Hypotheses testing and posterior concentration rates for semi-Markov processes

V S Barbu, Ghislaine Gayraud, Nikolaos Limnios, I. Votsi

Hypotheses testing and posterior concentration rates for semi-Markov processes

V.S. Barbu[*], G. Gayraud[†], N. Limnios[†], I. Votsi[‡]

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 finite 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 nonparametric statistics, posterior concentration rate, semi-Markov kernel, testing procedure, Hellinger distance

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 ([27],[35],[37]) 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 no memoryless distributions could be considered in a semi-Markov environment, duration effects could be reproduced. The duration effect firms that the time the semi-Markov system spends in a state

[*]Postal address: Laboratoire de Mathématiques Raphaël Salem, Université de Rouen-Normandie, UMR 6085, Avenue de l'Université, BP.12, F76801, Saint-Étienne-du-Rouvray, France

[†]Postal address: Sorbonne University, Université de Technologie de Compiègne, LMAC Laboratory of Applied Mathematics of Compiègne - CS 60 319 - 60 203 Compiègne cedex, France.

[‡]Postal address: Laboratoire Manceau de Mathématiques, Le Mans Université, 72000, Le Mans, France.

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 ([32], [33]). Since then, further significant results were obtained by Çinlar [13], Korolyuk et al. [23] and many others. We refer the interested reader to Limnios and Oprişan [28] 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 [4].

Although the statistical inference of SMPs has been extensively studied from a frequentist point of view, the Bayesian literature is rather limited. Except from some specific SMP models ([14],[15]), only a few papers have considered the nonparametric Bayesian theory supporting these models ([9],[31]). 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 $\mathcal{H}_n$ a trajectory of the SMP of length n and by Π the prior distribution of q, which in all generality, could depend on n, and thereafter will be denoted by Π_n. Given $\mathcal{H}_n$ and Π_n, the knowledge on q is updated by the posterior distribution, that is denoted by $\Pi_n^{\mathcal{H}_n}(\cdot) = \Pi_n(\cdot|\mathcal{H}_n)$. 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 $\mathcal{H}_n$. The main topic of the article is the study of the asymptotic behaviour of $\Pi_n^{\mathcal{H}_n}$ in a neighbourhood of Q_0.

Most of the known results in the asymptotic behaviour of posterior distributions in infinite-dimensional models address issues of the 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 [21] and [36] 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 ([1], [2], [12], [17], [19], [20]).

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 [17] 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 ([6], [24], [25], [26]) between two Hellinger balls for a given class of models. We refer the interested reader to [8] for improved results about the existence of such tests for the relevant estimation problems. Tang and Ghosal [38] 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 [16] 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. [30] 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 [40], Ghosh and Ramamoorthi [22] and Ghosal et al. [18].

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 finite state space [17] 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 Marov chains developed by Birgé [6] 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 problem in a semi-Markov context. Bath and Deshpande [5] developed a nonparametric test for testing Markov against semi-Markov processes. Banerjee and Bhattacharyya [3] 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 the Exponential family. Also in a parametric context, Malinovskii [29] considered that the probability distribution of an SMP depends on a real-valued parameter $\vartheta > 0$ and studied the simple hypothesis $H_0 : \vartheta = 0$ against $H_1 : \vartheta = hT^{-1/2}$, $0 < h \leq c$ (the SMP is observed up to time T). Chang et al. ([10], [11]) considered hypotheses testing problems for semi-Markov counting processes, in a survival analysis context. Tsai [39] 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 exchangeable. To the best of our knowledge, the present research is the first one that considers general robust hypotheses testing problems for SMPs in a nonparametric context.

We focus on SMPs since they are much more general and better adapted to applications than the Markov processes. In real-world systems, the state space of the under study processes could be $\{0, 1\}^{\mathbb{N}}$, (e.g., communication systems), where $\mathbb{N}$ is the set of nonnegative integers, 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 the 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 the 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 Section 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 Section 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 $\mathbf{Z} := (Z_t)_{t\in\mathbb{R}^+}$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The semi-Markov process $\mathbf{Z}$ corresponding to the Markov renewal process (MRP) $(\mathbf{J}, \mathbf{S}) := (J_n, S_n)_{n\in\mathbb{N}}$, is defined by

$$Z_t := J_{N(t)}, \quad t \in \mathbb{R}^+,$$

where $0 \le S_0 \le \ldots \le S_n \le \ldots$ are the successive $\mathbb{R}^+$-valued jump times of $\mathbf{Z}$, $(J_n)_{n\ge 0}$ denotes the successive visited states at these jump times (henceforth called *the embedded Markov chain (EMC)*) and

$$N(t) = \begin{cases} 0, & \text{if} \quad S_1 - S_0 > t, \\ \sup\{n \in \mathbb{N}^* : S_n \le t\}, & \text{if} \quad S_1 - S_0 \le t. \end{cases}$$

S_0 may be viewed as the first non-negative time at which a jump is observed. 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 $(\mathbf{J}, \mathbf{S})$ satisfies the following Markov property, i.e., for any $n \in \mathbb{N}$, any $t \in \mathbb{R}^+$ and any $B \in \mathcal{E}$:

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

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

$$Q_x(B, t) := \mathbb{P}(J_{n+1} \in B, S_{n+1} - S_n \le t | J_n = x), \ x \in E, \ t \in \mathbb{R}^+, \ B \in \mathcal{E}$$

Since we suppose that the distribution of $\mathbf{Z}$ is unknown, we focus our interest on the semi-Markov kernel. In particular the stochastic behavior of the SMP $\mathbf{Z}$ 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

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

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

It is worth mentioning that

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

The following assumptions have to be considered in the sequel.

A1 The embedded Markov chain $(J_n)_{n\in\mathbb{N}}$ is ergodic with stationary probability measure $\boldsymbol{\rho}$ (that is $\boldsymbol{\rho}P = \boldsymbol{\rho}$, with P the transition kernel of $\mathbf{J}$ and $\boldsymbol{\rho}(E) = 1$).

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

$$\int_E \boldsymbol{\rho}(dx)m(x) < \infty.$$

A3

$$\mathbb{P}(S_{n+1} - S_n \le t | J_n = x, J_{n+1} = y) \ne \mathbf{1}_{\mathbb{R}^+}(t), \forall n \in \mathbb{N},\ \forall t \in \mathbb{R}^+,\ \forall x, y \in E.$$

Note that **A2** and **A3** ensure that for all non negative t and $B \in \mathcal{E}$, $\mathbb{P}(Z_t \in B)$ is always well-defined and non-zero. However the conditional probability in Assumption **A3** may be defined as any Dirac measure on positive real numbers.

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.) ν, 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 $n \ge 1$, let $X_n := S_n - S_{n-1}$ be the successive sojourn times of $\mathbf{Z}$ and $0 \le X_0 = S_0$. On $\mathcal{E} \otimes \mathbb{B}^+$, we further define the measure $\widetilde{\boldsymbol{\rho}}$ as the distribution of $(\mathbf{J}, \mathbf{X}) := (J_n, X_n)_{n\in\mathbb{N}}$, where

$$\widetilde{\boldsymbol{\rho}}(A, \Gamma) = \int_E \boldsymbol{\rho}(dx)Q_x(A, \Gamma),\ \forall A \in \mathcal{E},\ \forall \Gamma \in \mathbb{B}^+. \tag{2}$$

Proposition 1. *The measure* $\widetilde{\boldsymbol{\rho}}$ *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 $(\mathbf{J}, \mathbf{X})$ its stationary distribution, $\widetilde{\boldsymbol{\rho}}$. To avoid complicated notation, we will also use $\widetilde{\boldsymbol{\rho}}$ to denote the density w.r.t. ν.

In the sequel, the hypotheses **A1**, **A2** and **A3** are considered to hold true.

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. ν, cf. Section 2. We suppose that q_0 belongs to a certain set of semi-Markov kernel densities $\mathcal{Q}$ defined by

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

which is equipped with a metric d that will be defined in the sequel. Next consider ϵ-neighborhoods around q_0 in $\mathcal{Q}$ w.r.t. d, that is

$$B_d(q_0, \epsilon) = \Big\{ q \in \mathcal{Q} : d(q_0, q) \leq \epsilon \Big\}.$$

To allow some flexibility, it is quite common to deal with $\mathcal{Q}_n$, a subset of $\mathcal{Q}$, that may depend on n, such that the prior distribution Π_n on $\mathcal{Q}$ assigns most of its mass on $\mathcal{Q}_n$ (see Assumption **H4** below). An ϵ-neighborhood around q_0 in $\mathcal{Q}_n$ w.r.t. d will be denoted by $B_{d,n}(q_0, \epsilon)$.

As noted by Birgé [6] in the setting of Markov chains, there exists a priori no "natural" distance d between two semi-Markov kernel densities. Nevertheless, a natural distance could be defined between two probability distributions $Q_{x;1}$ and $Q_{x;2}$ dominated by ν and corresponding to the same initial state $J_0 = x \in E$. Indeed, if we further denote by $q_{x;1}$ and $q_{x;2}$ their respective RN derivatives, and following the lines of Birgé [6], d could be defined in two steps. First by considering the squared Hellinger distance between $Q_{x;1}$ and $Q_{x;2}$, i.e.,

$$h_\nu^2(Q_{x;1}, Q_{x;2}) = \frac{1}{2} \int_{E \times \mathbb{R}^+} \left(\sqrt{q_{x;1}(y,t)} - \sqrt{q_{x;2}(y,t)} \right)^2 d\nu(y,t), \tag{3}$$

and second, given a measure on $\mathcal{E}$, say μ, by setting a semi-distance d_μ 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 Π_n on $\mathcal{Q}$. We aim to establish how fast the posterior distribution shrinks, in terms of d, the "true" semi-Markov kernel density, q_0. The precise definition of d will be given after the statement of Assumption **H1**, where the measure μ is fixed. More precisely, our objective is to find the minimal positive sequence ϵ_n tending to zero as n goes to infinity, such that under some assumptions on both $\mathcal{Q}$ and Π_n

$$\Pi_n^{\mathcal{H}_n}\left(B_d^{\complement}(q_0, \epsilon_n) \right) \xrightarrow{L_1(\mathbb{P}_0^{(n)})} 0 \text{ as } n \to 0,$$

where $B_d^{\complement}$ denotes the complementary of B_d in $\mathcal{Q}$ and $\mathbb{P}_0^{(n)}$ refers to the "true" distribution of $\mathcal{H}_n$.

Let us denote by $\mathbb{P}_q^{(n)}$ 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 $\mathrm{V}_q^{(n)}$ 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 [7]. Other examples of tests could be found in [17] and in [34]. However, to the best of our knowledge, no such tests exist for semi-Markov processes. 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

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

for some $\xi \in (0, 1)$.

In order to derive posterior concentration rates for SMK densities, one more assumption is required.

- **H1**: There exist two measures ν^* and η^* on $\mathcal{E}$ and two positive integers k, l such that for any $x \in E$,

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

where $P^{(\cdot)}$ is defined in (1). Note that **H1** implies the following inequalities which serve to prove Proposition 2:

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

Proposition 2. *Under Hypothesis **H1**, for any $n \in \mathbb{N}^*$, there exist universal positive constants $\xi \in (0, 1)$, K and $\tilde{K}$ such that for any $\epsilon > 0$ and any $q_1 \in \mathcal{Q}_n$ such that $d_{\nu *}(q_1, q_0) > \epsilon$, there exists a test $\psi_1(\mathcal{H}_n)$ satisfying*

$$\mathbb{E}_0^{(n)}[\psi_1(\mathcal{H}_n)] \leq e^{-Kn\epsilon^2} \quad \text{and} \quad \sup_{q \in \mathcal{Q}_n : d_{\eta *}(q_1, q) < \epsilon\xi} \mathbb{E}_q^{(n)}[1 - \psi_1(\mathcal{H}_n)] \leq e^{-\tilde{K}n\epsilon^2}. \tag{6}$$

The next corollary generalizes Proposition 2 to any $q_1 \in \mathcal{Q}_n$ which is $\epsilon-$distant from q_0 w.r.t. $d_{\nu*}$. It requires an additional assumption (see hereafter **H2**) to control the complexity of $\tilde{\mathcal{Q}}_n \subseteq \mathcal{Q}_n$. This assumption is based on the minimum number of $d_{\nu*}$-balls of radius $\tilde{\epsilon}$ needed to cover $\tilde{\mathcal{Q}}_n$, which is denoted by $N(\tilde{\epsilon}, \tilde{\mathcal{Q}}_n, d_{\nu*})$.

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 Hypothesis **H1**, assume that for a sequence ϵ_n of positive numbers such that $\lim_{n\to+\infty} \epsilon_n = 0$ and $\lim_{n\to+\infty} n\epsilon_n^2 = 0$, the following assumption holds true.*

- ***H2** For ξ in $(0,1)$,*

$$\sup_{\epsilon>\epsilon_n} \log N\big(\epsilon\xi, B_{d_{\nu*},n}(q_0,\epsilon), d_{\eta*}\big) \leq n\epsilon_n^2.$$

Then, there exists a test $\psi(\mathcal{H}_n)$ satisfying

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

and

$$\sup_{q\in\mathcal{Q}_n : d_{\nu*}(q_0,q)>\epsilon_n M} \mathbb{E}_q^{(n)}[1-\psi(\mathcal{H}_n)] \leq e^{-\tilde{K}n\epsilon_n^2 M^2}.$$

3.2 Particular cases

In this paper the results are rather generic 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.

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

- General state space

  Let us first denote by

$$q_x(y,k) = \mathbb{P}(J_{n+1} = y, X_{n+1} = k | J_n = x),$$

  the RN derivative of the SMK. Then for any $k \in \mathbb{N}$ and any $B \in \mathcal{E}$, the respective cumulative semi-Markov kernel is given by

$$Q_x(B,k) = \mathbb{P}(J_{n+1} \in B, X_{n+1} \leq k | J_n = x).$$

It should be noted that in this case ν in (3) is the product measure between a finite-measure μ on $(E, \mathcal{E})$ used in (4) and the counting measure on $\mathbb{N}$. Thus in this framework, the squared Hellinger distance becomes

$$h_\mu^2(Q_{x;1}, Q_{x;2}) \quad = \quad \frac{1}{2} \sum_{k \in \mathbb{N}} \int_E \left(\sqrt{q_{x;1}(y,k)} - \sqrt{q_{x;2}(y,k)} \right)^2 d\mu(y),$$

while the semi-distance d_μ between q_1 and q_2 is given in Equation (4).

- Finite state space

For any $k \in \mathbb{N}$ and any $y \in E$, we define by

$$q_x(y,k) = \mathbb{P}(J_{n+1} = y, X_{n+1} = k | J_n = x), \tag{7}$$

the semi-Markov kernel and by

$$Q_x(y,k) = \mathbb{P}(J_{n+1} = y, X_{n+1} \leq k | J_n = x)$$

the cumulative semi-Markov kernel, respectively.

Since in this framework μ is the counting measure on $(E, \mathcal{E})$, the squared Hellinger distance becomes

$$h^2(Q_{x;1}, Q_{x;2}) \quad = \quad \frac{1}{2} \sum_{k \in \mathbb{N}} \sum_{y \in E} \left(\sqrt{q_{x;1}(y,k)} - \sqrt{q_{x;2}(y,k)} \right)^2, \tag{8}$$

and the semi-distance d between q_1 and q_2 is given by

$$d^2(q_1, q_2) \quad = \quad \sum_{x \in E} h^2(Q_{x;1}, Q_{x;2}). \tag{9}$$

3.2.2 Continuous-time SMPs

- Finite state space

Let us first denote by

$$Q_x(y,t) = \mathbb{P}(J_{n+1} = y, X_{n+1} \leq t | J_n = x) \tag{10}$$

the semi-Markov kernel, for any $y \in E$ and any $t \in \mathbb{R}^+$.

In this context, the squared Hellinger distance becomes

$$h_{\nu_1}^2(Q_{x;1}, Q_{x;2}) \quad = \quad \frac{1}{2} \sum_{y \in E} \int_{\mathbb{R}^+} \left(\sqrt{q_{x;1}(y,t)} - \sqrt{q_{x;2}(y,t)} \right)^2 d\nu_1(t),$$

where ν_1 is the marginal on $(\mathbb{R}^+, \mathbb{B}^+)$ of the measure ν defined on $E \times \mathbb{R}^+$, and the semi-distance d between q_1 and q_2 is defined as in Eq. (9).

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

$$\begin{aligned}
&\tilde{H}_0 : Q_0 \text{ Markov kernel} \quad \text{vs} \\
&\tilde{H}_1 : Q_1 \text{ semi-Markov kernel } \epsilon \text{ distant from } Q_0 \text{ w.r.t. some pseudo-metric.}
\end{aligned}$$

Note that $\tilde{H}_1$ could be extended to any $\xi\epsilon-$ball around Q_1 with $\xi \in]0,1[$.

In this section, we are going to explain how the hypothesis testing problem $\tilde{H}_0$ versus $\tilde{H}_1$ can directly be handled from solving the hypothesis problem H_0 versus H_1 stated in (5).

First, for the discrete-time and finite state space case, assume that we have a Markov process with Markov transition matrix $\widetilde{p} = (\widetilde{p}_{xy})_{x,y\in E}$, $\widetilde{p}_{xx} \neq 1$ 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

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

Consequently, we can define the corresponding squared Hellinger distance as in (8) and construct the corresponding testing procedure.

Second, for the continuous-time and finite state space case, consider a regular jump Markov process with continuous transition semigroup $\widetilde{P} = \left(\widetilde{P}(t)\right)_{t\in\mathbb{R}^+}$ 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

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

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

Note that one can also consider the case where the null hypothesis is composite or the case where the alternative hypothesis is simple, with straightforward modifications.

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 $q \in \mathcal{Q}$ is given by

$$\mathcal{L}_n(q) \quad = \quad \widetilde{\boldsymbol{\rho}}(J_0, S_0) \prod_{\ell=1}^{n} q_{J_{\ell-1}}(J_\ell, X_\ell).$$

Let us introduce the tools that play a central role in asymptotic Bayesian nonparametrics: the Kullback-Liebler (KL) divergence between any two distributions $\mathbb{P}_{q_1}^{(n)}$ and $\mathbb{P}_{q_2}^{(n)}$ and the centered second moment of the integrand of the corresponding KL divergence, which are defined by

$$\begin{aligned} K(\mathbb{P}_{q_1}^{(n)}, \mathbb{P}_{q_2}^{(n)}) &:= \mathbb{E}_0^{(n)} \Big[\log \frac{\widetilde{\boldsymbol{\rho}}_1(J_0, S_0)}{\widetilde{\boldsymbol{\rho}}_2(J_0, S_0)} \prod_{l=1}^{n} \frac{q_{J_{l-1};1}(J_l, X_l)}{q_{J_{l-1};2}(J_l, X_l)} \Big], \\ V_0(\mathbb{P}_{q_1}^{(n)}, \mathbb{P}_{q_2}^{(n)}) &:= \mathbb{V}_0^{(n)} \Big[\log \frac{\widetilde{\boldsymbol{\rho}}_1(J_0, S_0)}{\widetilde{\boldsymbol{\rho}}_2(J_0, S_0)} \prod_{l=1}^{n} \frac{q_{J_{l-1};1}(J_l, X_l)}{q_{J_{l-1};2}(J_l, X_l)} \Big], \end{aligned}$$

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

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

$$\mathcal{U}(q_0, \epsilon) = \Big\{ q \in \mathcal{Q} : K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) \leq n\epsilon^2, V_0(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) \leq n\epsilon^2 \Big\}.$$

It is worth mentioning that although $\widetilde{\boldsymbol{\rho}}$ is not of primary interest, since it is unknown it should require a prior. But since any prior on $\widetilde{\boldsymbol{\rho}}$ that is independent of the prior on q would disappear upon marginalization of the posterior of $(\widetilde{\boldsymbol{\rho}}, q)$ relatively to $\widetilde{\boldsymbol{\rho}}$, 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 Π_n denotes a prior distribution on $\mathcal{Q}$.

Theorem 1. *Assume that* **H1** holds true and suppose that for a sequence of positive numbers ϵ_n such that $\lim_{n \to +\infty} \epsilon_n = 0$, $\lim_{n \to +\infty} n\epsilon_n^2 = 0$, **H2** and **H3-H4** defined hereafter, hold true.

- **H3** $\exists\, c > 0$, $\Pi_n\big(\mathcal{U}(q_0, \epsilon_n)\big) > e^{-cn\epsilon_n^2}$,
- **H4** $\mathcal{Q}_n \subset \mathcal{Q}$ is such that $\Pi_n\big(\mathcal{Q}_n^{\complement}\big) \leq e^{-2n(c+1)\epsilon_n^2}$.

Then for M large enough,

$$\Pi_n^{\mathcal{H}_n}\big(B_{d_{\nu*}}^{\complement}(q_0, \epsilon_n M)\big) \xrightarrow{L_1(\mathbb{P}_0^{(n)})} 0, \quad \text{as} \quad n \to \infty. \tag{11}$$

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

- Under **H1**, Theorem 1 guarantees that, for both a particular set of semi-Markov kernels $\mathcal{Q}$ containing some subset $\mathcal{Q}_n$ such that **H2** holds true for a sequence of positive numbers ϵ_n and a prior distribution Π_n on $\mathcal{Q}$ satisfying assumptions **H3**-**H4** with ϵ_n, the posterior distribution shrinks towards $q_0 \in \mathcal{Q}$ at a rate proportional to ϵ_n.
- Assumption **H3** is classical in Bayesian Nonparametrics; it states that the prior distribution puts enough mass around KL neighborhoods of q_0.
- As mentioned in Section 3.1, $\mathcal{Q}_n$ has to be almost the support of Π_n: it is guaranteed by Assumption **H4**, which in addition quantifies how Π_n covers $\mathcal{Q}_n$. If **H2** holds true with $B_{d_{\nu*}}(q_0, \epsilon)$ instead of $B_{d_{\nu*},n}(q_0, \epsilon)$, then $\mathcal{Q}_n$ coincides with $\mathcal{Q}$ and Assumption **H4** is no more needed.
- Although our semi-Markov framework differs from the Markov one, it is worth noticing that Assumption **H1** is similar to the one stated as Equation (4.1) in Ghosal and van Der Vaart [17]. In particular, for Markov chains, this assumption is related to the transition probabilities of the Markov chain, whereas in our context, **H1** is concerned with the SMK density.

Note also that Assumption **H1** could be replaced by the following:

- $\widetilde{\mathbf{H1}}$: There exists a strictly positive constant C and a strictly positive integer k such that for any $x \in E$,

$$\frac{1}{k}\sum_{u=1}^{k} P^{(u)}(x, \cdot) \geq C.$$

5 Proofs

5.1 Proof of Proposition 1

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 $\Gamma \in \mathbb{B}^+$, we have

$$\begin{aligned} \widetilde{\boldsymbol{\rho}}Q(A,\Gamma) &:= \int_{E\times\mathbb{R}^+} \widetilde{\boldsymbol{\rho}}(dy,ds)Q_y(A,\Gamma) \\ &= \int_{E\times E\times\mathbb{R}^+} \boldsymbol{\rho}(dx)Q_x(dy,ds)Q_y(A,\Gamma) \\ &= \int_E \boldsymbol{\rho}(dy)Q_y(A,\Gamma) \\ &= \widetilde{\boldsymbol{\rho}}(A,\Gamma). \end{aligned}$$

Second,

$$\widetilde{\boldsymbol{\rho}}(E,\mathbb{R}^+) = \int_E \boldsymbol{\rho}(dx)Q_x(E,\mathbb{R}^+) = 1.$$

5.2 Proof of Proposition 2

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.,

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

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

$$\widetilde{H}_{0,x} : q_{x;0}(\cdot,\cdot) \text{ against } \widetilde{H}_{1,x} : \ \left\{q_x(\cdot,\cdot) : h_\nu^2(Q_x,Q_{x;1}) \leq 1-\cos(\lambda\alpha_x)\right\}, \tag{12}$$

where λ is any value in $]0,1/4[$ and α_x belongs to $]0,\pi/2[$ such that

$$h_\nu^2(Q_{x;0},Q_{x;1}) = 1-\cos(\alpha_x). \tag{13}$$

Based on this least favorable pair of q_x's, we will then derive the construction of $\psi_1(\mathcal{H}_n)$ for the testing problem (5).

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.

Least favorable pair of q_x's for the testing problem (12)

For our purposes, we adapt the construction of Birgé [6] for Markov chains to the semi-Markov framework. Whatever is x in E, we attach to x a particular probability density function $q_{x;2} \in \tilde{H}_{1,x}$ defined by

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

By construction, the following relations hold:

$$\begin{aligned} \lambda^2 h_\nu^2(Q_{x;0}, Q_{x;1}) &\leq h_\nu^2(Q_{x;1}, Q_{x;2}) \leq h_\nu^2(Q_{x;0}, Q_{x;1}); && (14) \\ (1-\lambda)^2 h_\nu^2(Q_{x;0}, Q_{x;1}) &\leq h_\nu^2(Q_{x;0}, Q_{x;2}); && (15) \\ h_\nu^2(Q_{x;1}, Q_{x;2}) &= 1 - \cos(\lambda \alpha_x); && (16) \\ h_\nu^2(Q_{x;0}, Q_{x;2}) &= 1 - \cos((1-\lambda)\alpha_x). \end{aligned}$$

Construction of the test procedure for the testing problem (5)

Next, we set $\kappa = k + l$ and $N = [n/\kappa]$, where l and k are issued from Assumption **H1** and $\left[\,\cdot\,\right]$ denotes the integer part. We consider N i.i.d. random variables $Y_1, Y_2, \ldots, Y_N$, which are generated independently from $\mathcal{H}_n$ according to the discrete uniform distribution $\mathcal{U}_{\{1,\ldots,k\}}$.

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

$$\begin{cases} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}) &= \sqrt{\frac{q_{J_{\tau_i-1};2}(J_{\tau_i}, X_{\tau_i})}{q_{J_{\tau_i-1};0}(J_{\tau_i}, X_{\tau_i})}}, \\ \tau_i &= \kappa(i-1) + l + Y_i. \end{cases}$$

Our test procedure for the hypotheses problem (5) is then defined as follows

$$\psi_1(\mathcal{H}_n) = \mathbb{1}_{\{T(\mathcal{H}_n) > 0\}}. \tag{17}$$

- Test simple hypothesis vs simple hypothesis

Let us focus on the general SMPs and consider the following statistical test:

$$H_0 : q_0 \text{ against } H_1 : q_1 \text{ with } d_{\nu*}(q_0, q_1) \geq \epsilon.$$

To construct the testing procedure, the test statistic defined in (17), should be modified as follows:

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

where

$$\begin{cases} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i}) &= \sqrt{\frac{q_{J_{\tau_i-1};1}(J_{\tau_i}, X_{\tau_i})}{q_{J_{\tau_i-1};0}(J_{\tau_i}, X_{\tau_i})}}, \\ \tau_i &= \kappa(i-1)+1+Y_i \\ \kappa &= k+1 \\ Y_i &\overset{iid}{\sim} \mathcal{U}_{\{1,\ldots,k\}}. \end{cases}$$

In this case, Hypothesis **H1** reduces to $\mathbf{H1}^\sharp$:

– $\mathbf{H1}^\sharp$: There exist a measure ν^* on $\mathcal{E}$ and a positive integer k such that for any $x \in E$,

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

Then following the steps of the proof of the Proposition 2 and replacing the Assumption **H1** by $\mathbf{H1}^\sharp$ lead us to the desired result. It is worth mentioning that in this case the inequalities (14), (15), (16) and Lemma 1 are not used.

Note also that in Proposition 2, the upper-bound of both errors is the same, equal to $\exp(-Kn\epsilon^2)$.

Type I error probability

By means of the Markov property we obtain that

$$\begin{aligned} \mathbb{E}_0(\psi_1(\mathcal{H}_n)) &\leq \mathbb{E}_0\Big(\prod_{i=1}^{N-1} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i})\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N})\Big) \\ &= \mathbb{E}_0\Big(\prod_{i=1}^{N-1} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i})\mathbb{E}_0(\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N})|\mathcal{H}_{\kappa(N-1)})\Big) \\ &= \mathbb{E}_0\Big(\prod_{i=1}^{N-1} \Phi_{J_{\tau_i-1}}(J_{\tau_i}, X_{\tau_i})\mathbb{E}_0(\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N})|J_{\kappa(N-1)})\Big), \end{aligned} \quad (18)$$

where $\mathcal{H}_{\kappa(N-1)} = (J_0, \ldots, J_{\kappa(N-1)}, X_0, \ldots, X_{\kappa(N-1)},)$.

- **Step 1**

  Set $T_1 := \mathbb{E}_0(\Phi_{J_{\tau_N-1}}(J_{\tau_N}, X_{\tau_N})|J_{\kappa(N-1)})$. Since $\tau_i \sim U_{\{\kappa(i-1)+l+1,\ldots,\kappa i\}}$, we obtain

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

Next set $\Gamma_u := \mathbb{E}_0\big[\Phi_{J_{\kappa(N-1)+l+u-1}}(J_{\kappa(N-1)+l+u}, X_{\kappa(N-1)+l+u})|J_{\kappa(N-1)}\big]$ and rewrite Γ_u as follows,

$$\begin{aligned}
\Gamma_u &= \int_E \int_E \int_{\mathbb{R}^+} \Phi_x(y,t) P_0^{(l+u-1)}(J_{\kappa(N-1)}, dx) q_{x;0}(y,t) d\nu(y,t) \\
&= \int_E P_0^{(l+u-1)}(J_{\kappa(N-1)}, dx) \int_E \int_{\mathbb{R}^+} \Phi_x(y,t) q_{x;0}(y,t) d\nu(y,t) \\
&= \int_E P_0^{(l+u-1)}(J_{\kappa(N-1)}, dx) \Big(1 - h_\nu^2(Q_{x;0}, Q_{x;2})\Big),
\end{aligned}$$

where the last equality is due to

$$\int_E \int_{\mathbb{R}^+} \sqrt{q_{x;2} q_{x;0}} d\nu = 1 - h_\nu^2(Q_{x;0}, Q_{x;2}).$$

Assumption **H1** and Eq. (15) lead us to the following upper bound of T_1:

$$\begin{aligned}
T_1 &= 1 - \frac{1}{k} \sum_{u=1}^{k} \int_E P_0^{(l+u-1)}(J_{\kappa(N-1)}, dx) h_\nu^2(Q_{x;0}, Q_{x;2}) \\
&\leq 1 - \int_E h_\nu^2(Q_{x;0}, Q_{x;2}) d\nu^*(x) \\
&\leq 1 - (1-\lambda)^2 \int_E h_\nu^2(Q_{x;0}, Q_{x;1}) d\nu^*(x) \\
&= 1 - (1-\lambda)^2 d_{\nu^*}^2(q_0, q_1) \\
&\leq e^{-(1-\lambda)^2 d_{\nu^*}^2(q_0,q_1)} \\
&\leq e^{-(1-\lambda)^2 \epsilon^2}.
\end{aligned}$$

This latter inequality provides a first upper bound of $\mathbb{E}_0(\psi_1(\mathcal{H}_n))$ via the relation (18).

- Then, by setting $T_i := \mathbb{E}_0(\Phi_{J_{\tau_{N-i+1}-1}}(J_{\tau_{N-i+1}}, X_{\tau_{N-i+1}})|J_{\kappa(N-i)})$ for $i = 2, \ldots, N$, and by repeating **Step 1** for the successive T_i, we finally obtain

$$\mathbb{E}_0\big(\psi_1(\mathcal{H}_n)\big) \leq e^{-\frac{n}{\kappa}(1-\lambda)^2\epsilon^2} = e^{-Kn\epsilon^2}, \quad \text{with } K = \frac{(1-\lambda)^2}{\kappa}.$$

Type II error probability

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 $B_{d_{\eta^*},n}(q_1, \xi\epsilon)$.

This requires the definition of the subset G_q of E by

$$G_q := \{x \in E : h_\nu(Q_x, Q_{x;1}) \leq \lambda h_\nu(Q_{x;0}, Q_{x;1})\},$$

and the notation of its complementary into E by $G_q^{\complement}$.

Lemma 1. *For any* $\lambda \in]0, 1/4[$*, there exists* $\iota \in [0, \frac{3}{4}[$*, such that for all* $q \in B_{d_{\eta^*},n}(q_1, \xi\epsilon)$*,*

- *if* $x \in G_q$*, then*

$$\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] \leq 1 - h_\nu^2(Q_{x;0}, Q_{x;2}) \leq 1 - (1-\lambda)^2 h_\nu^2(Q_{x;0}, Q_{x;1}); \quad (19)$$

- *if* $x \in G_q^\complement$*, then*

$$\begin{aligned}\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] &< 1 + 8\frac{1-\lambda}{\lambda} h_\nu^2(Q_x, Q_{x;1}) \\ &- (1 - \frac{2\lambda}{1-\lambda})[1-\iota] h_\nu^2(Q_{x;0}, Q_{x;1}). \quad (20)\end{aligned}$$

The proof of Lemma 1 is postponed to Section 5.3.

Consider Φ^{-1} equal to one over Φ, that is $\Phi^{-1} = \sqrt{\frac{q_0}{q_2}}$. Similarly to the calculations of the type I error probability, we obtain that for any $q \in B_{d_{\eta*},n}(q_1, \xi\epsilon)$,

$$\mathbb{E}_q\big(1 - \psi_1(\mathcal{H}_n)\big) \leq \mathbb{E}_q\big(\prod_{i=1}^{N-1} \Phi_{J_{\tau_i - 1}}^{-1}(J_{\tau_i}, X_{\tau_i}) \mathbb{E}_q(\Phi_{J_{\tau_N - 1}}^{-1}(J_{\tau_N}, X_{\tau_N})|J_{\kappa(N-1)})\big).$$

Similarly to T_1, we further define W_1 by

$$\begin{aligned}W_1 &:= \mathbb{E}_q(\Phi_{J_{\tau_N - 1}}^{-1}(J_{\tau_N}, X_{\tau_N})|J_{\kappa(N-1)}) \\ &= \frac{1}{k}\sum_{u=1}^{k} \mathbb{E}_q\big[\Phi_{J_{\kappa(N-1)+l+u-1}}^{-1}(J_{\kappa(N-1)+l+u}, X_{\kappa(N-1)+l+u})|J_{\kappa(N-1)}\big].\end{aligned}$$

- **Step 2** Taking into account the partition of E into G_q and $G_q^\complement$, we obtain

$$\begin{aligned}W_1 &= \frac{1}{k}\sum_{u=1}^{k}\int_{\mathbb{R}^+}\int_E\int_E \Phi_x^{-1}(y,t) P_q^{(l+u-1)}(J_{\kappa(N-1)}, dx) q_x(y,t) d\nu(y,t) \\ &= \frac{1}{k}\sum_{u=1}^{k}\int_E P_q^{(l+u-1)}(J_{\kappa(N-1)}, dx)\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] \\ &= \frac{1}{k}\sum_{u=1}^{k}\int_{G_q} P_q^{(l+u-1)}(J_{\kappa(N-1)}, dx)\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] \\ &+ \frac{1}{k}\sum_{u=1}^{k}\int_{G_q^\complement} P_q^{(l+u-1)}(J_{\kappa(N-1)}, dx)\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x].\end{aligned}$$

Combining with $(1-\lambda)^2 > \dfrac{1-3\lambda}{1-\lambda}$, Assumption **H1** and Lemma 1 lead to,

$$\begin{aligned}
W_1 &\leq 1-\frac{1-3\lambda}{1-\lambda}[1-\iota]\frac{1}{k}\sum_{u=1}^{k}\int_E P_q^{(l+u-1)}(J_{\kappa(N-1)},dx)h_\nu^2(Q_{x;0},Q_{x;1}) \\
&+ 8\frac{1-\lambda}{\lambda}\frac{1}{k}\sum_{u=1}^{k}\int_{G_q^{\complement}} P_q^{(l+u-1)}(J_{\kappa(N-1)},dx)h_\nu^2(Q_x,Q_{x;1}) \\
&\leq 1-\frac{1-3\lambda}{1-\lambda}[1-\iota]\int_E h_\nu^2(Q_{x;0},Q_{x;1})d\nu^*(x)+8\frac{1-\lambda}{\lambda}\int_E h_\nu^2(Q_x,Q_{x;1})d\eta^*(x) \\
&= 1-\frac{1-3\lambda}{1-\lambda}[1-\iota]d^2_{\nu^*}(q_0,q_1)+8\frac{1-\lambda}{\lambda}d^2_{\eta^*}(q,q_1) \\
&\leq \exp\left(-\left\{\frac{1-3\lambda}{1-\lambda}[1-\iota]-8\frac{1-\lambda}{\lambda}\xi^2\right\}\epsilon^2\right)=\exp\left(-K(\lambda)\epsilon^2\right),
\end{aligned}$$

where $K(\lambda)$ is positive since there exists $\xi>0$ such that $\dfrac{1-3\lambda}{1-\lambda}[1-\iota] >$ $8\dfrac{(1-\lambda)}{\lambda}\xi^2$.

- To complete the proof, we consider $W_i := \mathbb{E}_q(\Phi^{-1}_{J_{\tau_{N-i+1}-1}}(J_{\tau_{N-i+1}},X_{\tau_{N-i+1}})|J_{\kappa(N-i)})$ for $i=2,\ldots,N$. We then repeat **Step 2** for the successive W_i, and finally deduce that for any $q\in B_{d_{\eta*},n}(q_1,\xi\epsilon)$,

$$\mathbb{E}_q^{(n)}\big(1-\psi_1(\mathcal{H}_n)\big) \;\leq\; \exp\Big(-n\tilde{K}(\lambda)\epsilon^2\Big),$$

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

5.3 Proof of Lemma 1

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

$$\varrho_\nu(P_1,P_2) \;:=\; \int_{\mathbb{R}^+}\int_E \sqrt{p_1p_2}d\nu = 1-h_\nu^2(P_1,P_2).$$

In the sequel, let q be an arbitrary element of $B_{d_{\eta^*},n}(q_1,\xi\epsilon)$.

When x belongs to G_q, the proof of (19) results directly from Theorem 2 in Birgé [8].

When x belongs to $G_q^{\complement}$, i.e., $x\in E$ such that $h_\nu(Q_x,Q_{x;1})>\lambda h_\nu(Q_{x;0},Q_{x;1})$, let us prove the statement (20).

We follow the lines of Birgé [6] and consider a real number A such that $A\geq \dfrac{2}{1-\lambda}$. We then decompose the term $\mathbb{E}_q[\Phi^{-1}_{J_0}(J_1,X_1)|J_0=x]$ into four terms:

$$\begin{aligned}\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] &\leq \mathbb{E}_{q_1}[\Phi_{J_0}^{-1}(J_1, X_1)|J_0 = x] + \sum_{i=1}^{3} \int_{\mathcal{A}_{x;i}} (\Phi_x^{-1} - 1)(q_x - q_{x;1}) d\nu \\ &:= T_0 + \sum_{i=1}^{3} T_i,\end{aligned}$$

where

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

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

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

For the sake of simplicity, set $r_x = \dfrac{q_x}{q_{x;1}}$ and start with T_0. Due to the definition of $\Phi_x^{-1}(\cdot,\cdot)$, to Equation (13) and to the concavity of the function $y \to \dfrac{\sin(\alpha_x) y}{\sin(\alpha_x \lambda) y + \sin(\alpha_x(1-\lambda))}$, we deduce that

$$\begin{aligned} T_0 &\leq \frac{\sin(\alpha_x)\rho_\nu(Q_{x;0}, Q_{x;1})}{\sin(\alpha_x\lambda)\rho_\nu(Q_{x;0}, Q_{x;1}) + \sin(\alpha_x(1-\lambda))} \\ &= \frac{\sin(\alpha_x)\cos(\alpha_x)}{\sin(\alpha_x\lambda)\cos(\alpha_x) + \sin(\alpha_x(1-\lambda))} \\ &= \frac{\cos(\alpha_x)}{\cos(\alpha_x\lambda)} \\ &\leq 1 - \left(1 - \frac{2\lambda}{1-\lambda}\right) h_\nu^2(Q_{x;0}, Q_{x;1}), \end{aligned} \tag{21}$$

where the last inequality results from both the convexity of the *tan* function on $]0, \pi/2[$ and $\lambda < 1/4$.

Let us now turn to T_1. First note that

$$\begin{aligned} \Phi_x^{-1} &= \frac{\sin(\alpha_x)\sqrt{\frac{q_{x;0}}{q_{x;1}}}}{\sin(\alpha_x\lambda)\sqrt{\frac{q_{x;0}}{q_{x;1}}} + \sin(\alpha_x(1-\lambda))} \\ &\leq \frac{\sin(\alpha_x)}{\sin(\alpha_x\lambda)} < \frac{1}{\lambda}, \end{aligned} \tag{22}$$

where (22) results from the following inequality

$$\forall\, \alpha \in]0, \pi/2[, \quad \forall \lambda \in]0,1[, \quad \frac{\sin(\lambda\alpha)}{\lambda \sin(\alpha)} > 1. \tag{23}$$

On $\mathcal{A}_{x;1}$, since $r_x - 1 < \frac{A}{A-2}(\sqrt{r_x}-1)^2$, then from (22) we obtain,

$$\begin{aligned} T_1 &\leq \frac{A}{A-2}\frac{1-\lambda}{\lambda}\int_{\mathcal{A}_{x;1}}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu \\ &\leq \frac{A}{A-2}\frac{1-\lambda}{\lambda}2h_\nu^2(Q_x,Q_{x;1}) - \frac{A}{A-2}\frac{1-\lambda}{\lambda}\int_{\mathcal{A}_x}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu, \end{aligned} \quad (24)$$

where $\mathcal{A}_x$ is a subset of $\mathcal{A}_{x;1}^{\complement}$.

Second we study the last two terms T_2 and T_3. On $\mathcal{A}_{x;2}$ and $\mathcal{A}_{x;3}$, we first apply the Cauchy-Schwarz inequality, i.e., $\forall i \in \{2,3\}$,

$$\left(\int_{\mathcal{A}_{x;i}}(\Phi_x^{-1}-1)(r_x-1)q_{x;1}d\nu\right)^2 \leq \int_{\mathcal{A}_{x;i}}(\Phi_x^{-1}-1)^2 q_{x;1}d\nu \int_{\mathcal{A}_{x;i}}(r_x-1)^2 q_{x;1}d\nu.$$

Second we note that

$$\begin{aligned} \int_{\mathcal{A}_{x;i}}(\Phi_x^{-1}(\cdot,\cdot)-1)^2 q_{x;1}d\nu &= \int_{\mathcal{A}_{x;i}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2\frac{q_{x;1}}{q_{x;2}}d\nu \\ &\leq \beta\int_{\mathcal{A}_{x;i}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu, \end{aligned} \quad (25)$$

where β, the upper bound of $\frac{q_{x;1}}{q_{x;2}}$, is given by $\beta = \begin{cases} 1 & \text{on} \quad \mathcal{A}_{x;2} \quad \text{since } \frac{q_{x;1}}{q_{x;0}} < 1, \\ \frac{1}{(1-\lambda)^2} & \text{on} \quad \mathcal{A}_{x;3} \quad \text{due to (23).} \end{cases}$

We further note that

$$\int_{\mathcal{A}_{x;i}}(r_x-1)^2 q_{x;1}(\cdot,\cdot)d\nu \leq \begin{cases} A^2\int_{\mathcal{A}_{x;2}}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu \\ 2^2\int_{\mathcal{A}_{x;3}}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu \end{cases}.$$

The latter combined with (25) and since $A > 2/(1-\lambda)$, entails

$$\begin{aligned} T_2+T_3 \leq\; & A\left(\int_{\mathcal{A}_{x;2}}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu\int_{\mathcal{A}_{x;2}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu\right)^{1/2} \\ & +\frac{2}{1-\lambda}\left(\int_{\mathcal{A}_{x;3}}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu\int_{\mathcal{A}_{x;3}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu\right)^{1/2} \\ \leq\; & A\left(\int_{\mathcal{A}_x}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu\int_{\mathcal{A}_{x;2}\cup\mathcal{A}_{x;3}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu\right)^{1/2}. \end{aligned} \quad (26)$$

From (21), (24) and (26), it follows that

$$\begin{aligned} \mathbb{E}[\Phi_{J_0}^{-1}(J_1,X_1)|J_0=x] \leq\; & 1-\left(1-\frac{2\lambda}{1-\lambda}\right)h_\nu^2(Q_{x;0},Q_{x;1}) + 2\frac{A}{A-2}\frac{1-\lambda}{\lambda}h_\nu^2(Q_x,Q_{x;1}) \\ & -\frac{A}{A-2}\frac{1-\lambda}{\lambda}\int_{\mathcal{A}_x}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu \\ & +A\left(\int_{\mathcal{A}_x}(\sqrt{q_x}-\sqrt{q_{x;1}})^2 d\nu\int_{\mathcal{A}_{x;2}\cup\mathcal{A}_{x;3}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu\right)^{1/2}. \end{aligned}$$

At a next step we consider the following function of z_x

$$z_x \to -\frac{A}{A-2}\frac{1-\lambda}{\lambda}z_x + z_x^{1/2}A\left(\int_{\mathcal{A}_{x;2}\cup\mathcal{A}_{x;3}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu\right)^{1/2},$$

whose maximum is reached at

$$z_{x;max} = \frac{1}{4}(A-2)^2\left(\frac{\lambda}{1-\lambda}\right)^2\int_{\mathcal{A}_{x;2}\cup\mathcal{A}_{x;3}}(\sqrt{q_{x;0}}-\sqrt{q_{x;2}})^2 d\nu.$$

Hence we obtain a new upper bound of $\mathbb{E}_1[\Phi_{J_0}^{-1}(J_1,X_1)|J_0=x]$, that is

$$\begin{aligned}
\mathbb{E}_q[\Phi_{J_0}^{-1}(J_1,X_1)|J_0=x] &\leq 1-\left(1-\frac{2\lambda}{1-\lambda}\right)h_\nu^2(Q_{x;0},Q_{x;1}) + 2\frac{A}{A-2}\frac{1-\lambda}{\lambda}h_\nu^2(Q_x,Q_{x;1}) \\
&\quad +\frac{A(A-2)}{2}\frac{\lambda}{1-\lambda}h_\nu^2(Q_{x;0},Q_{x;2}) \\
&\leq 1+2\frac{A}{A-2}\frac{1-\lambda}{\lambda}h_\nu^2(Q_x,Q_{x;1}) - (1-\frac{2\lambda}{1-\lambda})h_\nu^2(Q_{x;0},Q_{x;1}) \\
&\quad +A(A-2)\frac{\lambda}{1-\lambda}h_\nu^2(Q_{x;0},Q_{x;1})\sin^2\left((1-\lambda)\frac{\pi}{4}\right), \\
&\leq 1+2\frac{A}{A-2}\frac{1-\lambda}{\lambda}h_\nu^2(Q_x,Q_{x;1}) \\
&\quad -\left(1-\frac{2\lambda}{1-\lambda}\right)\left[1-\frac{A(A-2)\lambda}{(1-3\lambda)}\sin^2\left((1-\lambda)\frac{\pi}{4}\right)\right]h_\nu^2(Q_{x;0},Q_{x;1}),
\end{aligned}$$

where the penultimate inequality results from the increase of the function $x\in]0,\pi/2[\to\dfrac{\sin(\lambda x/2)}{\lambda\sin(x/2)}$ for any $\lambda\in]0,1]$.

Finally, by setting $A=8/3$ that satisfies $A\geq 2/(1-\lambda)$ and using both inequalities $\sin^2\left((1-\lambda)\frac{\pi}{4}\right) < (1-\lambda)^2\left(\frac{\pi}{4}\right)^2$ $\forall\lambda\in]0,1/4[$ and $\dfrac{\lambda(1-\lambda)^2}{1-3\lambda} < \dfrac{9}{16}$ $\forall\lambda\in]0,1/4[$, Lemma 1 is proved with $\iota=\frac{\pi^2}{16}<3/4$. □

5.4 Proof of Corollary 1

The proof of Corollary 1 is similar to the proof of Lemma 9 in [17]. However, we sketch it in order to define the statistical test procedure $\psi(\mathcal{H}_n)$. First, consider the partition:

$$\begin{aligned}
\{q\in\mathcal{Q}_n : d_{\nu*}(q_0,q)>\epsilon_n M\} &= \bigcup_{j\geq 1}\left\{q\in\mathcal{Q}_n : j\epsilon_n M < d_{\nu*}(q_0,q)\leq (j+1)\epsilon_n M\right\} \\
&=: \bigcup_{j\geq 1} H_j.
\end{aligned}$$

For $\xi\in]0,1[$, and any $j\geq 1$, we consider $\widetilde{H}_j$, a $j\epsilon_n\xi M$-net on H_j for the distance $d_{\eta*}$ satisfying three conditions:

- $\forall q \in \widetilde{H}_j$, $d_{\nu*}(q_0, q) \geq j\epsilon_n M$;
- $\forall q \in H_j$, $\exists q_j \in \widetilde{H}_j$ such that $d_{\eta*}(q, q_j) \leq j\epsilon_n \xi M$;
- $\log N\big(\epsilon_n M\xi, \widetilde{H}_j, d_{\eta*}\big) \leq n\epsilon_n^2$ (due to **H2**).

For $j \geq 1$ and any $q_{j,i} \in \widetilde{H}_j$, we then apply Proposition 2 with $\epsilon = jM\epsilon_n$ and $q_1 = q_{j,i}$; this implies the existence of a statistical procedure $\psi_{j,i}(\mathcal{H}_n)$ that satisfies (6).

We then define our test procedure

$$\psi(\mathcal{H}_n) := \max_{j\geq 1} \max_{q_{j,i}\in\widetilde{H}_j} \psi_{j,i}(\mathcal{H}_n). \tag{27}$$

We further combine Assumption **H2** and Proposition 2 to obtain for M large enough

$$\begin{aligned} \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] &\leq \sum_{j=1}^{\infty} \sum_{q_{j,i}\in\widetilde{H}_j} \mathbb{E}_0^{(n)}[\psi_{j,i}(\mathcal{H}_n)] \\ &\leq e^{n\epsilon_n^2} \frac{e^{-Kn\epsilon_n^2 M^2}}{1 - e^{-Kn\epsilon_n^2 M^2}} \leq e^{-Kn\epsilon_n^2 M/2^2}, \end{aligned}$$

and

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

5.5 Proof of Theorem 1

Let M be a positive constant. We first decompose the right-hand side of (11) in two parts

$$\begin{aligned} \Pi_n^{\mathcal{H}_n}\big(B^{\complement}_{d_{\nu^*}}(q_0, \epsilon_n M)\big) &= \Pi_n^{\mathcal{H}_n}\big(B^{\complement}_{d_{\nu^*}}(q_0, \epsilon_n M) \cap \mathcal{Q}_n\big) + \Pi_n^{\mathcal{H}_n}\big(B^{\complement}_{d_{\nu^*}}(q_0, \epsilon_n M) \cap \mathcal{Q}_n^{\complement}\big) \\ &=: A_1 + A_2. \end{aligned} \tag{28}$$

In the sequel, each term in the right-hand side of (28) is separately bounded from above: for A_1, we apply Corollory 1, whereas to upper bound A_2 we use **H3** and **H4**.

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

$$\mathcal{L}_n(q) = \widetilde{\boldsymbol{\rho}}(J_0, S_0) \prod_{l=1}^{n} q_{J_{l-1}}(J_l, X_l).$$

Then, A_1 could be written as follows:

$$\begin{aligned}
A_1 &= \frac{\int_{B^{\complement}_{\nu*}(q_0,\epsilon_n M)\cap\mathcal{Q}_n}\mathcal{L}_n(q)d\Pi_n(q)}{\int_{\mathcal{Q}}\mathcal{L}_n(q)d\Pi_n(q)}\\
&= \frac{\int_{B^{\complement}_{\nu*}(q_0,\epsilon_n M)\cap\mathcal{Q}_n}\frac{\mathcal{L}_n(q)}{\mathcal{L}_n(q_0)}d\Pi_n(q)}{\int_{\mathcal{Q}}\frac{\mathcal{L}_n(q)}{\mathcal{L}_n(q_0)}d\Pi_n(q)}\\
&:= \frac{N_n}{D_n}.
\end{aligned}$$

Moreover consider $\mathcal{D}_n$ as the following event:

$$\mathcal{D}_n = \left\{ D_n \leq \frac{e^{-n\epsilon_n^2}}{2}\Pi_n\left(\mathcal{U}(q_0,\epsilon_n)\right)\right\}.$$

By means of the test procedure defined in (27), $\psi(\mathcal{H}_n)$, $\mathbb{E}_0^{(n)}(A_1)$ could be written as follows

$$\begin{aligned}
\mathbb{E}_0^{(n)}(A_1) &= \mathbb{E}_0^{(n)}\Big(\frac{N_n}{D_n}\Big)\\
&\leq \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] + \mathbb{E}_0^{(n)}\Big[(1-\psi(\mathcal{H}_n))\frac{N_n}{D_n}\Big\{\mathbb{1}_{\mathcal{D}_n}+\mathbb{1}_{\mathcal{D}_n^{\complement}}\Big\}\Big]\\
&\leq \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] + \mathbb{E}_0^{(n)}\Big[(1-\psi(\mathcal{H}_n))\frac{N_n}{D_n}\mathbb{1}_{\mathcal{D}_n^{\complement}}\Big] + \mathbb{P}_0^{(n)}\Big(\mathcal{D}_n\Big)\\
&:= T_1+T_2+T_3. \qquad (29)
\end{aligned}$$

To bound from above $\mathbb{E}_0^{(n)}(A_1)$, it is sufficient to upper bound every term in the right-hand side of (29).

- Term T_1. We apply Corollary 1 and obtain that there exists $K>0$ such that

$$T_1 = \mathbb{E}_0^{(n)}[\psi(\mathcal{H}_n)] \leq e^{-Kn\epsilon_n^2M^2}. \qquad (30)$$

- Term T_2. We apply once again Corollary 1, which combined with **H3** entails that there exists $\tilde{K}>0$ such that

$$\begin{aligned}
T_2 &\leq \int_{B^{\complement}_{d_{\nu^*}}(q_0,\epsilon_n M)\cap\mathcal{Q}_n}\mathbb{E}_q^{(n)}[1-\psi(\mathcal{H}_n)]d\Pi_n(q)\frac{2}{e^{-n\epsilon_n^2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)}\\
&\leq \sup_{q\in B^{\complement}_{d_{\nu^*}}(q_0,\epsilon_n M)\cap\mathcal{Q}_n}\mathbb{E}_q^{(n)}[1-\psi(\mathcal{H}_n)]\frac{2}{e^{-n\epsilon_n^2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)}\\
&\leq e^{-\tilde{K}n\epsilon_n^2M^2}\frac{2}{e^{-n\epsilon_n^2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)}\\
&\leq 2e^{-(\tilde{K}M^2-1-c)n\epsilon_n^2} \leq 2e^{-\kappa n\epsilon_n^2}, \qquad (31)
\end{aligned}$$

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

- Term T_3. Consider the following subspace of $\mathcal{Q}$

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

and observe that

$$\begin{aligned} D_n &\geq \int_{\mathcal{U}(q_0,\epsilon_n)\cap\mathcal{V}_n} \exp\Big(\log \frac{\mathcal{L}_n(q)}{\mathcal{L}_n(q_0)} + K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) - K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)})\Big) d\Pi_n(q) \\ &\geq \exp\Big(\frac{-n\epsilon_n^2}{2}\Big)\Pi_n\Big(\mathcal{U}(q_0,\epsilon_n)\cap\mathcal{V}_n\Big). \end{aligned}$$

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

$$\begin{aligned} T_3 &\leq \mathbb{P}_0^{(n)}\Big(e^{\frac{-n\epsilon_n^2}{2}}\Pi_n\Big(\mathcal{U}(q_0,\epsilon_n)\cap\mathcal{V}_n\Big) \leq \frac{e^{-n\epsilon_n^2}}{2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)\Big) \\ &= \mathbb{P}_0^{(n)}\Big(\Pi_n\Big(\mathcal{U}(q_0,\epsilon_n)\cap\mathcal{V}_n^{\complement}\Big) \geq \big(1-\frac{1}{2}e^{\frac{-n\epsilon_n^2}{2}}\big)\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)\Big) \\ &\leq \frac{2}{\big(2-e^{\frac{-n\epsilon_n^2}{2}}\big)\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)}\mathbb{E}_0^{(n)}\Big(\Pi_n\big(\mathcal{V}_n^{\complement}\cap\mathcal{U}(q_0,\epsilon_n)\big)\Big) \\ &\leq \frac{2}{\big(2-e^{\frac{-n\epsilon_n^2}{2}}\big)\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)} \times \int_{\mathcal{U}(q_0,\epsilon_n)} \mathbb{P}_0^{(n)}\Big(\big|\log\frac{\mathcal{L}_n(q_0)}{\mathcal{L}_n(q)} - K(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)})\big| > \frac{n\epsilon_n^2}{2}\Big) d\Pi_n(q) \\ &\leq \frac{2}{\big(2-e^{\frac{-n\epsilon_n^2}{2}}\big)\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)}\int_{\mathcal{U}(q_0,\epsilon_n)} \mathbb{V}_0(\mathbb{P}_0^{(n)}, \mathbb{P}_q^{(n)}) d\Pi_n(q)\frac{4}{n^2\epsilon_n^4} \\ &\leq \frac{8}{n\epsilon_n^2\big(2-e^{\frac{-n\epsilon_n^2}{2}}\big)}. \end{aligned} \tag{32}$$

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

$$A_2 \quad = \quad \frac{\int_{B^{\complement}_{d_{\nu^*}}(q_0,\epsilon_n M)\cap\mathcal{Q}_n^{\complement}} \frac{\mathcal{L}_n(q)}{\mathcal{L}_n(q_0)} d\Pi_n(q)}{\int_{\mathcal{Q}} \frac{\mathcal{L}_n(q)}{\mathcal{L}_n(q_0)} d\Pi_n(q)} := \frac{\widetilde{N}_n}{D_n}.$$

Then, using Equation (32) and from Assumptions **H3** and **H4**, we obtain

$$\begin{aligned} \mathbb{E}_0^{(n)}(A_2) &= \mathbb{E}_0^{(n)}\Big(\frac{\widetilde{N}_n}{D_n}\Big\{\mathbb{I}_{D_n \leq \frac{e^{-n\epsilon_n^2}}{2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)} + \mathbb{I}_{D_n > \frac{e^{-n\epsilon_n^2}}{2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)}\Big\}\Big) \\ &\leq \mathbb{P}_0^{(n)}\Big(\mathcal{D}_n\Big) + \mathbb{E}_0^{(n)}\big(\widetilde{N}_n\big)\frac{2}{e^{-n\epsilon_n^2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)} \\ &\leq \mathbb{P}_0^{(n)}\Big(\mathcal{D}_n\Big) + \Pi_n\big(\mathcal{Q}_n^{\complement}\big)\frac{2}{e^{-n\epsilon_n^2}\Pi_n\big(\mathcal{U}(q_0,\epsilon_n)\big)} \\ &\leq \frac{8}{n\epsilon_n^2\big(2-e^{\frac{-n\epsilon_n^2}{2}}\big)} + 2e^{-(c+1)n\epsilon_n^2}. \end{aligned} \tag{33}$$

Finally, Inequalities (30)–(33) lead to the desired result (11).

Acknowledgments
The authors would like to thank the Fédération de Recherche Mathématiques des Pays de Loire (FR 2962) for their support.

References

[1] Amewou-Atisso, M., Ghosal, S., Ghosh, J.K. & Ramamoorthi, R.V. (2003). Posterior consistency for semiparametric regression problems. *Bernoulli* **9**, 291–312.

[2] Arbel, J., Gayraud, G. & Rousseau, J. (2013). Bayesian optimal adaptive estimation using a sieve prior. *Scand. J. Stat.* **40**, 549–570.

[3] Banerjee, A.K. & Bhattacharyya, G.K. (1976). Testing hypotheses in a two-state semi-Markov process. *Sankhya: The Indian Journal of Statistics*, Series A, **38 (4)** 340–356.

[4] Barbu, V. S. & Limnios, N. (2008). *Semi-Markov chains and hidden semi-Markov models toward applications. Their use in reliability and DNA analysis*, Springer-Verlag, New York.

[5] Bath, B.R. & Deshpande, S.K. (1986). Testing for Markov process VS semi-Markov process. *Commun. Statist.-Theor. Meth.* **15**, 2375–2382.

[6] Birgé, L. (1983a). *Robust testing for independent non identically distributed variables and Markov chains.* Specifying Statistical Models, 134–162. Springer-Verlag, Heidelberg.

[7] Birgé, L. (1983b). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65**, 181–237.

[8] 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* **9**, , 47–64, Institute of Mathematical Statistics.

[9] Bulla, P. & Muliere, P. (2007). Bayesian nonparametric estimation for reinforced Markov renewal processes. *Stat. Inference Stoch. Process.* **10 (3)**, 283–303.

[10] Chang, I.-S., Chuang, Y.-C. & Hsiung, C.A. (1999). A class of nonparametric K-sample tests for semi-Markov counting processes. *Statistica Sinica* **9**, 211–227.

[11] Chang, I.-S., Chuang, Y.-C. & Hsiung, C.A. (2001). Goodness-of-fit tests for semi-Markov and Markov survival models with one intermediate state. *Scand. J. Stat.* **28 (3)**, 505–525.

[12] Choudhuri, N., Ghosal, S. & Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99**, 1050–1059.

[13] Çinlar, E. (1969). Markov renewal theory. *Adv Appl Probab* **1**, 123—187.

[14] Economou, T., Bailey, T.C. & Kapelan, Z. (2014). MCMC implementation for Bayesian hidden semi-Markov models with illustrative applications. *Stat Comput* **24 (5)**, 739–752.

[15] 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.

[16] Gassiat, E. & Rousseau, J. (2013). About the posterior distribution in hidden Markov models with unknown number of states. *Ann Stat.* **35 (1)**, 192–223.

[17] Ghosal, S. & van Der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *Ann Stat.* **35 (1)**, 192–223.

[18] Ghosal, S. & van Der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.

[19] Ghosal, S. & Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann Stat.* **34 (5)**, 2413–2429.

[20] Ghosal, S., Ghosh, J.K. & Ramamoorthi, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann Stat.* **27 (1)**, 143–158.

[21] Ghosal, S., Ghosh, J.K. & van Der Vaart, A. (2000). Convergence rates of posterior distributions. *Ann Stat.* **28 (2)**, 500–531.

[22] Ghosh, J.K. & Ramamoorthi, R.V. (2003). *Bayesian Nonparametrics.* Springer-Verlag, New York.

[23] Korolyuk, V.S. & Limnios, N. (2005). *Stochastic systems in merging phase space*. World Scientific, Singapore.

[24] Le Cam, L.M. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

[25] Le Cam, L.M. (1975). *On local and global properties in the theory of asymptotic normality of experiments*. Stochastic processes and related topics, M.Puri ed. Academic Press.

[26] Le Cam, L.M. (1973). Convergence of estimates under dimensionality restrictions. *Ann Stat.* **1**, 38–53.

[27] Lévy, P. (1954). Processus semi-markoviens. *Proc. Int. Cong. Math. (Amsterdam*, 416–426.

[28] Limnios, N. & Oprişan, G. (2001). *Semi-Markov Processes and Reliability.* Birkhäuser, Boston.

[29] Malinovskii, V.K. (1992). Asymptotic optimality of criteria in the problem of testing hypotheses for a recurrent semi-Markov process. *Journal of Soviet Mathematics* **59 (4)**, 955–959.

[30] Pati, D., Dunsony, D. & Tokdary, S. (2013). Posterior consistency in conditional distribution estimation. *J Multivariate Anal* **116**, 456–472.

[31] Phelan, M.J. (1990). Bayes estimation from a Markov renewal process. *Ann Stat.* **18 (2)**, 603–616.

[32] Pyke, R. (1961a). Markov renewal processes: definitions and preliminary properties. *Ann. Math. Stat.* **32**, 1231—1242.

[33] Pyke, R. (1961b). Markov renewal processes with finitely many states. *Ann. Math. Stat.* **32**, 1243—1259.

[34] 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.

[35] Smith, W.L. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. London, Ser. A.* **232**, 6–31.

[36] Shen, X. & Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann Stat.* **29**, 687–714.

[37] Takács, L. (1954). Some investigations concerning recurrent stochastic processes of a certain type. *Magyar Tud. Akad. Mat. Kutato Int. Közl.* **3**, 115–128.

[38] Tang, Y. & Ghosal, S. (2007). Posterior consistency of Dirichlet mixtures for estimating a transition density. *J. Statist. Plann. Inference* **137**, 1711–1726.

[39] Tsai, W.-Y. (1985). Rank tests for a class of semi-Markov models with censored matched pairs. *Stat Probabil Lett* **3(5)**, 281–286.

[40] Wasserman, L. (1998). *Asymptotic properties of nonparametric Bayesian procedures.* Lecture Notes in Statistics 133, 293–304, Springer, New-York.