Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 1711 – 1726
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Posterior consistency of Dirichlet mixtures for estimating a transition density

Yongqiang Tang[a], Subhashis Ghosal[b,*,1]

[a]Department of Psychiatry, SUNY Health Sciences Center, 450 Clarkson Avenue, Box 1203, Brooklyn, NY 11203, USA
[b]Department of Statistics, North Carolina State University, 220 Patterson Hall, 2501 Founders Drive, Raleigh, NC 27695-8203, USA

Received 5 April 2005; accepted 27 March 2006
Available online 9 June 2006

Abstract

The Dirichlet process mixture of normal densities has been successfully used as a prior for Bayesian density estimation for independent and identically distributed (i.i.d.) observations. A Markov model, which generalizes the i.i.d. set up, may be thought of as a suitable framework for observations arising over time. The predictive density of the future observation is then given by the posterior expectation of the transition density given the observations. We consider a Dirichlet process mixture prior for the transition density and study posterior consistency. Like the i.i.d. case, posterior consistency is obtained if the Kullback–Leibler neighborhoods of the true transition density receive positive prior probabilities and uniformly exponentially consistent tests exist for testing the true density against the complement of its neighborhoods. We show that under reasonable conditions, the Kullback–Leibler property holds for the Dirichlet mixture prior. For certain topologies on the space of transition densities, we show consistency holds under appropriate conditions by constructing the required tests. This approach, however, may not always lead to the best possible results. By modifying a recent approach of Walker [2004. New approaches to Bayesian consistency. Ann. Statist. 32, 2028–2043] for the i.i.d. case, we also show that better conditions for consistency can be given for certain weaker topologies.

MSC: Primary 62G07; 62G20

Keywords: Dirichlet mixture; Markov process; Posterior consistency; Transition density; Uniformly exponentially consistent tests

1. Introduction

Let $\{X_n; n=0, 1, 2, \dots\}$ be a Markov process on $\mathbb{R}$ with transition density $f(y|x)$. In this paper, we consider the problem of estimating the function $f(y|x)$ using Bayesian smoothing techniques without assuming any parametric form. We shall restrict our attention to only the class of transition densities $\mathscr{F}=\{f : x \mapsto f(\cdot|x)$ is continuous from $\mathbb{R}$ to $L_1(\mathbb{R})\}$. The assumption is natural as smoothing-based methods can work only when the functions involved are at least continuous. Also, this assumption lets us avoid measurability concerns in a large space like that of all transition densities and work with the much smaller space of continuous functions. Further, under our assumed model (1.1) below, f is continuous in both of its arguments. From a non-Bayesian point of view, Prakasa Rao (1978) proposed a

* Corresponding author.
E-mail addresses: yongqiang_tang@yahoo.com (Y. Tang), ghosal@stat.ncsu.edu (S. Ghosal).
[1] Research of the second author is partially supported by NSF grant number DMS-0349111.

0378-3758/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2006.03.007

kernel based approach to estimating transition density and its invariant distribution function for stationary uniformly ergodic Markov process under the assumption that both the transition density and the density for invariant distribution are uniformly continuous; see also Rajarshi (1990) and Dorea (2002) for more recent work. Clemencon (2000) constructed a wavelet-based estimate which nearly attains the minimax rate. Bayesian approach to the estimation of transition density has not been considered in the literature so far. In this paper, we consider a Bayesian approach by putting a prior on the space of transition densities and study consistency of the posterior distribution.

The problem of Bayesian estimation of a transition density is naturally connected with the problem of predicting the future observations. In the absence of other factors, observations may be considered exchangeable and hence conditionally i.i.d. given the density f by de Finetti's theorem. The predictive density $p_{n+1}(\cdot|X_1,\ldots,X_n)$ of X_{n+1} at x is then given by the posterior expectation of $f(x)$. If, however, the observations arise over time, the future observations should also be dependent on the past given the parameter. A Markov process, where only the immediate past matters, is one of the most natural extensions of the i.i.d. structure. Although, given the present, the future will not further depend on the past, the dependence "propagates" and may reasonably capture the dependence structure of the observations. If a prior Π is put on the transition density $f(y|x)$ of the process, then as in the i.i.d. case, the predictive density of X_{n+1} at y given the observations is obtained as $E(f(y|X_n)|X_1,\ldots,X_n)$, the Bayes estimate of the transition density. The prediction problem therefore reduces to that of the estimation of the transition density.

Because we follow a likelihood-based approach, we assume a model for the unknown transition density. However, unlike the parametric models, we consider a nonparametric mixture model, allowing more flexibility. Because the data may be sparse in the thin strips one requires to consider in a classical kernel-based approach, model assisted inference may particularly be attractive. In our approach, the transition density is modeled as a mixture of normal kernels, where the center of the kernel is determined by the most recent observation through a link function H which is randomly chosen from a mixing distribution. Although, in reality, the link function could be arbitrary, we consider only the situation where link functions are indexed by a parameter $\theta \in \Theta \subset \mathbb{R}^d$. We also assume that $x \mapsto H(x;\theta)$ is continuous so that f defined in (1.1) belongs to $\mathscr{F}$. Therefore, the mixing can be described by a probability distribution P on Θ and the resulting transition density is given by

$$f(y|x) = f_{P,\sigma}(y|x) = \int \phi_\sigma(y - H(x;\theta))\,\mathrm{d}P(\theta). \tag{1.1}$$

Our model consists of a prior μ for σ and a prior $\tilde{\Pi}$ for P. The prior $\mu \times \tilde{\Pi}$ through the map $(\sigma, P) \mapsto f_{P,\sigma}$ induces a prior Π on $\mathscr{F}$, the space of the transition densities of the process. A Dirichlet process is particularly appealing as a prior for P, although other choices are also possible. Mixtures over σ may also be considered by assigning a joint mixing distribution to (θ, σ) which is then given a prior. An analysis similar to ours will lead to consistency results for this prior which can cover more true densities.

Our approach can be considered as a natural extension of the well studied Dirichlet mixture approach to density estimation. As in the case of a mixture model for density estimation, we may also represent the problem as that of a random effect (non-linear) autoregressive model, where the functional form of the autoregression function is considered to be known as below:

$$X_i \overset{\text{ind}}{\sim} \mathrm{N}(H(X_{i-1};\theta_i),\sigma^2), \quad \theta_i \overset{\text{i.i.d.}}{\sim} P. \tag{1.2}$$

Here, unlike that in a parametric model, the unknown parameters are varying along with the index of the observation, and are actually drawn as i.i.d. samples from an unknown distribution. Hence the model is "dynamic" as opposed to a "static" parametric mixture model. West and Harrison (1997) considered similar models as convenient forecasting tools, where the random effects can also have dependence. The representation (1.2) give rise to Markov chain Monte-Carlo methods for the computation of the posterior distributions as in Tang (2003). It may be noted that a lag of more than one in (1.2) may be considered (which leads to higher order Markov processes), but here we shall restrict attention to (1.2), mainly for notational convenience.

In this paper, we study posterior consistency of the above procedure. For i.i.d. data, a theory of posterior consistency has been developed by Schwartz (1965), Barron et al. (1999) and Ghosal et al. (1999), with the last one paying special attention to Dirichlet mixtures. The theory essentially requires showing that the prior probability of any ball around the true parameter in the sense of Kullback–Leibler divergence is positive, and that a uniformly exponentially consistent sequence of tests for testing the true value of the parameter, against the complement of any neighborhood of it intersected with a suitable sieve, exists. Beyond the i.i.d. set up, posterior consistency has been studied in the context of independent nonidentically distributed observations (Amewou-Atisso et al., 2003; Choudhuri et al., 2004). We shall first extend Schwartz's theory of posterior consistency to ergodic Markov processes and apply it in the context of Dirichlet mixture model for transition densities. As posterior consistency under dependence has not been addressed previously, we believe the techniques we develop here will be useful to subsequent investigations.

The notion of consistency naturally depends on the topology under consideration. We consider several different topologies on $\mathscr{F}$ of varying strength. Specifically, we consider the weak topology on the invariant density, the L_1-distance integrated with respect to a probability measure v given by

$$d_v(f_1, f_2) = \int \|f_1(\cdot|x) - f_2(\cdot|x)\| \, \mathrm{d}v(x) = \int\int |f_1(y|x) - f_2(y|x)| \, \mathrm{d}y \, \mathrm{d}v(x) \tag{1.3}$$

and the supremum of the L_1-distance on $f(\cdot|x)$ given by

$$d_s(f_1, f_2) = \sup_{x\in\mathbb{R}} \|f_1(\cdot|x) - f_2(\cdot|x)\| = \sup_{x\in\mathbb{R}} \int |f_1(y|x) - f_2(y|x)| \, \mathrm{d}y, \tag{1.4}$$

which we shall refer to as the sup-L_1 distance; here and below $\|\cdot\|$ will stand for the L_1-distance. Under d_v, $\mathscr{F}$ can be viewed as a closed subset of the separable Banach space $L_1(\mathrm{d}x \times v)$. Under d_s, $\mathscr{F}$ can be viewed as a closed subset of the separable Banach space of continuous functions from $\mathbb{R}$ to $L_1(\mathbb{R})$ which are norm bounded. Thus under both metrics, $\mathscr{F}$ can be viewed as a complete separable metric space and hence measurability difficulties will not arise.

The condition of Kullback–Leibler positivity needs to be shown irrespective of the topology under consideration. If the true transition density is itself a mixture of normal of the form (1.1)

$$f_0(y|x) = f_{P_0,\sigma_0}(y|x) = \int \phi_{\sigma_0}(y - H(x,\theta)) \, \mathrm{d}P_0(\theta), \tag{1.5}$$

where $\sigma_0 > 0$, we show that the Dirichlet mixture prior satisfies the Kullback–Leibler positivity condition provided that the base measure of the Dirichlet process has large enough support. In view of (1.2), the transition density (1.5) will obtain if the observations follow a dynamic nonlinear autoregressive model where the random effect coefficients are i.i.d. In particular, by allowing degenerate distributions for the random effects, we observe that the parametric nonlinear autoregressive model is included in our model. Mixtures increase the flexibility widening the applicability of the model. Verification of the condition on testing is more involved. For the weak topology, uniformly exponentially consistent tests can be constructed by a Hoeffding type inequality for the Markov processes under the assumption that the link function H is bounded. If further, the supports of P and P_0 are compact, we construct uniformly exponentially consistent tests for the strong topology induced by d_s. It is relatively easy to construct tests for a pair of singletons. The compactness assumption together with the boundedness of H ensures that the space of transitions probabilities in the support of the prior is compact with respect to the sup-L_1 distance and hence helps combine the individual tests for each pair to a single uniformly exponentially consistent test.

However, the condition on boundedness of H eliminates some important link function such as the linear link. Walker (2003, 2004) introduced a martingale-based approach to the study of posterior consistency which avoids construction of tests. By modifying Walker's approach to Markov processes, we shall obtain a posterior consistency theorem for v-integrated weak neighborhoods for $f(\cdot|x)$. Under slightly stronger conditions, similar results are also obtained for the metric d_v. However, the sup-L_1 distance seems to be too strong for this approach to work. It appears that the boundedness of H restricts the effect of the preceding observation when it is too wild, thus help restore stability leading to uniform ergodicity and the desired uniformly exponentially consistent test.

While it will be interesting to obtain convergence rates of the posterior distribution and it is not unreasonable to expect fast rates like those of Ghosal and van der Vaart (2001) under our model, dependence substantially complicates the analysis. In this paper, we have not attempted to obtain convergence rates.

The organization of the paper is as follows. A general consistency result for Markov processes is presented in the next section. Distances on the space of transition densities are considered in Section 3. In Section 4, some auxiliary results helpful for establishing consistency using Walker's approach are presented. The next four sections, respectively, present sufficient conditions for the Kullback–Leibler property, consistency under the weak, integrated L_1 and sup-L_1 distances for the posterior based on the Dirichlet mixture prior. Finally, Section 9 gives proofs of some auxiliary lemmas.

2. A general consistency theorem for ergodic Markov processes

We begin with an extension of Schwartz's (1965) theorem to ergodic Markov processes which is obtained relatively easily. For a Markov process $\{X_n; n \geqslant 0\}$ with transition density f, let P_f^∞ denote the distribution of the infinite sequence $(X_0, X_1, \ldots)$. We assume that X_0 is fixed or has a known initial distribution. We shall abbreviate the posterior $\Pi(\cdot|X_0, \ldots, X_n)$ by $\Pi_n(\cdot)$ and by a.s., we shall mean a.s. $[P_{f_0}^\infty]$.

Theorem 2.1. *Let $\{X_n, n \geqslant 0\}$ be an ergodic Markov process with transition density $f \in \mathscr{F}$ and stationary distribution π. Let Π be a prior on $\mathscr{F}$. Let $f_0 \in \mathscr{F}$ and π_0 be, respectively, the true values of f and π. Let U_n be a sequence of subsets of $\mathscr{F}$ containing f_0.*

Suppose that there exist a sequence of tests Φ_n, based on $X_0, X_1, \ldots, X_n$ for testing the pair of hypotheses $\mathrm{H}_0 : f = f_0$ against $\mathrm{H} : f \in U_n^c$, and subsets $V_n \subset \mathscr{F}$ such that

(i) *f_0 is in the Kullback–Leibler support of Π, that is $\Pi\{f : K(f_0, f) < \varepsilon\} > 0$ for all $\varepsilon > 0$, where $K(f_0, f) = \int\int \pi_0(x) f_0(y|x) \log(f_0(y|x)/f(y|x))\, \mathrm{d}y\, \mathrm{d}x$,*
(ii) *$\Phi_n \to 0$ a.s. $[P_{f_0}^\infty]$,*
(iii) *$\sup_{f \in U_n^c \cap V_n} E_f(1 - \Phi_n) \leqslant C_1 \mathrm{e}^{-n\beta_1}$ for some constants C_1 and β_1,*
(iv) *$\Pi(f \in V_n^c) \leqslant C_2 \mathrm{e}^{-n\beta_2}$ for some constants C_2 and β_2.*

Then $\Pi_n(U_n) \to 1$ a.s.

Remark 2.1. (a) The weaker condition $E_{f_0}\Phi_n \to 0$ is sufficient for convergence in probability.

(b) The result is intended for U_n a fixed neighborhood of f_0 for an appropriate topology.

(c) In practice, a test against the alternative $f \in U_n^c$ is often obtained by combining appropriate tests against small convex balls, while the number of balls may be controlled by bounds on covering numbers; see Ghosal et al. (1999, Theorem 2).

(d) The "sieves" V_n increase the scope of applicability of the result, particularly when $\mathscr{F}$ is not compact. In compact spaces, one may usually work with the trivial choice $V_n = \mathscr{F}$. Condition (iv) is only a convenient sufficient condition for $\Pi_n(V_n^c) \to 0$, which may be shown directly for some models.

(e) In the rest of the paper, Condition (i) will be referred to as the Kullback–Leibler property.

We shall apply the above result in the context of model (1.2). The following result shows that the resulting Markov process is ergodic, aperiodic and Harris positive recurrent, and hence has a stationary distribution.

Lemma 2.1. *Let X_n be a Markov process satisfying* (1.2) *and let the link function H be continuous in x.*

If H is bounded, then model (1.2) *is uniformly ergodic.*

If $E|H(x, \theta)| \leqslant b + a|x|$ for all $x \in \mathbb{R}$, where $0 \leqslant a < 1$, $b < \infty$ and $\theta \sim P$, then model (1.2) *is geometrically ergodic.*

Model (1.2) *is geometrically ergodic when $H(x, \theta) = \rho x + u$, $E|\rho| < 1$ and $E|u| < \infty$.*

Proof. First let H be bounded. Let a be the upper bound of $|H(x, \theta)|$. Then

$$f_{P,\sigma}(y|x) = \int \phi_\sigma(y - H(x; \theta))\, \mathrm{d}P(\theta) \geqslant g(y) = \phi_\sigma(|y| + a).$$

Let $c = \int g(y)\, \mathrm{d}y$. When $a > 0$, $c = \int \phi_\sigma(|y| + a)\, \mathrm{d}y < \int \phi_\sigma(|y|)\, \mathrm{d}y = 1$. Thus, $f_{P,\sigma}(y|x) \geqslant cq(y)$ where $q(y) = g(y)/c$ is a probability density function. By Theorem 16.2.4 of Meyn and Tweedie (1993), model (1.2) is uniformly ergodic.

Now, more generally, let H be dominated by a linear function as described in the statement. Model (1.2) is equivalent to the following random-coefficient model:

$$X_{i+1} = H(X_i, \theta_{i+1}) + \varepsilon_{i+1},$$

where $\theta_i \stackrel{\text{i.i.d.}}{\sim} P$. Let $V(x) = |x| + 1$ and $C = \{x : |x| + 1 \leqslant m\}$. Then

$$E[V(X_{i+1})|X_i] \leqslant E(|H(x_i, \theta_{i+1})|\, |x_i) + E(|\varepsilon_{i+1}|) + 1 \leqslant a|x_i| + b + E(|\varepsilon_{i+1}|) + 1.$$

Since $E|\varepsilon_{i+1}|<\infty$, there exist a sufficiently large m and $\beta_1>0$, $\beta_2<\infty$ such that

$$E[V(X_1)|X_0=x]-V(x)\leqslant -\beta_1 V(x)+\beta_2 I_C(x).$$

The result holds by Theorem 15.0.1 of Meyn and Tweedie (1993).

The final assertion follows similarly. □

Proof of Theorem 2.1. The proof is similar to that in the i.i.d. case. Bound the posterior probability of U_n^c by

$$\Phi_n+\frac{\int (I(U_n^c\cap V_n)(1-\Phi_n)+I(V_n^c)(1-\Phi_n))\prod_{i=1}^n(f(X_i|X_{i-1})/f_0(X_i|X_{i-1}))\,\mathrm{d}\Pi(f)}{\int\prod_{i=1}^n(f(X_i|X_{i-1})/f_0(X_i|X_{i-1}))\,\mathrm{d}\Pi(f)}.$$

The first term above goes to 0 a.s. by Assumption (ii). The expected values of the two terms in the numerator of the second term are exponentially small by Assumptions (iii) and (iv), respectively. Therefore, it remains only to show that

$$\text{for any } \beta>0,\quad \mathrm{e}^{n\beta}\int\prod_{i=1}^n\frac{f(X_i|X_{i-1})}{f_0(X_i|X_{i-1})}\,\mathrm{d}\Pi(f)\to\infty\quad\text{a.s.}\tag{2.1}$$

Because f_0 defines an ergodic, aperiodic Harris positive recurrent Markov process, by the strong law of large numbers (cf., Theorem 17.1.7 of Meyn and Tweedie, 1993), $-n^{-1}\sum_{i=1}^n\log(f_0(X_i|X_{i-1})/f(X_i|X_{i-1}))\to -K(f_0,f)>-\varepsilon$ a.s. for any f such that $K(f_0,f)<\varepsilon$. Choosing $\varepsilon=\beta/4$, the result now follows by Fatou's lemma as in the original proof of Schwartz (1965). □

3. Topologies on the space of transition densities

The notion of consistency on $\mathscr{F}$ depends on the notion of a topology. We consider three broad approaches—based on the distances on the invariant measures π for f, distances on $f(\cdot|x)$ integrated with respect a probability measure v and distances on $f(\cdot|x)$ maximized with respect to x. Among the first category, there are the weak topology, the L_1-distance and the Hellinger distance on π. This notion of distance is somewhat weak because it does not distinguish between many transition densities f that lead to the same invariant distribution π. In the last two approaches, we view $f(\cdot|x)$ as a family of densities indexed by x, where we have usual notions of distances for a given x. In the second approach, we average out the distances with respect to some probability measure v. The distance to be integrated could be, in principle, any distance such as a metric for the weak topology, the L_1-distance or the Hellinger distance, or distance like quantities such as the squared Hellinger distance or the Kullback–Leibler divergence. The probability measure v will generally be assumed to have a large support and dominate all $f(\cdot|x)$. For instance, Birgé (1983) considered the average squared Hellinger distance for this purpose. In the third approach, one measures distance based on the maximum of the distances for given x such as the sup-L_1 distance.

Suitable tests for the weak topology may be constructed by using essentially the same idea used in the case of i.i.d. observations; see the proof of Theorem 4.4.2 of Ghosh and Ramamoorthi (2003). However, we shall need to substitute Hoeffding's exponential inequality by a similar inequality for ergodic Markov processes. Such an inequality has been recently obtained by Tang (2006). However, the approach requires the assumption that H is bounded, while by using Walker's approach, consistency may be obtained without this assumption for a very similar topology. Therefore, we forgo the test construction, which may be found in Tang (200), and show consistency using Walker's method.

For the L_1-distance $\|\pi-\pi_0\|=\int|\pi(x)-\pi_0(x)|\,\mathrm{d}x$ between the stationary distributions, tests against a small ball may be obtained as in Theorem 2 of Ghosal et al. (1999) by the modification of Hoeffding's inequality for Markov processes mentioned above, assuming that the link function H is bounded. Test against the complement of a ball may then be obtained by a compactness condition, or more generally, from a bound on the covering numbers for appropriate sieves. Because we shall obtain tests for the much stronger sup-L_1 distance under essentially the same conditions, we do not describe the procedure in detail.

Birgé (1983) constructed tests on the space of transition densities with respect to the integrated squared Hellinger metric under the condition that for all x and some $c_1,c_2>0$, $c_1v(y)\leqslant f(y|x)\leqslant c_2v(y)$, that essentially leads to the condition that H is bounded and the support of P is compact in the context of Dirichlet mixtures. Following Walker's approach, we shall relax these assumptions.

For the sup-L_1 metric, we shall verify consistency by constructing appropriate uniformly exponentially consistent tests under the condition that H is bounded and the support of P is compact.

For technical reasons, we shall also consider the semimetric

$$d_C(f_1, f_2) = \sup_{x\in C} \| f_1(\cdot|x) - f_2(\cdot|x)\| = \sup_{x\in C} \int |f_1(y|x) - f_2(y|x)|\,\mathrm{d}y, \tag{3.1}$$

which reduces to the sup-L_1 metric d_s when $C = \mathbb{R}$.

4. Auxiliary results on consistency

In this section, we obtain some results on the behavior of posterior distribution of the transition density for a general Markov model extending a recent approach to consistency by Walker (2003, 2004). These auxiliary results will help derive consistency results for Dirichlet mixture models in the following sections. In this section, we do not assume model (1.2) but only that $f_0 \in \mathscr{F}$ and Π has support in $\mathscr{F}$.

Let A be the complement of a neighborhood of f_0. Let f_{nA} be the predictive density of X_{n+1} with posterior restricted to the set A. Then

$$f_{n,A} = \int_A f(X_{n+1}|X_n)\Pi_{n,A}(\mathrm{d}f) = \frac{\int_A f(X_{n+1}|X_n)R_n(f)\Pi(\mathrm{d}f)}{\int_A R_n(f)\Pi(\mathrm{d}f)}, \tag{4.1}$$

where $\Pi_{n,A}(\cdot)$ is the posterior Π_n restricted to A and

$$R_n(f) = \prod_{i=1}^{n} \frac{f(X_i|X_{i-1})}{f_0(X_i|X_{i-1})} \tag{4.2}$$

stands for the likelihood ratio. Let the squared Hellinger metric conditioning on x be

$$h^2(f_1, f_2|x) = \int (\sqrt{f_1(y|x)} - \sqrt{f_2(y|x)})^2\,\mathrm{d}y = 2\left(1 - \int \sqrt{f_1(y|x)f_2(y|x)}\,\mathrm{d}y\right).$$

The next theorem and the following two corollaries will be helpful in studying consistency with respect to the weak topology.

Theorem 4.1. *Assume that the prior Π on f satisfies the Kullback–Leibler property at f_0. Suppose that there exists C such that $\pi_0(C) > 0$ and for all $x \in C$, $f \in A$ and n, $h^2(f_{n,A}, f_0|x) \geqslant \gamma > 0$. Then $\Pi(A|X_0, \ldots, X_n) \to 0$ a.s. exponentially fast.*

Proof. Let $D_n^2 = \int_A R_n(f)\Pi_A(\mathrm{d}f) = \int_A R_n(f)\Pi(\mathrm{d}f)/\Pi(A)$ and $Z_n = 1 - \gamma I_{X_n}(C)/2$. By (2.1), it suffices to show that $D_n \to 0$ exponentially fast a.s. Now

$$E(D_{n+1}|X_0, \ldots, X_n) = D_n \int \sqrt{f_{n,A}(X_{n+1}|X_n) f_0(X_{n+1}|X_n)}\,\mathrm{d}X_{n+1} \leqslant D_n Z_n.$$

Applying successively and using $D_0 = 1$, we have

$$E(D_{n+1}) \leqslant E\left(\prod_{i=1}^{n} Z_n\right) \leqslant E(\mathrm{e}^{-n\gamma T_n/2}) \leqslant \mathrm{e}^{-n\gamma c/2} + \Pr(T_n < c),$$

where $T_n = n^{-1}\sum_{i=1}^{n} I_{X_i}(C) \to \pi_0(C) > 0$ a.s. The result follows by choosing $c < \pi_0(C)$ and a large deviation estimate such as Corollary 3 of Tang (2006). □

Corollary 4.1. *Suppose that*

$$A = \left\{ f : \left| \int g(y) f(y|x)\,\mathrm{d}y - \int g(y) f_0(y|x)\,\mathrm{d}y \right| \geqslant \delta > 0 \textit{ for all } x \in C \right\},$$

where g is a bounded continuous function and C is an interval with $\pi_0(C)>0$. Then $\Pi_n(A)\to 0$ a.s. if Π has the Kullback–Leibler property.

Proof. Without loss of generality, we assume that $|g|\leqslant 1$. Note that $A=A_1\cup A_2$, where $A_1=\{f\in\mathscr{F}:\int g(y)f(y|x)\,\mathrm{d}y-\int g(y)f_0(y|x)\,\mathrm{d}y\geqslant\delta$ for all $x\in C\}$ and $A_2=\{f\in\mathscr{F}:\int g(y)f(y|x)\,\mathrm{d}y-\int g(y)f_0(y|x)\,\mathrm{d}y\leqslant-\delta$ for all $x\in C\}$ by the fact that $\int g(y)f(y|x)\,\mathrm{d}y-\int g(y)f_0(y|x)\,\mathrm{d}y$ is a continuous function of x and cannot change sign given that $\inf\{|\int g(y)f(y|x)\,\mathrm{d}y-\int g(y)f_0(y|x)\,\mathrm{d}y|:x\in C\}\geqslant\delta$.

It is easy to show that for any $X_n\in C$,

$$\int g(X_{n+1})f_{n,A_1}(X_{n+1}|X_n)\,\mathrm{d}X_{n+1}\geqslant\int g(X_{n+1})f_0(X_{n+1}|X_n)\,\mathrm{d}X_{n+1}+\delta$$

and hence $h^2(f_{n,A_1},f_0|X_n)\geqslant\delta^2/4$ by the well-known relationships between the L_1-distance, integrals and the Hellinger distance. By Theorem 4.1, $\Pi_n(A_1)\to 0$ a.s. and similarly for $\Pi_n(A_2)$. □

Corollary 4.2. *Suppose that $A=\{f:\sup(\|f(\cdot|x)-f_*(\cdot|x)\|:x\in C)<\delta\}$, where f_* is a fixed transition density satisfying $\inf\{\|f_*(\cdot|x)-f_0(\cdot|x)\|:x\in C\}>2\sqrt{\delta}$ and $\pi_0(C)>0$. Then $\Pi(A|X_0,\ldots,X_n)\to 0$ a.s. if Π has the Kullback–Leibler property.*

Proof. For any $f\in A$, $x\in C$, we have $h^2(f,f_*|x)<\delta$, and hence for any $X_n\in C$, by the convexity of the squared Hellinger metric, $h^2(f_{n,A},f_*|X_n)<\delta$. Now for any $X_n\in C$, $h^2(f_0,f_*|X_n)\geqslant\varepsilon^2/4$ where $\varepsilon=\inf\{\|f_*(y|x)-f_0(y|x)\|:x\in C\}$. Thus, $\inf\{h^2(f_{n,A},f_0|X_n):X_n\in C\}\geqslant\varepsilon^2/4-\delta>0$. By Theorem 4.1, $\Pi_n(A)\to 0$ a.s. □

The following theorem uses a countable covering of the parameter space whose prior probabilities satisfy a summability condition and can be useful for studying consistency with respect to d_v.

Theorem 4.2. *Let $A=\bigcup_{j=1}^{\infty}A_j$ and assume that there exist sets $C_1,\ldots,C_k\subset\mathbb{R}$ and $\gamma>0$ such that $\pi_0(C_l)>0$ for each $l=1,\ldots,k$, and for each A_j and n, $\inf\{h^2(f_{n,A_j},f_0|x):x\in C_l\}>\gamma$ for some $j=1,\ldots,k$. Suppose that $\sum_{j=1}^{\infty}\sqrt{\Pi(A_j)}<\infty$ and the Kullback–Leibler property at f_0 holds. Then $\Pi_n(A)\to 0$ a.s.*

Proof. We have

$$\Pi_n(A)\leqslant\sum_j\sqrt{\Pi_n(A_j)}=\sum_j D_{nj}(\Pi(A_j))^{1/2}\left(\int R_n(f)\Pi(\mathrm{d}f)\right)^{-1/2},\tag{4.3}$$

where $D_{nj}^2=\int_{A_j}R_n(f)\Pi_{A_j}(\mathrm{d}f)$. Because of the Kullback–Leibler property, it suffices to show that the numerator goes to zero exponentially fast. As in the proof of Theorem 4.1, for any $j=1,2,\ldots$, $E(D_{nj})\leqslant\sum_{l=1}^{k}E(\mathrm{e}^{-n\gamma T_{nl}/2})$, where $T_{nl}=n^{-1}\sum_{i=1}^{n}I_{X_i}(C_l)$. Thus,

$$\Pr\left[\sum_{j=1}^{\infty}D_{nj}\sqrt{\Pi(A_j)}>\mathrm{e}^{-n\varepsilon}\right]\leqslant\mathrm{e}^{n\varepsilon}\sum_{l=1}^{k}E(\mathrm{e}^{-n\gamma T_{nl}/2})\sum_{j=1}^{\infty}\sqrt{\Pi(A_j)},\tag{4.4}$$

which converges to 0 exponentially fast. □

In the following result, along a more classical line, we use a condition of the existence of sieves with high prior probabilities instead of the condition on the summability of the square roots of prior probabilities of A_j's.

Theorem 4.3. *For any n, there exist sets W_n and V_n such that $A\subset W_n\cup V_n$, $V_n=\bigcup_{j=1}^{N_n}A_{nj}$, finitely many subsets $C_1,\ldots,C_k\subset\mathbb{R}$ and $\gamma>0$ such that W_n has exponentially small prior probability, and for each n and each $j=1,\ldots,N_n$, $\inf\{h^2(f_{n,A_{nj}},f_0|x):x\in C_l\}>\gamma$ for some $l=1,\ldots,k$. Suppose that $N_n\leqslant K\mathrm{e}^{nu}$ where $u<\min(\gamma\pi_0(C_l)/2:1\leqslant l\leqslant k)$ and $K<\infty$. If the Kullback–Leibler property holds at f_0, then $\Pi_n(A)\to 0$ a.s.*

Proof. As in Theorem 2.1, $\Pi_n(W_n) \to 0$ a.s. Now

$$\Pi_n(V_n) \leqslant \sum_{j=1}^{N_n} \sqrt{\Pi_n(A_{nj})} \leqslant \sum_{j=1}^{N_n} D_{nj} \left(\int R_n(f)\Pi(\mathrm{d}f) \right)^{-1/2}, \tag{4.5}$$

where $D_{nj} = \sqrt{\int_{A_{nj}} R_n(f)\Pi_{A_{nj}}(\mathrm{d}f)}$. As in the proof of Theorem 4.1 we have $E(D_{nj}) \leqslant \sum_{l=1}^{k} E(\mathrm{e}^{-n\gamma T_{nl}/2})$, where $T_{nl} = n^{-1}\sum_{i=1}^{n} I_{X_i}(C_l)$. Thus, $\Pr\left(\sum_{j=1}^{N_n} D_{nj} > \mathrm{e}^{-n\varepsilon}\right) \leqslant \mathrm{e}^{n\varepsilon} N_n \sum_{l=1}^{k} E(\mathrm{e}^{-n\gamma T_n/2})$, which converges to 0 a.s. whenever $\varepsilon + u < \min_{1\leqslant l\leqslant k} \gamma\pi_0(C)/2$. □

5. Kullback–Leibler positivity of Dirichlet mixtures

Consider the model (1.1), where P is given the Dirichlet process prior with precision M and base measure G_0, and σ is given a prior μ. Let $H(x, \theta)$ be continuous in x and $\theta \in \Theta$, where Θ stands for the support of G_0. By a well-known property of the Dirichlet process, for any realized P, we have $P(\Theta) = 1$. We further assume that for any compact $C \subset \mathbb{R}$, the family of functions $\{x \mapsto H(x, \theta) : x \in C, \theta \in \Theta\}$ is uniformly equicontinuous, that is, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sup_{\theta\in\Theta} |H(x_1, \theta) - H(x_2, \theta)| \leqslant \varepsilon, \quad \text{whenever } x_1, x_2 \in C,\ |x_1 - x_2| < \delta. \tag{5.1}$$

Remark 5.1. The assumption clearly holds for AR(1), where $H(x, \theta) = \rho x + b$ and $|\rho| < 1$. More generally, the assumption also holds when $(\partial/\partial x)H(x, \theta)$ is continuous in (x, θ) and Θ is compact.

Proposition 5.1. *Let the true transition density be given by* (1.5), *where* $\mathrm{supp}(P_0) \subset \mathrm{supp}(G_0)$ *and* $\sigma_0 \in \mathrm{supp}(\mu)$. *Assume that there exists a compact subset* $B \subset \Theta$ *such that* $P_0(B) = d_0 > 0$ *and* $\sup\{|H(x, \theta)| : \theta \in B\} \leqslant G(x)$, *where* $\int G^2(x)\pi_0(x)\,\mathrm{d}x < \infty$ *and* $\int x^2\pi_0(x)\,\mathrm{d}x < \infty$. *Then the Kullback–Leibler property holds at* f_0.

To prove Proposition 5.1, we shall use the following lemma whose proof is given in Section 9.

Lemma 5.1. *If* $f_{P_1,\sigma_2}, f_{P_1,\sigma_2} \in \mathscr{F}$, $P_1(B),\ P_2(B) \geqslant d_0 > 0$, *and* $\sigma_1, \sigma_2 \in [\underline{\sigma}, \overline{\sigma}] \subset (0, \infty)$, *then*

$$\left| \log \frac{f_{P_1,\sigma_1}(y|x)}{f_{P_2,\sigma_2}(y|x)} \right| \leqslant \log \frac{\overline{\sigma}}{d_0\underline{\sigma}} + \frac{y^2 + G^2(x)}{\underline{\sigma}^2}. \tag{5.2}$$

Proof of Proposition 5.1. Note that

$$K(f_0, f) \leqslant \int\int \pi_0(x) f_0(y|x) \log \frac{\int \phi_{\sigma_0}(y - H(x, \theta))\,\mathrm{d}P_0(\theta)}{\int \phi_{\sigma}(y - H(x, \theta))\,\mathrm{d}P_0(\theta)}\,\mathrm{d}y\,\mathrm{d}x$$
$$+ \int\int \pi_0(x) f_0(y|x) \log \frac{\int \phi_{\sigma}(y - H(x, \theta))\,\mathrm{d}P_0(\theta)}{\int \phi_{\sigma}(y - H(x, \theta))\,\mathrm{d}P(\theta)}\,\mathrm{d}y\,\mathrm{d}x.$$

The integrand in the first term goes to 0 as $\sigma \to \sigma_0$. Therefore, by a domination argument using Lemma 5.1, the first term can be made small by choosing σ in a small interval around σ_0. By the assumption on μ, the prior probability of such an event is positive. It therefore suffices to show that for any fixed σ lying in a set of positive μ-probability, the second term assumes arbitrarily small values with positive prior probability. Let $\mathscr{N}' = \{P : P(B) \geqslant d_0\}$ and $[\underline{\sigma}, \overline{\sigma}]$ be a compact subset of $(0, \infty)$ containing σ_0. It follows from the last lemma that, for all $P \in \mathscr{N}'$ and $\sigma \in (\underline{\sigma}, \overline{\sigma})$, the log likelihood ratio $|\log(f_{P_0,\sigma}(y|x)/f_{P,\sigma}(y|x))|$ is bounded by a $(\pi_0 \otimes f_0)$-integrable function. Hence, given $\varepsilon > 0$, we can find K_1 and K_2 such that

$$E_{\pi_0\otimes f_0}\left[\left| \log \frac{f_{P_0,\sigma}(Y|X)}{f_{P,\sigma}(Y|X)} \right| (I\{|Y| > K_2\} + I\{|X| > K_1\}) \right] < \varepsilon. \tag{5.3}$$

Clearly, there exist compact subsets $B \subset B_1 \subset B_2$, a continuous function $t(\theta)$, a positive c such that

(a) $c = \inf\{\phi_\sigma(y - H(x,\theta)) : \underline{\sigma} \leqslant \sigma \leqslant \overline{\sigma}, \theta \in B_2, |x| \leqslant K_1, |y| \leqslant K_2\} > 0$,
(b) $P_0(B_1^c) \leqslant c\varepsilon\sqrt{2\pi}\underline{\sigma} P_0(B_1)$,
(c) $I(B_1) \leqslant t(\theta) \leqslant I(B_2)$.

Thus,

$$f_{P_0,\sigma}(y|x) \leqslant c\varepsilon P_0(B_1) + \int_{B_1} \phi_\sigma(y - H(x,\theta))\,\mathrm{d}P_0 \leqslant (1+\varepsilon)\int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P_0$$

and

$$\log \frac{f_{P_0,\sigma}(y|x)}{f_{P,\sigma}(y|x)} \leqslant \log(1+\varepsilon) + \log \frac{\int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P_0}{\int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P}$$
$$\leqslant \varepsilon + \left| \frac{\int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P_0}{\int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P} - 1 \right|.$$

To estimate the contribution from the region $\{(x,y) : |x| \leqslant K_1, |y| \leqslant K_2\}$ to the Kullback–Leibler number, consider the class of functions $\{(\theta,\sigma) \mapsto \phi_\sigma(y - H(x,\theta)) : |x| \leqslant K_1, |y| \leqslant K_2\}$. Because ϕ is Lipschitz continuous, $\sigma \geqslant \underline{\sigma}$ and $\{H(x,\cdot) : |x| \leqslant K_1\}$ is equicontinuous, we can find finitely many points $(x_1,y_1),\ldots,(x_m,y_m)$ such that for any $|x| \leqslant K_1, |y| \leqslant K_2$, there exists l with the property that

$$\sup\{|\phi_\sigma(y - H(x,\theta)) - \phi_\sigma(y_l - H(x_l,\theta))| : \theta \in B_2, \sigma \in [\underline{\sigma},\overline{\sigma}]\} < cd_0\varepsilon. \tag{5.4}$$

Let $\mathscr{N}''_\sigma$ be defined as the set of all P such that

$$\left| \int \phi_\sigma(y_l - H(x_l,\theta))t(\theta)\,\mathrm{d}P(\theta) - \int \phi_\sigma(y_l - H(x_l,\theta))t(\theta)\,\mathrm{d}P_0(\theta) \right| < c\varepsilon \tag{5.5}$$

for all $l = 1,\ldots,m$. Clearly $\mathscr{N}''_\sigma$ is a weak neighborhood of P_0. By a triangulation argument, it easily follows that for all $|x| \leqslant K_1, |y| \leqslant K_2$,

$$\left| \int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P(\theta) - \int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P_0(\theta) \right| < 3cd_0\varepsilon. \tag{5.6}$$

Now, if $P \in \mathscr{N}'$, it follows that $\int \phi_\sigma(y - H(x,\theta))t(\theta)\,\mathrm{d}P(\theta) \geqslant cd_0$. Therefore, for any $\sigma \in [\underline{\sigma},\overline{\sigma}]$, we found $\mathscr{N}_\sigma = \mathscr{N}' \cap \mathscr{N}''_\sigma$ such that $K(f_{P_0,\sigma}, f_{P,\sigma}) < 5\varepsilon$ for any $P \in \mathscr{N}_\sigma$. Since $\mathrm{supp}(P_0) \subset \mathrm{supp}(G_0)$, the Dirichlet process prior has P_0 in the weak support. As $\mathscr{N}_\sigma$ is a weak neighborhood of P_0, it has positive prior probability. This completes the proof. □

6. Consistency of Dirichlet mixtures under weak topology

Assume the conditions on the link function H and the prior described in the last section. In this section, we study consistency with respect to a weak topology where a sub-base of the neighborhood system is described by

$$\left\{ f : \int \left| \int g(y) f(y|x)\,\mathrm{d}y - \int g(y) f^*(y|x)\,\mathrm{d}y \right| v(x)\,\mathrm{d}x < \varepsilon \right\}, \tag{6.1}$$

where v is a probability density, g is a bounded continuous function and $f^* \in \mathscr{F}$.

Theorem 6.1. *Assume that H satisfies* (5.1), *the Kullback–Leibler property holds at f_0 and further $\mu(\sigma \geqslant \underline{\sigma}) = 1$ for some $\underline{\sigma} > 0$. Then the posterior is consistent at f_0 for the weak topology.*

Remark 6.1. A topology on $\mathscr{F}$ often used in decision theory is given by the class of sub-basic open sets

$$\left\{f : \left|\int \int h(x)g(y)f(y|x)\,\mathrm{d}y\, v(x)\,\mathrm{d}x - \int \int h(x)g(y)f^*(y|x)\,\mathrm{d}y\, v(x)\,\mathrm{d}x\right| < \varepsilon\right\}, \tag{6.2}$$

where $h(x) \in L_1(v)$ and v, g and f^* are as before. Then clearly

$$\left|\int \int h(x)g(y)f(y|x)\,\mathrm{d}y\, v(x)\,\mathrm{d}x - \int \int h(x)g(y)f^*(y|x)\,\mathrm{d}y\, v(x)\,\mathrm{d}x\right| \leqslant \int \left|\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f^*(y|x)\,\mathrm{d}y\right| h(x)v(x)\,\mathrm{d}x.$$

Now replacing $v(x)\,\mathrm{d}x$ by the finite measure $h(x)v(x)\,\mathrm{d}x$, we observe that the neighborhoods considered in (6.1) is stronger. Therefore, consistency with respect to the topology generated by the sub-base (6.2) also follows from Theorem 6.1.

Proof of Theorem 6.1. By standard arguments, it suffices to show that $\Pi_n(A) \to 0$ a.s. for $A = \{f : \int |\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f_0(y|x)\,\mathrm{d}y| v(x)\,\mathrm{d}x > \varepsilon\}$, where $|g(y)| \leqslant 1$.

Given $\varepsilon > 0$, find a compact C such that $\int_{C^c} v(x)\,\mathrm{d}x < \varepsilon/4$. Put $M = 2\max(v(x) : x \in C)$. Then

$$\inf\left\{\int_C \left|\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f_0(y|x)\,\mathrm{d}y\right| v(x)\,\mathrm{d}x : f \in A\right\} > \varepsilon/2, \tag{6.3}$$

and hence $\sup\{|\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f_0(y|x)\,\mathrm{d}y| : x \in C\} > \varepsilon/M$.

For any $f \in A$, let $x_f = \arg\max\{|\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f_0(y|x)\,\mathrm{d}y| : x \in C\}$. Note for any $x_1, x_2 \in C$ and any transition density q in the support of the prior,

$$\begin{aligned}&\left|\int g(y)q(y|x_1)\,\mathrm{d}y - \int g(y)q(y|x_2)\,\mathrm{d}y\right| \leqslant \|q(\cdot|x_1) - q(\cdot|x_2)\| \\ &\quad = \left\|\int \phi_\sigma(\cdot - H(x_1,\theta))\,\mathrm{d}Q - \int \phi_\sigma(\cdot - H(x_2,\theta))\,\mathrm{d}Q\right\| \leqslant \sqrt{\frac{2}{\pi}}\frac{|H(x_1,\theta) - H(x_2,\theta)|}{\underline{\sigma}}.\end{aligned}$$

Now we can find $\delta > 0$ such that whenever $|x_1 - x_2| < \delta$, $x_1, x_2 \in C$ and $\theta \in \Theta$, we have $(2/\pi)^{1/2}|H(x_1,\theta) - H(x_2,\theta)|/\underline{\sigma} < \varepsilon/(4M)$. Choosing $q = f$ in one occasion and $q = f_0$ in another, and applying the triangle inequality, it follows that $|\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f_0(y|x)\,\mathrm{d}y| \geqslant \varepsilon/(2M)$ whenever $|x - x_f| < \delta$.

Thus, we can partition $C = \cup_{l=1}^m C_l$ and $A = \cup_{l=1}^m A_l$ such that the length of each interval C_l is at most δ and

$$\inf_{x \in C_l}\left|\int g(y)f(y|x)\,\mathrm{d}y - \int g(y)f_0(y|x)\,\mathrm{d}y\right| > \frac{\varepsilon}{2M}, \tag{6.4}$$

whenever $f \in A_l$. By Corollary 4.1, $\Pi_n(A_l) \to 0$ a.s. for each l and hence $\Pi_n(A) \to 0$ a.s. □

7. Consistency of Dirichlet mixtures under integrated-L_1 distance

Theorem 7.1. *Assume that H satisfies* (5.1), *the Kullback–Leibler property holds*, $\mu(\underline{\sigma} \leqslant \sigma \leqslant \overline{\sigma}) = 1$ *for some* $0 < \underline{\sigma} < \overline{\sigma} < \infty$ *and Θ is compact. Then the posterior is consistent at f_0 in d_v-metric.*

Proof. Let $A = \{f : \int \|f(\cdot|x) - f_0(\cdot|x)\| v(x)\,\mathrm{d}x > \varepsilon\}$. Similar to the proof of Theorem 6.1, we can find a compact interval C, and some $\varepsilon_1 > 0$ such that for any $f \in A$, $\sup\{\|f(\cdot|x) - f_0(\cdot|x)\| : x \in C\} > 2\varepsilon_1$. Then A can be partitioned into $\bigcup_{j=1}^m A_j$ such that for each A_j, there exists a compact interval $C_j \subset \mathbb{R}$ and $\varepsilon_1 > 0$ with the property that for any $f \in A_j$, $\inf\{\|f(\cdot|x) - f_0(\cdot|x)\| : x \in C_l\} > \varepsilon_1$.

By Lemma 8.1, the space of f is compact with respect to d_C for any compact C. Fix $0<\delta<\varepsilon_1^2/4$ and partition the support of the prior into $B_1,\dots,B_k$ such that for any $f_1, f_2 \in B_l$, $\sup\{\|f_1(\cdot|x)-f_2(\cdot|x)\|: x\in C_j\}<\delta$, $j=1,\dots,m$. By Corollary 4.2, $\Pi_n(A_j\cap B_l)\to 0$ a.s. for any $j=1,\dots,m$ and $l=1,\dots,k$, implying the result. □

For the stationary linear AR(1) process, compactness of Θ may be relaxed, as the following theorem shows.

Theorem 7.2. *Assume that* $H(x,\theta)=\rho x+b$, $\mu(\underline{\sigma}\leqslant\sigma\leqslant\overline{\sigma})=1$ *for some* $0<\underline{\sigma}<\overline{\sigma}<\infty$ *and the Kullback–Leibler property holds at* f_0. *If for any* $\beta_1,\beta_2>0$, *there exist* $\beta_0>0$ *and* $K_1<\infty$ *such that*

$$D_{P(\rho,b)}\{P: P[-n\beta_2\leqslant b\leqslant n\beta_2]<1-\beta_1\}<K_1\mathrm{e}^{-n\beta_0}, \tag{7.1}$$

then consistency holds at f_0 *for the distance* d_v.

Proof. Let A be as in the proof of Theorem 7.1. Similar to the proof of Theorem 6.1, we can find a compact interval C and $\varepsilon_1>0$ such that for any $f\in A$, $\sup\{\|f(\cdot|x)-f_0(\cdot|x)\|: x\in C\}>2\varepsilon_1$. Then A can be partitioned into $A_1,\dots,A_m$ as before.

Fix $\delta<\varepsilon_1^2/4$, $\beta_1=\delta/10$, $b=\sqrt{10\overline{\sigma}^2/\delta}$, $\eta_1=\underline{\sigma}\delta\sqrt{\pi/2}/(5a)$, $\eta_2=\underline{\sigma}\delta\sqrt{\pi/2}/5$, where $a=\sup\{|x|: x\in C\}$. Fix $\beta_2>0$ such that

$$\frac{32\beta_2}{\eta_1\eta_2}\left(1+\log\frac{1+\eta_2}{\eta_2}\right)<\beta=\left(\frac{\varepsilon_1^2}{8}-\frac{\delta}{2}\right)\min_{1\leqslant l\leqslant k}\pi_0(C_l). \tag{7.2}$$

According to the assumption, the sequence of sets $W_n=\{P: P[-n\beta_2\leqslant b\leqslant n\beta_2]<1-\beta_1\}$ has exponentially small prior probability. Define $V_n=\{P: P[-n\beta_2\leqslant b\leqslant n\beta_2]\geqslant 1-\beta_1\}\cap A$. Similar to the proof of Lemma 8.1, let $\mathscr{F}_b$ denote the set of all models $X_i=\rho_iX_{i-1}+b_i+\tau_i+\varepsilon_{i1}$ with corresponding transition density $f^*_{P,\underline{\sigma}}$, where $\tau_i=\varepsilon_{i2}I(|\varepsilon_{i2}|\leqslant b)$, ε_{i1} are i.i.d. from $\mathrm{N}(0,\underline{\sigma}^2)$ and ε_{i2} are i.i.d. from $\mathrm{N}(0,\sigma^2-\underline{\sigma}^2)$. Then for any $f_{P,\sigma}\in V_n$, $\sup\{\|f_{P,\sigma}(\cdot|x)-f^*_{P,\underline{\sigma}}(\cdot|x)\|: x\in\mathbb{R}\}\leqslant\delta/5$. Let P^* be the corresponding probability measure of P restricted to $[-n\beta_2\leqslant b\leqslant n\beta_2]$. If $f_{P,\sigma}\in V_n$, then $\sup\{\|f^*_{P,\underline{\sigma}}(\cdot|x)-f^*_{P^*,\underline{\sigma}}(\cdot|x)\|: x\in\mathbb{R}\}\leqslant\delta/5$.

Let $\mathscr{F}_d$ denote the subset of models $X_i=\rho_iX_{i-1}+d_i+\varepsilon_{i1}$, where $|d_i|\leqslant b+n\beta_2$. Divide $\rho\in(-1,1)$ into $N_1\leqslant 2/\eta_1+1\leqslant 4/\eta_1$ intervals with length no longer than η_1. Divide d into $N_2\leqslant 2(b+n\beta_2)/\eta_2+1\leqslant 4(b+n\beta_2)/\eta_2$ intervals with length no longer than η_2. Selecting a representative parameter value from each of these small rectangular region $R_1,\dots,R_N$ thus obtained, where $N=N_1N_2\leqslant 16(b+n\beta_2)/(\eta_1\eta_2)\leqslant 32n\beta_2/\eta_1\eta_2$, $n>b/\beta_2$, we construct a finite collection $\gamma_1^*,\dots,\gamma_N^*$. Let the corresponding transition densities be denoted by $f_l^*(y|x)=\phi_{\underline{\sigma}}(y-d_l^*-\rho_l^*x)$, $l=1,2,\dots,N$.

In the N-dimensional unit simplex $\Delta_N=\{(w_1,\dots,w_N): w_i\geqslant 0, \sum_{i=1}^N w_i=1\}$, we can get a finite set $\mathscr{S}$ such that for any $\mathbf{w}\in\Delta_N$, there exists a $\mathbf{w}^*\in\mathscr{S}$ such that $\|\mathbf{w}-\mathbf{w}^*\|_1<\delta/5$, where $\|\cdot\|_1$ stands for the ℓ_1-norm on Δ_N. Similar to the proof of Lemma 8.1, it is easy to show a $\delta/5$-net of $\mathscr{S}$ is a $3\delta/5$-net of $\mathscr{F}_d$ and hence 3δ-net of V_n with respect to d_C.

As in the proof of Lemma 1 of Ghosal et al. (1999), we have

$$J(\delta,V_n)\leqslant J(\delta/5,\mathscr{S})\leqslant N(1+\log\{(1+\delta/5)/(\delta/5)\})<n\beta \tag{7.3}$$

for $n>b/\beta_2$, where J stands for the metric entropy. An application of Theorem 4.3 now completes the proof. □

8. Consistency of Dirichlet mixtures under the sup-L_1 distance

For consistency in the sup-L_1 distance, we assume that (5.1) holds with $C=\mathbb{R}$, H is uniformly bounded and uniformly equicontinuous in each argument, Θ is compact, and $\mu(\underline{\sigma}<\sigma<\overline{\sigma})=1$ for some $0<\underline{\sigma}\leqslant\sigma_0\leqslant\overline{\sigma}<\infty$. For instance, if $H(x;\tau,\beta,\gamma,\delta)=\tau+\gamma(1+\exp(-\delta-\beta x))^{-1}$, then the assumptions hold provided that the possible values of all the parameters $(\beta,\gamma,\delta,\tau)$ lie in a compact set and β does not assume the value 0. Analogous result holds for other bounded link function satisfying the conditions of Theorem 7.1.

Theorem 8.1. *Under the above assumptions, if* $\operatorname{supp}(P_0) \subset \operatorname{supp}(G_0) \subset \Theta$ *and* $\sigma_0 \in \operatorname{supp}(\mu)$, *then the posterior distribution is consistent at* f_0 *under the* sup-L_1 *metric.*

To prove the theorem, we exploit the compactness obtained from Lemma 8.1 with $C = \mathbb{R}$.

Lemma 8.1. *If* H *is uniformly equicontinuous in* $x \in C$ *and* $\theta \in \Theta$ *and* Θ *is compact, then* $\mathscr{F} = \{f_{P,\sigma}(y|x) = \int \phi_\sigma(y - H(x;\theta))\,\mathrm{d}P(\theta) : P(\Theta) = 1\}$ *is compact with respect to* d_C.

If $H(x;\tau,\beta,\gamma,\delta) = \tau + \gamma\psi(\delta + \beta x)$, where $\psi(x) = (1 + \mathrm{e}^{-x})^{-1}$, then the condition that the support of β does not contain 0 may not be dropped as the following counterexample shows.

Remark 8.1. Consider the class of the degenerate measures $\mathscr{L} = \{\int \phi_\sigma(y - \tau - \gamma\psi(\delta + \beta x))\,\mathrm{d}P : \sigma = 1, P = \delta_{(\beta=b,\gamma=1,\delta=0,\tau=0)} : 0 < b \leqslant 1\}$. Put $P_b = \delta_{(b,1,0,0)}$. It easily follows that $d(f_{P_{b_1},\sigma}, f_{P_{b_2},\sigma}) = d(f_{P_1,\sigma}, f_{P_b,\sigma})$, where $b = b_2/b_1 > 1$. Given $\varepsilon > 0$, find b such that $d(f_{P_b,\sigma}, f_{P_1,\sigma}) = \varepsilon$. Now $\{f_{P_{b^{-j}},\sigma}; j = 0, 1, \ldots\}$ is an infinite sequence in $\mathscr{L}$, where any two elements are at least ε separated.

The following lemma shows how to construct exponentially consistent tests.

Lemma 8.2. *There exist subsets* $V_1, \ldots, V_k$ *of* $\mathscr{F}$ *covering the set* $V = \{f_{P,\sigma} : d_s(f_{P,\sigma}, f_0) > \varepsilon\}$ *and for each* $l = 1, \ldots, k$, *there exists a sequence of tests which is uniformly exponentially consistent for testing* $\mathrm{H}_0 : f_{P,\sigma} = f_0$ *against* $\mathrm{H}_l : f_{P,\sigma} \in V_l$.

The proofs of the above two lemmas are given in the next section.

Proof of Theorem 8.1. We verify the conditions of Theorem 2.1 with $U_n = \{f_{P,\sigma} : d(f_{P,\sigma}, f_0) > \varepsilon\}$ and $V_n = \mathscr{F}$. The condition (i) on Kullback–Leibler support holds by Proposition 5.1. The construction of the uniformly consistent tests required by (ii) and (iii) follows from Lemma 8.2. This completes the proof. □

9. Proof of the auxiliary lemmas

Proof of Lemma 5.1. Clearly $f_{P,\sigma}(y|x) \leqslant (2\pi)^{-1/2}\underline{\sigma}^{-1}$ and

$$f_{P,\sigma}(y|x) \geqslant \int_{\theta\in B} \frac{1}{\sqrt{2\pi}\overline{\sigma}} \exp\left[-\frac{(y - H(x,\theta))^2}{2\underline{\sigma}^2}\right] \mathrm{d}P(\theta) \geqslant \frac{1}{\sqrt{2\pi}\overline{\sigma}} \exp\left[-\frac{y^2 + G^2(x)}{\underline{\sigma}^2}\right] \mathrm{d}_0$$

for any $f_{P,\sigma}$ satisfying $P(B) \geqslant d_0$ and $\sigma \in [\underline{\sigma}, \overline{\sigma}]$. The conclusion follows. □

Proof of Lemma 8.1. Note that any Markov process having transition density $f_{P,\sigma}$ in $\mathscr{F} = \{f_{P,\sigma} : P(\Theta) = 1, \sigma \in [\underline{\sigma}, \overline{\sigma}]\}$ may be represented by

$$X_i = \varepsilon_{i2} + H(X_{i-1};\theta_i) + \varepsilon_{i1}, \tag{9.1}$$

where ε_{i1} are i.i.d. $\mathrm{N}(0, \underline{\sigma}^2)$, ε_{i2} are i.i.d. $\mathrm{N}(0, \sigma^2 - \underline{\sigma}^2)$, and ε_{i1} are independent of ε_{i2}. Let $\mathscr{F}_b^*$ denote the set of all models

$$X_i = \tau_i + H(X_{i-1};\theta_i) + \varepsilon_{i1}, \tag{9.2}$$

where $\tau_i = \varepsilon_{i2} I\{|\varepsilon_{i2}| \leqslant b\}$. Let P^* be the probability measure of (τ, θ) and $f^*_{P,\underline{\sigma}}$ be the transition density of model (9.2). By Lemma A.3 of Ghosal and van der Vaart (2001), it follows that

$$\sup_{x\in\mathbb{R}} \|f_{P,\sigma}(\cdot|x) - f^*_{P,\underline{\sigma}}(\cdot|x)\| \leqslant 2\Pr(|\varepsilon_{i2}| \geqslant b) \leqslant 2E(\varepsilon_{i2}^2)/b^2 \leqslant 2\overline{\sigma}^2/b^2. \tag{9.3}$$

We can find a sufficiently large b such that $2\overline{\sigma}^2/b^2 \leqslant \varepsilon$ for a given ε. Any ε-net for the auxiliary model will therefore be a 2ε-net for the original model.

By the equicontinuity of H and the compactness of Θ, for all $\theta_1, \theta_2 \in \Theta$ and $|\tau_1|, |\tau_2| \leqslant b$,

$$\|\phi_{\underline{\sigma}}(\cdot - \tau_1 - H(x;\theta_1)) - \phi_{\underline{\sigma}}(\cdot - \tau_2 - H(x;\theta_2))\| \leqslant D(|\tau_1 - \tau_2| + |H(x,\theta_1) - H(x,\theta_2)|) < \varepsilon$$

provided that $|\tau_1 - \tau_2| < \eta$, $\|\theta_1 - \theta_2\| < \eta$ for some suitable $\eta > 0$ and $D > 0$.

Divide the range of each component of the parameter (τ, θ) into N' intervals such that whenever a pair of parameter values share the same intervals in each component, they differ at most by η in each component. Selecting a representative parameter value from each of these small rectangular region $R_1, \ldots, R_N$ thus obtained, where $N \leqslant (N')^{d+1}$ and d is the dimension of θ, we have a finite collection $w_1 = (\tau_1^*, \theta_1^*), \ldots, w_N = (\tau_N^*, \theta_N^*)$. Let the corresponding transition densities be denoted by $f_l^*(y|x) = \phi_{\underline{\sigma}}(y - \tau_l^* - H(x;\theta_l^*))$, $l = 1, 2, \ldots, N$.

In the N-dimensional unit simplex $\Delta_N = \{(w_1, \ldots, w_N) : w_i \geqslant 0, \sum_{i=1}^N w_i = 1\}$, we can get a finite set $\mathscr{S}$ such that for any $\mathbf{w} \in \Delta_N$, there exists a $\mathbf{w}^* \in \mathscr{S}$ such that $\|\mathbf{w} - \mathbf{w}^*\|_1 < \varepsilon$, where $\|\cdot\|_1$ stands for the ℓ_1-norm on Δ_N. Let $\mathscr{F}^* = \{\sum_{l=1}^N w_l^* f_l^* : \mathbf{w}^* \in \mathscr{S}\}$. For any P^* satisfying $P^*(\Theta \times [-b \leqslant \tau \leqslant b]) = 1$, put $\mathbf{w} = (P^*(R_l) : l = 1, 2, \ldots, N)$ and find $\mathbf{w}^* \in \mathscr{S}$ such that $\|\mathbf{w} - \mathbf{w}^*\|_1 < \varepsilon$. Put $f^* = \sum_{l=1}^N w_l^* f_l^* \in \mathscr{F}^*$. Splitting $\int \phi_{\underline{\sigma}}(\cdot - \tau - H(x;\theta))\,\mathrm{d}P^*(\tau,\theta)$ into subregions $R_1, \ldots, R_N$ and noting that the integrand on R_l differs from $f_l^*(\cdot|x)$ by at most ε, thus for any $f^*_{P,\underline{\sigma}} \in \mathscr{F}_b^*$, there is $f^* \in \mathscr{F}^*$ such that $\sup_{x \in C}(f^*_{P,\underline{\sigma}}, f^*) < 2\varepsilon$. Thus, any ε-net for $\mathscr{F}^*$ will be a 2ε-net for $\mathscr{F}_b^*$ and hence a 3ε-net for $\mathscr{F}$. The result follows since $\mathscr{S}$ and $\mathscr{F}^*$ are compact. □

To prove Lemma 8.2, we need a few more lemmas.

Lemma 9.1. *For any P, P_1, P_2 supported on Θ and any $\sigma, \sigma_1, \sigma_2$ lying in $[\underline{\sigma}, \overline{\sigma}]$,*

$$\frac{1}{\sqrt{2\pi}\overline{\sigma}} \mathrm{e}^{-(y^2+L^2)/\underline{\sigma}^2} \leqslant f_{P,\sigma}(y|x) \leqslant \frac{1}{\sqrt{2\pi}\underline{\sigma}} \min(\mathrm{e}^{-y^2/(4\overline{\sigma}^2)} \mathrm{e}^{L^2/(2\underline{\sigma}^2)}, 1).$$

and hence $f_{P_1,\sigma_1}(y|x)/f_{P_2,\sigma_2}(y|x) \leqslant (\overline{\sigma}/\underline{\sigma})\mathrm{e}^{(y^2+L^2)/\underline{\sigma}^2}$.

Proof. Note $|H(x;\theta)| \leqslant L$. The first relation follows from $(a+b)^2 \leqslant 2(a^2+b^2)$. For the second, observe that

$$\frac{(y - H(x;\theta))^2}{2\sigma^2} = \frac{y^2/2 + (y/\sqrt{2} - \sqrt{2}H(x;\theta))^2 - H^2(x;\theta)}{2\sigma^2} \geqslant \frac{y^2}{4\overline{\sigma}^2} - \frac{L^2}{2\underline{\sigma}^2}. \tag{9.4}$$

The rest follows easily. □

Lemma 9.2. *If P_1, P_2, P_3 are supported on Θ and $\sigma_1, \sigma_2, \sigma_3$ lie in $[\underline{\sigma}, \overline{\sigma}]$, given $b_1 > 0$, there exists $\eta > 0$ such that for all x, with $\lambda = \underline{\sigma}^2/(8\overline{\sigma}^2)$, we have that*

$$\int \left[\frac{f_{P_1,\sigma_1}(y|x)}{f_{P_2,\sigma_2}(y|x)}\right]^\lambda [f_{P_3,\sigma_3}(y|x) - f_{P_2,\sigma_2}(y|x)]\,\mathrm{d}y \leqslant b_1, \tag{9.5}$$

whenever $d_s(f_{P_2,\sigma_2}, f_{P_3,\sigma_3}) < \eta$.

Proof. If we could find b such that

$$\int_{|y|>b} \left[\frac{f_{P_1,\sigma_1}(y|x)}{f_{P_2,\sigma_2}(y|x)}\right]^\lambda f_{P_3,\sigma_3}(y|x)\,\mathrm{d}y \leqslant b_1/2 \tag{9.6}$$

for all x, all P_1, P_2, P_3 supported on Θ and any $\sigma_1, \sigma_2, \sigma_3 \in [\underline{\sigma}, \overline{\sigma}]$, then the choice $\eta = b_1/T$ will work, where $T = 2(\overline{\sigma}/\underline{\sigma})^{\lambda} \exp((b^2 + L^2)/(8\overline{\sigma}^2))$. To see this, break the integral in integrals over the two regions $|y| > b$ and $|y| \leqslant b$. The second integral is then clearly less than $b_1/2$. To bound the first, note that the integrand is a constant multiple of $e^{-y^2/(8\overline{\sigma}^2)}$, which is integrable. This implies (9.6) for a sufficiently large b and thus proves the result. □

Lemma 9.3. *Let $V_l \subset \mathscr{F}$, $A_l \subset \mathbb{R}$, $\zeta, \delta_2 > 0$ such that*

$$\inf_{x \in A_l, f \in V_l} \| f_0(\cdot|x) - f(\cdot|x) \| \geqslant 2\delta_2 \quad \text{and} \quad \inf_{x, f \in V_l} \int_{A_l} f(y|x) \, dy \geqslant \zeta. \tag{9.7}$$

Let $\lambda = \underline{\sigma}^2/(8\overline{\sigma}^2)$, $b_1 = (1 - [1 - \delta_2^2]^{2\lambda})\zeta/2$ and η be the number obtained in Lemma 9.2 corresponding to b_1. If $\sup\{d(f_1, f_2) : f_1, f_2 \in V_l\} \leqslant \eta$, then there exists a sequence of uniformly exponentially consistent tests for the pair of hypotheses $H_0 : f = f_0$ against $H_l : f \in V_l$.

Proof. Fix $f_1 \in V_l$. Set

$$h(k) = \sum_{j=1}^{k} \log \frac{f_1(X_{2j}|X_{2j-1})}{f_0(X_{2j}|X_{2j-1})}, \quad \phi_n = I\{h(\lfloor n/2 \rfloor) > 0\}. \tag{9.8}$$

We shall show that ϕ_n is a sequence of uniformly exponentially consistent tests for testing H_0 against H_l.

For any two densities p and q, $\int p^{\alpha} q^{1-\alpha} \leqslant (1 - \|p - q\|^2/4)^{\min(\alpha, 1-\alpha)}$, $0 \leqslant \alpha \leqslant 1$. Thus,

$$\int \left[\frac{f_1(x_{n+1}|x_n)}{f_0(x_{n+1}|x_n)} \right]^{\lambda} f_0(x_{n+1}|x_n) \, dx_{n+1} \leqslant 1 + ([1 - \delta_2^2]^{\lambda} - 1) I(x_n \in A_l). \tag{9.9}$$

Similarly,

$$\int \left[\frac{f_0(x_{n+1}|x_n)}{f_1(x_{n+1}|x_n)} \right]^{\lambda} f_1(x_{n+1}|x_n) \, dx_{n+1} \leqslant 1 + ([1 - \delta_2^2]^{\lambda} - 1) I(x_n \in A_l). \tag{9.10}$$

Let $\kappa = b_1 + 1 + ([1 - \delta_2^2]^{2\lambda} - 1)\zeta = 1 + ([1 - \delta_2^2]^{2\lambda} - 1)\zeta/2 < 1$. We get by (9.9) that

$$\begin{aligned} &\int \int \left[\frac{f_1(x_{n+1}|x_n)}{f_0(x_{n+1}|x_n)} \right]^{\lambda} f_0(x_{n+1}|x_n) f_0(x_n|x_{n-1}) \, dx_{n+1} \, dx_n \\ &\leqslant 1 + ([1 - \delta_2^2]^{\lambda} - 1) \int I(x_n \in A_l) f_0(x_n|x_{n-1}) \, dx_n \leqslant 1 + ([1 - \delta_2^2]^{\lambda} - 1)\zeta \leqslant \kappa. \end{aligned} \tag{9.11}$$

For any $f \in V_l$, by Lemma 9.2 and (9.10),

$$\begin{aligned} &\int \left[\frac{f_0(x_{n+1}|x_n)}{f_1(x_{n+1}|x_n)} \right]^{\lambda} f(x_{n+1}|x_n) \, dx_{n+1} \\ &= \int \left[\frac{f_0(x_{n+1}|x_n)}{f_1(x_{n+1}|x_n)} \right]^{\lambda} [f(x_{n+1}|x_n) - f_1(x_{n+1}|x_n)] \, dx_{n+1} + \int \left[\frac{f_0(x_{n+1}|x_n)}{f_1(x_{n+1}|x_n)} \right]^{\lambda} f_1(x_{n+1}|x_n) \, dx_{n+1} \\ &\leqslant b_1 + 1 + ([1 - \delta_2^2]^{\lambda} - 1) I(x_n \in A_l). \end{aligned}$$

Therefore,

$$\begin{aligned} &\int \int \left[\frac{f_0(x_{n+1}|x_n)}{f_1(x_{n+1}|x_n)} \right]^{\lambda} f(x_{n+1}|x_n) f(x_n|x_{n-1}) \, dx_{n+1} \, dx_n \\ &\leqslant b_1 + 1 + ([1 - \delta_2^2]^{\lambda} - 1) \int I(x_n \in A_l) f(x_n|x_{n-1}) \, dx_n \leqslant b_1 + 1 + ([1 - \delta_2^2]^{\lambda} - 1)\zeta \leqslant \kappa. \end{aligned} \tag{9.12}$$

By an application of Markov's inequality, the type I error $E_{f_0}(\phi_n)$ of ϕ_n can therefore be bounded by $E_{f_0}(e^{\lambda h(k)}) \leqslant \kappa E_{f_0}(e^{\lambda h(k-1)})$, where $k = \lfloor n/2 \rfloor$. Applying repeatedly, we obtain $E_{f_0}\phi_n \leqslant \kappa^k \leqslant e^{-cn}$ for some $c > 0$.

The maximum type II error $\sup\{E_f(1-\phi_n) : f \in V_l\}$ of ϕ_n can similarly be bounded by

$$\sup_{f\in V_l} E_f(\exp[-\lambda h(k)]) \leqslant \kappa \sup_{f\in V_l} E_f(\exp[-\lambda h(k-1)]), \tag{9.13}$$

which leads again to the bound e^{-cn}. □

Proof of Lemma 8.2. We partition V twice. The first partition $V_1, \ldots, V_m$ is made in order to find intervals $A_1, \ldots, A_m$ satisfying

$$\inf_{x_n\in A_l,\, f_{P,\sigma}\in V_l} \| f_0(\cdot|x_n) - f_{P,\sigma}(\cdot|x_n)\| \geqslant \frac{\delta}{2} \quad \text{for } l = 1, \ldots, m. \tag{9.14}$$

Then refine the partition of each V_l in order to find uniformly exponentially consistent tests.

By Lemma 8.1, we can find a finite disjoint partition $\{V_1, \ldots, V_m\}$ of V such that for any two transition density f_1, f_2 in the same V_l $(l=1, \ldots, m)$, $d(f_1, f_2) \leqslant \delta/4$. Choose and fix $f_l = f_{P_l,\sigma_l} \in V_l$. Since $d(f_0, f_l) > \delta$ and $\|f_0(\cdot|x) - f_l(\cdot|x)\|$ is a continuous function of x, we can find intervals A_l such that

$$\inf_{x\in A_l} \|f_0(\cdot|x) - f_l(\cdot|x)\| \geqslant \frac{3\delta}{4} \quad \text{for } l = 1, \ldots, m. \tag{9.15}$$

Then $\|f_0(\cdot|x) - f(\cdot|x)\| \geqslant \delta/2$ for all $f \in V_l$ and $x \in A_l$, $l = 1, \ldots, m$.

By Lemma 9.1, for any x_n and any P supported on Θ,

$$f_{P,\sigma}(X_{n+1} \in A_l|x_n) \geqslant d_l := \frac{1}{\sqrt{2\pi}\overline{\sigma}} \int_{A_l} \exp[-(y^2 + L^2)/\underline{\sigma}^2]\,dy. \tag{9.16}$$

Put $\zeta = \min(d_1, \ldots, d_m) > 0$, $\delta_2 = \delta/4$, $\lambda = \underline{\sigma}^2/(8\overline{\sigma}^2)$, and $b_1 = (1 - [1 - \delta_2^2]^{\lambda})\zeta/2$. We shall subdivide each V_l to obtain the desired sets. Given b_1, let η be the value got as in Lemma 9.2 and $\eta_1 = \min(\eta, \delta/8)/2$. By the compactness of $\mathscr{F}$, we can find a finite η_1-net for each V_l. This leads to disjoint partitions V_{lj}, $j = 1, \ldots, v_l$, $l = 1, \ldots, m$ such that $d(f_1, f_2) \leqslant 2\eta_1 \leqslant \eta$ for all $f_1, f_2 \in V_{lj}$, $j = i, \ldots, v_l$, $l = 1, \ldots, m$. Uniformly exponentially consistent tests against the alternative V_{lj} can now be found by Lemma 9.3. □

References

Amewou-Atisso, M., Ghosal, S., Ghosh, J.K., Ramamoorthi, R.V., 2003. Posterior consistency for semiparametric regression problems. Bernoulli 9, 291–312.

Barron, A., Schervish, M., Wasserman, L., 1999. The consistency of posterior distributions in nonparametric problems. Ann. Statist. 27, 536–561.

Birgé, L., 1983. Robust testing for independent non identically distributed variables and Markov chains. In: J.P. Florens et al. (Eds.), Specifying Statistical Models. From Parametric to Non-Parametric. Using Bayesian or Non-Bayesian Approaches. Lecture Notes in Statistics, vol. 16 Springer, New York, pp. 134–162.

Choudhuri, N., Ghosal, S., Roy, A., 2004. Bayesian estimation of the spectral density of a time series. J. Amer. Statist. Assoc. 99, 1050–1059.

Clemencon, S.J.M., 2000. Adaptive estimation of the transition density of a regular Markov chain. Math. Methods Statist. 9, 323–357.

Dorea, C.C.Y., 2002. Strong consistency of kernel estimators for Markov transition densities. Bull. Braz. Math. Soc. 33, 409–418.

Ghosal, S., Ghosh, J.K., Ramamoorthi, R.V., 1999. Posterior consistency of Dirichlet mixtures in density estimation. Ann. Statist. 27, 143–158.

Ghosal, S., van der Vaart, A.W., 2001. Entropies and rates of convergence of maximum likelihood and Bayes estimation for mixtures of normal densities. Ann. Statist. 29, 1233–1263.

Ghosh, J.K., Ramamoorthi, R.V., 2003. Bayesian Nonparametrics. Springer, New York.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, New York.

Prakasa Rao, B.L.S., 1978. Density estimation for Markov processes using delta-sequences. An. Inst. Statist. Math. 30, 321–328.

Rajarshi, M.B., 1990. Bootstrap in Markov sequences based on estimates of transition density. An. Inst. Statist. Math. 42, 253–268.

Schwartz, L., 1965. On Bayes procedures. Z. Wahr. Verw. Gebiete 4, 10–26.

Tang, Y., 2003. Dirichlet process mixture models for Markov processes. Unpublished Ph.D. Thesis, North Carolina State University.

Tang, Y., 2006. A Hoeffding type inequality for ergodic time series. J. Theoretical Probab., to appear.

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.
West, M., Harrison, J., 1997. Bayesian forecasting and dynamic models. Second ed.. Springer, New York.