Bayesian Consistency for Markov Models

Isadora Antoniano-Villalobos*
Bocconi University, Milan, Italy.

Stephen G. Walker
University of Texas at Austin, USA.

This is a post-peer-review, pre-copyedit version of an article published in Sankhya A. *The final authenticated version is available online at:* https://doi.org/10.1007/s13171-014-0055-2

Abstract

We consider sufficient conditions for Bayesian consistency of the transition density of time homogeneous Markov processes. To date, this remains somewhat of an open problem, due to the lack of suitable metrics with which to work. Standard metrics seem inadequate, even for simple autoregressive models. Current results derive from generalizations of the i.i.d. case and additionally require some non-trivial model assumptions. We propose suitable neighborhoods with which to work and derive sufficient conditions for posterior consistency which can be applied in general settings. We illustrate the applicability of our result with some examples; in particular, we apply our result to a general family of nonparametric time series models.

AMS 2000 subject classifications: Primary 62G20; secondary 62M05.

Keywords: Nonparametric mixture; posterior consistency; transition density; Markov process; martingale sequence.

*PhD research funded by CONACyT.

1 Introduction

Consider an ergodic Markov process $\{X_n\}_{n\geq 0}$, defined on some separable filtered space $(\mathcal{X}, \mathcal{G}, \{\mathcal{G}_n\}_{n\geq 0})$. Denote by $\mathbb{P}_0$ the true law of the process. Throughout this paper, all probability statements will be made with respect to $\mathbb{P}_0$.

Assume the process is time homogeneous and let f_0 be the transition density for $\{X_n\}_{n\geq 0}$, with respect to some reference measure ν. Let ν_0 be the corresponding ergodic measure of the process. That is, for every $G \in \mathcal{G}$

$$\mathbb{P}_0[X_{n+1} \in G | X_n] = \int_G f_0(x|X_n) \mathrm{d}\nu(x),$$

and for every $h \in L^1(\nu_0)$

$$\frac{1}{n}\sum_{i=1}^{n} h(X_n) \to \int h(x)\mathrm{d}\nu_0(x) \quad \text{a.s. when } n \to \infty.$$

In particular, if the process has a stationary density g_0, then the integral in the limit is equal to $\int h(x)g_0(x)\mathrm{d}\nu(x)$.

If f_0 is fixed but unknown, Bayesian inference begins by constructing a prior distribution Π over the class $\mathcal{F}$ of transition densities on $(\mathcal{X}, \mathcal{G})$ with respect to ν. As in the case of i.i.d. observations, this prior combines with the data to define the posterior distribution Π^n. So, if A is a set of transition densities, the posterior mass assigned to it is given by

$$\Pi^n(A) = \frac{\int_A R_n(f)\mathrm{d}\Pi(f)}{\int R_n(f)\mathrm{d}\Pi(f)},$$

where

$$R_n(f) = \prod_{i=1}^{n} \frac{f(X_i|X_{i-1})}{f_0(X_i|X_{i-1})}$$

is the likelihood ratio. In order to simplify the notation, here and in the following, we assume that X_0 is either fixed or has a known initial distribution.

As is well known, the predictive density for X_{n+1}, given $X_0, \ldots, X_n$, is

$$f_n(\cdot \mid X_n) = \mathbb{E}[f(\cdot \mid X_n) \mid X_0, \ldots, X_n] = \int f(\cdot \mid X_n) \mathrm{d}\Pi^n(f),$$

and hence the importance of accurate estimation of f and the study of posterior consistency.

The general Markov process model is said to be consistent if the posterior mass accumulates around f_0 as n increases. More formally, Π^n is consistent at f_0 if for every suitable neighbourhood B of f_0, we have $\Pi^n(B^c \mid X_0, \ldots, X_n) \to 0$ a.s. as $n \to \infty$. The concept depends on the definition of such neighbourhoods, so different types of consistency can be considered.

Consistency results for transition densities of Markov processes extend the two main approaches regarding consistency for i.i.d. observations. The first approach, due to Schwartz [1965], Barron et al. [1999], and Ghosal et al. [1999], assumes the existence of an increasing sequence of sets, a sieve $\{\mathcal{F}_n\} \subset \mathcal{F}$, with $\Pi(\mathcal{F}_n^c)$ exponentially small; a sequence of uniformly consistent tests for testing $f = f_0$ against $f \in B^c \cap \mathcal{F}_n$; and the Kullback-Leibler property on the prior, which states that $\Pi[f : K(f, f_0) < \epsilon] > 0$ for every $\epsilon > 0$, where $K(f, f_0) = \int f_0 \log(f_0/f)$ is the Kullback-Leibler divergence between the densities f and f_0. In practice, the sieve is constructed as a finite union of balls around functions $\{f_j\}$, which then need to be counted to ensure a linear constraint on the entropy of $\mathcal{F}_n$.

An alternative approach provided by Walker [2003, 2004] relies on a martingale sequence to obtain sufficient conditions for consistency in the i.i.d. case. The construction turns out to be equivalent to the use of a suitable sieve, based on Hellinger balls, and which depends on the prior. Such a sieve satisfies the entropy condition automatically, simplifying the verification of necessary conditions for consistency.

For i.i.d. observations, a distinction is made between weak and strong consistency, the second of which is associated with the Hellinger distance on the space of densities, defined as $H\big(f_1, f_2\big) = \frac{1}{2}(\int \big(\sqrt{f_1} - \sqrt{f_2}\big)^2)^{1/2}$. The Hellinger distance is topologically equivalent to the L_1 distance, but the definition of the former is useful in the study of consistency. The literature concerning consistency for Markov processes is limited, due in great part to the difficulty in finding adequate topologies and distances between transition densities. It is not clear the Hellinger distance can be generalized for the space of transition densities and still be convenient in the context of consistency. However, a straightforward generalization of the Kullback-Leibler property for transition densities is possible due to the ergodic property (see Section 3.3).

To highlight the problem of extending the Hellinger distance between densities to the space of transition densities, consider the squared Hellinger distance between $f_1(\cdot|y)$ and $f_2(\cdot|y)$, given by

$$\begin{aligned} H^2\big(f_1(\cdot|y), f_2(\cdot|y)\big) &= \frac{1}{2}\int \Big(\sqrt{f_1(z|y)} - \sqrt{f_2(z|y)}\Big)^2 \,\mathrm{d}\nu(z) \\ &= 1 - \int \sqrt{f_1(z|y) f_2(z|y)}\,\mathrm{d}\nu(z). \end{aligned} \tag{1}$$

Here, H can not be used to define a topology on $\mathcal{F}$, as it depends on y. In order to adapt this and other quantities commonly used for densities, to define neighbourhoods in a space of transition densities, the dependence on y must somehow be eliminated.

A similar problem appears in the study of posterior consistency for regression models, where a distance between densities for the response variable z depends on the value of a covariate y. In this context, Ghosal and Roy [2006] and Choi and Schervish [2007] define a distance between two densities f_1 and f_2 as

$$h(f_1, f_2) = \int H\big(f_1(\cdot|y), f_2(\cdot|y)\big)\mathrm{d}Q(y),$$

where Q is the distribution for the covariate. The definition of an adequate metric in terms of the Hellinger distance is due to the availability of the measure Q when assuming the covariates are generated stochastically and i.i.d. In the Markov process case, an adequate choice of integrating measure is unclear, but the general idea can be applied for measures with a large enough support.

Tang and Ghosal [2007] propose different ways of defining a topology on a transition density space. The first is based on the notion of distances on the invariant measures associated with each transition, and results in a weak topology. Alternative ideas arise from using integrated and maximized distances between conditional densities respectively, resulting in strong types of neighbourhoods in both cases. In the same paper (Sections 7 and 8, respectively), the authors prove strong consistency in this sense for a specific family of transition densities based on Dirichlet mixtures, by generalizing the sieve and uniformly consistent tests approach. Ghosal and Tang (2006, Theorem 2.3) extend this result to a general family $\mathcal{F}$ of transitions, providing it is compact with respect to the supremum Hellinger distance,

$$H_s\big(f_1, f_2\big) = \sup_y H\big(f_1(\cdot|y), f_2(\cdot|y)\big). \tag{2}$$

Compactness with respect to H_s is a rather strong condition, as a simple example may show. Let $\mathcal{F} = \{N(\cdot|\theta y, 1) : \theta \in \Theta \subset \mathbb{R}\}$. The Hellinger distance between transition densities

in this case is given by

$$H^2\big(f_{\theta_1}(\cdot|y), f_{\theta_2}(\cdot|y)\big) = 1 - \exp\left\{-\frac{1}{8}y^2(\theta_1 - \theta_2)^2\right\}.$$

Therefore, $H_s\big(f_{\theta_1}, f_{\theta_2}\big) = 1$ for every $\theta_1 \neq \theta_2$, so the required compactness is achieved only when Θ is finite.

Constructing an adequate sieve and proving the existence of a set of uniformly consistent tests is difficult in general. Therefore, in order to remove the compactness assumption, Ghosal and Tang [2006] generalize the martingale approach of Walker [2003, 2004]. By assuming only the separability of $\mathcal{F}$ with respect to H_s, they are then able to prove consistency with respect to neighbourhoods of the type $\{f : \tilde{d}(f, f_0) < \epsilon\}$, where

$$\tilde{d}(f, f_0) = \inf_y H^2\big(f(\cdot|y), f_0(\cdot|y)\big). \tag{3}$$

Some families of transition densities may be found for which this type of consistency can be considered strong enough; however, in general, these neighbourhoods correspond to a weak topology. Once more, we illustrate this through the example, $\mathcal{F} = \{N(\cdot|\theta y, 1) : \theta \in \mathbb{R}\}$. When $y = 0$, for every $f \in \mathcal{F}$ we have $f(\cdot|0) = N(\cdot|0, 1)$, yielding

$$\inf_y H^2\big(f_1(\cdot|y), f_2(\cdot|y)\big) = 0 \text{ for any } f_1, f_2 \in \mathcal{F}.$$

Ghosal and Tang [2006] mention this problem, which extends to the nonlinear autoregressive model $f(y|x) = g(y - \psi(x))$, whenever g is a location shift of g_0. Throughout this paper, we develop a more general result for consistency which does not require compactness of $\mathcal{F}$ with respect to H_s, and applies to a wider family of processes.

Our main contribution is the definition of a neighbourhood around the true transition

density, f_0, using a natural adaptation of the Hellinger distance for bivariate densities. Each transition density $f \in \mathcal{F}$ is extended to a family of bivariate densities. A distance like-operator between f and f_0 is defined as the smallest distance between sets of extended bivariate densities (to be explained in section 2). This, as we shall see, guarantees the definition of strong neighbourhoods around f_0, for a relevant family of models.

We then find sufficient conditions for consistency by extending the martingale result from Walker [2004], assuming only separability of $\mathcal{F}$ with respect to the supremum Hellinger distance H_s.

The layout of the paper is as follows. In Section 2 we present the system of neighbourhoods which will provide the base for a strong form of consistency for the true transition density f_0. In Section 3, we present the basic notation and provide a set of sufficient conditions for posterior consistency. In Section 4 we present some examples, the first one to illustrate the contribution of our result; the second, a family of nonparametric mixture Markov models which has gained popularity in the literature over the last years; the third shows how the result can be useful even when an analytic expression for the transition density is not available.

2 A strong neighborhood around f_0

In this section, we define a non negative binary operator $d : \mathcal{F} \times \mathcal{F} \to [0, 1]$, and discuss the suitability of the neighbourhoods $\{f \in \mathcal{F} : d(f, f_0) < \epsilon\}$ in the study of posterior consistency for Markov processes.

Consider the set $\overline{\mathcal{F}}$ of bivariate densities on $(\mathcal{X} \times \mathcal{X}, \mathcal{G} \otimes \mathcal{G})$. For every $\bar{f}_1, \bar{f}_2 \in \overline{\mathcal{F}}$, the

squared Hellinger distance is given by

$$H^2(\bar{f}_1, \bar{f}_2) = \frac{1}{2}\int \left(\sqrt{\bar{f}_1} - \sqrt{\bar{f}_2}\right)^2 \mathrm{d}(\nu \times \nu) = 1 - \int\int \sqrt{\bar{f}_1 \bar{f}_2}\, \mathrm{d}(\nu \times \nu).$$

For each $x \in \mathcal{X}$, and $f \in \mathcal{F}$, a density in $\overline{\mathcal{F}}$ is defined by

$$\bar{f}(z, y|x) = f(z|y)f(y|x).$$

Therefore, for any $f_1, f_2 \in \mathcal{F}$, we can define

$$d(f_1, f_2) = \inf_x d_x(f_1, f_2), \tag{4}$$

where $d_x(\cdot, \cdot)$ denotes the Hellinger distance between the two corresponding bivariate densities in $\overline{\mathcal{F}}$. That is,

$$\begin{aligned} d_x^2(f_1, f_2) &= H^2\big(\bar{f}_1(\cdot, \cdot|x), \bar{f}_2(\cdot, \cdot|x)\big) \\ &= 1 - \int\int \sqrt{f_1(z|y)f_1(y|x)f_2(z|y)f_2(y|x)}\, \mathrm{d}\nu(z)\mathrm{d}\nu(y). \end{aligned}$$

In the context of strong consistency, we are interested in ensuring that the operator d allows us to identify the true transition density f_0 in a strong sense. Specifically, we want to ensure that, for every $f \in \mathcal{F}$,

$$d(f, f_0) = 0 \Leftrightarrow \int_{\mathcal{X}} \mathbb{P}(\cdot|x)\mathrm{d}\nu(x) = \int_{\mathcal{X}} \mathbb{P}_0(\cdot|x)\mathrm{d}\nu(x), \tag{5}$$

thus allowing for the two conditional probability measures, associated to the densities, $\mathbb{P}(\cdot|x)$ and $\mathbb{P}_0(\cdot|x)$, to differ at most for values of x on a set of ν-measure 0. Clearly, in the above

expression,

$$\mathbb{P}(A|x) = \int_A f(y|x)\mathrm{d}\nu(y) \text{ for every } A \in \mathcal{G}.$$

The implication to the left, in expression (5) is trivially satisfied, by the definition of d in terms of a Hellinger distance. We are therefore required to find conditions which ensure

$$\int_{\mathcal{X}} \mathbb{P}(\cdot|x)\mathrm{d}\nu(x) \neq \int_{\mathcal{X}} \mathbb{P}_0(\cdot|x)\mathrm{d}\nu(x) \Rightarrow d(f, f_0) > 0.$$

Equivalently,

$$\nu(S) > 0 \Rightarrow d(f, f_0) > 0, \text{ for } S = \{x \in \mathcal{X} : H\big(f(\cdot|x), f_0(\cdot|x)\big) > 0\}. \tag{6}$$

Notice that for every $x \in \mathcal{X}$ and every $f_1, f_2 \in \mathcal{F}$,

$$\begin{aligned} d_x^2(f_1, f_2) &= H^2\big(f_1(\cdot|x), f_2(\cdot|x)\big) + \int H^2\big(f_1(\cdot|y), f_2(\cdot|y)\big)\sqrt{f_1(y|x)f_2(y|x)}\,\mathrm{d}\nu(y), \\ &= H^2\big(f_1(\cdot|x), f_2(\cdot|x)\big) + \mathbb{E}_0\left[H^2\big(f_1(\cdot|y), f_2(\cdot|y)\big)\sqrt{\frac{f_1(y|x)}{f_2(y|x)}}\middle| x\right] \end{aligned}$$

so that, applying the Cauchy-Swartz inequality leads to

$$d_x^2(f_1, f_2) \leq \sqrt{\int_S f_2(y|x)\mathrm{d}\nu(y)} := \sqrt{\mathbb{P}_2(S|x)}$$

for every $x \in S' = \mathcal{X} \setminus S$. And, by the symmetry of d_x, the same is true for $\mathbb{P}_1(S|x)$. Therefore, applying this to f and f_0, we have that a necessary condition to satisfy (6) is

$$\inf_{x \in S'} \mathbb{P}(S|x) > 0 \text{ for every } f \in \mathcal{F} \text{ and } S \in \mathcal{G} \text{ such that } \nu(S) > 0. \tag{7}$$

When the family $\mathcal{F}$ under consideration is large enough, as is often the case, particularly in nonparametric settings, the sets S considered above can be arbitrarily small. In such cases, the above condition is practically equivalent to the sufficient condition provided by the following lemma.

Lemma 2.1. *Assume that for every $f \in \mathcal{F}$ and every $S \in \mathcal{G}$ with $\nu(S) > 0$,*

$$\inf_{x \in \mathcal{X}} \mathbb{P}(S|x) > 0. \tag{8}$$

Then, condition (6) is satisfied. In this case, d is a semimetric which induces a strong topology on $\mathcal{F}$.

Proof. Let $f \in \mathcal{F}$ and $x \in \mathcal{X}$. We first observe that, for each y,

$$H^2\big(f(\cdot|y), f_0(\cdot|y)\big) = 1 - \mathbb{E}_0\left[\sqrt{\frac{f(z|y)}{f_0(z|y)}}\,\middle|\, y\right],$$

where $\mathbb{E}_0$ denotes the expectation with respect to $\mathbb{P}_0$. We may also observe that for each x,

$$d_x^2(f, f_0) = 1 - \mathbb{E}_0\left[\sqrt{\frac{f(z|y)f(y|x)}{f_0(z|y)f_0(y|x)}}\,\middle|\, x\right].$$

Now,

$$\begin{aligned}\mathbb{E}_0\left[\sqrt{\frac{f(z|y)}{f_0(z|y)}}\sqrt{\frac{f(y|x)}{f_0(y|x)}}\,\middle|\, x\right] &= \mathbb{E}_0\left[\mathbb{E}_0\left[\sqrt{\frac{f(z|y)}{f_0(z|y)}}\,\middle|\, y\right]\sqrt{\frac{f(y|x)}{f_0(y|x)}}\,\middle|\, x\right] \\ &= \mathbb{E}_0\left[\Big(1 - H^2\big(f(\cdot|y), f_0(\cdot|y)\big)\Big)\sqrt{\frac{f(y|x)}{f_0(y|x)}}\,\middle|\, x\right].\end{aligned}$$

Therefore, since $0 \le 1 - H^2\big(f(\cdot|y), f_0(\cdot|y)\big) \le 1$, from the Cauchy-Schwartz inequality and

$\mathbb{E}_0\left[f(y|x)/f_0(y|x)\,|\,x\right] = 1$, we have

$$\mathbb{E}_0\left[\sqrt{\frac{f(z|y)f(y|x)}{f_0(z|y)f_0(y|x)}}\,\middle|\,x\right] \leq \sqrt{1-\mathbb{E}_0\Big[H^2\big(f(\cdot|y),f_0(\cdot|y)\big)\,\big|\,x\Big]}$$

and so

$$d_x^2(f,f_0) \geq 1-\sqrt{1-\mathbb{E}_0\Big[H^2\big(f(\cdot|y),f_0(\cdot|y)\big)\,\big|\,x\Big]} \geq 0. \tag{9}$$

For S defined as in (6),

$$\mathbb{E}_0\Big[H^2\big(f(\cdot|y),f_0(\cdot|y)\big)\,\big|\,x\Big] = \int_S H^2\big(f(\cdot|y),f_0(\cdot|y)\big)f_0(y|x)\mathrm{d}\nu(y).$$

Thus, $d(f,f_0)=0 \Rightarrow \inf_x \mathbb{E}_0\Big[H^2\big(f(\cdot|y),f_0(\cdot|y)\big)\,\big|\,x\Big] = 0$ and condition (8) ensures this only happens when $\nu(S)=0$. Since the operator d is symmetric and f_0 does not play a particular roll, the proof works for any two distinct densities in $\mathcal{F}$ and thus d defines a semimetric which induces a strong topology. □

For many Markov models found in the literature, rather than verifying condition (8) directly, we may use a simplified condition, provided in the following corollary.

Corollary 2.1. *Assume that for every transition density $f \in \mathcal{F}$, $f(\cdot|x)$ is a continuous function of x such that*

$$\inf_x f(y|x) > \beta g(y) \quad \forall\, y \in \mathcal{X}, \tag{10}$$

for some density function $g = g(f)$ with full support over $(\mathcal{X},\mathcal{G})$ and some constant $\beta = \beta(f) > 0$. Then, condition (6) is satisfied and d is a semimetric which induces a strong topology on $\mathcal{F}$.

Proof. For every f in $\mathcal{F}$ and $S \in \mathcal{G}$ such that $\nu(S) > 0$. It follows from the continuity of the

transition densities, the full support of g and condition (10) that

$$\inf_x \mathbb{P}(S|x) = \inf_x \int_S f(y|x)\mathrm{d}\nu(y) > \beta \int_S g(y)\mathrm{d}\nu(y) = \beta \mathbb{P}_g(S) > 0,$$

where $\mathbb{P}_g$ denotes the probability measure associated to the density g. The result follows from Lemma 2.1. □

It is important to emphasize that the definition of d in (4) differs from the one given by Ghosal and Tang [2006] (equation 3), in that the integral for the Hellinger distance in our definition is taken with respect to the product measure $\nu \times \nu$, and not with respect to ν. In other words, it minimizes the Hellinger distance between bivariate, rather than univariate conditional densities. The effect of this can be best understood by revisiting our example involving the simple normal regression model. If $f_\theta(\cdot|y) = N(\cdot|\theta y, 1)$ as before, the squared Hellinger distance between the univariate functions $f_{\theta_1}(\cdot|y)$ and $f_{\theta_2}(\cdot|y)$ is given by

$$H^2\big(f_{\theta_1}(\cdot|y), f_{\theta_2}(\cdot|y)\big) = 1 - \exp\left\{-\frac{1}{8}y^2(\theta_1 - \theta_2)^2\right\};$$

while the squared Hellinger distance between the bivariate functions $\bar{f}_{\theta_1}$ and $\bar{f}_{\theta_2}$ is

$$d_x^2(f_{\theta_1}, f_{\theta_2}) = 1 - \frac{2}{\sqrt{4 + (\theta_1 - \theta_2)^2}} \exp\left\{-\frac{x^2}{2}(\theta_1 - \theta_2)^2\Big[1 - \frac{2(1 - \theta_1\theta_2)}{4 + (\theta_1 - \theta_2)^2}\Big]\right\}.$$

Notice that $H^2\big(f_{\theta_1}(\cdot|0), f_{\theta_2}(\cdot|0)\big) = 0$ even if $\theta_1 \neq \theta_2$; while $\inf_x d_x^2(f_{\theta_1}, f_{\theta_2}) = d_0^2(f_{\theta_1}, f_{\theta_2}) = 1 - 2/\sqrt{4 + (\theta_1 - \theta_2)^2}$. This is actually a distance and therefore, we have $d(f_{\theta_1}, f_{\theta_2}) > 0$ whenever $|\theta_1 - \theta_2| > 0$.

In fact, d defines a distance on the family $\mathcal{F}$ of transition densities of interest whenever the infimum is reached at an internal point in $\mathcal{X}$, i.e., if $d(f_1, f_2) = d_{x^*}^2(f_1, f_2)$ for some $x^* \in \mathcal{X}$.

Although this is not true in general, we have shown that under adequate conditions, d defines a semimetric on the space of conditional probabilities with densities in a general space $\mathcal{F}$, thus inducing a strong topology on $\mathcal{F}$. Furthermore, using similar arguments to those in the proof of Lemma 2.1, we can see that d is bounded above by the supreme Hellinger distance. Additionally, if condition (10) is satisfied, uniformly, i.e. g and β do not depend on f, then it follows from (9) that d is bounded below by an integrated metric:

$$\frac{\beta}{2}\int H^2(f_1(\cdot|y), f_2(\cdot|y))\, g(y)\, \mathrm{d}\nu(y) \leq d^2(f_1, f_2) \leq d_x^2(f_1, f_2) \leq 2H_s^2(f_1, f_2). \tag{11}$$

This inequality will be useful for the consistency result in the next section.

3 Posterior Consistency

In this section we will establish the basic notation [following the setup of Walker, 2004, Ghosal and Tang, 2006] and present the main theorem regarding consistency.

3.1 Preliminaries and notation

Let $X^n = (X_1, \ldots, X_n)$ denote a sample of size n from $\mathbb{P}_0$ (formally, from the restriction of $\mathbb{P}_0$ to $\mathcal{G}_n$). The likelihood ratio for a transition density $f \in \mathcal{F}$ will be denoted by

$$R_n(f) = \prod_{i=1}^{n} \frac{f(X_i|X_{i-1})}{f_0(X_i|X_{i-1})}.$$

Let Π denote a prior on $\mathcal{F}$ and define the integrated likelihood ratio over a subset $A \subset \mathcal{F}$ as

$$L_n = L_{nA} = \int_A R_n(f)\mathrm{d}\Pi(f).$$

The posterior mass assigned to A is then given by

$$\Pi^n(A) = \frac{L_n}{I_n},$$

where $I_n = L_{n\mathcal{F}} = \int R_n(f)\mathrm{d}\Pi(f)$.

Finally, we define the bivariate predictive density, with posterior restricted to the set A as

$$f_{nA}(z, y|X_n) = \int_A f(z|y)f(y|X_n)\mathrm{d}\Pi_A^n(f),$$

where

$$\mathrm{d}\Pi_A^n(f) = \frac{\mathbf{1}(f \in A)\mathrm{d}\Pi^n(f)}{\int_A \mathrm{d}\Pi^n(f)}.$$

The key identity here is:

$$\frac{L_{n+2}}{L_n} = \frac{f_{nA}(X_{n+2}, X_{n+1}|X_n)}{f_0(X_{n+2}|X_{n+1})f_0(X_{n+1}|X_n)}. \tag{12}$$

Notice that in this case, the ratio is defined with a step of size 2, while in the i.i.d. case a size 1 step is sufficient. Now, $\mathbb{E}_0[L_2|\mathcal{G}_0] = L_0 = \Pi(A)$ and $\mathbb{E}_0[L_{2n+2}|\mathcal{G}_{2n}] = L_{2n}\mathbb{E}_0[L_{2n+2}/L_{2n}|\mathcal{G}_{2n}] = L_{2n}$ for every $n \geq 1$, therefore $\{L_{2n}\}$ is a martingale with respect to $\{\mathcal{G}_{2n}\}$. Analogously, $\{L_{2n+1}\}$ is a martingale with respect to $\{\mathcal{G}_{2n+1}\}$, since $\mathbb{E}_0[L_{2n+3}|\mathcal{G}_{2n+1}] = L_{2n+1}$ for every $n \geq 0$.

Recall that the posterior mass assigned to $A \subset \mathcal{F}$, given a sample of size n, is given by

$$\Pi^n(A) = \frac{L_n}{I_n}. \tag{13}$$

Different results regarding posterior consistency deal with the numerator and the denominator in this expression separately.

3.2 The numerator

The following lemma regards a general property, essential for the treatment of the numerator in equation (13).

Lemma 3.1. *For each* $n \geq 1$

$$\mathbb{E}_0\Big[\sqrt{L_{n+2}}\,\Big|\,\mathcal{G}_n\Big] \leq \sqrt{L_n}\,\Big[1 - d^2_{X_n}(f_{nA}, f_0)\Big].$$

Proof. Notice that L_n is $[\mathcal{G}_n]$-measurable, so

$$\frac{\mathbb{E}_0\big[\sqrt{L_{n+2}}\,\big|\,\mathcal{G}_n\big]}{\sqrt{L_n}} = \mathbb{E}_0\left[\sqrt{\frac{L_{n+2}}{L_n}}\,\middle|\,\mathcal{G}_n\right].$$

Applying the identity (12), and rearranging terms, we obtain

$$\mathbb{E}_0\Big[\sqrt{L_{n+2}}\,\Big|\,\mathcal{G}_n\Big] = \sqrt{L_n}\,\mathbb{E}_0\left[\sqrt{\frac{f_{nA}(X_{n+2}, X_{n+1}|X_n)}{f_0(X_{n+2}|X_{n+1})f_0(X_{n+1}|X_n)}}\,\middle|\,\mathcal{G}_n\right].$$

By applying the definition of d_{X_n},

$$\mathbb{E}_0\Big[\sqrt{L_{n+2}}\,\Big|\,\mathcal{G}_n\Big] = \sqrt{L_n}\,\big[1 - d^2_{X_n}(f_{nA}, f_0)\big].$$

This completes the proof. □

Consider a set A of transition densities. If we assume $\mathcal{F}$ is separable with respect to some distance, d^* then for every $\delta > 0$, we can find a d^*-cover for A of size δ. That is, a collection

$\{A_j\}_{j\geq 1}$ such that

$$A \subseteq \bigcup_j A_j$$

and for each j there exists $f_j \in A$ for which

$$A_j = \Big\{ f : d^*(f, f_j) < \delta \Big\}.$$

Lemma 3.2. *Let $A_\epsilon \subset \mathcal{F}$ be a set of transition densities d-bounded away from f_0, i.e.*

$$A_\epsilon = \Big\{ f \in \mathcal{F} : d(f, f_0) > \epsilon \Big\}.$$

Assume that a H_s-cover for A_ϵ of size $\delta < \frac{\epsilon}{\sqrt{2}}$ can be found such that

$$\sum_{j=1}^{\infty} \sqrt{\Pi(A_j)} < \infty. \tag{14}$$

Then, for some $b > 0$

$$\sum_{j=1}^{\infty} \sqrt{L_{nA_j}} < \exp(-nb) \quad a.s. \quad \forall \text{ large } n.$$

Proof. Let $\{A_j\}_{j\geq 1}$ be a cover satisfying assumption (14). Let $\gamma = \epsilon - \sqrt{2}\delta > 0$. For simplicity, denote $L_{nj} = L_{nA_j}$ and $f_{nj} = f_{nA_j}$.

Observe that $d(f, g) \leq d_x(f, g) \leq \sqrt{2} H_s\big(f, g\big)$, for any two densities $f, g \in \mathcal{F}$ and $x \in \mathcal{X}$, by the definition of d and inequality (11). Therefore, for each j,

$$d_{X_n}(f_{nj}, f_0) \geq d_{X_n}(f_j, f_0) - d_{X_n}(f_{nj}, f_j) \geq d(f_j, f_0) - \sqrt{2} H_s\big(f_{nj}, f_j\big) > \gamma.$$

We know from expression (12) that

$$L_{n+2j} = L_{nj} \, \frac{f_{nj}(X_{n+2}, X_{n+1}|X_n)}{f_0(X_{n+2}|X_{n+1}) f_0(X_{n+1}|X_n)},$$

with $L_{0j} = \Pi(A_j)$ by definition. Taking conditional expectations and applying Lemma 3.1, we get

$$\mathbb{E}_0 \left[\sqrt{L_{n+2j}} \, \middle| \mathcal{G}_n \right] \leq \sqrt{L_{nj}} \left\{ 1 - d^2_{X_n}(f_{nj}, f_0) \right\} < \sqrt{L_{nj}} \, (1 - \gamma^2).$$

Now, if we let k be the smallest integer larger than $n/2$, by iterating to remove the conditionality with respect to $\mathcal{G}_n$, we find

$$\mathbb{E}_0 \left[\sqrt{L_{n+2j}} \right] < \sqrt{L_{0j}} \, (1 - \gamma^2)^k < \sqrt{\Pi(A_j)} \, (1 - \gamma^2)^{(n+2)/2}.$$

Markov's inequality implies that, for any $b > 0$,

$$\mathbb{P}_0 \left[\sum_{j=1}^{\infty} \sqrt{L_{nj}} > \exp(-nb) \right] < \exp(nb) \, (1 - \gamma^2)^{n/2} \sum_{j=1}^{\infty} \sqrt{\Pi(A_j)}.$$

Finally, taking $b < -\log(1 - \gamma^2)/2$, by condition (14), we arrive at

$$\sum_{j=1}^{\infty} \sqrt{L_{nj}} < \exp(-nb) \quad \text{a.s.}$$

for all large n. □

Observe that if $\mathcal{F}$ is separable with respect to H_s, then it is separable with respect to d. This follows from the inequality (11). On the other hand, it is possible for $\mathcal{F}$ to be separable

with respect to d even if it is not so with respect to H_s. In particular, as mentioned before (Section 2), if $d(f_1, f_2) = d^2_{x^*}(f_1, f_2)$ for some $x^* \in \mathcal{X}$, then d satisfies the triangle inequality. Therefore, Lemma 3.2 still holds when the H_s-cover is substituted by a d-cover, since the critical inequality $d_{X_n}(f_{nj}, f_0) > \gamma$ in the proof is satisfied. This is fundamental for the verification of strong consistency in example 4.1.

3.3 The denominator

For every $y \in \mathcal{X}$, the Kullback-Leibler divergence from $f_0(\cdot|y)$ to $f(\cdot|y)$ is given by

$$K(f(\cdot|y), f_0(\cdot|y)) = \int \log\left(\frac{f_0(z|y)}{f(z|y)}\right) f_0(z|y)\mathrm{d}\nu(z). \tag{15}$$

When working with i.i.d. observations, an exponential bound can be found for the denominator I_n of the posterior $\Pi^n(A)$, given a condition on the prior known as the Kullback-Leibler property, which states that every Kullback-Leibler neighbourhood of the true density must have a positive prior probability. As is the case with the Hellinger distance, an adequate generalization of the semimetric must be found, to remove the random element x from (15). This time, it is convenient to define a semimetric on $\mathcal{F}$ by integration, exploiting the ergodicity of the process.

The integrated Kullback-Leibler divergence between f_0 and f is given by

$$K(f, f_0) = \int K(f(\cdot|y), f_0(\cdot|y))\mathrm{d}\nu_0(y).$$

In particular, if the stationary density g_0 is well defined, then

$$K(f, f_0) = \mathbb{E}_0\left[K(f(\cdot|y), f_0(\cdot|y))\right] = \int K(f(\cdot|y), f_0(\cdot|y))g_0(y)\mathrm{d}\nu(y).$$

Lemma 3.3. *Assume the prior* Π *has the Kullback-Leibler property at* f_0*, that is*

$$\Pi\Big(\{f : K(f, f_0) < \epsilon\}\Big) > 0 \quad \textit{for all} \quad \epsilon > 0.$$

Then for every $c > 0$ *and sufficiently large* n

$$I_n > \exp(-nc) \quad a.s.$$

The proof follows from Fatou's lemma and the law of large numbers for ergodic Markov processes [see Ghosal and Tang, 2006, Tang and Ghosal, 2007].

3.4 Posterior consistency

We now have everything we need to present our main result.

Theorem 3.1. *Let* A_ϵ *be a set of transition densities* d*-bounded away from* f_0*,*

$$A_\epsilon = \Big\{ f \in \mathcal{F} : d(f, f_0) > \epsilon \Big\}$$

with d *defined by (4). Assume* Π *has the Kullback-Leibler property and*

$$\sum_{j=1}^{\infty} \sqrt{\Pi(A_j)} < \infty, \tag{16}$$

where or $\{A_j\}_{j\geq 1}$ *is a countable cover for* A_ϵ *of* H_s*-size* $\delta < \epsilon/\sqrt{2}$*. Then*

$$\Pi^n(A_\epsilon) \to 0 \quad a.s.$$

Proof. Let $\{A_j\}$ be the cover satisfying condition (16), and denote $L_{nj} = L_{nA_j}$ for simplicity. Then

$$\begin{aligned}\Pi^n(A_\epsilon) &\leq \sum_{j=1}^{\infty} \Pi^n(A_j) \leq \sum_{j=1}^{\infty} \sqrt{\Pi^n(A_j)} \\ &= \sum_{j=1}^{\infty} \sqrt{L_{nj}/I_n} = I_n^{1/2} \sum_{j=1}^{\infty} \sqrt{L_{nj}}.\end{aligned}$$

Applying Lemmas 3.2 and 3.3, we have $\Pi^n(A_\epsilon) \leq \exp\{-nb\}/\exp\{-nc\}$ for every $c > 0$ and $b < -\log(1 + (\epsilon - \sqrt{2}\delta)^2)/2$. Therefore, $\Pi^n(A_\epsilon) \to 0$ as $n \to \infty$ exponentially fast a.s. □

Notice that, if condition (10) holds, then the convergence in theorem 3.1 implies strong consistency. Once again, if the infimum in the definition of d is reached at some $x^* \in \mathcal{X}$, the strong consistency follows from the Theorem, even in the absence of condition (10) by substituting the H_s-cover with a d-cover.

4 Illustrations

Here we present some examples. The first one, concerning a simple model, illustrates the features of our result. The second, includes a wide family of time series models found in the literature. The third shows how, under some conditions, our consistency result can be applied even when an analytic expression for the transition density is not available, as is the case, for example, when dealing with some discretely observed diffusions [see e.g. Beskos et al., 2006, 2009].

4.1 Normal Autoregressive Model

We recall once more the simple parametric model discussed before, with transition density given by $f_\theta(X_{n+1}|X_n) = N(\cdot|\theta X_n, 1)$, corresponding to the normal autoregressive AR(1) model,

$$X_{n+1} = \theta X_n + \epsilon_n; \quad \epsilon_n \overset{iid}{\sim} N(\cdot|0,1),$$

which is known to be stationary only for $|\theta| \in (0,1)$. This is one of the simplest and most common time series models, yet there is no straightforward result in the Bayesian literature, guaranteeing strong consistency for the transition densities that can be applied when the stationarity assumption is not satisfied. In particular, Ghosal and van der Vaart [2007] provide results for consistency only when the process is stationary, while Ghosal and Tang's 2006 results guarantee strong consistency only when Θ is compact. Other ideas, based on the construction of sieves and uniformly consistent tests for adequate metrics, would require a careful study for each proposed prior.

On the other hand, the separability of $\mathbb{R}$ makes it straightforward to check if a prior Π on Θ satisfies the conditions of Theorem 3.1 and, as mentioned before, the operator d defines a metric on the space $\mathcal{F}$ of transition densities for this particular model, even if $\Theta = \mathbb{R}$.

Recall that

$$d(f_{\theta_1}, f_{\theta_2}) = 1 - \frac{2}{\sqrt{4 + (\theta_1 - \theta_2)^2}}.$$

Therefore, for every $0 < \delta < 1$,

$$|\theta_1 - \theta_2| < \tilde{\delta} = 2\sqrt{(1-\delta^2)^{-2} - 1} \quad \Rightarrow \quad d(f_{\theta_1}, f_{\theta_2}) < \delta,$$

so a countable d-cover of size δ for $\mathcal{F}$ can be defined in terms of a cover of size $\tilde{\delta}$ for $\mathbb{R}$ in the

following way:

$$B_j = \left(j\tilde{\delta}, (j+1)\tilde{\delta}\right) \subset \mathbb{R}; \quad A_j = \{f_\theta : \theta \in B_j\}; \quad j \in \mathbb{Z}.$$

By symmetry, in order to prove

$$\sum_{j=-\infty}^{\infty} \sqrt{\Pi(f_\theta \in A_j)} = \sum_{j=-\infty}^{\infty} \sqrt{\Pi(\theta \in B_j)} < \infty$$

it is enough to show

$$\sum_{j=0}^{\infty} \sqrt{\Pi(\theta \in B_j)} < \infty,$$

which can be easily verified for any particular choice of Π.

In a more general setting, we could consider a transition density of the form

$$f(X_{n+1}|X_n) = N\left(X_{n+1}|\theta(X_n), \sigma^2\right)$$

for some function $\theta : \mathcal{X} \to \mathcal{X}$. If we consider functions such that $|\theta| < M$ for some finite M, then there exist a density g and a constant $\beta > 0$ such that $f(y|x) > \beta g(y)$ for all $x \in \mathcal{X}$ and all $f \in \mathcal{F}$. By Corollary 2.1, d defines a strong neighbourhood around f_0 and Theorem 3.1 can be applied to verify strong consistency of the model, once the structure of the $\theta \in \Theta$ and the prior have been specified.

4.2 Nonparametric Mixture Model

Consider a time series model with transition densities given by

$$f(X_{n+1}|X_n) = \int_\Theta K(X_{n+1}|X_n, \theta) \mathrm{d}P_{X_n}(\theta),$$

where $K(\cdot|x,\theta)$ is a parametric density on $\mathcal{X}$ for every $\theta \in \Theta$, and $x \in \mathcal{X}$; and $\{P_x\}_{x\in\mathcal{X}}$ is a family of mixing probability measures on Θ. In the most general case, the $\{P_x\}$ may be non parametric and the prior Π placed over them is usually some dependent measure valued process. Models of this type are becoming common in the literature; some of them can be found in e.g. Mena and Walker [2005, 2007] and Martínez-Ovando and Walker [2011]. The family $\mathcal{F}$ of transition densities of interest in this type of models is defined by the support of the prior, Π.

Assume a sequence of observations $\{X_n\}_{n\geq 0}$ is generated from a time homogeneous Markov process with transition density $f_0 \in \mathcal{F}$. In other words, there is some probability measure P_0 such that, for every n,

$$f_0(X_{n+1}|X_n) = \int_\Theta K(X_{n+1}|X_n,\theta)\mathrm{d}P_0(\theta|X_n).$$

Here are some noteworthy special cases:

(i) If, for example, f_0 is a finite normal mixture such as

$$f_0(y|x) = \sum_{j=1}^{M} w_j(x)N(y|\mu_j,\sigma_j^2)$$

for some finite M, then the condition of Corollary 2.1 is easily satisfied.

A particular case is obtained when

$$w_j(x) = \frac{\exp(\beta_j x)}{\sum_{j=1}^{M}\exp(\beta_j x)}.$$

By introducing adequate normalizing constants, we can consider truncated normal kernels, to make $\mathcal{X}$ equal to some bounded interval. In this case d is bounded below

by an integrated metric (see inequality 11). We can define the d-cover, i.e., the sets (A_k) for which $d(f, f') < \delta$ for any $f, f' \in A_k$, by ensuring that

$$\max_{j=1,\ldots,M} \{|\beta_j - \beta'_j|, |\mu_j - \mu'_j|, |1 - \sigma_j/\sigma'_j|\}$$

is suitably small. It is now a standard exercise to verify that

$$\sum_k \Pi(A_k)^{1/2} < \infty,$$

by splitting the range of each parameter and ensuring that the prior condition holds for each of them. For example, for μ_1, we split $(-\infty, +\infty)$ into a countable set of intervals $A_{k(\mu_1)}$ for which

$$\sum_k \Pi(A_{k(\mu_1)})^{1/2} < \infty.$$

(ii) Mena and Walker [2005] consider a transition density of the form

$$f_0(y|x) = \int K(y|\theta) \mathrm{d}P(\theta|x),$$

where $K(\cdot|\theta)$ is some (non degenerate) parametric density. For this type of transition density, Corollary 2.1 applies whenever $K(y|\theta) > \beta g(y)$ for some $\beta > 0$ and a density function g with full support on $\mathcal{X}$. For example, if $K(y|\theta) = \theta\, e^{-y\theta}$ then we may obtain $K(y|\theta) > \beta\, \phi\, e^{-y\phi}$ by truncating $\beta\phi < \theta < \phi$, and this is not overly restrictive since we can take β arbitrarily close to 0 and ϕ arbitrarily large.

(iii) In the more general setting, if the kernels satisfy

$$\inf_x K(y|x, \theta) > \beta g(y), \quad \forall y \in \mathcal{X}, \theta \in \Theta,$$

then, for every $y, x \in \mathcal{X}$,

$$f(y|x) > \beta \int_{\Theta} g(y) \mathrm{d}P(\theta|x) = \beta g(y).$$

If additionally, $f(\cdot|x)$ is continuous on x, then the conditions of Corollary 2.1 are once again satisfied, and the operator d can be used to define strong neighborhoods around f_0.

In all of these cases, strong consistency follows for any prior Π for which the conditions of Theorem 3.1 hold. The verification of consistency is therefore reduced to checking conditions on the prior.

4.3 Compact Support Model

If the state space $\mathcal{X}$ is compact, then for every f there is some $x_f \in \mathcal{X}$ which minimizes the Hellinger distance between $\bar{f}$ and $\bar{f}_0$. In this case,

$$d(f, f_0) = d_{x_f}(f, f_0)$$

defines a distance on $\mathcal{F}$ which induces a strong topology. Hence, strong consistency can be verified through Theorem 3.1.

An interesting particular case occurs when $\mathcal{X}$ is finite with cardinality N. Then we deal with a transition matrix with strictly positive entries, i.e. $f_0(y|x)$ is represented as $P_0 = (p_{i.j})_{i,j=1}^{N}$.

5 Discussion

Our main result, stated in Theorem 3.1, gives sufficient conditions for posterior consistency in the estimation of transition densities for ergodic (but not necessarily stationary) Markov processes. The key for this result lies in the definition of the neighbourhood around f_0. If the Hellinger distance $H^2\big(f(\cdot|y), f_0(\cdot|y)\big)$ is simply minimized over y, the resulting quantity defines weak neighbourhoods. An integrated Hellinger distance is a reasonable alternative, however it results in the additional condition of compactness of $\mathcal{F}$ in the supreme Hellinger distance.

We solve the issue of finding an appropriate distance by noticing that for every transition density $f \in \mathcal{F}$, and each point $x \in \mathcal{X}$, a bivariate density on $\mathcal{X} \times \mathcal{X}$ is uniquely defined by $\bar{f}(z, y|x) = f(z|y)f(y|x)$. We define the quantity d by looking at the distance between the bivariate densities they define, and subsequently minimizing over x. In this way d, under suitable conditions, defines a semimetric on $\mathcal{F}$which can be used to define strong neighbourhoods around the true transition f_0.Thus, a sufficient condition for consistency is found, involving the finiteness of sums of square roots of prior probabilities, which requires only the separability of the space $\mathcal{F}$ with respect to the supreme Hellinger distance.

It is is an interesting observation that for some families of transition densities, d does define a distance on the complete space $\mathcal{F}$. In this case, separability with respect to H_s can be substituted by separability with respect to d. This solves the problem of consistency for the simple example $f_\theta(\cdot|y) = N(\cdot|\theta y, 1)$ mentioned throughout the paper.

Future work should involve the study of posterior consistency for models with higher order Markov dependency, by extending the definition of d to a space of transition densities with an arbitrary number of dependent variables.

An important feature of our result is that it can be applied to some discretely observed

continuous processes, for which the transition density often does not have an analytic form. We believe this to be a relevant contribution, since consistency results had been found in the Bayesian setting only for continuously observed diffusions [van der Meulen et al., 2006]; while results for discrete observations are only available for maximum likelihood estimators [Beskos et al., 2009] or for approximate likelihood functions [Barndorff-Nielsen and Sorensen, 1994, Bibby and Sorensen, 1995, Kelly et al., 2004].

Acknowledgements

The authors would like to thank the Editor and a reviewer for their comments on a previous version of the paper.

References

O. E. Barndorff-Nielsen and M. Sorensen. A review of some aspects of asymptotic likelihood theory for stochastic processes. *International Statistical Review*, 62(1):133–165, 1994.

A. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2):536–561, 1999.

A. Beskos, O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society*, 68(3):333–382, June 2006.

A. Beskos, O. Papaspiliopoulos, and G. Roberts. Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *The Annals of Statistics*, 37(1):223–245, 2009.

B. M. Bibby and M. Sorensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1/2):17–39, 1995.

T. Choi and M. J. Schervish. On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis*, 98:1969–1987, 2007.

S. Ghosal and A. Roy. Consistency of Gaussian process prior for nonparametric binary regression. *The Annals of Statistics*, 34(5):2413–2429, 2006.

S. Ghosal and Y. Tang. Bayesian consistency for Markov processes. *The Indian Journal of Statistics*, 68(2):227–239, May 2006.

S. Ghosal and A. van der Vaart. Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–223, 2007.

S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi. Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 27(1):143–158, 1999.

L. Kelly, E. Platen, and M. Sorensen. Estimation for discretely observed diffusions using transform functions. *Journal of Applied Probability*, 41:99–118, 2004.

J. C. Martínez-Ovando and S. G. Walker. Time-series modelling, stationarity and Bayesian nonparametric methods. Technical report, Banco de México, 2011.

R. Mena and S. G. Walker. On the stationary version of the generalized hyperbolic ARCH model. *Annals of the Institute of Statistical Mathematics*, 59(2):325–348, 2007.

R. H. Mena and S. G. Walker. Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis*, 26(6):789–805, 2005.

L. Schwartz. On Bayes procedures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 4: 10–26, 1965.

Y. Tang and S. Ghosal. Posterior consistency of Dirichlet mixtures for estimating a transition density. *Journal of Statistical Planning and Inference*, 137:1711–1726, 2007.

F. H. van der Meulen, A. W. van der Vaart, and J. H. van Zanten. Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli*, (12):863–888, 2006.

S. Walker. On sufficient conditions for Bayesian consistency. *Biometrika*, 90(2):482–488, 2003.

S. Walker. New approaches to Bayesian consistency. *The Annals of Statistics*, 32(5):2028–2043, 2004.

Isadora Antoniano-Villalobos.
Department of Decision Sciences
Bocconi University
Milano, MI, 20136
Italy
E-mail: isadora.antoniano@unibocconi.it

Stephen G. Walker
School of Mathematics, Statistics and Actuarial Science
University of Kent Canterbury, Kent, CT27NF
United Kingdom
E-mail:S.G.Walker@kent.ac.uk