Hypotheses testing and posterior concentration rates for semi-Markov processes

I. Votsi1 · G. Gayraud2 · V. S. Barbu3 · N. Limnios2

Received: 13 July 2020 / Accepted: 21 June 2021 / Published online: 2 July 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
In this paper, we adopt a nonparametric Bayesian approach and investigate the asymptotic behavior of the posterior distribution in continuous-time and general state space semi-Markov processes. In particular, we obtain posterior concentration rates for semi-Markov kernels. For the purposes of this study, we construct robust statistical tests between Hellinger balls around semi-Markov kernels and present some specifications to particular cases, including discrete-time semi-Markov processes and countable state space Markov processes. The objective of this paper is to provide sufficient conditions on priors and semi-Markov kernels that enable us to establish posterior concentration rates.

Keywords Bayesian nonparametrics · Posterior concentration rates · Semi-Markov processes · Semi-Markov kernels · Robust statistical tests

1 Introduction

Semi-Markov processes (SMPs) are stochastic processes that are widely used to model real-life phenomena encountered in seismology, biology, reliability, survival analysis, wind energy, finance and other scientific fields. SMPs (Lévy 1954; Smith 1955; Takács 1954) generalize Markov processes in the sense that they allow the sojourn times in states to follow any distribution on $[0, +\infty)$, instead of the exponential distribution in the Markov case. Since non-memoryless distributions can be considered in a semi-Markov environment, duration effects can be reproduced. The duration effect firms that the time the semi-Markov system spends in a state influences its transition probabilities. Particular cases of SMPs include continuous and discrete-time Markov chains and ordinary, modified and alternating renewal processes.
The foundations of the theory of SMPs were laid by Pyke (1961a, b). Since then, further significant results were obtained by Çinlar (1969), Korolyuk and Limnios (2005) and many others. We refer the interested reader to Limnios and Oprişan (2001) for an approach to SMPs and their applications in reliability. For an overview in the theory on semi-Markov chains oriented toward applications in modeling and estimation see Barbu and Limnios (2008).

Although the statistical inference of SMPs has been extensively studied from a frequentist point of view, the Bayesian literature is rather limited. Except for some specific SMP models (Economou et al. 2014; Epifani et al. 2014), only a few papers have considered the nonparametric Bayesian theory supporting these models (Arfè et al. 2020; Bulla and Muliere 2007; Griffin and Li 2016; Johnson and Willsky 2013; Phelan 1990). Here we aim to close the aforementioned gap and follow a nonparametric Bayesian approach. The key quantity in the theory of SMPs is the semi-Markov kernel (SMK), $Q$. Our objective is to draw Bayesian inference on the Radon-Nikodym derivative of the SMK, $q$. Let us denote by $H_n$ a trajectory of the SMP of length $n$ and by $\Pi$ the prior distribution of $q$, which in all generality, could depend on $n$, and thereafter will be denoted by $\Pi_n$. Given $H_n$ and $\Pi_n$, the knowledge on $q$ is updated by the posterior distribution, that is denoted by $\Pi_n(H_n \cdot) = \Pi_n(\cdot | H_n)$. We shall stick to the last notation throughout the paper and further denote by $q_0$ the derivative of the “true” SMK, $Q_0$, which is the SMK that generated $H_n$. The main topic of the article is the study of the asymptotic behaviour of $\Pi_n(H_n \cdot q_n)$ in a neighbourhood of $Q_0$.

The first results on the asymptotic behaviour of posterior distributions in infinite-dimensional models addressed issues of posterior consistency and posterior concentration around the true distribution. In a nonparametric context, when the observations are i.i.d., such results were first derived in Ghosal et al. (2000) and Shen and Wasserman (2001) with a variety of examples. Beyond the i.i.d. setup, the asymptotic behaviour of the posterior has been studied in the context of independent nonidentically distributed observations (Amewou-Atisso et al. 2003; Arbel et al. 2013; Choudhuri et al. 2004; Ghosal and van der Vaart 2007; Ghosal and Roy 2006; Ghosal et al. 1999).

One of the most natural extensions of the i.i.d. structure is a Markov process, where only the immediate past matters. Although, given the present, the future will not further depend on the past, the dependence propagates and may reasonably capture the dependence structure of the observations. Ghosal and van der Vaart (2007) studied the asymptotic behaviour of posterior distributions to several classes of non-i.i.d. models including Markov chains. For their purposes the authors used previous results on the existence of statistical tests (Birgé 1983a; Le Cam 1986, 1975, 1973) between two Hellinger balls for a given class of models. We refer the interested reader to Birgé (2013) for improved results about the existence of such tests for the relevant estimation problems. Tang and Ghosal (2007) extended Schwartz’s theory of posterior consistency to ergodic Markov processes and applied it in the context of a Dirichlet mixture model for transition densities. More recently, Gassiat and Rousseau (2013) studied the posterior distribution in hidden Markov chains where both the observational and the state spaces are general. For nonparametric Bayesian estimation of conditional distributions, Pati et al. (2013) provided sufficient conditions on the prior under which the weak and various types of strong posterior consistency could be obtained.

For reviews on posterior consistency as well as posterior concentration in infinite dimensions, the interested reader can refer to Wasserman (1998), Ghosh and Ramamoorthi (2003) and Ghosal and van der Vaart (2017).

This paper aims to extend previous results by studying the convergence of the posterior distribution of $q$ for SMPs. Specifically, we generalize and extend previous results on
discrete-time Markov processes in countable state space (Ghosal and van der Vaart 2007) to continuous-time SMPs in general state space.

In order to apply the general theory to the semi-Markov framework, we demonstrate the existence of the relevant statistical tests. To this purpose, we extend the hypotheses testing results for Markov chains developed by Birgé (1983a) to continuous-time general state space SMPs. Such tests can also be used to distinguish Markov from semi-Markov models and decide which model could better describe the data, which is a crucial subject in real-world applications.

Very few researchers considered hypotheses testing problems in a semi-Markov context. Bath and Deshpande (1986) developed a nonparametric test for testing Markov against semi-Markov processes. Banerjee and Bhattacharyya (1976) considered a two-state SMP and proposed parametric tests for the equality of the sojourn time distributions, under the assumption that these distributions are absolutely continuous and belong to an exponential family. Also in a parametric context, Malinovskii (1992) considered the situation where the probability distribution of a SMP depends on a real-valued parameter \( \vartheta > 0 \) and studied the simple hypothesis \( H_0 : \vartheta = 0 \) against \( H_1 : \vartheta = h T^{-1/2}, 0 < h \leq c \) (the SMP is observed up to time \( T \)). Chang et al. (1999, 2001) considered hypotheses testing problems for semi-Markov counting processes, in a survival analysis context. Tsai (1985) proposed a rank test based on semi-Markov processes in order to test whether a pair of observation \((X, Y)\) has the same distribution as \((Y, X)\), i.e., \(X, Y\) are exchangeable. The aforementioned hypothesis testing results are derived for particular semi-Markov processes; to the best of our knowledge, the present paper is the first one that provides a performant and robust testing procedure for SMPs in a general nonparametric context. This procedure is “performant” w.r.t. both the first and second types of errors and “robust” since our theoretical results hold true even in case of misspecification (uniform on some balls).

We focus on SMPs since they are much more general and better adapted to applications than Markov processes. In real-world systems, the state space of the processes under study could be \(\{0, 1\}^\mathbb{N}\), (e.g., communication systems), where \(\mathbb{N}\) is the set of nonnegative integers (\(\mathbb{N}^* = \mathbb{N} \setminus \{0\}\)), or \([0, \infty)\) (e.g., fatigue crack growth modelling). This is the reason why we concentrate on general SMPs. On the other side, since in physical and biological applications time is usually considered to be continuous, discrete-time processes are not always appropriate for describing such phenomena. In such situations continuous-time processes are often more suitable than discrete-time ones. Therefore we focus our discussion on the continuous-time case rather than the discrete-time case. Nonetheless, note that our results on robust tests are very general and could also be applied to the discrete-time case, with the corresponding modifications.

The organization of the paper is as follows. In Sect. 2 the notation and preliminaries of semi-Markov processes are presented; the objectives of our paper are also presented. Section 3 describes the hypotheses testing for the processes under study and some particular cases. Section 4 discusses the derivation of the posterior concentration rate and the relative hypotheses. Finally, in Sect. 5, we give a detailed description of the proofs and some technical lemmas.
2 The semi-Markov framework and objectives

2.1 Semi-Markov processes

We consider \((E, \mathcal{E})\) a measurable space and an \((E, \mathcal{E})\)-valued semi-Markov process \(Z := (Z_t)_{t \in \mathbb{R}^+}\) defined on a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\). We denote by \((J, S) := (J_n, S_n)_{n \in \mathbb{N}}\) the Markov renewal process (MRP), with \(S_{n-1} \leq 0 \leq S_0 \leq \ldots \leq S_n \leq \ldots\) the successive \(\mathbb{R}^+\)-valued jump times of \(Z\) and \((J_n)_{n \geq 0}\) the successive visited states at these jump times (henceforth called the embedded Markov chain (EMC)). We further define the renewal process \(N(t) := \sup\{n \in \mathbb{N} : S_n \leq t\}, \ t \in \mathbb{R}^+\), and the semi-Markov process \(Z\) by \(Z_t := J_{N(t)}, \ t \in \mathbb{R}^+\).

In what follows, the EMC and MRP are considered to be homogeneous with respect to \(n \in \mathbb{N}\). It is worth noticing that the MRP \((J, S)\) satisfies the following Markov property, i.e., for any \(n \in \mathbb{N}\), any \(t \in \mathbb{R}^+\) and any \(B \in \mathcal{E}\):

\[
\mathbb{P}(J_{n+1} \in B, S_{n+1} - S_n \leq t | J_0, \ldots, J_n, S_0, \ldots, S_n) \overset{a.s.}{=} \mathbb{P}(J_{n+1} \in B, S_{n+1} - S_n \leq t | J_n).
\]

In the semi-Markov framework, of central importance is the semi-Markov kernel (SMK) defined as follows:

\[
Q_x(B, t) := \mathbb{P}(J_{n+1} \in B, S_{n+1} - S_n \leq t | J_0 = x), \ x \in E, \ t \in \mathbb{R}^+, \ B \in \mathcal{E}.
\]

Since we suppose that the distribution of \(Z\) is unknown, we focus our interest on the semi-Markov kernel. In particular the stochastic behavior of the SMP \(Z\) is determined completely by its SMK and its initial distribution.

Let us denote the \(n\)-step transition kernel of the EMC \((J_n)_{n \in \mathbb{N}}\) by

\[
P^{(n)}(x, B) := \mathbb{P}(J_n \in B | J_0 = x), \ x \in E, \ B \in \mathcal{E}, \quad (1)
\]

and the (one-step) transition kernel by \(P(x, B) = Q_x(B, \infty)\).

It is worth mentioning that

\[
Q_x(B, t) = \int_B P(x, dy) \mathbb{P}(S_{n+1} - S_n \leq t | J_n = x, J_{n+1} = y), \ t \in \mathbb{R}^+, \ B \in \mathcal{E}.
\]

In the sequel, we consider that the assumptions \textbf{A1}, \textbf{A2} and \textbf{A3} hold.

\textbf{A1} The embedded Markov chain \((J_n)_{n \in \mathbb{N}}\) is ergodic with stationary probability measure \(\rho\)

\[
\rho(B) = \int_E \rho(dy) P(y, B),
\]

with \(P\) defined in Eq. (1).

\textbf{A2} The mean sojourn times \(m(x) = \int_0^\infty \mathbb{P}(S_1 - S_0 > t | J_0 = x) dt\) satisfies

\[
\int_E \rho(dx) m(x) < \infty.
\]

\textbf{A3} \(n\)

\[
P_x(N(t) < \infty) = 1, \ t \in \mathbb{R}^+, \ x \in E.
\]
\[ \mathbb{P}(S_{n+1} - S_n \leq t | J_n = x, J_{n+1} = y) \neq 1_{\mathbb{R}^+}(t), \quad n \in \mathbb{N}, \quad t \in \mathbb{R}^+, \quad x, y \in E. \]

Note that \( A2 \) and \( A3 \) ensure that for all nonnegative \( t \) and \( B \in \mathcal{E} \), \( \mathbb{P}(Z_t \in B) \) is always well-defined, which means that the process cannot be explosive. However, the conditional probability in assumption \( A3 \) may be defined as any Dirac measure on positive real numbers. Moreover, under \( A3 \), we have that \( \ldots < S_{-1} < 0 < S_0 < \ldots < S_n < \ldots \), i.e. the sojourn time distribution is non-degenerate. Henceforth, we consider that the process starts being observed at \( S_0 \).

Denote also by \( \mathbb{B}^+ \) the Borelian \( \sigma \)-algebra on \( \mathbb{R}^+ \). We suppose that for any \( x \in E \), the SMK starting from \( x \) is absolutely continuous with respect to (w.r.t.) \( \nu \), a \( \sigma \)-finite measure \( (E \times \mathbb{R}^+, \mathcal{E} \otimes \mathbb{B}^+) \) and denote by \( q_x(\cdot, \cdot) \) its Radon-Nikodym (RN) derivative, i.e., \( Q_x(dy, dt) = q_x(y, t) d\nu(y, t) \). For \( n \geq 1 \), let \( X_n := S_n - S_{n-1} \) be the successive sojourn times of \( Z \) and set \( X_0 = S_0 \). In other words, \( X_n \) is the waiting time before jumping to the state \( J_n \). On \( \mathcal{E} \otimes \mathbb{B}^+ \), we further define the measure \( \tilde{\rho} \) as the distribution of \((J_n, X_n), n \in \mathbb{N} \), where

\[
\tilde{\rho}(A, \Gamma) = \int_E \rho(dx)Q_x(A, \Gamma), \quad A \in \mathcal{E}, \quad \Gamma \in \mathbb{B}^+. \tag{2}
\]

**Proposition 1** The measure \( \tilde{\rho} \) defined in (2) is the stationary distribution of \((J_n, X_n)_{n \in \mathbb{N}} \).

Since we are interested in obtaining asymptotic results, without loss of generality, we consider as initial distribution of the process \((J, X) := (J_n, X_n)_{n \in \mathbb{N}} \) its stationary distribution, \( \tilde{\rho} \). To avoid complicated notation, we will also use \( \tilde{\rho} \) to denote the density w.r.t. \( \nu \).

### 2.2 Objectives

Recall that we have denoted by \( Q_0 \) the true semi-Markov kernel and by \( q_0 \) its RN derivative w.r.t. \( \nu \), cf. Sect. 2.1. We suppose that \( q_0 \) belongs to \( \mathcal{Q} \), some set of semi-Markov kernel densities, i.e.,

\[ Q = \{ q = q_x(y, t) : x, y \in E, t \in \mathbb{R}^+ \}, \]

which is equipped with a metric \( d \) that will be defined in the sequel. Define also the \( \epsilon \)-neighborhoods around \( q_0 \) in \( \mathcal{Q} \) w.r.t. \( d \), that is

\[ B_d(q_0, \epsilon) = \left\{ q \in \mathcal{Q} : d(q_0, q) \leq \epsilon \right\}. \]

To allow some flexibility, it is quite common to deal with \( Q_n \), a subset of \( \mathcal{Q} \), that may depend on \( n \), such that the prior distribution \( \Pi_n \) on \( \mathcal{Q} \) assigns most of its mass on \( Q_n \) (see assumption \( A7 \) hereafter); an \( \epsilon \)-neighborhood around \( q_0 \) in \( Q_n \) w.r.t. \( d \) will be denoted by \( B_{d,n}(q_0, \epsilon) \).

As noted by Birgé (1983a) in the setting of Markov chains, there exists a priori no “natural” distance \( d \) between two semi-Markov kernel densities; however, it is possible to define one in two stages: first by starting with a natural distance between any two conditional probability distributions \( Q_x;1 \) and \( Q_x;2 \) dominated by \( \nu \) and corresponding to the same initial state \( J_0 = x \in E \) and second by integrating over \( x \). Indeed, if we further denote by \( q_{x;1} \) and \( q_{x;2} \) the respective RN derivatives of \( Q_x;1 \) and \( Q_x;2 \), we first consider the squared Hellinger distance between \( Q_x;1 \) and \( Q_x;2 \), i.e.,

\[
h^2_{\nu}(Q_{x;1}, Q_{x;2}) = \frac{1}{2} \int_{E \times \mathbb{R}^+} \left( \sqrt{q_{x;1}(y, t)} - \sqrt{q_{x;2}(y, t)} \right)^2 d\nu(y, t), \tag{3}
\]
and second, given a measure on \((\mathcal{E}, \mathcal{E})\), say \(\mu\), we define a semi-distance \(d_\mu\) between \(q_1\) and \(q_2\),

\[
d_\mu^2(q_1, q_2) = \int_E h_\nu^2(Q_{x;1}, Q_{x;2}) d\mu(x). \tag{4}
\]

Given a sample path of the SMP for a given number of jumps \(n \in \mathbb{N}^*\),

\[
\mathcal{H}_n = \{J_0, J_1, \ldots, J_n, S_0, S_1, \ldots, S_n\},
\]

we adopt a Bayesian point of view by considering a prior distribution \(\Pi_n\) on \(Q\). We aim to establish how fast the posterior distribution shrinks, in terms of \(d_\mu\), around the “true” semi-Markov kernel density, \(q_0\); the specification of \(\mu\) will be given after the statement of assumption \(A4\). More precisely, our objective is to find the minimal positive sequence \(\epsilon_n\) tending to zero as \(n\) goes to infinity, such that under some assumptions on both \(Q\) and \(\Pi_n\)

\[
\Pi_n^{\mathcal{H}_n} \left( B_{d_\mu}^\mathcal{E}(q_0, \epsilon_n) \right) \xrightarrow{L_1(\mathbb{P}_0^{(n)})} 0 \text{ as } n \to \infty,
\]

where \(B_{d_\mu}^\mathcal{E}\) denotes the complementary of \(B_{d_\mu}\) in \(Q\) and \(\mathbb{P}_0^{(n)}\) refers to the “true” distribution of \(\mathcal{H}_n\).

Let us denote by \(\mathbb{E}_n^{(n)}\) the distribution of \(\mathcal{H}_n\), when the density of the SMK is \(q\). We further denote by \(\mathbb{E}_q^{(n)}\) the expectation and by \(V_q^{(n)}\) the variance w.r.t. \(\mathbb{P}_q^{(n)}\), respectively. Every quantity (distribution, SMK, expectation, variance, ...) with an index 0 refers to the corresponding “true” quantity.

### 3 Hypotheses testing for semi-Markov processes

#### 3.1 Robust tests

One of the key ingredients needed to obtain posterior concentration rates is the construction of corresponding robust hypotheses tests. For a variety of models, depending on the semi-metric \(d\), some tests with exponential power do exist. For instance, in the case of density or conditional density estimation, Hellinger or \(L_1\) tests have been introduced in Birgé (1983b). Other examples of tests could be found in Ghosal and van der Vaart (2007) and in Rousseau et al. (2012). However, in a semi-Markov context, such tests are not explicitly defined in previous works. Therefore it is of paramount importance to build test procedures with exponentially small errors in the semi-Markov context. Thus in the sequel we will be interested in the following testing procedure

\[
H_0 : q_0 \text{ against } H_1 : q \in B_{d_{q_0,n}}(q_1, \xi \epsilon), \text{ with } d_{q_0}(q_0, q_1) \geq \epsilon, \tag{5}
\]

for some \(\xi \in (0, 1)\) and two measures \(\eta^*, \nu^*\) on \((E, \mathcal{E})\).

In order to solve the testing problem (5) with exponentially small error probabilities, two more assumptions are required.

- **A4**: There exist two measures \(\nu^*\) and \(\eta^*\) on \((E, \mathcal{E})\) and two positive integers \(k, l\) such that for any \(x \in E\),

\[
\frac{1}{k} \sum_{u=1}^{k} P^{(u)}(x, \cdot) \geq \nu^*(\cdot) \quad \text{and} \quad P^{(l)}(x, \cdot) \leq \eta^*(\cdot),
\]

\(\odot\) Springer
where \( P(\cdot) \) is defined in (1). Note that \( A_4 \) implies the following inequalities which serve to prove Proposition 2:

\[
\forall m \in \mathbb{N}, \quad \frac{1}{k} \sum_{u=1}^{k} P^{(u+m)}(x, \cdot) \geq v^*(\cdot) \quad \text{and} \quad P^{(l+m)}(x, \cdot) \leq \eta^*(\cdot).
\]

Moreover, if for all \( k \in \mathbb{N}^* \) and any \( x \in E \), \( v^*(\cdot) \leq P^{(k)}(x, \cdot) \leq \eta^*(\cdot) \), then assumption \( A_4 \) holds with \( k = 1 \) and \( l = 1 \); as a consequence, \( d_{v^*}(\tilde{q}, \tilde{q}) \leq d_{\eta^*}(\tilde{q}, \tilde{q}) \) for any two pairs \( \tilde{q} \) and \( \tilde{q} \) in \( Q \), and entails that hypotheses \( H_0 \) and \( H_1 \) in the testing problem (5) are disjoint.

Let us denote by \( \mathcal{N}(\epsilon, \mathcal{C}, d) \) the covering number associated to the coverage of the set \( \mathcal{C} \) with \( d \)-balls of radius \( \epsilon \).

- \( A_5 \) For \( \xi \) in \( ]0, 1[ \) and a sequence \( \epsilon_n \) of positive numbers

\[
\sup_{\epsilon > \epsilon_n} \log \mathcal{N} \left( \frac{\epsilon}{2}, B_{d_{v^*}n}(q_0, 2\epsilon), d_{\eta^*} \right) \leq n\epsilon_n^2.
\]

**Proposition 2** Under assumption \( A_4 \), for any \( n \in \mathbb{N}^* \), there exist \( \xi \in ]0, 1[, \ K > 0 \) and \( \tilde{K} > 0 \), universal constants, such that for any \( \epsilon > 0 \) and any \( q_1 \in Q_n \) with \( d_{v^*}(q_1, q_0) > \epsilon \), there exists a statistical test \( \psi_1(\mathcal{H}_n) \) satisfying

\[
\mathbb{E}_0^{(n)}[\psi_1(\mathcal{H}_n)] \leq e^{-K n \epsilon^2} \quad \text{and} \quad \sup_{q \in Q_n : d_{v^*}(q_1, q) < \xi} \mathbb{E}_q^{(n)}[1 - \psi_1(\mathcal{H}_n)] \leq e^{-\tilde{K} n \epsilon^2}.
\]

The next corollary generalizes Proposition 2 to any \( q_1 \in Q_n \), s.t. \( d_{v^*}(q_1, q_0) > \epsilon \); it requires assumption \( A_5 \) to control the complexity of \( \tilde{Q}_n \subseteq Q_n \), which is expressed in terms of an upper bound for the covering number of \( B_{d_{v^*}n}(q_0, \epsilon) \) w.r.t. \( d_{\eta^*} \).

Note that the case where the null hypothesis is composite could also be considered; the first type error in (6) would be written similarly to the second type error, with straightforward modifications.

**Corollary 1** Under assumptions \( A_4 \) and \( A_5 \), there exist \( M > 0 \) large enough, two universal positive constants \( K \) and \( \tilde{K} \) and a statistical test \( \psi(\mathcal{H}_n) \) for the hypotheses problem (5), satisfying

\[
\mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] \leq e^{-K n \epsilon_n^2 M^2} \quad \text{and} \quad \sup_{q \in B_{d_{v^*}n}(q_0, M \epsilon_n)} \mathbb{E}_q^{(n)}[1 - \psi(\mathcal{H}_n)] \leq e^{-\tilde{K} n \epsilon_n^2 M^2},
\]

where \( B_{d_{v^*}n}(q_0, M \epsilon_n) = \{q \in Q_n : d_{v^*}(q_0, q) > M \epsilon_n\} \).

### 3.2 Particular cases and examples

In this paper the results are rather general in the sense that they refer to continuous-time and general state space SMPs. In the sequel, we focus on some particular cases that could be of special interest either from an applicative point of view, or as a starting point for further research. We further present some examples.

First, note that the state space is considered to be finite in most of the applicative articles. Second, we would like to stress out that in some applications the state space is intrinsically continuous, due to the fact that the scale of the measures is continuous.
3.2.1 Discrete-time SMPs

In the case of discrete-time SMPs, the assumptions A1, A2 and A3 need to hold in their current form.

- **General state space**
  For any \( k \in \mathbb{N} \) and any \( B \in \mathcal{E} \), the semi-Markov kernel is defined by
  \[
  Q_x(B, k) = \mathbb{P}(J_{n+1} = B, X_{n+1} \leq k | J_n = x), \quad x \in \mathcal{E}.
  \]
  In this case the natural measure \( \nu \) is the product measure \( \nu = \nu_1 \otimes \nu_2 \), where \( \nu_1 \) is a \( \sigma \)-finite measure on \((\mathcal{E}, \mathcal{E})\) and \( \nu_2 \) is the counting measure on \( \mathbb{N} \). Let us denote by \( q_x \) the R.N. derivative of \( Q_x \) w.r.t. \( \nu \). Thus, in this framework, the squared Hellinger distance becomes
  \[
  h^2_{\nu}(Q_x; 1, Q_x; 2) = \frac{1}{2} \sum_{k \in \mathbb{N}} \int_{\mathcal{E}} \left( \sqrt{q_{x; 1}(y, k)} - \sqrt{q_{x; 2}(y, k)} \right)^2 d\nu_1(y),
  \]
  while the semi-distance \( d_\mu \), which is either \( d_{\nu^*} \) or \( d_{\eta^*} \), between \( q_1 \) and \( q_2 \) is given in Eq. (4).

- **Countable state space**
  First we define the natural measure \( \nu = \nu_1 \otimes \nu_2 \), where \( \nu_1 \) and \( \nu_2 \) are the counting measures on \( \mathcal{E} \) and \( \mathbb{N} \), respectively. Second for any \( k \in \mathbb{N} \) and any \( x, y \in \mathcal{E} \), we define the semi-Markov kernel
  \[
  Q_x(y, k) = \mathbb{P}(J_{n+1} = y, X_{n+1} \leq k | J_n = x)
  \]
  and its R.N. derivative w.r.t. \( \nu \) by
  \[
  q_x(y, k) = \mathbb{P}(J_{n+1} = y, X_{n+1} = k | J_n = x). \tag{7}
  \]
  Hence
  \[
  h^2_{\nu}(Q_x; 1, Q_x; 2) = \frac{1}{2} \sum_{k \in \mathbb{N}} \sum_{y \in \mathcal{E}} \left( \sqrt{q_{x; 1}(y, k)} - \sqrt{q_{x; 2}(y, k)} \right)^2, \tag{8}
  \]
  while
  \[
  d_\mu^2(q_1, q_2) = \sum_{x \in \mathcal{E}} h^2_{\nu}(Q_x; 1, Q_x; 2) \mu(\{x\}). \tag{9}
  \]

3.2.2 Continuous-time SMPs

In this context, the assumptions A1, A2 and A3 need to be valid.

- **Countable state space**
  For \( x \in \mathcal{E} \), the semi-Markov kernel is given by
  \[
  Q_x(y, t) = \mathbb{P}(J_{n+1} = y, X_{n+1} \leq t | J_n = x), \quad y \in \mathcal{E}, \quad t \in \mathbb{R}^+
  \]
  and its R.N. derivative w.r.t. \( \nu \) by
  \[
  q_x(y, t) = \frac{\partial Q_x(y, t)}{\partial t}. \tag{11}
  \]
In this context, the squared Hellinger distance becomes

$$h_{v_1}^2(Q_{x;1}, Q_{x;2}) = \frac{1}{2} \sum_{y \in E} \int_{\mathbb{R}^+} \left( \sqrt{q_{x;1}(y,t)} - \sqrt{q_{x;2}(y,t)} \right)^2 dv_1(t), \quad (12)$$

where $v_1$ is the marginal on $(\mathbb{R}^+, \mathbb{B}^+)$ of the measure $\nu$ defined on $(E \times \mathbb{R}^+, \mathcal{E} \otimes \mathbb{B}^+)$ and the semi-distance $d_{\mu}$ between $q_1$ and $q_2$ is defined as in Eq. (9).

**Example 1** An alternating renewal process, with up time distribution function $F$, and down time distribution function $G$, is a particular case of a semi-Markov process, with two states 1 (for up) and 0 (for down) and semi-Markov matrix $Q$ given by

$$Q(t) = \begin{pmatrix} 0 & F(t) \\ G(t) & 0 \end{pmatrix}, \quad t \in \mathbb{R}^+.$$ 

The up and down times are considered to have non-degenerate distributions and finite means, which means that the assumptions A2 and A3 hold. It is worth noticing that this process is not a Markov process, except in the very particular case where both distributions $F$ and $G$ are exponential. For the alternating renewal process we have, for any integer $m > 0$,

$$p^{(2m)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$p^{(2m-1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$ 

The reference measure $\nu$ used in Eq. (3) is defined as the product measure of $\nu_1$, the Lebesgue measure on $(\mathbb{R}^+, \mathbb{B}^+)$ and the counting measure on $(E = \{0, 1\}, \mathcal{P}(E))$. Assumption A4 is satisfied with suitable choices of positive integers $k$, $l$ and measures $\nu^*$, $\eta^*$, as for example, $k = 2, l = 1$, $\nu^* = (1/2, 1/2)$ and $\eta^* = (1, 1)$.

In this setting the set of semi-Markov kernel densities is given by

$$Q = \{q = q_x(y,t), x, y \in \{0, 1\}, t \in \mathbb{R}^+ : q_1(1, t) = q_0(0, t) = 0, q_1(0, t) = f(t), q_0(1, t) = g(t)\},$$

where $f$ and $g$ are the RN derivatives of $F$ and $G$ w.r.t. the measure $\nu_1$. Note also that for $\mu$, which is either $\nu^*$ or $\eta^*$, we have

$$d_{\mu}^2(\tilde{q}, \tilde{q}) = \frac{1}{2} \left( \int_{\mathbb{R}^+} \left( \sqrt{\tilde{f}(t)} - \sqrt{\tilde{f}(t)} \right)^2 dv_1(t) \right) \mu(\{1\})$$

$$+ \frac{1}{2} \left( \int_{\mathbb{R}^+} \left( \sqrt{\tilde{g}(t)} - \sqrt{\tilde{g}(t)} \right)^2 dv_1(t) \right) \mu(\{0\}),$$

where $\tilde{f}$ and $\tilde{g}$ are associated to $\tilde{q}$ while $\tilde{f}$ and $\tilde{g}$ are related to $\tilde{q}$.

**Example 2** The next example concerns a two-component series repairable system which is described by a three-state semi-Markov model. The continuous-time semi-Markov model is defined in a countable state space $E = \{1, 2, 3\}$, where states 1 and 2 indicate the failure of the system due to the failure of the first or the second component, respectively. On the other
side, state 3 describes the functioning of both components and therefore the functioning of the system. The semi-Markov matrix \( Q \) is given by

\[
Q(t) = \begin{pmatrix}
0 & 0 & Q_1([3], t) \\
0 & 0 & Q_2([3], t) \\
Q_3([1], t) & Q_3([2], t) & 0
\end{pmatrix}, \quad t \in \mathbb{R}^+.
\]

We further consider that the times to failure, i.e. the transition times from state 3 to state 1 (resp. 2) are exponentially distributed with corresponding parameter \( \lambda_1 \) (resp. \( \lambda_2 \)). On the other side, the times to repair for the component 1 (resp. 2) follow any arbitrary distribution defined by the cdf \( F_1(t) \) (resp. \( F_2(t) \)). The distributions \( F_1(t) \), \( F_2(t) \) are non-degenerate (A3), the corresponding times have finite means (A2) and assumption A1 holds. As for the previous example, this process is not a Markov process, except in the particular case where both distributions \( F_1 \) and \( F_2 \) are exponential.

In this case, the semi-Markov matrix becomes

\[
Q(t) = \begin{pmatrix}
\frac{\lambda_1}{\lambda_1 + \lambda_2}(1 - e^{-(\lambda_1 + \lambda_2)t}) & \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - e^{-(\lambda_1 + \lambda_2)t}) & F_1(t) \\
0 & 0 & F_2(t)
\end{pmatrix}, \quad t \in \mathbb{R}^+,
\]

and the transition probability matrix of the EMC satisfies for any positive \( m \),

\[
P^{(2m-1)} = \begin{pmatrix}
0 & 0 & 1 \\
0 & 0 & 1 \\
\frac{\lambda_1}{\lambda_1 + \lambda_2} & \frac{\lambda_2}{\lambda_1 + \lambda_2} & 0
\end{pmatrix}
\]

\[
P^{(2m)} = \begin{pmatrix}
0 & 0 & 1 \\
0 & 0 & 1 \\
\frac{\lambda_1}{\lambda_1 + \lambda_2} & \frac{\lambda_2}{\lambda_1 + \lambda_2} & 0
\end{pmatrix}.
\]

The measure \( \nu \) is naturally defined as the product measure of \( \nu_1 \), the Lebesgue measure on \((\mathbb{R}^+, \mathbb{B}^+)\) and the counting measure on \((E = \{1, 2, 3\}, \mathcal{P}(E))\). The assumption A4 is satisfied with suitable choices of positive integers \( k, l \) and measures \( \nu^*, \eta^* \), as for example, \( k = 2, l = 1 \), \( \nu^* = \left( \frac{\lambda_1}{2(\lambda_1 + \lambda_2)}, \frac{\lambda_2}{2(\lambda_1 + \lambda_2)}, \frac{1}{2} \right) \) and \( \eta^* = 2\nu^* \).

Note that in this case, the set of semi-Markov kernel densities under interest, is defined as follows

\[
Q = \{ q = q_\xi(3, t), x \in \{1, 2, 3\}, t \in \mathbb{R}^+ : q_1(3, t) = f_1(t), q_2(3, t) = f_2(t), q_3(3, t) = 0 \},
\]

where \( f_1 \) and \( f_2 \) are the RN derivatives of \( F_1 \) and \( F_2 \) w.r.t. the measure \( \nu_1 \). Moreover, for \( \mu \), which is either \( \nu^* \) or \( \eta^* \) and any \( \tilde{q} \) and \( \tilde{g} \) in \( Q \), we have

\[
d^2_\mu(\tilde{q}, \tilde{g}) = \frac{1}{2} \left( \int_{\mathbb{R}^+} (\sqrt{\tilde{f}_1(t)} - \sqrt{\tilde{f}_1(t)})^2 d\nu_1(t) \right) \mu([1])
\]

\[
+ \frac{1}{2} \left( \int_{\mathbb{R}^+} (\sqrt{\tilde{g}_1(t)} - \sqrt{\tilde{g}_1(t)})^2 d\nu_1(t) \right) \mu([2]),
\]

where \( \tilde{f}_1 \) and \( \tilde{g}_1 \) concern \( \tilde{q} \) whereas \( \tilde{f}_1 \) and \( \tilde{g}_1 \) are referred to \( \tilde{g} \).
3.3 Specification to the Markov case

Note that the previously obtained results on robust tests for SMPs could be adapted to the particular case of Markov processes. These tests are of great interest and could be used for real-life applications. In particular, they enable us to decide if an observed dataset would be better described by a Markov (null hypothesis) or a semi-Markov process (alternative hypothesis). More precisely suppose we are interested in the following testing problem:

\[ \tilde{H}_0 : Q_0 \text{ Markov kernel vs} \]
\[ \tilde{H}_1 : Q_1 \text{ semi-Markov kernel } \epsilon_n - \text{ distant from } Q_0 \text{ w.r.t. } d_{\nu^*}. \]

Note that \( \tilde{H}_1 \) is simple but could be extended to any \( \xi \epsilon - \)ball around \( q_1 \) w.r.t. \( d_{\eta^*} \) and with \( \xi \in [0, 1[. \)

In the sequel of this section, we explain how the hypothesis testing problem \( \tilde{H}_0 \) versus \( \tilde{H}_1 \) could be directly solved by means of the testing problem (5) with a simple alternative hypothesis.

Since \( \tilde{H}_1 \) is simple, we could notice that assumption A4 reduces to assumption A4*, whereas A5 is no more required to exponentially bound the error probabilities. In this case, both error probabilities are bounded by the same quantity \( e^{-Kn_{\epsilon^*}^2}; \)

- **A4**: There exist a measure \( \nu^* \) on \( E \) and a positive integer \( k \) such that for any \( x \in E, \)

\[
\frac{1}{k} \sum_{u=1}^{k} P(u)(x, \cdot) \geq \nu^*(\cdot).
\]

In what follows, and by construction, the only assumption that needs to hold is assumption A1.

First, we present the discrete-time and countable state space case, and consider a Markov process with transition matrix \( \tilde{p} = (\tilde{p}_{xy})_{x,y \in E}, (\tilde{p}_{xx} \neq 1 \text{ for all states } x \in E). \)

Note that a Markov process could be represented as a semi-Markov process with semi-Markov kernel given in (7) and expressed as

\[
q_{x;0}(y, k) = \begin{cases} 
\tilde{p}_{xy} (\tilde{p}_{xx})^{k-1}, & \text{if } x \neq y \text{ and } k \in \mathbb{N}^*; \\
0, & \text{otherwise.}
\end{cases}
\]

Consequently, we can define the corresponding squared Hellinger distance as in (8) and the \( d_{\nu^*}^2 \)-distance as in (9).

Second, for the continuous-time and countable state space case, consider a regular jump Markov process with continuous transition semigroup \( \tilde{P} = (\tilde{P}(t))_{t \in \mathbb{R}^+} \) and infinitesimal generator matrix \( A = (a_{xy})_{x,y \in E}. \)

In this context, we can represent the Markov process as a semi-Markov process with semi-Markov kernel given in (10) and expressed as

\[
Q_{x;0}(y, t) = \begin{cases} 
\frac{a_{xy}}{a_x} (1 - e^{-a_x t}), & \text{if } x \neq y \text{ and } t \in \mathbb{R}^+; \\
0, & \text{otherwise,}
\end{cases}
\]

where \( a_x := -a_{xx} < \infty, x \in E. \)

Finally, the \( q \)'s, the Hellinger distance between the \( Q \)'s and the semi-distance between the \( Q \)'s are defined by (11), (12) and (9) respectively.
4 Posterior concentration rates for semi-Markov kernels

In this part, we present the key assumptions and state our main result.

First note that the likelihood function of the sample path $\mathcal{H}_n$ evaluated at $q \in Q$ is given by

$$\mathcal{L}_n(q) = \tilde{\rho}(J_0, X_0) \prod_{\ell=1}^n q_{j_{\ell-1}}(J_\ell, X_\ell).$$

Let us introduce the tools that play a central role in asymptotic Bayesian nonparametrics: the Kullback-Leibler (KL) divergence between $p^{(n)}_0$ and $p^{(n)}_q$, $q \in Q$ (see Ghosal et al. 2000; Shen and Wasserman 2001), and its associated second centered moment, that is

$$K(p^{(n)}_0, p^{(n)}_q) := E^{(n)}_0 \left[ \log \frac{\tilde{\rho}_0(J_0, S_0)}{\tilde{\rho}_q(J_0, S_0)} \prod_{l=1}^n \frac{q_{j_{l-1};0}(J_l, X_l)}{q_{j_{l-1};q}(J_l, X_l)} \right],$$

$$V_0(p^{(n)}_0, p^{(n)}_q) := V^{(n)}_0 \left[ \log \frac{\tilde{\rho}_0(J_0, S_0)}{\tilde{\rho}_q(J_0, S_0)} \prod_{l=1}^n \frac{q_{j_{l-1};0}(J_l, X_l)}{q_{j_{l-1};q}(J_l, X_l)} \right],$$

where $E^{(n)}_0$ and $V^{(n)}_0$ denote respectively the expectation and the variance w.r.t. $p^{(n)}_0$.

Then, consider the subspace $\mathcal{U}(q_0, \epsilon)$ of $Q$, which represents the following Kullback-Leibler $\epsilon$-neighborhood of $p^{(n)}_0$, that is, for positive $\epsilon$,

$$\mathcal{U}(q_0, \epsilon) = \left\{ q \in Q : K(p^{(n)}_0, p^{(n)}_q) \leq \frac{n \epsilon^2}{2}, V_0(p^{(n)}_0, p^{(n)}_q) \leq \frac{n \epsilon^2}{2} \right\}. \quad (13)$$

It is worth mentioning that although $\tilde{\rho}$ is not of primary interest, since it is unknown it should require a prior. But since any prior on $\tilde{\rho}$ that is independent of the prior on $q$ would disappear upon marginalization of the posterior of $\tilde{\rho}$, in the sequel it will be dropped. Thus, it suffices to consider only a prior distribution on $q$.

Let us now state the main result. We recall that $\Pi_n$ denotes a prior distribution on $Q$.

**Theorem 1** Assume that $A4$ holds and suppose that for a sequence of positive numbers $\epsilon_n$ such $\lim_{n \to +\infty} \epsilon_n = 0$, $\lim_{n \to +\infty} n \epsilon_n^2 = \infty$, $A5$ and $A6$-$A7$ defined hereafter, hold.

- **A6** $\exists c > 0$, $\Pi_n(\mathcal{U}(q_0, \epsilon_n)) > e^{-cn \epsilon_n^2}$,
- **A7** $Q_n \subset Q$ is such that $\Pi_n(Q_n^c) \leq e^{-2n(c+1)\epsilon_n^2}$.

Then for $M$ large enough,

$$\Pi_n^\mathcal{H}_n\left(B_{d_{\ell^p}}(q_0, \epsilon_n M) \right) \xrightarrow{L_1(p^{(n)}_0)} 0, \quad \text{as } n \to \infty. \quad (14)$$

Some comments on the result of Theorem 1 as well as the hypotheses we deal with:

- Under $A4$, Theorem 1 guarantees that, for both a particular set of semi-Markov kernels $Q$ containing some subset $Q_n$ such that $A5$ holds for a sequence of positive numbers $\epsilon_n$ and a prior distribution $\Pi_n$ on $Q$ satisfying assumptions $A6$-$A7$ with $\epsilon_n$, the posterior distribution shrinks towards $q_0 \in Q$ at a rate proportional to $\epsilon_n$.
- Assumption $A6$ is classical to derive posterior consistency and posterior concentration rates in Bayesian Nonparametrics (see, for e.g., Shen and Wasserman (2001) and Ghosal et al. (2000)); it states that the prior distribution puts enough mass around KL neighborhoods of $q_0$. 

Springer
• As mentioned in Sect. 3.1, $Q_n$ has to be almost the support of $\Pi_{1n}$: this is guaranteed by assumption A7, which in addition quantifies how $\Pi_{1n}$ assigns mass to $Q_n$. If A5 holds with $B_{d_{\nu}}(q_0, 2\epsilon)$ instead of $B_{d_{\nu,n}}(q_0, 2\epsilon)$, then $Q_n$ coincides with $Q$ and assumption A7 is no more needed.

• Although our semi-Markov framework differs from the Markov one, it is worth noticing that assumption A4 is similar to the one stated as Eq. (4.1) in Ghosal and van der Vaart (2007). In particular, for Markov chains, this assumption is related to the transition probabilities of the Markov chain, whereas in our context, A4 is concerned with the SMK density.

5 Proofs

5.1 Proof of Proposition 1

Proof In order to prove Proposition 1, we prove that the right-hand side of Eq. (2) satisfies the two relevant conditions. First, for any $A \in \mathcal{E}$, any $\Gamma_1 \in \mathbb{B}^+$, we have

$$\tilde{\rho} Q(A, \Gamma) := \int_{E \times \mathbb{R}^+} \tilde{\rho}(dy, ds) Q_y(A, \Gamma)$$

$$= \int_{E \times E \times \mathbb{R}^+} \rho(dx) Q_x(dy, ds) Q_y(A, \Gamma)$$

$$= \int_{E} \rho(dy) Q_y(A, \Gamma)$$

$$= \tilde{\rho}(A, \Gamma).$$

Second,

$$\tilde{\rho}(E, \mathbb{R}^+) = \int_{E} \rho(dx) Q_x(E, \mathbb{R}^+) = 1.$$

5.2 Proof of Proposition 2

Proof Our proof is constructive; indeed, we are going to construct a suitable testing procedure, namely $\psi_1(\mathcal{H}_n)$, for the hypotheses testing problem given in (5), i.e.,

$$H_0: q_0 \ \text{against} \ H_1: q \in B_{d_{\nu,n}}(q_1, \xi \epsilon), \ \text{with} \ d_{\nu,n}(q_0, q_1) \geq \epsilon, \ \text{and some} \ \xi \in (0, 1).$$

To control exponentially both the type I and type II errors of $\psi_1(\mathcal{H}_n)$, we first fix some $x \in E$ for which we construct the “least favorable” pair of RN derivatives of semi-Markov kernels associated to the following auxiliary testing problem

$$\tilde{H}_0, x: q_{x;0}(\cdot, \cdot) \ \text{against} \ \tilde{H}_1, x: \{ q_{x}(\cdot, \cdot) : h^2_{\nu}(Q_x, Q_x;1) \leq 1 - \cos(\lambda \alpha_x) \},$$

where $\lambda$ is any value in $]0, 1/4[$ and $\alpha_x$ belonging to $]0, \pi/2[$ is such that

$$h^2_{\nu}(Q_x;0, Q_x;1) = 1 - \cos(\alpha_x).$$

Based on properties of this least favorable pair of $q_x$’s (some useful inequalities), we will derive the construction of $\psi_1(\mathcal{H}_n)$ for the testing problem (5) and then, study its performance.
For the sake of simplicity, let us denote by $q_x$ and $q_{x:j}$ for $j \in \mathbb{N}$ the probability density functions $q_x(\cdot, \cdot)$ and $q_{x:j}(\cdot, \cdot)$, respectively.

\section{Least favorable pair $(q_{x:0}, q_{x:2})$ for the testing problem (15)}

For our purposes, we adapt the construction of Birgé (1983a) for Markov chains to the semi-Markov framework. Whatever is $x$ in $E$, we attach to $x$ the particular probability density function $q_{x:2} \in \mathcal{H}_{1,x}$ defined by

$$q_{x:2} = \left(\frac{\sin((1 - \lambda)\alpha_x)}{\sin(\alpha_x)} q_{x:1} + \frac{\sin(\lambda\alpha_x)}{\sin(\alpha_x)} q_{x:0}\right)^2.$$

Since $q_{x:2}$ is a probability density w.r.t. the measure $\nu$, using

$$h_\nu^2(Q_{x:0}, Q_{x:1}) = 1 - \int_{E \times \mathbb{R}^+} \sqrt{q_{x:0}(y, t) q_{x:1}(y, t)} d\nu(y, t) = 1 - \cos(\alpha_x) = 2 \sin^2\left(\frac{\alpha_x}{2}\right),$$

and due to trigonometric calculations, we get

$$h_\nu^2(Q_{x:1}, Q_{x:2}) = 1 - \int_{E \times \mathbb{R}^+} \sqrt{q_{x:2}(q_{x:1})} d\nu = 1 - \cos((1 - \lambda)\alpha_x) = 2 \sin^2\left(\frac{(1 - \lambda)\alpha_x}{2}\right).$$

In a similar way, replacing $\lambda$ by $1 - \lambda$ in Eq. (17), one obtains

$$h_\nu^2(Q_{x:0}, Q_{x:2}) = 1 - \int_{E \times \mathbb{R}^+} \sqrt{q_{x:2}(q_{x:0})} d\nu = 1 - \cos(1 - \lambda)\alpha_x = 2 \sin^2\left(\frac{1 - \lambda}{2}\alpha_x\right).$$

At a next step, we aim to bound $h_\nu^2(Q_{x:1}, Q_{x:2})$ and $h_\nu^2(Q_{x:0}, Q_{x:2})$ with bounds that depend on $h_\nu^2(Q_{x:0}, Q_{x:1})$; to do this, we will use the following inequality:

$$\forall \lambda \in ]0, 1[, \quad \forall \alpha_x \in \left[0, \frac{\pi}{2}\right], \quad \frac{\sin(\lambda\alpha)}{\lambda \sin(\alpha)} > 1. \quad (19)$$

From Eq. (17) and (19), we obtain the right-hand side of (20) while the left-hand side immediately follows for $\lambda \in ]0, 1[$ and $\alpha_x \in \left[0, \frac{\pi}{2}\right]$:

$$2\lambda^2 \sin^2\left(\frac{\alpha_x}{2}\right) = \lambda^2 h_\nu^2(Q_{x:0}, Q_{x:1}) < h_\nu^2(Q_{x:1}, Q_{x:2}) < h_\nu^2(Q_{x:0}, Q_{x:1}) \quad (20)$$

In a similar way, replacing $\lambda$ by $1 - \lambda$ in the right-hand side of Eq. (20), leads to

$$2(1 - \lambda)^2 \sin^2\left(\frac{\alpha_x}{2}\right) = (1 - \lambda)^2 h_\nu^2(Q_{x:0}, Q_{x:1}) < h_\nu^2(Q_{x:0}, Q_{x:2}) < h_\nu^2(Q_{x:0}, Q_{x:1}). \quad (21)$$
II. Construction of the test procedure for the testing problem (5)

Set $\kappa = k + l$ and $N = [n/\kappa]$, where $l$ and $k$ are issued from assumption A4 and $[\cdot]$ denotes the integer part; since $n$ is large enough, note that $N > 1$. We consider $N$ i.i.d. random variables $Y_1, Y_2, \ldots, Y_N$, which are generated independently from $\mathcal{U}_n$ according to the discrete uniform distribution $\mathcal{U}_n[1, \ldots, k]$. We further define the test statistic

$$T(\mathcal{H}_n) = \sum_{i=1}^{N} \log \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}),$$

where

$$\left\{ \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}) = \sqrt{q_{J_{\tau_i-1}0}(J_{\tau_i}, X_{\tau_i})} \right\},$$

$$\tau_i = \kappa(i - 1) + l + Y_i.$$ 

Set $\Phi^{-1}$ being equal to one over $\Phi$, i.e., $\Phi^{-1} = \sqrt{q_{\kappa\kappa}/q_{00}}$. Our test procedure for the hypotheses problem (5) is then defined as follows

$$\psi_1(\mathcal{H}_n) = \mathbb{I}_{\{T(\mathcal{H}_n) > 0\}}.$$ (22)

III. Type I error probability of the test statistic defined by (22)

By means of the Markov property we obtain that

$$\mathbb{E}_0^{(n)}(\psi_1(\mathcal{H}_n)) \leq \mathbb{E}_0^{(n)}\left( \prod_{i=1}^{N-1} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}) \Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N}) \right)$$

$$= \mathbb{E}_0^{(n)}\left( \prod_{i=1}^{N-1} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}) \mathbb{E}_0(\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N}) | \mathcal{H}_k(N-1)) \right)$$

$$= \mathbb{E}_0^{(n)}\left( \prod_{i=1}^{N-1} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}) \mathbb{E}_0(\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N}) | \mathcal{H}_k(N-1)) \right),$$ (23)

where $\mathcal{H}_k(N-1) = (J_0, \ldots, J_k(N-1), X_0, \ldots, X_k(N-1))$.

- **Step “type I error”**

Denoting $T_1 := \mathbb{E}_0(\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N}) | \mathcal{H}_k(N-1))$ and since $\tau_i \sim U_{[\kappa(i-1)+l+1, \ldots, \kappa i]}$, we obtain

$$T_1 = \frac{1}{k} \sum_{u=1}^{k} \mathbb{E}_0[\Phi_{J_{\kappa(N-1)+l+u-1}}(J_{\kappa(N-1)+l+u}, X_{\kappa(N-1)+l+u}) | \mathcal{H}_k(N-1)]$$

$$= \frac{1}{k} \sum_{u=1}^{k} \Gamma_u,$$

where $\Gamma_u = \mathbb{E}_0[\Phi_{J_{\kappa(N-1)+l+u-1}}(J_{\kappa(N-1)+l+u}, X_{\kappa(N-1)+l+u}) | \mathcal{H}_k(N-1)]$; rewrite $\Gamma_u$ as follows,

$$\Gamma_u = \int_{E} \int_{E} \int_{\mathbb{R}^+} \Phi_{x}(y, t) P_0^{(l+u-1)}(J_{\kappa(N-1)}) dx q_X(0, y, t) d\nu(y, t),$$

where $\nu$ is the Lebesgue measure. The first-order expansion of $T_1$ is not difficult to obtain, and it is given by

$$T_1 \approx \sum_{u=1}^{k} \Gamma_u.$$
\[
\begin{align*}
&= \int_E P_0^{(l+u-1)}(J_{k(N-1)}, dx) \int E P_0^{(l+u-1)}(J_{k(N-1)}, dx) \int \Phi_x(y, t)q_{x;0}(y, t)d\nu(y, t) \\
&= \int_E P_0^{(l+u-1)}(J_{k(N-1)}, dx) \int E \sqrt{q_{x;2}(y, t)q_{x;0}(y, t)}d\nu(y, t) \\
&= \int_E P_0^{(l+u-1)}(J_{k(N-1)}, dx)(1 - h_v^2(Q_{x;0}, Q_{x;2})),
\end{align*}
\]

where the last equality follows from (18).

Assumption A4 and Eq. (21) lead to the following upper bound of \( T_1 \):

\[
T_1 = 1 - \frac{1}{k} \sum_{u=1}^{k} \int E P_0^{(l+u-1)}(J_{k(N-1)}, dx)h_v^2(Q_{x;0}, Q_{x;2})
\]

\[
\leq 1 - \int E h_v^2(Q_{x;0}, Q_{x;2})d\nu^*(x)
\]

\[
\leq 1 - (1 - \lambda)^2 \int E h_v^2(Q_{x;0}, Q_{x;1})d\nu^*(x)
\]

\[
= 1 - (1 - \lambda)^2 d_v^*(q_0, q_1)
\]

\[
\leq e^{-(1-\lambda)^2 d_v^*(q_0, q_1)}
\]

\[
\leq e^{-(1-\lambda)^2 e^2}.
\]

This latter inequality provides a first upper bound for \( \mathbb{E}_0(\psi_1(H_n)) \) via Relation (23).

- Then, by setting \( T_i := \mathbb{E}_0(\Phi_{J_{n,i+1}}^{-1}(J_{n,i+1}, X_{n,i+1})|J_{k(N-i)}) \) for \( i = 2, \ldots, N \), and by repeating Step “type I error” for the successive \( T_i \), we finally obtain

\[
\mathbb{E}_0^{(n)}(\psi_1(H_n)) \leq e^{\frac{1}{n}(1-\lambda)^2 e^2} = e^{-\frac{K}{n} e^2}, \quad \text{with } K = \frac{(1 - \lambda)^2}{2 \lambda}.
\]

**IV. Type II error probability of the test statistic defined by (22)**

To bound from above the type II error probability, we need an additional result stated as Lemma 1. This lemma provides upper bounds for a quantity which is similar to the \( T_1 \)-term appearing in the first type error probability. The main difference here is that this quantity should be bounded from above uniformly over \( q \) in \( B_{d_v^*,n}(q_1, \varepsilon) \). Recall that \( \Phi_x^{-1} = \frac{1}{\Phi_x} = \sqrt{q_{x;0}/q_{x;2}} \).

**Lemma 1** For \( \lambda = \frac{\pi^2}{2}, \) for any \( \lambda \in ]0, 1/11[ \) and for any \( q \in B_{d_v^*,n}(q_1, \varepsilon) \),

\[
\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] \leq 1 + 8 \frac{1 - \lambda}{\lambda} h_v^2(Q_x, Q_{x;1})
\]

\[
- (1 - \lambda)^2 (1 - \frac{\lambda}{1 - \lambda}) h_v^2(Q_{x;0}, Q_{x;1}).
\]

The proof of Lemma 1 is postponed to Sect. 5.3.

Next, let us make a partition of \( E \) into the subset \( G_q \) and its complementary \( G_q^c \), with \( G_q \) defined as follows :

\[
G_q := \{ x \in E : h_v(Q_x, Q_{x;1}) \leq \lambda h_v(Q_{x;0}, Q_{x;1}) \}.
\]
Then, applying Lemma 1 for \( x \in G_q \) leads to

\[
\mathbb{E}_q[\Phi^{-1}_0(J_1, X_1) | J_0 = x] \leq 1 - (1 - \lambda)^2 (1 - (t + 8) \frac{\lambda}{1 - \lambda}) h_\nu^2(Q_x;0, Q_x;1),
\]

where \( 1 - (t + 8) \frac{\lambda}{1 - \lambda} \) is positive whenever \( \lambda < 1/11 \).

Similarly to the calculations of the type I error probability and for any \( q \in B_{d_{q^*}}(q_1, \xi \epsilon) \), we start with

\[
\mathbb{E}_q^{(n)}(1 - \psi_1(H_n)) \leq \mathbb{E}_q^{(n)} \left( \prod_{i=1}^{N-1} \Phi^{-1}_{J^{-1}}(J_{\tau_i}, X_{\tau_i}) \right) \mathbb{E}_q(\Phi^{-1}_{J^{-1}}(J_{\tau_N}, X_{\tau_N}) | J_{k(N-1)})
\]

and we further define \( W_1 \) by

\[
W_1 := \mathbb{E}_q(\Phi^{-1}_{J^{-1}}(J_{\tau_N}, X_{\tau_N}) | J_{k(N-1)})
= \frac{1}{k} \sum_{u=1}^{k} \mathbb{E}_q[\Phi^{-1}_{J^{-1}}(J_{k(N-1)+t_u}, X_{k(N-1)+t_u}) | J_{k(N-1)}].
\]

**Step “type II error”**

Taking into account the partition of \( E \) into \( G_q \) and \( G_q^C \), we obtain

\[
W_1 \leq 1 - (1 - \lambda)^2 (1 - (t + 8) \frac{\lambda}{1 - \lambda}) \frac{1}{k} \sum_{u=1}^{k} \int_E p_q^{(l+u-1)}(J_{k(N-1)}, dx) q_x(y, t) d\nu(y, t)
+ 8 \frac{1}{\lambda} \frac{1}{k} \sum_{u=1}^{k} \int_{G_q^C} p_q^{(l+u-1)}(J_{k(N-1)}, dx) h_\nu^2(Q_x;0, Q_x;1)
\leq 1 - (1 - \lambda)^2 (1 - (t + 8) \frac{\lambda}{1 - \lambda}) \int_E h_\nu^2(Q_x;0, Q_x;1) d\nu^*(x)
+ 8 \frac{1}{\lambda} \int_E h_\nu^2(Q_x, Q_x;1) d\eta^*(x)
\leq 1 - (1 - \lambda)^2 (1 - (t + 8) \frac{\lambda}{1 - \lambda}) d_\nu^2(q_0, q_1) + 8 \frac{1}{\lambda} d_\eta^2(q, q_1)
\leq e^{-(1-\lambda)^2(1-(t+8)\frac{\lambda}{1-\lambda})-8\frac{1}{\lambda} \frac{\lambda}{1-\lambda} \xi^2} \leq e^{-K(\lambda)\xi^2},
\]

\( \odot \) Springer
where $K(\lambda)$ is positive since there exists $\xi > 0$ such that $(1 - \lambda)(1 - (\iota + 8) \frac{\lambda}{1 - \lambda})$ 

$> \frac{1}{2} \xi^2$.

- To complete the proof, we denote by

$$W_i := \mathbb{E}_q(\Phi_{J_{N-i+1}}^{-1}(J_{N-i+1}, X_{N-i+1})|J_{N-i}) \text{ for } i = 2, \ldots, N,$$

and then, repeating Step “type II error” for the successive $W_i$, we finally deduce that for any $q \in B_{d_q,n}(q_1, \xi)$,

$$\mathbb{E}_q^{(n)}\left(1 - \psi_1(\mathcal{H}_n)\right) \leq e^{-n \tilde{K}(\lambda)\xi^2},$$

with $\tilde{K}(\lambda) = K(\lambda)/2\kappa$.

\[\square\]

### 5.3 Proof of Lemma 1

**Proof** We define the Hellinger affinity between two distributions $P_1$ and $P_2$, absolutely continuous w.r.t. $v$, with derivatives $p_1$ and $p_2$ respectively, by

$$q_v(P_1, P_2) := \int \mathbb{R}^+ \int A \sqrt{p_1 p_2} dv = 1 - h_v^2(P_1, P_2).$$

We follow the lines of Birgé (1983a, 2013) and consider a positive real number $A$ such that $A \geq \frac{2}{1 - \lambda}$. We then decompose the term $\mathbb{E}_q[\Phi_{J_{0}}^{-1}(J_1, X_1)|J_0 = x]$ into four terms:

$$\mathbb{E}_q[\Phi_{J_{0}}^{-1}(J_1, X_1)|J_0 = x] \leq \mathbb{E}_q[\Phi_{J_{0}}^{-1}(J_1, X_1)|J_0 = x] + \sum_{i=1}^{3} \int A_{x;i} (\Phi_{x}^{-1} - 1)(q_x - q_{x;1}) dv$$

$$:= T_0 + \sum_{i=1}^{3} T_i,$$

where

$$A_{x;1} = \left\{(y, t) \in E \times \mathbb{R}^+ : \sqrt{\frac{q_{x}(y, t)}{q_{x;1}(y, t)}} > A - 1, \ \Phi_{x}^{-1}(y, t) > 1\right\}$$

$$A_{x;2} = \left\{(y, t) \in E \times \mathbb{R}^+ : 1 \leq \sqrt{\frac{q_{x}(y, t)}{q_{x;1}(y, t)}} \leq A - 1, \ \Phi_{x}^{-1}(y, t) > 1\right\}$$

$$A_{x;3} = \left\{(y, t) \in E \times \mathbb{R}^+ : \sqrt{\frac{q_{x}(y, t)}{q_{x;1}(y, t)}} < 1, \ \Phi_{x}^{-1}(y, t) < 1\right\}.$$

For the sake of simplicity, denote $q_{x;j}(y, t) = q_{x;j}$ for $j = 0, 1, 2$ and $q_{x}(y, t) = q_{x}$. Let us start with $T_0$. The definition of $q_{x;2}$ leads to

$$T_0 = \int \sqrt{\frac{q_{x;0}}{q_{x;2}}} q_{x;1} dv$$

$$= \int \frac{\sqrt{q_{x;0}} \sin(\alpha_x)}{\sin((1 - \lambda)\alpha_x)} \sqrt{q_{x;1}} + \sin(\lambda\alpha_x) \sqrt{q_{x;0}} q_{x;1} dv.$$
\[
= \int \frac{\sqrt{q_{x;0}} \sin(\alpha x)}{\sin((1-\lambda)\alpha x) + \sin(\lambda\alpha x)\sqrt{q_{x;1}}} q_{x;1} \nu. \tag{25}
\]

In (25), we first apply Jensen’s inequality to the concave function
\[
y \mapsto \frac{\sin(\alpha x)y}{\sin(\alpha x)\lambda y + \sin(\alpha x)(1-\lambda)},
\]
defined in \(\mathbb{R}^+\) then, due to Eq. (16), and trigonometric calculations, we deduce that
\[
T_0 \leq \frac{\sin(\alpha x)\rho_\nu(Q_{x;0}, Q_{x;1})}{\sin(\alpha x)\lambda \rho_\nu(Q_{x;0}, Q_{x;1}) + \sin(\alpha x)(1-\lambda))}
\sin(\alpha x) \cos(\alpha x)
\cos(\alpha x) 
= \cos((1-\lambda)\alpha x) - \tan(\lambda\alpha x) \sin((1-\lambda)\alpha x)
= 1 - h^2_\nu(Q_{x;0}, Q_{x;2}) - \tan(\lambda\alpha x) \sin((1-\lambda)\alpha x) \quad \text{(see (18))}
\leq 1 - (1-\lambda)^2 h^2_\nu(Q_{x;0}, Q_{x;1}), \tag{26}
\]

where the last inequality is due to (21) and the penultimate inequality holds true since \(\tan(\lambda\alpha x) \sin((1-\lambda)\alpha x) > 0\) whenever \(\alpha x \in ]0, \pi/2[\) and \(0 < \lambda < 1\).

Let us now turn to \(T_1\); from inequation (19), note that
\[
\Phi^{-1}_x = \sqrt{\frac{q_{x;0}}{q_{x;2}}}
\sin(\alpha x) \sqrt{\frac{q_{x;0}}{q_{x;1}}} \sin(\alpha x) \lambda \sqrt{\frac{q_{x;0}}{q_{x;1}}} + \sin(\alpha x)(1-\lambda)) \leq \frac{\sin(\alpha x)}{\sin(\alpha x) \lambda} < \frac{1}{\lambda}, \tag{27}
\]

On \(A_{x;1}\), we set \(r_x = \frac{q_x}{q_{x;1}}\) and since \(r_x - 1 < \frac{A}{A - 2}(\sqrt{r_x} - 1)^2\), then from (27) we obtain,
\[
T_1 \leq \frac{A}{A - 2} \frac{1-\lambda}{\lambda} \int_{A_{x;1}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 \nu
\leq \frac{A}{A - 2} \frac{1-\lambda}{\lambda} 2h^2_\nu(Q_x, Q_{x;1}) - \frac{A}{A - 2} \frac{1-\lambda}{\lambda} \int_{A_{x;1} \cup A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 \nu, \tag{28}
\]

where the last inequality holds true since \(A_{x;2} \cup A_{x;3} \subset A_{x;1}^G\).
Second we study the last two terms $T_2$ and $T_3$. On $A_{x;2}$ and $A_{x;3}$, we first apply the Cauchy-Schwarz inequality, i.e., \( \forall i \in \{2, 3\} \),
\[
\left( \int_{A_{x;i}} (\Phi_x^{-1} - 1)(r_x - 1)q_{x;1} \, dv \right)^2 \leq \int_{A_{x;i}} (\Phi_x^{-1} - 1)^2 q_{x;1} \, dv \int_{A_{x;i}} (r_x - 1)^2 q_{x;1} \, dv.
\]

Then we note that
\[
\int_{A_{x;i}} (\Phi_x^{-1} - 1)^2 q_{x;1} \, dv = \int_{A_{x;i}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 q_{x;2} \, dv
\leq \beta \int_{A_{x;i}} (\sqrt{q_x} - \sqrt{q_{x;2}})^2 \, dv,
\]
where $\beta$, the upper bound of $\frac{q_{x;1}}{q_{x;2}}$, is given by $\beta = \begin{cases} 1 & \text{on } A_{x;2} \text{ since } \frac{q_{x;1}}{q_{x;0}} < 1, \\ \frac{1}{(1-\lambda)^2} & \text{on } A_{x;3} \text{ due to } (19). \end{cases}$

To prove that $\beta = 1$ on $A_{x;2}$, we rely on the contrapositive statement of “$\frac{q_{x;1}}{q_{x;0}} \geq 1$ using (19)” $\Phi_x^{-1} < 1$ and hence $\Phi_x^{-1} \leq 1$” since on $A_{x;2}$, $\Phi_x^{-1} > 1$.

We further note that
\[
\int_{A_{x;i}} (r_x - 1)^2 q_{x;1} (\cdot, \cdot) \, dv \leq \begin{cases} A^2 \int_{A_{x;2}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 q_{x;2} \, dv \\ 2 \int_{A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;2}})^2 q_{x;1} \, dv \end{cases}.
\]

The latter combined with (29) and since $A > 2/(1-\lambda)$, entails
\[
T_2 + T_3 \leq A \left( \int_{A_{x;2}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 q_{x;2} \, dv \int_{A_{x;2}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 q_{x;1} \, dv \right)^{1/2}
+ \frac{2}{1-\lambda} \left( \int_{A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 q_{x;2} \, dv \int_{A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;2}})^2 q_{x;1} \, dv \right)^{1/2}
\leq \sqrt{2}A \left( \int_{A_{x;2} \cup A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 q_{x;2} \, dv \int_{A_{x;2} \cup A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;2}})^2 q_{x;1} \, dv \right)^{1/2}.
\]

From (26), (28) and (30), it follows that
\[
\mathbb{E}_q[\Phi_{j_0}^{-1}(J_1, X_1)|J_1 = x] \leq 1 - (1-\lambda)^2 h^2_{\nu}(Q_{x;0}, Q_{x;1}) + 2 \frac{A}{A-2} \frac{1-\lambda}{\lambda} h^2_{\nu}(Q_x, Q_{x;1})
- \frac{A}{A-2} \frac{1-\lambda}{\lambda} \int_{A_{x;2} \cup A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 \, dv
+ A\sqrt{2} \left( \int_{A_{x;2} \cup A_{x;3}} (\sqrt{q_x} - \sqrt{q_{x;1}})^2 \, dv \right)^{1/2}
\int_{A_{x;2} \cup A_{x;3}} (\sqrt{q_{x;0}} - \sqrt{q_{x;2}})^2 \, dv.
\]

At a next step we consider the function $f_x$ of $z_x$
\[
f_x: z_x \mapsto - \frac{A}{A-2} \frac{1-\lambda}{\lambda} z_x + \sqrt{2}z_x^{1/2} A \left( \int_{A_{x;2} \cup A_{x;3}} (\sqrt{q_{x;0}} - \sqrt{q_{x;2}})^2 \, dv \right)^{1/2}.
\]
where the last inequality holds true since \( \sin x \leq x \), \( x \neq 0 \).

Due to (16), (17), (32), (33), we get,

\[
\mathbb{E}_q \Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x) \leq 1 - (1 - \lambda)^2 h_v^2(Q_{x:0}, Q_{x:1}) + 2 \frac{A}{A - 2} \frac{1 - \lambda}{\lambda} h_v^2(Q_x, Q_{x:1}) + f_x(z_{x;\text{max}})
\]

\[
= 1 - (1 - \lambda)^2 h_v^2(Q_{x:0}, Q_{x:1}) + 2 \frac{A}{A - 2} \frac{1 - \lambda}{\lambda} h_v^2(Q_x, Q_{x:1}) + \frac{A(A - 2)}{2} \frac{1 - \lambda}{\lambda} h_v^2(Q_{x:0}, Q_{x:2})
\]

\[
\leq 1 - (1 - \lambda)^2 h_v^2(Q_{x:0}, Q_{x:1}) + 2 \frac{A}{A - 2} \frac{1 - \lambda}{\lambda} h_v^2(Q_x, Q_{x:1}) + \frac{A(A - 2)}{2} \frac{1 - \lambda}{\lambda} h_v^2(Q_{x:0}, Q_{x:2})
\]
\[-(1 - \lambda)^2 h_\nu^2 (Q_{x:0}, Q_{x:1}) \left( 1 - A(A - 2) \frac{\lambda}{1 - \lambda} \left( \frac{\pi}{4} \right)^2 \right).\]

Setting \( A = 8/3 \) ensures that \( A \geq 2/(1 - \lambda) \) for \( \lambda < 1/11 \), then, Lemma 1 is proved with \( t = \frac{\pi^2}{9} \) since \( A(A - 2)(\frac{\pi}{4})^2 = \frac{\pi^2}{9} \).

\[\square\]

### 5.4 Proof of Corollary 1

**Proof** The proof of Corollary 1 is similar to the proof of Lemma 9 in Ghosal and van der Vaart (2007). However, we sketch it in order to define the statistical test procedure \( \psi(\mathcal{H}_n) \).

First, consider the partition:

\[\{ q \in \mathcal{Q}_n : d_{v*}(q_0, q) > \epsilon_n M \} = \bigcup_{j \geq 1} \left\{ q \in \mathcal{Q}_n : j \epsilon_n M < d_{v*}(q_0, q) \leq (j + 1) \epsilon_n M \right\}\]

\[=: \bigcup_{j \geq 1} H_j.\]

For \( j \geq 1, \epsilon_n M(j + 1) \leq 2 \epsilon_n M j \) and hence, let us set \( \epsilon \) in assumption \( A5 \) \( \epsilon = \epsilon_n M j \); note that this \( \epsilon \) is convenient since \( \epsilon \geq \epsilon_n \) for \( j \geq 1 \) and \( M > 0 \) large enough.

Then, define \( \tilde{H}_j \) as the \( \xi \epsilon \)-separated set in \( H_j \) w.r.t \( \eta^* \); this means that the following two relations are valid:

\[\forall q, \tilde{q} \in \tilde{H}_j, q \neq \tilde{q}, d_{\eta^*}(q, \tilde{q}) > \xi \epsilon, \]

\[\tilde{H}_j \subset H_j \subset B_{d_{\eta^*}, \epsilon_n}(q_0, 2\epsilon).\]  

(34)

Since \( H_j \subset B_{d_{\eta^*}, \epsilon_n}(q_0, 2\epsilon), \tilde{H}_j \) is also an \( \xi \epsilon \)-separated set of \( B_{d_{\eta^*}, \epsilon_n}(q_0, 2\epsilon) \). The order relation between the \( \xi \epsilon \)-packing number and the \( (\xi \epsilon)/2 \)-covering number of \( B_{d_{\eta^*}, \epsilon_n}(q_0, 2\epsilon) \) implies that

\[|\tilde{H}_j| < N \left( \frac{\epsilon}{2}, B_{d_{\eta^*}, \epsilon_n}(q_0, 2\epsilon) \right).\]

Therefore, assumption \( A5 \) entails that \( |\tilde{H}_j| \leq e^{n \epsilon_n^2} \).

In addition, (34) implies that for each \( q_{j,i} \in \tilde{H}_j, d_{\eta^*}(q_0, q_{j,i}) > \epsilon \), and hence, there exists a statistical test \( \psi_{j,i} \), for the testing problem (5) with \( q_1 = q_{j,i} \), whose performance is given in Proposition 2.

Thus, this last remark enables us to define the test procedure and derive its performance:

\[\psi(\mathcal{H}_n) := \max_{j \geq 1} \max_{q_{j,i} \in \tilde{H}_j} \psi_{j,i}(\mathcal{H}_n).\]

(35)

We further combine assumption \( A5 \) and Proposition 2 to obtain for \( M \) large enough

\[\mathbb{E}^{(n)}_0 [\psi(\mathcal{H}_n)] \leq \sum_{j=1}^{\infty} \sum_{q_{j,i} \in \tilde{H}_j} \mathbb{E}^{(n)}_0 [\psi_{j,i}(\mathcal{H}_n)] \leq e^{n \epsilon_n^2} e^{-K n \epsilon_n^2 M^2} \leq e^{-K n \epsilon_n^2 M^2/2},\]

and

\[\sup_{q \in \bigcup_{j \geq 1} H_j} \mathbb{E}_q^{(n)}[1 - \psi(\mathcal{H}_n)] \leq \sup_{j \geq 1} e^{-K n j^2 \epsilon_n^2 M^2} \leq e^{-K n \epsilon_n^2 M^2}.\]

\[\square\]
5.5 Proof of Theorem 1

Proof Let $M$ be a positive constant. We first decompose the left-hand side of (14) in two parts

$$
\Pi_n^{\mathcal{H}_n}(B_{d,v}^c(q_0, \epsilon_n M)) = \Pi_n^{\mathcal{H}_n}(B_{d,v}^c(q_0, \epsilon_n M) \cap Q_n) + \Pi_n^{\mathcal{H}_n}(B_{d,v}^c(q_0, \epsilon_n M) \cap Q_n^c)
$$

(36)

In the sequel, each term in the right-hand side of (36) is separately bounded from above: for $A_1$, we apply Corollary 1, whereas to upper bound $A_2$ we use $A_6$ and $A_7$.

First, let us focus on $A_1$. Recall that $L_n(q)$, the likelihood function of the sample path $\mathcal{H}_n$ evaluated at $q \in Q$, is given by

$$
L_n(q) = \tilde{\rho}(J_0, X_0) \prod_{i=1}^n q_{j_{i-1}}(J_i, X_i).
$$

Then, $A_1$ could be written as follows:

$$
A_1 = \frac{\int_{B_{d,v}^c(q_0, \epsilon_n M) \cap Q_n} L_n(q) d\Pi_n(q)}{\int_{Q_n} L_n(q) d\Pi_n(q)}
$$

$$
= \frac{\int_{B_{d,v}^c(q_0, \epsilon_n M) \cap Q_n} \frac{L_n(q)}{L_n(q_0)} d\Pi_n(q)}{\int_{Q} \frac{L_n(q)}{L_n(q_0)} d\Pi_n(q)}
$$

$$
= \frac{N_n}{D_n}.
$$

Moreover consider $D_n$ as the following event

$$
D_n = \left\{ D_n \leq e^{-n\epsilon_n^2} \prod_n (\mathcal{U}(q_0, \epsilon_n)) \right\},
$$

where $\mathcal{U}(q_0, \epsilon_n)$ is defined according to (13).

By means of the test procedure defined in (35), $\psi(\mathcal{H}_n)$, $\mathbb{E}_0^{(n)}(A_1)$ could be written as follows

$$
\mathbb{E}_0^{(n)}(A_1) = \mathbb{E}_0^{(n)} \left( \frac{N_n}{D_n} \right)
$$

$$
\leq \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] + \mathbb{E}_0^{(n)} \left[ (1 - \psi(\mathcal{H}_n)) \frac{N_n}{D_n} \left( I_{D_n} + I_{D_n^c} \right) \right]
$$

$$
\leq \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] + \mathbb{E}_0^{(n)} \left[ (1 - \psi(\mathcal{H}_n)) \frac{N_n}{D_n} I_{D_n^c} \right] + \mathbb{E}_0^{(n)}(D_n)
$$

$$
:= T_1 + T_2 + T_3.
$$

(37)

Second, to bound from above $\mathbb{E}_0^{(n)}(A_1)$, it is sufficient to upper bound every term on the right-hand side of (37).

- **TERM T1.** We apply Corollary 1 and obtain that there exists $K > 0$ such that

$$
T_1 = \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] \leq e^{-K n \epsilon_n^2 M^2}.
$$

(38)
• **TERM T2.** We apply once again Corollary 1, which combined with A6 entails that there exists $\tilde{K} > 0$ such that

$$T_2 \leq \int_{B_{d_n}(q_0, \epsilon_n) \cap Q_n} \mathbb{E}_q^{(n)} [1 - \psi(H_n)] d\Pi_n(q) \frac{2}{e^{-n\epsilon_n^2} \Pi_n(U(q_0, \epsilon_n))} \leq \sup_{q \in B_{d_n}(q_0, \epsilon_n) \cap Q_n} \mathbb{E}_q^{(n)} [1 - \psi(H_n)] \frac{2}{e^{-n\epsilon_n^2} \Pi_n(U(q_0, \epsilon_n))} \leq e^{-\tilde{K}n\epsilon_n^2} \frac{2}{\Pi_n(U(q_0, \epsilon_n))} \leq 2e^{-(\tilde{K}M^2 - 1 - c)n\epsilon_n^2}, \quad (39)$$

where $\kappa := \tilde{K}M^2 - 1 - c$ is positive under the condition that $M$ is sufficiently large.

• **TERM T3.** Consider the following subspace of $Q$

$$\mathcal{V}_n := \{ q \in Q : \log \frac{L_n(q)}{L_n(q_0)} + K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) \geq - \frac{n\epsilon_n^2}{2} \},$$

and observe that

$$D_n \geq \int_{U(q_0, \epsilon_n) \cap \mathcal{V}_n} e^{\log \frac{L_n(q)}{L_n(q_0)} + K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)})} - K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) \Pi_n(q) d\Pi_n(q) \geq e^{-n\epsilon_n^2} \Pi_n(U(q_0, \epsilon_n) \cap \mathcal{V}_n).$$

It then follows from Fubini’s theorem and Markov’s inequality that

$$T_3 \leq \mathbb{P}_0^{(n)} \left( e^{-n\epsilon_n^2} \Pi_n(U(q_0, \epsilon_n) \cap \mathcal{V}_n) \right) \leq \frac{e^{-n\epsilon_n^2}}{2} \Pi_n(U(q_0, \epsilon_n))$$

$$= \mathbb{P}_0^{(n)} \left( \frac{\Pi_n(U(q_0, \epsilon_n))}{2} \leq \Pi_n(U(q_0, \epsilon_n) \cap \mathcal{V}_n) \right)$$

$$\leq \frac{2}{\Pi_n(U(q_0, \epsilon_n))} \mathbb{P}_0^{(n)} \left( \Pi_n(\mathcal{V}_n \cap U(q_0, \epsilon_n)) \right)$$

$$= \frac{2}{\Pi_n(U(q_0, \epsilon_n))} \int_{U(q_0, \epsilon_n)} \mathbb{P}_0^{(n)} \left( |\log \frac{L_n(q)}{L_n(q_0)} - K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)})| > \frac{n\epsilon_n^2}{2} \right) d\Pi_n(q)$$

$$\leq \frac{2}{\Pi_n(U(q_0, \epsilon_n))} \int_{U(q_0, \epsilon_n)} \frac{4}{n^2 \epsilon_n^4} \mathbb{V}_0^{(n)}(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) d\Pi_n(q)$$

$$\leq \frac{2}{\Pi_n(U(q_0, \epsilon_n))} \frac{n\epsilon_n^2}{2} \frac{4}{n^2 \epsilon_n^4} \Pi_n(U(q_0, \epsilon_n)) = \frac{4}{n\epsilon_n^2} \to +\infty. \quad (40)$$

Third, let us turn to $A_2$ which is rewritten as follows

$$A_2 = \frac{\int_{B_{d_n}(q_0, \epsilon_n) \cap Q_n} \frac{L_n(q)}{L_n(q_0)} d\Pi_n(q)}{\int_Q \frac{L_n(q)}{L_n(q_0)} d\Pi_n(q)} := \frac{\tilde{N}_n}{D_n}.$$ 

Then, using Eq. (40) and from assumptions A6 and A7, we obtain

$$\mathbb{P}_0^{(n)}(A_2) = \frac{\mathbb{P}_0^{(n)}}{D_n} \left\{ \mathbb{1}_{D_n \leq \frac{e^{-n\epsilon_n^2}}{2} \Pi_n(U(q_0, \epsilon_n))} + \mathbb{1}_{D_n > \frac{e^{-n\epsilon_n^2}}{2} \Pi_n(U(q_0, \epsilon_n))} \right\}$$

\(\square\) Springer
\[ \begin{align*}
\left( D_n \right) + \mathbb{E}_0^{(n)} \left( \hat{N}_n \right) & \leq \frac{2}{\pi^{1/2}} e^{-n\epsilon_n^2} \Pi_n \left( U(q_0, \epsilon_n) \right) \\
\left( D_n \right) + \Pi_n \left( \mathbb{Q}_n^{(n)} \right) & \leq \frac{4}{n\epsilon_n^2} + 2e^{-(c+1)n\epsilon_n^2}.
\end{align*} \]

Finally, Inequalities (38)–(41) lead to the desired result (14).

\[ \square \]

Acknowledgements The authors are grateful to the two anonymous referees and the associate editor. Their valuable comments and suggestions improved considerably this paper. This work is supported by the research projects \textit{DEFIMATHS} and \textit{PANORisk}, région Pays de la Loire.

References

Amewou-Atisso M, Ghosal S, Ghosh JK, Ramamoorthi RV (2003) Posterior consistency for semiparametric regression problems. Bernoulli 9:291–312
Arbel J, Gayraud G, Rousseau J (2013) Bayesian optimal adaptive estimation using a sieve prior. Scand J Stat 40:549–570
Arfè A, Peluso S, Muliere P (2020) The semi-Markov beta-Stacy process: a Bayesian non-parametric prior for semi-Markov processes. Stat Inference Stoch Process. https://doi.org/10.1007/s11203-020-09224-2
Banerjee AK, Bhattacharyya GK (1976) Testing hypotheses in a two-state semi-Markov process. Sankhya: The Indian Journal of Statistics, Series A 38(4):340–356
Barbu VS, Limnios N (2008) Semi-Markov chains and hidden semi-Markov models toward applications. Their use in reliability and DNA analysis. Springer-Verlag, New York
Bath BR, Deshpande SK (1986) Testing for Markov process VS semi-Markov process. Commun Statist-Theor Meth 15:2375–2382
Birgé L (1983a) Robust testing for independent non identically distributed variables and Markov chains. Statistical models. Springer-Verlag, Heidelberg, pp 134–162
Birgé L (1983b) Approximation dans les espaces métriques et théorie de l’estimation. Z Wahrsch Verw Gebiete 65:181–237
Birgé L (2013) Robust tests for model selection, IMS Collections, From Probability to Statistics and Back: High-Dimensional Models and Processes – A Festschrift in Honor of Jon A. Wellner. Inst Math Stat 9:47–64
Bulla P, Muliere P (2007) Bayesian nonparametric estimation for reinforced Markov renewal processes. Stat Inference Stoch Process 10(3):283–303
Chang I-S, Chuang Y-C, Hsiung CA (1999) A class of nonparametric \( K \)-sample tests for semi-Markov counting processes. Statistica Sinica 9:211–227
Chang I-S, Chuang Y-C, Hsiung CA (2001) Goodness-of-fit tests for semi-Markov and Markov survival models with one intermediate state. Scand J Stat 28(3):505–525
Choudhuri N, Ghosal S, Roy A (2004) Bayesian estimation of the spectral density of a time series. J Am Statist Assoc 99:1050–1059
Çinlar E (1969) Markov renewal theory. Adv Appl Probab 1:123–187
Economou T, Bailey TC, Kapelan Z (2014) MCMC implementation for Bayesian hidden semi-Markov models with illustrative applications. Stat Comput 24(5):739–752
Epifani I, Ladelli L, Pievatolo A (2014) Bayesian estimation for a parametric Markov renewal model applied to seismic data. Electron J Stat 8(2):2264–2295
Gassiat E, Rousseau J (2013) About the posterior distribution in hidden Markov models with unknown number of states. Ann Stat 35(1):192–223
Ghosal S, van der Vaart A (2007) Convergence rates of posterior distributions for noniid observations. Ann Stat 35(1):192–223
Ghosal S, van der Vaart A (2017) Fundamentals of nonparametric bayesian inference. Cambridge University Press, Cambridge
Ghosal S, Roy A (2006) Posterior consistency of Gaussian process prior for nonparametric binary regression. Ann Stat 34(5):2413–2429
Ghosal S, Ghosh JK, Ramamoorthi RV (1999) Posterior consistency of Dirichlet mixtures in density estimation. Ann Stat 27(1):143–158
Ghosal S, Ghosh JK, van der Vaart A (2000) Convergence rates of posterior distributions. Ann Stat 28(2):500–531
Ghosh JK, Ramamoorthi RV (2003) Bayesian nonparametrics. Springer-Verlag, New York
Griffin WA, Li X (2016) Using bayesian nonparametric hidden semi-Markov models to disentangle affect processes during marital interaction. PLoS ONE. https://doi.org/10.1371/journal.pone.0155706
Johnson MJ, Willsky AS (2013) Bayesian nonparametric hidden semi-Markov models. J Mach Learn Res 14:673–701
Korolyuk VS, Limnios N (2005) Stochastic systems in merging phase space. World Scientific, Singapore
Le Cam LM (1986) Asymptotic methods in statistical decision theory. Springer, New York
Le Cam LM (1975) On local and global properties in the theory of asymptotic normality of experiments. In: Puri M (ed) Stochastic processes and related topics. Academic Press, London
Le Cam LM (1973) Convergence of estimates under dimensionality restrictions. Ann Stat 1:38–53
Lévy P (1954) Processus semi-markoviens. Proc Int Cong Math (Amsterdam 416–426
Limnios N, Oprișan G (2001) Semi-Markov processes and reliability. Birkhäuser, Boston
Malinovskii VK (1992) Asymptotic optimality of criteria in the problem of testing hypotheses for a recurrent semi-Markov process. J Soviet Math 59(4):955–959
Pati D, Dunsony D, Tokdary S (2013) Posterior consistency in conditional distribution estimation. J Multiv Anal 116:456–472
Phelan MJ (1990) Bayes estimation from a Markov renewal process. Ann Stat 18(2):603–616
Pyke R (1961a) Markov renewal processes: definitions and preliminary properties. Ann Math Stat 32:1231–1242
Pyke R (1961b) Markov renewal processes with finitely many states. Ann Math Stat 32:1243–1259
Rousseau J, Chopin N, Liseo B (2012) Bayesian nonparametric estimation of the spectral density of a long or intermediate memory Gaussian process. Ann Stat 40(2):964–995
Smith WL (1955) Regenerative stochastic processes. Proc R Soc Lond Ser A 232:6–31
Shen X, Wasserman L (2001) Rates of convergence of posterior distributions. Ann Stat 29:687–714
Takács L (1954) Some investigations concerning recurrent stochastic processes of a certain type. Magyar Tud Akad Mat Kutato Int Közl 3:115–128
Tang Y, Ghosal S (2007) Posterior consistency of Dirichlet mixtures for estimating a transition density. J Statist Plann Inference 137:1711–1726
Tsai W-Y (1985) Rank tests for a class of semi-Markov models with censored matched pairs. Stat Probabil Lett 3(5):281–286
Wasserman L (1998) Asymptotic properties of nonparametric Bayesian procedures. Lecture Notes in Statistics 133. Springer, New-York, pp 293–304

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