Posterior consistency for partially observed Markov models

Randal Douc, Jimmy Olsson, François Roueff

Posterior consistency for partially observed Markov models

Randal Douc * **Jimmy Olsson** † **and François Roueff** ‡

Abstract: In this work we establish the posterior consistency for a parametrized family of partially observed, fully dominated Markov models. As a main assumption, we suppose that the prior distribution assigns positive probability to all neighborhoods of the true parameter, for a distance induced by the expected Kullback-Leibler divergence between the family members' Markov transition densities. This assumption is easily checked in general. In addition, under some additional, mild assumptions we show that the posterior consistency is implied by the consistency of the maximum likelihood estimator. The latter has recently been established also for models with non-compact state space. The result is then extended to possibly non-compact parameter spaces and non-stationary observations. Finally, we check our assumptions on examples including the partially observed Gaussian linear model with correlated noise and a widely used stochastic volatility model.

1. Introduction

We consider a very general framework where a bivariate Markov chain $(Z_n)_{n\in\mathbb{N}}$ taking on values in some product state space $\mathsf{Z} = \mathsf{X} \times \mathsf{Y}$, i.e., $Z_n = (X_n, Y_n)$, is only partially observed through the second component $(Y_n)_{n\in\mathbb{N}}$. In this model, which we refer to as a *partially observed Markov model* (POMM) ([33] uses the alternative term *pairwise Markov chain*), any statistical inference has to be carried through on the basis of the observations $(Y_n)_{n\in\mathbb{N}}$ only, which is generally far from straightforward due to the fact that the observation process $(Y_n)_{n\in\mathbb{N}}$ is, on the contrary to $(Z_n)_{n\in\mathbb{N}}$, generally non-Markovian. Of particular interest are the *hidden Markov models* (HMMs) (alternatively termed *state-space models* in the case where X is continuous), which constitute a special case of the POMMs in which the process $(X_n)_{n\in\mathbb{N}}$ is itself a Markov chain, referred to as the *state process*, and the observations are conditionally independent given the states, such that the marginal conditional distribution of each observation Y_n depends only on the corresponding state X_n. The use of unobservable states provides the HMMs with a relatively simple dependence structure which is still generic enough to handle complex, real-world time series in a large variety of scientific and engineering disciplines (such as financial econometrics [21, 29], speech recognition [23], biology [7], neurophysiology [15], etc.; see also the monographs

*SAMOVAR, CNRS UMR 5157, Institut Télécom/Télécom SudParis, 9 rue Charles Fourier, 91000 Evry.

†KTH Royal Institute of Technology, Stockholm, Sweden. Jimmy Olsson is supported by the Swedish Research Council, Grant 2011-5577.

‡LTCI, CNRS 5141, Institut Télécom/Télécom Paristech, 42 rue Barrault, 75000 Paris.

[28] and [6] for introductive and state of the art treatments of the topic, respectively), and the POMMs can be viewed as a natural extension and generalization of this model class.

In this paper, we will consider a parameterized family of POMMs with parameter space Θ, where the latter is assumed to be furnished with some metric. For each $\theta \in \Theta$, the dynamics of the model is specified by the transition kernel $\mathbf{Q}_\theta$ of $(Z_n)_{n\in\mathbb{N}}$ on $\mathsf{X} \times \mathsf{Y}$, and we will in this work restrict ourselves to the *fully dominated case* where $\mathbf{Q}_\theta$ has a transition density q_θ w.r.t. some dominating measure (all these objects will be defined rigorously in the next section). Each transition kernel $\mathbf{Q}_\theta$ is assumed to have a unique invariant distribution π_θ.

We assume that we have access to a single observation trajectory $(Y_n)_{n\in\mathbb{N}}$ sampled from the canonical law $\mathbb{P}$ induced by $\mathbf{Q}_{\theta_\star}$ and some initial distribution $\eta_\star$ on $\mathsf{X} \times \mathsf{Y}$, where $\theta_\star \in \Theta$ is a distinguished parameter interpreted as the *true* parameter and $\eta_\star$ is generally different from $\pi_{\theta_\star}$. In order to estimate $\theta_\star$ via the observations we adopt a Bayesian framework and introduce a possibly improper *prior* distribution λ on Θ, reflecting our a priori belief concerning θ, and compute the conditional—*posterior*—distribution $\lambda\langle Y_{1:n}\rangle$ of θ given the observations $Y_{1:n} = (Y_1, \ldots, Y_n)$, which is, for measurable $A \subseteq \Theta$ and $y_{1:n} \in \mathsf{Y}^n$, given by

$$\lambda\langle y_{1:n}\rangle(A) = \frac{\int_A p_\theta(y_{1:n})\,\lambda(\mathrm{d}\theta)}{\int_\Theta p_\theta(y_{1:n})\,\lambda(\mathrm{d}\theta)},$$

where $p_\theta(y_{1:n})$ denotes the density of $Y_{1:n}$ given θ. In this general setting, we examine the asymptotics of the posterior distribution and identify model conditions under which the *posterior consistency*

$$\mathbb{P}\left(\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}\right) = 1, \tag{1}$$

holds true, where $\Longrightarrow$ denotes weak convergence and $\delta_{\theta_\star}$ denotes the Dirac mass located at the true parameter $\theta_\star$. In (1), $\mathbb{P}$ denotes the distribution of $(Y_n)_{n\in\mathbb{N}}$ corresponding to the true parameter $\theta_\star$; in this sense we adopt, by proving (1), a frequentist point of view for the asymptotic behavior of the posterior distribution. However, establishing that the influence of the prior is overwhelmed by the data as the sample size n grows to infinity is of fundamental interest in Bayesian analysis.

1.1. Previous work

From the frequentist inference point of view, POMMs have been subjected to extensive research during the last decades. For the important subclass formed by HMMs with finite state space, the asymptotic consistency of the *maximum likelihood estimator* (MLE) was established by [5, 32] and [26] in the cases of finite and general observation spaces, respectively, and these results were generalized gradually to more general HMMs in [10, 13, 18]. The first MLE consistency result for general HMMs with possibly non-compact state space was obtained in [12], and [11] extended further this result to misspecified models. For POMMs that

fall outside the HMM class, [13] established the MLE consistency for autoregressive models with Markov regimes by applying strong mixing assumptions requiring typically the state space of the latent Markov chain to be compact. Recently, [9] established the MLE consistency for general observation-driven time series with possibly non-compact state space and [14] established the analogous result for general partially dominated POMMs, covering the observation-driven models as a special case. The latter work can be viewed as the state of the art when it concerns MLE analysis for POMMs. The mentioned works demonstrate a variety of techniques, but share the assumption that the parameter space Θ is a compact set.

On the other hand, when it concerns Bayesian asymptotic analysis of POMMs, there are only a handful results of which all treat exclusively HMMs. In the case of HMMs with a finite state space, [16] provides the posterior consistency (with rates) for parametric models with an unknown number of states and the recent paper [36] deals with posterior concentration in the non-parametric case. For more general HMMs, [8] establishes, along the now classical lines of [25, Theorem 8.3], a Bernstein-von Mises-type result under the assumption that the model satisfies, first, a law of large numbers for the log-likelihood function, second, a central limit theorem for the score function and, third, a law of large numbers for the observed Fisher information. As these asymptotic properties, which are the cornerstones of the proof of the asymptotic normality of the MLE, can be established for models satisfying the strong mixing assumption (see [13] and [6, Chapter 12]), the result holds, in principle, true for HMMs with a compact state space. A more direct approach to the posterior consistency for HMMs is taken in [31], where the author works with a large deviation bound for the observation process; nevertheless, the analysis is driven by very restrictive model assumptions in terms of strong mixing and additive observation noise.

In conclusion, all available results on posterior concentration for parametric HMMs rest on very stringent model assumptions, in the sense that the state space of the unobservation process is assumed to be compact. Needless to say, this is not the case for many models met in practical applications (such as the linear Gaussian state-space models). In addition, the mentioned results require, without exception, also the parameter space to be compact, which is generally a severe restriction for the Bayesian. Consequently, a general posterior consistency result for parametric POMMs (and, in particular, parametric HMMs) has hitherto been lacking. In the light of the widespread and ever increasing interest in Bayesian inference in models of this sort, which is boosted by novel achievements in computational statistics (especially in the form of *particle* [20] and *particle Markov chain Monte Carlo methods* [1]), this is indeed remarkable, and the goal of the present paper is to fill this gap.

1.2. Our approach

In this paper, we establish the posterior consistency (1) under very mild assumptions which can be checked for a large class of POMMs used in practice.

The result is stated in Theorem 3 for general POMMs with positive transition density (Theorem 1 deals with the case of a compact parameter space) and in Theorem 2 for HMMs under an alternative set of assumptions requiring, e.g., only the emission density to be positive. The starting point of our analysis is the observation that it is, by the Portmanteau lemma, enough to show that for all $A_p = \{\theta \in \Theta : d(\theta, \theta_\star) \geq 1/p\}$, $p \in \mathbb{N}^*$,

$$\limsup_{n\to\infty} \lambda\langle Y_{1:n}\rangle(A_p) = 0, \quad \mathbb{P}\text{-a.s.} \tag{2}$$

(see Remark 10 for details). Now, by expressing the posterior as

$$\lambda\langle Y_{1:n}\rangle(A) = \frac{\int_A p_\theta(Y_{1:n})/p_{\theta_\star}(Y_{1:n})\,\lambda(\mathrm{d}\theta)}{\int_\Theta p_\theta(Y_{1:n})/p_{\theta_\star}(Y_{1:n})\,\lambda(\mathrm{d}\theta)}, \tag{3}$$

we conclude that (2) will hold if

- all closed sets A not containing $\theta_\star$ are $\mathbb{P}$-*remote* from $\theta_\star$ in the sense that the numerator of (3) tends to zero exponentially fast under $\mathbb{P}$;
- for all $\delta > 0$, there exists some subset Θ_δ of Θ which is charged by the prior, i.e., $\lambda(\Theta_\delta) > 0$, and such that for all $\theta \in \Theta_\delta$, the ratio $p_\theta(Y_{1:n})/p_{\theta_\star}(Y_{1:n})$ is, $\mathbb{P}$-a.s., eventually bounded from below by $\mathrm{e}^{-\delta n}$. This *asymptotic merging property* forces the numerator of (3) to vanish at a faster rate than the denominator for all $\mathbb{P}$-remote sets A, implying (2).

This machinery, which is adopted from [4] and described generally in Section 4.1, does not require the model under consideration to be a POMM; it is hence of independent interest. As we will see, the situation of a non-compact parameter space calls for a refined notion of $\mathbb{P}$-remoteness; indeed, by operating under the assumption that the sequence of posterior distributions is *tight*, it is enough to require $\mathbb{P}$-remoteness to hold on a sufficiently large compact subset of Θ.

The $\mathbb{P}$-remoteness, the asymptotic merging property and the tightness of the posterior are the fundamental building blocks of our analysis of the posterior concentration. Interestingly, a key finding of us is that the $\mathbb{P}$-remoteness is closely related to the MLE consistency; more specifically, in Proposition 9 we establish that if all sequences of *approximate MLEs* (see Definition 11) on some compact subset K of Θ (with $\lambda(K) < \infty$) containing $\theta_\star$ are strongly consistent, then $A \cap K$ is $\mathbb{P}$-remote for all closed sets A not containing $\theta_\star$. As mentioned in the literature review above, the MLE can, under the assumption that the parameter space is compact, be proven to be consistent under very mild model assumptions satisfied for most fully dominated POMMs, and we will hence obtain the remoteness for free for a large set of models.

When it concerns the asymptotic merging property, we derive the instrumental bound

$$\liminf_{n\to\infty} n^{-1} \log \frac{p_\theta(Y_{1:n})}{p_{\theta_\star}(Y_{1:n})} \geq \liminf_{n\to\infty} n^{-1} \log \frac{\bar{p}_\theta(Z_{1:n})}{\bar{p}_{\theta_\star}(Z_{1:n})}, \tag{4}$$

where $\bar{p}_\theta(z_{1:n})$, $z_{1:n} \in \mathsf{Z}^n$, denotes the density of the *complete data* $Z_{1:n}$ given θ (see Lemma 8 for a more general formulation). In the stationary mode, i.e., when $\eta_\star = \pi_{\theta_\star}$, the right hand side of (4) tends, by Birkhoff's ergodic theorem, to minus the expectation $\Delta(\theta_\star, \theta)$ of the Kullback-Leibler divergence (KLD) between $\mathbf{Q}_{\theta_\star}$ and $\mathbf{Q}_\theta$ under the stationary distribution. As a consequence, the asymptotic merging property holds true as long as the prior is *information dense* at $\theta_\star$ in the sense that $\lambda(\{\theta \in \Theta : \Delta(\theta_\star, \theta) \geq \delta\}) > 0$ for all $\delta > 0$. This condition can however be checked straightforwardly in general, since $\Delta(\theta_\star, \theta)$ involves a KLD between perfectly known *transition kernels*. Without access to the bound (4), an alternative strategy would have been to study directly the limit of the left hand side of (4) by, e.g., going to "the infinite past" in the spirit of [13, 11]; however, this approach would require the analysis of an expected KLD between *ergodic limits* $p_\theta(Y_0 \mid Y_{-\infty:-1})$ and $p_{\theta_\star}(Y_0 \mid Y_{-\infty:-1})$ (we refer to the mentioned works for the meaning of these quantities) under the stationary distribution, which is infeasible in general.

As described above, our technique of handling models with a non-compact parameter space is based on the assumption that the sequence of posterior distributions is tight. Recalling that our objective is the establishment of the gradual concentration of these distributions around the true parameter as n increases, we may expect this assumption to be mild. Indeed, by operating in the stationary mode using Kingman's subadditive theorem, we are able to derive handy assumptions under which the posterior tightness holds at an exponential rate (see Theorem 3 and Proposition 18). As far as is known to us, this is the first result ever in this direction for models of this sort, and we believe that a similar approach may be used also for extending existing results on MLE consistency to the setting of a non-compact parameter space.

We remark that Birkhoff's ergodic theorem and Kingman's subadditive theorem require the observation process to be stationary (i.e. $\eta_\star = \pi_{\theta_\star}$). Nevertheless, if for all parameters θ and initial distributions η, the distribution of $Y_{1:\infty}$, when initialized according to η and evolving according to θ, admits a positive density w.r.t. the distribution of $Y_{1:\infty}$ under the stationary distribution π_θ, one may prove that any property that holds a.s. under the latter distribution holds a.s. under the former as well. That such positive densities exist can be established for POMMs in general under the assumption that the transition density is positive (Lemma 12) and for HMMs in particular under the weaker assumption that the emission density is positive and the hidden chain is geometrically ergodic (Lemma 14), and we are consequently able to treat also the non-stationary case. As far as is known to us, this efficient approach to non-stationarity results has never been taken before.

Finally, we demonstrate the flexibility of our results by checking carefully our assumptions on a partially observed linear Gaussian Markov model as well as the widely used stochastic volatility model proposed by [21] (the latter falls into the framework of HMMs).

To sum up, our contribution is fourfold, since we

– establish the posterior consistency for very general POMMs under mild

assumptions, which allow the state space of the latent part of the model to be non-compact and which can be checked for a large number of models used in practice.
- link, via the concept of $\mathbb{P}$-remoteness, the posterior consistency to the consistency of the MLE.
- are able to treat also the case of a non-compact parameter space.
- treat efficiently the case of non-stationary observations.

The paper is structured as follows. In Section 2, we introduce the POMM framework under consideration and state, in Section 2.2, our main results (Theorems 1–3) and assumptions. In particular, we provide an alternative set of assumptions that are taylor-made for the special case of HMMs. Section 3 treats the two examples mentioned previously and discusses generally our assumptions in the light of nonlinear state-space models. In Section 4 we embed the problem of posterior concentration into the general framework outlined above, serving as a machinery for the proofs of our main results. The latter proofs are found in Section 5 and in Section 6 we conclude the paper.

2. Fully dominated partially observed Markov models (fdPOMM)

2.1. Setting

Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be general measurable spaces referred to as *state space* and *observation space*, respectively. The product space $\mathsf{Z} = \mathsf{X} \times \mathsf{Y}$ is then endowed with the product σ-field $\mathcal{Z} = \mathcal{X} \otimes \mathcal{Y}$, and we set $\Omega = \mathsf{Z}^{\mathbb{N}}$ and $\mathcal{F} = \mathcal{Z}^{\otimes \mathbb{N}}$. Let further $(Z_n)_{n\in\mathbb{N}}$, $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ denote the canonical processes taking on values in the spaces $(\mathsf{Z}, \mathcal{Z})$, $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$, respectively, and defined by $Z_n(\omega) = (x_n, y_n)$, $X_n(\omega) = x_n$ and $Y_n(\omega) = y_n$, where $\omega = ((x_k, y_k))_{k\in\mathbb{N}}$. Now, for all $n \in \mathbb{N}^*$, only $(Y_n)_{n\in\mathbb{N}^*}$ is observable, and for this reason we refer to the model as *partially observed*. Let us define $\mathcal{F}_n = \sigma(Y_{1:n})$ for all $n \in \mathbb{N}$, with $Y_{1:n} = (Y_1, \ldots, Y_n)$ serving as our general notation for vectors. In addition, let $(\Theta, \mathcal{T})$ be a measurable space and let $\{\mathbf{Q}_\theta, \theta \in \Theta\}$ be a collection of Markov transition kernels on $(\mathsf{Z}, \mathcal{Z})$. Denote, for all $\theta \in \Theta$, by $\mathbb{P}_\eta^\theta$ the law of the canonical Markov chain $(Z_n)_{n\in\mathbb{N}}$ induced by the Markov transition kernel $\mathbf{Q}_\theta$ and the initial distribution η. We say that the model is *fully dominated* if there exist two σ-finite measures μ and ν on $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$, respectively, such that for all $\theta \in \Theta$ and $z \in \mathsf{Z}$, the probability measure $\mathbf{Q}_\theta(z, \cdot)$ is dominated by $\mu \otimes \nu$, and we denote by $q_\theta(z, \cdot)$ the corresponding density. We may now introduce the main class of models studied in this article.

Definition 1 *We say that $(Y_n)_{n\in\mathbb{N}^*}$ follows a* fully dominated partially observed Markov model *(fdPOMM) if, for all $n \in \mathbb{N}^*$, under the previous definitions and assumptions, the distribution of $Y_{1:n}$ is given by $\mathbb{P}_\eta^\theta(Y_{1:n} \in \cdot)$ for some $\theta \in \Theta$ and initial distribution η on Z.*

We denote, with a slight abuse of notation, by $y_{1:n} \mapsto p_{\theta,\eta}(y_{1:n})$ the density of $Y_{1:n}$ with respect to $\nu^{\otimes n}$ under $\mathbb{P}_\eta^\theta$, i.e.,

$$p_{\theta,\eta}(y_{1:n}) \coloneqq \int q_\theta(z_0,(x_1,y_1)) \prod_{\ell=1}^{n-1} q_\theta((x_\ell,y_\ell),(x_{\ell+1},y_{\ell+1}))\, \eta(\mathrm{d}z_0)\, \mu^{\otimes n}(\mathrm{d}x_{1:n}).$$

Now, let λ be some measure, called the *prior distribution*, on $(\Theta,\mathcal{T})$. We will always assume that Θ is endowed with some metric d and that $\mathcal{T}$ is taken to be the corresponding Borel σ-field.

Given observations $y_{1:n} \in \mathsf{Y}^n$, the *posterior distribution* associated with the initial probability distribution η is defined by

$$\lambda\langle y_{1:n}\rangle(A) \coloneqq \frac{\int_A p_{\theta,\eta}(y_{1:n})\, \lambda(\mathrm{d}\theta)}{\int_\Theta p_{\theta,\eta}(y_{1:n})\, \lambda(\mathrm{d}\theta)}, \quad \text{for all } A \in \mathcal{T}. \tag{5}$$

For the numerator and denominator of (5) to be well-defined, we will always assume that $(\theta,z,z') \mapsto q_\theta(z,z')$ is measurable on $\Theta \times \mathsf{Z}^2$. However, at this point it is not guaranteed that the ratio itself is well-defined (and does not degenerate into $0/0$ or ∞/∞). In fact, we will only be interested in the case where $\lambda\langle Y_{1:n}\rangle$ is $\mathbb{P}$-a.s. a probability distribution for n large enough, where $\mathbb{P}$ denotes the *true distribution* of $(Y_n)_{n\in\mathbb{N}}$ and is introduced below.

We always assume the following.

(B1) For all $\theta \in \Theta$, the Markov transition kernel $\mathbf{Q}_\theta$ has a unique stationary distribution π_θ.

Under (**B1**) it is typically assumed that the law of the observations is given by $\mathbb{P} = \mathbb{P}_{\pi_{\theta_\star}}^{\theta_\star}$ for some distinguished parameter $\theta_\star \in \Theta$ interpreted as the *true* parameter (which is not known a priori). We proceed similarly and set $\pi_\star = \pi_{\theta_\star}$. However, in the present paper we will also consider the more general case where $\mathbb{P} = \mathbb{P}_{\eta_\star}^{\theta_\star}$ for some possibly unknown initial distribution $\eta_\star \neq \pi_\star$, and since the initial distribution η appearing in (5) is designed arbitrarily by the user, we cannot assume that $\eta = \eta_\star$. See also Remark 13 for further comments concerning this.

Remark 2 *Under (**B1**), since $\mathbf{Q}_\theta$ is dominated by $\mu \otimes \nu$, the stationary probability measure π_θ is also dominated by $\mu \otimes \nu$, and by abuse of notation, we still denote by π_θ the associated density.*

2.2. Main results

We now state the main results of this contribution, which consist in providing general sufficient conditions for the posterior consistency

$$\mathbb{P}\left(\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}\right) = 1,$$

where $\Longrightarrow$ denotes weak convergence and δ_θ denotes a Dirac point mass located at θ. The proof of this result is based on basically two main ingredients. The

first is to ensure that only parameters θ close to $\theta_\star$ have a large likelihood as the number n of observations tends to infinity. This is formalized by the following assumption.

(B2) If $K \in \mathcal{T}$ is a compact set containing $\theta_\star$, then all K-valued, $(\mathcal{F}_n)_{n\in\mathbb{N}}$-adapted random sequences $(\hat{\theta}_n)_{n\in\mathbb{N}^*}$ such that for all $n \in \mathbb{N}^*$,

$$n^{-1} \log p_{\hat{\theta}_n,\eta}(Y_{1:n}) \geq n^{-1} \log p_{\theta_\star,\pi_\star}(Y_{1:n}) + \epsilon_n,$$

with

$$\lim_{n\to\infty} \epsilon_n = 0, \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.},$$

converges $\mathbb{P}^{\theta_\star}_{\pi_\star}$-a.s. to $\theta_\star$.

For identifiable models, this assumption follows directly from standard consistency results on the maximum likelihood estimator for ergodic partially observed Markov chains; see [14] and the references therein. In other words, it does not require a specific treatment from a Bayesian point of view.

Remark 3 *Note that in the case where Θ is compact, (**B2**) only needs to be checked for $K = \Theta$.*

The second ingredient is to ensure that the prior distribution does not concentrate around parameters whose likelihood is too small asymptotically. We provide below sufficient conditions that are easily checked in two specific situations.

The fully dominated case. For all $\theta \in \Theta$ we set

$$\Delta(\theta_\star, \theta) := \mathbb{E}^{\theta_\star}_{\pi_\star} \left[\mathrm{KL}\left(\mathbf{Q}_{\theta_\star}(Z_0, \cdot) \| \mathbf{Q}_\theta(Z_0, \cdot) \right) \right] \in [0, \infty], \tag{6}$$

where for any two probability measures P and Q defined on the same probability space, $\mathrm{KL}(P\|Q)$ denotes the KLD of Q from P defined by

$$\mathrm{KL}(P\|Q) := \begin{cases} \int \log \frac{\mathrm{d}P}{\mathrm{d}Q} \, \mathrm{d}P, & \text{if } P \ll Q, \\ \infty, & \text{otherwise.} \end{cases}$$

Theorem 1 *Consider a fdPOMM satisfying (**B1**–**2**) with a compact parameter space Θ. Let λ be a finite measure on Θ and η an arbitrary distribution on $(\mathsf{Z}, \mathcal{Z})$. Then the conditions*

(B3) for all $\theta \in \Theta$, $q_\theta(z, z') > 0$ $\mu \otimes \nu$-a.s.,

(B4) for all $\delta > 0$, $\lambda(\{\theta \in \Theta : \Delta(\theta_\star, \theta) \leq \delta\}) > 0$,

imply that for all initial distributions $\eta_\star$ on $(\mathsf{Z}, \mathcal{Z})$,

$$\mathbb{P}^{\theta_\star}_{\eta_\star}\left(\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star} \right) = 1.$$

The proof of this result can be found in Section 5.2.

An alternative set of conditions for HMMs. The HMMs can be viewed as a subclass of the fdPOMMs defined by the following assumption.

(C1) For all $\theta \in \Theta$, $z = (x, y) \in \mathsf{Z}$ and $z = (x', y') \in \mathsf{Z}$,

$$q_\theta(z, z') = k_\theta(x, x') g_\theta(x', y'),$$

where k_θ and g_θ are kernel densities on $\mathsf{X} \times \mathsf{X}$ and $\mathsf{X} \times \mathsf{Y}$, respectively.

Under (**C1**) we denote by $\mathbf{K}_\theta$ the Markov transition kernel on $(\mathsf{X}, \mathcal{X})$ associated with the transition density k_θ. In this subclass of models it may happen that the positiveness condition (**B3**) does not hold, but only since k_θ vanishes. In this case, we rely on the following weaker assumption.

(C2) For all $\theta \in \Theta$ and $x \in \mathsf{X}$, $g_\theta(x, \cdot) > 0$ ν-a.s.

In this context, we define

$$\overline{\Delta}(\theta_\star, \theta) \coloneqq \int \mathrm{KL}\left(g_{\theta_\star}(x, \cdot) \| g_\theta(x', \cdot)\right) \pi_\star(\mathrm{d}x \times \mathsf{Y}) \, \pi_\theta(\mathrm{d}x' \times \mathsf{Y}) \tag{7}$$

and consider the following assumption replacing (**B4**).

(C3) For all $\delta > 0$, $\lambda(\{\theta \in \Theta : \overline{\Delta}(\theta_\star, \theta) \leq \delta\}) > 0$.

Finally, we impose the following condition.

(C4) For all $\theta \in \Theta$ and all initial distributions η on $(\mathsf{X}, \mathcal{X})$, $(\eta \mathbf{K}_\theta^n)_{n \in \mathbb{N}^*}$ converges to the first marginal of π_θ in the total variation norm, i.e.,

$$\lim_{n \to \infty} \|\eta \mathbf{K}_\theta^n - \pi_\theta(\cdot \times \mathsf{Y})\|_{\mathrm{TV}} = 0.$$

The kernel notation in (**C4**) is standard and described in detail in Section A.

We can now state the following result, whose proof can be found in Section 5.3.

Theorem 2 *Theorem 1 holds still true when (**B3–4**) are replaced by (**C1–4**).*

Remark 4 *It is interesting to note that in the case of i.i.d. observations, which corresponds to the HMM case (**C1**) with $k_\theta(x, x')$ arbitrary (say equal to 1 with μ an arbitrary probability measure) and $g_\theta(x', y')$ not depending on x', simply denoted by $g_\theta(y')$ hereafter, we have*

$$\Delta(\theta_\star, \theta) = \overline{\Delta}(\theta_\star, \theta) = \mathrm{KL}\left(g_{\theta_\star} \| g_\theta\right) .$$

*Hence (**B4**) and (**C3**) boil down to the well known condition of the i.i.d. setting introduced by [34], see also [19, Eqn. (1)].*

Non-compact parameter space. Needless to say, a drawback of Theorems 1 and 2 is that Θ is assumed to be compact. This assumption is standard in the frequentist setting, e.g., when studying the maximum likelihood estimator, but can be problematic in the Bayesian setting, where it is often convenient to work

with priors defined on non-compact spaces for computational reasons. We now derive some additional conditions dealing with the non-compact case.

Define for all $A \in \mathcal{T}$, $n \in \mathbb{N}^*$ and $y_{0:n} \in \mathsf{Y}^{n+1}$,

$$\hat{p}_A(y_{0:n}) \coloneqq \sup_{(\theta,x_0)\in A\times\mathsf{X}} \int \prod_{\ell=0}^{n-1} q_\theta(z_\ell, z_{\ell+1})\, \mu^{\otimes n}(\mathrm{d}x_{1:n}) \tag{8}$$

(with $z_\ell = (x_\ell, y_\ell)$). Now the following result holds true also for a possibly non-compact Θ.

Theorem 3 *Consider a fdPOMM satisfying (**B1**) and (**B2**). Let λ be a Radon measure on Θ and η an arbitrary distribution on $(\mathsf{Z}, \mathcal{Z})$. In addition, suppose that the following conditions hold true.*

(B5) There exist $\ell \in \mathbb{N}^*$ and a non-decreasing sequence $(C_m)_{m\in\mathbb{N}}$ of compact sets in $\mathcal{T}$ such that

$$\limsup_{m\to\infty} \hat{p}_{C_m^c}(Y_{0:\ell}) = 0 \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.}, \tag{9}$$

$$\mathbb{E}^{\theta_\star}_{\pi_\star}[\log^+ \hat{p}_\Theta(Y_{0:\ell})] < \infty. \tag{10}$$

(B6) There exists $n_0 \in \mathbb{N}^*$ such that

$$\int \lambda(\mathrm{d}\theta)\, p_{\theta,\eta}(Y_{1:n_0}) < \infty \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.}, \tag{11}$$

$$\mathbb{E}^{\theta_\star}_{\pi_\star}[\log p_{\theta_\star,\pi_\star}(Y_{n_0} \mid Y_{1:n_0-1})] > -\infty. \tag{12}$$

*Then (**B3**–**4**) or (**C1**–**4**) imply that for all initial distributions $\eta_\star$ on $(\mathsf{Z}, \mathcal{Z})$,*

$$\mathbb{P}^{\theta_\star}_{\eta_\star}\left(\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}\right) = 1.$$

The proof is postponed to Section 5.4. In (12), $p_{\theta_\star,\pi_\star}(Y_{n_0} \mid Y_{1:n_0-1})$ is, as usual, defined as the ratio $p_{\theta_\star,\pi_\star}(Y_{1:n_0})/p_{\theta_\star,\pi_\star}(Y_{1:n_0-1})$, with the convention that the denominator is unity if $n_0 = 1$. Note that this ratio is always well defined under (**B3**) or (**C1**–**2**).

Remark 5 *It is interesting to observe that, as detailed in Lemma 16, each condition in (**B6**) implies the same condition with n_0 replaced by any $n \in \mathbb{N}$ larger than n_0. It is therefore sufficient to check the conditions independently with two possibly different n_0. The fact that (11) holds for all $n \geq n_0$ is of particular interest, since it guaranties that both the numerator and denominator in the definition of the posterior $\lambda\langle Y_{1:n}\rangle$ in (5) are finite. On the other hand, by (**B3**) or (**C1**–**2**) the denominator is positive. Hence, if (11) holds, the posterior $\lambda\langle Y_{1:n}\rangle$ is well defined as a probability distribution for n large enough.*

3. Examples

3.1. Partially observed Gaussian linear Markov model

First, we consider a linear Gaussian fdPOMM defined on $\mathsf{Z} = \mathbb{R}^p \times \mathbb{R}^q$ by

$$Z_{k+1} = \Phi_\theta Z_k + \epsilon_{k+1}, \tag{13}$$

where Φ_θ is $(p+q)\times(p+q)$ matrix and $(\epsilon_n)_{n\in\mathbb{N}^*}$ is a sequence of i.i.d. centered Gaussian vectors with $(p+q)\times(p+q)$ covariance matrix R_θ. In the following we assume that Θ is a compact subset of $\mathbb{R}^d$ and that for all $\theta \in \Theta$, Φ_θ has spectral radius strictly less than unity and R_θ is positive definite. Then $(Z_n)_{n\in\mathbb{N}}$ is a vector auto-regressive process with transition density

$$q_\theta(z_0, z_1) = (2\pi)^{-(p+q)/2}(\det(R_\theta))^{-1} \exp\left(-\frac{1}{2}(z_1 - \Phi_\theta z_0)^\intercal R_\theta^{-1}(z_1 - \Phi_\theta z_0)\right),$$

with $(z_0, z_1) \in \mathsf{Z}^2$, satisfying (**B1**). This framework includes the widespread *linear Gaussian state-space model*

$$\begin{aligned} X_{k+1} &= A_\theta X_k + \zeta_{k+1}, \\ Y_k &= B_\theta X_k + \xi_k, \end{aligned} \qquad k \in \mathbb{N}, \tag{14}$$

corresponding to

$$\Phi_\theta = \begin{bmatrix} A_\theta & 0 \\ B_\theta A_\theta & 0 \end{bmatrix} \quad \text{and} \quad \epsilon_k = \begin{bmatrix} \zeta_k \\ B_\theta \zeta_k + \xi_k \end{bmatrix}.$$

Note that the model (14) is an HMM only if ζ_k and ξ_k are uncorrelated for all k, which is not assumed in the model (13). Assumption (**B3**) is trivially satisfied. The expected KLD in (6) is easily computed; indeed, for all $\theta_\star = (\Phi_*, R_*)$ and $\theta = (\Phi, R)$ in Θ,

$$\begin{aligned} \Delta(\theta_\star, \theta) = \frac{1}{2}\big[&\mathrm{tr}(R^{-1}R_*) - p - q - \log\det(R^{-1}R_*) \\ &+ \mathrm{tr}(R^{-1}(\Phi - \Phi_*)^\intercal(\Phi - \Phi_*)\Gamma_*)\big], \end{aligned}$$

where

$$\Gamma_* := \mathbb{E}_{\pi_\star}^{\theta_\star}\left[Z_0 Z_0^\intercal\right] = \sum_{k=0}^{\infty} \Phi_*^k R (\Phi_*^\intercal)^k.$$

It follows that $\theta \mapsto \Delta(\theta_\star, \theta)$ is continuous at $\theta_\star$ (where it always vanishes) and thus that (**B4**) is satisfied whenever λ is strictly positive on Θ (i.e., $\lambda(A) > 0$ for every non-empty open set A). Theorem 1 hence applies as soon as Assumption (**B2**) holds true (examples are treated in, e.g., [18] or [12, Section 3.3]).

3.2. General HMMs

In this section, we consider the case of HMMs, i.e., we assume (**C1**). Up to our knowledge, the following set of assumptions, which are borrowed from [12, Theorem 1], are the weakest available for obtaining the strong consistency of the (approximate) MLE for well-specified HMMs.

(D1) For all $\theta \in \Theta$, the Markov kernel $\mathbf{K}_\theta$ is aperiodic positive Harris recurrent.

Note that under (**C1**), this assumption is sufficient for (**B1**), i.e., the existence of a unique stationary probability measure for each complete chain kernel $\mathbf{Q}_\theta$, $\theta \in \Theta$.

(D2) $\mathbb{E}^{\theta_\star}_{\pi_\star}[\sup_{x\in\mathsf{X}} (\log g_{\theta_\star}(x, Y_0))^+] < \infty$, $\mathbb{E}^{\theta_\star}_{\pi_\star}\left[\left|\log \int g_{\theta_\star}(x, Y_0)\, \pi_{\theta_\star}(\mathrm{d}x)\right|\right] < \infty$.

(D3) For all $\theta \neq \theta_\star$, there are a neighborhood $\mathcal{U}_\theta$ of θ such that

$$\sup_{\theta'\in\mathcal{U}_\theta} \sup_{(x,x')\in\mathsf{X}^2} k_{\theta'}(x, x') < \infty, \qquad \mathbb{E}^{\theta_\star}_{\pi_\star}\left[\sup_{\theta'\in\mathcal{U}_\theta} \sup_{x\in\mathsf{X}} (\log g_{\theta'}(x, Y_0))^+\right] < \infty$$

and an integer r_θ such that

$$\mathbb{E}^{\theta_\star}_{\pi_\star}\left[\sup_{\theta'\in\mathcal{U}_\theta} (\log p_{\theta',\eta}(Y_{1:r_\theta}))^+\right] < \infty.$$

(D4) For all $\theta \neq \theta_\star$ and $n \geq r_\theta$, the function $\theta' \mapsto p_{\theta',\eta}(Y_{1:n})$ is upper-semicontinuous at θ, $\mathbb{P}^{\theta_\star}_{\pi_\star}$-a.s.

(D5) For all $\theta \neq \theta_\star$ such that $p_{\theta,\eta}(Y_{1:r_\theta}) > 0$ $\mathbb{P}^{\theta_\star}_{\pi_\star}$-a.s., we have

$$\liminf_{n\to\infty} \mathbb{P}^{\theta_\star}_{\pi_\star}(Y_{1:n} \in A_n) > 0, \qquad \limsup_{n\to\infty} n^{-1} \log \mathbb{P}^{\theta}_{\eta}(Y_{1:n} \in A_n) < 0$$

for some sequence of sets $A_n \in \mathcal{Y}^{\otimes(n+1)}$.

As a consequence of Theorems 1 and 2, the only requirement on the prior λ in the case of general HMMs with a compact parameter space is given by (**B4**) or (**C3**), depending on the positivity assumption on the kernel densities ((**B3**) or (**C2**)).

Theorem 4 *Assume (**C1**) and (**D1–5**) with a compact parameter space Θ. If (**B3–4**) or (**C2–3**) hold, then, for all initial distributions $\eta_\star$ on $(\mathsf{Z}, \mathcal{Z})$,*

$$\mathbb{P}^{\theta_\star}_{\eta_\star}\left(\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}\right) = 1.$$

Proof. Assumption (**D1**) implies (**B1**) and (**C4**) (see [30, Theorem 13.0.1]). Moreover, the strong consistency of the AMLE has been proved in [12, Theorem 1] under (**D1–5**), showing that (**B2**) is satisfied. Thus, all the assumptions of Theorem 1 and Theorem 2 are satisfied and the proof follows. □

Thus, for HMMs, one may, in order to apply Theorem 4, choose to check (**B3**–**4**) or (**C2**–**3**), depending on the model. Consider for example the nonlinear state-space model on $\mathsf{X} = \mathbb{R}^p$,

$$X_{k+1} = T_\theta(X_k) + \Sigma_\theta(X_k)\zeta_{k+1}, \quad k \in \mathbb{N}, \tag{15}$$

where θ is an m-dimensional parameter on a compact space $\Theta \subset \mathbb{R}^m$, $(\zeta_k)_{k\in\mathbb{N}^*}$ is an i.i.d. sequence of d-dimensional random vectors with density ρ_ζ with respect to Lebesgue measure λ^{Leb} and $T_\theta : \mathbb{R}^p \to \mathbb{R}^p$ and $\Sigma_\theta : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ are given (measurable) matrix-valued functions. Conditions on T_θ, σ_θ, ρ_θ and g_θ to ensure (**D1**–**5**) can be found in [12, Section 3.3]. These conditions are stated in the case where ρ_ζ is positive over $\mathbb{R}^d$ but can be easily adapted to the case where the support of ρ_ζ is compact. However, in the latter case, (**B3**) does not hold, and assuming (**C2**), we can rely on (**C3**) as an alternative for (**B4**). In what follows, we explain how to deal with π_θ appearing in the definition (7) of $\overline{\Delta}(\theta_\star, \theta)$ using standard conditions.

Assume that there exist a measurable function $V : \mathsf{X} \to [1, \infty)$ and constants $(C, \rho) \in \mathbb{R}_+^* \times (0, 1)$ such that for all $n \in \mathbb{N}$ and all $(x, x') \in \mathsf{X}^2$,

$$\sup_{\theta\in\Theta} \|\mathbf{K}_\theta^n(x, \cdot) - \mathbf{K}_\theta^n(x', \cdot)\|_V \le C(V(x) + V(x'))\rho^n, \tag{16}$$

where for any signed measure χ on $(\mathsf{X}, \mathcal{X})$, $\|\chi\|_V := \sup \chi f$, where the supremum is taken over all measurable functions $f : \mathsf{X} \to \mathbb{R}$ such that $|f|_V := \mu - \operatorname{esssup}_{x\in\mathsf{X}}[|f(x)|/V(x)] \le 1$ and χf denotes the integral of f w.r.t. χ (see Section A for details).

Proposition 5 *Assume that* (16) *holds. Moreover, suppose that*

(i) for all $\theta_\star \in \Theta$, there exists $C_\star > 0$ such that, for μ-a.e. $x \in \mathsf{X}$ and μ-a.e. $x' \in \mathsf{X}$, $\theta \mapsto \mathrm{KL}\left(g_{\theta_\star}(x, \cdot)\|g_\theta(x', \cdot)\right)$ is continuous at $\theta_\star$ and

$$\sup_{\theta\in\Theta} \mathrm{KL}\left(g_{\theta_\star}(x, \cdot)\|g_\theta(x', \cdot)\right) \le C_\star\, V(x)V(x');$$

(ii) there exists a constant $M < \infty$ such that for μ-a.e. $x \in \mathsf{X}$, $\sup_{\theta\in\Theta} \mathbf{K}_\theta V(x) \le MV(x)$;

(iii) for μ-a.e. $x \in \mathsf{X}$, $\lim_{\theta\to\theta_\star} \|\mathbf{K}_\theta(x, \cdot) - \mathbf{K}_{\theta_\star}(x, \cdot)\|_V = 0$.

*Then, (**C3**) is satisfied for all strictly positive prior measures λ on $(\Theta, \mathcal{T})$.*

Proof. Let $\theta_\star \in \Theta$. It is sufficient to show that the function $\theta \mapsto \overline{\Delta}(\theta_\star, \theta)$ is continuous at $\theta_\star$, where it takes on the value zero. For all $\theta \in \Theta$, denote by π_θ^X the marginal probability measure on $(\mathsf{X}, \mathcal{X})$ defined by $\pi_\theta^X(A) = \pi_\theta(A \times \mathsf{Y})$ for all $A \in \mathcal{X}$, and let the function $\phi_\theta : \mathsf{X} \to \mathbb{R}_+^*$ be defined by

$$\phi_\theta(x') := \int \mathrm{KL}\left(g_{\theta_\star}(x, \cdot)\|g_\theta(x', \cdot)\right) \pi_{\theta_\star}^X(\mathrm{d}x).$$

We may then write

$$\overline{\Delta}(\theta_\star, \theta) = [\pi_\theta^X \phi_\theta - \pi_{\theta_\star}^X \phi_\theta] + \pi_{\theta_\star}^X \phi_\theta \tag{17}$$

(where, as usual, $\pi_\theta^X \phi_\theta$ denotes the integral of ϕ_θ w.r.t. π_θ^X; see Section A). We proceed stepwise.

Step 1. We first show that $\pi_{\theta_\star}^X V < \infty$. Denoting by $\mathcal{M}_V$ the Banach space of signed measures μ such that $|\mu|V < \infty$, equipped with the V-norm, we will actually prove the following more precise assertion.

(iv) For all $x \in \mathsf{X}$, $\mathbf{K}_\theta^n(x,\cdot)$ converges to π_θ^X in $\mathcal{M}_V$, uniformly over $\theta \in \Theta$.

(Hence $\pi_\theta^X V < \infty$ for all $\theta \in \Theta$.) For all probability measures μ_1, μ_2 on $(\mathsf{X}, \mathcal{X})$, all $\theta \in \Theta$ and all f such that $|f|_V \le 1$, we have

$$|\mu_1 \mathbf{K}_\theta^n f - \mu_2 \mathbf{K}_\theta^n f| = \left| \int \mu_1(\mathrm{d}x)\, \mu_2(\mathrm{d}x')\, (\mathbf{K}_\theta^n f(x) - \mathbf{K}_\theta^n f(x')) \right| \le \int \mu_1(\mathrm{d}x)\, \mu_2(\mathrm{d}x')\, \|\mathbf{K}_\theta^n(x,\cdot) - \mathbf{K}_\theta^n(x',\cdot)\|_V .$$

Thus, (16) provides a constant $C > 0$ such that

$$\|\mu_1 \mathbf{K}_\theta^n - \mu_2 \mathbf{K}_\theta^n\|_V \le C(\mu_1 V + \mu_2 V)\rho^n .$$

Taking $\mu_1 = \delta_x$ and $\mu_2 = \mathbf{K}_\theta(x,\cdot)$ and combining with (ii), we get that

$$\sum_{n=0}^{\infty} \sup_{\theta\in\Theta} \|\mathbf{K}_\theta^n(x,\cdot) - \mathbf{K}_\theta^{n+1}(x,\cdot)\|_V \le \frac{C(1+M)}{1-\rho} V(x),$$

and since $\mathcal{M}_V$ is complete, we obtain that $\mathbf{K}_\theta^n(x,\cdot)$ converges uniformly in $\mathcal{M}_V$ over Θ. Denoting by $\tilde{\pi}_\theta$ this limit, we are only required to show that $\tilde{\pi}_\theta = \pi_\theta^X$ in order to establish (iv). Now, since all bounded functions have finite V-norm, for all $A \in \mathcal{X}$,

$$\tilde{\pi}_\theta(A) = \lim_{n\to\infty} \mathbf{K}_\theta^n \mathbb{1}_A(x) = \lim_{n\to\infty} \mathbf{K}_\theta^{n+1} \mathbb{1}_A(x) = \tilde{\pi}_\theta \mathbf{K}_\theta(A),$$

so that $\tilde{\pi}_\theta$ is an invariant probability measure for $\mathbf{K}_\theta$. It is thus π_θ^X as a consequence of (**B1**).

Step 2. Next, we show that $\lim_{\theta\to\theta_\star} \pi_{\theta_\star}^X \phi_\theta = 0$. Assumption (i), $\pi_{\theta_\star}^X V < \infty$ and the dominated convergence theorem give immediately that $\theta \mapsto \phi_\theta(x)$ is continuous at $\theta_\star$, for μ-a.e. x. Moreover, $\sup_{\theta\in\Theta} |\phi_\theta|_V < \infty$. Consequently, using again the dominated convergence theorem, the last term on the right hand side of (17) converges to $\pi_{\theta_\star}^X \phi_{\theta_\star} = 0$ as θ tends to $\theta_\star$.

Step 3. Finally, we consider the term between brackets in (17) and show that it converges to zero as $\theta \to \theta_\star$. Since we just showed that $\sup_{\theta\in\Theta} |\phi_\theta|_V < \infty$, it suffices to show that π_θ^X converges to $\pi_{\theta_\star}^X$ in $\mathcal{M}_V$. By (iv), this boils down to proving that for all $n \in \mathbb{N}^*$, $\mathbf{K}_\theta^n(x,\cdot)$ converges to $\mathbf{K}_{\theta_\star}^n(x,\cdot)$ in $\mathcal{M}_V$. This can be done by induction on n. The base case $n = 1$ corresponds to (iii). The induction follows easily from the following decomposition, valid for all $f : \mathsf{X} \to \mathbb{R}$ such that $|f|_V \le 1$:

$$\begin{aligned} &\left|\mathbf{K}_\theta^{n+1}(x,f) - \mathbf{K}_{\theta_\star}^{n+1}(x,f)\right| \\ &\quad \le \left|\mathbf{K}_\theta^n(x,\mathbf{K}_\theta f) - \mathbf{K}_{\theta_\star}^n(x,\mathbf{K}_\theta f)\right| + \left|\mathbf{K}_{\theta_\star}^n(x,\mathbf{K}_\theta f - \mathbf{K}_{\theta_\star} f)\right| \\ &\quad \le \|\mathbf{K}_\theta^n - \mathbf{K}_{\theta_\star}^n\|_V \sup_{\theta\in\Theta} \|\mathbf{K}_\theta\|_V + \|\mathbf{K}_{\theta_\star}^n\|_V \, \|\mathbf{K}_\theta - \mathbf{K}_{\theta_\star}\|_V . \end{aligned}$$

By observing that (ii) implies $\sup_{\theta\in\Theta} \|\mathbf{K}_\theta^n\|_V < \infty$ for all $n \in \mathbb{N}^*$, we may conclude the proof. □

3.3. Stochastic volatility models

Consider the stochastic volatility model

$$\begin{aligned} X_{k+1} &= \varphi X_k + \sigma\zeta_{k+1}, \\ Y_k &= \beta \exp(X_k/2)\epsilon_k, \end{aligned} \qquad k \in \mathbb{N}, \tag{18}$$

where $(\zeta_k)_{k\in\mathbb{N}^*}$ and $(\epsilon_k)_{k\in\mathbb{N}}$ are independent sequences of i.i.d. Gaussian random vectors in $\mathbb{R}^2$ with zero mean and identity covariance matrix; see [21]. A general description of stochastic volatility models as HMMs is provided in [17]. In this case, $\mathsf{X} = \mathsf{Y} = \mathbb{R}$. If the parameter vector $\theta = (\beta, \sigma, \varphi)$ belongs to a compact parameter space, we may apply the theory developed in the previous section. However, in this example, θ is assumed to belong to the *non-compact* parameter space

$$\Theta \coloneqq \{(\beta, \sigma, \varphi) : \beta \geq \beta_-,\ \sigma \geq \sigma_-,\ |\varphi| \leq \varphi_+\},$$

where $\beta_- > 0$, $\sigma_- > 0$ and $\varphi_+ \in (0, 1)$. Denote by $\theta_\star = (\beta_\star, \sigma_\star, \varphi_\star)$ the true value of the parameter. In this model,

$$\begin{aligned} k_\theta(x, x') &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x' - \varphi x)^2}{2\sigma^2}\right\}, \\ g_\theta(x, y) &= \frac{1}{\sqrt{2\pi\beta^2}} \exp\left\{-\frac{x}{2} - \frac{y^2}{2\beta^2}\mathrm{e}^{-x}\right\}. \end{aligned} \tag{19}$$

Assumption (**B1**) is clearly satisfied with

$$\pi_\theta(B) = \int \frac{1}{2\pi}\sqrt{\frac{1-\varphi^2}{\sigma^2}} \exp\left(-\frac{(1-\varphi^2)x^2}{2\sigma^2} - \frac{u^2}{2}\right) \mathbb{1}_B(x, \beta\mathrm{e}^{x/2}u)\,\mathrm{d}u\,\mathrm{d}x. \tag{20}$$

Assumption (**B2**) follows from [12, Section 3.3] and Assumption (**B3**) is immediate. Moreover, straightforward algebra yields

$$\begin{aligned} \Delta(\theta_\star, \theta) = \log\frac{\sigma\beta}{\sigma_\star\beta_\star} &+ \frac{1}{2}\mathbb{E}_{\pi_\star}^{\theta_\star}\left[X_1^2\right]\left(\sigma^{-2} - \sigma_\star^{-2}\right) + \mathbb{E}_{\pi_\star}^{\theta_\star}\left[X_0X_1\right]\left(\frac{\phi_\star}{\sigma_\star^2} - \frac{\phi}{\sigma^2}\right) \\ &+ \frac{1}{2}\mathbb{E}_{\pi_\star}^{\theta_\star}\left[X_0^2\right]\left(\frac{\phi^2}{\sigma^2} - \frac{\phi_\star^2}{\sigma_\star^2}\right) + \frac{1}{2}\mathbb{E}_{\pi_\star}^{\theta_\star}\left[Y_1^2\mathrm{e}^{-X_1}\right]\left(\beta^{-2} - \beta_\star^{-2}\right). \end{aligned}$$

Note that $\mathbb{E}_{\pi_\star}^{\theta_\star}\left[\mathrm{e}^{a|X_0|}\right]$ and $\mathbb{E}_{\pi_\star}^{\theta_\star}\left[|Y_0|^a\right]$ are finite for all $a > 0$. Hence by the Cauchy-Schwarz inequality all the expectations appearing in the previous display are finite and we conclude that, for all $\theta_\star \in \Theta$, $\theta \mapsto \Delta(\theta_\star, \theta)$ is a continuous function. Thus, (**B4**) holds for all priors being strictly positive (possibly unnormalized) measures λ on $(\Theta, \mathcal{T})$. Since Θ is non-compact, the posterior consistency needs

to be established via Theorem 3. To this end, it only remains to check (**B5**) and (**B6**).

We check (**B5**) with $\ell = 2$. Write, for all $A \in \mathcal{T}$ and $y_{0:2} \in \mathbb{R}^3$,

$$\hat{p}_A(y_{0:2}) = \sup_{(\theta,x_0)\in A\times\mathbb{R}} D_{\theta,x_0}(y_{0:2}), \tag{21}$$

with

$$D_{\theta,x_0}(y_{0:2}) \coloneqq \iint k_\theta(x_0,x_1)g_\theta(x_1,y_1)k_\theta(x_1,x_2)g_\theta(x_2,y_2)\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

We will use the following bounds obtained by straightforward algebra: for all $\theta \in \Theta$,

$$\sup_{x\in\mathbb{R}} g_\theta(x,y) = \frac{1}{|y|\sqrt{2\pi\mathrm{e}}}, \quad \int g_\theta(x,y)\,\mathrm{d}x = \frac{1}{|y|}, \quad g_\theta(x,y) \le \frac{\mathrm{e}^{-x/2}}{\beta\sqrt{2\pi}} \tag{22}$$

and

$$\sup_{(x,x')\in\mathbb{R}^2} k_\theta(x,x') = \frac{1}{\sqrt{2\pi\sigma^2}}. \tag{23}$$

Then, using (22) and (23),

$$\begin{aligned} D_{\theta,x_0}(y_{0:2}) &\le \sup_{(x_0,x_1)\in\mathbb{R}^2} k_\theta(x_0,x_1)\int g_\theta(x_1,y_1)\int k_\theta(x_1,x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1 \sup_{x_2\in\mathbb{R}} g_\theta(x_2,y_2) \\ &\le \frac{1}{|y_1||y_2|\sqrt{2\pi\sigma^2}\sqrt{2\pi\mathrm{e}}}. \end{aligned} \tag{24}$$

Moreover, using the definitions of $g_\theta(x_1,y_1)$, $k_\theta(x_1,x_2)$ and the bound on $g_\theta(x_2,y_2)$ given in (22), we get, by standard calculations,

$$\begin{aligned} D_{\theta,x_0}(y_{0:2}) &\le \left(\int k_\theta(x_0,x_1)\,\mathrm{d}x_1\right)\sup_{x_1\in\mathbb{R}}\left(g_\theta(x_1,y_1)\int k_\theta(x_1,x_2)g_\theta(x_2,y_2)\,\mathrm{d}x_2\right) \\ &\le \sup_{x_1\in\mathbb{R}}\left(\frac{1}{\sqrt{2\pi\beta^2}}\exp\left\{-\frac{x_1}{2}-\frac{y_1^2}{2\beta^2}\mathrm{e}^{-x_1}\right\}\right. \\ &\qquad \left.\times\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{\sqrt{2\pi\beta^2}}\int\exp\left\{-\frac{(x_2-\varphi x_1)^2}{2\sigma^2}-\frac{x_2}{2}\right\}\mathrm{d}x_2\right) \\ &= \sup_{x_1\in\mathbb{R}}\left(\frac{1}{\sqrt{2\pi\beta^2}}\exp\left\{-\frac{x_1}{2}-\frac{y_1^2}{2\beta^2}\mathrm{e}^{-x_1}\right\}\frac{1}{\sqrt{2\pi\beta^2}}\exp\left\{-\frac{1}{2}\varphi x_1+\frac{\sigma^2}{8}\right\}\right) \\ &= \sup_{x_1\in\mathbb{R}}\left(\frac{1}{2\pi\beta^2}\exp\left\{-\frac{x_1}{2}(1+\varphi)-\frac{y_1^2}{2\beta^2}\mathrm{e}^{-x_1}+\frac{\sigma^2}{8}\right\}\right) \\ &= \frac{\mathrm{e}^{\frac{\sigma^2}{8}}}{2\pi\beta^{1-\varphi}}\left[\frac{(1+\varphi)\mathrm{e}^{-1}}{y_1^2}\right]^{\frac{1+\varphi}{2}}. \end{aligned} \tag{25}$$

We then set $C_m = \{\theta \in \Theta : \sigma^2 \le \log m,\ \beta \le \mathrm{e}^m\}$, so that

$$C_m^c \subset \{\theta \in \Theta :\ \sigma^2 > \log m\} \cup \{\theta \in \Theta :\ \sigma^2 \le \log m,\ \beta > \mathrm{e}^m\}.$$

Combining this inclusion with (21), (24) and (25) yields

$$\limsup_{m\to\infty} \hat{p}_{C_m^c}(Y_{0:2}) = 0 \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.},$$

implying that (9) is satisfied with $\ell = 2$. Now, by (24), for all $y_{0:2} \in \mathbb{R}^3$,

$$\log^+ \hat{p}_\Theta(y_{0:2}) = \sup_{(\theta,x_0)\in\Theta\times\mathbb{R}} \log^+ D_{\theta,x_0}(y_{0:2}) \le \frac{1}{2}\log^+(4\pi^2\sigma^2\mathrm{e}) + \log^+|y_1| + \log^+|y_2|,$$

and using (20), this implies that (10) is satisfied with $\ell = 2$. Thus, we may conclude that (**B5**) holds.

Condition (11) in (**B6**) holds if λ is a probability measure on $(\Theta, \mathcal{T})$. Alternatively, one may, e.g., use (24) and (25) to obtain that for all $y_{1:2} \in \mathbb{R}^2$,

$$p_{\theta,\eta}(y_{1:2}) \le \left(\frac{1}{|y_1||y_2|\sqrt{2\pi\sigma^2}\sqrt{2\pi\mathrm{e}}}\right) \wedge \left(\frac{\mathrm{e}^{\frac{\sigma^2}{8}}}{2\pi\beta^{1-\varphi}}\left[\frac{(1+\varphi)\mathrm{e}^{-1}}{y_1^2}\right]^{\frac{1+\varphi}{2}}\right).$$

Condition (11) in (**B6**) is then implied by less restrictive condition

$$\int \sigma^{-1} \wedge (\mathrm{e}^{\sigma^2/8}/\beta^{1-\varphi_+})\,\lambda(\mathrm{d}\theta) < \infty.$$

We now prove that (12) in Assumption (**B6**) holds true with $n_0 = 1$. By Jensen's inequality and (19),

$$\begin{aligned}
\mathbb{E}_{\pi_\star}^{\theta_\star}[\log p_{\theta_\star,\pi_\star}(Y_1)] &= \mathbb{E}_{\pi_\star}^{\theta_\star}\left[\log \iint \pi_\star(\mathrm{d}x_0)\, k_{\theta_\star}(x_0,x_1) g_{\theta_\star}(x_1,Y_1)\right] \\
&\ge \mathbb{E}_{\pi_\star}^{\theta_\star}\left[\iint \pi_\star(\mathrm{d}x_0)\, k_{\theta_\star}(x_0,x_1)\log g_{\theta_\star}(x_1,Y_1)\right] \\
&= -\frac{1}{2}\log(2\pi\beta_\star^2) - \frac{\mathbb{E}_{\pi_\star}^{\theta_\star}[X_1]}{2} - \frac{\mathbb{E}_{\pi_\star}^{\theta_\star}[Y_1^2\mathrm{e}^{-X_1}]}{2\beta_\star^2} > -\infty,
\end{aligned}$$

where we used again that $\mathbb{E}_{\pi_\star}^{\theta_\star}[\mathrm{e}^{a|X_0|}]$ and $\mathbb{E}_{\pi_\star}^{\theta_\star}[|Y_0|^a]$ are both finite for all $a > 0$. Finally, Theorem 3 applies, establishing the posterior consistency for the model (18), e.g., for all strictly positive probability measures λ on Θ and all initial distributions $\eta_\star$ on $(\mathsf{Z}, \mathcal{Z})$,

$$\mathbb{P}_{\eta_\star}^{\theta_\star}\left(\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}\right) = 1.$$

4. A general approach to posterior consistency

It turns out to be convenient to embed the problem of posterior consistency for fdPOMMs into a more general setting. This widened perspective allows a number of universal steps to be identified, by which the posterior consistency can be established in general. Importantly, this machinery is not at all specific to the framework of fdPOMMs and is thus of independent interest.

4.1. General setting

Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathbb{P})$ be a filtered probability space. If for all $n \in \mathbb{N}$, ν_n is a σ-finite measure on $(\Omega, \mathcal{F}_n)$ and the restriction $\mathbb{P}|_{\mathcal{F}_n}$ of $\mathbb{P}$ to $\mathcal{F}_n$ is absolutely continuous with respect to ν_n, then we say that $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}}, \mathbb{P})$ is a *progressively dominated filtered probability space*.

Moreover, let $\{p_{\theta,n}, \theta \in \Theta\}$ be a collection of probability densities with respect to ν_n. Let p_n^* denote the Radon-Nikodym derivative

$$p_n^* \coloneqq \frac{\mathrm{d}\mathbb{P}|_{\mathcal{F}_n}}{\mathrm{d}\nu_n}.$$

In Section 5, it is shown how the fdPOMMs can be cast into this general setting; see also Appendix B for a treatment of the more simple i.i.d. case.

We now introduce the prior and posterior distributions, denoted by λ and λ_n, respectively, in this general setting. Let λ be a non-zero σ-finite measure on $(\Theta, \mathcal{T})$. Then, for all $A \in \mathcal{T}$, the posterior "probability" $\lambda_n(A)$ with prior λ is defined by

$$\lambda_n(A) \coloneqq \frac{\int_A \lambda(\mathrm{d}\theta)\, p_{\theta,n}}{\int_\Theta \lambda(\mathrm{d}\theta)\, p_{\theta,n}} \tag{26}$$

whenever this ratio is well-defined. In what follows, we will always assume that $(\theta, \omega) \mapsto p_{\theta,n}(\omega)$ is measurable from $(\Theta \times \Omega, \mathcal{T} \otimes \mathcal{F}_n)$ to $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, which implies that the numerator and denominator are (non-negative) $\mathcal{F}_n$-measurable random variables.

We now introduce the main assumption on the model. It says that there are "sufficiently many" θ for which the likelihood ratio $p_{\theta,n}/p_n^*$ is not decreasing exponentially fast under $\mathbb{P}$.

(A1) For all $\delta > 0$, there exists $\Theta_\delta \in \mathcal{T}$ such that $\lambda(\Theta_\delta) > 0$ and for all $\theta \in \Theta_\delta$,

$$\liminf_{n\to\infty} n^{-1} \log \frac{p_{\theta,n}}{p_n^*} \geq -\delta \quad \mathbb{P}\text{-a.s.}$$

We expect the set Θ_δ in (**A1**) to contain parameters θ whose corresponding densities $p_{\theta,n}$ are "asymptotically close" to the true ones p_n^*. In contrast, the following definition addresses the case of parameters indexing densities remaining "far away" from the true ones.

Definition 6 *We say that a set $A \in \mathcal{T}$ is* $\mathbb{P}$-remote *if and only if*

$$\limsup_{n\to\infty} n^{-1} \log\left(\int_A \frac{p_{\theta,n}}{p_n^*}\,\lambda(\mathrm{d}\theta)\right) < 0 \quad \mathbb{P}\text{-a.s.}$$

Moreover, we say that a set A is approximately $\mathbb{P}$-remote *if and only if for all $\varepsilon > 0$ there exists a set $K_\varepsilon \in \mathcal{T}$ such that*

(i) $A \cap K_\varepsilon$ *is $\mathbb{P}$-remote;*
(ii) $\limsup_{n\to\infty} \lambda_n(K_\varepsilon^c) \le \varepsilon \quad \mathbb{P}$-a.s.

We will denote by $\mathcal{A}_{\mathbb{P}}$ the class of all approximately $\mathbb{P}$-remote sets.

Here "remote" refers to the fact that the likelihood ratio averaged over the parameters within A decreases exponentially fast to zero under $\mathbb{P}$.

Remark 7 *Typically, for all $\varepsilon > 0$, Property (ii) in Definition 6 is satisfied for a well chosen compact set K_ε. We refer to this as the* uniform $\mathbb{P}$-a.s. tightness property of the posterior distribution. *In this case a set A is approximately $\mathbb{P}$-remote whenever $A \cap K$ is $\mathbb{P}$-remote for all compact sets K.*

Let $\mathbb{P}_{\theta,n}$ be the probability on $(\Omega, \mathcal{F}_n)$ defined by

$$\mathbb{P}_{\theta,n}(B) := \int_B p_{\theta,n}\,\mathrm{d}\nu_n, \quad B \in \mathcal{F}_n.$$

We have the following characterization of $\mathbb{P}$-remote sets, which is closely related to [3, Theorem 5(2)], although in a simplified form and without relying on the asymptotic-merging condition, thanks to the normalization by p_n^*.

Proposition 6 *The set $A \in \mathcal{T}$ is $\mathbb{P}$-remote if and only if there exists a sequence $(B_n)_{n\in\mathbb{N}}$ of sets in $\mathcal{F}$ such that $B_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$ and*

$$\limsup_{n\to\infty} n^{-1} \log \int_A \lambda(\mathrm{d}\theta)\,\mathbb{P}_{\theta,n}(B_n) < 0, \tag{27}$$

$$\mathbb{P}\left(\liminf_{n\to\infty} B_n\right) = 1. \tag{28}$$

Proof. See Appendix C.1. □

Following the approach of [4] we have the following general result, which extends the i.i.d. framework used in that reference.

Theorem 7 *Assume that (**A1**) holds. Then all approximately $\mathbb{P}$-remote sets $A \in \mathcal{A}_{\mathbb{P}}$ satisfy*

$$\lim_{n\to\infty} \lambda_n(A) = 0 \quad \mathbb{P}\text{-a.s.}, \tag{29}$$

where $\lambda_n(A)$ is defined by (26).

Proof. See Section 4.3. □

Remark 8 *It may happen that the ratio that defines $\lambda_n(A)$ in (26) is not well-defined. However, in (29) it should be understood that $\lambda_n(A)$ is well-defined, $\mathbb{P}$-a.s., for n large enough (otherwise the limit would not be defined).*

Remark 9 *Note that λ_n is, $\mathbb{P}$-a.s., a well-defined (posterior) probability on $(\Theta, \mathcal{T})$ if*

$$\mathbb{P}\left(0 < \int \lambda(\mathrm{d}\theta) p_{\theta,n} < \infty\right) = 1. \tag{30}$$

There are simple although restrictive sufficient conditions to ensure (30). For instance, note that when the prior is proper, i.e., when λ is a finite measure, then $\int_\Theta \lambda(\mathrm{d}\theta) p_{\theta,n} < \infty$, ν_n-almost everywhere and thus $\mathbb{P}$-a.s. Moreover, if for all $\theta \in \Theta$, $p_{\theta,n} > 0$ $\mathbb{P}$-a.s., then $\mathbb{P}(\int_\Theta \lambda(\mathrm{d}\theta) p_{\theta,n} > 0) = 1$.

From now on we suppose that (Θ, d) is a metric space, and let $\mathcal{T}$ be the Borel σ-field.

Remark 10 *Assume that λ_n is, $\mathbb{P}$-a.s., a well-defined probability measure on $(\Theta, \mathcal{T})$ for n large enough. Suppose in addition that there exists $\theta_\star \in \Theta$ such that $A_p = \{\theta \in \Theta : d(\theta, \theta_\star) \geq 1/p\}$ is approximately $\mathbb{P}$-remote for all $p \in \mathbb{N}^*$. Then Theorem 7 implies*

$$\mathbb{P}\left(\lambda_n \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}\right) = 1. \tag{31}$$

This stems from the fact that by the Portmanteau lemma, $\{\lambda_n \Longrightarrow_n \delta_{\theta_\star}\}$ is implied by $\{\lim_n \lambda_n(A_p) = 0 \text{ for all } p \in \mathbb{N}^*\}$. Indeed, suppose that the latter event has occurred and let F be a closed set. If $\theta_\star \notin F$ then there exists $p \in \mathbb{N}^*$ such that $F \subset A_p$ and $\limsup_n \lambda_n(F) \leq \limsup_n \lambda_n(A_p) = 0 = \delta_{\theta_\star}(F)$. On the other hand, since we assumed λ_n to be a probability measure for n large enough, if $\theta_\star \in F$, $\limsup_n \lambda_n(F) \leq 1 = \delta_{\theta_\star}(F)$.

The following lemma will be useful for checking (**A1**) with an explicit expression of Θ_δ.

Lemma 8 *Let $(\Omega, \mathcal{F}, (\overline{\mathcal{F}}_n)_{n\in\mathbb{N}}, (\overline{\nu}_n)_{n\in\mathbb{N}}, \mathbb{P})$ be a progressively dominated filtered probability space. For all $n \in \mathbb{N}$, let $\overline{p}_n$ be a probability density function with respect to $\overline{\nu}_n$. Denote by $\overline{p}_n^*$ the density of $\overline{\mathbb{P}}_n^* = \mathbb{P}|_{\overline{\mathcal{F}}_n}$ with respect to $\overline{\nu}_n$ and by $\overline{\mathbb{P}}_n$ the probability measure having density $\overline{p}_n$ with respect to $\overline{\nu}_n$. Suppose that for all $n \in \mathbb{N}$, $\tilde{\mathcal{F}}_n$ is a sub-σ-field of $\overline{\mathcal{F}}_n$ and define by*

$$\tilde{p}_n := \frac{\mathrm{d}[\overline{\mathbb{P}}_{n|\tilde{\mathcal{F}}_n}]}{\mathrm{d}[\overline{\nu}_{n|\tilde{\mathcal{F}}_n}]} \quad \text{and} \quad \tilde{p}_n^* := \frac{\mathrm{d}[\overline{\mathbb{P}}^*_{n|\tilde{\mathcal{F}}_n}]}{\mathrm{d}[\overline{\nu}_{n|\tilde{\mathcal{F}}_n}]} \tag{32}$$

the Radon-Nikodym derivatives of the $\tilde{\mathcal{F}}_n$-restrictions of $\overline{\mathbb{P}}_n$ and $\overline{\mathbb{P}}_n^*$ w.r.t. the $\tilde{\mathcal{F}}_n$-restriction of $\overline{\nu}_n$, respectively. Then it holds that

$$n^{-1} \log \frac{\tilde{p}_n}{\tilde{p}_n^*} \geq n^{-1} \log \frac{\overline{p}_n}{\overline{p}_n^*} + \epsilon_n \quad \text{with} \quad \lim_{n\to\infty} \epsilon_n = 0 \quad \mathbb{P}\text{-a.s.} \tag{33}$$

Proof. Let $n \in \mathbb{N}$. The result follows from the identity

$$\frac{\tilde{p}_n}{\tilde{p}_n^*} = \overline{\mathbb{E}}_n^* \left[\frac{\overline{p}_n}{\overline{p}_n^*} \middle| \tilde{\mathcal{F}}_n \right] \quad \mathbb{P}_n^*\text{-a.s.;} \tag{34}$$

indeed, since, by (34),

$$\overline{\mathbb{P}}_n^* \left(\frac{\tilde{p}_n}{\tilde{p}_n^*} = 0, \frac{\overline{p}_n}{\overline{p}_n^*} > 0 \right) = 0,$$

it holds that

$$\overline{\mathbb{P}}_n^* \left(\frac{\overline{p}_n}{\overline{p}_n^*} > n^2 \frac{\tilde{p}_n}{\tilde{p}_n^*} \right) = \overline{\mathbb{P}}_n^* \left(\frac{\overline{p}_n}{\overline{p}_n^*} > n^2 \frac{\tilde{p}_n}{\tilde{p}_n^*}, \frac{\tilde{p}_n}{\tilde{p}_n^*} \neq 0 \right).$$

Hence, using the Markov inequality, we get

$$\overline{\mathbb{P}}_n^* \left(\frac{\overline{p}_n}{\overline{p}_n^*} > n^2 \frac{\tilde{p}_n}{\tilde{p}_n^*} \right) \leq n^{-2} \overline{\mathbb{E}}_n^* \left[\left(\frac{\overline{p}_n}{\overline{p}_n^*} \Big/ \frac{\tilde{p}_n}{\tilde{p}_n^*} \right) \mathbb{1} \left\{ \frac{\tilde{p}_n}{\tilde{p}_n^*} \neq 0 \right\} \right].$$

By conditioning on $\tilde{\mathcal{F}}_n$ and reapplying (34), we conclude that the expectation on the right hand side of the previous display is equal to $\mathbb{P}_n^* \left(\frac{\tilde{p}_n}{\tilde{p}_n^*} \neq 0 \right) \leq 1$. We thus get that

$$\overline{\mathbb{P}}_n^* \left(\frac{\overline{p}_n}{\overline{p}_n^*} > n^2 \frac{\tilde{p}_n}{\tilde{p}_n^*} \right) \leq n^{-2}.$$

Since $\overline{\mathbb{P}}_n^*$ coincides with $\mathbb{P}$ on $\overline{\mathcal{F}}_n$, the Borel-Cantelli lemma gives that

$$\mathbb{P} \left(\frac{\overline{p}_n}{\overline{p}_n^*} > n^2 \frac{\tilde{p}_n}{\tilde{p}_n^*} \quad \text{i.o.} \right) = 0.$$

This implies (33) and concludes the proof. □

4.2. $\mathbb{P}$-remoteness of $\theta_\star$-missing compact sets

Note that Theorem 7 does not rely on the existence of a *true parameter* $\theta_\star$. This only appears in Remark 10 where we assume the existence of a parameter $\theta_\star$ such that closed sets not containing this parameter are approximately $\mathbb{P}$-remote. In this section we relate the notion of $\mathbb{P}$-remoteness to the consistency of approximate maximum likelihood estimators. The first step is to relate the true density p_n^* to a true parameter by assuming that p_n^* and $p_{\theta_\star,n}$ *merge with probability one* in the sense of [3, Definition 1], i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{p_{\theta_\star,n}}{p_n^*} = 0 \qquad \mathbb{P}\text{-a.s.} \tag{35}$$

Let $A \in \mathcal{T}$ be a closed set that does not contain $\theta_\star$. In this section we show that $A \cap K$ is $\mathbb{P}$-remote for every compact $K \in \mathcal{T}$ on which one is able to

establish the consistency of approximate maximum likelihood estimators. For most models of interest, this is actually possible for all compact sets K, and consequently A is approximately $\mathbb{P}$-remote whenever the uniform $\mathbb{P}$-a.s. tightness property of the posterior distribution holds true (see Remark 7).

Definition 11 *Let $K \in \mathcal{T}$ be compact. We say that $(\hat{\theta}_n)_{n\in\mathbb{N}}$ is a sequence of* approximate maximum likelihood estimators *(AMLEs) on K if it is $(\mathcal{F}_n)_{n\in\mathbb{N}}$-adapted and for all $n \in \mathbb{N}$, $\hat{\theta}_n \in K$ and*

$$n^{-1} \log p_{\hat{\theta}_n,n} \geq n^{-1} \log p^*_n + \epsilon_n \quad \textit{with} \quad \lim_{n\to\infty} \epsilon_n = 0 \quad \mathbb{P}\text{-a.s.}$$

Proposition 9 *Let $K \in \mathcal{T}$ be compact and suppose that there exists $\theta_\star \in K$ such that (35) holds. Then the two following assertions are equivalent.*

(i) All sequences $(\hat{\theta}_n)_{n\in\mathbb{N}}$ of AMLEs on K are strongly consistent.
(ii) For all closed sets A not containing $\theta_\star$,

$$\limsup_{n\to\infty} \sup_{\theta\in A\cap K} \frac{1}{n} \log \frac{p_{\theta,n}}{p^*_n} < 0 \qquad \mathbb{P}\text{-a.s.} \tag{36}$$

Suppose that one of these assertions holds true and, in addition, that $\lambda(K) < \infty$. Then, for all closed sets A not containing $\theta_\star$, the set $A \cap K$ is $\mathbb{P}$-remote.

Proof. We first show that (i) implies (ii). Suppose that (ii) is false; then there exists a closed set A not containing $\theta_\star$ such that (36) does not hold. For all n, let $\tilde{\theta}_n \in A \cap K$ be such that

$$\log \frac{p_{\tilde{\theta}_n,n}}{p^*_n} \geq \sup_{\theta\in A\cap K} \log \frac{p_{\theta,n}}{p^*_n} - 1.$$

Since (36) does not hold, there exists, on a set Ω^* with $\mathbb{P}(\Omega^*) > 0$, an increasing sequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \frac{1}{n_k} \log \frac{p_{\tilde{\theta}_{n_k},n_k}}{p^*_{n_k}} \geq 0.$$

Define the random variables $\hat{\theta}_n$ by setting, on Ω^* and if $n = n_k$ for some $k \in \mathbb{N}$, $\hat{\theta}_n = \tilde{\theta}_n$ and $\hat{\theta}_n = \theta_\star$ otherwise. Then by (35) and the previous display, $(\hat{\theta}_n)_{n\in\mathbb{N}}$ is a sequence of AMLEs; however, it is not strongly consistent, since $\hat{\theta}_{n_k} = \tilde{\theta}_{n_k} \in A \cap K \not\ni \theta_\star$ on Ω^* for all $k \in \mathbb{N}$. Hence (i) does not hold.

We now show that (ii) implies (i). Let A a closed set not containing $\theta_\star$, and let $(\hat{\theta}_n)_{n\in\mathbb{N}}$ be a sequence of AMLEs on K. Then $\lim_{n\to\infty} \epsilon_n = 0$ $\mathbb{P}$-a.s. with

$$\epsilon_n \leq \frac{1}{n} \log \frac{p_{\hat{\theta}_n,n}}{p^*_n}.$$

We thus have that

$$\hat{\theta}_n \in A \cap K \Longrightarrow \epsilon_n \leq \sup_{\theta\in A\cap K} \frac{1}{n} \log \frac{p_{\theta,n}}{p^*_n}.$$

Now (ii) and the limit $\lim_{n\to\infty} \epsilon_n = 0$ $\mathbb{P}$-a.s. imply that

$$\epsilon_n > \sup_{\theta\in A\cap K} \frac{1}{n} \log \frac{p_{\theta,n}}{p_n^*}$$

eventually, $\mathbb{P}$-a.s. The previous implication therefore shows that $\hat{\theta}_n \in K \setminus A$ eventually, $\mathbb{P}$-a.s. The proof is completed by taking $A = \{\theta \in \Theta : d(\theta, \theta_\star) < 1/p\}^c$ for any positive integer p, which shows that $\hat{\theta}_n$ is strongly consistent.

The last assertion of Proposition 9 follows immediately from the bound

$$\log\left(\int_{A\cap K} \frac{p_{\theta,n}}{p_n^*}\, \lambda(\mathrm{d}\theta)\right) \le \log \lambda(K) + \sup_{\theta\in A\cap K} \log \frac{p_{\theta,n}}{p_n^*}\,.$$

□

4.3. Proof of Theorem 7

We preface the proof of Theorem 7 by the following lemma.

Lemma 10 *Under (A1), for all $\varepsilon > 0$,*

$$\mathbb{P}\left(\int \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} \le \mathrm{e}^{-\varepsilon n} \quad i.o.\right) = 0. \tag{37}$$

Remark 12 *According to the terminology used in [3, Definition 1], the property (37) says that $\int \lambda(\mathrm{d}\theta)\, p_{\theta,n}$ and p_n^* merge with probability one.*

Proof. Pick $\varepsilon > 0$ and write

$$\mathrm{e}^{\varepsilon n} \int \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} \ge \int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, F_{\theta,n},$$

where $\Theta_{\varepsilon/2}$ is defined in (**A1**) and for all $\theta \in \Theta$ and $n \in \mathbb{N}^*$,

$$F_{\theta,n} := \exp\left(\varepsilon n + \log \frac{p_{\theta,n}}{p_n^*}\right).$$

By assumption, for all $\theta \in \Theta_{\varepsilon/2}$, $\liminf_n F_{\theta,n} = \infty$ $\mathbb{P}$-a.s., and the proof is completed by establishing that, $\mathbb{P}$-a.s.,

$$\liminf_{n\to\infty} \int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, F_{\theta,n} = \infty. \tag{38}$$

To check (38), note that, by Fubini's theorem,

$$\mathbb{E}\left[\int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, \mathbb{1}\left\{\liminf_{n\to\infty} F_{\theta,n} < \infty\right\}\right] = \int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, \mathbb{P}\left(\liminf_{n\to\infty} F_{\theta,n} < \infty\right) = 0.$$

Hence, $\mathbb{P}$-a.s.,

$$\int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, \mathbb{1}\left\{\liminf_{n\to\infty} F_{\theta,n} < \infty\right\} = 0,$$

and, consequently, by Fatou's lemma and the fact that $\lambda(\Theta_{\varepsilon/2}) > 0$,

$$\liminf_{n\to\infty} \int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, F_{\theta,n} \geq \int_{\Theta_{\varepsilon/2}} \lambda(\mathrm{d}\theta)\, \liminf_{n\to\infty} F_{\theta,n} \mathbb{1}\left\{\liminf_{n\to\infty} F_{\theta,n} = \infty\right\} = \infty,$$

$\mathbb{P}$-a.s., which completes the proof. □

Using the previous lemma, the proof of Theorem 7 is straightforward. Indeed, let $A \in \mathcal{T}$ be approximately $\mathbb{P}$-remote; then for all $\varepsilon > 0$ there exists $K_\varepsilon \in \mathcal{T}$ such that $\mathbb{P}$-a.s.,

$$\limsup_{n\to\infty} \lambda_n(A) \leq \limsup_{n\to\infty} \lambda_n(A \cap K_\varepsilon) + \varepsilon.$$

To treat further the right hand side above, let

$$\alpha := -\limsup_{n\to\infty} n^{-1} \log\left(\int_{A\cap K_\varepsilon} \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*}\right), \tag{39}$$

which is $\mathbb{P}$-a.s. positive by Definition 6(i), and write

$$\lambda_n(A \cap K_\varepsilon) = \left(\mathrm{e}^{\alpha n/2} \int_{A\cap K_\varepsilon} \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*}\right) \left(\mathrm{e}^{\alpha n/2} \int \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*}\right)^{-1}. \tag{40}$$

Here, by Lemma 10,

$$\limsup_{n\to\infty} \left(\mathrm{e}^{\alpha n/2} \int \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*}\right)^{-1} \leq 1 \qquad \mathbb{P}\text{-a.s.},$$

and applying this bound together with (39) and (40) yields, $\mathbb{P}$-a.s.,

$$\limsup_{n\to\infty} \lambda_n(A \cap K_\varepsilon) = 0.$$

Consequently, $\limsup_n \lambda_n(A) \leq \varepsilon$ $\mathbb{P}$-a.s., and as ε was picked arbitrarily we may conclude the proof.

5. Proof of main results

5.1. Preliminaries

We use the general setting detailed in Section 4.1 for proving the results in Section 2.2. In Appendix B, also the i.i.d. case is embedded into the general setting for completeness. Here the σ-finite measure ν_n is defined similarly (see (56)), but in the present case, for a given initial distribution η, we define, for all

$\theta \in \Theta$ and $n \in \mathbb{N}^*$, $p_{\theta,n}$ as the density of $\mathbb{P}_\eta^\theta|_{\mathcal{F}_n}$ w.r.t. ν_n, i.e., it satisfies, for all $B = [Y_{1:n}]^{-1}(A) \in \mathcal{F}_n$ with $A \in \mathcal{Y}^{\otimes n}$,

$$\int_B p_{\theta,n} \, \mathrm{d}\nu_n = \mathbb{P}_\eta^\theta(B) = \mathbb{P}_\eta^\theta(Y_{1:n} \in A). \tag{41}$$

This density is simply given by

$$p_{\theta,n} = p_{\theta,\eta}(Y_{1:n}).$$

Remark 13 *In Appendix B the true density is among the targeted ones,* $p_n^* = p_{\theta_\star,n}$. *Although we here assume a* true parameter $\theta_\star \in \Theta$, *we do not suppose that* $p_n^* = p_{\theta_\star,n}$. *The main reason is that in the case of fdPOMMs, the initial distribution* η *in* (41) *is chosen arbitrarily, often for computational convenience. More specifically, here* $p_{\theta_\star,n} = p_{\theta_\star,\eta}(Y_{1:n})$, *where the initial distribution* η *used in practice when computing the likelihood is chosen arbitrarily as one generally different from the* true *initial distribution. Concerning the latter we will instead consider* $p_n^* = p_{\theta_\star,\pi_\star}(Y_{1:n})$ *(i.e. the true initial distribution is the invariant one).*

Following this remark, the first thing to check is the merging property (35) with $p_n^* = p_{\theta_\star,\pi_\star}(Y_{1:n})$ and $p_{\theta_\star,n} = p_{\theta_\star,\eta}(Y_{1:n})$, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{p_{\theta_\star,\eta}(Y_{1:n})}{p_{\theta_\star,\pi_\star}(Y_{1:n})} = 0 \qquad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.} \tag{42}$$

As we will see below, this condition is implied by the following assumption.

(B7) For all $\theta \in \Theta$ and all initial distributions η, the distribution of $Y_{1:\infty}$ under $\mathbb{P}_\eta^\theta$ admits a positive density with respect to the distribution of $Y_{1:\infty}$ under $\mathbb{P}_{\pi_\theta}^\theta$, i.e.,

$$R_\eta^\theta := \frac{\mathrm{d}\mathbb{P}_\eta^\theta(Y_{1:\infty} \in \cdot)}{\mathrm{d}\mathbb{P}_{\pi_\theta}^\theta(Y_{1:\infty} \in \cdot)} > 0 \quad \mathbb{P}_{\pi_\theta}^\theta\text{-a.s.}$$

We now state a result that will be shown to apply under the various sets of assumptions in Section 5.2 and Section 5.3

Proposition 11 *Consider a fdPOMM satisfying (**B1–2**) and (**B7**). Let* λ *be a Radon measure on* Θ, $\theta_\star \in \Theta$ *and define, for all* $y_{1:n} \in \mathsf{Y}^n$, $\lambda\langle y_{1:n}\rangle$ *by* (5). *Let us consider the following conditions.*

(i) For all $\epsilon > 0$, *there exists a compact set* $K_\epsilon \in \mathcal{T}$ *such that*

$$\limsup_{n\to\infty} \lambda\langle Y_{1:n}\rangle(K_\epsilon^c) \le \epsilon \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.}$$

*(ii) Assumption (**A1**) holds with* $p_{\theta,n} = p_{\theta,\eta}(Y_{1:n})$, $p_n^* = p_{\theta_\star,\pi_\star}(Y_{1:n})$ *and* $\mathbb{P} = \mathbb{P}_{\pi_\star}^{\theta_\star}$.

(iii) For n *large enough, we have* $p_{\theta,\eta}(y_{1:n}) > 0$ *for* $\nu^{\otimes n}$-a.e. $y_{1:n} \in \mathsf{Y}^n$ *and* λ-a.e. $\theta \in \Theta$.

Then (i) implies

(a) All closed sets $A \in \mathcal{T}$ that do not contain $\theta_\star$ are approximately $\mathbb{P}^{\theta_\star}_{\pi_\star}$-remote.

and (i)–(iii) imply

(b) For all initial distributions $\eta_\star$, $\lambda\langle Y_{1:n}\rangle \underset{n\to\infty}{\Longrightarrow} \delta_{\theta_\star}$, $\mathbb{P}^{\theta_\star}_{\eta_\star}$-a.s.

Proof. If (42) holds, then, setting $\mathbb{P} = \mathbb{P}^{\theta_\star}_{\pi_\star}$, Proposition 9 and (**B2**) give immediately that $A \cap K$ is $\mathbb{P}$-remote for all closed sets A not containing $\theta_\star$. Hence, in order to establish (a), we only need to check that (**B7**) implies (42). For this purpose, assume (**B7**) and set $R^* := R^{\theta_\star}_{\eta}$, which then is $\mathbb{P}$-a.s. positive Then $\mathbb{E}^{\theta_\star}_{\pi_\star}[R^*] = 1$ and

$$\frac{p_{\theta_\star,\eta}(Y_{1:n})}{p_{\theta_\star,\pi_\star}(Y_{1:n})} = \mathbb{E}^{\theta_\star}_{\pi_\star}\left[R^* \mid Y_{1:n}\right] \quad \mathbb{P}\text{-a.s.}, \tag{43}$$

from which we conclude that the left hand side converges to $R^* > 0$ $\mathbb{P}$-a.s. and in L^1 (see, e.g., [22, Theorem 27.3]). This implies (42) and (a) follows.

We now assume, additionally, (ii) and (iii). Then, by Theorem 7 and Remark 10, it suffices, in order to obtain (31), to check that $\lambda_n = \lambda\langle Y_{1:n}\rangle$ is, $\mathbb{P}$-a.s., a well-defined probability for n large enough. By (i) there is a compact set K such that $\int_{K^c} \lambda(\mathrm{d}\theta)\, p_{\theta,n} < \infty$ for n large enough $\mathbb{P}$-a.s., and since λ is a Radon measure, it also holds that $\int_K \lambda(\mathrm{d}\theta)\, p_{\theta,n} < \infty$ ν_n-almost everywhere and thus $\mathbb{P}$-a.s. Hence $\int_\Theta \lambda(\mathrm{d}\theta)\, p_{\theta,n} < \infty$ for n large enough, $\mathbb{P}$-a.s. By Remark 9, it only remains to check that $\int_\Theta \lambda(\mathrm{d}\theta)\, p_{\theta,n} > 0$ $\mathbb{P}$-a.s., which is directly implied by (iii).

Hence, we get that $\lambda\langle Y_{1:n}\rangle$ is a well-defined probability for n large enough $\mathbb{P}$-a.s. This yields (31), which corresponds to Assertion (b) in the special case $\eta_\star = \pi_\star$. Under (**B7**), this also implies Assertion (b) for all initial distributions $\eta_\star$. □

5.2. Proof of Theorem 1

Before proving Theorem 1, we need two preliminary results. The first is a change of probability formula (see Lemma 12 below), which will be used for extending the posterior consistency property to a non-stationary sequence. The second result allows (**A1**) to be checked in order to apply Theorem 7 (see Lemma 13 below).

Lemma 12 *If a fdPOMM satisfies (**B1**) and (**B3**), then it also satisfies (**B7**).*

Proof. Let $\theta \in \Theta$ and η be an initial distribution on $(\mathsf{Z}, \mathcal{Z})$. Under $\mathbb{P}^{\theta}_{\eta}$, $(Z_k)_{k\in\mathbb{N}^*}$ is a Markov chain with transition kernel $\mathbf{Q}_\theta$ and initial distribution having the density

$$z_1 \mapsto \eta_1(z_1) = \int \eta(\mathrm{d}z)\, q_\theta(z, z_1)$$

with respect to $\mu \otimes \nu$. Similarly, under $\mathbb{P}^{\theta}_{\pi_\theta}$, $(Z_k)_{k\in\mathbb{N}^*}$ is a Markov chain with the same transition kernel $\mathbf{Q}_\theta$ as above but with another initial distribution having the density

$$z_1 \mapsto \pi_\theta(z_1) = \int \pi_\theta(\mathrm{d}z)\, q_\theta(z, z_1)$$

with respect to $\mu \otimes \nu$.

Under (**B3**), these two densities are positive and $z_{1:\infty} \mapsto \eta_1(z_1)/\pi_\theta(z_1)$ is thus the Radon-Nikodym ratio between $\mathbb{P}^{\theta}_{\eta}$ and $\mathbb{P}^{\theta}_{\pi_\theta}$ restricted to $\sigma(Z_{1:\infty})$. The result follows. □

Lemma 13 *Consider a fdPOMM satisfying (**B1**). Then Assumptions (**B3**) and (**B4**) imply (**A1**) with* $p_{\theta,n} = p_{\theta,\eta}(Y_{1:n})$, $p^*_n = p_{\theta_\star,\pi_\star}(Y_{1:n})$, $\mathbb{P} = \mathbb{P}^{\theta_\star}_{\pi_\star}$ *and* $\Theta_\delta = \{\theta \in \Theta : \Delta(\theta_\star, \theta) \le \delta\}$.

Proof. Pick $\delta > 0$ and take any $\theta \in \Theta$ such that $\Delta(\theta_\star, \theta) \le \delta$. To prove the result, we need to show that

$$\liminf_{n\to\infty} n^{-1} \log \frac{p_{\theta,n}}{p^*_n} \ge -\delta \quad \mathbb{P}\text{-a.s.}, \tag{44}$$

where $p_{\theta,n} = p_{\theta,\eta}(Y_{1:n})$ and $p^*_n = p_{\theta_\star,\pi_\star}(Y_{1:n})$. We apply Lemma 8 with $\Omega = \mathsf{Z}^{\mathbb{N}}$, $\mathcal{F} = \mathcal{Z}^{\otimes\mathbb{N}}$, $\overline{\mathcal{F}}_n = \sigma(Z_{1:n})$, $\mathbb{P} = \mathbb{P}^{\theta_\star}_{\pi_\star}$ and $\overline{\nu}_n$ given (as in (56)) by

$$\overline{\nu}_n(B) = (\mu \otimes \nu)^{\otimes n}(A), \;\; B \in \overline{\mathcal{F}}_n \text{ with } B = [Z_{1:n}]^{-1}(A) \text{ and } A \in \mathcal{Z}^{\otimes n}.$$

Then $\overline{p}_n$ and $\overline{p}^*_n$ are the corresponding densities

$$\overline{p}_n = \int q_\theta(z_0, Z_1)\, \eta(\mathrm{d}z_0) \prod_{k=1}^{n-1} q_\theta(Z_k, Z_{k+1}),$$

and

$$\overline{p}^*_n = \pi_\star(Z_1) \prod_{k=1}^{n-1} q_{\theta_\star}(Z_k, Z_{k+1}).$$

Moreover, set $\tilde{\mathcal{F}}_n = \sigma(Y_{1:n})$, so that ν_n is the restriction of $\overline{\nu}_n$ to $\tilde{\mathcal{F}}_n$. In this case that the densities introduced in (32) are $\tilde{p}_n = p_{\theta,n}$ and $\tilde{p}^*_n = p^*_n$. Thus, applying Lemma 8, we get that (44) is implied by

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{\overline{p}_n}{\overline{p}^*_n} \ge -\delta \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.} \tag{45}$$

Now, observe that, $\mathbb{P}^{\theta_\star}_{\pi_\star}$-a.s.,

$$\log \frac{\overline{p}_n}{\overline{p}^*_n} = \log \frac{\int q_\theta(z_0, Z_1)\, \eta(\mathrm{d}z_0)}{\pi_\star(Z_1)} + \sum_{\ell=1}^{n-1} \log \frac{q_\theta(Z_\ell, Z_{\ell+1})}{q_{\theta_\star}(Z_\ell, Z_{\ell+1})}.$$

Since the transition density q_θ is assumed to be positive, the first term is a finite number and tends to zero when divided by n, $\mathbb{P}^{\theta_\star}_{\pi_\star}$-a.s. Moreover, by (**B1**), Z is ergodic under $\mathbb{P}^{\theta_\star}_{\pi_\star}$, and we obtain

$$\liminf_{n\to\infty} n^{-1} \sum_{\ell=1}^{n-1} \log \frac{q_\theta(Z_\ell, Z_{\ell+1})}{q_{\theta_\star}(Z_\ell, Z_{\ell+1})} = -\Delta(\theta_\star, \theta) \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.}$$

Since we have assumed that $\Delta(\theta_\star, \theta) \le \delta$, (45) holds true and the proof is completed. □

The proof of Theorem 1 is now completed by observing that Lemmas 12 and 13 show that the assumptions of Theorem 1 imply those of Proposition 11.

Note that (i) in Proposition 11 is trivially satisfied when Θ is compact, and that (iii) directly follows from (**B3**).

5.3. *Proof of Theorem 2*

The only modification of the proof consists in observing that Condition (iii) in Proposition 11 now directly follows from (**C1**–**2**) and in showing that the conclusions of Lemma 12 and Lemma 13 hold true under the new set of assumptions. This is done in Lemma 14 and Lemma 15 below.

Lemma 14 *If a fdPOMM satisfies (**B1**), (**C1**), (**C2**) and (**C4**), then it also satisfies (**B7**).*

Proof. Let $\theta \in \Theta$ and η be an initial distribution on $(\mathsf{Z}, \mathcal{Z})$. Then under Assumptions (**B1**), (**C1**) and (**C4**) there exists a sequence $((X'_k, X''_k, Y_k))_{k\in\mathbb{N}}$ such that

1. $((X'_k, Y_k))_{k\in\mathbb{N}}$ is distributed according to $\mathbb{P}^\theta_{\pi_\theta}$,
2. $((X''_k, Y_k))_{k\in\mathbb{N}}$ is distributed according to $\mathbb{P}^\theta_\eta$,
3. there is a $\mathbb{P}^\theta_{\pi_\theta}$-a.s. finite stopping time τ such that $X'_k = X''_k$ for all $k > \tau$.

See [35, Lemma 3.7] and also the end of the proof of Lemma 15 where the same construction is used. Then, using (**C2**) and mimicking [35, Lemma 3.7] yields that the laws of $((X'_k, Y_k))_{k\in\mathbb{N}}$ and $((X''_k, Y_k))_{k\in\mathbb{N}}$ are equivalent, which in its turn imply (**B7**). □

Lemma 15 *Consider a fdPOMM with initial distribution η on $(\mathsf{Z}, \mathcal{Z})$. Assume (**B1**) and (**C1**) and set $\mathbb{P} = \mathbb{P}^{\theta_\star}_{\pi_\star}$. Then (**C2**–**4**) imply (**A1**) with $p_{\theta,n} = p_{\theta,\eta}(Y_{1:n})$, $p^*_n = p_{\theta_\star,\pi_\star}(Y_{1:n})$, $\mathbb{P} = \mathbb{P}^{\theta_\star}_{\pi_\star}$ and $\Theta_\delta = \{\theta \in \Theta : \overline{\Delta}(\theta_\star, \theta) \le \delta\}$.*

Proof. Pick $\delta > 0$ and take any $\theta \in \Theta$ such that $\overline{\Delta}(\theta_\star, \theta) \le \delta$. By (**C3**) it is sufficient to show that

$$\liminf_{n\to\infty} n^{-1} \log \frac{p_{\theta,n}}{p^*_n} \ge -\delta \quad \mathbb{P}\text{-a.s.}, \tag{46}$$

where $p_{\theta,n} = p_{\theta,\eta}(Y_{1:n})$ and $p_n^* = p_{\theta_\star,\pi_\star}(Y_{1:n})$. The idea of the proof is now similar to that of Lemma 13, but with a completely different augmented set of variables. Instead of augmenting the data by just the unobserved sequence (X_k), we now add one more sequence (X'_k) to the data as follows. Let $\mathsf{Z}' = \mathsf{X}^2 \times \mathsf{Y}$ and $\mathcal{Z}' = \mathcal{X}^{\otimes 2} \otimes \mathcal{Y}$ and denote, for all $k \in \mathbb{N}$, by $Z_k = (X_k, Y_k)$ and $Z'_k = (X_k, X'_k, Y_k)$ the members of the corresponding canonical sequences. We define $\mathbb{P}$ as the distribution of $(Z'_n)_{n\in\mathbb{N}}$ when $(Z_n)_{n\in\mathbb{N}}$ is distributed according to $\mathbb{P}_{\pi_\star}^{\theta_\star}$, $(X'_n)_{n\in\mathbb{N}}$ is the canonical Markov chain with initial distribution η and kernel $\mathbf{K}_\theta$ and, moreover, $(X'_n)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ are independent.

We apply Lemma 8 with $\Omega = \mathsf{Z}'^{\mathbb{N}}$, $\mathcal{F} = \mathcal{Z}'^{\otimes\mathbb{N}}$, $\overline{\mathcal{F}}_n = \sigma(Z'_{1:n})$ and $\overline{\nu}_n$ given (as in (56)) by

$$\overline{\nu}_n(B) = (\mu^{\otimes 2} \otimes \nu)^{\otimes n}(A), \ B \in \overline{\mathcal{F}}_n \text{ with } B = [Z'_{1:n}]^{-1}(A) \text{ and } A \in \mathcal{Z}'^{\otimes n}.$$

In this particular setting, $\overline{p}_n^*$ takes the form

$$\overline{p}_n^* = \pi_\star(Z_1) \int \mathbf{K}_\theta(x'_0, X'_1)\, \eta(\mathrm{d}x'_0) \prod_{k=1}^{n-1} q_{\theta_\star}(Z_k, Z_{k+1}) k_\theta(X'_k, X'_{k+1}).$$

Now, define $\mathbb{P}_\theta$ as the distribution of $(Z'_n)_{n\in\mathbb{N}}$ when $(X_n)_{n\in\mathbb{N}}$ and $(X'_n)_{n\in\mathbb{N}}$ are distributed exactly as under $\mathbb{P}$ (i.e., two independent Markov chains with kernels $\mathbf{K}_{\theta_\star}$ and $\mathbf{K}_\theta$ and initial distributions $\pi_\star$ and η, respectively), but, conditionally on these sequences, $(Y_n)_{n\in\mathbb{N}}$ has the law of a sequence of independent random variables such that for all k, Y_k has density $g_\theta(X'_k, \cdot)$ with respect to ν. Hence, $((X'_k, Y_k))_{k\in\mathbb{N}}$ is an HMM with parameter θ and initial distribution η. Recall that we define $\overline{p}_n$ as the density of the distribution $\mathbb{P}_{\theta,n}$, which in its turn is defined as the restriction of $\mathbb{P}_\theta$ on $\overline{\mathcal{F}}_n$. Consequently,

$$\begin{aligned}\overline{p}_n = \pi_\star(X_1, \mathsf{Y}) \int \mathbf{K}_\theta(x'_0, X'_1)\, \eta(\mathrm{d}x'_0) \\ \times \left(\prod_{k=1}^{n-1} k_{\theta_\star}(X_k, X_{k+1}) k_\theta(X'_k, X'_{k+1}) g_\theta(X'_k, Y_k) \right) g_\theta(X'_n, Y_n),\end{aligned}$$

where $\pi_\star(\cdot, \mathsf{Y})$ denotes the marginal of the density $\pi_\star$ w.r.t. the X component. It now holds that

$$\frac{\overline{p}_n}{\overline{p}_n^*} = \prod_{k=1}^{n} \frac{g_\theta(X'_k, Y_k)}{g_{\theta_\star}(X_k, Y_k)}.$$

Now, set $\tilde{\mathcal{F}}_n = \sigma(Y_{1:n})$; then, defining $\tilde{p}_n$ and $\tilde{p}_n^*$ as in (32) yields straightforwardly that $\tilde{p}_n = p_{\theta,n}$ and $\tilde{p}_n^* = p_n^*$. Hence by Lemma 8, Eqn. (44) follows from

$$\liminf_{n\to\infty} n^{-1} \log \frac{\overline{p}_n}{\overline{p}_n^*} \geq -\delta \quad \mathbb{P}\text{-a.s.}, \tag{47}$$

whose establishment is the object of the remainder of the proof.

We prove (47) by means of a coupling argument used in [35, Lemma 3.7]. Recall that under $\mathbb{P}$, $(Z_n)_{n\in\mathbb{N}}$ and $(X'_n)_{n\in\mathbb{N}}$ are independent. Thus, by Condition (**C4**), following [27, Theorem III.14.10], we extend $\mathbb{P}$ by adding an X-valued

process $(X''_n)_{n\in\mathbb{N}}$ to $(Z'_n)_{n\in\mathbb{N}}$ such that $(X''_n)_{n\in\mathbb{N}}$ is independent of $(Z_n)_{n\in\mathbb{N}}$, has distribution $\mathbb{P}^{\theta}_{\pi_\theta}$ and

$$\tau := \min\{k \in \mathbb{N} : X''_\ell = X'_\ell \text{ for all } \ell \geq k\} < \infty \quad \mathbb{P}\text{-a.s.}$$

Then by (**C2**) we have, for all $n \geq \tau$,

$$\frac{\overline{p}_n}{\overline{p}^*_n} = \left[\prod_{k=1}^{\tau} \frac{g_\theta(X'_k, Y_k)}{g_{\theta_\star}(X_k, Y_k)} \prod_{k=1}^{\tau} \frac{g_{\theta_\star}(X_k, Y_k)}{g_\theta(X''_k, Y_k)}\right] \prod_{k=1}^{n} \frac{g_\theta(X''_k, Y_k)}{g_{\theta_\star}(X_k, Y_k)},$$

where the term within the brackets is positive $\mathbb{P}$-a.s. Now, by Lemma 20, $((X_k, Y_k, X''_k))_{k\in\mathbb{N}}$ is a stationary ergodic Markov chain under $\mathbb{P}$. Hence,

$$\lim_{n\to\infty} n^{-1} \log \frac{\overline{p}_n}{\overline{p}^*_n} = \mathbb{E}\left[\log \frac{g_\theta(X''_0, Y_0)}{g_{\theta_\star}(X_0, Y_0)}\right] \quad \mathbb{P}\text{-a.s.},$$

where $\mathbb{E}$ denotes the expectation under $\mathbb{P}$. To conclude, we observe that the latter expectation is exactly minus $\overline{\Delta}(\theta_\star, \theta)$ as defined in (7), and thus the choice of θ at the beginning of this proof gives (47). □

5.4. *Proof of Theorem 3*

Since Condition (i) in Proposition 11 is trivially satisfied in the case where Θ is assumed to be compact, the proofs of Theorem 1 and Theorem 2 only required (**B7**) and Condition (ii) of Proposition 11 to be checked. The latter two assumptions are still implied by those of Theorem 3; however, since Θ is no longer assumed to be compact, it remains to show that Condition (i) in Proposition 11 still holds under the assumptions of Theorem 3. This will be done in Proposition 18 below.

We preface this result with two lemmas.

Lemma 16 *Consider a fdPOMM satisfying (**B1**) and let λ be a positive measure on $(\Theta, \mathcal{T})$. Then each condition in (**B6**) implies the same condition with n_0 replaced by any $n \geq n_0$.*

Proof. Let $n > n_0$. Using that for all $y_{1:n_0} \in \mathsf{Y}^{n_0}$,

$$\iint \lambda(\mathrm{d}\theta)\, p_{\theta,\eta}(y_{1:n})\, \nu^{\otimes(n-n_0)}(\mathrm{d}y_{n_0+1:n}) = \int \lambda(\mathrm{d}\theta)\, p_{\theta,\eta}(y_{1:n_0}), \tag{48}$$

and that, under $\mathbb{P}^{\theta_\star}_{\pi_\star}$, $p_{\theta_\star,\pi_\star}(y_{n_0+1:n} \mid Y_{1:n_0}) = p_{\theta_\star,\pi_\star}(Y_{1:n_0}, y_{n_0+1:n})/p_{\theta_\star,\pi_\star}(Y_{1:n_0})$ is the density, with respect to $\nu^{\otimes(n-n_0)}$, of the conditional distribution of $Y_{n_0+1:n}$ given $Y_{1:n_0}$, we get that

$$\mathbb{E}^{\theta_\star}_{\pi_\star}\left[\frac{\int \lambda(\mathrm{d}\theta)\, p_{\theta,\eta}(Y_{1:n})}{p_{\theta_\star,\pi_\star}(Y_{n_0+1:n} \mid Y_{1:n_0})}\middle| Y_{1:n_0}\right] = \int \lambda(\mathrm{d}\theta)\, p_{\theta,\eta}(Y_{1:n_0}).$$

Thus, by Lemma 19, if (11) holds, it also holds with n replacing n_0.

Now, concerning (12), the comments before [2, Theorem 1] show that, using that the observations are stationary under $\mathbb{P}^{\theta_\star}_{\pi_\star}$, $(\mathbb{E}^{\theta_\star}_{\pi_\star}[\log p_{\theta_\star,\pi_\star}(Y_{n+1} \mid Y_{1:n})])_{n\in\mathbb{N}^*}$ is a nondecreasing sequence. Hence if (12) holds, it continues to hold true when n_0 is replaced by any $n \geq n_0$. □

Lemma 17 *Consider a fdPOMM satisfying (**B1**) and (**B5**). Then for all $\delta > 0$ there exists a compact $K \in \mathcal{T}$ such that*

$$\limsup_{n\to\infty} n^{-1} \log \hat{p}_{K^c}(Y_{0:n}) \leq -\delta \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.},$$

where $\hat{p}_{K^c}(Y_{0:n})$ is defined in (8).

Proof. We first show that

$$\limsup_{m\to\infty} \mathbb{E}^{\theta_\star}_{\pi_\star}[\log \hat{p}_{C_m^c}(Y_{0:\ell})] = -\infty. \tag{49}$$

Set $U \coloneqq \log^+ \hat{p}_\Theta(Y_{0:\ell})$ and $U_m \coloneqq \log \hat{p}_{C_m^c}(Y_{0:\ell})$. First note that $\mathbb{E}^{\theta_\star}_{\pi_\star}[U_m]$ is well-defined since by (10) in (**B5**), $\mathbb{E}^{\theta_\star}_{\pi_\star}[U_m^+] \leq \mathbb{E}^{\theta_\star}_{\pi_\star}[U] < \infty$. Now, since $U - U_m \geq 0$, Fatou's lemma yields

$$\begin{aligned}\liminf_{m\to\infty} \mathbb{E}^{\theta_\star}_{\pi_\star}[U - U_m] &= \mathbb{E}^{\theta_\star}_{\pi_\star}[U] - \limsup_{m\to\infty} \mathbb{E}^{\theta_\star}_{\pi_\star}[U_m] \\ &\geq \mathbb{E}^{\theta_\star}_{\pi_\star}\left[\liminf_{m\to\infty}(U - U_m)\right] = \mathbb{E}^{\theta_\star}_{\pi_\star}[U] - \mathbb{E}^{\theta_\star}_{\pi_\star}\left[\limsup_{m\to\infty} U_m\right].\end{aligned}$$

Combining with (9) in (**B5**) yields

$$\limsup_{m\to\infty} \mathbb{E}^{\theta_\star}_{\pi_\star}[\log \hat{p}_{C_m^c}(Y_{0:\ell})] \leq \mathbb{E}^{\theta_\star}_{\pi_\star}\left[\limsup_{m\to\infty} \log \hat{p}_{C_m^c}(Y_{0:\ell})\right] = -\infty,$$

and (49) is shown. Now, let $\delta > 0$. According to (49), one may pick $m \in \mathbb{N}$ sufficiently large such that

$$\ell^{-1}\mathbb{E}^{\theta_\star}_{\pi_\star}[\log \hat{p}_{C_m^c}(Y_{0:\ell})] \leq -\delta.$$

Now, set $K \coloneqq C_m$ and define, for $(r,s) \in \mathbb{N}^2$ such that $r \leq s$, $W_{r,s} \coloneqq \hat{p}_{K^c}(Y_{r:s})$. By (10) in (**B5**), $\mathbb{E}[\log^+ W_{0,\ell}] < \infty$ and for all $r \leq s \leq t$,

$$W_{r,t} \leq W_{r,s} W_{s,t}.$$

Since under $\mathbb{P}^{\theta_\star}_{\pi_\star}$, the sequence $(Y_n)_{n\in\mathbb{N}}$ is stationary and ergodic, the Kingman subadditive theorem ([24]) applies. Thus, $\lim_{n\to\infty} n^{-1} \log W_{0,n}$ exists $\mathbb{P}^{\theta_\star}_{\pi_\star}$-a.s. and

$$\lim_{n\to\infty} n^{-1} \log W_{0,n} = \inf_{n\geq\ell} n^{-1}\mathbb{E}^{\theta_\star}_{\pi_\star} \log W_{0,n} \leq \ell^{-1}\mathbb{E}^{\theta_\star}_{\pi_\star}[\log \hat{p}_{K^c}(Y_{0:\ell})] \leq -\delta \quad \mathbb{P}^{\theta_\star}_{\pi_\star}\text{-a.s.}$$

The proof is completed. □

Proposition 18 *Consider a fdPOMM satisfying (**B1**), (**B5**) and (**B6**). Assume (**A1**) with $p_{\theta,n} = p_{\theta,\eta}(Y_{1:n})$, $p_n^* = p_{\theta_\star,\pi_\star}(Y_{1:n})$ and $\mathbb{P} = \mathbb{P}_{\pi_\star}^{\theta_\star}$. Then there exists a compact set $K \in \mathcal{T}$ such that*

$$\limsup_{n\to\infty} n^{-1} \log \lambda\langle Y_{1:n}\rangle(K^c) < 0 \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.}$$

Proof. Following [2], let us define, for all $n \in \mathbb{N}^*$,

$$V_n \coloneqq p_{\theta_\star,\pi_\star}(Y_{1:n}), \quad \tilde{V}_n \coloneqq \mathbb{E}_{\pi_\star}^{\theta_\star}[\log(V_n/V_{n-1})].$$

By (12) in (**B6**), we have $\tilde{V}_{n_0} > -\infty$. As explained in the comments before [2, Theorem 1], $(\tilde{V}_n)_{n\in\mathbb{N}^*}$ is a non-decreasing sequence, and denoting by $\tilde{V}$ its limit in $(-\infty, \infty]$, by [2, Theorem 1], we have

$$\lim_{n\to\infty} n^{-1} \log V_n = \tilde{V} \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.} \tag{50}$$

Pick $\delta > 0$ and $\epsilon > 0$ such that

$$-\delta - \tilde{V} + \epsilon < 0.$$

According to Lemma 17, there exists a compact set $K \in \mathcal{T}$ such that

$$\limsup_{n\to\infty} n^{-1} \log \hat{p}_{K^c}(Y_{0:n}) \le -\delta \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.} \tag{51}$$

Now, write for all $n \in \mathbb{N}$ strictly larger than n_0,

$$\lambda\langle Y_{1:n}\rangle(K^c) \le \left(\int_{K^c} p_{\theta,\eta}(Y_{1:n_0})\lambda(\mathrm{d}\theta)\right) \frac{\hat{p}_{K^c}(Y_{n_0:n})}{p_{\lambda,\eta}(Y_{1:n})} \le p_{\lambda,\eta}(Y_{1:n_0}) \frac{U_n}{V_n W_n}, \tag{52}$$

where

$$U_n \coloneqq \hat{p}_{K^c}(Y_{n_0:n}), \quad W_n \coloneqq p_{\lambda,\eta}(Y_{1:n})/p_{\theta_\star,\pi_\star}(Y_{1:n}).$$

By the first condition in (**B6**),

$$\limsup_{n\to\infty} n^{-1} \log p_{\lambda,\eta}(Y_{1:n_0}) = 0 \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.} \tag{53}$$

Moreover, according to Lemma 10, (**A1**) implies

$$\liminf_{n\to\infty} n^{-1} \log W_n \ge -\epsilon \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.} \tag{54}$$

Finally, combining (50), (51), (53), (54) with (52) yields:

$$\limsup_{n\to\infty} n^{-1} \log \lambda\langle Y_{1:n}\rangle(K^c) \le -\delta - \tilde{V} + \epsilon < 0 \quad \mathbb{P}_{\pi_\star}^{\theta_\star}\text{-a.s.}$$

□

6. Conclusion

We have established that the posterior consistency for fdPOMMs is a consequence of the consistency of the AMLE under—what we believe—minimal assumptions that can be checked for a variety of models used in practice. Importantly, our assumptions can be checked for models where the state space of the latent process is non-compact, which is most often the case in applications (including, e.g., the linear Gaussian state-space models). Moreover, we allow also the parameter space to be non-compact, which is essential in the Bayesian setting (where many prior distributions of fundamental importance have infinite support), and the prior to be improper. Thus, our results generalize substantially existing results in this direction, which focus exclusively on the special case of HMMs and require both the state space of the hidden chain and the parameter space to be compact. Our proofs rely on a machinery revolving around the general concept of $\mathbb{P}$-remoteness introduced in Section 4, which is naturally linked to the consistency of the MLE in the frequentist setting. As far as known to the authors, this link has not been explored before in the literature.

Our analysis relies substantially on the assumption that the model is fully dominated, which is certainly a restriction. A natural direction for research is hence the relaxation of this assumption, and incorporating techniques developed in [14] into the analysis could allow our results to be extended to observation-driven models (including the GARCH framework). Moreover, as we do not provide any rate of convergence, supplementing our results with a Bernstein-von Mises-type theorem would of course be desirable; nevertheless, establishing such a theorem involves typically, *inter alia*, a law of large numbers for the observed Fisher information, which appears to be a real challenge in the non-compact setting (the proof of a Bernstein-von Mises-type theorem is hence expected to be at least on the same level of difficulty as the proof of the asymptotic normality of the MLE, which is still an open problem in the non-compact case). Finally, another natural topic for future research is the extension of our results in the direction of nonparametric Bayesian modeling.

Appendix A: Some kernel notation

Let μ be a signed measure on some measurable space $(\mathsf{X}, \mathcal{X})$. For any $|\mu|$-integrable function h, we denote by

$$\mu h := \int h(x)\, \mu(\mathrm{d}x)$$

the Lebesgue integral of h w.r.t. μ.

In addition, let $(\mathsf{Y}, \mathcal{Y})$ be some other measurable space and $\mathbf{K}$ some possibly unnormalized transition kernel $\mathbf{K} : \mathsf{X} \times \mathcal{Y} \to \mathbb{R}_+$. The kernel $\mathbf{K}$ induces two integral operators, one acting on functions and the other on measures. More specifically, given a measure ν on $(\mathsf{X}, \mathcal{X})$ and a measurable function h on $(\mathsf{Y}, \mathcal{Y})$,

we define the measure

$$\nu\mathbf{K} : \mathcal{Y} \ni A \mapsto \int \mathbf{K}(x, A)\, \nu(\mathrm{d}x)$$

and the function

$$\mathbf{K}h : \mathsf{X} \ni x \mapsto \int h(y)\, \mathbf{K}(x, \mathrm{d}y)$$

whenever these quantities are well-defined. For the latter, we will, whenever convenient, use also the alternative notation $\mathbf{K}(\cdot, h)$.

Finally, given a third measurable space $(\mathsf{Z}, \mathcal{Z})$ and another kernel $\mathbf{L} : \mathsf{Y} \times \mathcal{Z} \to \mathbb{R}_+$ we define, with $\mathbf{K}$ as above, the *product kernel*

$$\mathbf{KL} : \mathsf{X} \times \mathcal{Z} \ni (x, B) \mapsto \int \mathbf{L}(y, B)\, \mathbf{K}(x, \mathrm{d}y),$$

whenever this is well-defined. When $\mathbf{K}$ describes transitions within the *same* space $(\mathsf{X}, \mathcal{X})$, its *iterates* are defined inductively by

$$\mathbf{K}^0(x, \cdot) \coloneqq \delta_x \text{ for all } x \in \mathsf{X} \text{ and } \mathbf{K}^n \coloneqq \mathbf{K}^{n-1}\mathbf{K} \text{ for all } n \in \mathbb{N}^*.$$

Appendix B: Dominated i.i.d. model

The general setting in 4.1 is a bit unusual in the sense that all densities and dominating measures are defined directly on the same space Ω (endowed however with different σ-fields picked among the members of the sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$). The advantage is to avoid writing all likelihoods as functions of the observations, yielding more compact expressions and arguments. In order to illustrate better how this setting can be used in practice, we describe it in the simple i.i.d. case.

Consider the dominated i.i.d. case comprising an n-sample $Y_{1:n} = (Y_1, \ldots, Y_n)$ of i.i.d. observations taking on values in $(\mathsf{Y}, \mathcal{Y})$ and having marginal density $q_{\theta_\star}$, depending on some unknown parameter $\theta_\star \in \Theta$, w.r.t. a given σ-finite dominating measure ν on $(\mathsf{Y}, \mathcal{Y})$. In this case, we set $(Y_n)_{n\in\mathbb{N}^*}$ as the canonical process defined on $\Omega = \mathsf{Y}^{\mathbb{N}^*}$ endowed with $\mathcal{F} = \mathcal{Y}^{\otimes\mathbb{N}^*}$, the σ-field generated by cylinder sets. Then $\mathcal{F}_n = \sigma(Y_{1:n})$ and $p_{\theta,n} = q_\theta^{\otimes n}(Y_{1:n})$, where $q_\theta^{\otimes n}$ denotes the nth self tensor product of q_θ defined on Y^n. While $q_\theta^{\otimes n}$ is a density w.r.t. the product measure $\nu^{\otimes n}$, $p_{\theta,n}$ is a density with respect to the σ-finite measure ν_n given by

$$\nu_n(B) = \nu^{\otimes n}(A), \;\; B \in \mathcal{F}_n \text{ with } B = [Y_{1:n}]^{-1}(A) \text{ and } A \in \mathcal{Y}^{\otimes n}. \tag{55}$$

In the i.i.d. case, one assumes that there is a true parameter $\theta_\star \in \Theta$ and $\mathbb{P}$ is the corresponding distribution of $(Y_n)_{n\in\mathbb{N}^*}$, implying that $p_n^* = p_{\theta_\star,n}$.

Note that the $\mathcal{F}_n \to \mathcal{Y}^{\otimes n}$-mapping defined by $B \mapsto Y_{1:n}(B) = \{Y_{1:n}(\omega), \omega \in B\}$ is bijective and can be seen as the (unique) reciprocal of the preimage mapping $[Y_{1:n}]^{-1} : \mathcal{Y}^{\otimes n} \to \mathcal{F}_n$. Hence ν_n in (55) can be equivalently defined as

$$\nu_n = \nu^{\otimes n} \circ Y_{1:n} \text{ on } \mathcal{F}_n \iff \nu_n \circ [Y_{1:n}]^{-1} = \nu^{\otimes n} \text{ on } \mathcal{Y}^{\otimes n}. \tag{56}$$

Appendix C: Postponed proof of general results

C.1. Proof of Proposition 6

Let $A \in \mathcal{T}$ and suppose that $B_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$ and that (27) and (28) hold. Then there exists $\rho > 1$ such that

$$\mathbb{E}\left(\sum_{n=1}^{\infty} \rho^n \int_A \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} \mathbb{1}_{B_n}\right) = \sum_{n=1}^{\infty} \rho^n \int_A \lambda(\mathrm{d}\theta)\, \mathbb{P}_{\theta,n}(B_n) < \infty.$$

This implies

$$\sum_{n=1}^{\infty} \rho^n \int_A \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} \mathbb{1}_{B_n} < \infty, \quad \mathbb{P}\text{-a.s.}$$

Now, the set $\Omega_0 = \liminf_{n\to\infty} B_n \in \mathcal{F}$ of $\mathbb{P}$-probability one in (28) is such that for all $\omega \in \Omega_0$, $\{n \in \mathbb{N} : \mathbb{1}_{B_n}(\omega) = 0\}$ is finite. Thus, the series $\sum_{n=1}^{\infty} \rho^n \int_A \lambda(\mathrm{d}\theta)\, p_{\theta,n}/p_n^*$ is convergent $\mathbb{P}$-a.s., establishing that A is $\mathbb{P}$-remote.

Conversely, if $A \in \mathcal{T}$ is $\mathbb{P}$-remote, then choose $\rho > 1$ such that

$$\limsup_{n\to\infty} n^{-1} \log\left(\int_A \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*}\right) \le -\log\rho < 0 \quad \mathbb{P}\text{-a.s.} \tag{57}$$

Pick $\tilde{\rho} \in (1, \rho)$ and $\varrho \in (1/\tilde{\rho}, 1)$. Set

$$B_n \coloneqq \left\{\int_A \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} \le \varrho^n\right\}.$$

Then $B_n \in \mathcal{F}_n$ and, by Tonelli's theorem,

$$\int_A \lambda(\mathrm{d}\theta)\, \mathbb{P}_{\theta,n}(B_n) \le \varrho^n \mathbb{P}(B_n) \le \varrho^n \quad \mathbb{P}\text{-a.s.};$$

thus, (27) is satisfied. Since $\tilde{\rho} < \rho$, (57) implies

$$\sup_{n\in\mathbb{N}} \tilde{\rho}^n \int_A \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} < \infty \quad \mathbb{P}\text{-a.s.},$$

so that

$$\mathbb{P}(\{\omega \in \Omega : \omega \in B_n^c \text{ i.o.}\}) = \mathbb{P}\left(\tilde{\rho}^n \int_A \lambda(\mathrm{d}\theta)\, \frac{p_{\theta,n}}{p_n^*} > (\tilde{\rho}\varrho)^n \text{ i.o.}\right) = 0,$$

and (28) holds.

Appendix D: Useful lemmas

Lemma 19 *Let Z be a $[0,\infty]$-valued variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. If $\mathbb{E}[Z|\mathcal{G}] < \infty$ $\mathbb{P}$-a.s., then $Z < \infty$ $\mathbb{P}$-a.s.*

Proof. Let $B = \{Z = \infty\}$. Then we have

$$\mathbb{E}\left[Z \,|\, \mathcal{G}\right] \geq \mathbb{E}\left[Z \mathbb{1}_B \,|\, \mathcal{G}\right] = \infty \text{ on } \left\{\mathbb{P}(B \,|\, \mathcal{G}) > 0\right\}.$$

Hence, if $\mathbb{E}\left[Z \,|\, \mathcal{G}\right] < \infty$ $\mathbb{P}$-a.s., then $\mathbb{P}(B \,|\, \mathcal{G}) = 0$ $\mathbb{P}$-a.s. and so $\mathbb{P}(B) = 0$. □

Lemma 20 *Let Q and Q' be two Markov kernels on $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{X}', \mathcal{X}')$, respectively. Let $\bar{Q}$ the Markov kernel on $(\mathsf{X}\times\mathsf{X}', \mathcal{X}\otimes\mathcal{X}')$ defined by, for all $A \in \mathcal{X}$ and $B \in \mathcal{X}'$,*

$$\bar{Q}((x, x'), A \times B) = Q(x, A)Q'(x, B).$$

Suppose that Q and Q' are ergodic with stationary distributions π and π', respectively, in the sense that for all initial distributions η and η' on $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{X}', \mathcal{X}')$, respectively, it holds that

$$\lim_{n\to\infty} \|\eta Q^n - \pi\|_{\mathrm{TV}} = 0 \text{ and } \lim_{n\to\infty} \|\eta' Q'^n - \pi'\|_{\mathrm{TV}} = 0.$$

Then $\bar{Q}$ is ergodic with stationary distribution $\pi \otimes \pi'$.

Proof. Let h be a measurable function on $\mathsf{X} \times \mathsf{X}'$ such that $|h| \leq 1$. For all $n \in \mathbb{N}^*$, we may write, for $(x, x') \in \mathsf{X} \times \mathsf{X}'$, $\bar{Q}^n h(x, x') - (\pi \otimes \pi')h$ as

$$\begin{aligned}\int \left(\int h(x_n, x'_n)\, Q^n(x, \mathrm{d}x_n) - \int h(x_n, x'_n)\, \pi(\mathrm{d}x_n) \right) Q'^n(x', \mathrm{d}x'_n) \\ + \int \left(\int h(x_n, x'_n)\, Q'^n(x', \mathrm{d}x'_n) - \int h(x_n, x'_n)\, \pi'(\mathrm{d}x'_n) \right) \pi(\mathrm{d}x_n),\end{aligned}$$

from which we immediately deduce that

$$\lim_{n\to\infty} \|\delta_{(x,x')} Q'^n - \pi \otimes \pi'\|_{\mathrm{TV}} = 0.$$

The ergodicity of $\bar{Q}$ follows. □

References

[1] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B*, 72(3):269–342, 2010.

[2] A. Barron. The strong ergodic theorem for densities; generalized Shannon-McMillan-Breiman theorem. *Ann. Probab.*, 13:1292–1303, 1985.

[3] A Barron. The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical report, Department of Statistics, University of Illinois, Champaign, Illinois, 1988.

[4] A. Barron, M. Chervish, and L. Wasserman. The consistency of posterior distributions in non parametric problems. *Ann. Statist.*, 27(2):536–561, 1999.

[5] L. E. Baum and T. P. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37:1554–1563, 1966.
[6] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.
[7] G. Churchill. Hidden Markov chains and the analysis of genome structure. *Computers & Chemistry*, 16(2):107–115, 1992.
[8] M. C. M. de Gunst and O. Shcherbakova. Asymptotic behavior of Bayes estimators for hidden Markov models with application to ion channels. *Mathematical Methods of Statistics*, 17(4):342–356, 2008.
[9] R. Douc, P. Doukhan, and E. Moulines. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Process. Appl.*, 123(7):2620–2647, 2013.
[10] R. Douc and C. Matias. Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli*, 7(3):381–420, 2001.
[11] R. Douc and E. Moulines. Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. *Ann. Statist.*, 40(5):2697–2732, 10 2012.
[12] R. Douc, E. Moulines, J. Olsson, and R. van Handel. Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, 39(1):474–513, 2011.
[13] R. Douc, E. Moulines, and T. Rydén. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.*, 32(5):2254–2304, 2004.
[14] R. Douc, F. Roueff, and T. Sim. The maximizing set of the asymptotic normalized log-likelihood for partially observed Markov chains. Technical report, Institut Mines-Telecom, 2015. To appear, preprint available at [HAL] or [arXiv].
[15] D. Fredkin and J. Rice. Correlation functions of a function of a finite-state Markov process with application to channel kinetics. *Math. Biosci.*, 87:161–172, 1987.
[16] E. Gassiat and J. Rousseau. About the posterior distribution in hidden Markov models with unknown number of states. *Bernoulli*, 20(4):2039–2075, 2014.
[17] V. Genon-Catalot, T. Jeantheau, and C. Larédo. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6(6):1051–1079, 12 2000.
[18] V. Genon-Catalot and C. Laredo. Leroux's method for general hidden Markov models. *Stochastic Process. Appl.*, 116(2):222–243, 2006.
[19] Subhashis Ghosal. A review of consistency and convergence of posterior distribution. In *Varanashi Symposium in Bayesian Inference, Banaras Hindu University*, 1997.
[20] N. Gordon, D. Salmond, and A. F. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.*, 140:107–113, 1993.
[21] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42:281–300, 1987.

[22] J. Jacod and P. Protter. *Probability Essentials*. Springer, 2000.
[23] B. Juang and L. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33:251–272, 1991.
[24] J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probability*, 1:883–909, 1973. With discussion by D. L. Burkholder, Daryl Daley, H. Kesten, P. Ney, Frank Spitzer and J. M. Hammersley, and a reply by the author.
[25] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New-York, 2nd edition, 1998.
[26] B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.*, 40:127–143, 1992.
[27] T. Lindvall. *Lectures on the Coupling Method*. Wiley, New-York, 1992.
[28] I. MacDonald and W. Zucchini. *Hidden Markov models for time series: an introduction using R*. Chapman, London, 2009.
[29] Rogemar S. Mamon and Robert J. Elliott. *Hidden Markov Models in Finance*, volume 104 of *International Series in Operations Research & Management Science*. Springer, Berlin, 2007.
[30] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
[31] A. Papavasiliou. Parameter estimation and asymptotic stability in stochastic filtering. *Stochastic Process. Appl.*, 116(7):1048–1065, 2006.
[32] T. Petrie. Probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 40:97–115, 1969.
[33] W. Pieczynski. Pairwise Markov chains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):634–639, 2003.
[34] Lorraine Schwartz. On bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4(1):10–26, 1965.
[35] R. van Handel. The stability of conditional Markov processes and Markov chains in random environments. *Ann. Probab.*, 37:1876–1925, 2009.
[36] E. Vernet. Posterior consistency for nonparametric hidden markov models with finite state space. *Electron. J. Statist.*, 9(1):717–752, 2015.