Sankhyā : The Indian Journal of Statistics
2006, Volume 68, Part 2, pp. 227-239

Bayesian Consistency for Markov Processes

Subhashis Ghosal
North Carolina State University, U.S.A.
Yongqiang Tang
SUNY Brooklyn, U.S.A.

Abstract

Recently, Walker (2004) developed a martingale based technique for posterior consistency giving a useful alternative to the classical technique based on hypothesis testing for i.i.d. data. In this article, we extend some of his results to ergodic Markov processes.

AMS (2000) *subject classification.* Primary 62G07, 62G20.
Keywords and phrases. Markov process, martingale, posterior consistency, transition density.

1 Introduction

Let $\{X_n; n = 0, 1, 2, \ldots\}$ be an ergodic Markov process on a state space S with transition density $f(y|x)$ and stationary density $\pi(x)$ with respect to some σ-finite measure ν. The process $\{X_n\}$ need not be stationary. Let f belong to a class $\mathcal{F}$ of transition densities, which need not be a parametric family. Let Π be a prior distribution for f. Let $\Pi(\cdot|X_0, \ldots, X_n)$ stand for the posterior distribution of f given $X_0, \ldots, X_n$. In our set up, we assume that either X_0 is fixed or it has a known initial distribution. By the Bayes theorem, the posterior is given by the relation

$$\Pi(B|X_0, \ldots, X_n) = \frac{\int_B \prod_{i=1}^n f(X_i|X_{i-1}) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^n f(X_i|X_{i-1}) d\Pi(f)}. \tag{1.1}$$

Note that the predictive density of X_{n+1} at y given $X_0, \ldots, X_n$ is obtained as $\mathrm{E}(f(y|X_n)|X_0, \ldots, X_n)$, and hence accurate estimation of f is important. Let f_0 stand for the true transition density and π_0 the true stationary density. All the probability statements are given relative to the true distribution.

Suppose that a topology on $\mathcal{F}$ has been specified. Then posterior distribution is said to be consistent at f_0 if for every neighborhood U of f_0, we have that $\Pi(U^c|X_0, \ldots, X_n) \to 0$ a.s. For i.i.d. observations, a classical theory of posterior consistency has been developed by Schwartz (1965), Barron et al. (1999) and Ghosal et al. (1999) based on the assumptions that the prior assigns positive probabilities to each neghborhood of the true density defined by the Kullback-Leibler divergence number and there exists a sequence of uniformly consistent test for testing $f = f_0$ against $f \in U^c \cap \mathcal{F}_n$, where $\mathcal{F}_n \subset \mathcal{F}$ are sieves with $\Pi(\mathcal{F}_n^c)$ exponentially small. The theory of consistency has been recently extended to independent non-identically distributed case by Amewou-Atisso et al. (2003) and Choudhuri et al. (2004), and to Markov process by Tang and Ghosal (2007). Rates of convergence for general dependent cases have been investigated by Ghosal and van der Vaart (2007). However, the construction of appropriate sieves and uniformly consistent tests on them is often difficult. It is therefore of interest to explore alternative sufficient conditions for consistency. Walker (2003, 2004) recently formulated a martingale based approach for the study of consistency for i.i.d. observations. In this article, we extend some of his results to ergodic Markov processes.

It may be observed that consistency results easily generalize to higher order processes, where the conditional distribution given history depends only on the k most recent values. All (non-linear) autoregressive time series satisfy the condition. If X_n is a process satisfying the condition, then it is well known that $Y_n = (X_n, \ldots, X_{n-k+1})$ is a Markov process with state space S^k, where S is the state space of X_n. Marginalizing to the first component, conclusions can be obtained about the conditional distribution of X_n given $(X_{n-1}, \ldots, X_{n-k+1})$.

2 Main Results

Let

$$R_n(f) = \prod_{i=1}^{n} \frac{f(X_i|X_{i-1})}{f_0(X_i|X_{i-1})}, \quad n \geq 1, \tag{2.1}$$

stand for the likelihood ratio and

$$I_n = \int_{\mathcal{F}} R_n(f) d\Pi(f) \tag{2.2}$$

stand for the integrated likelihood ratio. Set $R_0(f) = 1 = I_0$. Let

$$K(f_0, \epsilon) = \left\{ f : \int \int \pi_0(x) f_0(y|x) \log(f(y|x)/f_0(y|x)) d\nu(y)\, d\nu(x) < \epsilon \right\} \tag{2.3}$$

stand for the Kullback-Leibler neighborhood of f_0. We say that f_0 is in the *Kullback-Leibler support* of Π or the *Kullback-Leibler property* holds at f_0 if $\Pi(K(f_0, \epsilon)) > 0$ for all $\epsilon > 0$.

Note that

$$\Pi(U^c|X_0, \ldots, X_n) = I_n^{-1} \int_{U^c} R_n(f) d\Pi(f). \tag{2.4}$$

Assume that the Kullback-Leibler property holds at f_0. Similar to the i.i.d. case, it follows from the strong law of large numbers for ergodic Markov processes and Fatou's lemma that

$$\text{for all } \beta > 0, \qquad e^{n\beta} I_n \to \infty \text{ a.s.}; \tag{2.5}$$

see (2.1) of Tang and Ghosal (2007). In other words,

$$\text{for all } c > 0, \qquad I_n > e^{-nc} \text{ for all sufficiently large } n \text{ a.s.}, \tag{2.6}$$

or

$$\liminf_{n\to\infty} n^{-1} \log I_n \geq 0 \text{ a.s.} \tag{2.7}$$

In order to show that $\Pi(A|X_0, \ldots, X_n) \to 0$ a.s. for some subset A of $\mathcal{F}$, it therefore suffices to show that

$$L_n = L_{n,A} = \int_A R_n(f) d\Pi(f) \leq e^{-nc} \tag{2.8}$$

for all sufficiently large n a.s. for some $c > 0$. Following Walker's (2004) approach, we give a sufficient condition for (2.8).

For a given set A, let

$$f_{n,A}(y|x) = \frac{\int_A f(y|x) \prod_{i=1}^n f(X_i|X_{i-1}) d\Pi(f)}{\int_A \prod_{i=1}^n f(X_i|X_{i-1}) d\Pi(f)} = \frac{\int_A f(y|x) R_n(f) d\Pi(f)}{\int_A R_n(f) d\Pi(f)}. \tag{2.9}$$

Note that, $f_{n,A}$ is also the posterior expectation of f given that f belongs to A, that is, the posterior expectation of f when the prior is Π_A — the prior

Π restricted to A. In particular, if A is closed and convex, $f_{n,A} \in A$. Let $\mathcal{A}_n$ stand for the σ-field generated by the observations $X_0, \ldots, X_n$.

Similar to Walker (2004), consistency results are obtained from the basic identity

$$\frac{L_{n+1}}{L_n} = \frac{\int_A \frac{f(X_{n+1}|X_n)}{f_0(X_{n+1}|X_n)} R_n(f) d\Pi(f)}{\int_A R_n(f) d\Pi(f)} = \frac{f_{n,A}(X_{n+1}|X_n)}{f_0(X_{n+1}|X_n)}, \tag{2.10}$$

for $n = 0, 1, \ldots$.

The following theorem generalizes the conclusions of Lemma 1 and Theorems 1, 2 and 2* of Walker (2004) from i.i.d. observations to Markov processes.

THEOREM 2.1. *Let $T : [0, \infty) \to \mathbb{R}$ be a strictly increasing function such that $T(x) \geq b \log x$ for some constant $b > 0$ and*

$$\mathrm{E}(T(f(X_{n+1}|X_n)/f_0(X_{n+1}|X_n))|\mathcal{A}_n) = -d(f_0(\cdot|X_n), f(\cdot|X_n)|X_n), \tag{2.11}$$

where $d(f_1(\cdot|x), f_2(\cdot|x)|x)$ is a distance-like measure on the conditional densities given x. Assume that the Kullback-Leibler property holds at f_0 and

$$\sum_{n=1}^{\infty} n^{-2} \mathrm{var}(T(L_{n+1}/L_n)) < \infty. \tag{2.12}$$

Then

$$N^{-1} \sum_{n=0}^{N-1} \{T(L_{n+1}/L_n) + d(f_0(\cdot|X_n), f_{n,A}(\cdot|X_n)|X_n)\} \to 0 \ a.s. \tag{2.13}$$

Consequently,

$$\limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} T(L_{n+1}/L_n) < 0 \ a.s. \tag{2.14}$$

if and only if

$$\liminf_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} d(f_0(\cdot|X_n), f_{n,A}(\cdot|X_n)|X_n) > 0 \ a.s. \tag{2.15}$$

In particular, $\Pi(A|X_0, \ldots, X_n) \to 0$ a.s. if (2.15) holds.

REMARK 2.1. By a distance-like measure d on $\mathcal{F}$, we mean a non-negative function on $\mathcal{F} \times \mathcal{F}$ such that $d(f_0(\cdot|x), f(\cdot|x)|x) = 0$ if and only if $f(\cdot|x) = f_0(\cdot|x)$. The common choice $T(x) = 2(\sqrt{x} - 1)$ leads to

$$d(f_0(\cdot|x), f(\cdot|x)|x) = \int (\sqrt{f_0(y|x)} - \sqrt{f(y|x)})^2 d\nu(y), \tag{2.16}$$

the squared Hellinger distance between conditional densities, while $T(x) = \log x$ leads to

$$d(f_0(\cdot|x), f(\cdot|x)|x) = \int f_0(y|x) \log(f_0(y|x)/f(y|x)) d\nu(y), \tag{2.17}$$

the Kullback-Leibler divergence between the conditional densities.

More generally, for $T(x) = (x^\alpha - 1)/\alpha$, we have $d(f_0(\cdot|x), f(\cdot|x)|x) = (1 - \int f_0^{1-\alpha}(y|x) f^\alpha(y|x) d\nu(y))/\alpha$. The choice $\alpha = -1$ yields $T(x) = 1 - 1/x$, and the obtained distance is the chi-squared distance $\int (f_0^2(y|x)/f(y|x)) d\nu(x) - 1$. However, this T does not dominate $b \log x$ for any $b > 0$.

REMARK 2.2. If $T(x) = \sqrt{x} - 1$, the condition (2.12) is automatically satisfied because by (2.10), for $n = 0, 1, \ldots,$

$$\operatorname{var}(T(L_{n+1}/L_n)) \leq \mathrm{E}(L_{n+1}/L_n) = \int f_{n,A}(y|X_n) d\nu(y) = 1.$$

REMARK 2.3. The Kullback-Leibler property is a classical condition appearing in Schwartz's (1965) theory of consistency for i.i.d. observations. Priors such as Dirichlet mixtures or Polya trees for densities satisfy the Kullback-Leibler property under appropriate conditions in the i.i.d. case. In the context of Markov processes, the condition has been verified in Tang and Ghosal (2007) for a specific type of Dirichlet mixture of normal prior for transition densities.

PROOF OF THEOREM 2.1. By (2.10),

$$\begin{aligned} \mathrm{E}[T(L_{n+1}/L_n)|\mathcal{A}_n] &= \mathrm{E}[T(f_{n,A}(X_{n+1}|X_n)/f_0(X_{n+1}|X_n))|\mathcal{A}_n] \\ &= -d(f_0(\cdot|X_n), f_{n,A}(\cdot|X_n)|X_n) \end{aligned} \tag{2.18}$$

and therefore

$$M_N = \sum_{n=0}^{N-1} [T(L_{n+1}/L_n) + d(f_0(\cdot|X_n), f_{n,A}(\cdot|X_n)|X_n)] \tag{2.19}$$

is an $\{\mathcal{A}_N\}$-martingale. Note that

$$\sum_{n=1}^{\infty} \frac{\mathrm{E}((M_{n+1}-M_n)^2|\mathcal{A}_n)}{n^2} = \sum_{n=1}^{\infty} \frac{\mathrm{var}(T(L_{n+1}/L_n)|\mathcal{A}_n)}{n^2} < \infty \tag{2.20}$$

a.s., because by (2.12),

$$\begin{aligned} \mathrm{E}\left[\sum_{n=1}^{\infty} n^{-2}\mathrm{var}(T(L_{n+1}/L_n)|\mathcal{A}_n)\right] &= \sum_{n=1}^{\infty} n^{-2}\mathrm{E}[\mathrm{var}(T(L_{n+1}/L_n)|\mathcal{A}_n)] \\ &\leq \sum_{n=1}^{\infty} n^{-2}\mathrm{var}(T(L_{n+1}/L_n)) < \infty. \end{aligned}$$

Hence, by a result on convergence of a martingale (see, for instance, Shiryayev, 1984, page 487), we have that $M_N/N \to 0$ a.s., which is a re-statement of (2.13). Clearly, the equivalence of (2.14) and (2.15) is now immediate.

Now, if (2.15) holds, we have that

$$\begin{aligned} 0 > \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} T(L_{n+1}/L_n) &\geq b \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} \log(L_{n+1}/L_n) \\ &\geq b \limsup_{N\to\infty} N^{-1} \log L_N, \end{aligned}$$

which implies that, for some $c > 0$, we have $L_n = \int_A R_n(f)d\Pi(f) < e^{-nc}$ for all sufficiently large n a.s. As discussed in the beginning of this section, this implies $\Pi(A|X_0, \ldots, X_n) \to 0$ a.s. □

Corollary 2.1 (Predictive consistency). *Let*

$$f_n(y|X_n) = E(f(y|X_n)|X_0, \ldots, X_n) \tag{2.21}$$

stand for the predictive density of X_{n+1} at y given $X_0, \ldots, X_n$. Under the conditions of Theorem 2.1, *f_n converges to the true transition density f_0 in the sense that*

$$N^{-1} \sum_{n=0}^{N-1} d(f_0(\cdot|X_n), f_n(\cdot|X_n)|X_n) \to 0 \ \ a.s. \tag{2.22}$$

In particular, when d is the squared Hellinger distance, consistency of the predictive density in this sense holds only under the Kullback-Leibler property.

Proof. Note that $f_n = f_{n,A}$ for $A = \mathcal{F}$, and $L_n = I_n = \int_{\mathcal{F}} R_n(f)d\Pi(f)$. By (2.13) and non-negativity of d, we have

$$\limsup_{N\to\infty} N^{-1}\sum_{n=0}^{N-1} T(I_{n+1}/I_n) = -\liminf_{N\to\infty} N^{-1}\sum_{n=0}^{N-1} d(f_0(\cdot|X_n), f_n(\cdot|X_n)|X_n) \\ \leq 0 \text{ a.s.}$$

However, as $T(x) \geq b\log x$,

$$\liminf_{N\to\infty} N^{-1}\sum_{n=0}^{N-1} T(I_{n+1}/I_n) \geq b\liminf_{N\to\infty} N^{-1}\sum_{n=0}^{N-1}\log(I_{n+1}/I_n) \quad (2.23)$$

$$= b\liminf_{N\to\infty} N^{-1}\log I_N \geq 0 \quad (2.24)$$

by (2.7). Thus $N^{-1}\sum_{n=0}^{N-1} T(I_{n+1}/I_n) \to 0$ a.s., and hence (2.22) follows from (2.13). □

In some situtaions, it may be reasonable to consider consistency relative to a system of neighborhoods which depends on the observations. Note that the sense of convergence implied by (2.22) is also based on a sample dependent neighborhood system. In the following result, we allow the set A in Theorem 2.1 to depend on n as well as on the sample $X_0, \dots, X_n$. The result is new even in the context of i.i.d. data.

Let $h^2(f_1(\cdot|x), f_2(\cdot|x)|x) = \int(\sqrt{f_1(y|x)} - \sqrt{f_2(y|x)})^2 d\nu(y)$ stand for the conditional squared Hellinger distance between transition densities.

Theorem 2.2. *Let the Kullback-Leibler property hold at f_0. Assume that $A_n \supset A_{n+1}$ and $I\{f \in A_n\}$ is an $\mathcal{A}_n$-measurable random variable for all n. If*

$$\liminf_{N\to\infty} N^{-1}\sum_{n=0}^{N-1} h^2(f_0(\cdot|X_n), f_{n,A_n}(\cdot|X_n)|X_n) > 0 \text{ a.s.}, \quad (2.25)$$

then $\Pi(A_n|X_0, \dots, X_n) \to 0$ a.s.

Proof. Put $L_0 = 1$ and $L_n = \int_{A_n} R_n(f)d\Pi(f)$, $n \geq 1$. As $A_{n+1} \subset A_n$, it follows that

$$\frac{L_{n+1}}{L_n} \leq \frac{\int_{A_n} R_{n+1}(f)d\Pi(f)}{\int_{A_n} R_n(f)d\Pi(f)} = \frac{f_{n,A_n}(X_{n+1}|X_n)}{f_0(X_{n+1}|X_n)}. \quad (2.26)$$

Hence, with $T(x) = 2(\sqrt{x} - 1)$, we have

$$\mathrm{E}[T(L_{n+1}/L_n)|\mathcal{A}_n] \leq -h^2(f_0(\cdot|X_n), f_{n,A_n}(\cdot|X_n)|X_n) \tag{2.27}$$

and that

$$M_N = \sum_{n=0}^{N-1} \left[T\left(\frac{f_{n,A_n}(X_{n+1}|X_n)}{f_0(X_{n+1}|X_n)}\right) + h^2(f_0(\cdot|X_n), f_{n,A_n}(\cdot|X_n)|X_n) \right] \tag{2.28}$$

is an $\mathcal{A}_N$-martingale. Also, $M_N/N \to 0$ a.s. by the arguments given in Theorem 2.1 and Remark 2.2. As in Theorem 2.1, this implies that

$$\limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} T(L_{n+1}/L_n) \leq \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} T\left(\frac{f_{n,A_n}(X_{n+1}|X_n)}{f_0(X_{n+1}|X_n)}\right) < 0$$

whenever (2.25) holds. □

In applications to consistency, where A is typically the complement of a neighborhood, it is often difficult to satisfy (2.15) because of the lack of convexity in A. The result can be modified easily to the case when A is the union of finitely many convex sets. However, unless compactness holds, the complement of a neighborhood may not be expressible as a finite union of small balls.

Walker (2004) considered a countable covering with an additional restriction on the prior probabilities of the sets in the cover. The following result gives an analog of Theorem 4 of Walker (2004) for Markov processes. In the i.i.d. case, if A is the complement of the Hellinger ϵ-neighborhood of f_0 and A_j's are Hellinger δ-balls centered in A with $\delta < \epsilon$, then (2.29) below holds with $\gamma = \epsilon - \delta$. Thus our result is a generalization of Walker's (2004) Theorem 4.

THEOREM 2.3. *Let the Kullback-Leibler property hold. Let A be a set of transition densities such that $A \subset \cup_{j=1}^{\infty} A_j$, where $\sum_{j=1}^{\infty} \sqrt{\Pi(A_j)} < \infty$ and for some $\gamma > 0$,*

$$\inf_x \inf_j h^2(f_0(\cdot|x), f_{n,A_j}(\cdot|x)|x) \geq \gamma. \tag{2.29}$$

Then $\Pi(A|X_0, \ldots, X_n) \to 0$ a.s.

REMARK 2.4. When x ranges over a compact set, $\inf_x h^2(f_0(\cdot|x), f(\cdot|x)|x) \geq \gamma$ may often follow by continuity in x. If the transition densities are obtained by non-linear autoregressive model $f(y|x) = g(y - \psi(x))$ and $f_0(y|x) = g_0(y - \psi_0(x))$, then $h^2(f_0(\cdot|x), f(\cdot|x)|x) = h^2(g_0(\cdot), g(\cdot - \psi(x) + \psi_0(x)))$ which remains bounded away from zero if g_0 is not a location shift of g.

Proof of Theorem 2.3.

As in the proof of Theorem 4 of Walker (2004), we have

$$\begin{aligned}\Pi(A|X_0,\ldots,X_n) &\leq \sum_{j=1}^{\infty}\sqrt{\Pi(A_j|X_0,\ldots,X_n)}\\ &= I_n^{-1/2}\sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}\left(\int_{A_j}R_n(f)d\Pi_{A_j}(f)\right)^{1/2}\\ &= I_n^{-1/2}\sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}L_{n,A_j}^{1/2},\end{aligned}$$

so it suffices to show that $\sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}L_{n,A_j}^{1/2}$ tends to 0 exponentially fast a.s. To this end, note that by an application of (2.18) with $T(x)=\sqrt{x}$, we have

$$\begin{aligned}&\mathrm{E}\left[\sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}L_{n,A_j}^{1/2}\right]\\ &= \sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}\mathrm{E}[\mathrm{E}(L_{n,A_j}^{1/2}|\mathcal{A}_{n-1})]\\ &\leq \sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}\mathrm{E}[(1-h^2(f_0(\cdot|X_{n-1}),f_{n-1,A_j}(\cdot|X_{n-1})|X_{n-1}))L_{n-1,A_j}^{1/2}]\\ &\leq \sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}\mathrm{E}[(1-\gamma)L_{n-1,A_j}^{1/2}] \;\leq\; \sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}(1-\gamma)^n,\end{aligned}$$

which concludes the proof. □

3 Applications

Example 3.1. Let the π_0-integrated weak topology be defined by the basic open sets which are finite intersections of sets of the form

$$\left\{f:\left|\int\int\psi(y)f(y|x)d\nu(y)\pi_0(x)d\nu(x)\right.\right.$$
$$\left.\left.-\int\int\psi(y)f_0(y|x)d\nu(y)\pi_0(x)d\nu(x)\right|<\epsilon\right\},$$

where ψ is a bounded continuous function and $\epsilon > 0$. Let $\mathcal{F}$ be compact with respect to the sup-Hellinger metric defined by

$$d_{s,H}(f_1, f_2) = \sup_{x \in \mathbb{R}} \left[\int (\sqrt{f_1(y|x)} - \sqrt{f_2(y|x)})^2 d\nu(y) \right]^{1/2}, \qquad (3.1)$$

and let the Kullback-Leibler property hold at f_0. Then the posterior is consistent with respect to the π_0-integrated weak topology.

To show consistency, clearly it is enough to show that $\Pi(A|X_0, \ldots, X_n) \to 0$ a.s., where

$$A = \left\{ f : \int \int \psi(y) f(y|x) d\nu(y) \pi_0(x) d\nu(x) \right. \\ \left. \geq \int \int \psi(y) f_0(y|x) d\nu(y) \pi_0(x) d\nu(x) + \epsilon \right\}.$$

As A is closed and convex, we only need to check that

$$\liminf_{N \to \infty} N^{-1} \sum_{n=0}^{N-1} h^2(f_0(\cdot|X_n), f(\cdot|X_n)|X_n) > 0 \text{ a.s.} \qquad (3.2)$$

uniformly for all $f \in A$.

For a fixed $f \in A$, let $Z(f, x) = h(f_0(\cdot|x), f(\cdot|x)|x)$. We have

$$n^{-1} \sum_{i=1}^{n} Z(f, X_i) \to \int \int (\sqrt{f(y|x)} - \sqrt{f_0(y|x)})^2 d\nu(y) \pi_0(x) d\nu(x) \text{ a.s.} \quad (3.3)$$

by the strong law of large numbers for ergodic Markov processes. Now

$$\int \left| \int \psi(y) f(y|x) d\nu(y) - \int \psi(y) f_0(y|x) d\nu(y) \right| \pi_0(x) d\nu(x) \\ \leq \|\psi\|_\infty \int \int |f(y|x) - f_0(y|x)| d\nu(y) \pi_0(x) d\nu(x) \\ \leq \|\psi\|_\infty \left(\int \int |\sqrt{f(y|x)} - \sqrt{f_0(y|x)}|^2 d\nu(y) \pi_0(x) d\nu(x) \right)^{1/2},$$

so that for every $f \in A$, the right hand side of (3.3) is bounded away from 0. To conclude (3.2), it thus only remains to show that the convergence in (3.3) is uniform.

We shall apply the uniform strong law of large numbers under bracketing given by Theorem 2.4.1 of van der Vaart and Wellner (1996). Note that the argument of the theorem is applicable even though the variables are not i.i.d., because the mechanism only helps to uniformize a function-wise strong law. As $\mathcal{F}$ is compact under the sup-Hellinger metric $d_{s,H}$ defined by (3.1), given any $\epsilon > 0$, there exist finitely many transition densities $f_1, \ldots, f_k \in \mathcal{F}$ such that for any $f \in \mathcal{F}$, there is l, $l = 1, \ldots, k$, with the property that $h(f(\cdot|x), f_l(\cdot|x)|x) \le d_{s,H}(f, f_l) < \epsilon$ for all x. Therefore, as $Z(f, x) \le 2$ for all f,

$$\begin{aligned} |Z(f,x) - Z(f_l,x)| &= |Z^{1/2}(f,x) - Z^{1/2}(f_l,x)||Z^{1/2}(f,x) + Z^{1/2}(f_l,x)| \\ &\le 2\sqrt{2}|h(f(\cdot|x), f_0(\cdot|x)|x) - h(f_l(\cdot|x), f_0(\cdot|x)|x)| \\ &\le 2\sqrt{2}\epsilon \end{aligned}$$

for all x by the traingle inequality. This gives that $\{Z(f_l, x) \pm 2\sqrt{2}\epsilon : l = 1, \ldots, k\}$ is a $L_1(\pi_0)$-bracketing of size $4\sqrt{2}\epsilon$ for $\{Z(f, \cdot) : f \in \mathcal{F}\}$. Thus $\{Z(f, \cdot) : f \in \mathcal{F}\}$ is a Glivenko-Cantelli class of functions, and hence the required uniform convergence in (3.3) holds.

In fact, as $\mathcal{F}$ is compact with respect to $d_{s,H}$ and the topology of $d_{s,H}$ is stronger than the π_0-integrated weak topology, it easily follows that the two topologies coincide on $\mathcal{F}$. Thus the convergence actually holds under $d_{s,H}$.

Tang and Ghosal (2007) showed that the compactness conditions hold for the Dirichlet mixture prior if its dynamic auto-regressive function is bounded and the random effect coefficients lie in a compact set.

EXAMPLE 3.2. The compactness assumption with respect to $d_{s,H}$ in the last example is arguably strong. It is therefore of interest to know whether the assumption could be relaxed to give some similar conclusion. Below we consider a random sequence of neighborhoods to define a convergence.

Let m be arbitrary but fixed. Define $A_N = \cap_{n=1}^N B_n$, where

$$B_n = \left\{ f : n^{-1} \sum_{i=1}^n \int \psi(y) f(y|X_i)\nu(y)dy \right. \\ \left. \ge n^{-1} \sum_{i=1}^n \int \psi(y) f_0(y|X_i)\nu(y)dy + \epsilon \right\}$$

for $n > m$ and the whole space of transition densities for $n \le m$. Clearly, A_N's are decreasing, closed, convex random sets such that $I(f \in A_N)$ is

$\mathcal{A}_N$-measurable. With $\mathbb{P}_n(\cdot)$ the empirical distribution of $X_1, \ldots, X_n$, we have

$$\begin{aligned}\int \left| \int \psi(y) f(y|x) d\nu(y) - \int \psi(y) f_0(y|x) d\nu(y) \right| d\mathbb{P}_n(x) \\ \leq \|\psi\|_\infty \int\int |f(y|x) - f_0(y|x)| d\nu(y) d\mathbb{P}_n(x) \\ \leq \|\psi\|_\infty \left(\int\int |\sqrt{f(y|x)} - \sqrt{f_0(y|x)}|^2 d\nu(y) d\mathbb{P}_n(x) \right)^{1/2}.\end{aligned}$$

Hence $N^{-1} \sum_{n=0}^{N-1} h^2(f(\cdot|X_n), f_0(\cdot|X_n)|X_n)$ is bounded away from 0 for all $f \in A_N$, $N > m$. As A_n is convex, $f_{n,A_n} \in A_n \subset B_n$ for all n so that (2.25) follows. In contrast, derivation of the analogous condition (2.15) for a non-random neighborhood required compactness in the last example. However, the conclusion of Theorem 2.2 is weaker in that it only concludes $\Pi(A_n|X_0, \ldots, X_n) \to 0$ instead of the more desirable conclusion $\Pi(B_n|X_0, \ldots, X_n) \to 0$.

EXAMPLE 3.3. Consider a Markov process with transition density $f(y|x) = g(y - \rho x)$, where g is a density function on $[0, 1]$ and $-1 < \rho < 1$ is known. Let g_0 be the true value of g, assumed to be continuous and bounded away from 0. The corresponding Markov process is then ergodic.

Let $\phi_0(u) = 1$ and $\phi_j(u) = \sqrt{2}\cos(j\pi x)$, $j \geq 1$. Clearly, $\{\phi_j\}$ is an orthonormal basis for $L_2[0, 1]$. Consider a prior for g given by the infinite dimensional exponential family

$$g(u) = \exp\left\{ \sum_{j=0}^{\infty} \theta_j \phi_j(u) - c(\theta) \right\}, \tag{3.4}$$

where independently $\theta_j \sim N(0, j^{-2-q})$, $q > 0$, and $c(\theta)$ is a normalizer. By Barron et al. (1999) and Walker (2004), and the fact that the Kullback-Leibler and Hellinger distances are unaffected by transformations, it follows that

(i) f is a.s. a transition density,

(ii) the Kullback-Leibler property holds at $f_0(y|x) = g_0(y - \rho x)$,

(iii) $\sum_{j=0}^{\infty} \sqrt{\Pr(f \in B_j)} < \infty$ for $B_j = \{g(y - \rho x) : g \in A_j\}$, where A_j's are constructed as in pages 2037–2038 of Walker (2004) for densities.

Then by Theorem 2.3, it follows that for any given $\epsilon > 0$, the posterior probability of the set $\{f(y|x) = g(y - \rho x) : h(g_0, g) > \epsilon\}$ goes to 0 a.s.

References

AMEWOU-ATISSO, M., GHOSAL, S., GHOSH, J.K. and RAMAMOORTHI, R.V. (2003). Posterior consistency for semiparametric regression problems. *Bernoulli*, **9**, 291–312.

BARRON, A., SCHERVISH, M. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.*, **27**, 536–561.

CHOUDHURI, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.*, **99**, 1050–1059.

GHOSAL, S., GHOSH, J.K. and RAMAMOORTHI, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.*, **27**, 143–158.

GHOSAL, S. and VAN DER VAART, A.W. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35** (to appear).

SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **4**, 10–26.

SHIRYAYEV, A.N. (1984). *Probability.* Springer-Verlag, New York.

TANG, Y. and GHOSAL, S. (2007). Posterior consistency of Dirichlet mixtures for estimating a transition density. *J. Statist. Plann. Inf.* (to appear).

VAN DER VAART, A.W. and WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

WALKER, S.G. (2003). On sufficient conditions for Bayesian consistency. *Biometrika*, **90**, 482–488.

WALKER, S.G. (2004). New approaches to Bayesian consistency. *Ann. Statist.*, **32**, 2028–2043.

SUBHASHIS GHOSAL
DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
2501 FOUNDERS DRIVE
RALEIGH, NC 27695-8203, U.S.A
E-mail: ghosal@stat.ncsu.edu

YONGQIANG TANG
DEPARTMENT OF PSYCHIATRY
SUNY HEALTH SCIENCES CENTER
450 CLARKSON AVENUE
BROOKLYN, NY 11203-1203, U.S.A.
E-mail: yongqiang_tang@yahoo.com

Paper received March 2005; revised November 2005.