Econometric Theory, **23**, 2007, 749–759. Printed in the United States of America.
DOI: 10.1017/S0266466607070314

BAYESIAN CONSISTENCY FOR STATIONARY MODELS

Antonio Lijoi
University of Pavia
and
CNR-IMATI Milan

Igor Prünster
University of Turin
College Carlo Alberto, and ICER

Stephen G. Walker
University of Kent

In this paper, we provide a Doob-style consistency theorem for stationary models. Many applications involving Bayesian inference deal with non independent and identically distributed data, in particular, with stationary data. However, for such models, there is still a theoretical gap to be filled regarding the asymptotic properties of Bayesian procedures. The primary goal to be achieved is establishing consistency of the sequence of posterior distributions. Here we provide an answer to the problem. Bayesian methods have recently gained growing popularity in economic modeling, thus implying the timeliness of the present paper. Indeed, we secure Bayesian procedures against possible inconsistencies. No results of such a generality are known up to now.

1. INTRODUCTION

Most commonly employed econometric models rely on the assumption of stationarity of the sequence of observations. Statistical inference involving this dependence structure should feature good asymptotic properties, and, among these, consistency plays a major role. From a classical point of view, there is a wealth of literature on the topic providing consistency results in specific situations. From a Bayesian point of view, no Doob-type results are available beyond the exchangeable case. However, because Bayesian methods are now routinely applied to problems in which the dependence structure for the observations is

The authors are grateful for the comments and suggestions of two referees. Antonio Lijoi and Igor Prünster were supported by the Italian Ministry of University and Research, grants 2006134525 and 2006133449, respectively. The research of Stephen G. Walker was funded by an EPSRC Advanced Research Fellowship. Address correspondence to Stephen G. Walker, Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Kent CT2 7NZ, UK; e-mail: S.G.Walker@kent.ac.uk.

 0266-4666/07 $15.00

more complicated than exchangeability, there is a need for the extension of Doob's theorem to these cases.

Here we will deal with stationary sequences of observations collected in time whose probability distribution is representable as a mixture of conditionally stationary and ergodic laws. In other words we suppose that, conditionally on some random element, the observations are stationary, with sample means obeying the ergodic theorem. The representation theorem we will need is given in Maitra (1977) and further analyzed in Aldous (1985).

Consider a stationary sequence of random variables $(X_n)_{n\geq 1}$ defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in a Polish space $\mathbb{X}$, endowed with the Borel σ-algebra $\mathcal{X}$. Conditionally on a random element, say, θ, the sequence is stationary and ergodic and its law is characterized by the family $(P_n(\cdot;\theta))_{n\geq 1}$ of one-step transition distributions, that is,

$$P_n(A;\theta) = \mathbb{P}[X_n \in A | X_1, \ldots, X_{n-1}; \theta] \quad \forall A \in \mathcal{X},$$

with $P_1(A;\theta) = \mathbb{P}[X_1 \in A|\theta]$. The random parameter θ taking values in the parameter space can be either finite- or infinite-dimensional, and one is typically interested in estimating θ or a functional of θ, given an n-sample of observation $(X_1, \ldots, X_n)$. Such estimates are known as a posterior estimates. Here we will deal with asymptotic properties of the sequence of predictive distributions $(P_n)_{n\geq 1}$. The models we will consider are characterized by the fact that the probability distribution of the vector $(X_1, \ldots, X_n)$ admits a mixture representation of the type

$$\begin{aligned}\mathbb{P}[(X_1, \ldots, X_n) \in A] &= \int_\Theta \mathbb{P}[(X_1, \ldots, X_n) \in A|\theta]\Pi(\mathrm{d}\theta) \\ &= \int_\Theta \left\{ \int_A \prod_{i=1}^n p(\mathrm{d}x_i | x_1, \ldots, x_{i-1}; \theta) \right\} \Pi(\mathrm{d}\theta)\end{aligned}$$

for any set A in $\mathcal{X}^n$ and for some probability distribution Π on the parameter space Θ, with $p(\mathrm{d}x_1|x_0;\theta) = p(\mathrm{d}x_1|\theta)$. Using the terminology commonly employed in Bayesian statistics, Π acts as a prior distribution for θ. Moreover, for any θ, $\prod_{i=1}^n p(\mathrm{d}x_i|x_1, \ldots, x_{i-1};\theta)$ is a probability distribution on $(\mathbb{X}^n, \mathcal{X}^n)$ that can be seen as the distribution of $(X_1, \ldots, X_n)$ given θ. If we let $P^\infty(\cdot|\theta)$ denote the distribution of the whole sequence $(X_n)_{n\geq 1}$ this means that, for any $n \geq 1$,

$$P^\infty(A \times \mathbb{X}^\infty|\theta) = \int_A \prod_{i=1}^n p(\mathrm{d}x_i|x_1, \ldots, x_{i-1};\theta) \quad \forall A \in \mathcal{X}^n. \tag{1}$$

Suppose that each $p(\mathrm{d}x|x_1, \ldots, x_{i-1};\theta)$ admits a density with respect to some σ-finite measure λ on $\mathbb{X}$, that is, $p(\mathrm{d}x|x_1, \ldots, x_{i-1};\theta) = f(x|x_1, \ldots, x_{i-1};\theta)$ $\lambda(\mathrm{d}x)$. With respect to these transition densities, we ask for an *identifiability* condition to hold true. This implies that $f(\cdot\ |x_1, \ldots, x_{i-1};\theta) \neq f(\cdot\ |x_1, \ldots, x_{i-1};\theta')$

for any $\theta \neq \theta'$, $x_1, \ldots, x_{i-1} \in \mathbb{X}$ and $i \geq 1$. This is a fundamental requirement for all statistical models. Hence it should be always satisfied, and it will be valid in all the examples we are going to list in the following sections.

As one can see, a number of processes of common use in applications may play the role of the mixing model in the representation theorem for the sequence of observations. Here we highlight some interesting and useful examples.

1.1. Exchangeable Sequences

A noteworthy special example is represented by a sequence of exchangeable observations. In this case, the de Finetti representation theorem states that the distribution of a vector of n of those random variables can be represented as a mixture of independent and identically distributed (i.i.d.) random variables, that is,

$$\mathbb{P}[(X_1, \ldots, X_n) \in A] = \int_A \int_\Theta \left\{ \prod_{i=1}^n p(\mathrm{d}x_i | \theta) \right\} \Pi(\mathrm{d}\theta).$$

Hence

$$p(\mathrm{d}x_i | x_1, \ldots, x_{i-1}; \theta) = p(\mathrm{d}x_i | \theta).$$

This describes the common setting for Bayesian applications, which parallels the classical assumption of i.i.d. observations. The first Bayesian study of consistency for this class of models dates back to the paper by Doob (1949). More recent work has been done by Lijoi, Prünster, and Walker (2004) that inspires the proof to the much more general theorem in the present paper.

1.2. Discretely Observed Diffusion Processes

Let the stochastic process $X = \{X_t : t \geq 0\}$ be a diffusion process that is the solution of the following stochastic differential equation:

$$\mathrm{d}X_t = b(X_t; \theta)\mathrm{d}t + \sigma(X_t; \theta)\mathrm{d}B_t, \tag{2}$$

where $\{B_t : t \geq 0\}$ is a standard Brownian motion and b and σ satisfy suitable conditions ensuring existence and uniqueness of the solution to (2). More specifically, introduce the scale density $s(x;\theta) = \exp\{-2\int_{x^*}^{x} [b(y;\theta)/\sigma^2(y;\theta)]\mathrm{d}y\}$, for some x^* in the interior of the domain, I, of X_t. We suppose $I = [l, r]$, where $-\infty \leq l < r \leq +\infty$. If the following conditions are satisfied:

$$\int_l^{x^*} s(x;\theta)\mathrm{d}x = -\infty, \qquad \int_{x^*}^{r} s(x;\theta)\mathrm{d}x = +\infty$$

$$\int_l^r \frac{1}{s(x;\theta)\sigma^2(x;\theta)} \mathrm{d}x < +\infty,$$

then X is strongly stationary (see, e.g., Karatzas and Shreve, 1991, Sect. 5.5) and ergodic.

Take, now, an increasing sequence of positive numbers $(t_n)_{n\geq 1}$ with $t_n \to +\infty$ as $n \to +\infty$ and set $Y_n = X_{t_n}$. The interest of a discretely observed diffusion process is crucial to a number of problems in economics and finance. Indeed, models of the type (2) are routinely used to describe the behavior of stock prices, exchange rates, and interest rates, among other interesting economic phenomena. In all these applications, data are recorded at discrete time points (e.g., monthly, weekly, daily), and then it is important to study properties of estimation procedures that involve the sequence $(Y_n)_{n\geq 1}$. Within the frequentist setting, an area of research has focused on methods of estimation based on the use of contrast functions or approximate likelihood functions, and remarkable papers are those by Dacunha-Castelle and Florens-Zmirou (1986), Bibby and Sørensen (1995), Gourieroux, Monfort, and Renault (1993), Pedersen (1995), and Aït-Sahalia (2002). Some of these papers also consider the consistency of the suggested estimation methods. Techniques for Bayesian inference in this context have been developed in Elerian, Chib, and Shephard (2001), Eraker (2001), Beskos, Papaspiliopoulos, Roberts, and Fernhead (2006), and Roberts and Stramer (2001).

1.3. Markov Processes

Another noteworthy example that falls within the class of models we deal with is represented by stationary Markov processes. Hence, we let $(X_n)_{n\geq 1}$ be a stationary Markov process taking values in a general state space $\mathbb{X}$, and we denote its one-step transition density by $p(\cdot,\cdot)$. If the process is aperiodic and irreducible, it is also ergodic. Thus our result includes such a model. Note that the parameter θ might in this case coincide with the transition density itself, thus being infinite-dimensional. Classical approaches typically involve maximum likelihood procedures. See, for example, Billingsley (1961). As far as the Bayesian approach is concerned, interest in inference for stationary and ergodic Markov chains has grown in a variety of settings and applied problems.

Here we consider a Markovian state space model based on a time series $\{Y_t : t = 1, \ldots, n\}$ whose observations are conditionally independent given an unobserved sufficient state $\{X_t : t = 1, \ldots, n\}$, assumed to be Markovian. The aim is to learn about the state X_t, given contemporaneously available information. See, for example, Pitt and Shephard (1999). Moreover, and still in a Bayesian setting, popular Markov chain Monte Carlo (MCMC) techniques are based on the simulation of stationary and ergodic Markov chains. See, for example, Chib (1996) and Chib and Greenberg (1996).

More recently an interesting approach for constructing strictly stationary AR(1) models has been introduced by Mena and Walker (2005). The idea is to use one-step Bayesian nonparametric predictive distributions to define the transition density of the process.

1.4. ARCH(1) Models

Suppose $(X_n)_{n\geq 1}$ is a sequence of random variables defined by

$$X_n = \sigma_n^2 Z_n, \tag{3}$$

where $\sigma_n^2 = \theta_1 + \theta_2 X_{n-1}^2$ and the Z_n's are i.i.d. with a zero mean and unit variance. If $\theta_2 < \exp\{-\psi(0.5)\}$, where ψ is the so-called ψ-function, then the process $(X_n)_{n\geq 1}$ is stationary ergodic. The parameter is, in this case, the vector $\theta = (\theta_1, \theta_2)$. Both a classical likelihood-based approach and Bayesian techniques are considered, for instance, in Fiorentini, Sentana, and Shephard (2004).

1.5. Linear Processes

Let $(a_n(\theta))_{n\geq 1}$ be a sequence of real numbers that is squared-summable, that is,

$$\sum_{n\geq 1} |a_n(\theta)|^2 < +\infty \qquad \forall \theta \in \Theta,$$

where Θ is possibly an infinite-dimensional space. Denote by $\{\epsilon_n(\theta) : n = -\infty, \ldots, +\infty\}$ a stochastic process such that

$$E(\epsilon_n(\theta)) = 0 \qquad \mathrm{Cov}(\epsilon_i(\theta), \epsilon_j(\theta)) = \sigma^2 \delta_{i,j},$$

where $\delta_{i,j} = 1$ if $i = j$ and it is zero otherwise. Then the process $(X_n)_{n\geq 1}$ defined by

$$X_n = \sum_{j=-\infty}^{+\infty} a_j(\theta)\epsilon_{n-j}(\theta)$$

is stationary and ergodic. Classical inferential procedures for these processes are accounted for in Brockwell and Davis (1996). A description of Bayesian solutions to problem estimation can be found in West and Harrison (1997) and Huerta and West (1999a, 1999b).

2. CONSISTENCY FOR STATIONARY MODELS

Here we wish to prove an asymptotic result that is known to hold in the exchangeable case. Let us first set the parameter space Θ to be complete and separable with respect to some suitable metric and indicate with $\sigma(\Theta)$ the corresponding Borel σ-algebra. According to the assumptions outlined in Section 1, for any $n \geq 1$ and $\theta \in \Theta$, the probability distribution $\mathbb{P}[(X_1, \ldots, X_n) \in \cdot|\theta]$ admits a density with respect to the product measure λ^n on $\mathbb{X}^n$, and we denote such a density by $h_n(\cdot|\theta)$, that is,

$$\mathbb{P}[(X_1, \ldots, X_n) \in A|\theta] = \int_A h_n(x_1, \ldots, x_n|\theta)\lambda^n(\mathrm{d}x_1, \ldots, \mathrm{d}x_n),$$

and $h_n(x_1,\dots,x_n|\theta) = \prod_{i=1}^n f(x_i|x_1,\dots,x_{i-1};\theta)$. To evaluate distance between any two densities on $\mathbb{X}$, we consider the Hellinger distance, which is well known to be equivalent to the L_1-metric, defined as follows:

$$d_H(f,g) = \left\{\int_{\mathbb{X}} \left[\sqrt{f(x)} - \sqrt{g(x)}\right]^2 \lambda(\mathrm{d}x)\right\}^{1/2}.$$

We simply denote by P_θ the law of the whole sequence of observations when the value of the parameter is θ. Moreover, if Π_n denotes the conditional distribution of θ, given the sample $x_1,\dots,x_n$, it can be represented as

$$\Pi_n(\mathrm{d}\theta) = \frac{\prod_{i=1}^n f(x_i|x_1,\dots,x_{i-1};\theta)\Pi(\mathrm{d}\theta)}{\int_\Theta \prod_{i=1}^n f(x_i|x_1,\dots,x_{i-1};\theta)\Pi(\mathrm{d}\theta)}.$$

Finally, let δ_x stand for the point mass at x.

THEOREM 1. *Suppose the model in (1) is identifiable. Then, there exists a random element $\tilde\theta$ such that*

$$\Pi_n(A) \to \delta_{\tilde\theta}(A) \quad a.s.$$

as $n \to \infty$, for any A in $\sigma(\Theta)$, and the distribution of $\tilde\theta$ coincides with Π. *Moreover, $\tilde\theta$ is essentially unique.*

Proof. Before entering the details of the proof we outline the strategy we will follow. As a first step, we work with a general prior on a space of density functions and show that the posterior admits an almost sure weak limit. This result is used to show that the sequence of predictive densities converges. Next, we translate the problem to fit it to the model (1), where the prior is defined on the (possibly infinite-dimensional) parameter space Θ. This allows the identification of a function $\tilde\theta : \Omega \to \Theta$ that is shown to be measurable. Finally, using a representation theorem, we show that the distribution of $\tilde\theta$ is the distribution Π appearing in (1).

Let Q be a distribution on the space $\mathbb{F}$ of density functions (with respect to some σ-finite measure λ) on the complete and separable metric space $\mathbb{X}$. Moreover, let $\mathcal{F}$ denote the Borel σ-algebra on $\mathbb{F}$. If $X_1,\dots,X_n$ is a sequence of observations generated by

$$\prod_{i=1}^n f(x_i|x_1,\dots,x_{i-1}),$$

the posterior distribution of f is given by

$$Q_n(A) = \frac{\int_A \prod_{i=1}^{n} f(x_i|x_1,\ldots,x_{i-1})Q(\mathrm{d}f)}{\int_{\mathbb{F}} \prod_{i=1}^{n} f(x_i|x_1,\ldots,x_{i-1})Q(\mathrm{d}f)}.$$

Let $f_n(x)$ denote the predictive density (with respect to λ) of the $(n+1)$th observation given the previous n, that is,

$$f_n(x) = \int_{\mathbb{F}} f(x|x_1,\ldots,x_n)Q_n(\mathrm{d}f).$$

Moreover, f_{nA} is the predictive density restricted to set $A \subset \mathbb{F}$, namely,

$$f_{nA}(x) = \frac{\int_A f(x|x_1,\ldots,x_n)Q_n(\mathrm{d}f)}{Q_n(A)} \qquad \forall x \in \mathbb{X}.$$

Note that

$$\begin{aligned} f_{nA}(x_{n+1})Q_n(A) &= \int_A f(x_{n+1}|x_1,\ldots,x_n)Q_n(\mathrm{d}f) \\ &= \frac{\int_A \prod_{i=1}^{n+1} f(x_i|x_1,\ldots,x_{i-1})Q(\mathrm{d}f)}{m(x_1,\ldots,x_{n+1})} \frac{m(x_1,\ldots,x_{n+1})}{m(x_1,\ldots,x_n)}, \end{aligned}$$

where $m(x_1,\ldots,x_n) = \int_{\mathbb{F}} \prod_{i=1}^{n} f(x_i|x_1,\ldots,x_{i-1})Q(\mathrm{d}f)$ is the marginal density of the vector of observations $(X_1,\ldots,X_n)$. Hence it follows that

$$\frac{Q_{n+1}(A)}{Q_n(A)} = \frac{f_{nA}(x_{n+1})}{f_n(x_{n+1})}. \tag{4}$$

Because for any set A in $\mathcal{F}$, $\mathbb{E}[Q_n(A)|X_1,\ldots,X_{n-1}] = Q_{n-1}(A)$, almost surely, the martingale convergence theorem implies

$$Q_n(A) \to Q_\infty(A) \quad \text{a.s.}$$

Thus, exploiting tightness of $\mathbb{E}[Q_n] = Q$ (for all $n \geq 1$), by Theorem 2.2 in Berti, Pratelli, and Rigo (2006), one has that Q_∞ is a random probability measure and Q_n converges weakly (almost surely) to Q_∞. At this point we will need to introduce a slight modification of the Hellinger distance between any two densities, f and g, which coincides with

$$h(f,g) = 1 - \int_{\mathbb{X}} \sqrt{f(x)g(x)}\lambda(\mathrm{d}x).$$

Set σ_n to be the σ-algebra generated by $(X_1,\dots,X_n)$. By virtue of (4) one has

$$\mathbb{E}\{Q_{n+1}^{1/2}(A)|\sigma_n\} = Q_n^{1/2}(A)\{1 - h(f_{nA},f_n)\}.$$

Consider, now, the martingale $(S_N,\sigma_N)_{N\geq 1}$ defined by

$$S_N = \sum_{n=1}^{N} [Q_n^{1/2}(A) - Q_{n-1}^{1/2}(A)\{1 - h(f_{nA},f_n)\}] \tag{5}$$

$$= Q_N^{1/2}(A) - Q^{1/2}(A) + \sum_{n=1}^{N} Q_{n-1}^{1/2}(A)\, h(f_{(n-1)A},f_{n-1}),$$

where we have obviously set $Q_0(A) := Q(A)$. Because $\mathbb{E}(S_n) = 0$ we have

$$\mathbb{E}\left\{\sum_n Q_n^{1/2}(A)\, h(f_{nA},f_n)\right\} \leq 1, \tag{6}$$

from which it follows that

$$Q_n^{1/2}(A^\epsilon)\, h(f_{nA^\epsilon},f_n) \to 0 \quad \text{a.s.}$$

as n tends to $+\infty$, which in turn yields

$$h(f_{nA^\epsilon},f_n) \to 0 \quad \text{a.s.}$$

as $n \to +\infty$. Because Q_∞ is a random probability measure, for any $\omega \in \Omega$ there exists a density f^* in $\mathbb{F}$ such that $Q_\infty(A^\epsilon) > 0$ for all $\epsilon > 0$, where $A^\epsilon = \{f \in \mathbb{F} : h(f,f^*) < \epsilon\}$. In particular, A^ϵ can be chosen in such a way that $Q_\infty(\partial A^\epsilon) = 0$, where ∂A^ϵ is the boundary of set A^ϵ. If $f_n^*(x_{n+1}) = f^*(x_{n+1}|x_1,\dots,x_n)$, by virtue of the triangular inequality one has

$$h(f_n,f_n^*) \leq h(f_n,f_{nA^\epsilon}) + h(f_{nA^\epsilon},f_n^*).$$

Because the first summand tends to zero, we only need to care about the second one. To this end, using convexity of $h(\cdot,\cdot)$, one obtains

$$h(f_{nA^\epsilon},f_n^*) \leq \int_{A^\epsilon} h(f,f_n^*)\,\frac{Q_n(\mathrm{d}f)}{Q_n(A^\epsilon)},$$

which implies that

$$h(f_n,f_n^*) \to 0 \quad \text{a.s.} \tag{7}$$

To exploit the last limiting result in the setting of the model (1), one has to interpret Q as the prior on $\mathbb{F}$ induced by Π on Θ, and Π_∞ is the (almost sure) weak limit of the sequence Π_n corresponding to Q_∞. Hence, working now on the space

Θ, we have $f_n = \int_\Theta f(\cdot|x_1,\dots,x_n,\theta)\Pi(\mathrm{d}\theta)$ and $f_n^* = f(\cdot|x_1,\dots,x_n,\theta^*)$. It can now be shown that such a θ^* is unique. To this end, suppose that the convergence in (7) holds true also for some other $\theta^{**} \in \Theta$. If $f_n^{**} = f(\cdot|x_1,\dots,x_n,\theta^{**})$, from

$$h(f_n^*, f_n^{**}) \le h(f_n^*, f_n) + h(f_n^{**}, f_n)$$

one has $h(f_n^*, f_n^{**}) \to 0$, almost surely, and by the identifiability assumption, it follows that $\theta^* = \theta^{**}$.

Now, let $\tilde{\theta} : \Omega \to \Theta$ be a function that associates to each ω a parameter value θ such that $\Pi_\infty(A^\epsilon) > 0$, for any $\epsilon > 0$. Such a function is measurable. Indeed, for any $B \in \sigma(\Theta)$,

$$\tilde{\theta}^{-1}(B) \subset \{\omega \in \Omega : \Pi_\infty(B) > 0\}.$$

On the other hand, if $\omega \in \Omega$ is such that $\Pi_\infty(B) > 0$, there exists a parameter θ in B such that θ is in the support of Π_∞. This means that

$$\tilde{\theta}^{-1}(B) \supset \{\omega \in \Omega : \Pi_\infty(B) > 0\},$$

and measurability of $\tilde{\theta}$ follows from the fact that Π_∞ is a random probability measure. Moreover, equality between the preceding two sets implies

$$\Pi_\infty = \delta_{\tilde{\theta}}.$$

By virtue of the stationarity assumption, a theorem due to Maitra (1977) implies that there exists some random element $\tilde{\eta}$ with values in Θ such that, conditional on $\tilde{\eta}$, the distribution of the observations can be represented (almost surely) as

$$\prod_{i=1}^{n} f(x_i|x_1,\dots,x_{i-1};\tilde{\eta}).$$

We aim to show that $\tilde{\theta} = \tilde{\eta}$, almost surely. If $P_{\tilde{\eta}}(B|x_1,\dots,x_n) = \int_B f(x|x_1,\dots,x_n;\tilde{\eta})\lambda(\mathrm{d}x)$ for any B in $\mathcal{X}$, then

$$\begin{aligned}
\mathbb{E}\{P_{\tilde{\eta}}(B|x_1,\dots,x_n)\} = \mathbb{P}[X_{n+1} \in B|x_1,\dots,x_n] &= \int_\Theta P_\theta(B|x_1,\dots,x_n)\Pi_n(\mathrm{d}\theta) \\
&= \int_\Theta P_\theta(B|x_1,\dots,x_n)\mathbb{E}[\Pi_\infty(\mathrm{d}\theta)|x_1,\dots,x_n] \\
&= \mathbb{E}\left[\int_\Theta P_\theta(B|x_1,\dots,x_n)\Pi_\infty(\mathrm{d}\theta)\,\middle|\, x_1,\dots,x_n\right],
\end{aligned}$$

where the last equality follows from the definition of conditional expectation. Because $P_{\tilde{\eta}}(B|x_1,\ldots,x_n)$ and $\int_\Theta P_\theta(B|x_1,\ldots,x_n)\Pi_\infty(\mathrm{d}\theta)$ are bounded, one has

$$\mathbb{E}[P_{\tilde{\eta}}(B|x_1,\ldots,x_n)|x^{(\infty)}] = \mathbb{E}\left[\int_\Theta P_\theta(B|x_1,\ldots,x_n)\Pi_\infty(\mathrm{d}\theta)\middle| x^{(\infty)}\right]$$

$$= \int_\Theta P_\theta(B|x_1,\ldots,x_n)\Pi_\infty(\mathrm{d}\theta) = P_{\tilde{\theta}}(B|x_1,\ldots,x_n),$$

where the last equality follows from the fact that $\Pi_\infty = \delta_{\tilde{\theta}}$ and $X^{(\infty)}$ denotes the whole sequence of observations. Moreover, $\tilde{\eta}$ is measurable with respect to the σ-algebra generated by $X^{(\infty)}$. Hence

$$P_{\tilde{\eta}}(B) = P_{\tilde{\theta}}(B) \quad \text{a.s.}$$

for every B in $\mathcal{X}$. This entails $\tilde{\theta} = \tilde{\eta}$ (almost surely). Hence the result follows. ■

3. CONCLUDING REMARKS

We have obtained a quite remarkable result regarding the asymptotic properties of Bayesian procedures for stationary models. Bayesians can be confident in carrying out analysis of models much more general than those based on the common assumption of exchangeability. An open question that remains is the case of nonstationary data, for which we are not able to extend the result of the present paper because of the lack of a suitable representation theorem. The proof to Theorem 1 can still be used to state that the sequence of posterior distributions converges to a point mass at a random parameter, but there is no guarantee that this random parameter is the one that generates the sequence. Indeed, in this more general case one can still assume that the joint distribution of the observations is written as a mixture with respect to the distribution of some parameter. However, the representation theorem does not apply, and the mixing parameter from which the data are generated is not identifiable. This is an important remark because it means that a Bayesian analysis of nonstationary data could lead to an answer in the form of the posterior converging to a point mass at the wrong parameter.

REFERENCES

Aït-Sahalia, Y. (2002) Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* 70, 223–262.

Aldous, D. (1985) *Exchangeability and Related Topics.* Lecture Notes in Mathematics 1117. Springer-Verlag.

Berti, P., L. Pratelli, & P. Rigo (2006) Almost sure weak convergence of random probability measures. *Stochastics* 78, 91–97.

Beskos, A., O. Papaspiliopoulos, G.O. Roberts, & P. Fernhead (2006) Exact and efficient likelihood-based estimation for diffusion processes. With discussion. *Journal of the Royal Statistical Society, Series B* 68, 333–382.

Bibby, B.M. & M. Sørensen (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* 1, 17–39.

Billingsley, P. (1961) *Statistical Inference for Markov Processes.* Statistical Research Monographs. University of Chicago Press.

Brockwell, P.J. & R.A. Davis (1996) *Time Series: Theory and Methods*. 3rd ed. Springer-Verlag.

Chib, S. (1996) Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* 75, 79–97.

Chib, S. & E. Greenberg (1996) Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409–431.

Dacunha-Castelle, D. & D. Florens-Zmirou (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* 19, 263–284.

Doob, J.L. (1949) Application of the theory of martingales. In *Le Calcul des Probabilités et ses Applications*, pp. 23–27. Colloques Internationaux du Centre National de la Recherche Scientifique, no. 13, Paris.

Elerian, O., S. Chib, & N. Shephard (2001) Likelihood inference for discretely observed non-linear diffusions. *Econometrica* 69, 959–993.

Eraker, B. (2001) MCMC analysis of diffusion models with applications to finance. *Journal of Business & Economic Statistics* 19, 177–191.

Fiorentini, G., E. Sentana, & N. Shephard (2004) Likelihood-based estimation of latent generalised ARCH structures. *Econometrica* 72, 1481–1517.

Gourieroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, 85–118.

Huerta, G. & M. West (1999a) Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society, Series B* 61, 881–899.

Huerta, G. & M. West (1999b) Bayesian inference on periodicities and component spectral structure in time series. *Journal of Time Series Analysis* 20, 401–416.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*. Springer-Verlag.

Lijoi, A., I. Prünster, & S.G. Walker (2004) Extending Doob's consistency theorem to nonparametric densities. *Bernoulli* 10, 651–663.

Maitra, A. (1977) Integral representations of invariant measures. *Transactions of the American Mathematical Society* 229, 209–225.

Mena, R.H. & S.G. Walker (2005) Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis* 26, 789–805.

Pedersen, A.R. (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* 22, 55–71.

Pitt, M.K. & N. Shephard (1999) Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 94, 590–599.

Roberts, G.O. & O. Stramer (2001) On inference for partially observed non-linear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* 88, 603–621.

West, M. & J. Harrison (1997) *Bayesian Forecasting and Dynamic Models*. Springer-Verlag.