The Annals of Statistics
2012, Vol. 40, No. 5, 2697–2732
DOI: 10.1214/12-AOS1047

ASYMPTOTIC PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATION IN MISSPECIFIED HIDDEN MARKOV MODELS[1]

BY RANDAL DOUC AND ERIC MOULINES

Télécom SudParis and Télécom ParisTech

Let $(Y_k)_{k\in\mathbb{Z}}$ be a stationary sequence on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in a standard Borel space Y. Consider the associated maximum likelihood estimator with respect to a parametrized family of hidden Markov models such that the law of the observations $(Y_k)_{k\in\mathbb{Z}}$ is *not* assumed to be described by any of the hidden Markov models of this family. In this paper we investigate the consistency of this estimator in such misspecified models under mild assumptions.

1. Introduction. An assumption underlying most of the classical theory of maximum likelihood is that the "true" distribution of the observations is known to lie within a specified parametric family of distributions. In many settings, it is doubtful that this assumption is satisfied. It is therefore natural to investigate the convergence of the maximum likelihood estimator (MLE) and to identify the possible limit for misspecified models. Such questions have been mainly investigated for models in which observations are independent; see [15, 29]. Much less is known on the behavior of the MLE estimate for dependent observations; see [10] and the references therein.

For independent observations, under mild additional technical conditions, the MLE converges to the parameter which minimizes the relative entropy rate; see [15]. The purpose of this paper is to show that such a result remains true when the observations are from an ergodic process and for classes of parametric distributions associated to hidden Markov models (HMM). A HMM is a bivariate stochastic process $(X_k, Y_k)_{k\geq 0}$, where $(X_k)_{k\geq 0}$ is a Markov chain (often referred to as the state sequence) in a state space X and, conditionally on $(X_k)_{k\geq 0}$, $(Y_k)_{k\geq 0}$ is a sequence of independent random variables in a state space Y such that the conditional distribution of Y_k given the state sequence depends on X_k only. The key feature of HMMs is that the state sequence $(X_k)_{k\geq 0}$ is not observable, so that statistical inference has to be carried out by means of the observations $(Y_k)_{k\geq 0}$ only. Such problems are far from straightforward due to the fact that the observation process $(Y_k)_{k\geq 0}$ is generally a dependent, non-Markovian time series [despite that the bivariate process $(X_k, Y_k)_{k\geq 0}$ is itself a Markov chain].

Received October 2011; revised July 2012.
[1]Supported by the Agence Nationale de la Recherche through the 2009-2012 project Big MC.
MSC2010 subject classifications. Primary 62M09; secondary 62F12.
Key words and phrases. Strong consistency, hidden Markov models, maximum likelihood estimator, misspecified models, state space models.

HMMs have been intensively used in many scientific disciplines including econometrics [16, 23], biology [5], engineering [18], neurophysiology [11] and the statistical inference is therefore of significant practical importance [4]. In all these applications, misspecified models are the rule, so it is worthwhile to understand the behavior of MLE under such regime.

This work extends previous results in this direction obtained by Mevel and Finesso [24], but which are restricted to discrete state-space Markov chains. Our main result of consistency of the MLE in misspecified HMMs is derived under assumptions which are quite weak, covering general state-space HMMs under conditions which are much weaker than [9], where a strong mixing condition was imposed on the transition kernels of the hidden chain. Therefore our results can be applied to many models of practical interest, including the Gaussian linear state space model, the discrete state-space HMM and more general nonlinear state-space models.

The paper is organized as follows. In Section 2, we first introduce the setting and notations that are used throughout the paper. In Section 3, we state our main assumptions and results. In Section 4, our main result is used to establish consistency in three general classes of models: linear-Gaussian state space models, finite state models and nonlinear state space models of the vector ARCH type (this includes the stochastic volatility model and many other models of interest in time series analysis and financial econometrics). Section 5 is devoted to the proof of our main result.

Notation. Some notation pertaining to transition kernels is required. Let L be a (possibly unnormalized) transition kernel on $(\mathsf{X}, \mathcal{X})$, that is, for any $x \in \mathsf{X}$, $L(x, \cdot)$ is a finite measure on $(\mathsf{X}, \mathcal{X})$ and for any $A \in \mathcal{X}$, $x \mapsto L(x, A)$ is measurable function from $(\mathsf{X}, \mathcal{X})$ to $([0, 1], \mathcal{B}([0, 1]))$. L acts on bounded functions f on X and on σ-finite positive measures μ on $(\mathsf{X}, \mathcal{X})$ via

$$Lf(x) = \delta_x Lf \triangleq \int L(x, \mathrm{d}y) f(y), \qquad \mu L(A) = \mu L \mathbb{1}_A \triangleq \int \mu(\mathrm{d}x) L(x, A).$$

If L_1 and L_2 are two transition kernels on $(\mathsf{X}, \mathcal{X})$, then $L_1 L_2$ is the transition kernel on $(\mathsf{X}, \mathcal{X})$, given, for any $x \in \mathsf{X}$ and $A \in \mathcal{X}$ by

$$L_1 L_2(x, A) = \int L_1(x, \mathrm{d}y) L_2(y, A).$$

2. Problem statement. We consider a parameterized family of HMMs with parameter space Θ, assumed to be a compact metric space. For each parameter $\theta \in \Theta$, the distribution of the HMM is specified by the transition kernel Q^θ of the Markov chain $(X_k)_{k \geq 0}$, and by the conditional distribution g^θ of the observation Y_k given the hidden state X_k, referred to as the likelihood of the observation.

For any $m \leq n$ and any sequence $\{a_k\}_{k \in \mathbb{Z}}$, denote $a_m^n \triangleq (a_m, \dots, a_n)$, and for any probability measure χ on $(\mathsf{X}, \mathcal{X})$, define the likelihood of the observations by

$$p_\chi^\theta(y_m^n) \triangleq \int \cdots \int \chi(\mathrm{d}x_m) g^\theta(x_m, y_m) \prod_{p=m+1}^{n} Q^\theta(x_{p-1}, \mathrm{d}x_p) g^\theta(x_p, y_p),$$

$$p_\chi^\theta(y_p^n | y_m^{p-1}) \triangleq p_\chi^\theta(y_m^n) / p_\chi^\theta(y_m^{p-1}), \qquad m < p \leq n,$$

with the standard convention $\prod_{p=m}^n a_p = 1$ if $m > n$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(Y_k)_{k\in\mathbb{Z}}$ be a *stationary ergodic* stochastic process taking value in $(\mathsf{Y}, \mathcal{Y})$. We denote by $\mathbb{P}_Y$ the image probability of $\mathbb{P}$ by $(Y_k)_{k\in\mathbb{Z}}$ on the product space $(\mathsf{Y}^{\mathbb{Z}}, \mathcal{Y}^{\otimes\mathbb{Z}})$, and $\mathbb{E}_Y$ the associated expectation. We stress that the distribution $\mathbb{P}_Y$ may or may not belong to the parametric family of distributions specified by the transition kernels $\{(Q^\theta, g^\theta), \theta \in \Theta\}$. If $\mathbb{P}_Y$ does not belong to $\mathcal{G}$, the model is said to be misspecified.

If χ is a probability measure $(\mathsf{X}, \mathcal{X})$, we define the maximum likelihood estimator (MLE) associated to the initial distribution χ by

$$\hat{\theta}_{\chi,n} \triangleq \arg\max_{\theta\in\Theta} \ln p_\chi^\theta(Y_0^{n-1}). \tag{1}$$

The study of asymptotic properties of the MLE in HMMs was initiated in the seminal work of Baum and Petrie [2, 26] in the 1960s. In these papers, the model is assumed to be well specified, and the state space X and the observation space Y were both presumed to be finite sets. More than two decades later, Leroux [22] proved consistency for well-specified models in the case that X is a finite set, and Y is a general state space. The consistency of the MLE in more general HMMs has subsequently been investigated for well-specified models in a series of contributions [7, 9, 14, 20, 21] using different methods. A general consistency result for HMMs has been developed in [8].

Though the consistency results above differ in the details of their proofs, all proofs have a common thread which serves also as the starting point for this paper. Denote by $p_\chi^\theta(Y_0^n)$ the likelihood of the observations Y_0^n for the HMM with parameter $\theta \in \Theta$ and initial distribution χ. The first step of the proof aims to establish that for any $\theta \in \Theta$, there is a constant $\ell(\theta)$ such that

$$\lim_{n\to\infty} n^{-1} \log p_\chi^\theta(Y_0^{n-1}) = \lim_{n\to\infty} n^{-1} \mathbb{E}[\log p_\chi^\theta(Y_0^{n-1})] = \ell(\theta), \qquad \mathbb{P}\text{-a.s.}$$

Up to an additive constant, $\theta \mapsto \ell(\theta)$ is the negated relative entropy rate between the distribution of the observations and $p_\chi^\theta(\cdot)$, respectively. When the model is well-specified and $\theta = \theta_\star$ is the true value of the parameter, this convergence follows from the generalized Shannon–Breiman–McMillan theorem [1]; for misspecified models or for well-specified models with $\theta \neq \theta_\star$ the existence of the limit is far from obvious.

The second step of the proof aims to prove that the maximizer of the likelihood $\theta \mapsto n^{-1} \log p_\chi^\theta(Y_0^n)$ converges $\mathbb{P}$-a.s. to the maximizer of $\theta \mapsto \ell(\theta)$, that is, to the minimizer of the relative entropy rate. Together, these two steps show that the MLE is a natural estimator for the parameters which minimizes the relative entropy rate in the parametric family $\{(Q^\theta, g^\theta), \theta \in \Theta\}$.

Let us note that one could write the likelihood as

$$n^{-1} \log p_\chi^\theta(Y_0^{n-1}) = \frac{1}{n} \sum_{k=0}^{n-1} \log p_\chi^\theta(Y_k | Y_0^{k-1}),$$

where $p_\chi^\theta(Y_k | Y_0^{k-1})$ denotes the conditional density of Y_k given Y_0^{k-1} under the misspecified model with parameter θ (i.e., the one-step predictive density). If the limit of $p_\chi^\theta(Y_k | Y_0^{k-1}) \to \pi_Y^\theta(Y_{-\infty}^k)$ as $k \to \infty$ can be shown to exist $\mathbb{P}$-a.s. and to be $\mathbb{P}$-integrable, the convergence of the log-likelihood to the relative entropy rate follows from the Birkhoff ergodic theorem, since the process $\{Y_k\}_{k \in \mathbb{Z}}$ is assumed to be ergodic. This result provides an explicit representation of the relative entropy rate $\ell(\theta)$ as the expectation of the limit $\ell(\theta) = \mathbb{E}[\log \pi_Y^\theta(Y_{-\infty}^0)]$. The limit $\pi_Y^\theta(Y_{-\infty}^k)$ might be interpreted as the conditional likelihood of Y_k given the whole past $Y_{-\infty}^{k-1}$, but we must refrain ourselves of considering this quantity as a conditional density.

Such an approach was used in [2] for finite state-space, and was later extended by Douc, Moulines and Rydén [9] to general state-space, but under stringent technical conditions (uniform mixing of the Markov kernel, which more or less restricts the validity of the results to compact state-spaces, leaving aside important models, such as Linear Gaussian state-space models).

Alternatively, the predictive distribution $p_\chi^\theta(Y_k | Y_0^{k-1})$ can be expressed as a component of the state of a measure-valued Markov chain; in this approach, the existence of the limiting relative entropy rate $\ell(\theta)$, follows from the ergodic theorem for Markov chains, provided that this Markov chain can be shown to be ergodic. This approach was used in [7, 20, 21] and was later extended to misspecified models by White [24]. This technique is adequate for finite state-space Markov chains, but does not extend easily to general state-space Markov chains; see [7].

In [22], the existence of the relative entropy rate is established by means of Kingman's subadditive ergodic theorem (the same approach is used indirectly in [26], which invokes the Furstenberg–Kesten theory of random matrix products). After some additional work, an explicit representation of the relative enropy rate is again obtained. However, as is noted in [22], page 136, the latter is surprisingly difficult, as Kingman's ergodic theorem does not directly yield a representation of the limit as an expectation.

For completeness, we note that a recent attempt [12] to prove consistency of the MLE for general HMMs contains very serious problems in the proof [17] (not addressed in [13]), and therefore fails to establish the claimed results.

In this paper, we prove consistency of the MLE for general HMMs in misspecified models under quite general assumptions. Our proof follows broadly the original approach of Baum and Petrie [2] and Douc, Moulines and Rydén [9], but relaxes the very restrictive technical conditions used in these works and extends the analysis to misspecified models. The key technique to obtain this result is to establish the exponential forgetting of the filtering distribution; this result is obtained by using an original coupling technique originally introduced in [19] and refined in [6].

3. Assumptions and main results. For any integer $t \geq 1$, $\theta \in \Theta$ and any sequence $y_0^{t-1} \in \mathsf{Y}^t$, consider the unnormalized kernel $\mathbf{L}^\theta \langle y_0^{t-1} \rangle$ on $(\mathsf{X}, \mathcal{X})$ defined for all $x_0 \in \mathsf{X}$ and $\mathsf{A} \in \mathcal{X}$, by

$$\mathbf{L}^\theta\langle y_0^{t-1}\rangle(x_0, \mathsf{A}) = \int \cdots \int \left[\prod_{i=0}^{t-1} g^\theta(x_i, y_i) Q^\theta(x_i, \mathrm{d}x_{i+1})\right] \mathbb{1}_{\mathsf{A}}(x_t). \tag{2}$$

Note that, for any $t \geq 1$, $\theta \in \Theta$, $x_0 \in \mathsf{X}$, and $y_0^{t-1} \in \mathsf{Y}^t$,

$$\mathbf{L}^\theta\langle y_0^{t-1}\rangle(x_0, \mathsf{X}) = p_{x_0}^\theta\big(y_0^{t-1}\big), \tag{3}$$

where for $x \in \mathsf{X}$, $s \leq t$, $p_x^\theta(y_s^t)$, the likelihood of the observation y_s^t starting from state x, is a shorthand notation for $p_{\delta_x}^\theta(y_s^t)$.

DEFINITION 1. Let r be an integer. A set $\mathsf{C} \in \mathcal{X}$ is a *r-local Doeblin set* with respect to the family $\{Q^\theta, g^\theta\}_{\theta\in\Theta}$, if there exist positive functions $\epsilon_{\mathsf{C}}^- : \mathsf{Y}^r \to \mathbb{R}^+$, $\epsilon_{\mathsf{C}}^+ : \mathsf{Y}^r \to \mathbb{R}^+$ and a family of probability measures $\{\lambda_{\mathsf{C}}^\theta\langle z\rangle\}_{\theta\in\Theta, z\in\mathsf{Y}^r}$ and of positive functions $\{\varphi_{\mathsf{C}}^\theta\langle z\rangle\}_{\theta\in\Theta, z\in\mathsf{Y}^r}$ such that for any $\theta \in \Theta$, $z \in \mathsf{Y}^r$, $\lambda_{\mathsf{C}}^\theta\langle z\rangle(\mathsf{C}) = 1$ and, for any $\mathsf{A} \in \mathcal{X}$, and $x \in \mathsf{C}$,

$$\epsilon_{\mathsf{C}}^-(z)\varphi_{\mathsf{C}}^\theta\langle z\rangle(x)\lambda_{\mathsf{C}}^\theta\langle z\rangle(\mathsf{A}) \leq \mathbf{L}^\theta\langle z\rangle(x, \mathsf{A}\cap\mathsf{C}) \leq \epsilon_{\mathsf{C}}^+(z)\varphi_{\mathsf{C}}^\theta\langle z\rangle(x)\lambda_{\mathsf{C}}^\theta\langle z\rangle(\mathsf{A}). \tag{4}$$

This implies that for any measurable nonnegative function f on $(\mathsf{X}, \mathcal{X})$, $x \in \mathsf{C}$ and any $z \in \mathsf{Y}^r$,

$$\epsilon_{\mathsf{C}}^-(z)\varphi_{\mathsf{C}}^\theta\langle z\rangle(x)\lambda_{\mathsf{C}}^\theta\langle z\rangle(\mathbb{1}_{\mathsf{C}} f) \leq \delta_x \mathbf{L}^\theta\langle z\rangle(\mathbb{1}_{\mathsf{C}} f) \leq \epsilon_{\mathsf{C}}^+(z)\varphi_{\mathsf{C}}^\theta\langle z\rangle(x)\lambda_{\mathsf{C}}^\theta\langle z\rangle(\mathbb{1}_{\mathsf{C}} f).$$

We require that the condition is satisfied for any $\theta \in \Theta$, but this is not a serious restriction since Θ is assumed to be compact.

REMARK 1. To illustrate this condition, consider the case $r = 1$. Assume that for some set C, there exist positive constants $\epsilon_{\mathsf{C}}^-, \epsilon_{\mathsf{C}}^+$ and a family of probability measures $\{\lambda_{\mathsf{C}}^\theta\}_{\theta\in\Theta}$ such that for any $\theta \in \Theta$, $\lambda_{\mathsf{C}}^\theta(\mathsf{C}) = 1$ and, for any $\mathsf{A} \in \mathcal{X}$, and $x \in \mathsf{C}$,

$$\epsilon_{\mathsf{C}}^-\lambda_{\mathsf{C}}^\theta(\mathsf{A}) \leq Q^\theta(x, \mathsf{A}\cap\mathsf{C}) \leq \epsilon_{\mathsf{C}}^+\lambda_{\mathsf{C}}^\theta(\mathsf{A}).$$

Then, clearly $\mathbf{L}^\theta\langle y\rangle(x, \mathsf{A}) = g^\theta(x, y)Q^\theta(x, \mathsf{A})$ satisfies (4) where ϵ_{C}^- and ϵ_{C}^+ are positive *constants*. In this case C is a 1-local Doeblin set with respect to Q^θ; see [6] and [19].

REMARK 2. Local Doeblin sets share some similarities with 1-small set in the theory of Markov chains over general state spaces; see [25], Chapter 5. Recall that a set C is 1-small for the kernel Q^θ, $\theta \in \Theta$ if there exists a probability measure $\tilde{\lambda}^\theta_{\mathsf{C}}$ and a constant $\tilde{\epsilon}_{\mathsf{C}} > 0$, such that $\tilde{\lambda}^\theta_{\mathsf{C}}(\mathsf{C}) = 1$, and for all $x \in \mathsf{C}$ and $A \in \mathcal{X}$, $Q^\theta(x, A \cap \mathsf{C}) \geq \tilde{\epsilon}_{\mathsf{C}} \tilde{\lambda}^\theta_{\mathsf{C}}(A \cap \mathsf{C})$. In particular, a local Doeblin set is 1-small with $\tilde{\epsilon}_{\mathsf{C}} = \epsilon^-_{\mathsf{C}}$ and $\tilde{\lambda}^\theta_{\mathsf{C}} = \lambda^\theta_{\mathsf{C}}$. The main difference stems from the fact that we impose both a lower and an upper bound, and we impose that the minorizing and the majorizing measures are the same.

(A1) There exist an integer $r \geq 1$ and a set $\mathsf{K} \in \mathcal{Y}^{\otimes r}$ such that:

(i) $\mathbb{P}[Y_0^{r-1} \in \mathsf{K}] > 2/3$.

(ii) For all $\eta > 0$, there exists a r-local Doeblin set $\mathsf{C} \in \mathcal{X}$ such that for all $\theta \in \Theta$ and for all $y_0^{r-1} \in \mathsf{K}$,

$$\sup_{x_0 \in \mathsf{C}^c} p^\theta_{x_0}(y_0^{r-1}) \leq \eta \sup_{x_0 \in \mathsf{X}} p^\theta_{x_0}(y_0^{r-1}) < \infty \tag{5}$$

and

$$\inf_{y_0^{r-1} \in \mathsf{K}} \frac{\epsilon^-_{\mathsf{C}}(y_0^{r-1})}{\epsilon^+_{\mathsf{C}}(y_0^{r-1})} > 0, \tag{6}$$

where the functions ϵ^+_{C} and ϵ^-_{C} are defined in Definition 1.

(iii) There exists a set D such that

$$\mathbb{E}\Big[\ln^- \inf_{\theta \in \Theta} \inf_{x \in \mathsf{D}} \mathbf{L}^\theta \langle Y_0^{r-1} \rangle (x, \mathsf{D})\Big] < \infty. \tag{7}$$

(A2) (i) For any $\theta \in \Theta$, the function $g^\theta : (x, y) \in \mathsf{X} \times \mathsf{Y} \mapsto g^\theta(x, y)$ is positive,

(ii) $\mathbb{E}[\ln^+ \sup_{\theta \in \Theta} \sup_{x \in \mathsf{X}} g^\theta(x, Y_0)] < \infty$.

(A3) There exists $p \in \mathbb{N}$ such that for any $x \in \mathsf{X}$ and $n \geq p$, $\mathbb{P}$-a.s. the function $\theta \mapsto p^\theta_x(Y_0^n)$ is continuous on Θ.

REMARK 3. Assumption (A2) assumes that the conditional likelihood g^θ is positive. The case where g^θ can vanish typically requires different conditions; see [3, 27]. The second condition can be read as a generalized moment condition on Y_0. It is satisfied in many examples of interest.

REMARK 4. To check (A1)(iii), one may, for example, check that:

(i) $\inf_{x \in \mathsf{D}} \inf_{\theta \in \Theta} Q^\theta(x, \mathsf{D}) > 0$;

(ii) $\mathbb{E}[\ln^- \inf_{\theta \in \Theta} \inf_{x \in \mathsf{D}} g^\theta(x, Y_0)] < \infty$.

This condition is satisfied if $(x, \theta) \mapsto g^\theta(x, y)$ is continuous and D is a compact small set for all $\theta \in \Theta$, there exists a probability measure ν^θ such that $\nu^\theta(\mathsf{D}) = 1$

and a constant $\delta > 0$, such that, for all $x \in \mathsf{D}$ and $A \in \mathcal{X}$, $Q^\theta(x, A) \geq \delta \nu^\theta(A)$. Note, however, that (A1)(iii) is far weaker than imposing that the set D is 1-small. This is important to deal with examples for which the transition kernel $Q^\theta(x, \cdot)$ does not admit a density with respect to to some fixed dominating measure; see, for example, Section 4.1.

REMARK 5. Assumption (A3) is in general the consequence of the continuity of the kernel $\theta \mapsto Q^\theta(x, \cdot)$ and of the function $\theta \mapsto g^\theta(x, \cdot)$, using classical techniques to deal with integrals depending on a parameter.

REMARK 6. According to (3), bound (5) may also be rewritten in terms of the kernel $\mathbf{L}^\theta\langle y_0^{r-1}\rangle$ as

$$\sup_{x_0 \in \mathsf{C}^c} \mathbf{L}^\theta\langle y_0^{r-1}\rangle(x_0, \mathsf{X}) \leq \eta \sup_{x_0 \in \mathsf{X}} \mathbf{L}^\theta\langle y_0^{r-1}\rangle(x_0, \mathsf{X}) < \infty.$$

The convergence of the relative entropy is achieved for initial distributions belonging to a particular class of initial probability distributions. For the integer r and the set $\mathsf{D} \in \mathcal{X}$ defined in (A1), let $\mathcal{M}(\mathsf{D}, r)$ be the subset $\mathcal{P}(\mathsf{X}, \mathcal{X})$ of probability measures on $(\mathsf{X}, \mathcal{X})$ satisfying

$$\begin{aligned} \mathcal{M}(\mathsf{D}, r) = \Big\{ &\chi \in \mathcal{P}(\mathsf{X}, \mathcal{X}), \\ &\mathbb{E}\Big[\ln^- \inf_{\theta \in \Theta} \chi \mathbf{L}^\theta\langle Y_0^{u-1}\rangle \mathbb{1}_{\mathsf{D}}\Big] < \infty \text{ for all } u \in \{1, \dots, r\}\Big\}. \end{aligned} \tag{8}$$

PROPOSITION 1. *Assume* (A1) *and* (A2). *Then*:

(i) *for any* $\theta \in \Theta$, *there exists a measurable function* $\pi_Y^\theta : \mathsf{Y}^{\mathbb{Z}^-} \to \mathbb{R}$ *such that for any probability measure* $\chi \in \mathcal{M}(\mathsf{D}, r)$,

$$\mathbb{P}\Big[\lim_{m \to \infty} p_\chi^\theta(Y_0 | Y_{-m}^{-1}) = \pi_Y^\theta(Y_{-\infty}^0)\Big] = 1;$$

moreover,

$$\mathbb{E}\big[|\ln \pi_Y^\theta(Y_{-\infty}^0)|\big] < \infty; \tag{9}$$

(ii) *for any* $\theta \in \Theta$ *and any probability measure* $\chi \in \mathcal{M}(\mathsf{D}, r)$,

$$\lim_{n \to \infty} n^{-1} \ln p_\chi^\theta(Y_0^{n-1}) = \ell(\theta), \qquad \mathbb{P}\text{-}a.s.,$$

where $\ell(\theta) \triangleq \mathbb{E}[\ln \pi_Y^\theta(Y_{-\infty}^0)]$.

THEOREM 2. *Assume* (A1)–(A3). *Then*, $\theta \mapsto \ell(\theta)$ *is upper semi-continuous and defining* $\Theta^\star \subset \Theta$ *by* $\Theta^\star \triangleq \arg\max_{\theta \in \Theta} \ell(\theta)$, *we have for any probability measure* $\chi \in \mathcal{M}(\mathsf{D}, r)$,

$$\lim_{n \to \infty} d(\hat{\theta}_{\chi,n}, \Theta^\star) = 0, \qquad \mathbb{P}\text{-}a.s.$$

REMARK 7. When the model is well specified, the law of the observations belongs to the parametric family of distributions on which the maximization occurs and is therefore associated to a specific parameter θ^*. In this particular case, under some appropriate assumptions, the set Θ^* is reduced to the singleton $\{\theta^*\}$, and the consistency result of the MLE in well specified models can then be written as (see [8])

$$\lim_{n\to\infty} d(\hat{\theta}_{\chi,n}, \theta^\star) = 0, \qquad \mathbb{P}\text{-a.s.}$$

A simple sufficient condition can be proposed to ensure that $\chi \in \mathcal{M}(\mathsf{D}, r)$.

PROPOSITION 3. *Assume there exist a sequence of sets* $\mathsf{D}_u \in \mathcal{X}$, $u \in \{0, \dots, r-1\}$, *such that* (*setting* $\mathsf{D}_r = \mathsf{D}$ *for notational convenience*), *for some* $\delta > 0$,

$$\inf_{x_{u-1}\in\mathsf{D}_{u-1}} \inf_{\theta\in\Theta} Q^\theta(x_{u-1}, \mathsf{D}_u) \ge \delta, \qquad u \in \{1, \dots, r\}, \tag{10}$$

and

$$\mathbb{E}\Big[\ln^- \inf_{\theta\in\Theta} \inf_{x\in\mathsf{D}_u} g^\theta(x, Y_0)\Big] < \infty \qquad \text{for } u \in \{0, \dots, r\}. \tag{11}$$

Then, *any initial distribution* χ *on* $(\mathsf{X}, \mathcal{X})$ *satisfying* $\chi(\mathsf{D}_0) > 0$ *belongs to* $\mathcal{M}(\mathsf{D}, r)$.

REMARK 8. To check (11), we typically assume that, for any given $y \in \mathsf{Y}$, the function $(x, \theta) \mapsto g^\theta(x, y)$ is continuous and that $\mathsf{D}_i \times \Theta$ is a compact set, $i \in \{0, \dots, r-1\}$. This condition then translates into an assumption on some generalized moments of the process Y.

To check (10), the following lemma is useful.

LEMMA 4. *Assume that* $\mathsf{X} = \mathbb{R}^d$ *for some integer* $d > 0$ *and that* $\mathcal{X}$ *is the associated Borel* σ*-field. Assume in addition that, for any open subset* $\mathsf{O} \in \mathcal{X}$, *the function* $(x, \theta) \to Q^\theta(x, \mathsf{O})$ *is lower semi-continuous on the product space* $\mathsf{X} \times \Theta$. *Then*, *for any* $\delta > 0$ *and any compact subset* $\mathsf{D}_0 \in \mathcal{X}$, *there exists a sequence of compact subsets* D_u, $u \in \{0, \dots, r-1\}$ *satisfying* (10).

4. Applications. In this section, we develop three classes of examples. In Section 4.1 we consider linear Gaussian state space models. This is obviously a very important model, which is used routinely to analyze time-series models. We analyze this model under assumptions which are very general and might serve to illustrate the stated assumptions. In Section 4.2, we consider the classic case where state space of the underlying Markov chain is a finite set. Finally, in Section 4.3, we develop a general class of nonlinear state space models. In all these examples, we will find that the assumptions of Theorem 2 are satisfied under general assumptions.

4.1. *Gaussian linear state space models.* Gaussian linear state space models form an important class of HMMs. In this setting, let $\mathsf{X} = \mathbb{R}^{d_x}$, and $\mathsf{Y} = \mathbb{R}^{d_y}$ for some integers and let Θ be a compact parameter space. The model is specified by

$$X_{k+1} = A_\theta X_k + R_\theta U_k, \tag{12}$$

$$Y_k = B_\theta X_k + S_\theta V_k, \tag{13}$$

where $\{(U_k, V_k)\}_{k \geq 0}$ is an i.i.d. sequence of Gaussian vectors with zero mean and identity covariance matrix, independent of X_0. Here U_k is d_u-dimensional, V_k is d_y-dimensional and the matrices $A_\theta, R_\theta, B_\theta, S_\theta$ have the appropriate dimensions.

For any integer n, define $\mathcal{O}_{\theta,n}$ and $\mathcal{C}_{\theta,n}$ the observability matrix and the controllability matrices

$$\mathcal{O}_{\theta,n} \triangleq \begin{bmatrix} B_\theta \\ B_\theta A_\theta \\ B_\theta A_\theta^2 \\ \vdots \\ B_\theta A_\theta^{n-1} \end{bmatrix} \quad \text{and} \quad \mathcal{C}_{\theta,n} \triangleq \left[A_\theta^{n-1} R_\theta A_\theta^{n-2} R_\theta \cdots R_\theta \right]. \tag{14}$$

It is assumed in the sequel that for any $\theta \in \Theta$, the following hold:

(L1) The pair $[A_\theta, B_\theta]$ is observable, and the pair $[A_\theta, R_\theta]$ is controllable; that is, there exists an integer r such that, the observability matrix $\mathcal{O}_{\theta,r}$ and the controllability matrix $\mathcal{C}_{\theta,r}$ are full rank.

(L2) The measurement noise covariance matrix S_θ is full rank.

(L3) The functions $\theta \mapsto A_\theta$, $\theta \mapsto R_\theta$, $\theta \mapsto B_\theta$ and $\theta \mapsto S_\theta$ are continuous on Θ.

(L4) $\mathbb{E}[\|Y_0\|^2] < \infty$.

We now check the assumptions of Theorem 2.

The dimension d_u of the state noise vector U_k is in many situations smaller than the dimension d_x of the state vector X_k and hence $R_\theta {}^t R_\theta$ (where ${}^t A$ is the transpose of the matrix A) may be rank deficient.

Some additional notation is needed. For any positive matrix A and any vector z of appropriate dimension, denote $\|z\|_A^2 = {}^t z A^{-1} z$. Define for any integer n

$$\mathcal{F}_{\theta,n} = \mathcal{D}_{\theta,n} {}^t\mathcal{D}_{\theta,n} + \mathcal{S}_{\theta,n} {}^t\mathcal{S}_{\theta,n}, \tag{15}$$

where ^{t} denotes the transpose and

$$\mathcal{D}_{\theta,n} \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 \\ B_\theta R_\theta & \ddots & & 0 \\ B_\theta A_\theta R_\theta & B_\theta R_\theta & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ B_\theta A_\theta^{n-2} R_\theta & B_\theta A_\theta^{n-3} R_\theta & \cdots & B_\theta R_\theta \end{bmatrix},$$

$$\mathcal{S}_{\theta,n} \triangleq \begin{bmatrix} S_\theta & 0 & \cdots & 0 \\ 0 & S_\theta & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & S_\theta \end{bmatrix}.$$

Under (L2), for any $n \geq r$, the matrix $\mathcal{F}_{\theta,n}$ is positive definite. The likelihood of the observations $y_0^{n-1} \in \mathsf{Y}^n$ starting from x_0 is given by

$$p_{x_0}^\theta(y_0^{n-1}) = (2\pi)^{-nd_y} \det{}^{-1/2}(\mathcal{F}_{\theta,n}) \exp\bigl(-\tfrac{1}{2}\|\mathbf{y}_{n-1} - \mathcal{O}_{\theta,n}x_0\|^2_{\mathcal{F}_{\theta,n}}\bigr), \tag{16}$$

where $\mathbf{y}_{n-1} = {}^t[{}^t y_0, {}^t y_1, \ldots, {}^t y_{n-1}]$, and $\mathcal{O}_{\theta,n}$ is defined in (14).

Consider first (A1). Under (L1), the observability matrix $\mathcal{O}_{\theta,r}$ is full rank, we have, for any compact subset $\mathsf{K} \subset \mathsf{Y}^r$,

$$\lim_{\|x_0\|\to\infty} \inf_{y_0^{r-1}\in\mathsf{K}} \|\mathbf{y}_{r-1} - \mathcal{O}_{\theta,r}x_0\|_{\mathcal{F}_{\theta,r}} = \infty,$$

showing that, for all $\eta > 0$, we may choose a compact set C in such a way that (5) is satisfied. It remains to prove that any compact set C is a r-local Doeblin satisfying the condition (6). For any $y_0^{r-1} \in \mathsf{Y}^{r-1}$ and $x_0 \in \mathsf{X}$ the measure $\mathbf{L}^\theta\langle y_0^{r-1}\rangle(x_0, \cdot)$ is absolutely continuous with respect to the Lebesgue measure on X with Radon–Nikodym denoted $\boldsymbol{\ell}^\theta\langle y_0^{r-1}\rangle(x_0, x_r)$ given (up to an irrelevant multiplicative factor) by

$$\boldsymbol{\ell}^\theta\langle y_0^{r-1}\rangle(x_0, x_r) \propto \det{}^{-1/2}(\mathcal{G}_{\theta,r}) \exp\left(-\frac{1}{2}\left\| \begin{bmatrix} \mathbf{y}_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} \mathcal{O}_{\theta,r} \\ A_\theta^r \end{bmatrix} x_0 \right\|^2_{\mathcal{G}_{\theta,r}}\right), \tag{17}$$

where the covariance matrix $\mathcal{G}_{\theta,r}$ is given by

$$\mathcal{G}_{\theta,r} = \begin{bmatrix} \mathcal{D}_{\theta,r} \\ \mathcal{C}_{\theta,r} \end{bmatrix} [{}^t\mathcal{D}_{\theta,r} {}^t\mathcal{C}_{\theta,r}] + \begin{bmatrix} \mathcal{S}_{\theta,r} \\ \mathbf{0} \end{bmatrix} [{}^t\mathcal{S}_{\theta,r} {}^t\mathbf{0}].$$

The proof of (17) relies on the positivity of $\mathcal{G}_{\theta,r}$, which requires further discussion. By construction, the matrix $\mathcal{G}_{\theta,r}$ is nonnegative. For any $\mathbf{y}_{r-1} \in \mathsf{Y}^r$ and $x \in \mathsf{X}$, the equation

$$[{}^t\mathbf{y}_{r-1} {}^t x]\mathcal{G}_{\theta,r} \begin{bmatrix} \mathbf{y}_{r-1} \\ x \end{bmatrix} = \|{}^t\mathcal{D}_{\theta,r}\mathbf{y}_{r-1} + {}^t\mathcal{C}_{\theta,r}x\|^2 + \|{}^t\mathcal{S}_{\theta,r}\mathbf{y}_{r-1}\|^2 = 0$$

implies that $\|{}^t\mathcal{D}_{\theta,r}\mathbf{y}_{r-1} + {}^t\mathcal{C}_{\theta,r}x\|^2 = 0$ and $\|{}^t\mathcal{S}_{\theta,r}\mathbf{y}_{r-1}\|^2 = 0$. Since the matrix $\mathcal{S}_{\theta,r}$ is full rank, this implies that $\mathbf{y}_{r-1} = 0$. Since $\mathcal{C}_{\theta,r}$ is full-rank (the pair $[A_\theta, R_\theta]$ is controllable), this implies that $x = 0$. Therefore, the matrix $\mathcal{G}_{\theta,r}$ is positive definite and, for any $\mathbf{y}_{r-1}$, the function

$$(x_0, x_r) \mapsto \left\| \begin{bmatrix} \mathbf{y}_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} \mathcal{O}_{\theta,r} \\ A_\theta^r \end{bmatrix} x_0 \right\|^2_{\mathcal{G}_{\theta,r}}$$

is continuous, and is therefore bounded on any compact subset of $\mathsf{X} \times \mathsf{X}$. This implies that every nonempty compact set $\mathsf{C} \subset \mathbb{R}^{d_x}$ is a r-local Doeblin set, with $\lambda_{\mathsf{C}}^{\theta}(\cdot) = \lambda^{\mathrm{Leb}}(\cdot)/\lambda^{\mathrm{Leb}}(\mathsf{C})$ and

$$\epsilon_{\mathsf{C}}^{-}(y_0^{r-1}) = (\lambda^{\mathrm{Leb}}(\mathsf{C}))^{-1} \inf_{\theta \in \Theta} \inf_{(x_0, x_r) \in \mathsf{C} \times \mathsf{C}} \boldsymbol{\ell}^{\theta}\langle y_0^{r-1} \rangle (x_0, x_r),$$

$$\epsilon_{\mathsf{C}}^{+}(y_0^{r-1}) = (\lambda^{\mathrm{Leb}}(\mathsf{C}))^{-1} \sup_{\theta \in \Theta} \sup_{(x_0, x_r) \in \mathsf{C} \times \mathsf{C}} \boldsymbol{\ell}^{\theta}\langle y_0^{r-1} \rangle (x_0, x_r).$$

Therefore, condition (6) is satisfied with any compact set $\mathsf{K} \subseteq \mathsf{Y}^{r-1}$. It remains to show (A1)(iii). Under (L1), $\mathbf{L}^{\theta}\langle y_0^{r-1} \rangle (x_0, \cdot)$ is absolutely continuous with respect to the Lebesgue measure λ^{Leb}. Therefore, for any set D,

$$\inf_{\theta \in \Theta} \inf_{x_0 \in \mathsf{D}} \mathbf{L}^{\theta}\langle y_0^{r-1} \rangle (x_0, \mathsf{D}) \geq \inf_{\theta \in \Theta} \inf_{(x_0, x_r) \in \mathsf{D} \times \mathsf{D}} \boldsymbol{\ell}^{\theta}\langle y_0^{r-1} \rangle (x_0, x_r) \lambda^{\mathrm{Leb}}(\mathsf{D}).$$

Take D to be any compact set with positive Lebesgue measure.

$$\begin{aligned} &\sup_{\theta \in \Theta} \sup_{(x_0, x_r) \in \mathsf{D} \times \mathsf{D}} \left\| \begin{bmatrix} \mathbf{y}_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} \mathcal{O}_{\theta,r} \\ A_{\theta}^{r} \end{bmatrix} x_0 \right\|_{\mathcal{G}_{\theta,r}}^2 \\ &\quad \leq 2\lambda_{\max}(\mathcal{G}_{\theta,r}) \Big\{ \|\mathbf{y}_{r-1}\|^2 + \max_{x \in \mathsf{D}} \|x\|^2 \big[1 + \lambda_{\max}\big({}^t\mathcal{O}_{\theta,r}\mathcal{O}_{\theta,r} + {}^t A_{\theta}^{r} A_{\theta}^{r}\big)\big] \Big\}, \end{aligned}$$

where $\lambda_{\max}(A)$ is the largest eigenvalue of A. Under (L3), $\theta \mapsto \lambda_{\max}(\mathcal{G}_{\theta,r})$ and $\theta \mapsto \lambda_{\max}({}^t\mathcal{O}_{\theta,r}\mathcal{O}_{\theta,r} + {}^t A_{\theta}^{r} A_{\theta}^{r})$ are bounded. Under (L4), $\mathbb{E}[\|Y_0\|^2] < \infty$, then (A1)(iii) is satisfied for any compact set.

Consider now (A2). Under (L2), S_θ is full rank, and choosing the reference measure μ to be the Lebesgue measure on Y, we find that $g^{\theta}(x, y)$ is a Gaussian density for each $x \in \mathsf{X}$ with covariance matrix $S_{\theta}{}^t S_{\theta}$. We therefore have

$$\sup_{\theta \in \Theta} \sup_{x \in \mathsf{X}} g^{\theta}(x, y) = (2\pi)^{-d_y/2} \sup_{\theta \in \Theta} \det{}^{-1/2}\big(S_{\theta}{}^t S_{\theta}\big) < \infty,$$

so that (A2)(i) and (ii) are satisfied.

We finally check (A3). For any $n \geq r$, and $x \in \mathsf{X}$ the function $\theta \mapsto p_{x_0}^{\theta}(y_0^{n-1})$ is given by (16). Under (L3), the functions $\theta \mapsto \mathcal{O}_{\theta,n}$ [where $\mathcal{O}_{\theta,n}$ is the observability matrix defined in (14)] and $\theta \mapsto \det^{-1/2}(\mathcal{F}_{\theta,n})$ [where $\mathcal{F}_{\theta,n}$ is the covariance matrix defined in (15)], are continuous on Θ for any $n \geq r$. Thus, for any $x \in \mathsf{X}$, $\theta \mapsto p_x^{\theta}(y_0^{n-1})$ is continuous for every $n \geq r$, showing (A3).

To conclude this discussion, we need to specify more explicitly the set $\mathcal{M}(\mathsf{D}, r)$ [see (8)] of possible initial distributions. Using Proposition 3, we have to check the sufficient conditions (10) and (11). To check (10), we use Lemma 4. Note that, for any open subset O,

$$Q^{\theta}(x, \mathsf{O}) = \mathbb{E}\big[\mathbb{1}_{\mathsf{O}}(A_{\theta}x + R_{\theta}U)\big],$$

where the expectation is taken with respect to the standard normal random variable U. Let $\{(x_n, \theta_n)\}_{n=1}^{\infty}$ be any sequence converging to (x, θ). By the Fatou lemma, using that function $\mathbb{1}_{\mathsf{O}}$ is lower semi-continuous and that $\theta \mapsto A_\theta$ is continuous under (L3), we have

$$\liminf_{n\to\infty} Q^{\theta_n}(x_n, \mathsf{O}) \geq \mathbb{E}\Big[\liminf_{n\to\infty} \mathbb{1}_{\mathsf{O}}(A_{\theta_n} x + R_{\theta_n} U)\Big]$$
$$\geq \mathbb{E}\Big[\liminf_{n\to\infty} \mathbb{1}_{\mathsf{O}}(A_{\theta_n} x + R_{\theta_n} U)\Big] = Q^\theta(x, \mathsf{O}),$$

showing that, for any open subset O, the function $(x,\theta) \mapsto Q^\theta(x,\mathsf{O})$ is lower semi-continuous.

Assumption (L2) implies that, for all $(x, y) \in \mathsf{X} \times \mathsf{Y}$,

$$\ln g^\theta(x, y) \geq -\frac{d_y}{2} \ln(2\pi) - \frac{1}{2} \inf_{\theta\in\Theta} \ln \det{}^{-1/2}(S_\theta{}^t S_\theta)$$
$$- \Big[\inf_{\theta\in\Theta} \lambda_{\min}(S_\theta{}^t S_\theta)\Big]^{-1} \Big[\|y\|^2 + \sup_{\theta\in\Theta} \|B_\theta x\|^2\Big],$$

where $\lambda_{\min}(S_\theta{}^t S_\theta)$ is the minimal eigenvalue of $S_\theta{}^t S_\theta$. Therefore, under (L4), (11) is satisfied because D_u is a compact set, $u \in \{0, \ldots, r\}$.

We can therefore apply Theorem 2 to show that the MLE is consistent for any initial measure χ as soon as the process $\{Y_k\}_{k\in\mathbb{Z}}$ is stationary ergodic and $\mathbb{E}[|Y_0|^2] < \infty$.

4.2. *Finite state models*. One of the most widely used classes of HMMs is obtained when the state-space is finite, that is, $\mathsf{X} = \{1, \ldots, d\}$ for some integer d, Y is any Polish space and Θ is a compact metric space. For each parameter $\theta \in \Theta$, the transition kernel Q^θ is determined by the corresponding transition probability matrix $\mathcal{Q}_\theta$, while the observation density g^θ is given as in the general setting of this paper.

It is assumed in the sequel that:

(F1) There exists an integer $r > 0$, such that, $\inf_{\theta\in\Theta} \inf_{(x,x')\in\mathsf{X}\times\mathsf{X}} \mathcal{Q}_\theta^r(x, x') > 0$.

(F2) There exists a set $\mathsf{M} \subset \mathsf{Y}$ such that $\inf_{\theta\in\Theta} \inf_{y\in\mathsf{M}} \inf_{x\in\mathsf{X}} g^\theta(x, y) > 0$ and $\sup_{\theta\in\Theta} \sup_{y\in\mathsf{M}} \sup_{x\in\mathsf{X}} g^\theta(x, y) < \infty$.

(F3) For any $\theta \in \Theta$, the function $g^\theta : (x, y) \in \mathsf{X} \times \mathsf{Y} \mapsto g^\theta(x, y)$ is positive and

$$\mathbb{E}\Big[\ln^+ \sup_{\theta\in\Theta} \sup_{x\in\mathsf{X}} g^\theta(x, Y_0)\Big] < \infty.$$

(F4) $\mathbb{E}[\ln^- \inf_{\theta\in\Theta} \inf_{x\in\mathsf{X}} g^\theta(x, Y_0)] < \infty$.

(F5) $\theta \mapsto \mathcal{Q}_\theta$ and $\theta \mapsto g_\theta(x, y)$ are continuous for any $x \in \mathsf{X}$, $y \in \mathsf{Y}$.

Consider first (A1). We set $\mathsf{C} = \mathsf{X}$. Since $\mathsf{C}^c = \varnothing$, (5) is trivially satisfied. Under (F1), equation (4) is satisfied with $\varphi_{\mathsf{X}}\langle y_0^{r-1}\rangle(x) \equiv 1$, $\lambda_{\mathsf{X}}^\theta = d^{-1} \sum_{i=1}^d \delta_i$, and

$$\epsilon_{\mathsf{X}}^-[y_0^{r-1}] = d \prod_{i=0}^{d-1} \inf_{\theta\in\Theta} \inf_{x\in\mathsf{X}} g^\theta(x, y_i) \times \inf_{\theta\in\Theta} \inf_{(x,x')\in\mathsf{X}\times\mathsf{X}} \mathcal{Q}_\theta^r(x, x'),$$

$$\epsilon^+_{\mathsf{X}}[y_0^{r-1}] = d \prod_{i=0}^{d-1} \sup_{\theta\in\Theta}\sup_{x\in\mathsf{X}} g^\theta(x, y_i) \times \sup_{\theta\in\Theta} \sup_{(x,x')\in\mathsf{X}\times\mathsf{X}} \mathcal{Q}^r_\theta(x, x').$$

Hence, the state space X is a r-local Doeblin set. Assumption (F2) implies that (6) is satisfied with $\mathsf{K} = \mathsf{M}^r$. Now, note that for all $u \in \{1, \dots, r\}$ and $y_0^{u-1} \in \mathsf{Y}^r$,

$$\inf_{\theta\in\Theta}\inf_{x\in\mathsf{X}} \mathbf{L}^\theta\langle y_0^{u-1}\rangle \geq \prod_{i=0}^{u-1} \inf_{\theta\in\Theta}\inf_{x\in\mathsf{X}} g^\theta(x, y_i). \tag{18}$$

Using the previous inequality with $u = r$ and noting that (F4) implies that $\mathbb{E}[\ln^- \inf_{\theta\in\Theta}\inf_{x\in\mathsf{X}} g^\theta(x, Y_0)] < \infty$, we see that equation (7) is satisfied with $\mathsf{D} = \mathsf{X}$. The same argument for any $u \in 1, \dots, r$ shows that all the probability measures on $(\mathsf{X}, \mathcal{X})$ belong to the set $\mathcal{M}(\mathsf{X}, r)$, defined in (8).

Assumption (A2) is a direct consequence of (F3). Finally, we note that the continuity of $\theta \mapsto \mathcal{Q}_\theta$ and $\theta \mapsto g_\theta(x, y)$ yield immediately that $\theta \mapsto p^\theta_x(y_0^n)$ is a continuous function for every $n \geq 0$ and $y_0^n \in \mathsf{Y}^{n+1}$, establishing (A3).

We can therefore apply Theorem 2 under (F1)–(F5) to show that the MLE is consistent for any initial measure χ as soon as the process $\{Y_k\}_{k\in\mathbb{Z}}$ is stationary ergodic.

4.3. *Nonlinear state space models*. In this section, we consider a class of nonlinear state space models. Let $\mathsf{X} = \mathbb{R}^d$, $\mathsf{Y} = \mathbb{R}^\ell$ and $\mathcal{X}$ and $\mathcal{Y}$ be the associated Borel σ-fields. Let Θ be a compact metric space. For each $\theta \in \Theta$ and each $x \in \mathsf{X}$, the Markov kernel $Q_\theta(x, \cdot)$ has a density $q_\theta(x, \cdot)$ with respect to the Lebesgue measure on X.

For example, $(X_k)_{k\geq 0}$ may be defined through the nonlinear recursion

$$X_k = T_\theta(X_{k-1}) + \Sigma_\theta(X_{k-1})\zeta_k,$$

where $(\zeta_k)_{k\geq 1}$ is an i.i.d. sequence of d-dimensional random vectors which are assumed to possess a density ρ_ζ with respect to the Lebesgue measure λ^{Leb} on $\mathbb{R}^d$, and $T_\theta : \mathbb{R}^d \to \mathbb{R}^d$, $\Sigma_\theta : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are given (measurable) matrix-valued functions such that for each $\theta \in \Theta$ and $x \in \mathsf{X}$, $\Sigma_\theta(x)$ is full-rank. Such a model for $(X_k)_{k\geq 0}$ is sometimes known as a vector ARCH model, and covers many models of interest in time series analysis and financial econometrics. We let the reference measure μ be the Lebesgue measure on $\mathbb{R}^\ell$, and define the observed process $(Y_k)_{k\geq 0}$ by means of a given observation density $g^\theta(x, y)$.

We now introduce the basic assumptions of this section.

(NL1) The function $(x, x', \theta) \mapsto q^\theta(x, x')$ is a positive continuous function on $\mathsf{X} \times \mathsf{X} \times \Theta$. In addition, $\sup_{\theta\in\Theta}\sup_{(x,x')\in\mathsf{X}\times\mathsf{X}} q^\theta(x, x') < \infty$.

(NL2) For any compact subset $\mathsf{K} \subset \mathsf{Y}$, and $\theta \in \Theta$,

$$\lim_{|x|\to\infty} \sup_{y\in\mathsf{K}} \frac{g^\theta(x, y)}{\sup_{x'\in\mathsf{X}} g^\theta(x', y)} = 0.$$

(NL3) For each $(x, y) \in \mathsf{X} \to \mathsf{Y}$, the function $\theta \mapsto g^\theta(x, y)$ is positive and continuous on Θ. Moreover,

$$\mathbb{E}\Big[\ln^+ \sup_{\theta\in\Theta} \sup_{x\in\mathsf{X}} g^\theta(x, Y_0)\Big] < \infty.$$

(NL4) There exists a compact subset $\mathsf{D} \subset \mathsf{Y}$ such that

$$\mathbb{E}\Big[\ln^- \inf_{\theta\in\Theta} \inf_{x\in\mathsf{D}} g^\theta(x, Y_0)\Big] < \infty.$$

We have made no attempt at generality here: for sake of example, we have chosen a set of conditions under which the assumptions of Theorem 2 are easily verified. Of course, the applicability of Theorem 2 extends far beyond the simple assumptions imposed in this section.

REMARK 9. Nonetheless, the present assumptions already cover a broad class of nonlinear models. Consider, for example, the stochastic volatility model [16]

$$\begin{cases} X_{k+1} = \phi_\theta X_k + \sigma_\theta \zeta_k, \\ Y_k = \beta_\theta \exp(X_k/2)\varepsilon_k, \end{cases} \tag{19}$$

where (ζ_k, ε_k) are i.i.d. Gaussian random variables in $\mathbb{R}^2$ with zero mean and identity covariance matrix, $\beta_\theta > 0$, $\sigma_\theta > 0$ for every $\theta \in \Theta$, and the functions $\theta \mapsto \phi_\theta$, $\theta \mapsto \sigma_\theta$, and $\theta \mapsto \beta_\theta$ are continuous. Then, assumptions (NL1)–(NL4) are satisfied as noted by Douc et al. [8], Remark 10.

Under (NL1), every compact set $\mathsf{C} \subset \mathsf{X} = \mathbb{R}^d$ with $\lambda^{\mathrm{Leb}}(\mathsf{C}) > 0$ is a 1-small set and therefore a local Doeblin with $\lambda_{\mathsf{C}}^\theta(\cdot) = \lambda^{\mathrm{Leb}}(\cdot \cap \mathsf{C})/\lambda^{\mathrm{Leb}}(\mathsf{C})$, $\varphi_{\mathsf{C}}^\theta\langle y_0\rangle = \lambda^{\mathrm{Leb}}(\mathsf{C})$ and

$$\epsilon_{\mathsf{C}}^- = \inf_{\theta\in\Theta} \inf_{(x,x')\in\mathsf{C}\times\mathsf{C}} q^\theta(x, x'),$$

$$\epsilon_{\mathsf{C}}^+ = \sup_{\theta\in\Theta} \sup_{(x,x')\in\mathsf{C}\times\mathsf{C}} q^\theta(x, x').$$

Under (NL1) and (NL2), (5) and (6) are satisfied with $r = 1$; equation (7) follows from (NL1) and (NL4). Thus assumption (A1) holds.

Assumption (A2) follows directly from (NL3). To establish (A3), it suffices to note that, under (NL1), for any $(x, x') \in \mathsf{X} \times \mathsf{X}$, $\theta \mapsto q^\theta(x, x')$ is continuous, under (NL3), for any $(x, y) \in \mathsf{X} \times \mathsf{Y}$, $\theta \mapsto g^\theta(x, y)$ is continuous and for any $n \in \mathbb{N}$, $\sup_{\theta\in\Theta} \sup_{x\in\mathsf{X}} \prod_{k=0}^n g^\theta(x, Y_k) < \infty$, $\mathbb{P}$-a.s. The bounded convergence theorem shows that, $\mathbb{P}$-a.s. the function $\theta \mapsto p_x^\theta(Y_0^n)$ is continuous.

Finally, under (NL1)–(NL4) according to Theorem 2 and Proposition 3 the MLE is consistent for any initial measure χ such that $\chi(\mathsf{D}) > 0$.

5. Proofs of Proposition 1 and Theorem 2.

5.1. *Block decomposition.* The first step of the proof consists of splitting the observations into blocks of size r where r is defined in (A1). More precisely, we will first show the equivalent of Proposition 1 and Theorem 2 with Y_i replaced by $Z_i \triangleq Y_{ir}^{(i+1)r-1}$. With this notation,

$$\hat{\theta}_{\chi,nr} = \underset{\theta\in\Theta}{\arg\max} \ln p_\chi^\theta(Y_0^{nr-1}) = \underset{\theta\in\Theta}{\arg\max} \ln p_\chi^\theta(Z_0^{n-1}).$$

In the following, $\hat{\theta}_{\chi,nr}$ is called the *block maximum likelihood estimator* (denoted hereafter as the block MLE) associated to the observations $Z_0, \dots, Z_{n-1}$.

5.1.1. *Forgetting of the initial distribution for the block conditional likelihood.* Denote, for $i \in \mathbb{Z}$,

$$z_i = y_{ir}^{(i+1)r-1} \in \mathsf{Y}^r. \tag{20}$$

Then, the likelihood $p_\chi^\theta(z_0^{n-1})$ may be rewritten as

$$p_\chi^\theta(z_0^{n-1}) = p_\chi^\theta(y_0^{nr-1}) = \chi \mathbf{L}^\theta\langle z_0\rangle \cdots \mathbf{L}^\theta\langle z_{n-1}\rangle \mathbb{1}_\mathsf{X} = \chi \mathbf{L}^\theta\langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}, \tag{21}$$

where $\mathbf{L}^\theta\langle z_0^{n-1}\rangle = \mathbf{L}^\theta\langle y_0^{nr-1}\rangle$ is defined in (2).

For any sequence $\{z_i\}_{i\ge 0} \in \mathsf{Z}^\mathbb{N}$ where $\mathsf{Z} \triangleq \mathsf{Y}^r$, any probability measures χ and χ' on $(\mathsf{X}, \mathcal{X})$ and any measurable nonnegative functions f and h from X to $\mathbb{R}^+$, define

$$\begin{aligned}\Delta_{\chi,\chi'}^\theta\langle z_0^{n-1}\rangle(f,h) &= \big(\chi\mathbf{L}^\theta\langle z_0^{n-1}\rangle f\big)\big(\chi'\mathbf{L}^\theta\langle z_0^{n-1}\rangle h\big)\\ &\quad - \big(\chi\mathbf{L}^\theta\langle z_0^{n-1}\rangle h\big)\big(\chi'\mathbf{L}^\theta\langle z_0^{n-1}\rangle f\big).\end{aligned} \tag{22}$$

Let $\bar{\mathsf{X}} = \mathsf{X}\times\mathsf{X}$ and $\bar{\mathcal{X}} = \mathcal{X}\otimes\mathcal{X}$. For P a (possibly unnormalized) kernel on $(\mathsf{X}, \mathcal{X})$, we denote by $\bar{P}$ the transition kernel on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$ defined, for any $(x, x') \in \bar{\mathsf{X}}$ and A, $\mathsf{A}' \in \mathcal{X}$, by

$$\bar{P}\big[(x,x'), \mathsf{A}\times\mathsf{A}'\big] = P(x,\mathsf{A})P(x',\mathsf{A}'). \tag{23}$$

If χ and χ' are two probability measures on $(\mathsf{X}, \mathcal{X})$ and f, g are real valued measurable functions on $(\mathsf{X}, \mathcal{X})$, define for $\bar{\mathsf{A}} \in \bar{\mathcal{X}}$ and $\bar{w} = (w, w') \in \bar{\mathsf{X}}$,

$$\chi\otimes\chi'(\bar{\mathsf{A}}) = \iint \chi(\mathrm{d}x)\chi'(\mathrm{d}x')\mathbb{1}_{\bar{\mathsf{A}}}(x,x'), \qquad f\otimes h(\bar{w}) = f(w)g(w'). \tag{24}$$

With the notation introduced above, (22) can be rewritten as follows:

$$\begin{aligned}\Delta_{\chi,\chi'}^\theta\langle z_0^{n-1}\rangle(f,h) &= \int\cdots\int \chi\otimes\chi'(\mathrm{d}\bar{w}_0)\left(\prod_{i=0}^{n-1}\bar{\mathbf{L}}^\theta\langle z_i\rangle(\bar{w}_i, \mathrm{d}\bar{w}_{i+1})\right)\\ &\quad \times \{f\otimes h - h\otimes f\}(\bar{w}_n).\end{aligned} \tag{25}$$

The following proposition extends [6], Proposition 12.

PROPOSITION 5. *Assume* (A1). *Let* $0 \le \gamma^- < \gamma^+ \le 1$. *Then, for any* $\eta > 0$, *there exists* $\rho \in (0,1)$ *such that, for any sequence* $(z_i)_{i \ge 0} \in \mathsf{Z}^{\mathbb{N}}$ *satisfying*

$$n^{-1} \sum_{i=0}^{n-1} \mathbb{1}_{\mathsf{K}}(z_i) \ge \max\bigl(1-\gamma^-, (1+\gamma^+)/2\bigr) \tag{26}$$

for any $\beta \in (\gamma^-, \gamma^+)$, *any nonnegative bounded functions* f *and* h, *any probability measures* χ *and* χ' *on* $(\mathsf{X}, \mathcal{X})$ *and any* $\theta \in \Theta$,

$$\begin{aligned} &\bigl|\Delta^{\theta}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h)\bigr| \\ &\quad \le \rho^{\lfloor n(\beta-\gamma^-)\rfloor}\bigl\{\bigl(\chi \mathbf{L}^{\theta}\langle z_0^{n-1}\rangle f\bigr)\bigl(\chi' \mathbf{L}^{\theta}\langle z_0^{n-1}\rangle g\bigr) + \bigl(\chi' \mathbf{L}^{\theta}\langle z_0^{n-1}\rangle f\bigr)\bigl(\chi \mathbf{L}^{\theta}\langle z_0^{n-1}\rangle g\bigr)\bigr\} \\ &\qquad + 2\eta^{\lfloor n(\gamma^+-\beta)\rfloor/2}\left[\prod_{i=0}^{n-1} \bigl|\mathbf{L}^{\theta}\langle z_i\rangle(\cdot, \mathsf{X})\bigr|_\infty^2\right] |f|_\infty |h|_\infty. \end{aligned}$$

PROOF. Let $\eta > 0$. According to (A1), there exists a set $\mathsf{C} \subset \mathsf{Y}$ such that (5) and (6) hold. Denote $\bar{\mathsf{C}} \triangleq \mathsf{C} \times \mathsf{C}$ and for $z = y_0^{r-1}$, set $\bar{\varphi}^{\theta}_{\mathsf{C}}\langle z\rangle = \varphi^{\theta}_{\mathsf{C}}\langle z\rangle \otimes \varphi^{\theta}_{\mathsf{C}}\langle z\rangle$ and $\bar{\lambda}^{\theta}_{\mathsf{C}}\langle z\rangle \triangleq \lambda^{\theta}_{\mathsf{C}}\langle z\rangle \otimes \lambda^{\theta}_{\mathsf{C}}\langle z\rangle$ where $\varphi^{\theta}_{\mathsf{C}}\langle z\rangle$ and $\lambda^{\theta}_{\mathsf{C}}\langle z\rangle$ are defined in Definition 1. For any measurable nonnegative function $\bar{f}$ on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$, $\theta \in \Theta$ and $\bar{x} \in \bar{\mathsf{C}}$,

$$\begin{aligned} &\bigl(\epsilon^-_{\mathsf{C}}(z)\bigr)^2 \bar{\varphi}^{\theta}_{\mathsf{C}}\langle z\rangle(\bar{x}) \bar{\lambda}^{\theta}_{\mathsf{C}}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}) \\ &\qquad \le \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}) \le \bigl(\epsilon^+_{\mathsf{C}}(z)\bigr)^2 \bar{\varphi}^{\theta}_{\mathsf{C}}\langle z\rangle(\bar{x}) \bar{\lambda}^{\theta}_{\mathsf{C}}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}). \end{aligned} \tag{27}$$

Define the unnormalized kernel $\bar{\mathbf{L}}^{\theta,0}\langle z\rangle$ and $\bar{\mathbf{L}}^{\theta,1}\langle z\rangle$ on $(\bar{\mathsf{X}}, \bar{\mathcal{X}})$ as follows: for all $\bar{x} \in \bar{\mathsf{X}}$ and $\bar{A} \in \bar{\mathcal{X}}$,

$$\bar{\mathbf{L}}^{\theta,0}\langle z\rangle(\bar{x}, \bar{A}) \triangleq \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x})\bigl(\epsilon^-_{\mathsf{C}}(z)\bigr)^2 \bar{\varphi}^{\theta}_{\mathsf{C}}\langle z\rangle(\bar{x}) \bar{\lambda}^{\theta}_{\mathsf{C}}\langle z\rangle(\bar{\mathsf{C}} \cap \bar{A}), \tag{28}$$

$$\bar{\mathbf{L}}^{\theta,1}\langle z\rangle(\bar{x}, \bar{A}) \triangleq \bar{\mathbf{L}}^{\theta}\langle z\rangle(\bar{x}, \bar{A}) - \bar{\mathbf{L}}^{\theta,0}\langle z\rangle(\bar{x}, \bar{A}). \tag{29}$$

Equation (27) implies that, for all $\bar{x} \in \bar{\mathsf{C}}$, and any measurable nonnegative function $\bar{f}$,

$$0 \le \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta,1}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}) \le r_{\mathsf{C}}(z) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}),$$

where $r_{\mathsf{C}}(z) \triangleq 1 - (\epsilon^-_{\mathsf{C}}(z)/\epsilon^+_{\mathsf{C}}(z))^2$. It then follows

$$\begin{aligned} &\delta_{\bar{x}} \bar{\mathbf{L}}^{\theta,1}\langle z\rangle(\bar{f}) \\ &\quad = \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x}) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta,1}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x}) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta,1}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}^c} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x}) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta,1}\langle z\rangle(\bar{f}) \\ &\quad \le r_{\mathsf{C}}(z) \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x}) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}}(\bar{x}) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta}\langle z\rangle(\mathbb{1}_{\bar{\mathsf{C}}^c} \bar{f}) + \mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x}) \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta}\langle z\rangle(\bar{f}) \\ &\quad \le \delta_{\bar{x}} \bar{\mathbf{L}}^{\theta}\langle z\rangle\bigl(r_{\mathsf{C}}(z)^{\mathbb{1}_{\bar{\mathsf{C}}}(\bar{x}) \mathbb{1}_{\bar{\mathsf{C}}}} \bar{f}\bigr). \end{aligned} \tag{30}$$

Note that $\Delta^{\theta}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h)$ may be decomposed as

$$\Delta^{\theta}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h) = \sum_{t_0^{n-1}\in\{0,1\}^n} \Delta^{\theta,t_0^{n-1}}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h),$$

where

$$\Delta^{\theta,t_0^{n-1}}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h) = \int\cdots\int \chi\otimes\chi'(\mathrm{d}\bar{w}_0')\left(\prod_{i=0}^{n-1}\bar{\mathbf{L}}^{\theta,t_i}\langle z_i\rangle(\bar{w}_i,\mathrm{d}\bar{w}_{i+1})\right)\Phi(\bar{w}_n)$$

with $\Phi \triangleq f\otimes h - h\otimes f$. First assume that there exists an index $i\in\{0,\dots,n-1\}$ such that $t_i = 0$. Then

$$\begin{aligned}\Delta^{\theta,t_0^{n-1}}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h) &= \chi\otimes\chi'\big(\bar{\mathbf{L}}^{\theta,t_0}\langle z_0\rangle\cdots\bar{\mathbf{L}}^{\theta,t_{i-1}}\langle z_{i-1}\rangle\big(\mathbb{1}_{\bar{\mathsf{C}}}\times\bar{\varphi}^{\theta}_{\mathsf{C}}\langle z_i\rangle\big)\big)\\ &\quad\times\big(\epsilon^-_{\mathsf{C}}(z_i)\big)^2\bar{\lambda}^{\theta}_{\mathsf{C}}\langle z_i\rangle\big(\mathbb{1}_{\bar{\mathsf{C}}}\bar{\mathbf{L}}^{\theta,t_{i+1}}\langle z_{i+1}\rangle\cdots\bar{\mathbf{L}}^{\theta,t_{n-1}}\langle z_{n-1}\rangle\Phi\big).\end{aligned}$$

By symmetry,

$$\bar{\lambda}^{\theta}_{\mathsf{C}}\langle z_i\rangle\big(\mathbb{1}_{\bar{\mathsf{C}}}\bar{\mathbf{L}}^{\theta,t_{i+1}}\langle z_{i+1}\rangle\cdots\bar{\mathbf{L}}^{\theta,t_{n-1}}\langle z_{n-1}\rangle\Phi\big) = 0,$$

showing that $\Delta^{\theta,t_0^{n-1}}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h) = 0$ except if for all $i\in\{0,\dots,n-1\}$, $t_i = 1$. Therefore,

$$\Delta^{\theta}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h) = \chi\otimes\chi'\big(\bar{\mathbf{L}}^{\theta,1}\langle z_0\rangle\cdots\bar{\mathbf{L}}^{\theta,1}\langle z_{n-1}\rangle\Phi\big).$$

This implies, using (30), that

$$\begin{aligned}&\big|\Delta^{\theta}_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h)\big|\\ &\quad\le \chi\otimes\chi'\big(\bar{\mathbf{L}}^{\theta,1}\langle z_0\rangle\cdots\bar{\mathbf{L}}^{\theta,1}\langle z_{n-1}\rangle|\Phi|\big)\\ &\quad\le \int\cdots\int\chi\otimes\chi'(\mathrm{d}\bar{w}_0)\left(\prod_{i=0}^{n-1}\bar{\mathbf{L}}^{\theta}\langle z_i\rangle(\bar{w}_i,\mathrm{d}\bar{w}_{i+1})\big(r_{\mathsf{C}}(z_i)\big)^{\mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i,\bar{w}_{i+1})}\right)\\ &\qquad\times|\Phi|(\bar{w}_n).\end{aligned}\tag{31}$$

Note that

$$\prod_{i=0}^{n-1}\big(r_{\mathsf{C}}(z_i)\big)^{\mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i,\bar{w}_{i+1})} \le \varrho_{\mathsf{C}}^{\sum_{i=0}^{n-1}\mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i,\bar{w}_{i+1})\mathbb{1}_{\mathsf{K}}(z_i)},\tag{32}$$

where $\varrho_{\mathsf{C}} \triangleq \sup_{z\in\mathsf{K}} r_{\mathsf{C}}(z) < 1$ under (A1). For any sequence z_0^{n-1} such that $n^{-1}\sum_{i=0}^{n-1}\mathbb{1}_{\mathsf{K}}(z_i)\ge(1-\gamma^-)$, we have $\sum_{i=0}^{n-1}\mathbb{1}_{\mathsf{K}^c}(z_i)\le n\gamma^-$, so that

$$\sum_{i=0}^{n-1}\mathbb{1}_{\mathsf{K}^c}(z_i)\le\lfloor n\gamma^-\rfloor.$$

Moreover, we have

$$\begin{aligned}
&\sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i, \bar{w}_{i+1})\mathbb{1}_{\mathsf{K}}(z_i)\\
&\qquad = \sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i, \bar{w}_{i+1}) - \sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i, \bar{w}_{i+1})\mathbb{1}_{\mathsf{K}^c}(z_i)\\
&\qquad \geq N_{\bar{\mathsf{C}},n}(\bar{w}_0^n) - \sum_{i=0}^{n-1} \mathbb{1}_{\mathsf{K}^c}(z_i)\\
&\qquad \geq N_{\bar{\mathsf{C}},n}(\bar{w}_0^n) - \lfloor n\gamma^- \rfloor,
\end{aligned} \tag{33}$$

where, for any set $\bar{\mathsf{A}} \in \bar{\mathcal{X}}$, $N_{\bar{\mathsf{A}},n}(\bar{w}_0^n) = \sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{A}}\times\bar{\mathsf{A}}}(\bar{w}_i, \bar{w}_{i+1})$. By combining (32) and (33) and using that $\lfloor n\beta \rfloor - \lfloor n\gamma^- \rfloor \geq \lfloor n(\beta - \gamma^-)\rfloor$, we therefore obtain, for any $\beta \in (\gamma^-, 1]$,

$$\prod_{i=0}^{n-1} \big(r_{\mathsf{C}}(z_i)\big)^{\mathbb{1}_{\bar{\mathsf{C}}\times\bar{\mathsf{C}}}(\bar{w}_i,\bar{w}_{i+1})} \leq \varrho_{\mathsf{C}}^{\lfloor n(\beta-\gamma^-)\rfloor} + \mathbb{1}\{N_{\bar{\mathsf{C}},n}(\bar{w}_0^n) < \lfloor n\beta \rfloor\}. \tag{34}$$

For any sequence $\bar{w}_0^{n-1} \in \bar{\mathsf{X}}^n$ and any $\bar{\mathsf{A}} \in \bar{\mathcal{X}}$, denote

$$M_{\bar{\mathsf{A}},n}(\bar{w}_0^{n-1}) \triangleq \sum_{i=0}^{n-1} \mathbb{1}_{\bar{\mathsf{A}}}(\bar{w}_i).$$

Using [6], Lemma 17, for any sequence $\bar{w}_0^n$ satisfying $N_{\bar{\mathsf{C}},n}(\bar{w}_0^n) < \lfloor n\beta \rfloor$ which is equivalent to $N_{\bar{\mathsf{C}},n}(\bar{w}_0^n) \leq \lfloor n\beta \rfloor - 1$, we have $M_{\bar{\mathsf{C}},n}(\bar{w}_0^{n-1}) \leq (\lfloor n\beta \rfloor + n)/2$, so that

$$N_{\bar{\mathsf{C}},n}(\bar{w}_0^n) < \lfloor n\beta \rfloor \quad \Rightarrow \quad M_{\bar{\mathsf{C}}^c,n}(\bar{w}_0^{n-1}) \geq a_n \triangleq \frac{n - \lfloor n\beta \rfloor}{2}. \tag{35}$$

In words, either the number of consecutive visits to the set $\bar{\mathsf{C}}$ at most $\lfloor n\beta \rfloor$, or the number of visits to the complementary of the set $\bar{\mathsf{C}}$ is larger than a_n. Plugging (35) into (34) and combining it with (31) yields

$$\begin{aligned}
\big|\Delta^\theta_{\chi,\chi'}\langle z_0^n\rangle(f,h)\big| &\leq \varrho_{\mathsf{C}}^{\lfloor n(\beta-\gamma^-)\rfloor} \chi \otimes \chi'\big(\bar{\mathbf{L}}^\theta\langle z_0\rangle \cdots \bar{\mathbf{L}}^\theta\langle z_{n-1}\rangle |\Phi|\big)\\
&\quad + 2|f|_\infty |h|_\infty \Gamma^\theta_{\chi,\chi'}(z_0^{n-1}),
\end{aligned}$$

where

$$\Gamma^\theta_{\chi,\chi'}(z_0^{n-1}) \triangleq \int \cdots \int \chi \otimes \chi'(\mathrm{d}\bar{w}_0) \prod_{i=0}^{n-1} \bar{\mathbf{L}}^\theta\langle z_i\rangle(\bar{w}_i, \mathrm{d}\bar{w}_{i+1}) \mathbb{1}\{M_{\bar{\mathsf{C}}^c,n}(\bar{w}_0^{n-1}) \geq a_n\}.$$

We finally have to bound this last term. First rewrite $\Gamma^\theta_{\chi,\chi'}(z_0^{n-1})$ as follows:

$$\Gamma^\theta_{\chi,\chi'}(z_0^{n-1}) = \left(\prod_{i=0}^{n-1} \left|\mathbf{L}^\theta\langle z_i\rangle(\cdot,\mathsf{X})\right|_\infty^2\right) \int \chi\otimes\chi'(\mathrm{d}\bar{w}_0)\left(\eta^{\sum_{i=0}^{n-1}\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{w}_i)\mathbb{1}_{\mathsf{K}}(z_i)}\right)$$
$$\times\left(\prod_{i=0}^{n-1}\frac{\bar{\mathbf{L}}^\theta\langle z_i\rangle(\bar{w}_i,\mathrm{d}\bar{w}_{i+1})}{\eta^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{w}_i)\mathbb{1}_{\mathsf{K}}(z_i)}|\mathbf{L}^\theta\langle z_i\rangle(\cdot,\mathsf{X})|_\infty^2}\right)\mathbb{1}\{M_{\bar{\mathsf{C}}^c,n}(\bar{w}_0^{n-1})\geq a_n\}.$$

Note that (26) implies that $\sum_{i=0}^{n-1}\mathbb{1}_{\mathsf{K}}(z_i)\geq (n+\lfloor n\gamma^+\rfloor)/2$. Then, for any $\gamma^+>\beta$, the inequality $M_{\bar{\mathsf{C}}^c,n}(\bar{w}_0^{n-1})\geq a_n$ implies that

$$\sum_{i=0}^{n-1}\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x}_i)\mathbb{1}_{\mathsf{K}}(z_i)\geq\sum_{i=0}^{n-1}\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x}_i)-\sum_{i=0}^{n-1}\mathbb{1}_{\mathsf{K}^c}(z_i)\geq\frac{\lfloor n\gamma^+\rfloor-\lfloor n\beta\rfloor}{2}\geq\frac{\lfloor n(\gamma^+-\beta)\rfloor}{2},$$

showing that

$$\left(\eta^{\sum_{i=0}^{n-1}\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{x}_i)\mathbb{1}_{\mathsf{K}}(z_i)}\right)\mathbb{1}\{M_{\bar{\mathsf{C}}^c,n}(\bar{x}_0^{n-1})\geq a_n\}\leq\eta^{\lfloor n(\gamma^+-\beta)\rfloor/2}.$$

The proof follows noting that, for any $\bar{w}=(w,w')\in\bar{\mathsf{X}}$ and $z\in\mathsf{Y}^r$, (3) and (5) imply

$$\iint\frac{\bar{\mathbf{L}}^\theta\langle z\rangle(\bar{w},\mathrm{d}\bar{w}_{i+1})}{\eta^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{w})\mathbb{1}_{\mathsf{K}}(z)}|\mathbf{L}^\theta\langle z\rangle(\cdot,\mathsf{X})|_\infty^2}=\frac{\mathbf{L}^\theta\langle z\rangle(w,\mathsf{X})\mathbf{L}^\theta\langle z\rangle(w',\mathsf{X})}{\eta^{\mathbb{1}_{\bar{\mathsf{C}}^c}(\bar{w})\mathbb{1}_{\mathsf{K}}(z)}|\mathbf{L}^\theta\langle z\rangle(\cdot,\mathsf{X})|_\infty^2}\leq 1. \qquad \square$$

LEMMA 6. *Let $(U_k)_{k\in\mathbb{Z}}$, $(V_k)_{k\in\mathbb{Z}}$, $(W_k)_{k\in\mathbb{Z}}$ be stationary sequences such that*

$$\mathbb{E}[\ln^+ U_0]<\infty,\qquad \mathbb{E}[\ln^+ V_0]<\infty,\qquad \mathbb{E}[\ln^+ W_0]<\infty.$$

Then, for all η,ρ in $(0,1)$ such that $-\ln\eta>\mathbb{E}[\ln^+ V_0]$, there exists a $\mathbb{P}$-a.s. finite random variable D and a constant $\varrho\in(0,1)$ such that for all $k\geq 1, m\geq 0$,

$$\rho^{k+m}+\eta^{k+m}W_{-m}\left(\prod_{i=-m}^{k-1}V_i\right)U_k\leq\varrho^{k+m}D,\qquad \mathbb{P}\text{-a.s.}$$

PROOF. Let $\alpha\in(0,1)$ such that $\mathbb{E}[\ln^+ V_0]<-\ln\alpha<-\ln\eta$, and let $\tilde{\alpha}>0$ such that $(\eta/\alpha)\vee\rho<\tilde{\alpha}<1$. Then

$$\rho^{k+m}+\eta^{k+m}W_{-m}\left(\prod_{i=-m}^{k-1}V_i\right)U_k$$
$$=\left[\left(\frac{\rho}{\tilde{\alpha}}\right)^{k+m}\tilde{\alpha}^m+\left(\frac{\eta}{\alpha\tilde{\alpha}}\right)^{k+m}(\tilde{\alpha}^mW_{-m})\left(\prod_{i=-m}^{k-1}(V_i\alpha)\right)(\tilde{\alpha}^kU_k)\right]$$
$$\leq\left(\frac{\rho}{\tilde{\alpha}}\vee\frac{\eta}{\alpha\tilde{\alpha}}\right)^{k+m}D$$

with

$$D \triangleq 1 + \Big(\sup_{m\ge 0} \tilde{\alpha}^m W_{-m}\Big)\Bigg(\sup_{m\ge 0} \prod_{i=-m}^{0} (V_i\alpha)\Bigg)\Bigg(\sup_{k\ge 1} \prod_{i=1}^{k-1} (V_i\alpha)\Bigg)\Big(\sup_{k\ge 1} \tilde{\alpha}^k U_k\Big).$$

We now show that D is $\mathbb{P}$-a.s. finite. First note that combining the bound $\mathbb{E}[\ln^+ U_0 < \infty]$ with Lemma 7 (stated and proved below), we obtain that the random variable $\sup_{k\ge 1} \tilde{\alpha}^k U_k$ is $\mathbb{P}$-a.s. finite; in the same way, $\sup_{m\ge 0} \tilde{\alpha}^m W_{-m}$ is $\mathbb{P}$-a.s. finite. Moreover, since $\mathbb{E}[\ln^+ V_0] < \infty$, Birkoff's ergodic theorem ensures that

$$\frac{1}{k-1} \sum_{i=1}^{k-1} \ln^+ V_i \to_{k\to\infty} \mathbb{E}[\ln^+ V_0] < -\ln\alpha, \qquad \mathbb{P}\text{-a.s.}$$

By taking the exponential function in the previous limit, we obtain that

$$\prod_{i=1}^{k-1} (V_i\alpha) \le \exp\Bigg\{(k-1)\Bigg(\frac{1}{k-1}\sum_{i=1}^{k-1} \ln^+ V_i + \ln\alpha\Bigg)\Bigg\} \to_{k\to\infty} 0, \qquad \mathbb{P}\text{-a.s.}$$

so that $\sup_{k\ge 1} \prod_{i=1}^{k-1} (V_i\alpha)$ is $\mathbb{P}$-a.s. finite. Following the same arguments,

$$\sup_{m\ge 0} \prod_{i=-m}^{0} (V_i\alpha)$$

is $\mathbb{P}$-a.s. finite. Finally D is $\mathbb{P}$-a.s. finite. The proof is complete. $\square$

LEMMA 7. *Let $\{Z_k\}_{k\in\mathbb{Z}}$ be a sequence of nonnegative random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ having the same marginal distribution, that is, for any $k \in \mathbb{Z}$ and any measurable nonnegative function f, $\mathbb{E}[f(Z_k)] = \mathbb{E}[f(Z_0)]$.*

(i) *Assume that $\mathbb{E}[(\ln Z_0)^+] < \infty$. Then, for all $\beta \in (0, 1)$, $\sup_{k\ge 0} \beta^k Z_k < \infty$, $\mathbb{P}$-a.s.*

(ii) *Assume that $\mathbb{E}[|\ln Z_0|] < \infty$. Then, for all $\beta \in (0, 1)$, $\sup_{k\in\mathbb{Z}} \beta^{|k|} Z_k < \infty$ and $\inf_{k\in\mathbb{Z}} \beta^{-|k|} Z_k > 0$, $\mathbb{P}$-a.s.*

PROOF. Let $\beta \in (0, 1)$. Since

$$\mathbb{P}\big[\beta^k Z_k > 1\big] = \mathbb{P}\big[\ln Z_k/(-\ln\beta) \ge k\big] = \mathbb{P}\big[\ln Z_0/(-\ln\beta) \ge k\big],$$

it follows that

$$\sum_{k=0}^{\infty} \mathbb{P}\big[\beta^k Z_k > 1\big] = \sum_{k=0}^{\infty} \mathbb{P}\big[\ln Z_0/(-\ln\beta) \ge k\big] \le \mathbb{E}\big[(\ln Z_0)^+\big]/(-\ln\beta) < \infty.$$

The proof of (i) is completed by using the Borel–Cantelli lemma. Now, (ii) can be easily derived by noting that if $\mathbb{E}[|\ln Z_0|] < \infty$, then one may use twice (i), first by replacing Z_k by Z_{-k} and then by replacing Z_k by $1/Z_k$. $\square$

PROPOSITION 8. *Assume (A1) and (A2). There exist a constant* $\kappa \in (0,1)$, *an integer-valued random variable* K *satisfying* $\mathbb{P}_Y[K < \infty] = 1$ *such that, for any initial distributions* $\chi, \chi' \in \mathcal{M}(\mathsf{D}, r)$ [*where* $\mathcal{M}(\mathsf{D}, r)$ *is defined in* (8)],

$$\sup_{\theta\in\Theta}\sup_{k\geq K}\sup_{m\geq 0}\kappa^{-(m+k)}\big|\ln p_\chi^\theta(Z_k|Z_{-m}^{k-1}) - \ln p_{\chi'}^\theta(Z_k|Z_{-m}^{k-1})\big| < \infty, \quad \mathbb{P}\text{-}a.s., \tag{36}$$

$$\sup_{\theta\in\Theta}\sup_{k\geq K}\sup_{m\geq 0}\kappa^{-(m+k)}\big|\ln p_\chi^\theta(Z_k|Z_{-m}^{k-1}) - \ln p_{\chi}^\theta(Z_k|Z_{-m-1}^{k-1})\big| < \infty, \quad \mathbb{P}\text{-}a.s., \tag{37}$$

$$\sup_{\theta\in\Theta}\sup_{m\geq 0}\kappa^{-m}\big|\ln p_\chi^\theta(Z_0|Z_{-m}^{-1}) - \ln p_{\chi}^\theta(Z_0|Z_{-m-1}^{-1})\big| < \infty, \quad \mathbb{P}\text{-}a.s. \tag{38}$$

PROOF. *Proof of* (36). It follows from (21) that, for any integer $(m,k) \in \mathbb{N}$ and any sequence z_{-m}^k,

$$p_\chi^\theta(z_k|z_{-m}^{k-1}) = \frac{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X})}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X})}.$$

Since, for any $a, b > 0$, $\ln(a) - \ln(b) \leq (a-b)/b$, definition (22) implies that

$$\begin{aligned}&\ln p_\chi^\theta(z_k|z_{-m}^{k-1}) - \ln p_{\chi'}^\theta(z_k|z_{-m}^{k-1})\\ &\quad\leq \frac{\Delta_{\chi,\chi'}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X}, \mathbb{1}_\mathsf{X})}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X})}.\end{aligned}\tag{39}$$

Let $0 \leq \gamma^- < \gamma^+ \leq 1$. By Proposition 5, for any $\eta > 0$ and $\beta \in (\gamma^-, \gamma^+)$ there exists $\varrho \in (0,1)$ such that, for any sequence z_{-m}^{k-1} satisfying

$$(m+k)^{-1}\sum_{i=-m}^{k-1}\mathbb{1}_\mathsf{K}(z_i) \geq \max\big(1-\gamma^-, (1+\gamma^+)/2\big), \tag{40}$$

we have

$$\begin{aligned}&\frac{\Delta_{\chi,\chi'}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X}, \mathbb{1}_\mathsf{X})}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X})}\\ &\qquad\leq \varrho^{a(m+k)}\bigg[1 + \frac{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X})}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X})}\bigg]\\ &\qquad\quad + 2\eta^{b(m+k)}C_{m,k},\end{aligned}\tag{41}$$

where $a(n) = \lfloor n(\beta - \gamma^-)\rfloor$, $b(n) = \lfloor n(\gamma^+ - \beta)\rfloor/2$ and

$$C_{m,k} \triangleq \frac{\prod_{i=-m}^{k-1} |\mathbf{L}^\theta\langle z_i\rangle(\cdot, \mathsf{X})|_\infty^2}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X})}\left|\mathbf{L}^\theta\langle z_k\rangle(\cdot, \mathsf{X})\right|_\infty. \tag{42}$$

Moreover, by (22),

$$\begin{aligned}
&\frac{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X})}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X})}\\
&\qquad = \frac{\Delta^\theta_{\chi,\chi'}\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X}, \mathbb{1}_\mathsf{X})}{\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) \times \chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X})} + 1.
\end{aligned}$$

Plugging this identity into (41) and then using (39) yields

$$\begin{aligned}
&\ln p^\theta_\chi(z_k|z_{-m}^{k-1}) - \ln p^\theta_{\chi'}(z_k|z_{-m}^{k-1})\\
&\qquad \le 2(1 - \varrho^{a(m+k)})^{-1}[\varrho^{a(m+k)} + \eta^{b(m+k)}C_{m,k}].
\end{aligned} \tag{43}$$

For any sequence z_{-m}^{k-1}, we have

$$\begin{aligned}
\chi\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbb{1}_\mathsf{X}) &\ge \chi(\mathsf{D})\prod_{i=-m}^{k-1}\Big\{\inf_{x\in\mathsf{D}}\mathbf{L}^\theta\langle z_i\rangle(x, \mathsf{D})\Big\},\\
\chi'\mathbf{L}^\theta\langle z_{-m}^{k-1}\rangle(\mathbf{L}^\theta\langle z_k\rangle\mathbb{1}_\mathsf{X}) &\ge \chi'(\mathsf{D})\prod_{i=-m}^{k}\Big\{\inf_{x\in\mathsf{D}}\mathbf{L}^\theta\langle z_i\rangle(x, \mathsf{D})\Big\}.
\end{aligned} \tag{44}$$

Exchanging χ and χ' in (43) allows us to obtain an upper bound for $|\ln p^\theta_\chi(z_k| z_{-m}^{k-1}) - \ln p^\theta_{\chi'}(z_k|z_{-m}^{k-1})|$. More precisely, for any sequence z_{-m}^{k-1} satisfying (40), we have

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\ln p^\theta_\chi(z_k|z_{-m}^{k-1}) - \ln p^\theta_{\chi'}(z_k|z_{-m}^{k-1})\right|\\
&\qquad \le 2(1 - \varrho^{a(m+k)})^{-1}\\
&\qquad\qquad \times \left\{\varrho^{a(m+k)} + \frac{\eta^{b(m+k)}}{\chi(\mathsf{D})\chi'(\mathsf{D})}\left[\prod_{j=-m}^{k-1}(D_{z_j})^2\right]D_{z_k}\right\},
\end{aligned} \tag{45}$$

where, for $z \in \mathsf{Y}^r$,

$$D_z = \frac{\sup_{\theta\in\Theta}|\mathbf{L}^\theta\langle z\rangle(\cdot, \mathsf{X})|_\infty}{\inf_{\theta\in\Theta}\inf_{x\in\mathsf{D}}\mathbf{L}^\theta\langle z\rangle(x, \mathsf{D})}. \tag{46}$$

Assume that $\mathbb{E}[\ln^+(D_{Z_0})] < \infty$, and set η small enough so that $\mathbb{E}[\ln^+(D_{Z_0})] \le -\ln\eta$. By Lemma 6, there exists a $\mathbb{P}$-a.s. finite random variable C, and a constant

$\kappa \in (0,1)$ such that, for all $k \geq 1$, $m \geq 0$,

$$\frac{2}{1-\varrho^{a(m+k)}}\left\{\varrho^{a(m+k)} + \frac{\eta^{b(m+k)}}{\chi(\mathsf{D})\chi'(\mathsf{D})}\left[\prod_{j=-m}^{k-1}(D_{z_j})^2\right]D_{z_k}\right\} \leq C\kappa^{k+m}, \qquad \mathbb{P}\text{-a.s.}$$

It remains to show that $\mathbb{E}[\ln^+(D_{Z_0})] < \infty$. Since for any $a, b > 0$, $\ln^+(a/b) \leq \ln^+(a) + \ln^-(b)$,

$$\ln^+(D_z) \leq \ln^+\Big(\sup_{\theta\in\Theta}\big|\mathbf{L}^\theta\langle z\rangle(\cdot,\mathsf{X})\big|_\infty\Big) + \ln^-\Big(\inf_{\theta\in\Theta}\inf_{x\in\mathsf{D}}\mathbf{L}^\theta\langle z\rangle(x,\mathsf{D})\Big). \tag{47}$$

Since, for any $z = y_0^{r-1} \in \mathsf{Y}^r$, $\sup_{\theta\in\Theta}|\mathbf{L}^\theta\langle z\rangle(\cdot,\mathsf{X})|_\infty \leq \prod_{i=0}^{r-1}\sup_{\theta\in\Theta}|g^\theta(\cdot,y_i)|_\infty$, (A1)(iii) and (A2) imply that $\mathbb{E}[\ln^+(D_{Z_0})] < \infty$. Finally, according to (45),

$$\sup_{\theta\in\Theta}\big|\ln p_\chi^\theta(Z_k|Z_{-m}^{k-1}) - \ln p_{\chi'}^\theta(Z_k|Z_{-m}^{k-1})\big| \leq C\kappa^{m+k}, \qquad \mathbb{P}\text{-a.s.},$$

provided that

$$(m+k)^{-1}\sum_{j=-m}^{k-1}\mathbb{1}_{\mathsf{K}}(Z_j) \geq \max\big(1-\gamma^-, (1+\gamma^+)/2\big), \qquad \mathbb{P}\text{-a.s.} \tag{48}$$

It thus remains to show the existence of a $\mathbb{P}$-a.s. finite random variable K such that for any $k \geq K$ and any $m \geq 0$, (48) holds $\mathbb{P}$-a.s. Under (A1)(i), $1 - \mathbb{P}[Z_0 \in \mathsf{K}] < 2\mathbb{P}[Z_0 \in \mathsf{K}] - 1$. Then, choose $\tilde{\gamma}^-$, γ^-, γ^+ and $\tilde{\gamma}^+$ such that

$$1 - \mathbb{P}[Z_0 \in \mathsf{K}] < \tilde{\gamma}^- < \gamma^- < \gamma^+ < \tilde{\gamma}^+ < 2\mathbb{P}[Z_0 \in \mathsf{K}] - 1. \tag{49}$$

By construction $(1+\tilde{\gamma}^+)/2 < \mathbb{P}_Y[Z_0 \in \mathsf{K}]$ and $1 - \tilde{\gamma}^- < \mathbb{P}[Z_0 \in \mathsf{K}]$. Since $(Z_k)_{k\in\mathbb{Z}}$ is stationary and ergodic, the Birkhoff ergodic theorem ensures that there exists a $\mathbb{P}$-a.s. finite random variable B such that for any $k \geq B$ and $m \geq B$, $\mathbb{P}$-a.s.,

$$\max\Big(1-\tilde{\gamma}^-, \frac{1+\tilde{\gamma}^+}{2}\Big) < k^{-1}\sum_{i=0}^{k-1}\mathbb{1}_{\mathsf{K}}(Z_i), \tag{50}$$

$$\max\Big(1-\tilde{\gamma}^-, \frac{1+\tilde{\gamma}^+}{2}\Big) < m^{-1}\sum_{i=-m}^{-1}\mathbb{1}_{\mathsf{K}}(Z_i). \tag{51}$$

Set $K^+ \triangleq B(1+\gamma^+)/(\tilde{\gamma}^+ - \gamma^+)$. If $m \geq B$ and $k \geq K^+$, then using that $K^+ \geq B$, $\mathbb{P}$-a.s.,

$$\frac{\sum_{i=-m}^{k-1}\mathbb{1}_{\mathsf{K}}(Z_i)}{k+m} > \frac{k(1+\tilde{\gamma}^+)/2 + m(1+\tilde{\gamma}^+)/2}{k+m} = \big(1+\tilde{\gamma}^+\big)/2 > \big(1+\gamma^+\big)/2.$$

Now, if $0 \leq m < B$ and $k \geq K^+$,

$$\begin{aligned}\frac{\sum_{i=-m}^{k-1}\mathbb{1}_{\mathsf{K}}(Z_i)}{k+m} &\geq \frac{\sum_{i=0}^{k-1}\mathbb{1}_{\mathsf{K}}(Z_i)}{k+m} > \frac{k(1+\tilde{\gamma}^+)/2}{k+m}\\ &> \frac{K^+(1+\tilde{\gamma}^+)/2}{K^+ + B} = (1+\gamma^+)/2.\end{aligned}$$

Similarly, setting $K^- \triangleq B(1-\gamma^-)/(\tilde{\gamma}^- - \gamma^-)$, we obtain, for all $m \ge 0$ and all $k \ge K^-$ that, $\mathbb{P}$-a.s.,

$$\frac{\sum_{i=-m}^{k-1} \mathbb{1}_{\mathsf{K}}(Z_i)}{k+m} \ge 1-\gamma^-.$$

The proof of (36) is now completed by setting $K = K^+ \vee K^-$.

Proof of (37). Note that

$$p_\chi^\theta(z_k|z_{-m-1}^{k-1}) = p_{\chi'}^\theta(z_k|z_{-m}^{k-1})$$

with $\chi'(\mathsf{A}) = \chi(\mathbf{L}^\theta\langle z_{-m-1}\rangle \mathbb{1}_{\mathsf{A}})/\chi(\mathbf{L}^\theta\langle z_{-m-1}\rangle \mathbb{1}_{\mathsf{X}})$. Since

$$\frac{1}{\chi'(\mathsf{D})} = \frac{\chi(\mathbf{L}^\theta\langle z_{-m-1}\rangle \mathbb{1}_{\mathsf{X}})}{\chi(\mathbf{L}^\theta\langle z_{-m-1}\rangle \mathbb{1}_{\mathsf{D}})} \le \frac{D_{z_{-m-1}}}{\chi(\mathsf{D})},$$

where D_z is defined in (46), (45) writes

$$\begin{aligned}
\sup_{\theta\in\Theta} &\left|\ln p_\chi^\theta(z_k|z_{-m}^{k-1}) - \ln p_\chi^\theta(z_k|z_{-m-1}^{k-1})\right| \\
&\le 2(1-\varrho^{a(m+k)})^{-1} \\
&\quad \times \left[\varrho^{a(m+k)} + \frac{\eta^{b(m+k)}}{[\chi(\mathsf{D})]^2} D_{z_{-m-1}} \prod_{j=-m}^{k-1} (D_{z_j})^2 D_{z_k}\right].
\end{aligned}$$

And the rest of the proof of (37) follows the same lines as (36) and is omitted for brevity.

Proof of (38). Noting that, when $k=0$, equation (48) follows immediately from (51), the proof of (38) follows the same lines as the proof of (37) and is omitted for brevity. □

COROLLARY 9 (Corollary of Proposition 8). *Assume* (A1) *and* (A2). *For any* $\theta \in \Theta$, *there exists a measurable function* $\pi_Z^\theta : \mathsf{Z}^{\mathbb{Z}^-} \to \mathbb{R}$ *such that for any probability measure* χ *satisfying* $\chi(\mathsf{D}) \in \mathcal{M}(\mathsf{D}, r)$ [*where* $\mathcal{M}(\mathsf{D}, r)$ *is defined in* (8)],

$$\mathbb{P}_Y\Big[\lim_{m\to\infty} p_\chi^\theta(Z_0|Z_{-m}^{-1}) = \pi_Z^\theta(Z_{-\infty}^0)\Big] = 1. \tag{52}$$

In the sequel, we denote $p^\theta(Z_0|Z_{-\infty}^{-1}) \triangleq \pi_Z^\theta(Z_{-\infty}^0)$ and for $n \ge 0$, $p^\theta(Z_0^n|Z_{-\infty}^{-1}) \triangleq \prod_{i=0}^n \pi_Z^\theta(Z_{-\infty}^i)$.

5.1.2. *Consistency of the block MLE.*

PROPOSITION 10. *Assume* (A1) *and* (A2). *Then*:

(i) *For any* $\theta \in \Theta$,

$$\mathbb{E}\big[\big|\ln p^{\theta}(Z_0|Z_{-\infty}^{-1})\big|\big] < \infty. \tag{53}$$

(ii) *For any probability measure* $\chi \in \mathcal{M}(\mathsf{D}, r)$ [*where* $\mathcal{M}(\mathsf{D}, r)$ *is defined in* (8)],

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} \big|n^{-1} \ln p_{\chi}^{\theta}(Z_0^{n-1}) - n^{-1} \ln p^{\theta}(Z_0^{n-1}|Z_{-\infty}^{-1})\big| = 0, \qquad \mathbb{P}\text{-}a.s.$$

(iii) *For any* $\theta \in \Theta$, *and for any probability measure* $\chi \in \mathcal{M}(\mathsf{D}, r)$,

$$\lim_{n\to\infty} n^{-1} \ln p_{\chi}^{\theta}(Z_0^{n-1}) = \mathbb{E}[\ln p^{\theta}(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-}a.s.$$

PROOF. *Proof of* (i). It follows from (52) that, $\mathbb{P}$-a.s.,

$$p^{\theta}(Z_0|Z_{-\infty}^{-1}) = \lim_{m\to\infty} p_{\chi}^{\theta}(Z_0|Z_{-m}^{-1}) \le \big|\mathbf{L}^{\theta}\langle Z_0\rangle(\cdot, \mathsf{X})\big|_{\infty} \le \prod_{i=0}^{r-1} \big|g^{\theta}(\cdot, Y_i)\big|_{\infty}. \tag{54}$$

Then, (A2) shows that

$$\mathbb{E}\big[\ln^{+} p^{\theta}(Z_0|Z_{-\infty}^{-1})\big] \le \mathbb{E}\big[\ln^{+}\big|\mathbf{L}^{\theta}\langle Z_0\rangle(\cdot, \mathsf{X})\big|_{\infty}\big] < \infty.$$

We now show that $\mathbb{E}[\ln^{-} p^{\theta}(Z_0|Z_{-\infty}^{-1})] < \infty$ by establishing that $\mathbb{E}[\ln p^{\theta}(Z_0| Z_{-\infty}^{-1})] > -\infty$. For that purpose, introduce the sequence

$$L_m^{\theta} \triangleq m^{-1} \sum_{\ell=1}^{m} \big[\ln^{+}\big|\mathbf{L}^{\theta}\langle Z_0\rangle(\cdot, \mathsf{X})\big|_{\infty} - \ln p_{\chi}^{\theta}(Z_0|Z_{-\ell}^{-1})\big].$$

By (54), the sequence $(L_m^{\theta})_{m\ge 0}$ is nonnegative and the Fatou lemma implies that

$$\liminf_{m\to\infty} \mathbb{E}\big[L_m^{\theta}\big] \ge \mathbb{E}\Big[\liminf_{m\to\infty} L_m^{\theta}\Big]. \tag{55}$$

By definition,

$$\begin{aligned}\liminf_{m\to\infty} \mathbb{E}\big[L_m^{\theta}\big] &= \mathbb{E}\big[\ln^{+}\big|\mathbf{L}^{\theta}\langle Z_0\rangle(\cdot, \mathsf{X})\big|_{\infty}\big] \\ &\quad - \limsup_{m\to\infty} m^{-1} \sum_{\ell=1}^{m} \mathbb{E}\big[\ln p_{\chi}^{\theta}(Z_0|Z_{-\ell}^{-1})\big]\end{aligned} \tag{56}$$

and

$$\begin{aligned}\mathbb{E}\Big[\liminf_{m\to\infty} L_m^{\theta}\Big] &= \mathbb{E}\big[\ln^{+}\big|\mathbf{L}^{\theta}\langle Z_0\rangle(\cdot, \mathsf{X})\big|_{\infty}\big] \\ &\quad - \mathbb{E}\Bigg[\limsup_{m\to\infty} m^{-1} \sum_{\ell=1}^{m} \ln p_{\chi}^{\theta}(Z_0|Z_{-\ell}^{-1})\Bigg].\end{aligned} \tag{57}$$

Since $(Y_k)_{k\in\mathbb{Z}}$ is stationary, for any $\ell \in \mathbb{N}$, $\mathbb{E}[\ln p_\chi^\theta(Z_0|Z_{-\ell}^{-1})] = \mathbb{E}[\ln p_\chi^\theta(Z_\ell|Z_0^{\ell-1})]$ showing that

$$m^{-1}\sum_{\ell=1}^m \mathbb{E}[\ln p_\chi^\theta(Z_0|Z_{-\ell}^{-1})] = m^{-1}\sum_{\ell=1}^m \mathbb{E}[\ln p_\chi^\theta(Z_\ell|Z_0^{\ell-1})]. \tag{58}$$

The Cesaro mean convergence lemma implies that, $\mathbb{P}$-a.s.,

$$\limsup_{m\to\infty} m^{-1}\sum_{\ell=1}^m \ln p_\chi^\theta(Z_0|Z_{-\ell}^{-1}) = \lim_{\ell\to\infty} \ln p_\chi^\theta(Z_0|Z_{-\ell}^{-1}) = \ln p^\theta(Z_0|Z_{-\infty}^{-1}). \tag{59}$$

Combining (55), (56), (57), (58) and (59) yields to

$$\begin{aligned} &\mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] \\ &\quad \ge \limsup_{m\to\infty} m^{-1}\sum_{\ell=1}^m \mathbb{E}[\ln p_\chi^\theta(Z_\ell|Z_0^{\ell-1})] \\ &\quad = \limsup_{m\to\infty}\{\mathbb{E}[m^{-1}\ln p_\chi^\theta(Z_0^m)] - m^{-1}\mathbb{E}[\ln p_\chi^\theta(Z_0)]\} > -\infty, \end{aligned} \tag{60}$$

where the last bound follows from (A1)(iii) and the minorization

$$\ln p_\chi^\theta(Z_0^m) \ge \ln \chi(\mathsf{D}) + \sum_{i=0}^m \ln \inf_{x\in\mathsf{D}} \mathbf{L}^\theta\langle Z_i\rangle(x,\mathsf{D}).$$

The proof of (i) follows.

Proof of (ii). According to Proposition 8 (36), there exists a random variable C satisfying $\mathbb{P}_Y[C<\infty]=1$ such that for all $k \ge K$ and $m \ge 0$,

$$\sup_{\theta\in\Theta}|\ln p_\chi^\theta(Z_k|Z_{-m}^{k-1}) - \ln p_\chi^\theta(Z_k|Z_{-m-1}^{k-1})| \le C\kappa^{k+m}, \qquad \mathbb{P}\text{-a.s.},$$

which implies that

$$\sup_{\theta\in\Theta}|\ln p_\chi^\theta(Z_k|Z_0^{k-1}) - \ln p^\theta(Z_k|Z_{-\infty}^{k-1})| \le C\kappa^k/(1-\kappa), \qquad \mathbb{P}\text{-a.s.}$$

The proof of (ii) follows from the obvious decomposition

$$\begin{aligned} n^{-1}\ln p_\chi^\theta(Z_0^{n-1}) &= n^{-1}\sum_{k=1}^{n-1}\ln p_\chi^\theta(Z_k|Z_0^{k-1}) + n^{-1}\ln p_\chi^\theta(Z_0), \\ n^{-1}\ln p^\theta(Z_0^{n-1}|Z_{-\infty}^{-1}) &= n^{-1}\sum_{k=0}^{n-1}\ln p^\theta(Z_k|Z_{-\infty}^{k-1}). \end{aligned} \tag{61}$$

The proof of (iii) follows from (53) and (61) using the Birkhoff theorem; see, for example, [28], Theorem 1.14. $\square$

PROPOSITION 11. *Assume (A1)–(A3). Let χ be a probability measure such that $\chi \in \mathcal{M}(\mathsf{D}, r)$ [where $\mathcal{M}(\mathsf{D}, r)$ is defined in (8)].*

(i) *For any $\theta_0 \in \Theta$ and any $\rho > 0$,*

$$\limsup_{n\to\infty} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \le \mathbb{E}\Big[\sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1})\Big], \qquad \mathbb{P}\text{-}a.s.$$

(ii) *The function $\theta \mapsto \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})]$ is upper semi-continuous.*

(iii) *For any compact set $\Xi \subset \Theta$, the sequence $(\sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}))_{n\ge 0}$ converges $\mathbb{P}$-a.s. and*

$$\lim_{n\to\infty} \sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) = \sup_{\theta\in\Xi} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-}a.s.$$

PROOF. *Proof of* (i). Proposition 10(ii) shows that

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \\ &\qquad \le \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_i|Z_{-\infty}^{i-1}), \qquad \mathbb{P}\text{-a.s.} \end{aligned} \tag{62}$$

By (54), for any $\theta_0 \in \Theta$ and $\rho > 0$,

$$\begin{aligned} \ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1}) &\le \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1}) \\ &\le \sum_{i=0}^{r-1} \sup_{\theta\in\Theta} \ln^+ |g(\cdot, Y_i)|_\infty, \qquad \mathbb{P}\text{-a.s.}, \end{aligned} \tag{63}$$

which shows using (53) and (A2) that

$$\mathbb{E}\Big[\Big|\sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1})\Big|\Big] < \infty.$$

The Birkhoff theorem therefore implies

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_i|Z_{-\infty}^{i-1}) \\ &\qquad = \mathbb{E}\Big[\sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1})\Big], \qquad \mathbb{P}\text{-a.s.}, \end{aligned} \tag{64}$$

which completes the proof of (i).

Proof of (ii). First note that

$$\sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] \le \mathbb{E}\Big[\sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1})\Big]. \tag{65}$$

Now, since under (A3), for any $m \geq p$, $\mathbb{P}$-a.s., the function $\theta \mapsto \ln p_\chi^\theta(Z_0|Z_{-m}^{-1})$ is continuous, then $\mathbb{P}$-a.s., the function $\theta \mapsto \ln p^\theta(Z_0|Z_{-\infty}^{-1})$ is continuous as a uniform limit of continuous functions. Using (63),

$$\sum_{i=0}^{r-1} \sup_{\theta\in\Theta} \ln^+ |g(\cdot, Y_i)|_\infty - \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1}) \geq 0,$$

the monotone convergence theorem therefore implies that

$$\begin{aligned} \lim_{\rho\downarrow 0} \mathbb{E}\Big[\sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1})\Big] &= \mathbb{E}\Big[\lim_{\rho\downarrow 0} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \ln p^\theta(Z_0|Z_{-\infty}^{-1})\Big] \\ &= \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})]. \end{aligned} \tag{66}$$

Combining (65) and (66) shows that

$$\lim_{\rho\downarrow 0} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] \leq \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})].$$

Proof of (iii). By taking the limit of both sides of (i) with respect to $\rho \downarrow 0$, (66) shows that for any $\theta_0 \in \Theta$,

$$\lim_{\rho\downarrow 0} \limsup_{n\to\infty} \sup_{\theta\in\mathcal{B}(\theta_0,\rho)} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \leq \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-a.s.} \tag{67}$$

Therefore, for any $\delta > 0$ and $\theta_0 \in \Xi$, there exists $\rho_{\theta_0} > 0$ such that

$$\limsup_{n\to\infty} \sup_{\theta\in\mathcal{B}(\theta_0,\rho_{\theta_0})} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \leq \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})] + \delta, \qquad \mathbb{P}\text{-a.s.}$$

Since Ξ is compact, by extracting a finite covering, the latter inequality shows that

$$\limsup_{n\to\infty} \sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \leq \sup_{\theta_0\in\Xi} \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})] + \delta, \qquad \mathbb{P}\text{-a.s.}$$

Since δ is arbitrary, we therefore have

$$\limsup_{n\to\infty} \sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \leq \sup_{\theta_0\in\Xi} \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})]. \tag{68}$$

Now, since for any $\theta_0 \in \Xi$,

$$\sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \geq \frac{1}{n} \ln p_\chi^{\theta_0}(Z_0^{n-1}).$$

Proposition 10(iii) yields

$$\liminf_{n\to\infty} \sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \geq \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-a.s.}$$

θ_0 being arbitrary in Ξ, we finally obtain

$$\liminf_{n\to\infty} \sup_{\theta\in\Xi} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) \geq \sup_{\theta_0\in\Xi} \mathbb{E}[\ln p^{\theta_0}(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-a.s.}$$

Combining this inequality with (68) completes the proof. $\square$

THEOREM 12. *Assume (A1)–(A3). Then, for any probability measure* $\chi \in \mathcal{M}(\mathsf{D}, r)$,

$$\lim_{n\to\infty} d(\hat{\theta}_{\chi,nr}, \Theta_b^\star) = 0, \qquad \mathbb{P}\text{-}a.s.,$$

where $\Theta_b^\star \subset \Theta$ *is defined by* $\Theta_b^\star \triangleq \arg\max_{\theta\in\Theta} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})]$.

PROOF. By Proposition 11(ii) the function $\theta \mapsto \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})]$ is upper semi-continuous. Therefore the set $\Theta_b^\star$ is compact as a closed subset of a the compact set Θ so that for any $\delta > 0$, $\Xi_\delta = \{\theta \in \Theta; d(\theta, \Theta_b^\star) \geq \delta\}$ is also a compact set. In addition, as a upper semi-continuous function, $\theta \mapsto \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})]$ restricted to Ξ_δ attains its maximum which implies that

$$\sup_{\theta\in\Xi_\delta} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] = \max_{\theta\in\Xi_\delta} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] < \mathbb{E}[\ln p^{\theta^\star}(Z_0|Z_{-\infty}^{-1})],$$

where $\theta^\star$ is any point in $\Theta_b^\star$. Combining this with Proposition 10(iii) yields

$$\lim_{n\to\infty} \sup_{\theta\in\Xi_\delta} \frac{1}{n} \ln p_\chi^\theta(Z_0^{n-1}) < \mathbb{E}[\ln p^{\theta^\star}(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-a.s.}$$

Using that

$$\lim_{n\to\infty} \frac{1}{n} \ln p_\chi^{\theta^\star}(Z_0^{n-1}) = \mathbb{E}[\ln p^{\theta^\star}(Z_0|Z_{-\infty}^{-1})], \qquad \mathbb{P}\text{-a.s.}$$

we finally obtain that $\mathbb{P}$-a.s., $\hat{\theta}_{\chi,n} \in \Xi_\delta$ finitely many times. The proof is complete. $\square$

5.2. *Proofs of Proposition* 1 *and Theorem* 2. We have now all the tools for obtaining the consistency of the MLE as a byproduct of the results obtained for the block MLE. We first state and prove the forgetting of the initial distribution for the predictive filter.

LEMMA 13. *Assume (A1). Let* $0 < \gamma^- < \gamma^+ \leq 1$. *Then, for all* $\eta > 0$, *there exists* $\rho_\eta \in (0, 1)$ *such that, for all sequence* $(z_i)_{i\geq 0}$ *satisfying*

$$n^{-1} \sum_{i=0}^{n-1} \mathbb{1}_{\mathsf{K}}(z_i) \geq \max(1 - \gamma^-, (1 + \gamma^+)/2), \tag{69}$$

all $\beta \in (\gamma^-, \gamma^+)$, all measurable function f, all probability measures χ and χ' and all $\theta \in \Theta$,

$$\left|\frac{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle f}{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}} - \frac{\chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle f}{\chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}}\right| \\ \leq 2\left\{\rho^{\lfloor n(\beta-\gamma^-)\rfloor} + \frac{\eta^{\lfloor n(\gamma^+-\beta)\rfloor/2}}{\chi(\mathsf{D})\chi'(\mathsf{D})}\left[\prod_{i=0}^{n-1} D_{z_i}^2\right]\right\}|f|_\infty,$$

where D_z is defined in (46).

PROOF. By Proposition 5,

$$\left|\frac{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle f}{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}} - \frac{\chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle f}{\chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}}\right| \\ = \frac{|\Delta^\theta_{\chi,\chi'}\langle z_0^{n-1}\rangle(f, \mathbb{1}_\mathsf{X})|}{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X} \times \chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}} \\ \leq 2\rho^{\lfloor n(\beta-\gamma^-)\rfloor}|f|_\infty + 2\eta^{\lfloor n(\gamma^+-\beta)\rfloor/2}\frac{\prod_{i=0}^{n-1}|\mathbf{L}^\theta\langle z_i\rangle(\cdot,\mathsf{X})|_\infty^2}{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X} \times \chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}}|f|_\infty,$$

where we have used that

$$\frac{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle f}{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}} \vee \frac{\chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle f}{\chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}} \leq |f|_\infty.$$

The proof follows by noting that (44) implies that

$$\frac{\prod_{i=0}^{n-1}|\mathbf{L}^\theta\langle z_i\rangle(\cdot,\mathsf{X})|_\infty^2}{\chi \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X} \times \chi' \mathbf{L}^\theta \langle z_0^{n-1}\rangle \mathbb{1}_\mathsf{X}} \leq \frac{[\prod_{i=0}^{n-1} D_{z_i}^2]}{\chi(\mathsf{D})\chi'(\mathsf{D})}.$$ □

PROOF OF PROPOSITION 1. *Proof of* (i). Let χ a probability measure such that $\chi(\mathsf{D}) > 0$. The first step of the proof consists of using the forgetting property obtained in Lemma 13 to show that $\mathbb{P}$-a.s., the sequence $(p_\chi^\theta(Y_0|Y_{-\ell}^{-1}))_{\ell\geq 0}$ converges. Denote for any $t \in \{1,\dots,r\}$,

$$\chi_{m,t}^\theta(\mathsf{A}) = \frac{\chi \mathbf{L}^\theta \langle y_{-mr-t}^{-mr-1}\rangle \mathbb{1}_A}{\chi \mathbf{L}^\theta \langle y_{-mr-t}^{-mr-1}\rangle \mathbb{1}_\mathsf{X}}.$$

Then, write for any $m \geq 0$, $t \in \{1,\dots,r\}$ and any $y_{-mr-t}^0 \in \mathsf{Y}^{mr+t+1}$,

$$p_\chi^\theta(y_0|y_{-mr-t}^{-1}) = p_{\chi_{m,t}^\theta}^\theta(y_0|z_{-m}^{-1}) = \frac{\chi_{m,t}^\theta \mathbf{L}^\theta\langle z_{-m}^{-1}\rangle(g^\theta(\cdot, y_0))}{\chi_{m,t}^\theta \mathbf{L}^\theta\langle z_{-m}^{-1}\rangle(\mathbb{1}_\mathsf{X})}.$$

Let $0 < \gamma^- < \gamma^+ < 1$. Lemma 13 shows that for any $t \in \{1, \dots, r\}$ and $\eta > 0$, there exists $\rho \in (0, 1)$ such that, if

$$m^{-1} \sum_{i=-m}^{-1} \mathbb{1}_{\mathsf{K}}(z_i) \geq \max\bigl(1 - \gamma^-, (1 + \gamma^+)/2\bigr),$$

then for all $\beta \in (\gamma^-, \gamma^+)$, and $\theta \in \Theta$,

$$\begin{aligned}
&\bigl|p_\chi^\theta\bigl(y_0 | y_{-mr-t}^{-1}\bigr) - p_\chi^\theta\bigl(y_0 | y_{-mr}^{-1}\bigr)\bigr| \\
&\quad \leq 2\left(\rho^{\lfloor m(\beta-\gamma^-)\rfloor} + \frac{\eta^{\lfloor m(\gamma^+-\beta)\rfloor/2}}{\chi_{m,t}^\theta(\mathsf{D})\chi(\mathsf{D})} \prod_{j=-m}^{-1} (D_{z_j})^2\right) \sup_{\theta\in\Theta} \bigl|g^\theta(\cdot, y_0)\bigr|_\infty \\
&\quad \leq 2\left(\rho^{\lfloor m(\beta-\gamma^-)\rfloor} + \eta^{\lfloor m(\gamma^+-\beta)\rfloor/2} D'_{-m} \prod_{j=-m}^{-1} (D_{z_j})^2\right) \sup_{\theta\in\Theta} \bigl|g^\theta(\cdot, y_0)\bigr|_\infty,
\end{aligned}$$

where

$$D'_{-m} = \max_{t=1,\dots,r-1} \frac{1}{\inf_{\theta\in\Theta} \chi_{m,t}^\theta(\mathsf{D})\chi(\mathsf{D})}.$$

$(D'_{-m})_{m\geq 0}$ is a stationary sequence. Using the same argument as in the proof of (47), the condition $\chi \in \mathcal{M}(\mathsf{D}, r)$ [defined in (8)], we have $\mathbb{E}[\ln^+ D'_{-m}] < \infty$. By choosing γ^+ and γ^- such that $\mathbb{P}_Y[Z_0 \in \mathsf{K}] > \max(1 - \gamma^-, (1 + \gamma^+)/2)$ and by applying Lemma 6, it follows that there exist $\varrho_\chi \in (0, 1)$ and a $\mathbb{P}$-a.s. finite random variable C_χ such that for any $\ell \geq 1$,

$$\bigl|p_\chi^\theta\bigl(Y_0 | Y_{-\ell}^{-1}\bigr) - p_\chi^\theta\bigl(Y_0 | Y_{-\ell-1}^{-1}\bigr)\bigr| \leq C_\chi \varrho_\chi^\ell, \qquad \mathbb{P}\text{-a.s.}$$

Similarly, for any probability measure χ' such that $\chi'(\mathsf{D}) > 0$, there exist $\varrho_{\chi,\chi'} \in (0, 1)$ and a $\mathbb{P}$-a.s. finite random variable $C_{\chi,\chi'}$ such that for any $\ell \geq 0$,

$$\bigl|p_\chi^\theta\bigl(Y_0 | Y_{-\ell}^{-1}\bigr) - p_{\chi'}^\theta\bigl(Y_0 | Y_{-\ell}^{-1}\bigr)\bigr| \leq C_{\chi,\chi'} \varrho_{\chi,\chi'}^\ell, \qquad \mathbb{P}\text{-a.s.}$$

This implies that for any probability measure χ satisfying $\chi(\mathsf{D}) > 0$, the sequence $(p_\chi^\theta(Y_0 | Y_{-\ell}^{-1}))_{\ell\geq 0}$ converges $\mathbb{P}$-a.s. and that the limit denoted by $p^\theta(Y_0 | Y_{-\infty}^{-1})$ does not depend on χ. Then, by stationarity of $(Y_\ell)_{\ell\in\mathbb{Z}}$, we obtain that for all $k \geq 0$ and $\theta \in \Theta$,

$$\lim_{m\to\infty} p_\chi^\theta\bigl(Y_k | Y_{-m}^{k-1}\bigr) = p^\theta\bigl(Y_k | Y_{-\infty}^{k-1}\bigr), \qquad \mathbb{P}\text{-a.s.},$$

which shows the first part of (i). To complete the proof of (i), it remains to prove that $\mathbb{E}[|\ln p^\theta(Y_k | Y_{-\infty}^{k-1})|] < \infty$. Since $p_\chi^\theta(Y_k | Y_{-m}^{k-1}) \leq \sup_{x\in\mathsf{X}} g^\theta(x, Y_k)$, we have

$$\ln^+ p_\chi^\theta\bigl(Y_k | Y_{-\infty}^{k-1}\bigr) \leq \ln^+ \sup_{x\in\mathsf{X}} g^\theta(x, Y_k),$$

which shows, under (A2), that

$$\mathbb{E}[\ln^+ p^\theta(Y_k|Y_{-\infty}^{k-1})] < \infty. \tag{70}$$

This allows us to define $\mathbb{E}[\ln p^\theta(Y_k|Y_{-\infty}^{k-1})]$ as

$$\mathbb{E}[\ln p^\theta(Y_k|Y_{-\infty}^{k-1})] = \mathbb{E}[\ln^+ p^\theta(Y_k|Y_{-\infty}^{k-1})] - \mathbb{E}[\ln^- p^\theta(Y_k|Y_{-\infty}^{k-1})],$$

so that $\mathbb{E}[\ln^- p^\theta(Y_k|Y_{-\infty}^{k-1})] < \infty$ provided that we have shown $\mathbb{E}[\ln p^\theta(Y_k|Y_{-\infty}^{k-1})] > -\infty$. By stationarity of $(Y_k)_{k\in\mathbb{Z}}$,

$$\begin{aligned}
r\mathbb{E}[\ln p^\theta(Y_0|Y_{-\infty}^{-1})] &= r\{\mathbb{E}[\ln^+ p^\theta(Y_0|Y_{-\infty}^{-1})] - \mathbb{E}[\ln^- p^\theta(Y_0|Y_{-\infty}^{-1})]\} \\
&= \mathbb{E}\left[\sum_{k=0}^{r-1} \ln^+ p^\theta(Y_k|Y_{-\infty}^{k-1})\right] - \mathbb{E}\left[\sum_{k=0}^{r-1} \ln^- p^\theta(Y_k|Y_{-\infty}^{k-1})\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{r-1} \ln p^\theta(Y_k|Y_{-\infty}^{k-1})\right],
\end{aligned} \tag{71}$$

where the last equality follows by applying $\mathbb{E}(A-B) = \mathbb{E}(A) - \mathbb{E}(B)$ for nonnegative random variables A, B such that $\mathbb{E}(A) < \infty$. Now, note that

$$\begin{aligned}
\prod_{k=0}^{r-1} p^\theta(Y_k|Y_{-\infty}^{k-1}) &= \prod_{k=0}^{r-1} \lim_{m\to\infty} p_\chi^\theta(Y_k|Y_{-mr}^{k-1}) = \lim_{m\to\infty} \prod_{k=0}^{r-1} p_\chi^\theta(Y_k|Y_{-mr}^{k-1}) \\
&= \lim_{m\to\infty} p_\chi^\theta(Y_0^{r-1}|Y_{-mr}^{-1}) = \lim_{m\to\infty} p_\chi^\theta(Z_0|Z_{-m}^{-1}) \\
&= p^\theta(Z_0|Z_{-\infty}^{-1}).
\end{aligned}$$

By plugging this expression into (71) and using $\mathbb{E}[|\ln p_\chi^\theta(Z_0|Z_{-\infty}^{-1})|] < \infty$ (see Proposition 10), we finally obtain

$$r\mathbb{E}[\ln p^\theta(Y_0|Y_{-\infty}^{-1})] = \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] > -\infty, \tag{72}$$

which completes the proof of (i).

Proof of (ii). Let χ be a probability measure such that $\chi(\mathsf{D}) > 0$ and let $t \in \{0, \dots, r-1\}$. Then, for any $m \geq 0$,

$$\begin{aligned}
m^{-1}\ln p_\chi^\theta(Z_0^{m+1}) &\leq m^{-1}\ln p_\chi^\theta(Y_0^{mr+t}) + m^{-1}\ln^+ A_{m,t} \\
&\leq m^{-1}\ln p_\chi^\theta(Z_0^m) + m^{-1}\ln^+ B_{m,t} + m^{-1}\ln^+ A_{m,t},
\end{aligned} \tag{73}$$

where

$$A_{m,t} \triangleq \sup_{\theta\in\Theta}\sup_x p^\theta_{Q^\theta(x,\cdot)}(Y_{mr+t+1}^{(m+1)r-1}), \qquad B_{m,t} \triangleq \sup_{\theta\in\Theta}\sup_x p^\theta_{\delta_x}(Y_{mr}^{mr+t}).$$

Note that $(A_{m,t})_{m\geq 0}$ and $(B_{m,t})_{m\geq 0}$ are stationary. Moreover, using (A2), it can be easily checked that

$$\mathbb{E}[\ln^+ A_{m,t}] < \infty, \qquad \mathbb{E}[\ln^+ B_{m,t}] < \infty.$$

Then, Lemma 7 may apply and for any $\beta \in (0,1)$, there exist $\mathbb{P}$-a.s. finite random variables A, B such that for all $m \geq 0$,

$$A_{m,t} \leq A\beta^{-m}, \qquad B_{m,t} \leq B\beta^{-m}, \qquad \mathbb{P}\text{-a.s.}$$

so that, $\mathbb{P}$-a.s.,

$$0 \leq \limsup_{m\to\infty} m^{-1} \ln^+ A_{m,t} \leq -\ln\beta,$$

$$0 \leq \limsup_{m\to\infty} m^{-1} \ln^+ B_{m,t} \leq -\ln\beta.$$

By letting $\beta \uparrow 1$,

$$\lim_{m\to\infty} m^{-1} \ln^+ A_{m,t} = 0, \qquad \lim_{m\to\infty} m^{-1} \ln^+ B_{m,t} = 0, \qquad \mathbb{P}\text{-a.s.} \tag{74}$$

Now, note that $(A_{m,t})_{m\geq 0}$ and $(B_{m,t})_{m\geq 0}$ do not depend on $\theta \in \Theta$ so that (74) together with (73) yields

$$\limsup_{m\to\infty} \sup_{\theta\in\Theta} m^{-1} |\ln p_\chi^\theta(Y_0^{mr+t}) - \ln p_\chi^\theta(Z_0^m)| = 0, \qquad \mathbb{P}\text{-a.s.} \tag{75}$$

Since t is chosen arbitrarily in $\{0,\dots,r-1\}$, we finally obtain using Proposition 10(ii),

$$\begin{aligned}\lim_{n\to\infty} n^{-1} \ln p_\chi^\theta(Y_0^n) &= r^{-1} \lim_{m\to\infty} m^{-1} \ln p_\chi^\theta(Z_0^m) \\ &= r^{-1}\mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] \\ &= \mathbb{E}[\ln p^\theta(Y_0|Y_{-\infty}^{-1})], \qquad \mathbb{P}\text{-a.s.,}\end{aligned}$$

which completes the proof of Proposition 1. □

PROOF OF THEOREM 2. By Proposition 11(ii) and (72), the function $\theta \mapsto \ell(\theta)$ is upper semi-continuous. Moreover, (72) also implies

$$\Theta^\star = \arg\max_{\theta\in\Theta} \mathbb{E}[\ln p^\theta(Y_0|Y_{-\infty}^{-1})] = \arg\max_{\theta\in\Theta} \mathbb{E}[\ln p^\theta(Z_0|Z_{-\infty}^{-1})] = \Theta_b^\star.$$

Now let t in $\{0,\dots,r-1\}$ and recall that $Z_0^m = Y_0^{mr-1}$. Theorem 12 together with (75) shows that

$$\lim_{n\to\infty} d(\hat{\theta}_{\chi,nr+t}, \Theta^\star) = 0, \qquad \mathbb{P}\text{-a.s.} \tag{76}$$

The proof of Theorem 2 is then complete since t is arbitrary in $\{0,\dots,r-1\}$. □

PROOF OF PROPOSITION 3. Under these two conditions, for any $u \in \{1, \dots, r\}$, and $\theta \in \Theta$,

$$\begin{aligned}\chi \mathbf{L}^\theta \langle y_0^{u-1} \rangle \mathbb{1}_{\mathsf{D}} &\geq \left(\prod_{i=0}^{u-1} \inf_{x_i \in \mathsf{D}_i} g^\theta(x_i, y_i) \right) \int \cdots \int \chi(\mathrm{d}x_0) \mathbb{1}_{\mathsf{D}}(x_u) \prod_{i=1}^{u} \mathbb{1}_{\mathsf{D}_{i-1}}(x_{i-1}) Q^\theta(x_{i-1}, \mathrm{d}x_i) \\ &\geq \left(\prod_{i=0}^{u-1} \inf_{x_i \in \mathsf{D}_i} g^\theta(x_i, y_i) \right) \chi(\mathsf{D}_0) \delta^u.\end{aligned}$$

□

PROOF OF LEMMA 4. The proof proceeds by induction on $u \in \{1, \dots, r\}$. Assume that D_{u-1} is a compact subset; we show that there exists a compact set D_u such that $\inf_{x_{u-1} \in \mathsf{D}_{u-1}} \inf_{\theta \in \Theta} Q^\theta(x_{u-1}, \mathsf{D}_u) \geq \delta$.

Let $(x, \theta) \in \mathsf{D}_{u-1} \times \Theta$ and set $\delta < \delta' < 1$. Since $\mathsf{X} = \mathbb{R}^d$ is a complete separable metric space and $\mathcal{X}$ is the associated Borel σ-field, there exists a sequence $B_1^{x,\theta}, B_2^{x,\theta}, \dots,$ of open balls of radius 1 covering X. Choose $N_{x,\theta}$ large enough so that $Q^\theta(x, \mathsf{O}_{x,\theta}) \geq \delta'$, where $\mathsf{O}_{x,\theta} = \bigcup_{i \leq N_{x,\theta}} B_i^{x,\theta}$. Since for any open set O the function $(x', \theta') \mapsto Q^{\theta'}(x', \mathsf{O})$ is lower semi-continuous, there exists a neighborhood $\mathcal{V}_{x,\theta}$ (for the product topology on $\mathsf{X} \times \Theta$), such that for all $(x', \theta') \in \mathcal{V}_{x,\theta}$, $Q^{\theta'}(x', \mathsf{O}_{x,\theta}) \geq \delta$. Since $\mathsf{O}_{x,\theta}$ is totally bounded its closure, denoted $\mathsf{K}_{x,\theta}$, is a compact subset, which satisfies, for any $(x', \theta') \in \mathcal{V}_{x,\theta}$ that $Q^\theta(x, \mathsf{K}_{x,\theta}) \geq \delta$.

Then, $\bigcup_{(x,\theta) \in \mathsf{D}_{u-1} \times \Theta} \mathcal{V}_{x,\theta}$ is a covering of $\mathsf{D}_{u-1} \times \Theta$. Since the set $\mathsf{D}_{u-1} \times \Theta$ is compact, we may extract a finite subcover $\mathsf{D}_{u-1} \times \Theta \subseteq \bigcup_{i=1}^{I} \mathcal{V}_{x_i,\theta_i}$. Take $\mathsf{D}_u = \bigcup_{i=1}^{I} \mathsf{K}_{x_i,\theta_i}$. As a finite union of compact sets, D_u is a compact set, which satisfies, for all $(x, \theta) \in \mathsf{D}_{u-1} \times \Theta$, $Q^\theta(x, \mathsf{D}_u) \geq \delta$. This completes the proof. □

REFERENCES

[1] BARRON, A. R. (1985). The strong ergodic theorem for densities: Generalized Shannon–McMillan–Breiman theorem. *Ann. Probab.* **13** 1292–1303. MR0806226

[2] BAUM, L. E. and PETRIE, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **37** 1554–1563. MR0202264

[3] BUDHIRAJA, A. and OCONE, D. (1997). Exponential stability of discrete-time filters for bounded observation noise. *Systems Control Lett.* **30** 185–193. MR1455877

[4] CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

[5] CHURCHILL, G. (1992). Hidden Markov chains and the analysis of genome structure. *Computers and Chemistry* **16** 107–115.

[6] DOUC, R., FORT, G., MOULINES, E. and PRIOURET, P. (2009). Forgetting the initial distribution for hidden Markov models. *Stochastic Process. Appl.* **119** 1235–1256. MR2508572

[7] DOUC, R. and MATIAS, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7** 381–420. MR1836737

[8] DOUC, R., MOULINES, E., OLSSON, J. and VAN HANDEL, R. (2011). Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* **39** 474–513. MR2797854

[9] DOUC, R., MOULINES, É. and RYDÉN, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.* **32** 2254–2304. MR2102510

[10] FOMBY, T. B. and HILL, R. C., eds. (2003). *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later. Advances in Econometrics* **17**. Elsevier, Amsterdam. MR2531667

[11] FREDKIN, D. R. and RICE, J. A. (1987). Correlation functions of a function of a finite-state Markov process with application to channel kinetics. *Math. Biosci.* **87** 161–172. MR0929996

[12] FUH, C.-D. (2006). Efficient likelihood estimation in state space models. *Ann. Statist.* **34** 2026–2068. MR2283726

[13] FUH, C.-D. (2010). Reply to "On some problems in the article Efficient likelihood estimation in state space models" by Cheng-Der Fuh [Ann. Statist. **34** (2006) 2026–2068] [MR2604693]. *Ann. Statist.* **38** 1282–1285. MR2604694

[14] GENON-CATALOT, V. and LAREDO, C. (2006). Leroux's method for general hidden Markov models. *Stochastic Process. Appl.* **116** 222–243. MR2197975

[15] HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

[16] HULL, J. and WHITE, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42** 281–300.

[17] JENSEN, J. L. (2010). On some problems in the article Efficient likelihood estimation in state space models by Cheng-Der Fuh [Ann. Statist. **34** (2006) 2026–2068] [MR2283726]. *Ann. Statist.* **38** 1279–1281. MR2604693

[18] JUANG, B. H. and RABINER, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics* **33** 251–272. MR1132665

[19] KLEPTSYNA, M. L. and VERETENNIKOV, A. Y. (2008). On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields* **141** 411–444. MR2391160

[20] LE GLAND, F. and MEVEL, L. (2000). Basic properties of the projective product with application to products of column-allowable nonnegative matrices. *Math. Control Signals Systems* **13** 41–62. MR1742139

[21] LE GLAND, F. and MEVEL, L. (2000). Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems* **13** 63–93. MR1742140

[22] LEROUX, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* **40** 127–143. MR1145463

[23] MAMON, R. S. and ELLIOTT, R. J., eds. (2007). *Hidden Markov Models in Finance. International Series in Operations Research & Management Science* **104**. Springer, New York. MR2407726

[24] MEVEL, L. and FINESSO, L. (2004). Asymptotical statistics of misspecified hidden Markov models. *IEEE Trans. Automat. Control* **49** 1123–1132. MR2071939

[25] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

[26] PETRIE, T. (1969). Probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **40** 97–115. MR0239662

[27] VAN HANDEL, R. (2008). Discrete time nonlinear filters with informative observations are stable. *Electron. Commun. Probab.* **13** 562–575. MR2461532

[28] WALTERS, P. (1982). *An Introduction to Ergodic Theory. Graduate Texts in Mathematics* **79**. Springer, New York. MR0648108

[29] WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163

SAMOVAR
CNRS UMR 5157
INSTITUT TÉLÉCOM/TÉLÉCOM SUDPARIS
9 RUE CHARLES FOURIER
91000 EVRY
FRANCE
E-MAIL: randal.douc@it-sudparis.eu

LTCI
CNRS UMR 5141
INSTITUT TÉLÉCOM /TÉLÉCOM PARISTECH
46 RUE BARRAULT
75634 PARIS CEDEX 13
FRANCE
E-MAIL: eric.moulines@telecom-paristech.fr