



HAL
open science

Estimating the parameters of a seasonal Markov-modulated Poisson process

Armelle Guillou, Stéphane Loisel, Gilles Stupfler

► **To cite this version:**

Armelle Guillou, Stéphane Loisel, Gilles Stupfler. Estimating the parameters of a seasonal Markov-modulated Poisson process. 2014. hal-00965279v1

HAL Id: hal-00965279

<https://hal.science/hal-00965279v1>

Preprint submitted on 24 Mar 2014 (v1), last revised 24 Feb 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating the parameters of a seasonal Markov-modulated Poisson process

Armelle Guillou⁽¹⁾, Stéphane Loisel⁽²⁾ & Gilles Stupfler⁽³⁾

⁽¹⁾ Université de Strasbourg & CNRS, IRMA, UMR 7501,
7 rue René Descartes, 67084 Strasbourg Cedex, France

⁽²⁾ Université Lyon 1, Institut de Science Financière et d'Assurances,
50 avenue Tony Garnier, 69007 Lyon, France

⁽³⁾ Aix Marseille Université, CERGAM, EA 4225,
15-19 allée Claude Forbin, 13628 Aix-en-Provence Cedex 1, France

Abstract. We present a new model of counting processes in insurance. The process is a Markov-modulated Poisson process featuring seasonality. We prove the strong consistency and the asymptotic normality of a maximum split-time likelihood estimator of the parameters of this model, and present an algorithm to compute it in practice. The method is illustrated on a simulation study.

Keywords: Markov-modulated Poisson process, seasonality, split-time likelihood, strong consistency, asymptotic normality.

MSC 2010 Subject Classifications: Primary 62M05, 62F12; Secondary 60F05, 60F15.

1 Introduction

It is often the case that the insurance claim frequency is impacted by environment variables. For instance, flood risk is higher in a period of frequent heavy rains, and fire risk is more intense when the weather is particularly dry. Such environment variables may be hidden to

some extent to the practitioner: for instance, it is now accepted that the probabilities of severe floods in Australia, strong snowstorms in North America or hurricanes on the East Coast of the United States increase during La Niña episodes (see Neumann *et al.* [13], Cole and Pfaff [3], Parisi and Lund [15] and Landreneau [6]). This is now taken seriously by most reinsurers as well as Lloyd's and the UK Met Office [9]. However, observing and understanding the role of those variables is not easy, which makes it realistic to consider these variables as unobserved so far.

To take such a dependency into account, one may for instance assume that the underlying environment process is a Markov process J in continuous time and that in each state of J , the claim counting process N is a Poisson process. The resulting bivariate process (J, N) is then called a Markov-Modulated Poisson Process (MMPP). The idea of considering a Markov modulation was first introduced by Asmussen [2]; the obtained model can capture the fact that the insurance claim frequency may be modified if climatic, political or economic factors change. Such a model has gained considerable attention recently: see for instance Lu and Li [11], Ng and Yang [14], Zhu and Yang [22] and Wei *et al.* [21]. The parameters of an MMPP are often estimated using a Maximum Likelihood Estimator (MLE), whose consistency was proved in Rydén [17]. Various methods have been suggested to compute the MLE; a standard tool is the Expectation-Maximization (EM) algorithm, see Rydén [20] for the implementation of this procedure for the estimation of the parameters of an MMPP. We finally mention that in a recent paper, Guillou *et al.* [4] introduced a new MMPP-driven loss process in insurance with several lines of business, showed the strong consistency of the MLE and fitted their model to real sets of insurance data using an adaptation of the EM algorithm.

Furthermore, many examples of practical applications in insurance display some sort of seasonal variation. For example, theft in garages are more frequent before Christmas as people tend to store Christmas gifts in them, fire risk is more intense in the summer, and hurricanes occur mostly between June and November on the East Coast of the United States. These random, cyclic factors and their impact on insurance risk, which need to be taken into account to carry out a proper regime switching analysis, are yet to be understood and forecasted. In an inhomogeneous context with deterministic intensity function, Lu and Garrido [10] have fitted double-periodic Poisson intensity rates to hurricane data, for particular parametric forms (like

double-beta and sine-beta intensities) to hurricane data. Helmers *et al.* [5] have provided an in-depth theoretical statistical analysis of such doubly periodic intensities. We aim at carrying out a theoretical statistical analysis in a stochastic intensity framework with seasonality.

In this paper, we thus consider an MMPP featuring seasonality, and study estimation issues for this process when the environment process is unobserved. A problem with this type of process is that contrary to when there is no seasonality, the random sequence of the inter-event times is not ergodic. Studying the asymptotic properties of the MLE, as Rydén [17] and Guillou *et al.* [4] do, is thus very difficult; furthermore, the particular structure of the likelihood makes it hard to compute the MLE in practice. To tackle this issue, we borrow an idea of Rydén [18, 19] and we introduce a Split-Time Likelihood (STL). The logarithm of this quantity is shown to be a sum of ergodic random variables; maximizing the STL then yields a Maximum Split-Time Likelihood Estimator (MSTLE) whose consistency and asymptotic normality can be proven using regenerative theory.

The outline of the paper is as follows. We give a precise definition of our model in Section 2. In Section 3, we define our estimator, whose asymptotic properties are studied in Section 4. We explain how to implement our estimation technique in practice in Section 5, and we study the numerical behavior of this estimator on a simulation study in Section 6. Proofs of the main results are deferred to Appendix A and those of the preliminary results to Appendix B.

2 The model

We consider an irreducible continuous-time Markov process J with generator L on the state space $\{1, \dots, r\}$, where $r \geq 2$. Consider further a counting process N and real numbers $\tau_0 = 0 < \tau_1 < \dots < \tau_{k-1} < 1 = \tau_k$ such that for every $q \in \mathbb{N}$, on the intervals $[q + \tau_{s-1}, q + \tau_s)$, if J is in state i , then N is a Poisson process with jump rate $\lambda_i^{(s)}$, where $\lambda_i^{(s)} \geq 0$. The time interval $[q + \tau_{s-1}, q + \tau_s)$ represents season s of the period $q + 1$.

The context of our work is the following: let us assume that the process N has been observed until time $n \in \mathbb{N} \setminus \{0\}$, so that the available data consists of

1. the number r of states of J and the times τ_j ,

2. the full knowledge of the process N between time 0 and time n .

The goal is to estimate the unknown parameters of the model, namely the elements ℓ_{ij} of the transition intensity matrix L of J and the jump intensities $\lambda_i^{(s)}$ of N . Since the process J is not observed, estimating the parameters of this model is not straightforward. For the sake of shortness, we let Φ be the global parameter of the model, the distribution of the process with parameter Φ is denoted by \mathbb{P}_Φ and we let \mathcal{E} be the parameter space

$$\mathcal{E} = \left\{ \Phi \mid L(\Phi) \text{ is irreducible and } \min_{1 \leq i \leq r} \min_{1 \leq s \leq k} \lambda_i^{(s)}(\Phi) > 0 \right\}.$$

The space \mathcal{E} can be seen as the set of those parameters for which in any state of J and in any season, an event can occur with positive probability. Note that an element of \mathcal{E} consists in $|\mathcal{E}| = r(r-1) + rs$ parameters. Before proceeding, we give some useful notation: for $1 \leq s \leq k$ and $0 \leq y \leq \tau_s - \tau_{s-1}$, let

$$\begin{aligned} f(y, s, \Phi) &= \exp(y(L(\Phi) - \Lambda^{(s)}(\Phi)))\Lambda^{(s)}(\Phi) \\ \text{and } \bar{F}(y, s, \Phi) &= \exp(y(L(\Phi) - \Lambda^{(s)}(\Phi))), \end{aligned}$$

where $\Lambda^{(s)}(\Phi) = \text{diag}(\lambda_1^{(s)}(\Phi), \dots, \lambda_r^{(s)}(\Phi))$.

3 Estimating the parameters with an MSTLE

It is impossible to apply any known results on MMPPs here since the jump intensities change as time goes by; especially, given that J is in state j , N is an inhomogeneous Poisson process. To overcome this issue, we introduce some notation: let $W_{q,s}$ be the number of jumps of N during season s of the period q , and let $T_1^{(q,s)}, \dots, T_{W_{q,s}}^{(q,s)}$ be the successive jump times of N during this season. Let $Y_l^{(q,s)} = T_l^{(q,s)} - T_{l-1}^{(q,s)}$ be the inter-event times (with $T_0^{(q,s)} = (q-1) + \tau_{s-1}$). It is assumed that the starting distribution of J is its unique stationary distribution a on $\{1, \dots, r\}$, that is, the only row vector a such that $aL = 0$ and the sum of the entries of a is equal to 1.

Let $Z_q = (W_{q,s}, Y_1^{(q,s)}, \dots, Y_{W_{q,s}}^{(q,s)})_{1 \leq s \leq k}$ be the random vector representing the information available for period number q . With this notation, the process $(Z_q)_{q \geq 1}$ is stationary. Besides, given the states of the irreducible Markov chain $(J(q))_{q \in \mathbb{N}}$, the random variables $Z_q, q \geq 1$ are independent, so that applying Lemma 1 in Leroux [8] shows that the process $(Z_q)_{q \geq 1}$ is ergodic.

We denote by $Z = (W_s, Y_1^{(s)}, \dots, Y_{w_s}^{(s)})_{1 \leq s \leq k}$ a random vector which shares the distribution of the Z_q , $q \geq 1$.

Let then $\mathcal{L}_1(Z, \Phi)$ be the likelihood of the observations over one period, computed under the parameter Φ : if $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$,

$$\begin{aligned} \mathcal{L}_1(z, \Phi) = & \sum_{i_0, \dots, i_{k+1}} a_{i_0}(\Phi) \prod_{s=1}^k \left[e'_{i_{s-1}} \prod_{l=1}^{w_s} f(y_l^{(s)}, s, \Phi) \right. \\ & \left. \times \bar{F} \left(\tau_s - \tau_{s-1} - \sum_{l=1}^{w_s} y_l^{(s)}, s, \Phi \right) e_{i_s} \right] \end{aligned} \quad (1)$$

where e_j is the column vector of size r having all entries equal to 0 except the j th one. Following Rydén [19], we let

$$\mathcal{STL}_n(\Phi) = \prod_{q=1}^n \mathcal{L}_1(Z_q, \Phi)$$

be the split-time likelihood (STL) of the observations. If the random variables Z_q , $q \geq 1$ were independent, the STL would be the total likelihood of the observations; we shall in fact show that the Z_q are independent given a suitable sequence of increasing random times, which is the key idea to the proofs of our main results, see Appendix A. An MSTL is then any parameter that maximizes the STL, or equivalently the log-STL

$$\log \mathcal{STL}_n(\Phi) = \sum_{q=1}^n \log \mathcal{L}_1(Z_q, \Phi)$$

and an MSTLE is any estimator of an MSTL.

4 Asymptotic results

We shall write $\Phi \sim \Phi'$ whenever the distributions of Z under \mathbb{P}_Φ and $\mathbb{P}_{\Phi'}$ agree. The following consistency result then holds:

Theorem 1. *Let $\Phi_0 \in \mathcal{E}$ be the true value of the parameter and $\overline{\Phi_0}$ be the equivalence class of Φ_0 modulo \sim . Let K be a compact subset of \mathcal{E} such that $\Phi_0 \in K$ and $\widehat{\Phi}_n$ be the MSTLE for Φ_0 on K , computed over n periods. Then if $O \subset K$ is an open set containing $K \cap \overline{\Phi_0}$, one has $\widehat{\Phi}_n \in O$ almost surely for n large enough.*

We now wish to obtain an asymptotic normality result for our estimator. In what follows, we pick $i_0 \in \{1, \dots, r\}$ and we let ω_k be the successive times when the Markov chain $(J(q))$ reaches i_0 :

$$\omega_1 = \min\{q \geq 1 \mid J(q) = i_0\} \quad \text{and} \quad \forall k \geq 1, \quad \omega_{k+1} = \min\{q > \omega_k \mid J(q) = i_0\}.$$

Let further $\mathbb{P}_\Phi^{i_0}(\cdot) = \mathbb{P}_\Phi(\cdot \mid J(0) = i_0)$ be the probability measure deduced from \mathbb{P}_Φ given that J starts at i_0 .

Note that for every $\Phi \in \mathcal{E}$, since $L(\Phi)$ is the generator of an irreducible continuous-time Markov chain on a finite state space, then 0 is an eigenvalue of the transpose $L'(\Phi)$ of $L(\Phi)$ with multiplicity 1 and related normalized eigenvector $a'(\Phi)$. Since the map $\varphi \mapsto L'(\varphi)$ is infinitely continuously differentiable in a neighborhood of Φ , a straightforward extension of Theorem 8 in Chapter 9 of Lax [7] shows that the map $\varphi \mapsto a'(\varphi)$ is infinitely continuously differentiable in a neighborhood of Φ . The function $\varphi \mapsto \log \mathcal{L}_1(Z, \varphi)$ is thus infinitely continuously differentiable in a neighborhood of the true parameter Φ_0 ; if $\mathbb{E}_\Phi^{i_0}$ denotes the expectation under the measure $\mathbb{P}_\Phi^{i_0}$, we can set for Φ close enough to Φ_0

$$\begin{aligned} h_i(z, \Phi) &= \frac{\partial \log \mathcal{L}_1}{\partial \varphi_i}(z, \Phi), \\ A_{ij}(\Phi) &= \mathbb{E}_\Phi^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi) h_j(Z_q, \Phi) \right), \\ V_{ij}(\Phi) &= \mathbb{E}_\Phi^{i_0} \left(\sum_{p,q=1}^{\omega_1} h_i(Z_p, \Phi) h_j(Z_q, \Phi) \right). \end{aligned}$$

We assume that the matrix $A(\Phi) = (A_{ij}(\Phi))$ is invertible for $\Phi = \Phi_0$ and we let $V(\Phi) = (V_{ij}(\Phi))$, $C(\Phi) = \frac{1}{a_{i_0}(\Phi)} A^{-1}(\Phi) V(\Phi) A^{-1}(\Phi)$ for Φ in a neighborhood of Φ_0 .

Our asymptotic normality result for $\widehat{\Phi}_n$ is as follows:

Theorem 2. *Let $\Phi_0 \in \mathcal{E}$ be the true value of the parameter and $\overline{\Phi_0}$ be the equivalence class of Φ_0 modulo \sim . Assume that Φ_0 lies in the interior of \mathcal{E} and let K be a compact subset of \mathcal{E} , whose interior contains Φ_0 , such that $K \cap \overline{\Phi_0} = \{\Phi_0\}$. Let $\widehat{\Phi}_n$ be the MSTLE computed on K over n periods. Then*

$$\sqrt{n}(\widehat{\Phi}_n - \Phi_0) \xrightarrow{d} \mathcal{N}(0, C(\Phi_0)) \quad \text{as } n \rightarrow \infty.$$

5 Practical computations

The MSTLE is typically computed using a (quasi-)Newton algorithm. Such an algorithm being iterative, we give a method to start it in practice. Let $w_{q,s}$ be the number of jumps of N during season s of the period q , $t_1^{(q,s)}, \dots, t_{w_{q,s}}^{(q,s)}$ be the successive jump times of N during this season and let $u_l^{(q,s)} = t_l^{(q,s)} - \tau_{s-1} - (q-1)(1 - (\tau_s - \tau_{s-1}))$, $1 \leq l \leq w_{q,s}$.

For all $1 \leq s \leq k$, the data $D_s = (u_l^{(q,s)})_{q,l}$ is regarded as a sample of a univariate MMPP with parameters $\Phi^{(s)} = (L, \lambda_1^{(s)}, \dots, \lambda_r^{(s)})$. An EM algorithm is then used to provide a first estimate of $\Phi^{(s)}$; the initial estimation procedure is a straightforward adaptation of the one in Guillou *et al.* [4].

Let $(\tilde{L}^{(s)}, \tilde{\lambda}_1^{(s)}, \dots, \tilde{\lambda}_r^{(s)})$ be the EM estimations obtained by this way. The EM algorithms involved may have switched some of the parameters: we thus note, if $M = (m_{ij})$ is a square matrix having size r and σ is a permutation of $\{1, \dots, r\}$, $M \circ \sigma = (m_{\sigma(i)\sigma(j)})$. Let, for such permutations $\sigma_1, \dots, \sigma_k$,

$$\tilde{L}^{(\sigma_1, \dots, \sigma_k)} = \frac{1}{k} \sum_{s=1}^k \tilde{L}^{(s)} \circ \sigma_s$$

and

$$\tilde{\Phi}^{(\sigma_1, \dots, \sigma_k)} = (\tilde{L}^{(\sigma_1, \dots, \sigma_k)}, (\tilde{\lambda}_{\sigma_s(1)}^{(s)}, \dots, \tilde{\lambda}_{\sigma_s(r)}^{(s)})_{1 \leq s \leq k}).$$

We then run the optimization algorithm starting from each of these values: in doing so, we obtain estimates $\hat{\Phi}^{(\sigma_1, \dots, \sigma_k)}$ of the parameters. Finally, we compute

$$(\sigma_1^0, \dots, \sigma_k^0) = \arg \max_{\sigma_1, \dots, \sigma_k} \mathcal{STL}_n(\hat{\Phi}^{(\sigma_1, \dots, \sigma_k)})$$

and our MSTLE is $\hat{\Phi} = \hat{\Phi}^{(\sigma_1^0, \dots, \sigma_k^0)}$.

6 Simulation study and discussion of results

In this section, we examine how our estimator behaves on a finite sample situation. We are motivated by the impact of El Niño-La Niña cycle on hurricane risk. We choose reasonable parameters for a simple illustration, but cannot claim that those parameters are estimated on a real dataset, simply because it is very hard to understand the impact of this effect on

hurricane frequency. We consider the situation $r = 2$, that is, the underlying Markov process J is a two-state continuous-time Markov process. We further choose $k = 2$ and $\tau_1 = 1/2$, so that any period is divided into two seasons of equal length. The following cases are considered:

- Case 1: $\ell_{12} = 1, \ell_{21} = 2, \lambda_1^{(1)} = 1, \lambda_2^{(1)} = 5, \lambda_1^{(2)} = 5, \lambda_2^{(2)} = 25$.
- Case 2: $\ell_{12} = 3, \ell_{21} = 10, \lambda_1^{(1)} = 1, \lambda_2^{(1)} = 5, \lambda_1^{(2)} = 5, \lambda_2^{(2)} = 25$.
- Case 3: $\ell_{12} = 1, \ell_{21} = 2, \lambda_1^{(1)} = 1/2, \lambda_2^{(1)} = 5, \lambda_1^{(2)} = 5/2, \lambda_2^{(2)} = 25$.

For each of these cases, we consider the situations $n = 50$ and $n = 100$. Our estimation procedure is carried out on $S = 100$ replications of the considered process. In each case, we compute the mean and median L^1 -error related to each parameter. The results are shown in Table 1.

One can see that if an insurance company had around 50 or 100 years of data, which would be an ideal situation, for intensities like the ones we chose (5 times as many hurricanes during the hurricane season as in the other one, and 5 times as many hurricanes during a bad El Niño-La Niña phase as in a favorable one), it is feasible to estimate the parameters with an average relative error of around 10 to 20%. Following the intuition, parameters are estimated more accurately when regime switching is more frequently observed and when the ratio of the claim frequencies in different states is further from 1. Of course, we have taken here a quite favorable case, as we have only two states of the environment and two seasons with fixed intensities. When considering 5 or 10 states and 4 seasons, we would need a period of observation too long for the approach to be reasonable. However, in most hidden Markov models in insurance and finance, the number of states is 2 or 3. In the near future, one could probably refine these estimations thanks to partial observations of the environment process, as climatologists understand better and better El Niño-La Niña phenomenon and its impact on insurance perils. Eventually, let us note that the Markov assumption is of course questionable, as well as the absence of impact of climate change on the parameters. Extending the number of states in order to get a Markov process is feasible in theory, but would lead to estimation problems that would not be tractable in practice. Another problem is that in different states of the environment, the hurricane season could be longer or shorter, as conditions are more or less met for hurricanes to form. This is not considered here and would require further theoretical

analysis.

References

- [1] Asmussen, S. (1987) *Applied probability and queues*, Wiley, New York.
- [2] Asmussen, S. (1989) Risk theory in a Markovian environment, *Scandinavian Actuarial Journal* **2**: 69–100.
- [3] Cole, J.D. and Pfaff, S.R. (1997) A climatology of tropical cyclones affecting the Texas coast during El Niño/non-El Niño years: 1990-1996, *Technical Attachment SR/SSD 97-37*, National Weather Service Office, Corpus Christi, Texas. Available at <http://www.srh.noaa.gov/topics/attach/html/ssd97-37.htm>.
- [4] Guillou, A., Loisel, S., Stupfler, G. (2013) Estimation of the parameters of a Markov-modulated loss process in insurance, *Insurance: Mathematics and Economics* **53**: 388–404.
- [5] Helmers, R., Mangku, I.W., Zitikis, R. (2007) A non-parametric estimator for the doubly periodic Poisson intensity function, *Statistical Methodology* **4**: 481–492.
- [6] Landreneau, D. (2001) Atlantic tropical storms and hurricanes affecting the United States: 1899-2000, *NOAA Technical Memorandum NWS SR-206 (updated through 2002)*, National Weather Service Office, Lake Charles, Louisiana. Available at <http://www.srh.noaa.gov/lch/?n=tropical>.
- [7] Lax, P.D. (2007) *Linear Algebra and its applications*, Wiley, New York.
- [8] Leroux, B.G. (1992) Maximum-likelihood estimation for hidden Markov models, *Stochastic Processes and their Applications* **40**: 127–143.
- [9] Lloyd's (2010) *Forecasting risk. The value of long-range forecasting for the insurance industry*, joint report with the UK Met Office.
- [10] Lu, Y., Garrido, J. (2005) Doubly periodic non-homogeneous models for hurricane data, *Statistical Methodology* **2**: 17–35.

- [11] Lu, Y., Li, S. (2005) On the probability of ruin in a Markov-modulated risk model, *Insurance: Mathematics and Economics* **37**: 522–532.
- [12] Meier-Hellstern, K.S. (1987) A fitting algorithm for Markov-modulated Poisson processes having two arrival rates, *European Journal of Operational Research* **29**: 970–977.
- [13] Neumann, C.J., Jarvinen, B.R., McAdie, C.J., Elms, J.D. (1993) Tropical Cyclones of the North Atlantic Ocean, 1871-1992, *Historical Climatology Series 6-2*, National Climatic Data Center, Asheville, North Carolina.
- [14] Ng, A., Yang, H. (2006) On the joint distribution of surplus prior and immediately after ruin under a Markovian regime switching model, *Stochastic Processes and their Applications* **116**: 244–266.
- [15] Parisi, F., Lund, R. (2000) Seasonality and return periods of landfalling Atlantic basin hurricanes, *Australian & New Zealand Journal of Statistics* **42**(3): 271–282.
- [16] Rao, C.R. (1973) *Linear statistical inference and its applications*, Wiley, New York.
- [17] Rydén, T. (1994) Parameter estimation for Markov modulated Poisson processes, *Communications in Statistics. Stochastic Models* **10**(4): 795–829.
- [18] Rydén, T. (1994) Consistent and asymptotically normal parameter estimates for hidden Markov models, *Annals of Statistics* **22**(4): 1884–1895.
- [19] Rydén, T. (1995) Consistent and asymptotically normal parameter estimates for Markov modulated Poisson processes, *Scandinavian Journal of Statistics* **22**(3): 295–303.
- [20] Rydén, T. (1996) An EM algorithm for estimation in Markov-modulated Poisson processes, *Computational Statistics and Data Analysis* **21**(4): 431–447.
- [21] Wei, J., Yang, H., Wang, R. (2010) On the Markov-modulated insurance risk model with tax, *Blätter der DGVFM* **31**(1): 65–78.
- [22] Zhu, J., Yang, H. (2008) Ruin theory for a Markov regime-switching model under a threshold dividend strategy, *Insurance: Mathematics and Economics* **42**: 311–318.

Appendix A: proofs of the main results

Proof of Theorem 1. The proof is similar to that of Theorem 1 in Rydén [19]: let $\Phi \in \mathcal{E}$ be such that $\Phi \not\sim \Phi_0$ and G_Φ be a neighborhood of Φ as in Lemma 2. If $B(\Phi, 1/q)$ denotes the open ball with center Φ and radius $1/q$, the continuity of the map $\varphi \mapsto \mathcal{L}_1(Z, \varphi)$ yields

$$\sup_{\varphi \in G_\Phi \cap B(\Phi, 1/q)} \log \mathcal{L}_1(Z, \varphi) \rightarrow \log \mathcal{L}_1(Z, \Phi) \text{ as } q \rightarrow \infty.$$

Noticing that

$$\left| \sup_{\varphi \in G_\Phi \cap B(\Phi, 1/q)} \log \mathcal{L}_1(Z, \varphi) \right| \leq \left| \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right| + |\log \mathcal{L}_1(Z, \Phi)|$$

the dominated convergence theorem implies

$$\mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi \cap B(\Phi, 1/q)} \log \mathcal{L}_1(Z, \varphi) \right] \rightarrow \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi)] \text{ as } q \rightarrow \infty. \quad (2)$$

Since $\Phi \not\sim \Phi_0$, the information inequality (see Rao [16]) gives

$$\mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi)] + 2\varepsilon < \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] \quad (3)$$

for some $\varepsilon > 0$. It is thus a consequence of (2) and (3) that there exists a (possibly different) neighborhood G_Φ of Φ with

$$\mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right] \leq \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] - \varepsilon. \quad (4)$$

Besides, since $(Z_q)_{q \geq 1}$ is ergodic,

$$\begin{aligned} \frac{1}{n} \sup_{\varphi \in G_\Phi} \log \mathcal{STL}_n(\varphi) &\leq \frac{1}{n} \sum_{q=1}^n \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z_q, \varphi) \rightarrow \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right] \\ \text{and } \frac{1}{n} \log \mathcal{STL}_n(\Phi_0) &= \frac{1}{n} \sum_{q=1}^n \log \mathcal{L}_1(Z_q, \Phi_0) \rightarrow \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] \end{aligned}$$

almost surely as $n \rightarrow \infty$. Hence

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sup_{\varphi \in G_\Phi} \log \mathcal{STL}_n(\varphi) \leq \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right] \leq \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] - \varepsilon$$

almost surely as $n \rightarrow \infty$, by (4). Finally, remark that the compact set $O^c \cap K$, where O^c is the complement of O , may be covered by a finite number of such neighborhoods G_{Φ_i} , $1 \leq i \leq d$;

this yields

$$\begin{aligned} & \sup_{\varphi \in O^c \cap K} \{\log \mathcal{STL}_n(\varphi) - \log \mathcal{STL}_n(\Phi_0)\} \\ & \leq \max_{1 \leq i \leq d} \left\{ \sup_{\varphi \in G_{\Phi_i}} \log \mathcal{STL}_n(\varphi) - \log \mathcal{STL}_n(\Phi_0) \right\} \rightarrow -\infty \end{aligned}$$

almost surely as $n \rightarrow \infty$. As a consequence, necessarily $\widehat{\Phi}_n \in O$ for n large enough, which completes the proof. \blacksquare

Before proving Theorem 2, we highlight that the random process $(Z_q)_{q \geq 1}$ is regenerative with associated cycle lengths $(C_q = \omega_q - \omega_{q-1})_{q \geq 1}$ (where we set $\omega_0 = 0$ for convenience), that is:

- for each $l \geq 2$, the random process $(C_{l+1+q}, Z_{\omega_{l+q}})_{q \geq 1}$ is independent of the ω_j , $1 \leq j \leq l-1$ and its distribution does not depend on l ;
- given $(\omega_q)_{q \geq 1}$, $q \geq 1$, the random vectors Z_j , $\omega_{q-1} \leq j \leq \omega_q - 1$ are independent.

Let $E = (\bigcup_{m=0}^{\infty} \{m\} \times \mathbb{R}^m)^k$ be the set in which Z_1 takes its values. Note that every $\mathbb{P}_{\Phi_0}^{i_0}$ -moment of ω_1 is finite (see Lemma 3) and that $\mathbb{E}_{\Phi_0}^{i_0}(\omega_1) = 1/a_{i_0}(\Phi_0) > 0$ because $a(\Phi_0)$ is the stationary distribution of the Markov chain $(J(q))$. We may now state a law of large numbers and a central limit theorem which are direct consequences of Theorems 2 and 3 in Rydén [19]; similar results can be found in Asmussen [1].

Proposition 1. *Let $g = (g_1, \dots, g_r) : E \rightarrow \mathbb{R}^r$, $r \geq 1$ be a measurable function. If $\mathbb{E}_{\Phi_0}^{i_0} \left| \sum_{q=1}^{\omega_1} g_i(Z_q) \right| < \infty$ for every $i \in \{1, \dots, r\}$, then*

$$\frac{1}{n} \sum_{q=1}^n g(Z_q) \xrightarrow{\mathbb{P}} a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} g(Z_q) \right) \text{ as } n \rightarrow \infty.$$

Proposition 2. *Let $g = (g_1, \dots, g_r) : E \rightarrow \mathbb{R}^r$, $r \geq 1$ be a measurable function. If $\mathbb{E}_{\Phi_0}^{i_0} \left| \sum_{q=1}^{\omega_1} g_i(Z_q) \right|^2 < \infty$ for every $i \in \{1, \dots, r\}$, then*

$$\frac{1}{\sqrt{n}} \sum_{q=1}^n \left[g(Z_q) - a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} g(Z_q) \right) \right] \xrightarrow{d} \mathcal{N}(0, a_{i_0}(\Phi_0) \Sigma^{i_0}(\Phi_0)) \text{ as } n \rightarrow \infty$$

where for every $i, j \in \{1, \dots, r\}$, the (i, j) -th element of the matrix $\Sigma^{i_0}(\Phi_0)$ is equal to

$$\text{Cov}_{\Phi_0}^{i_0} \left[\sum_{q=1}^{\omega_1} g_i(Z_q) - a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} g_i(Z_q) \right), \sum_{q=1}^{\omega_1} g_j(Z_q) - a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} g_j(Z_q) \right) \right].$$

We may now prove Theorem 2.

Proof of Theorem 2. We start as in the proof of Theorem 4 in Rydén [19]. Condition $K \cap \overline{\Phi_0} = \{\Phi_0\}$ ensures that $\widehat{\Phi}_n \rightarrow \Phi_0$ almost surely as $n \rightarrow \infty$. Especially, with probability 1, one has $\widehat{\Phi}_n \in K$ for n large enough by Theorem 1. For such n , a Taylor expansion of the i -th partial derivative of $\Phi \mapsto \log \mathcal{STL}_n(\Phi)$ at Φ_0 is

$$\begin{aligned} 0 = \frac{\partial \log \mathcal{STL}_n}{\partial \varphi_i}(\widehat{\Phi}_n) &= \frac{\partial \log \mathcal{STL}_n}{\partial \varphi_i}(\Phi_0) + \sum_{j=1}^{|\mathcal{E}|} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \frac{\partial^2 \log \mathcal{STL}_n}{\partial \varphi_j \partial \varphi_i}(\Phi_0) \\ &\quad + \frac{1}{2} \sum_{j,k=1}^{|\mathcal{E}|} (\widehat{\Phi}_{n,j} - \Phi_{0,j})(\widehat{\Phi}_{n,k} - \Phi_{0,k}) \frac{\partial^3 \log \mathcal{STL}_n}{\partial \varphi_k \partial \varphi_j \partial \varphi_i}(\tilde{\Phi}_n) \end{aligned}$$

where $\tilde{\Phi}_n$ is some point on the line connecting Φ_0 and $\widehat{\Phi}_n$. In other words,

$$\begin{aligned} T_{1,n} &= T_{2,n} + T_{3,n} \tag{5} \\ \text{with } T_{1,n} &= \frac{1}{\sqrt{n}} \sum_{q=1}^n h_i(Z_q, \Phi_0), \\ T_{2,n} &= - \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \left[\frac{1}{n} \sum_{q=1}^n \frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i}(Z_q, \Phi_0) \right] \\ \text{and } T_{3,n} &= - \frac{1}{2} \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \left[\sum_{k=1}^{|\mathcal{E}|} (\widehat{\Phi}_{n,k} - \Phi_{0,k}) \frac{1}{n} \sum_{q=1}^n \frac{\partial^3 \log \mathcal{L}_1}{\partial \varphi_k \partial \varphi_j \partial \varphi_i}(Z_q, \tilde{\Phi}_n) \right]. \end{aligned}$$

We start by dealing with the right-hand side of equality (5). To this end, use Lemma 4 and the fact that the process $(Z_q)_{q \geq 1}$ is ergodic to obtain

$$T_{2,n} = - \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0} \left(\frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i}(Z, \Phi_0) \right) (1 + o_{\mathbb{P}}(1)).$$

Moreover, since

$$\frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i} = \frac{1}{\mathcal{L}_1} \frac{\partial^2 \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i} - \frac{1}{\mathcal{L}_1^2} \frac{\partial \mathcal{L}_1}{\partial \varphi_i} \frac{\partial \mathcal{L}_1}{\partial \varphi_j},$$

it is a consequence of Lemma 4 and of a differentiation under the expectation sign that

$$T_{2,n} = - \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0} (h_i(Z, \Phi_0) h_j(Z, \Phi_0)) (1 + o_{\mathbb{P}}(1)).$$

Lemma 6 thus yields

$$T_{2,n} = -a_{i_0}(\Phi_0) \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) h_j(Z_q, \Phi_0) \right) (1 + o_{\mathbb{P}}(1)). \tag{6}$$

Besides, Lemma 4, the ergodicity of $(Z_q)_{q \geq 1}$ and the consistency of $\widehat{\Phi}_n$ entail

$$T_{3,n} = o_{\mathbb{P}} \left(\sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \right) \quad (7)$$

as $n \rightarrow \infty$. Collecting (6) and (7), we obtain

$$T_{2,n} + T_{3,n} = a_{i_0}(\Phi_0) \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) h_j(Z_q, \Phi_0) \right) (1 + o_{\mathbb{P}}(1)) \quad (8)$$

as $n \rightarrow \infty$. Finally, we note that thanks to Lemma 4, we may once again differentiate under the integral sign to obtain $\mathbb{E}_{\Phi_0}(h_i(Z, \Phi_0)) = 0$. Lemma 4 and the ergodicity of $(Z_q)_{q \geq 1}$ thus yield

$$\frac{1}{\sqrt{n}} T_{1,n} = \frac{1}{n} \sum_{q=1}^n h_i(Z_q, \Phi_0) \rightarrow 0 \text{ almost surely as } n \rightarrow \infty.$$

Besides, Proposition 1 entails

$$\frac{1}{\sqrt{n}} T_{1,n} = \frac{1}{n} \sum_{q=1}^n h_i(Z_q, \Phi_0) \xrightarrow{\mathbb{P}} a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) \right) \text{ as } n \rightarrow \infty$$

so that the right-hand side in this convergence must be zero; finally

$$\mathbb{E}_{\Phi_0}^{i_0} \left| \sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) \right|^2 < \infty,$$

see Lemmas 4 and 5; applying Proposition 2 then gives

$$T_{1,n} \xrightarrow{d} \mathcal{N}(0, a_{i_0}(\Phi_0) V(\Phi_0)) \text{ as } n \rightarrow \infty. \quad (9)$$

Collecting (5), (8) and (9) concludes the proof. ■

Appendix B: proofs of the preliminary results

The first result is a necessary step to obtain the strong consistency of our estimator.

Lemma 1. *For every $\varphi \in \mathcal{E}$ and $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$, one has*

$$m(rm)^{k+2} (rm)^{w_1 + \dots + w_k} \leq \mathcal{L}_1(z, \varphi) \leq M(rM)^{k+2} (rM)^{w_1 + \dots + w_k}$$

where

$$M := \max \left\{ 1, \max_{j,s} \lambda_j^{(s)}(\varphi) \right\} < \infty$$

and if $K_s := [0, \tau_s - \tau_{s-1}]$,

$$m := \min \left\{ \min_i a_i(\varphi), \min_{i,j,s} \min_{y \in K_s} f_{ij}(y, s, \Phi), \min_{i,j,s} \min_{y \in K_s} \bar{F}_{ij}(y, s, \Phi) \right\} > 0.$$

Proof of Lemma 1. Start by remarking that

$$\bar{F}_{\alpha\beta}(y, s, \varphi) = [\exp(y(L(\varphi) - \Lambda^{(s)}(\varphi)))]_{\alpha\beta} = \mathbb{P}(J'(y) = \beta, N'(y) = 0 \mid J'(0) = \alpha) \leq 1$$

when (J', N') is an MMPP with transition intensity matrix $L(\varphi)$ and jump intensity matrix $\Lambda^{(s)}(\varphi)$, see Meier-Hellstern [12]. Therefore, for every $y \geq 0$, one has $\bar{F}_{\alpha\beta}(y, s, \varphi) \leq M$ and $f_{\alpha\beta}(y, s, \varphi) \leq M$. Using (1), we immediately obtain

$$\mathcal{L}_1(z, \varphi) \leq M(rM)^{k+2}(rM)^{w_1+\dots+w_k}.$$

Besides, the compactness of K_s and the continuity of the maps involved entail

$$\min_{i,j,s} \min_{y \in K_s} f_{ij}(y, s, \Phi) > 0 \quad \text{and} \quad \min_{i,j,s} \min_{y \in K_s} \bar{F}_{ij}(y, s, \Phi) > 0$$

so that $m > 0$. Hence, using (1), the inequality

$$\mathcal{L}_1(z, \varphi) \geq m(rm)^{k+2}(rm)^{w_1+\dots+w_k}$$

which completes the proof. ■

A second pivotal tool is the following technical lemma:

Lemma 2. *For every $\Phi \in \mathcal{E}$, there exists a neighborhood G_Φ of Φ in \mathcal{E} such that*

$$\mathbb{E}_{\Phi_0} \left| \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right| < \infty.$$

Proof of Lemma 2. Let G_Φ be a neighborhood of Φ , $A = \left\{ \sup_{\varphi \in G_\Phi} \mathcal{L}_1(Z, \varphi) \leq 1 \right\}$ and write

$$\begin{aligned} \mathbb{E}_{\Phi_0} \left| \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right| &= \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_{A^c} \right] - \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_A \right] \\ &\leq \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_{A^c} \right] - \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi) \mathbb{1}_A] \end{aligned} \quad (10)$$

where A^c is the complement of A . The goal is to show that the quantity in the right-hand side of this inequality is finite for a suitable choice of the neighborhood G_Φ of Φ in \mathcal{E} . Let $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$ and G_Φ be a compact neighborhood of Φ such that $\varphi \mapsto a(\varphi)$ is continuous on G_Φ and

$$M := \sup_{\varphi \in G_\Phi} \max \left\{ 1, \max_{j,s} \lambda_j^{(s)}(\varphi) \right\} < \infty.$$

Lemma 1 entails

$$\sup_{\varphi \in G_\Phi} \mathcal{L}_1(z, \varphi) \leq M(rM)^{k+2}(rM)^{w_1+\dots+w_k}.$$

Therefore, since the logarithm function is increasing,

$$\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(z, \varphi) \leq \log M + (k+2+w_1+\dots+w_k) \log rM.$$

This inequality yields

$$\mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_{A^c} \right] \leq \log M + \left[k+2 + \sum_{s=1}^k \mathbb{E}_{\Phi_0}(W_s) \right] \log rM < \infty \quad (11)$$

since $\mathbb{E}_{\Phi_0}(W_s) < \infty$ for all s . Furthermore, Lemma 1 yields for some $m > 0$:

$$\mathcal{L}_1(z, \Phi) \geq m(rm)^{k+2}(rm)^{w_1+\dots+w_k}.$$

Especially

$$- \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi) \mathbb{1}_A] \leq |\log m| + \left[k+2 + \sum_{s=1}^k \mathbb{E}_{\Phi_0}(W_s) \right] |\log rm| < \infty. \quad (12)$$

Using together (10), (11) and (12) concludes the proof. \blacksquare

The next result shows that for every $i_0 \in \{1, \dots, r\}$, the survival function of the random time $\omega_1 = \min\{q \geq 1 \mid J(q) = i_0\}$ converges to 0 geometrically fast.

Lemma 3. *For every $i_0 \in \{1, \dots, r\}$, there exists a neighborhood G of Φ_0 and a constant $c \in (0, 1)$ such that*

$$\forall k \in \mathbb{N}, \sup_{\Phi \in G} \mathbb{P}_\Phi^{i_0}(\omega_1 > k) \leq c^k.$$

In particular, $\mathbb{E}_{\Phi_0}^{i_0}(\omega_1^k) < \infty$ for every $k \geq 1$.

Proof of Lemma 3. The result is obvious for $k = 0$. Pick $k \geq 1$ and note that

$$\begin{aligned} \mathbb{P}_{\Phi}^{i_0}(\omega_1 > k) &= \sum_{j \neq i_0} \mathbb{P}_{\Phi}^{i_0}(J(1) \neq i_0, \dots, J(k-2) \neq i_0, J(k-1) = j, J(k) \neq i_0) \\ &= \sum_{j \neq i_0} \mathbb{P}_{\Phi}^{i_0}(J(1) \neq i_0, \dots, J(k-2) \neq i_0, J(k-1) = j) \mathbb{P}_{\Phi}^j(J(1) \neq i_0). \end{aligned} \quad (13)$$

Set, for $i, j \in \{1, \dots, r\}$,

$$P_{ij}(\Phi) = \mathbb{P}_{\Phi}^i(J(1) = j) = [\exp(L(\Phi))]_{ij}.$$

In particular, the maps $\Phi \mapsto P_{ij}(\Phi)$ are continuous. Moreover, since the Markov process J is irreducible, it holds that $P_{ij}(\Phi) > 0$ for all i and j . Consequently

$$0 < c = \max_{j \neq i_0} \sup_{\Phi \in G} \mathbb{P}_{\Phi}^j(J(1) \neq i_0) < 1.$$

Using (13) entails

$$\sup_{\Phi \in G} \mathbb{P}_{\Phi}^{i_0}(\omega_1 > k) \leq c \sup_{\Phi \in G} \mathbb{P}_{\Phi}^{i_0}(\omega_1 > k-1)$$

which gives the desired result by induction on k . ■

An important part of the proof of Theorem 2 is to prove the following technical result:

Lemma 4. *There exists a neighborhood G of Φ_0 in \mathcal{E} and positive constants C, C' such that for any i, j, k :*

$$\begin{aligned} &\sup_{\varphi \in G} \max \left\{ \frac{1}{\mathcal{L}_1(Z, \varphi)}, \mathcal{L}_1(Z, \varphi), \left| \frac{\partial \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_k}(Z, \varphi) \right| \right\} \\ &\leq C \exp \left(C' \sum_{s=1}^k W_s \right). \end{aligned}$$

Epecially, there exist (possibly different) positive constants C, C' such that for any i, j, k :

$$\begin{aligned} &\sup_{\varphi \in G} \max \left\{ |\log \mathcal{L}_1(Z, \varphi)|, \left| \frac{\partial \log \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_k}(Z, \varphi) \right| \right\} \\ &\leq C \exp \left(C' \sum_{s=1}^k W_s \right), \end{aligned}$$

the right-hand side of the above inequality defines a random variable with finite \mathbb{P}_{Φ_0} -moments and for all i_1, i and j ,

$$\mathbb{E}_{\Phi_0}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi) h_j(Z, \varphi)| \right] < \infty.$$

Proof of Lemma 4. Assume that G is a neighborhood of Φ_0 such that $\varphi \mapsto a(\varphi)$ is continuous on G and $\inf_{\varphi \in G} \min_{j,s} \lambda_j^{(s)}(\varphi) > 0$. It was shown in the proof of Lemma 2 that if $K_s = [0, \tau_s - \tau_{s-1}]$ then for any $\varphi \in G$,

$$\min \left\{ \min_i a_i(\varphi), \min_{y \in K_s} \min_{i,j,s} f_{ij}(y, s, \Phi), \min_{y \in K_s} \min_{i,j,s} \bar{F}_{ij}(y, s, \Phi) \right\} > 0.$$

Consequently,

$$m := \inf_{\varphi \in G} \left\{ \min_i a_i(\varphi), \min_{y \in K_s} \min_{i,j,s} f_{ij}(y, s, \Phi), \min_{y \in K_s} \min_{i,j,s} \bar{F}_{ij}(y, s, \Phi) \right\} > 0.$$

For all $\varphi \in G$ and every $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$, Lemma 2 entails

$$\mathcal{L}_1(z, \varphi) \geq m(rm)^{k+2}(rm)^{w_1 + \dots + w_k}. \quad (14)$$

If moreover G is a neighborhood of Φ_0 with $\sup_{\varphi \in G} \max_{i,j} \ell_{ij}(\varphi) < \infty$ and $\sup_{\varphi \in G} \max_{j,s} \lambda_j^{(s)}(\varphi) < \infty$ then similarly

$$M_1 := \sup_{\varphi \in G} \left\{ \max_i a_i(\varphi), \max_{i,j,s} \max_{y \in K_s} f_{ij}(y, s, \Phi), \max_{i,j,s} \max_{y \in K_s} \bar{F}_{ij}(y, s, \Phi) \right\} < \infty.$$

Let further

$$\begin{aligned} M_2 &= \sup_{\varphi \in G} \max_{n,p,s} \max_k \max_{y \in K_s} \left\{ \left| \frac{\partial f_{np}}{\partial \varphi_k}(y, s, \varphi) \right|, \left| \frac{\partial \bar{F}_{np}}{\partial \varphi_k}(y, s, \varphi) \right| \right\}, \\ M_3 &= \sup_{\varphi \in G} \max_{n,p,s} \max_{k,l} \max_{y \in K_s} \left\{ \left| \frac{\partial^2 f_{np}}{\partial \varphi_k \partial \varphi_l}(y, s, \varphi) \right|, \left| \frac{\partial^2 \bar{F}_{np}}{\partial \varphi_k \partial \varphi_l}(y, s, \varphi) \right| \right\} \\ \text{and } M_4 &= \sup_{\varphi \in G} \max_{n,p,s} \max_{k,l,m} \max_{y \in K_s} \left\{ \left| \frac{\partial^3 f_{np}}{\partial \varphi_k \partial \varphi_l \partial \varphi_m}(y, s, \varphi) \right|, \left| \frac{\partial^3 \bar{F}_{np}}{\partial \varphi_k \partial \varphi_l \partial \varphi_m}(y, s, \varphi) \right| \right\} \end{aligned}$$

and note that $M = \max\{M_1, M_2, M_3, M_4\} < \infty$. Equation (1), inequality (14) and tedious computations may then be used to show that one may find positive constants C, C' such that

$$\begin{aligned} & \sup_{\varphi \in G} \max \left\{ \frac{1}{\mathcal{L}_1(Z, \varphi)}, \mathcal{L}_1(Z, \varphi), \left| \frac{\partial \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_k}(Z, \varphi) \right| \right\} \\ & \leq C \exp \left(C' \sum_{s=1}^k W_s \right). \end{aligned} \quad (15)$$

Besides, for every function $\varphi \mapsto G(\varphi)$ such that $\log G$ is three times continuously differentiable,

$$\begin{aligned} \frac{\partial^2 \log G}{\partial \varphi_i \partial \varphi_j} &= \frac{1}{G} \frac{\partial^2 G}{\partial \varphi_i \partial \varphi_j} - \frac{1}{G^2} \frac{\partial G}{\partial \varphi_i} \frac{\partial G}{\partial \varphi_j}, \\ \frac{\partial^3 \log G}{\partial \varphi_i \partial \varphi_j \partial \varphi_k} &= \frac{1}{G} \frac{\partial^3 G}{\partial \varphi_i \partial \varphi_j \partial \varphi_k} - \frac{1}{G^2} \frac{\partial^2 G}{\partial \varphi_i \partial \varphi_j} \frac{\partial G}{\partial \varphi_k} \\ &\quad - \frac{1}{G^2} \left[\frac{\partial^2 G}{\partial \varphi_i \partial \varphi_k} \frac{\partial G}{\partial \varphi_j} + \frac{\partial G}{\partial \varphi_i} \frac{\partial^2 G}{\partial \varphi_j \partial \varphi_k} \right] + 2 \frac{1}{G^3} \frac{\partial G}{\partial \varphi_i} \frac{\partial G}{\partial \varphi_j} \frac{\partial G}{\partial \varphi_k}. \end{aligned}$$

Inequality (15) therefore yields

$$\begin{aligned} & \sup_{\varphi \in G} \max \left\{ \left| \log \mathcal{L}_1(Z, \varphi) \right|, \left| \frac{\partial \log \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_k}(Z, \varphi) \right| \right\} \\ & \leq C \exp \left(C' \sum_{s=1}^k W_s \right), \end{aligned}$$

for (possibly different) positive constants C, C' . Finally, since for every $s \in \{1, \dots, k\}$, W_s is a Poisson distributed random variable, one has $\mathbb{E}_{\Phi_0}(x^{W_s}) < \infty$ for all s and $x > 0$, which proves that the right-hand side of the inequality above defines a random variable having finite \mathbb{P}_{Φ_0} -moments. The second part of the lemma is thus a straightforward consequence of the above inequality. \blacksquare

The next result shall be used to check a couple of integrability conditions used in the proof of Theorem 2.

Lemma 5. *Assume that ψ is a Borel measurable nonnegative function. We consider the random variable $U = \sum_{q=1}^{\omega_1} \psi(Z_q)$.*

- If $\mathbb{E}_{\Phi_0}(\psi(Z)) < \infty$ then for every $i_0 \in \{1, \dots, r\}$, $\mathbb{E}_{\Phi_0}^{i_0}(U) < \infty$.
- If $\mathbb{E}_{\Phi_0}(\psi^2(Z)) < \infty$ then for every $i_0 \in \{1, \dots, r\}$, $\mathbb{E}_{\Phi_0}^{i_0}(U^2) < \infty$.

Proof of Lemma 5. We start by remarking that for any $l \geq 1$, if $\mathbb{E}_{\Phi_0}(\psi^l(Z)) < \infty$ then for any i_1 :

$$\mathbb{E}_{\Phi_0}^{i_1} |\psi^l(Z)| \leq \frac{\mathbb{E}_{\Phi_0} |\psi^l(Z)|}{a_{i_1}(\Phi_0)} < \infty$$

since $a_{i_1}(\Phi_0) > 0$. To prove the first statement, we then write

$$\mathbb{E}_{\Phi_0}^{i_0}(U) = \sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^N \psi(Z_q) \mathbb{1}_{\{\omega_1=N\}} \right) = \sum_{q=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0}(\psi(Z_q) \mathbb{1}_{\{\omega_1 \geq q\}}).$$

Since $\{\omega_1 \geq q\} = \bigcap_{l=0}^{q-1} \{J(l) \neq i_0\}$, the hidden Markov structure of (Z_q) yields

$$\begin{aligned} \mathbb{E}_{\Phi_0}^{i_0}(U) &= \sum_{q=1}^{\infty} \sum_{i_1 \neq i_0} \mathbb{E}_{\Phi_0}^{i_1}(\psi(Z)) \mathbb{P}_{\Phi_0}^{i_0}(\omega_1 \geq q, J(q) = i_1) \\ &\leq \mathbb{E}_{\Phi_0}^{i_0}(\omega_1) \max_{i_1} \mathbb{E}_{\Phi_0}^{i_1}(\psi(Z)). \end{aligned}$$

Since the Markov chain $(J(q))$ is irreducible on $\{1, \dots, r\}$ and has stationary distribution $a(\Phi_0)$, one has $\mathbb{E}_{\Phi_0}^{i_0}(\omega_1) = 1/a_{i_0}(\Phi_0) < \infty$, from which we deduce that the right-hand side is finite.

We now turn to the second part of the lemma. Note that

$$\begin{aligned} \mathbb{E}_{\Phi_0}^{i_0}(U^2) &= \sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{p,q=1}^N \psi(Z_p)\psi(Z_q)\mathbb{1}_{\{\omega_1=N\}} \right) \\ &= \sum_{N=1}^{\infty} \sum_{q=1}^N \mathbb{E}_{\Phi_0}^{i_0}(\psi^2(Z_q)\mathbb{1}_{\{\omega_1=N\}}) + 2 \sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{\substack{p,q=1 \\ p < q}}^N \psi(Z_p)\psi(Z_q)\mathbb{1}_{\{\omega_1=N\}} \right). \end{aligned}$$

We already know from the first statement of the lemma that

$$\sum_{N=1}^{\infty} \sum_{q=1}^N \mathbb{E}_{\Phi_0}^{i_0}(\psi^2(Z_q)\mathbb{1}_{\{\omega_1=N\}}) = \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} \psi^2(Z_q) \right) < \infty.$$

Further, for all Borel nonnegative functions f_1, \dots, f_N and all i_1, \dots, i_N , the hidden Markov structure of (Z_q) entails

$$\begin{aligned} &\mathbb{E}_{\Phi_0}^{i_0} (f_1(Z_1) \cdots f_N(Z_N)\mathbb{1}_{\{J(1)=i_1, \dots, J(N)=i_N\}}) \\ &= \left[\prod_{q=1}^N \mathbb{E}_{\Phi_0}^{i_0} (f_q(Z_q) | J(q-1) = i_{q-1}, J(q) = i_q) \right] \mathbb{P}_{\Phi_0}^{i_0} \left(\bigcap_{q=1}^N \{J(q) = i_q\} \right) \\ &= \left[\prod_{q=1}^N \mathbb{E}_{\Phi_0}^{i_{q-1}} (f_q(Z_1) | J(1) = i_q) \right] \mathbb{P}_{\Phi_0}^{i_0} \left(\bigcap_{q=1}^N \{J(q) = i_q\} \right). \end{aligned}$$

Since we may write

$$\{\omega_1 = N\} = \left(\bigcup_{j_1, \dots, j_{N-1} \neq i_0} \bigcap_{q=1}^{N-1} \{J(q) = j_q\} \right) \cap \{J(N) = i_0\}$$

it is straightforward that

$$\sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{\substack{p,q=1 \\ p < q}}^N \psi(Z_p)\psi(Z_q)\mathbb{1}_{\{\omega_1=N\}} \right) \leq \left[\max_{i_1, i_2} \mathbb{E}_{\Phi_0}^{i_1}(\psi(Z) | J(1) = i_2) \right]^2 \mathbb{E}_{\Phi_0}^{i_0}(\omega_1^2)$$

which is finite since $\mathbb{E}_{\Phi_0}^{i_0}(\omega_1^2) < \infty$, see Lemma 3. ■

Lemma 6 is the key to the conclusion of the proof of Theorem 2.

Lemma 6. *Assume that ψ is a Borel measurable function such that $\mathbb{E}_{\Phi_0}|\psi(Z)| < \infty$. Then*

$$\mathbb{E}_{\Phi_0}(\psi(Z)) = a_{i_0}(\Phi_0)\mathbb{E}_{\Phi_0}^{i_0}\left(\sum_{q=1}^{\omega_1}\psi(Z_q)\right).$$

Proof of Lemma 6. The ergodicity of the process $(Z_q)_{q \geq 1}$ entails

$$\frac{1}{n}\sum_{q=1}^n\psi(Z_q) \rightarrow \mathbb{E}_{\Phi_0}(\psi(Z)) \text{ almost surely as } n \rightarrow \infty.$$

Besides, Lemma 5 yields $\mathbb{E}|\sum_{q=1}^{\omega_1}\psi(Z_q)| < \infty$, so that Proposition 1 gives

$$\frac{1}{n}\sum_{q=1}^n\psi(Z_q) \xrightarrow{\mathbb{P}} a_{i_0}(\Phi_0)\mathbb{E}_{\Phi_0}^{i_0}\left(\sum_{q=1}^{\omega_1}\psi(Z_q)\right) \text{ as } n \rightarrow \infty$$

from which the result follows. ■

	ℓ_{12}	ℓ_{21}	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\lambda_1^{(2)}$	$\lambda_2^{(2)}$
Case 1						
$n = 50$						
Mean L^1 -error	0.345	0.663	0.419	0.909	0.673	2.041
Median L^1 -error	0.259	0.449	0.312	0.612	0.497	1.637
$n = 100$						
Mean L^1 -error	0.242	0.455	0.249	0.573	0.508	1.294
Median L^1 -error	0.202	0.367	0.224	0.492	0.459	1.150
Case 2						
$n = 50$						
Mean L^1 -error	1.760	5.266	0.616	2.210	1.098	4.882
Median L^1 -error	1.352	3.830	0.454	1.499	0.766	3.772
$n = 100$						
Mean L^1 -error	1.038	3.099	0.455	1.768	0.805	3.611
Median L^1 -error	0.716	2.833	0.273	1.153	0.622	3.117
Case 3						
$n = 50$						
Mean L^1 -error	0.277	0.577	0.235	0.781	0.501	2.075
Median L^1 -error	0.194	0.451	0.177	0.513	0.438	1.699
$n = 100$						
Mean L^1 -error	0.172	0.363	0.174	0.629	0.273	1.337
Median L^1 -error	0.118	0.301	0.143	0.516	0.238	1.103

Table 1: Mean and median L^1 -errors associated to the estimators in all cases.