-dimensional discrete-time Markov chain with state-space $\widetilde{E} = E \times \mathcal{Y}$ and transition matrix $\widetilde{\mathbf{P}}$ as deduced in Section 3. Let us denote $\mathcal{N}$ the first time the system visits the set of down states D, i.e., the hitting time of set D, that is, $\mathcal{N} = \min\{n > 0 : \widetilde{\mathcal{X}} = (X_n, Y_n) \in \widetilde{\mathcal{D}}\}$. Therefore the reliability of the system can be defined as $\widetilde{R}^d(n) = \mathbf{P}(\mathcal{N} > n)$, for $n \geq 0$. Conditioning on the initial state $(i, y) \in \widetilde{U} = U \times \mathcal{Y}_1$, we write

$$\widetilde{R}^d_{(i,y)}(n) = \mathbb{P}_i(\mathcal{N} > n) = \mathbb{P}_i(X_m \in U, Y_m \in \mathcal{Y}_1, 0 < m \leq n)$$

$$= \mathbb{P}_i((X_n, Y_n) \in \widetilde{U}, 0 < m \leq n), \quad (10)$$

and then

$$\widetilde{R}^{\mathrm{d}}(n) = \sum_{(i,y)\in\widetilde{U}} \widetilde{R}^{\mathrm{d}}_{(i,y)}(n),$$

for $n = 1, 2 \ldots$. Using matrix notation, we can write

$$\widetilde{R}^{\mathrm{d}}(n) = \widetilde{\alpha}(\widetilde{\mathbf{P}}_{\widetilde{U}\widetilde{U}})^n \mathbf{1}_{\widetilde{U}},$$

where $\widetilde{\mathbf{P}}_{\widetilde{U}\widetilde{U}}$ is the sub-matrix of $\widetilde{\mathbf{P}}$ with all transition probabilities among states of subset $\widetilde{U}$, considering a block partition of matrix $\mathbf{P}$ similar to 8. Based on a sample $Y_1, \ldots, Y_N$, we define the following estimator of the reliability function in discrete time as

$$\widehat{\widetilde{R}}^{\mathrm{d}}(n) = \widetilde{\alpha}(\widehat{\widetilde{\mathbf{P}}}_{\widetilde{U}\widetilde{U}})^n \mathbf{1}_{\widetilde{U}}, \quad n \geq 0.$$

Using similar arguments as in [7], the following result can be obtained.

Proposition 2. *The estimator $\widehat{\widetilde{R}}^{\mathrm{d}}(n)$ is asymptotically Normal, as $N \to \infty$, for any $n \geq 1$, i.e.,*

$$\sqrt{N}(\widehat{\widetilde{R}}^{\mathrm{d}}(n; N) - \widetilde{R}^{\mathrm{d}}(n)) \xrightarrow{d} N(0, \Sigma_{R^{\mathrm{d}}}).$$

5. Maintenance

Since the observations are registered in a discrete grid of time-points, let us formulate the problem of maintenance strategies based on the discrete-time HMM model defined above. That is, let (X_n, Y_n) be defined as in Section 3. Following Section 4 we assume that the system state-space can be split into two subsets, that is $E = U \cup D$, with U the *working* states, and D the *down* states. As well, the set of signals is split into two subsets $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$, with $\mathcal{Y}_1$ the *safe* signals, and $\mathcal{Y}_2$ the *warning* signals. We consider maintenance policies as follows.

The system is inspected at regular intervals to detect any problem and intervene if necessary. The natural strategy would consider to maintain when a *warning* of the subset $\mathcal{Y}_2$ is received. However, even when a *warning* arrives, there is a non-zero probability that the system is working properly. Then an unnecessary maintenance action would be carried out incurring the corresponding cost of intervention. On the other hand, when a safe signal is received, i.e. the subset $\mathcal{Y}_1$, there is a positive probability that the system has failed nevertheless. Thus, the system is not inspected although the operation is stopped which means a production loss. In any of the two cases, undesirable economic consequences can be triggered for the system environment.

In order to prevent these situations, we propose to implement a preventive maintenance policy that minimizes the probabilities of error defined in the next section.

5.1. Sensitivity analysis

Two types of error can be committed when a decision is taken based on the registered signal. On the one hand, there is a positive probability of receiving a signal of system failure (alarm) while the system is functioning properly, we call this *False Positive*. On the other hand, a signal can be emitted indicating good performance while the system is out of work. This is a *False Negative* case. What is expected is that the mechanism that emits the signals works well in the sense that the probability of getting false positive signals (FPP) and the probability of getting false negative signals (FNP) are small enough. In what follows we discuss the discrete-time model only.

Definition 1. Types of error

1. The probability of False Positive (FPP) is defined as

$$FPP(n) := \mathbb{P}(Y_n \in \mathcal{Y}_2 \mid X_n \in U), \quad n = 1, 2, \ldots. \quad (11)$$

2. The probability of False Negative (FNP) is defined as

$$FNP(n) := \mathbb{P}(Y_n \in \mathcal{Y}_1 \mid X_n \in D), \quad n = 1, 2, \ldots. \quad (12)$$

It can be checked that

$$FPP(n) = 1 - \frac{\widetilde{A}_v(n)}{A_v(n)}, \quad (13)$$

and

$$FNP(n) = \frac{\mathbb{P}(Y_n \in \mathcal{Y}_1) - \widetilde{A}_v(n)}{1 - A_v(n)}, \quad (14)$$

where $A_v(n) = \mathbb{P}(X_n \in U)$, for $n = 1, 2, \ldots$, gives the (internal) availability of the system at the nth observation, and $\widetilde{A}_v(n) = \mathbb{P}((X_n, Y_n) \in \widetilde{U})$ is the probability that the system is operative and the signal received is indicating good performance.

5.2. Signal-runs

While an efficient mechanism of control for false positive is crucial to establish the ability of our model in order to detect the (unobserved) failure of the system, for the sake of a competitive maintenance plan, it is very important to precisely measure the degree of confidence that we have when recording a warning signal (i.e. $Y_n \in \mathcal{Y}_2$) that the system state is indeed in failure (i.e. $X_n \in U$). So we introduce the concept of *signal-runs*.

At the nth observation time t_n let us assume that the signal received is an alarm, i.e. $Y_n \in \mathcal{Y}_2$, meaning that the system is not working properly. The question is to know how much we trust this observation. In other words, we want to know whether the system is truly in failure, i.e. $X_n \in U$. This question is analogous to the evaluation of the accuracy of diagnostic testing. A key issue in clinical studies that are aimed at determining the true state of a patient with respect to a disease based on the result of a screening or diagnostic test (see [32]).

Definition 2. Signal-run of length k.

For $n \geq 1$ and $1 \leq k \leq n$, let us define the following function

$$\gamma(i, y^n_{n-k+1}, k, n) := \mathbb{P}(X_n = i \mid Y^n_{n-k+1} = y^n_{n-k+1}),$$

for $i \in E$, and $y^n_{n-k+1} \in \mathcal{Y}^k$. In particular, we define a ***positive*** k- ***run*** as

$$\gamma_+(k, n) = \mathbb{P}(X_n \in D \mid Y_{n-k+1} \in \mathcal{Y}_2, \ldots, Y_n \in \mathcal{Y}_2),$$

for $n \geq k - 1$, where $D \subset E$ is the subset of failure states and $\mathcal{Y}_2 \subset \mathcal{Y}$ is the subset of warning signals.

On the other hand, we define a ***negative*** k- ***run*** as

$$\gamma_-(k, n) = \mathbb{P}(X_n \in U \mid Y_{n-k+1} \in \mathcal{Y}_1, \ldots, Y_n \in \mathcal{Y}_1),$$

for $U \subset E$, the subset of operative states and $\mathcal{Y}_1 \subset \mathcal{Y}$ is the subset of safe signals that indicate correct functioning of the system.

Remark.

Since the main objective is to predict system failures, we are specially interested in signal-runs of successive warnings, that is $\gamma_+(k, n)$, for $1 \leq k \leq n$. For simplicity we will refer to them as k-runs.

As particular cases, for any $n > 0$, based on runs of length $k = 1$ we obtain a probability that has a similar meaning as the concept of predictive values that are used in biostatistics (see [32]). Specifically,

$$\gamma_+(1, n) = \mathbb{P}(X_n \in D \mid Y_n \in \mathcal{Y}_2), \quad n = 1, 2, \ldots,$$

is the probability that the system is failed when a warning signal is observed. We call this probability the positive predictive value at the nth observation time, and

$$\gamma_-(1, n) = \mathbb{P}(X_n \in U \mid Y_n \in \mathcal{Y}_1), \quad n = 1, 2, \ldots,$$

which is the probability that the system is working when a safe signal is registered, and we will call it the negative predictive value.

To calculate these values we consider the following expressions

$$\gamma_{+}(1,n) = \frac{\mathbb{P}(Y_n \in \mathcal{Y}_2 \mid X_n \in D)\mathbb{P}(X_n \in D)}{\mathbb{P}(Y_n \in \mathcal{Y}_2 \mid X_n \in D)\mathbb{P}(X_n \in D) + \mathbb{P}(Y_n \in \mathcal{Y}_2 \mid X_n \in U)\mathbb{P}(X_n \in U)} = \frac{(1-FNP(n))(1-A_v(n))}{(1-FNP(n))(1-A_v(n)) + FPP(t)A_v(n)},$$

and,

$$\gamma_{-}(1,n) = \frac{\mathbb{P}(Y_n \in \mathcal{Y}_1 \mid X_n \in U)\mathbb{P}(X_n \in U)}{\mathbb{P}(Y_n \in \mathcal{Y}_1 \mid X_n \in U)\mathbb{P}(X_n \in U) + \mathbb{P}(Y_n \in \mathcal{Y}_1 \mid X_n \in D)\mathbb{P}(X_n \in D)} = \frac{(1-FPP(n))A_v(n)}{(1-FPP(n))A_v(n) + FNP(n)(1-A_v(n))}.$$

Using the assumptions A1-A2 given in Appendix A, we can obtain the stationary distribution of the chain X_n as well as the stationary distribution Y_n, and then we can define the corresponding stationary measures as follows.

Definition 3. Limiting values.
Let us denote $\{\widetilde{\pi}(i,y); (i,y) \in \widetilde{E}\}$ the stationary distribution of the process (X_n, Y_n), then the stationary positive predictive value is obtained as

$$\gamma_{+} = \frac{\sum_{(i,y)\in D\times\mathcal{Y}_2} \widetilde{\pi}(i,y)}{\sum_{(i,y)\in E\times\mathcal{Y}_2} \widetilde{\pi}(i,y)},$$

and the stationary negative predictive value is

$$\gamma_{-} = \frac{\sum_{(i,y)\in U\times\mathcal{Y}_1} \widetilde{\pi}(i,y)}{\sum_{(i,y)\in E\times\mathcal{Y}_1} \widetilde{\pi}(i,y)}.$$

5.3. Maintenance strategy based on k-runs

Then, the frequency of alarm signals has to be studied which might depend on many factors. For example, if we wish to monitor remote sites we could be receiving continuous monitoring signals or on the contrary, we might need to access the site to collect some information about the system performance. This will depend on the importance of the equipment that is being monitored, the budget, the existence or not of a communication infrastructure etc. The implementation of low cost, wireless monitoring sensors and other monitoring equipment is becoming common in many areas, therefore, receiving continuous information about the system health is becoming usual.

Let us consider that a warning signal is emitted by the system environment. Should we maintain always in this situation? That is, a warning signal emitted at a single time point is sufficient to believe that the system is actually in failure and must be repaired or, on the contrary, an uninterrupted sequence of warnings of a certain length is required to believe that the system is really not working.

Given the M1-M1 structure of our model, the following result holds.

Proposition 3. *Let (X_n, Y_n) a DT-M1M1-HMM defined as above. Then, for any $i \in E$, any $k \ge 2$, and $y_{n-k}^n \in \mathcal{Y}^{k+1}$, we have that*

$$\gamma(i, y_{n-k+1}^n, k, n) = \gamma(i, y_{n-1}^n, 2, n).$$

Proof. Let us assume that the initial state of the system is i_0 and the first signal emitted is fixed as y_0, then we can prove that

$$\gamma(i, y_{n-k+1}^n, k, n) = \mathbb{P}(X_n = i \mid Y_{n-k+1}^n = y_{n-k+1}^n) = \frac{\mathbb{P}(X_n = i, Y_{n-k+1}^n = y_{n-k+1}^n)}{\mathbb{P}(Y_{n-k+1}^n = y_{n-k+1}^n)} = \frac{\mathbb{P}(Y_n = y_n \mid X_n = i, Y_{n-k+1}^{n-1} = y_{n-k+1}^{n-1})\mathbb{P}(X_n = i \mid Y_{n-k+1}^{n-1} = y_{n-k+1}^{n-1})}{\mathbb{P}(Y_n = y_n \mid Y_{n-k+1}^{n-1} = y_{n-k+1}^{n-1})}.$$

Using that

$$\mathbb{P}(X_n = i \mid Y_{n-k+1}^{n-1} = y_{n-k+1}^{n-1}) = \sum_{j\in E} \mathbb{P}(X_n = i \mid X_{n-1} = j)\mathbb{P}(X_{n-1} = j) = \mathbb{P}(X_n = i),$$

and the Markov property of $\{Y_n\}$, we obtain

$$\gamma(i, y_{n-k}^n, k, n) = \frac{\mathbb{P}(X_n = i, Y_n = y_n, Y_{n-1} = y_{n-1})}{\mathbb{P}(Y_n = y_n, Y_{n-1} = y_{n-1})} = \gamma(i, y_{n-1}^n, 2, n). \quad \square$$

We can conclude that under the M1-M1 assumption, we only need to inspect the system when runs of alarm signals of length 1 or 2 are registered. In these cases, when the probability of a down state is above a certain threshold, the system should be maintained. Specifically, we propose to inspect the system regularly at time intervals of length τ_q where

$$\tau_q = \min\{n \le 2 : \max\{\gamma_{+}(1,n), \gamma_{+}(2,n)\} \ge q\}, \tag{15}$$

with $0 < q < 1$ previously fixed.

6. Numerical examples

In the first practical case a simulated example modelled using a CT-M1M1-HMM is discussed while in the second one a real dataset is analysed using a DT-M1M1-HMM.

6.1. Simulations: A two-unit system with hot redundancy

For the example presented in Section 2.1 we take the following particular values of the elements of matrix $\mathbf{A}$ in (1), $\lambda_1 = 0.1$, $\lambda_2 = 0.2$, $\mu_1 = \mu_2 = 0.5$.

As in the previous sections, at regular instants of time, one has access to some indicators (signals) related somehow to the level of performance of the system. In particular, each time, we receive a signal of good functioning (i.e., $1 = ON$) or a signal of failure in the system (i.e., $2 = OFF$). The matrices ruling the conditional behaviour of the observable process Y_t, over the event $\{X_t = i\}$, for $i = 2,3$ are taken as

$$\mathbf{B}_2 = \begin{pmatrix} -0.2 & 0.2 \\ 0.7 & -0.7 \end{pmatrix}, \qquad \mathbf{B}_3 = \begin{pmatrix} -0.4 & 0.4 \\ 0.4 & -0.4 \end{pmatrix}.$$

In this case we assume that $\mathbb{P}(Y_t = 1 \mid X_t = 1) = 1$, and $\mathbb{P}(Y_t = 2 \mid X_t = 4) = 1$ for all $t > 0$, and then when $X_t = 1$ the state 2 of MC Y is such that $B_{1;22} = +\infty$, that is, state 2 is an instantaneous state as defined in [33]. In a similar way, we assume that when the two units of the system have failed, i.e. $i = 4$, the only possible output is 2, and then, in this case state 1 is an instantaneous states for the MC Y over the event $X_t = 4$. So, we re-define the state space of the coupled process as the set $\widetilde{E} = \{(1,1),(2,1),(2,2),(3,1),(3,2),(4,2)\}$, with transition intensity matrix given by

$$\widetilde{\mathbf{A}} = \begin{pmatrix} -0.3 & 0.2 & 0 & 0.1 & 0 & 0 \\ 0.5 & -0.8 & 0.2 & 0 & 0 & 0.1 \\ 0.5 & 0.7 & -1.3 & 0 & 0 & 0.1 \\ 0.5 & 0 & 0 & -1.1 & 0.4 & 0.2 \\ 0.5 & 0 & 0 & 0.4 & -1.1 & 0.2 \\ 0 & 0 & 0.5 & 0 & 0.5 & -1 \end{pmatrix}.$$

Let us denote $X_0, X_1, \dots, X_N$ the successive (unobserved) states of the system, taking values in the set $E = \{1,2,3,4\}$; and, $Y_0, Y_1, \dots, Y_N$ the successive observed indicators, which are assumed to range in the set $\mathcal{Y} = \{1,2\}$. We consider that inspections are carried out at times $n = 0, \Delta, 2\Delta, \dots$, for simplicity we take $\Delta = 1$. At time $n = 0$ we assume that the system is new so that the initial state is $X_0 = 1$, and also that $Y_0 = 1$.

For each system we have simulated Markovian sample paths of size $N = 50, 100, 500$ using the corresponding true model $(\alpha, \mathbf{A})$. Then, to generate the sequence of observed signals, we consider at each time, the generated state, that is if $X_n = i$ then we simulate the next output Y_n, using the intensity matrix $\mathbf{B}_i$, for any $n = 1, \dots, N$.

To avoid wrong conclusions due to the randomness in the simulation process the experiment has been repeated a total of 500 times for each sample size. The estimation results are represented in Fig. 2. The

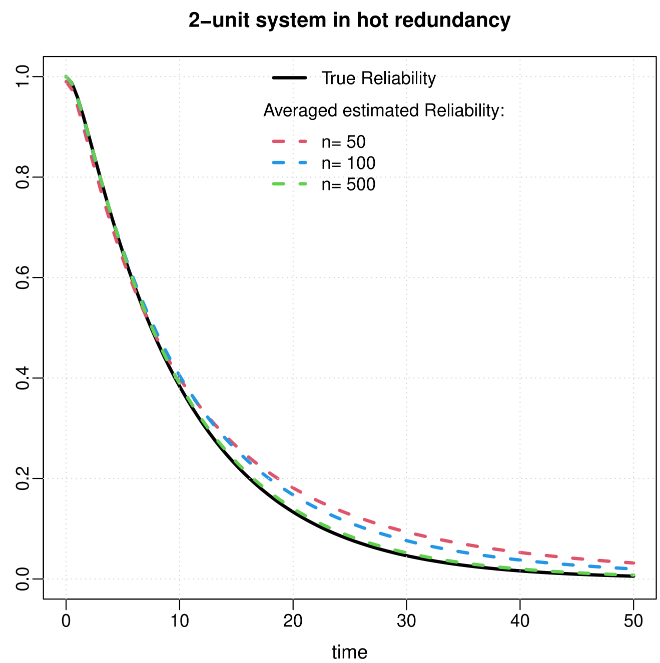

Fig. 2. Reliability estimation for the two-unit system in hot redundancy.

true reliability is given by the black curve. For each sample we have estimated the reliability function based on the HMM model. The results have been summarized through averaging. That is, we consider the following

$$\hat{R}_{av;N}(t) = \frac{1}{500}\sum_{r=1}^{500} \hat{R}_N^{(r)}(t),$$

where $\hat{R}_N^{(r)}$ is the estimated reliability function based on the rth sample, for $r = 1, \ldots, 500$, and for the sample size N.

The red curve represents the average of the estimated curves along the 500 replications for each case $N = 50, 100, 500$. As expected, the accuracy of the estimator decreases with the sample size.

6.2. A real case: Water pump sensor data

We analyse a dataset related to the functioning of a water pump of a small area. The data have been taken from the data platform www.kaggle.com/ and a further analysis can be found in [34].

Very few technical details about the system are available from the website. The information provided consists of timestamp measurements recorded by 50 sensors and machine status every minute from 01-04-2018 to 31-08-2018. In total there are 220320 data points and 51 variables.

Sensors are used to record temperature, pressure, vibration, load capacity, volume, flow density, and others. We have a longitudinal follow up of one single system with observations taken with a time span of just one minute between two consecutive data points. We do not take all the records in the sample but consider only daily registers since 01-04-2018 to 31-08-2018. Then, we have a sample of size $N = 153$ that consists of all records taken at 00:00 every day. We have considered the model in discrete time for this example.

First of all, we have performed auto-correlation tests for the sequence of measurements provided by the sensors. The results indicate dependence between successive observations, so that the dependence structure considered in the model is justified in this case. We have explored every series of measurements reported in the full dataset, and the results are very similar in all cases. For illustration, let us consider for example sensor labelled with number 08 in the dataset. The Durbin–Watson test has reported a test statistic $DW = 1.2079$ and then the p-value= 4.084e−07 is significant. So, we reject the null hypothesis and then the values in the series are autocorrelated. Also, from the ACF plot displayed in Fig. 3 we conclude that autocorrelation of first order cannot be rejected.

The goal is to find out which sensors provide the more useful information about the true machine status. That is we perform a sensitivity analysis to determine which sensors report a lower rate of error when predicting the machine state. Let us focus on sensor labelled # that measures a particular characteristic, $Y_{\#}$ of the machine taking values in $\mathcal{Y} \in \mathbb{R}$. For all sensors considered in the study, lower values reported indicate good machine performance, while higher values warn of a machine failure. Then, for the machine status we consider the 1 represents the operative state and 2 is the failure state.

6.2.1. Sensitivity analysis

We proceed as follows:

1. Let $y_1 < \cdots < y_{100}$, a set of possible values in the range of measurements of sensor #.
2. For each $j = 1, \ldots, 100$, let us define Y_j taking value 1 if $Y_{\#} > y_j$, and 2, otherwise. Then we have $Y_{j,1}, \ldots, Y_{j,N}$ a sample of the DT-M1M1-HMM.
3. Fit the model following Section 3.3 to obtain estimations of $\mathbf{P}$ and $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Then we calculate an estimation of the transition matrix of the coupled process. Let us denote $\widetilde{\mathbf{P}}_j$ such estimation, for $j = 1, \ldots, 100$.
4. Estimate the two error probabilities defined in Section 5.1, which are
$$FPP_j = \frac{\widetilde{\mathbf{P}}_j(1,2)}{\widetilde{\mathbf{P}}_j(1,1) + \widetilde{\mathbf{P}}_j(1,2)},$$
and
$$FNP_j = \frac{\widetilde{\mathbf{P}}_j(2,1)}{\widetilde{\mathbf{P}}_j(2,1) + \widetilde{\mathbf{P}}_j(2,2)}.$$
5. Build the ROC curve as the set of points $\{(FPP_j, 1 - FNP_j), j = 1, \ldots, 100\}$;
6. Let $(FPP_0, 1 - FNP_0)$ the point of the curve that minimizes the distance to the point (0,1).
7. Finally define the sensitivity index of sensor # as $Se_{\#} = 1 - FNP_0$, and the specificity index is given by $Sp_{\#} = 1 - FPP_0$.

The most informative sensor will be chosen as the one displaying the highest combination of sensitivity and specificity. We have considered in our study the complete network of 50 sensors. As can be seen from the results in Table 2 the subset of sensors with labels from 17 to 26 are the most sensible to the real state of the water pump. As we have already mentioned, the technical information about this system and in particular referring to the location and mission of each sensor is not given in the website that provides the dataset. However posterior analyses by the users have shed some light on the matter. In particular it is specified that this group of sensors (17 to 26) have to do with certain characteristics of the impeller of the pump.

6.2.2. Maintenance based on k-runs

For each sensor we have calculated the corresponding probability of system failure based on runs of length $k = 1$ and $k = 2$.

The results indicated that the estimated probability that the system is failed conditioning on k-runs with $k = 2$ is smaller than with $k = 1$ for all sensors except for sensor number 37. In this example we are considering only daily sensor measurements, however in the real problem sensors are continuously monitoring the water pump. Our results are likely due to an intervention in the system previous to the next observation that we consider on the following day. In other words,

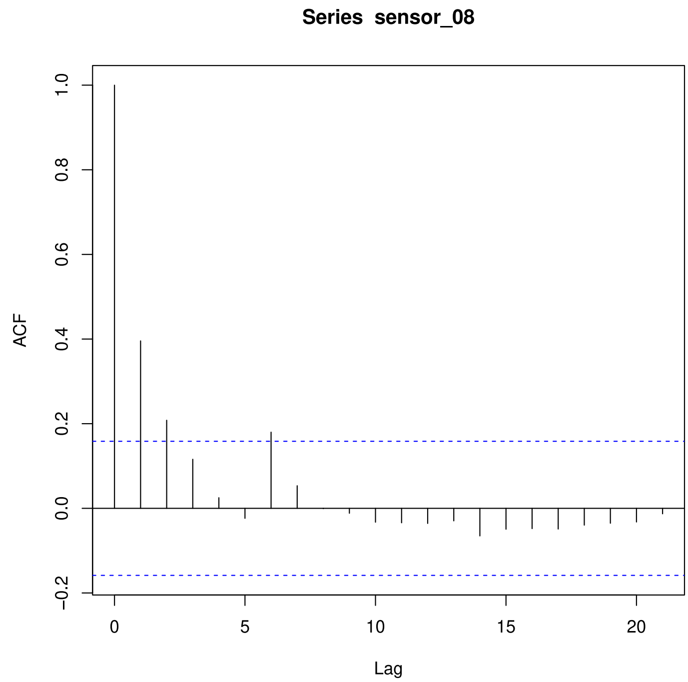

Fig. 3. Autocorrelation test for measurements from $sensor_0 8$.

Table 2
Sensitivity analysis of the sensor network for a water pump system.

Sensor	00	01	02	03	04	05	06	07	08	09
Se	0.465	0.650	0.606	0.537	0.493	0.485	0.581	0.678	0.558	0.510
Sp	0.908	0.929	0.900	0.880	0.943	0.908	0.757	0.878	0.881	0.889
Sensor	10	11	12	13	14	16	17	18	19	20
Se	0.616	0.616	0.656	0.695	0.700	0.632	0.804	0.804	0.809	0.809
Sp	0.970	0.970	0.767	0.899	0.940	0.941	0.937	0.937	0.944	0.944
Sensor	21	22	23	24	25	26	27	28	29	30
Se	0.809	0.816	0.809	0.809	0.809	0.809	0.578	0.809	0.716	0.817
Sp	0.944	0.937	0.944	0.944	0.944	0.944	0.831	0.937	0.911	0.932
Sensor	31	32	33	34	35	36	37	38	39	40
Se	0.727	0.713	0.727	0.760	0.760	0.760	0.750	0.492	0.714	0.560
Sp	0.931	0.881	0.930	0.933	0.933	0.933	0.967	0.909	0.797	0.756
Sensor	41	42	43	44	45	46	47	48	49	50
Se	0.479	0.503	0.509	0.616	0.469	0.474	0.483	0.674	0.465	0.710
Sp	0.902	0.861	0.882	0.970	0.914	0.888	0.862	0.807	0.914	0.771

it could mean that a maintenance action took place or the system was reset between the two successive observations that we are considering here.

Let us focus for example on sensor number 08, although results regarding the complete set of sensors are available from the authors. In Fig. 4 the estimated probability of system failure conditioned to k-runs with $k = 1, 2$ are represented based on the observations of this sensor. As we can observe from the plot, in the beginning, the sensor network seems to be not well tuned, because there is very little probability of system failure even when warning signals are being emitted by a sensor. However, it is clear that after some time the sensors are better adjusted so that information provided is more reliable.

A reference line has been plotted for a threshold probability equal to 0.5. Using Eq. (15), in this case $\tau = 13$ is estimated, meaning that according to this criterion, with $q = 0.5$, the system should be maintained every 13 days according to the information provided by this sensor because after 13 days, the probability of system failure given a warning signal has been produced is bigger than 0.5. Here a threshold of 0.5 has been taken just for illustrative purposes, but in a real situation a more appropriate value for q should be chosen using certain criteria based on cost optimization for instance.

7. Conclusions

The main objective of this paper is to build a stochastic model based on Hidden Markov processes to describe the evolution-in-time of a system and to estimate some dependability functions when no direct observation of the true current state of the system is available but only some indicators or signals.

The deployment of low cost, wireless monitoring sensors is becoming usual in many areas, in particular in many engineering applications where observations related to the health state of the system are recorded over a period of time. The challenge is to interpret the measurements provided by a sensor (which can record temperature, pressure, vibration, among others) in terms of the internal state of the system. Besides, when a longitudinal follow-up is carried out one of the key consequences is that measurements over time might be correlated.

To account for this dependence between observations, we propose a Hidden Markov model with structure of type M1-M1 to infer the state of the system. That is, both the hidden process as well as the signal (observable) process are Markov chains. Thus, we generalize the model M1-M0 that is considered more usually in practice.

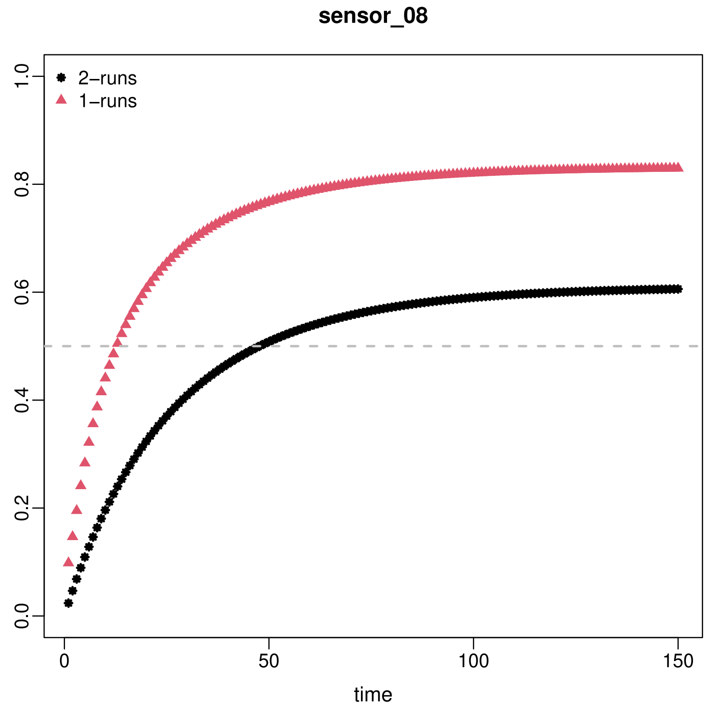

Fig. 4. Probability of system failure conditioned on k-runs for sensor_08, and $k = 1, 2$.

The main contributions of the paper are: 1. The proposed model and some dependability measures have been estimated by maximum-likelihood and some theoretical properties have been studied. The estimation has been approached by means of two versions of the model, in discrete-time as well as in continuous-time. 2. From the point of view of the practical applications in reliability engineering, we have introduced a sensitivity analysis in the context of HMMs. In particular we have defined two types of errors that can be incurred when the hidden state of the system is diagnosed based on observations (signals), these are false positive and false negative errors. Additionally, we have introduced the concepts of sensitivity, specificity and predictive values (these are usual concepts in clinical essays in biostatistics). Finally, we have defined the concept of signal runs. All these measurements can help the decision makers to adopt a better maintenance strategy, avoiding maintaining the system before it is really necessary.

As a future work, we plan to extend the research on maintenance strategies. Furthermore, we will consider a multidimensional observation process. The motivation for this problem can be also inspired by sensor applications. Any industrial equipment is monitored by a set of sensors, more than just one, and the information provided by all of them must be processed together to determine the state of health of the machine. We believe that the insights in this paper and future research can result of great use for engineers.

On the other hand, a deeper knowledge of the real system that is being investigated will be also of great help for us since we will be able improve the accuracy of our model on the basis of expert opinion. So, while it is not always easy, a better communication with engineers would be mutually beneficial.

CRediT authorship contribution statement

M.L. Gámiz: Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Formal analysis, Conceptualization. **F. Navas-Gómez:** Writing – original draft, Software, Investigation, Formal analysis, Conceptualization. **R. Raya-Miranda:** Writing – original draft, Software, Investigation, Formal analysis, Conceptualization. **M.C. Segovia-García:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

All authors have seen and approved the final version of the manuscript being submitted. They warrant that the article is the authors' original work, has not received prior publication and is not under consideration for publication else where.

Data availability

Data will be made available on request.

Acknowledgements

The authors are grateful to three anonymous reviewers and the Editor for many valuable comments and suggestions, which have helped to improve the quality of the article. This work is jointly supported by the Spanish Ministry of Science and Innovation—State Research Agency through grants numbered PID2020-120217RB-I00 and PID2021-123737NB-I00, and by the Spanish Junta de Andalucia through grant number B-FQM-284-UGR20 and the IMAG Maria de Maeztu, Spain grant CEX2020-001105-/AEI/10.13039/501100011033. All authors read and approved the final manuscript.

Appendix A. Maximum-likelihood estimation of the DT-M1M1-HMM

The following conditions are assumed [35]:

A1 The Markov chain X is ergodic, i.e., irreducible and aperiodic, and stationary;

A2 The Markov chain Y is ergodic, i.e., irreducible and aperiodic, and stationary;

A3 There exists an integer $n \in \mathbb{N}$ such that the Fisher information matrix

$$I_n(\theta_0) = -E_{\theta_0}\left(\left.\frac{\partial^2 \log p_\theta(Y_0^n)}{\partial\theta_i \partial\theta_j}\right|_{\theta=\theta_0} \right)_{i,j}$$

is nonsingular, where $\log p_\theta(Y_0^N)$ is the log-likelihood function defined from Eq. (6).

Based on [29,30,35], and, [7] the following result is straightforward.

Theorem 1. *Under assumptions A1 - A3, given a sample of observations $\{Y_1^N\}$, the maximum-likelihood estimator $\hat{\theta}_N = (\hat{\theta}_1, \hat{\theta}_2)_N$ of $\theta = (\theta_1, \theta_2)$ is strongly consistent as N tends to infinity. Moreover, the random vector*

$$\sqrt{N}\,(\hat{\theta}_N - \theta_0) = \sqrt{N}\,[((\hat{P}(i,j)_{1\le i\le d, 1\le j<d}), (\hat{Q}_i(y_l, y_h)_{1\le i\le d; 1\le l\le s; 1\le h<s})) - ((P^0(i,j)_{1\le i\le d, 1\le j<d}), (Q_i^0(y_l, y_h)_{1\le i\le d; 1\le l\le s; 1\le h<s}))]$$

is asymptotically Normal, as $N \to +\infty$, with zero mean and covariance matrix the inverse of the asymptotic Fisher information matrix $I(\theta_0)$.

The asymptotic Fisher information matrix is given by

$$I(\theta_0) = -E_{\theta_0}\left(\left.\frac{\partial^2 \log \mathbb{P}_\theta(Y_0|Y_{-1}, Y_{-2}, \ldots)}{\partial\theta_i \partial\theta_j}\right|_{\theta=\theta_0} \right)_{i,j},$$

see [29], and in [36] it is shown that $I(\theta_0)$ is nonsingular under assumption A2.

From Theorem 1 we immediately obtain the consistency and the asymptotic normality of the estimator of matrix $\tilde{\mathbf{P}}$.

Consistency

Proposition 4. *Under the Assumptions A1–A3, given a sample of observations $\{Y_1^N\}$, the maximum likelihood estimator of $\left(\tilde{P}((i,y_l),(j,y_h))\right)_{(i,y_l),(j,Y_h)\in\tilde{E}}$, that is $\left(\hat{\tilde{P}}((i,y_l),(j,y_h))\right)_{(i,y_l),(j,y_h)\in\tilde{E}}$, is strongly consistent as N tends to infinity.*

Proof. The transition probabilities for the two-dimensional process (X,Y) are obtained as $\tilde{P}((i,y_l),(j,y_h)) = P(i,j)Q_j(y_l,y_h)$ for all $i,j \in E$ and $y_l, y_h \in \mathcal{Y}$. From the vector of parameters θ we define the function

$$\Phi \;:\; [0,1]^{d\cdot(d-1)+d\cdot s\cdot(s-1)} \longrightarrow [0,1]^{d^2\cdot s^2}$$

such that $\Phi = \left(\Phi_{i,j,l,h}; i,j = 1,2,\ldots,d; l,h = 1,\ldots,s\right) \in [0,1]^{d^2\cdot s^2}$ with the following:

- $1 \le i,j \le d,\ j \ne d;\ 1 \le l,h \le s,\ h \ne s,$

$$\Phi_{i,j,l,h}(\theta) = P_{ij} Q_j(y_l, y_h); \tag{A.1}$$

- $1 \le i,j \le d,\ j \ne d;\ 1 \le l \le s,\ h = s,$

$$\Phi_{i,j,l,s}(\theta) = P_{ij}\left(1 - \sum_{h=1}^{s-1} Q_j(y_l, y_h)\right); \tag{A.2}$$

- $1 \le i \le d,\ j = d;\ 1 \le l,h \le s,\ h \ne s,$

$$\Phi_{i,d,l,h}(\theta) = \left(1 - \sum_{j=1}^{d-1} P_{ij}\right) Q_d(y_l, y_h); \tag{A.3}$$

- $1 \le i \le d,\ j = d;\ 1 \le l,h \le s,\ h = s,$

$$\Phi_{i,d,l,s}(\theta) = \left(1 - \sum_{j=1}^{d-1} P_{ij}\right)\left(1 - \sum_{h=1}^{s-1} Q_j(y_l, y_h)\right); \tag{A.4}$$

This function returns a vector whose components are the elements of matrix $\tilde{\mathbf{P}}$ conveniently sorted. Then, using the consistency of the estimator $\hat{\theta}_N$, which is expressed in Theorem 1 above, and applying the continuous mapping theorem to the function Φ defined above, we obtain the desired result. □

Asymptotic normality

Proposition 5. *Under the Assumptions A1 – A3, given a sample of observations $\{Y_1^N\}$, the random vector $\mathbf{F}_N = \left(F_{(i,l),(j,h)}; _{i,j\in E; l,h\in\mathcal{Y}}\right)$ such that*

$$F_{(i,l),(j,h)} = \sqrt{N}\left[\left(\hat{\tilde{P}}((i,y_l),(j,y_h))\right)_{i,j\in E; l,h\in\mathcal{Y}} - \left(\tilde{P}((i,y_l),(j,y_h))\right)_{i,j\in E; l,h\in\mathcal{Y}}\right]$$

is asymptotically Normal, as $N \to +\infty$ with 0 mean and covariance matrix $\Sigma_{\tilde{P}} = \Phi' \cdot \Sigma_\theta \cdot \Phi'^{\top}$, where Σ_θ is the covariance matrix of the random vector $\hat{\theta}_N$ and Φ is the function defined in (A.1)–(A.4) whose partial derivative matrix is denoted by Φ'.

Appendix B. The EM algorithm for the M1M1-DTHMM

Instead of directly solving the optimization problem (7), we consider a version of the EM-algorithm [7,10,29,37] for our context of M1M1-HMM, which can be summarized as follows.

Let us denote Θ the set of possible parameters of the model, and define the following function

$$\psi(\theta, \theta^{(0)}) = \mathbb{E}_{\theta(0)}\left[\log \mathbb{P}(X_1^N, Y_1^N \mid \theta, Y_1^N)\right] \tag{B.1}$$

with $\theta, \theta^{(0)} \in \Theta$. For $\theta^{(0)}$ given, the aim is to obtain $\theta^{(1)}$ such that

$$\theta^{(1)} = \arg\max_\theta\ \psi(\theta \mid \theta^{(0)}).$$

$$\begin{aligned}
\psi\left(\theta \mid \theta^{(0)}\right) &= \mathbb{E}_{\theta^{(0)}}\left[\log\left(\prod_{k=1}^N \mathbb{P}_\theta(X_{k-1}, X_k \mid Y_1^N)\mathbb{P}_\theta(X_k, Y_{k-1}, Y_k \mid Y_1^N)\right)\right] \\
&= \sum_{k=1}^N \sum_{i\in E}\sum_{j\in E} \log P_{ij}\mathbb{P}_{\theta^{(0)}}\left[X_{k-1} = i, X_k = j \mid Y_1^N\right] \\
&\quad + \sum_{k=1}^N \sum_{i\in E} \log Q_i(Y_{k-1}, Y_k)\mathbb{P}_{\theta^{(0)}}(X_k = i \mid Y_1^N) \\
&= \psi_1\left(\theta_1^{(1)} \mid \theta^{(0)}\right) + \psi_2\left(\theta_2^{(1)} \mid \theta^{(0)}\right).
\end{aligned}$$

We can split the ψ function and maximize separately in $\theta_1^{(1)}$ and $\theta_2^{(1)}$, and define $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)})$.

B.1. The M-step

At the m iteration of the algorithm we have an estimation of the unknown vector θ given by $\theta^{(m-1)} = \left(\theta_1^{(m-1)}, \theta_2^{(m-1)}\right)$. Then, we have to maximize $\psi\left(\theta \mid \theta^{(m-1)}\right)$ and obtain $\theta^{(m)}$.

First we consider $\psi_1\left(\theta_1 \mid \theta^{(m-1)}\right)$, that is,

$$\theta_1^{(m)} = \arg\max_{\theta_1} \sum_{k=1}^N \sum_{i,j\in E} \left(\log P_{ij}\right)\mathbb{P}_{\theta^{(m-1)}}\left[X_{k-1} = i, X_k = j \mid Y_1^N\right],$$

subject to $\sum_{j\in E} P_{ij} = 1$.

Using the method of Lagrange multipliers we obtain

$$P_{ij}^{(m)} = \frac{\sum_{k=1}^N \mathbb{P}_{\theta^{(m-1)}}\left(X_{k-1} = i, X_k = j \mid Y_1^N\right)}{\sum_{k=1}^N \mathbb{P}_{\theta^{(m-1)}}\left(X_{k-1} = i \mid Y_1^N\right)}$$

Second, let us consider $\psi_2\left(\theta_2 \mid \theta^{(m-1)}\right)$

$$\psi_2\left(\theta_2 \mid \theta^{(m-1)}\right) = \sum_{k=1}^N \sum_{i\in E}\sum_{y,y'\in\mathcal{Y}} \log Q_i(y,y')\mathbf{1}_{\{Y_{k-1}=y, Y_k=y'\}}\mathbb{P}_{\theta^{(m-1)}}(X_k = i \mid Y_1^N).$$

Again using Lagrange multipliers, we get

$$Q_i^{(m)}(y,y') = \frac{\sum_{k=1}^{N} \mathbb{P}_{\theta^{(m-1)}}\left(X_k = i \mid Y_1^N\right) \mathbf{1}_{\{Y_{k-1}=y, Y_k=y'\}}}{\sum_{k=1}^{n} \mathbb{P}_{\theta^{(m-1)}}\left(X_k = i \mid Y_1^N\right) \mathbf{1}_{\{Y_{k-1}=y\}}},$$

where $\mathbf{1}_{\{\cdot\}}$ takes value 1 if condition $\{\cdot\}$ is true and 0 otherwise.

B.2. The E-step

At the m iteration of the algorithm we have to calculate the following probabilities:

- $\mathbb{P}_{\theta^{(m)}}\left(X_{k-1} = i, X_k = j \mid Y_1^N\right)$; $\forall i, j \in E$, $\forall k = 1, \dots, N$
- $\mathbb{P}_{\theta^{(m)}}\left(X_k = i \mid Y_1^N\right)$; $\forall i \in E$, $\forall k = 1, \dots, N$

To do it we define the family of backward and forward probabilities:

1. **Forward probabilities**
   Let $k = 0, 1, \dots, n$, $i \in E$, we define
   $$F_k^{(m)}(i) = \mathbb{P}_{\theta^{(m)}}\left(Y_0^k, X_k = i\right).$$
   Then we have $\left\{F_k^{(m)}(i),\ i \in E;\ k = 0, 1, \dots, N\right\}$, that satisfies the following recurrence equation
   $$F_k^{(m)}(i) = \sum_{j \in E} F_{k-1}^{(m)}(j) P_{ji}^{(m)} Q_i^{(m)}\left(Y_{k-1}, Y_k\right);\ \forall k = 1, \dots, N,\ \forall i \in E.$$
   For $k = 0$ we get that $F_0^{(m)}(i) = \alpha(i)\beta_i(Y_0)$.
2. **Backward probabilities**
   For $k = 0, 1, \dots, N-1$, $i \in E$, let us define
   $$B_k^{(m)}(i) = \mathbb{P}_{\theta^{(m)}}\left(Y_{k+1}^N \mid X_k = i, Y_k\right).$$
   Then we have a family of probabilities $\left\{B_k^{(m)}(i),\ i \in E,\ k = 0, \dots, N\right\}$, satisfying
   $$B_k^{(m)}(i) = \sum_{j \in E} P_{ij}^{(m)} Q_j^{(m)}\left(Y_k, Y_{k+1}\right) B_{k+1}^{(m)}(j).$$
   for $k = 0, \dots, N-1$ and $B_N^{(m)}(i) = 1$, $\forall i \in E$.
   Also we have that $\mathbb{P}_{\theta^{(m)}}(Y_1^N) = \sum_{i \in E} B_k^{(m)}(i) F_k^{(m)}(i)$, for any $k = 0, \dots, N$, and, in particular, $\mathbb{P}_{\theta^{(m)}}(Y_1^N) = \sum_{i \in E} F_N(i)$.
   Finally, for $k = 1, \dots, N$,
   $$\mathbb{P}_{\theta^{(m)}}\left(X_k = i \mid Y_1^N\right) = \frac{F_k^{(m)}(i) B_k^{(m)}(i)}{\sum_{j \in E} F_k^{(m)}(j) B_k^{(m)}(j)},$$
   and
   $$\mathbb{P}_{\theta^{(m)}}\left(X_{k-1} = i, X_k = j \mid Y_1^N\right) = \frac{F_{k-1}^{(m)}(i) P_{ij}^{(m)} Q_j^{(m)}\left(Y_{k-1}, Y_k\right) B_k^{(m)}(j)}{\sum_{h \in E} B_k^{(m)}(h) F_k^{(m)}(h)}.$$

References

[1] Duan C, Li Y, Pu H, Luo J. Adaptative monitoring scheme of stochastically failing systems under hidden degradation processes. Reliab Eng Syst Saf 2022;221:108322.
[2] Wang Y, Lee IC, Hong Y, Deng X. Building degradation index with variable selection for multivariate sensory data. Reliab Eng Syst Saf 2022;227:108704.
[3] Li N, Gebraeel N, Lei Y, Fang X, Cai X, Yan T. Remaining useful life prediction based on a multi-sensor data fusion model. Reliab Eng Syst Saf 2021;208:107249.
[4] Wu H, Li YF. A multi-sensor fusion-based prognostic model for systems with partially observable failure modes. IISE Trans 2023;1–21.
[5] Mor B, Garhwal S, Kumar A. A systematic review of hidden Markov models and their applications. Arch Comput Method Eng 2021;28:1429–48.
[6] Ramezani SB, Killen B, Cummins L, Rahimi S, Amirlatifi A, Seale M. A survey of HMM-based algorithms in machinery fault prediction. In: IEEE symposium series on computational intelligence. 2021, p. 1–9.
[7] Gamiz ML, Limnios N, Segovia-Garcia MC. Hidden Markov models in reliability and maintenance. Eur J Oper Res 2023;304:1242–55.
[8] Gamiz ML, Limnios N, Segovia-Garcia MC. The continuous-time hidden Markov model based on discretization. Properties of estimators and applications. In: Stat inference stoch process. 2023 [in press].
[9] Chen Z, Li Z, Xia T, Pan E. Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy. Reliab Eng Syst Saf 2019;184:123–36.
[10] Berchtold A. The double chain Markov model. Commun Stat Theory Methods 1999;28(11):2569–89.
[11] Cui L, Li Y, Shen J, Lin C. Reliability for discrete state systems with cyclic missions periods. Appl Math Model 2016;40:10783–99.
[12] Nagy I, Suzdaleva E. Mixture estimation with state-space components and Markov model of switching. Appl Math Model 2013;37(44):9970–84.
[13] Castillo C, Castillo E, Fernández-Canteli A, Gómez R, Molina R. Rainflow analysis in coastal engineering using switching second order Markov models. Appl Math Model 2012;36(9):4286–303.
[14] Cheng G, Li L, Shangguan C, Yang N, Jiang B, Tao N. Optimal joint inspection and mission abort policy for a partially observable system. Reliab Eng Syst Saf 2023;229:108870.
[15] Coraça EM, Ferreira JV, Nobrega EGO. An unsupervised structural health monitoring framework based on variational autoencoders and hidden Markov models. Reliab Eng Syst Saf 2023;231:109025.
[16] Danisman O, Kocer UU. Hidden Markov models with binary dependence. Physica A 2021;567:125668.
[17] Ghasvarian Jahromi K, Gharavian D, Mahdiani HR. Wind power prediction based on wind speed forecast using hidden Markov model. J Forecast 2023;42(1):101–23.
[18] Gualeni P, Vairo T. A prediction tool for maintenance costs estimation during the design process of a ship engine room. J Ocean Eng Mar Energy 2023;1–11.
[19] Guo C, Liang Z. A predictive Markov decision process for optimizing inspection and maintenance strategies of partially observable multi-state systems. Reliab Eng Syst Saf 2022;226:108683.
[20] Habayeb M, Murtaza SS, Miranskyy A, Bener AB. On the use of hidden Markov model to predict the time to fix bugs. IEEE Trans Softw Eng 2018;44(12):1224–44.
[21] Khan MA, Abuhasel KA. An evolutionary multi-hidden Markov model for intelligent threat sensing in industrial internet of things. Journal Supercomput 2021;77(6):6236–50.
[22] Li R, Arzaghi E, Abbassi R, Chen D, Li C, Li H, et al. Dynamic maintenance planning of a hydro-turbine in operational life cycle. Reliab Eng Syst Saf 2020;204:107129.
[23] Lin M, You Y, Wang W, Wu J. Battery health prognosis with gated recurrent unit neural networks and hidden Markov model considering uncertainty quantification. Reliab Eng Syst Saf 2023;230:108978.
[24] Martindale CF, Hoenig F, Strohrmann C, Eskofier BM. Smart annotation of cyclic data using hierarchical hidden Markov models. Sensors 2021;17(10):1–16.
[25] Soleimani M, Campean F, Neagu D. Integration of hidden Markov modelling and Bayesian network for fault detection and prediction of complex engineered systems. Reliab Eng Syst Saf 2021;215:107808.
[26] Zhao Y, Gao W, Smidts C. Sequential Bayesian inference of transition rates in the hidden Markov model for multi-state system degradation. Reliab Eng Syst Saf 2021;214:107662.
[27] Liu S, Zheng K, Zhao L, Fan P. A driving intention prediction method based on hidden Markov model for autonomous driving. Comput Commun 2020;157:143–9.
[28] Xu R, Chen X, Zhang F. Green technology innovation ecosystem based on hidden Markov model. Ekoloji 2019;28(107):1729–36.
[29] Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains. Ann Math Stat 1966;37(6):1554–63.
[30] Bickel BJ, Ritov Y, Rydén T. Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. Ann Stat 1998;26:1614–35.
[31] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989;77(2):257–86.
[32] Pepe MS. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press; 2003.
[33] Smith G. Instantaneous states of Markov processes. Trans Amer Math Soc 1964;110:185–95.
[34] Alagarsamy P. Predict pump failure before it happens using deep learning model. 2021, In https://becominghuman.ai/predict-pump-failure-before-it-happens-using-deep-learning-model-dc886bfa073e. [Last Accessed 2022].
[35] Barbu VS, Limnios N. Semi-Markov chains and hidden semi-Markov models toward applications. Springer; 2008.
[36] Douc R. Non singularity of the asymptotic fisher information matrix in hidden Markov models. École Polytechnique; 2005, preprint.
[37] Cappé O, Moulines E, Rydén T. Inference in hidden Markov models. Springer Ser. Stat., Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2005.