Posterior consistency for the spectral density of non-Gaussian stationary time series

Yifu Tang[*]   Claudia Kirch[†]   Jeong Eun Lee[‡]   Renate Meyer[§]

Abstract

Various nonparametric approaches for Bayesian spectral density estimation of stationary time series have been suggested in the literature, mostly based on the Whittle likelihood (Whittle, 1957) approximation. A generalization of this approximation has been proposed in Kirch et al. (2019) who prove posterior consistency for spectral density estimation in combination with the Bernstein-Dirichlet process prior for Gaussian time series. In this paper, we will extend the posterior consistency result to non-Gaussian time series by employing a general consistency theorem of Shalizi (2009) for dependent data and misspecified models. As a special case, posterior consistency for the spectral density under the Whittle likelihood as proposed by Choudhuri et al. (2004) is also extended to non-Gaussian time series. Small sample properties of this approach are illustrated with several examples of non-Gaussian time series.

1   Introduction

Many ecological, epidemiological, and physical time series show periodic behaviour. Spectral analysis of stationary time series in the frequency domain is useful in detecting cycles and modelling the second order dependence structure of the data. However, a Bayesian approach generally requires an evaluation of the likelihood function. Even in the Gaussian case, the likelihood function can be computationally intractable due to the necessary inversion of a high-dimensional covariance matrix. Therefore, most Bayesian nonparametric methods for spectral density estimation (e.g. Gangopadhyay et al. 1999; Liseo et al. 2001; Choudhuri et al. 2004; Hermansen 2008; Chopin et al. 2013; Cadonna et al. 2017) have used the Whittle likelihood approximation (Whittle, 1957) which was originally suggested as an approximation

---

[*]Department of Statistics, The University of Auckland, Auckland, New Zealand

[†]Institute for Mathematical Stochastics, Department of Mathematics, Otto-von-Guericke University, Magdeburg, Germany

[‡]Department of Statistics, The University of Auckland, Auckland, New Zealand

[§]Department of Statistics, The University of Auckland, Auckland, New Zealand

for Gaussian time series but later extended to non-Gaussian time series by Hannan (1973). Apart from avoiding matrix inversion, the Whittle likelihood has the advantage that it depends directly on the spectral density instead of indirectly via an inverse Fourier transform. However, using the Whittle likelihood approximation can be inefficient for non-Gaussian time series with moderate sample sizes (Contereras-Cristán et al., 2006) and even for Gaussian time series with high autocorrelations and small sample sizes. To address this issue, Chopin et al. (2013) used importance sampling to correct a posterior sample based on the Whittle likelihood, Sykulski et al. (2019) introduced the de-biased Whittle likelihood, Rao and Yang (2020) suggested a new frequency-domain quasi-likelihood, and Kirch et al. (2019) the so-called corrected likelihood. The latter is a generalization of the Whittle likelihood based on the idea of utilizing the efficiencies of a parametric working model but correcting this nonparametrically in the frequency domain to make it robust with respect to violations of the parametric model assumptions.

When using any of these pseudo-likelihoods, one has to deal with model misspecification. An ensuing question is: When one adopts a pseudo-likelihood, will the posterior distribution associated with this pseudo-likelihood still be consistent in the sense that the posterior distribution eventually concentrates in the neighbourhood of the true parameter, namely the underlying spectral density function in this case, as the sample size gets large? For parametric Bayesian modelling using the Whittle likelihood, the Bernstein-von Mises theorem (Tamaki, 2008) guarantees that under some regularity conditions, Bayesian and frequentist inference procedures can produce the same asymptotic result. In the case of nonparametric Bayesian modelling, however, posterior consistency needs to be proved explicitly. While in the frequentist literature on nonparametric spectral density estimation (e.g. Pawitan and O'sullivan 1994; Kakizawa 2006) it is common to prove asymptotic results beyond the Gaussianity assumption, the same cannot be said for Bayesian nonparametric spectral density estimation. Prior works such as Choudhuri et al. (2004) and Kirch et al. (2019) established posterior consistency by utilizing the result that the Gaussian measure and the distribution associated with the pseudo-likelihood are mutually contiguous. However, in practice, it is often the case that the time series data at hand does not look anywhere close to a Gaussian time series. This calls for a theoretical justification of posterior consistency under mild assumptions on the time series without having to assume Gaussianity.

The difficulty of extending the consistency result to non-Gaussian time series is due to the fact that the contiguity argument, on which prior works relied heavily, is no longer viable for general non-Gaussian time series as shown in Meier (2018, Section 4.3). Therefore, a novel method is needed to establish posterior consistency under mild conditions on the time series. To this end, we adopt the general consistency theorem of Shalizi (2009).

The goal of this paper is to prove that under rather general conditions regarding the underlying time series, the posterior distribution associated with the corrected Whittle likelihood will still be consistent. As the Whittle likelihood is a special case of the corrected likelihood, the result will carry over to the posterior distribution based on the Whittle likelihood.

The article is organized as follows. A modification of the general consistency theorem of Shalizi (2009), which is the main tool used to prove consistency in this paper, is introduced in Section 2. The assumptions on the time series, together with the working model, the parameter space and the prior are detailed in Section 3. In Section 4 the conditions on time series are summarized with some examples. The proof of posterior consistency is illustrated in Section 5. Section 6 discusses the tools and assumptions employed in this paper and outlines possible future directions. The proof of Theorem 2.1, the modified theorem of Shalizi (2009), is given in Appendix A. The proof of Proposition 3.1 as well as some results regarding Lipschitz norm and Bernstein polynomials are stated in Appendix B. Finally, Appendix C contains the proof of Proposition 3.2.

2   Modification of the general consistency theorem of Shalizi

Historically, people tend to prove posterior consistency by gauging the discrepancy between the probability distribution of the sample and the 'proxy model' via Kullback-Leibler (KL) divergence. Examples are Schwartz (1965), Choudhuri et al. (2004), Choi and Schervish (2007), Kleijn and van der Vaart (2006) and Shalizi (2009), to name but a few. Most of these papers require that the KL divergence be finite, which is reasonable as it ensures that even in the case of misspecification, the discrepancy between the true distribution and the proxy model is not too severe. Nevertheless, an interesting question is: Will we obtain a consistent result even if we use some proxy model which drastically deviates from the true distribution in terms of the KL divergence? If the answer is affirmative, then is it possible to prove the consistency without resorting to the KL divergence?

Often, we just want to estimate the value of a parameter pertaining to the probability distribution of the sample. For instance, the parameter of interest may be the expectation, the variance-covariance matrix, the quantile function or the spectral density function. It is possible that the proxy model will result in a consistent posterior even though it deviates from the true distribution severely. We will use a rather simple example to illustrate this idea. Suppose we observe an i.i.d. real sequence $Y_1, Y_2, \cdots, Y_n$. We know nothing about the true probability distribution except $E|Y_1| < \infty$. The parameter of interest is the expectation $\mu_0 = EY_1$. From a frequentist perspective, one can always use the arithmetic mean $\hat{\mu}_n = \frac{Y_1 + \cdots + Y_n}{n}$ as the

3

estimator because the Law of Large Numbers (LLN) tells us that $\hat{\mu}_n$ is always a consistent estimator no matter what the true probability distribution is. From a Bayesian point of view, however, it becomes complicated in that we have to adopt some (possibly) misspecified proxy model when using Bayes theorem in light of the absence of the true distribution. Suppose in this case we use the normal distribution with mean μ and variance 1 as proxy model and assume $\mu \sim N(0, 1)$, i.e. the prior is the standard normal distribution. The resulting posterior distribution will be

$$N\left(\frac{n}{n+1}\hat{\mu}_n, \frac{1}{n+1}\right).$$

From the LLN as $n \to \infty$, the posterior will concentrate on μ_0, which is the true value of the parameter. In other words, posterior consistency is achieved by using a normal proxy model, even though this is not the actual likelihood associated with the sample. Even if the data come from the Cantor distribution (cf. e.g. Schmidt (1991)), posterior consistency holds despite the fact that the normal proxy model is not a good approximation in this case because the Cantor distribution does not even possess probability density function. In fact, the KL divergence between the Cantor distribution and the normal proxy model $N(\mu, 1)$ is infinity for all $\mu \in \mathbb{R}$ due to the mutual singularity of the two probability measures.

There are some attempts to prove posterior consistency without invoking the KL divergence in the literature. For example, Sriram et al. (2013) considered Bayesian quantile regression with proxy model being the asymmetric Laplacian distribution and proved consistency without evaluating the KL divergence between the true distribution and the proxy model when the prior is proper. When extending the result to an improper prior, however, they did use the KL assumption mainly to ensure that the posterior was a probability measure. Another example is Syring et al. (2019) in the context of the quantile estimation in combination with a carefully chosen proxy model, where posterior consistency is achieved without using the KL assumption.

Despite the fact that Shalizi (2009) proved the consistency theorem by evaluating the KL divergence, those techniques can be adapted so that the evaluation of the KL divergence is no longer required. In this section we derive a version of the consistency theorem of Shalizi (2009) that will then be used to prove posterior consistency for a spectral density estimator in Section 3, where we both strengthen and weaken certain assumptions. For instance, apart from the avoidance of the KL assumption, we consider convergence in probability while Shalizi (2009) adopts almost sure convergence. Also, instead of considering a large parameter space, we confine ourselves to a smaller parameter set on which some desired properties hold.

Let $X_1, X_2, \cdots$, for short X_1^∞, be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$, taking values in the measurable space $(\Xi, \mathcal{X})$. The sigma field generated by $X_1^n := (X_1, X_2, \cdots, X_n)$ is denoted by $\sigma(X_1^n)$.

Unless otherwise noted, all probabilities are taken with respect to P, and $\mathrm{E}[\cdot]$ always means expectation under this probability measure.

We begin with a prior probability measure Π_0 over $(\Theta, \mathcal{T})$, and update this using Bayes rule after seeing $x_1^n = X_1^n(\omega)$. This updating is done based on a proxy model via a linking function

$$R_n\left(\theta, \omega\right) \coloneqq f_\theta\left(x_1^n\right),$$

where θ is the parameter of interest taking values in the parameter space, a measurable space $(\Theta, \mathcal{T})$. In classical Bayesian analysis the likelihood function $f_\theta(x_1^n)$ is used but here this is not required, rather a quasi-likelihood or some exponentiated loss or discrepancy function (see Bissiri et al. 2016; Syring 2017) can be used. In the next section we will use a generalization of the Whittle likelihood to this end.

In analogy to classical Bayesian statistics, for any $A \in \mathcal{T}$, we define the so-called (pseudo) posterior measure after seeing $x_1^n = X_1^n(\omega)$ to be

$$\Pi_n\left(A\right) \coloneqq \frac{\int_A R_n\left(\theta, \omega\right) \Pi_0\left(d\theta\right)}{\int_\Theta R_n\left(\theta, \omega\right) \Pi_0\left(d\theta\right)} = \frac{\int_A f_\theta\left(x_1^n\right) \Pi_0\left(d\theta\right)}{\int_\Theta f_\theta\left(x_1^n\right) \Pi_0\left(d\theta\right)}. \tag{2.1}$$

The parameter space can either be finite-dimensional such as $\mathbb{R}^n$ or infinite-dimensional such as $C([0,1])$.

In the following, we will often suppress the arguments θ and/or ω for brevity if this is possible without causing ambiguity.

The assumptions required for the proof of the main theorem, Theorem 2.1, follow:

Assumption 2.1. *For each $n \in \mathbb{N}$, the proxy model $R_n\left(\theta, \omega\right)$ is positive, $\mathcal{T} \otimes \sigma\left(X_1^n\right)$ -measurable and,*

$$0 < \int_\Theta R_n\left(\theta, \omega\right) \Pi_0\left(d\theta\right) < \infty \quad P\text{-a.s.}$$

As already stated we do not require $f_\theta(\cdot)$ to be the probability density function of some distribution, rather the inequalities in Assumption 2.1 ensures that the posterior measure Π_n is well-defined.

Let $R_n(\theta)$ denotes the random variable $\omega \mapsto R_n(\theta, \omega)$.

Assumption 2.2. *For each $\theta \in \Theta$ and $n \in \mathbb{N}$, $\mathrm{E}\left|\ln(R_n(\theta))\right| < \infty$ and,*

$$h\left(\theta\right) \coloneqq - \lim_{n \to \infty} \frac{1}{n} \mathrm{E}(\ln(R_n(\theta)))$$

exists pointwise in θ such that h is $\mathcal{T}$-measurable. Moreover, $|h(\Theta)| < \infty$, where $h(A)$ denotes the essential infimum of h on $A \in \mathcal{T}$ with respect to Π_0.

The function $h(\cdot)$ defined in Assumption 2.2 can be negative. Note that the definition of $h(\theta)$ is different to Assumption 2 of Shalizi (2009) where $h(\theta)$ denotes the KL divergence rate from P.

Assumption 2.3. *As a function of ω,*

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \ln(R_n\left(\theta, \omega\right)) + h\left(\theta\right) \right|$$

is $\sigma\left(X_1^n\right)$ -measurable and,

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \ln(R_n\left(\theta\right)) + h\left(\theta\right) \right| \xrightarrow{P} 0, \quad n \to \infty.$$

Assumption 2.3 effectively states a Uniform Weak Law of Large Numbers.

Now we state the modified general consistency theorem in which the proof is closely related to Theorem 3 of Shalizi (2009), and is given in Appendix A.

Theorem 2.1. *Suppose that Assumptions 2.1 –2.3 hold and in addition $h(A) > h(\Theta)$ for any $A \in \mathcal{T}$ such that $\Pi_0(A) > 0$. Then*

$$\Pi_n\left(A\right) \xrightarrow{P} 0, \quad n \to \infty.$$

The proof of the above theorem shows $P(\Pi_n(A) \geq \exp(-n\alpha)) \to 0$ for some $\alpha > 0$ as $n \to \infty$.

3 Bayesian modelling and posterior consistency

In this section, we turn back to our original problem of interest and introduce the assumptions on the observed time series, the proposed approximate likelihood function with its parameter space, as well as the nonparametric prior on the space of spectral density functions.

3.1 Assumptions on the time series

Let $\{X_k\}_{k \in \mathbb{Z}}$ be a real-valued, zero-mean strongly stationary time series with finite variance on $(\Omega, \mathcal{F}, P)$. It is assumed that

$$\mathrm{E}(X_0|\mathcal{U}_{-\infty}) = 0, \ \mathrm{E}(X_0 X_k|\mathcal{U}_{-\infty}) = \mathrm{E}(X_0 X_k) =: \gamma_0(k), \ \forall k \in \mathbb{Z}, \quad (3.1)$$

where $\mathcal{U}_k = \sigma(X_s, s \leqslant k)$, $\mathcal{U}_{-\infty} = \cap_{k=-\infty}^{\infty} \mathcal{U}_k$ and $\sum_{k \in \mathbb{Z}} |\gamma_0(k)| < \infty$. Then, the spectral density exists, is continuous, 2π-periodic and symmetric at 0, such that it is fully defined if known on the interval $[0, \pi]$. For convenience, we will rescale the spectral density to $[0, 1]$, such that

$$\varphi_0(x) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_0(k) \exp\left(-ik\pi x\right) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_0(k) \cos\left(k\pi x\right), \quad x \in [0, 1].$$

Here the second representation is due to the real and symmetric autocovariance function. The above assumptions also guarantee that a law of large numbers for the quadratic form $(X_1^n)' A_n X_1^n$ for all $\|A_n\|_s \leqslant 1$ (see Corollary 1 in Yaskov (2018)) where X_1^n now denotes the column vector $(X_1, \ldots, X_n)'$ and $\|\cdot\|_s$ denotes the matrix spectral norm. This is closely related to spectral density estimation (see Yaskov (2018) for some further discussion) and will also be a key element of the below proof of consistency.

We further assume that

$$\varphi_0(x) \geqslant \beta \text{ for some } \beta > 0, \tag{3.2}$$

and that φ_0 is Lipschitz continuous with Lipschitz constant $L(\varphi_0)$, i.e.

$$L(\varphi_0) := \sup_{x,y \in [0,1], x \neq y} \left| \frac{\varphi_0(x) - \varphi_0(y)}{x - y} \right| < \infty.$$

The assumption of a Lipschitz continuous spectral density is stronger than absolute summability of the autocovariance sequence; see e.g. Bernstein's Theorem (Katznelson, 2004, Theorem 6.3).

However, it is weaker than the assumption made in Choudhuri et al. (2004) and Kirch et al. (2019), that the auto-covariance function $\gamma_0(\cdot)$ satisfies

$$\sum_{k \in \mathbb{Z}} |k|^\alpha |\gamma_0(k)| < \infty \text{ for some } \alpha > 1.$$

The latter implies that φ_0 is differentiable with $(\alpha - 1)$-Hölder-continuous derivative (see e.g. Remark 8.5 in page 46 of Serov (2017)), which in turn implies Lipschitz continuity. This stronger assumption is required to show contiguity of the proxy and true likelihood functions, which cannot be expected to hold beyond Gaussian time series. Because we will prove consistency using Theorem 2.1 above, we no longer require contiguity to hold.

We will present some examples of time series that satisfy all of our assumptions in Section 4.

3.2 Proxy model

The Whittle likelihood (Whittle, 1957) has been used in a variety of situations in both frequentist and Bayesian time series analysis. Whittle's likelihood assumes that the periodograms are exactly independent and exponentially distributed, which is well known to be true under quite general assumptions in some asymptotic sense. For applications concerning the spectral density alone this approximation is usually good enough, however, higher-order quantities (in the non-Gaussian case) often depend on the information coded in this dependency as observed by Dahlhaus et al. (1996). To mitigate this effect Kreiss et al. (2003) and Kirch et al. (2019) propose a related likelihood that inherits the dependency between periodogram

ordinates from a parametric working model in the time domain, while at the same time 'nonparametrically correcting' the inherited spectral density coded in the variances of the periodograms. In this context Whittle's likelihood becomes a 'nonparametrically corrected' version of the Gaussian white noise working model.

Here, we use this corrected Whittle likelihood approximation with a Gaussian mean-zero parametric working model as proxy model. We assume that the (rescaled) spectral density of the Gaussian parametric working model $\varphi_{\mathrm{par}}(x)$, $x \in [0,1]$ is Lipschitz continuous and satisfies

$$\varphi_{\mathrm{par}}(x) \geqslant \beta_{\mathrm{par}} \text{ for some } \beta_{\mathrm{par}} > 0$$

with autocovariance matrix $\Gamma_{\mathrm{par}} = \Gamma_{n,\mathrm{par}} = (\gamma_{\mathrm{par}}(k-j))_{k,j=1}^{n}$, where $\gamma_{\mathrm{par}}(\cdot)$ is the autocovariance function of the parametric working model.

The corrected Whittle likelihood is then defined as

$$f_\varphi(x_1^n) = (2\pi)^{-\frac{n}{2}} \det\left(\Gamma_{\mathrm{par}}^{-1} C_n^2(\varphi)\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}(x_1^n)' C_n(\varphi)\Gamma_{\mathrm{par}}^{-1} C_n(\varphi) x_1^n\right),$$
$$(3.3)$$

where $C_n(\varphi) = F_n' D_n^{-1/2}\left(\frac{\varphi}{\varphi_{\mathrm{par}}}\right) F_n$, F_n is an orthogonal matrix called the real Discrete Fourier Transform matrix (see e.g. formula (4.5.4) of Brockwell and Davis (1991) and Section 2.1 of Kirch et al. (2019) for details), ξ is a function on $[0,1]$ and $D_n(\cdot)$ the following corresponding diagonal matrix

$$D_n(\xi) = \begin{cases} \operatorname{diag}\left(\xi(0), \xi(y_1), \xi(y_1), \cdots, \xi(y_N), \xi(y_N), \xi\left(y_{\frac{n}{2}}\right)\right), & n \text{ even.} \\ \operatorname{diag}\left(\xi(0), \xi(y_1), \xi(y_1), \cdots, \xi(y_N), \xi(y_N)\right), & n \text{ odd.} \end{cases}$$

where $y_j = \frac{2j}{n}$, $j = 1, \cdots, N = \lfloor \frac{n-1}{2} \rfloor$ are scaled Fourier frequencies. The unknown spectral density φ is the parameter of interest.

The name *corrected* Whittle likelihood comes from the fact that a potentially misspecified parametric working model is nonparametrically corrected in the frequency domain by the matrix $D_n\left(\frac{\varphi}{\varphi_{\mathrm{par}}}\right)$. When a Gaussian white noise model is used as parametric working model, i.e. $\varphi_{\mathrm{par}}(x)$ is a constant function, then the corresponding corrected Whittle likelihood coincides with the original Whittle likelihood.

3.3 Parameter space

The true value of parameter is often assumed to lie in a compact set when proving consistency results. The parameter space, $(C([0,1]), \|\cdot\|_\infty)$, is the collection of all continuous functions on $[0,1]$ endowed with the sup-norm $\|\cdot\|_\infty$. It is well-known that the closed unit ball of $C([0,1])$ is not compact (Searcòid, 2006, Theorem 12.10.2) and, compact subsets of $C([0,1])$ is obtained by using the following Lipschitz norm $\|\cdot\|_{\mathrm{Lip}}$:

For any $\varphi \in C([0,1])$, the Lipschitz norm is defined by

$$\|\varphi\|_{\mathrm{Lip}} = \|\varphi\|_{\infty} + L(\varphi),$$

with Lipschitz constant of φ,

$$L(\varphi) = \sup_{0 \leqslant x \neq y \leqslant 1} \left| \frac{\varphi(x) - \varphi(y)}{x - y} \right|.$$

The parameter space is defined as

$$\Theta = \left\{ \varphi \in C[0,1] : \|\varphi\|_{\mathrm{Lip}} \leqslant M, \quad \inf_{0 \leqslant x \leqslant 1} \varphi(x) \geqslant m \right\} \tag{3.4}$$

for some suitably chosen values $0 < m < M < \infty$ and the true spectral density φ_0 is contained in Θ. The choices for m and M will be shown later in Assumption 3.1.

Let the metric $d_{\infty}(\cdot, \cdot)$ be induced by the sup-norm $\|\cdot\|_{\infty}$,

$$d_{\infty}(\phi, \psi) = \|\phi - \psi\|_{\infty} = \sup_{x \in [0,1]} |\phi(x) - \psi(x)|, \quad \text{for all } \phi, \psi \in \Theta.$$

In this paper, we will work on the space $(\Theta, d_{\infty}(\cdot, \cdot))$ equipped with the Borel σ-algebra $\mathcal{T}$. The following proposition shows that this space is a compact metric space and, the proof is given in Appendix B.

Proposition 3.1. Θ, equipped with $d_{\infty}(\cdot, \cdot)$, is a compact metric space.

3.4 Bernstein-Dirichlet process prior

For the Bayesian analysis, we will use the Bernstein-Dirichlet process prior Π_{BDP} proposed by Choudhuri et al. (2004) for spectral density estimation.

As in Kirch et al. (2019) we will not put the prior directly on the spectral density $\varphi \in \Theta$ but rather on the correction $c(x)$, which relates to the spectral density via

$$\varphi(x) = c(x)\varphi_{\mathrm{par}}^{\delta}(x) \tag{3.5}$$

for some $0 \leqslant \delta \leqslant 1$. This implies that only for $\delta = 0$ the Bernstein Dirichlet process prior is actually specified directly for the spectral density φ (as in (Choudhuri et al., 2004)), while for $\delta > 0$ the spectral density of the parametric working model provides some pre-smoothing. The latter is important because Bernstein polynomials are better for smoother functions (see e.g. Farouki 2012, Section 3.3 and Ghosal and van der Vaart 2017, Appendix E). Indeed for an (almost) correctly specified parametric working model a choice close to $\delta = 1$ is ideal in terms of smoothness of the correction term while for a completely misspecified working model a choice of $\delta = 0$ is better. For this reason, the parameter δ is called *model confidence parameter*. In this

work we assume δ to be fixed a priori. However, if a uniform prior is put on δ, the MCMC algorithm implemented in the beyondWhittle R-package (Meier et al., 2017) does provide estimates of δ with values according to the above intuition (see the simulations in Kirch et al. (2019)).

The Bernstein-Dirichlet process prior approximates the unknown correction function c by Bernstein polynomials, which have a uniform approximation property (see e.g. Kruijer 2008, p. 20): For any $c \sim \Pi_{BDP}$, define for $x \in [0,1]$

$$c(x) = \tau b(x; K, G), \tag{3.6}$$

$$\text{with} \quad b(x; K, G) = b(x; K, w_K(G)) := \sum_{j=1}^{K} w_{j,K}(G)\beta\left(x; j, K - j + 1\right),$$

where $\beta\left(x; a, b\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ are Beta-densities. Priors are then put on both the variance parameter τ, the order of the polynomial K, as well as the weights of the Bernstein polynomial. The weight vector $w_K(G) = (w_{1,K}(G), \cdots, w_{K,K}(G))$ is obtained from a Dirichlet process $G \sim$ Dirichlet Process(MG_0), where $M > 0$ and G_0 is a probability distribution function on $[0, 1]$ via $w_{j,K}(G) = G\left(\frac{j-1}{K}, \frac{j}{K}\right]$, $j = 1, 2, \cdots, K$. Samples from the Dirichlet process can be readily obtained using the truncated stick-breaking representation Sethuraman (1994).

The Dirichlet process G, the order K as well as the variance parameter τ are assumed to be a priori independent, where K has a probability mass function, $\rho(k) > 0$, for $k = 1, 2, \cdots$, and τ is a continuous random variable with support $(0, \infty)$.

Some properties of Bernstein polynomials that are important for the proofs in this work are summarized in Appendix B.

Clearly, $c \sim \Pi_{BDP}$ puts an implicit prior denoted by Π_C onto φ via (3.5). We will further restrict this prior to the parameter set of interest Θ as follows:

$$\Pi_0(A) := \frac{\Pi_C(A \cap \Theta)}{\Pi_C(\Theta)},$$

for any $A \in \mathcal{T}$.

This is well defined by the next proposition (proven in Appendix C) which shows prior positivity in any ball around φ_0 for appropriately chosen m, M in (3.4).

Assumption 3.1. *Let* $\inf_{0 \leqslant x \leqslant 1} \varphi_0(x) > m$ *and* $\left\|\frac{\varphi_0}{\varphi_{\text{par}}^\delta}\right\|_{\text{Lip}} \cdot \left\|\varphi_{\text{par}}^\delta\right\|_{\text{Lip}} < M$.

Assumption 3.1 guarantees that φ as in (3.5) with c as in (3.6) that are close enough to φ_0 in the supremum norm are also contained in Θ.

Remark. (a) For $\delta = 0$ the second part of Assumption 3.1 clearly simplifies to $\|\varphi_0\|_{\text{Lip}} < M$.

10

(b) If we make the slightly stronger assumption that φ_0 and φ_{par} are continuously differentiable, then the second assertion of Assumption 3.1 can be weakend to $\|\varphi_0\|_{\mathrm{Lip}} < M$ for any $0 \leqslant \delta \leqslant 1$. For a proof we refer to Appendix C.

Proposition 3.2. *Under Assumption 3.1 it holds $\Pi_C(B(\varphi_0, r)) > 0$ for all $r > 0$, where*

$$B(\varphi_0, r) = \{\varphi \in \Theta : \|\varphi - \varphi_0\|_\infty < r\}. \qquad (3.7)$$

Thus $\Pi_0(\Theta) = 1$ and Π_0 will assign positive probability to any neighbourhood of φ_0.

3.5 Posterior consistency

The posterior with respect to Π_0 and the proxy likelihood after updating X_1^n is denoted by Π_n and is defined in (2.1).

The following theorem is the main result of this paper, showing that with increasing sample size the posterior Π_n contracts towards a Dirac measure centred at the true spectral density φ_0. In particular, any (reasonable) point estimator obtained from the posterior will be a consistent estimator for the true spectral density even for non-Gaussian time series.

Theorem 3.1 (Posterior consistency)**.** *Under the assumptions on the time series and the prior in addition to Assumption 3.1, for any $r > 0$,*

$$\Pi_n(\{\varphi \in \Theta : \|\varphi - \varphi_0\|_\infty \geqslant r\}) = \Pi_n(B^c(\varphi_0, r)) \xrightarrow{P} 0, \ t \to \infty,$$

where $B^c(\varphi_0, r)$ denotes the complement of $B(\varphi_0, r)$ with respect to Θ and $B(\varphi_0, r)$ is as in (3.7).

Before we come to the proof of this theorem in Section 5, we will present some non-Gaussian time series in which our posterior consistency result is applicable.

4 Examples

To obtain posterior consistency, some assumptions on the underlying time series as outlined in Section 3.1 are required. A time series $\{X_n\}_{n \in \mathbb{Z}}$ is called regular (Bradley, 2007, Definition 2.13) if its past tail σ-field $\cap_{n=-\infty}^\infty \sigma(X_k, k \leqslant n)$ is (almost) trivial. If the time series is regular, then (3.1) automatically holds. Peligrad and Wu (2010) presents several examples of regular processes, including mixing sequences, functions of Gaussian processes, functions of i.i.d. random variables and reversible Markov chains. Apart from the above assumptions, we also require that the spectral density function of the time series $\varphi_0(\lambda)$, $\lambda \in [0, 1]$ is strictly positive and Lipschitz continuous. Consequently, any regular time series with strictly positive and Lipschitz

continuous spectral density function satisfies all our assumptions. Below are two concrete examples of this category: The first one is non-Gaussian with a singular distribution and the second one has a Lipschitz continuous but not continuously differentiable spectral density.

4.1 Causal and invertible ARMA processes

Causal and invertible $ARMA(p,q)$ model is powerful and popular in time series analysis. As a result of Pourahmadi (2001, Theorem 5.22), any such process $\{X_n\}_{n\in\mathbb{Z}}$ is regular. It is well-known that the spectral density function of $\{X_n\}_{n\in\mathbb{Z}}$ is strictly positive and continuously differentiable (cf. e.g. Brockwell and Davis 1991, Theorem 4.4.2).
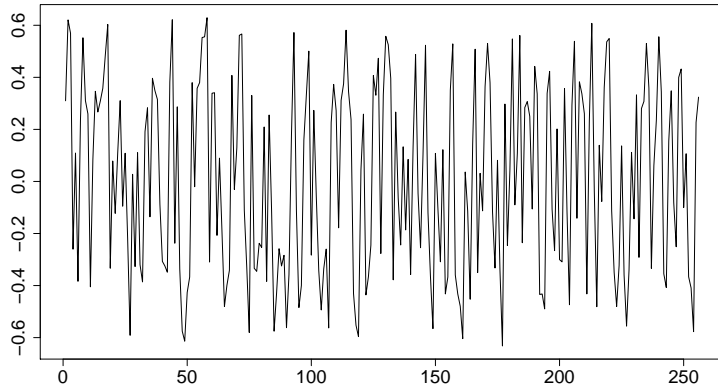

Figure 1: Simulated MA(1) time series with standard Cantor distributed innovations, sample size $n=256$.

Some simulation results of posterior consistency for causal and invertible ARMA process with i.i.d non-Gaussian innovation are reported in the supplementary material of Kirch et al. (2019), indicating that posterior consistency does hold, which corroborates our finding. Here, we complement these finding with a more extreme innovation sequence, namely an MA(1) process with i.i.d. innovation obeying a centred standard Cantor distribution (see e.g. Dovgoshey et al. 2006; Schmidt 1991):

$$X_n = \epsilon_n + \frac{1}{3}\epsilon_{n-1}\,, \quad \epsilon_k = \sum_{n=1}^{\infty} \frac{2B_{n,k}}{3^n} - \frac{1}{2},$$

where $\{B_{n,k}\}_{n,k=1}^{\infty}$ are i.i.d. Bernoulli random variables with expectation $1/2$. The standard Cantor distribution has a cumulative distribution function which is continuous on $[0,1]$, but not absolutely continuous with respect
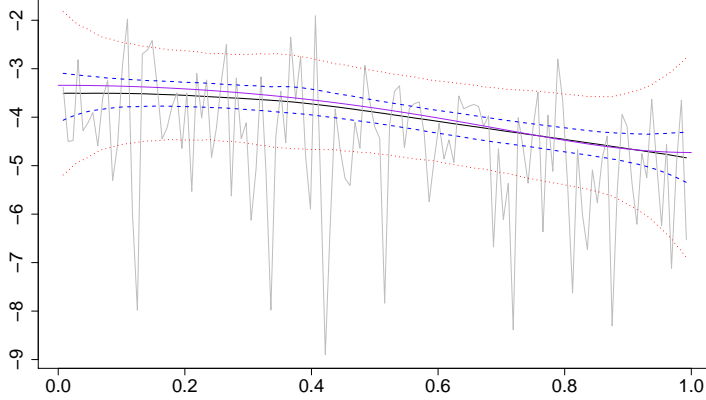
Figure 2: Log-periodogram of the time series in Figure 1, overlaid by the true log-spectral density functions scaled to $[0, 1]$ in solid purple, the estimated log-spectral density in solid black, pointwise 90% credible bands in dashed blue, and uniform 90% credible bands in dashed red.

to the Lebesgue measure (Dovgoshey et al., 2006, Proposition 2.1). Therefore, it is a so-called singular distribution (Dovgoshey et al., 2006, Remark 2.2), a property that is then inherited by X_n. Figure 1 shows a realization of this time series with length $n = 256$. In Figure 2, the corresponding estimated spectral density function is represented by the black solid line, the blue dashed lines represent 90% pointwise credible bands, the red dashed lines represent 90% uniform credible bands and the purple solid line is the true spectral density function. The y-axis in Figure 2 is on the log scale. The corresponding Bayesian analysis via an MCMC algorithm as described in Kirch et al. (2019) is conducted using the beyondWhittle R package (Meier et al., 2017).

4.2 Strong mixing sequences

A stationary time series $\{X_t\}_{t \in \mathbb{Z}}$ is said to be strong mixing if

$$\lim_{n \to \infty} \alpha(\mathcal{F}_0, \mathcal{F}^n) = 0$$

where $\mathcal{F}_0 = \sigma(X_k, k \leqslant 0)$, $\mathcal{F}^n = \sigma(X_k, k \geqslant n)$, and α is the strong mixing coefficient defined by

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup \{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{A}, B \in \mathcal{B}\}$$

for two σ-algebras $\mathcal{A}$ and $\mathcal{B}$.

By Bradley (2007, Theorem 2.14), any strong mixing sequence is also a regular sequence. Therefore, any strong mixing sequence with strictly
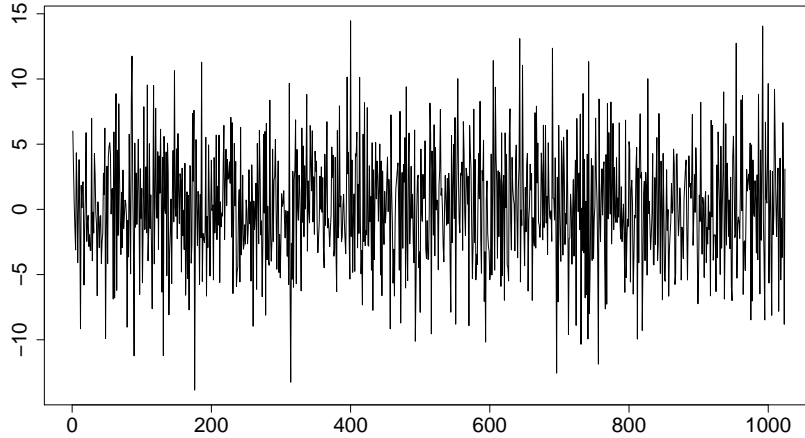
Figure 3: Realization of a Gaussian stationary time series with spectral density function (4.1) and sample size $n = 1024$.
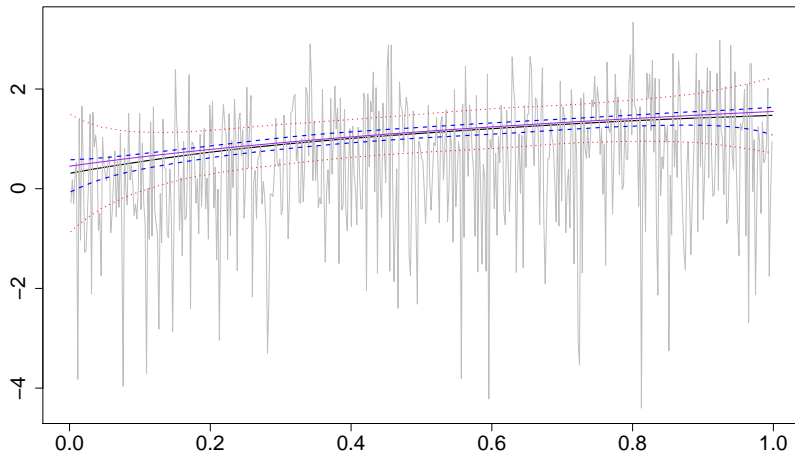

Figure 4: Log-periodogram of the time series in Figure 3, overlaid by true log-spectral density function scaled to $[0, 1]$ in solid purple, its estimate in solid black, and pointwise 90% credible bands in dashed blue, and uniform 90% credible bands in dashed red.

positive and Lipschitz continuous spectral density function satisfies all our assumptions. Many time series models used in practice, including countable-state ergodic Markov chains (Bradley, 2005) and some of the ARCH and GARCH models (Francq and Zakoian, 2019, Section 3.2), are strong mix-

14

ing. Specifically, a stationary Gaussian time series with strictly positive and Lipschitz continuous spectral density is also strong mixing (Ibragimov and Linnik, 1971, Section 3, Chapter 17). Nevertheless, this property does not hold even for some AR(1) models with discrete innovations, see e.g. Andrews (1984).

Figures 3 and 4 show the trace plot, the spectral density function and its estimate for a mean-zero Gaussian process with spectral density function

$$\varphi(x) = \pi|x| + \frac{\pi}{2}, \quad x \in [0, 1]. \tag{4.1}$$

The spectral density $\varphi(x)$ is chosen to be strictly positive and Lipschitz continuous but not continuously differentiable. Therefore, this time series satisfies our assumptions but not the assumptions in Choudhuri et al. (2004) and Kirch et al. (2019). In fact, φ_0 admits the following representation.

$$\varphi_0(\lambda) = \frac{1}{2\pi} \left(2\pi^2 - 8 \sum_{n=1}^{\infty} \frac{\cos(2n-1)\pi x}{(2n-1)^2} \right).$$

Hence, the auto-covariance sequence is

$$\gamma_0(h) = \begin{cases} 2\pi^2, & \text{if } h = 0. \\ 0, & \text{if } h \text{ is non-zero and even.} \\ -\frac{4}{h^2}, & \text{if } h \text{ is odd.} \end{cases}$$

such that for all $\alpha > 1$

$$\sum_{h \in \mathbb{Z}} |h|^{\alpha} |\gamma_0(h)| = 8 \sum_{n=1}^{\infty} (2n-1)^{\alpha-2} = \infty.$$

5 Proof of the posterior consistency

We start with verification of Assumptions 2.1 – 2.3 then show our main proof of Theorem 3.1 using Theorem 2.1.

Verification of Assumption 2.1

As a result of Proposition 3.1, (Θ, d_{∞}) is a separable metric space (Aliprantis and Border, 2006, Lemma 3.26). $\mathcal{T}$ is the corresponding Borel σ-algebra on Θ. Denote by $\mathcal{B}(\mathbb{R}^n)$ the Borel σ-algebra on $\mathbb{R}^n$. In order to prove that the proxy model $R_n(\varphi, \omega) = f_{\varphi}(x_1^n)$ defined in (3.3) is $\mathcal{T} \otimes \sigma(X_1^n)$-measurable for all n, it suffices to prove that for any fixed $n \in \mathbb{N}$, $R_n(\varphi, \omega)$ is a Carathéodory function (Aliprantis and Border, 2006, Lemma 4.51). That is,

C.1 for each $\omega \in \Omega$, $R_n(\cdot, \omega)$ is a continuous functional on Θ,

C.2 for each $\varphi \in \Theta$, $R_n(\varphi, \cdot)$ is $\sigma(X_1^n)$-measurable.

For any fixed n and ω the corrected likelihood defined in (3.3) is clearly a continuous functional in ϕ with respect to uniform convergence, confirming C.1.

Furthermore, as a composition of X_1^n and an n-dimensional normal density function f_φ, clearly $R_n(\varphi, \cdot)$ is also $\sigma(X_1^n)$-measurable, showing C.2.

Next, we will show that the inequalities in Assumption 2.1 are satisfied: Clearly, for each $\varphi \in \Theta$ and each $n \in \mathbb{N}$, $R_n(\varphi, \omega)$ is P-a.s. strictly positive noting that $(X_1^n)'X_1^n < \infty$ a.s. by the existence of second moments. This fact, along with Proposition 3.2, implies that

$$\int_\Theta R_n(\theta, \omega) \, \Pi_0(d\theta) > 0 \quad P\text{-a.s.}$$

Furthermore, because $\varphi \in \Theta$ has the uniform lower bound m, it holds that $\det(C_n^2(\varphi)) \leq \frac{1}{m} \det(D_n(\varphi_{\mathrm{par}})) < \infty$ such that

$$\int_\Theta R_n(\theta, \omega) \, \Pi_0(d\theta) < \infty \quad P\text{-a.s.}$$

Verification of Assumption 2.2

We will first verify the integrability of $\ln R_n(\varphi)$. By direct calculation, we have

$$\ln R_n(\varphi, \omega) = -\frac{n \ln(2\pi)}{2} + \frac{1}{2} \ln \det\left(\Gamma_{\mathrm{par}}^{-1} C_n^2(\varphi)\right) - \frac{1}{2}(x_1^n)' C_n(\varphi) \Gamma_{\mathrm{par}}^{-1} C_n(\varphi) x_1^n. \tag{5.1}$$

Therefore, for each $\varphi \in \Theta$ and each $n \in \mathbb{N}$,

$$\mathrm{E}\,|\ln R_n(\varphi)| \leqslant \frac{n \ln(2\pi)}{2} + \frac{1}{2}\left|\ln \det\left(\Gamma_{\mathrm{par}}^{-1} C_n^2(\varphi)\right)\right| + \frac{1}{2}\,\mathrm{E}\left|(X_1^n)' C_n(\varphi) \Gamma_{\mathrm{par}}^{-1} C_n(\varphi) X_1^n\right|.$$

Because $C_n(\varphi)$ and $\Gamma_{\mathrm{par}}^{-1}$ are positive definite, we have

$$\mathrm{E}\left|(X_1^n)' C_n(\varphi) \Gamma_{\mathrm{par}}^{-1} C_n(\varphi) X_1^n\right| = \mathrm{E}\left[(X_1^n)' C_n(\varphi) \Gamma_{\mathrm{par}}^{-1} C_n(\varphi) X_1^n\right]$$
$$= \mathrm{tr}\left(C_n(\varphi) \Gamma_{\mathrm{par}}^{-1} C_n(\varphi) \Gamma_n\right),$$

where $\Gamma_n = \mathrm{E}\left[(X_1^n)(X_1^n)'\right]$ is the autocovariance matrix of the time series.

We are now ready to calculate $h(\varphi)$. The functions $\varphi \in \Theta$, φ_0, φ_{par}, $\sqrt{\frac{\varphi_{\mathrm{par}}}{\varphi}}$, $\frac{\varphi}{\varphi_{\mathrm{par}}}$ as well as their products are all Lipschitz continuous due to Proposition B.1. Thus, by the Bernstein's Theorem (Katznelson, 2004, Theorem 6.3) they are in the class of functions with absolutely summable Fourier (i.e. autocovariance) functions (Wiener algebra). Let F be a continuous function, f from the above class and $\tau_{n,k}$, $k = 1, \cdots, n$, be the eigenvalues of an $n \times n$ matrix T_n. Furthermore, let $T_n \sim T_n(f)$ in the sense of Section

2.3 in Gray (2006), where $T_n(f)$ is the $n \times n$ Toeplitz matrix associated with f. Then, the fundamental eigenvalue distribution theorem of Szegö (cf. e.g. Gray 2006, Theorem 4.2), in conjunction with Theorem 2.4 of Gray (2006), yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} F(\tau_{n,k}) = \int_0^1 F(f(x))dx.$$

Theorems 5.2 and 5.3 of Gray (2006) imply:

$$\Gamma_{\text{par}} C_n^{-2}(\varphi) \sim T_n \left(2\pi\varphi_{\text{par}} \frac{2\pi\varphi}{2\pi\varphi_{\text{par}}} \right) = T_n(2\pi\varphi),$$

$$C_n(\varphi)\Gamma_{\text{par}}^{-1}C_n(\varphi)\Gamma_n \sim T_n \left(\sqrt{\frac{2\pi\varphi_{\text{par}}}{2\pi\varphi}} \frac{1}{2\pi\varphi_{\text{par}}} \sqrt{\frac{2\pi\varphi_{\text{par}}}{2\pi\varphi}} 2\pi\varphi_0 \right) = T_n \left(\frac{\varphi_0}{\varphi} \right).$$

Thus, a choice of $F(x) = \ln x$ yields

$$\lim_{n\to\infty} \frac{\ln \det \left(\Gamma_{\text{par}} C_n^{-2}(\varphi) \right)}{n} = \int_0^1 \ln \varphi(x)dx + \ln(2\pi),$$

while choosing $F(x) = x$ yields

$$\lim_{n\to\infty} \frac{1}{n} \text{tr} \left(C_n(\varphi)\Gamma_{\text{par}}^{-1}C_n(\varphi)\Gamma_n \right) = \int_0^1 \frac{\varphi_0(x)}{\varphi(x)} dx.$$

Consequently, for each $\varphi \in \Theta$, we have by (5.1)

$$h(\varphi) = -\lim_{n\to\infty} \frac{1}{n} \text{E} \left[\ln R_n(\varphi) \right]$$

$$= \ln(2\pi) + \frac{1}{2} \int_0^1 \ln \varphi_0(x)dx + \frac{1}{2} \int_0^1 \left(\frac{\varphi_0(x)}{\varphi(x)} - \ln \frac{\varphi_0(x)}{\varphi(x)} \right) dx. \quad (5.2)$$

Next, we show that h is a continuous functional on Θ by verifying the Lipschitz condition. For any $0 < u < v$, by the mean value theorem, there exists $w \in (u, v)$ such that

$$|(u - \ln u) - (v - \ln v)| = \left| 1 - \frac{1}{w} \right| |u - v| \leqslant \frac{v+1}{u} |u - v|.$$

For all $\xi \in \Theta$, we have $\frac{m}{M} \leqslant \frac{\varphi_0(x)}{\xi(x)} \leqslant \frac{M}{m}$ for all $0 \leqslant x \leqslant 1$. Therefore, for any $\phi, \xi \in \Theta$

$$|h(\xi) - h(\phi)| \leqslant \int_0^1 \left| \left(\frac{\varphi_0(x)}{\xi(x)} - \ln \frac{\varphi_0(x)}{\xi(x)} \right) - \left(\frac{\varphi_0(x)}{\phi(x)} - \ln \frac{\varphi_0(x)}{\phi(x)} \right) \right| dx$$

$$\leqslant \frac{\frac{M}{m} + 1}{\frac{m}{M}} \int_0^1 \varphi_0(x) \left| \frac{1}{\xi(x)} - \frac{1}{\phi(x)} \right| dx$$

$$= \frac{\frac{M}{m} + 1}{\frac{m}{M}} \int_0^1 \frac{\varphi_0(x)}{\xi(x)\phi(x)} |\xi(x) - \phi(x)| dx$$

$$\leqslant \frac{\left(\frac{M}{m} + 1 \right) M}{\frac{m}{M} m^2} \|\xi - \phi\|_\infty.$$

17

This indicates that h is continuous on Θ, which in turn implies h is $\mathcal{T}$-measurable.

Finally, we will show that $|h|$ is bounded from above: Using the inequality $y - \ln y \geqslant 1$ on $y > 0$, we have for all $\xi \in \Theta$,

$$1 \leqslant \int_0^1 \left(\frac{\varphi_0(x)}{\xi(x)} - \ln \frac{\varphi_0(x)}{\xi(x)} \right) dx \leqslant \frac{M}{m} + \ln \frac{M}{m},$$

such that

$$\sup_{\xi \in \Theta} |h(\xi)| \leqslant \ln(2\pi) + \frac{1}{2} \left| \int_0^1 \ln \varphi_0(x) dx \right| + \frac{1}{2} \left(\frac{M}{m} + \ln \frac{M}{m} \right) < \infty.$$

Verification of Assumption 2.3

Let

$$l_n(\varphi, \omega) \coloneqq \frac{1}{n} \ln R_n(\varphi, \omega) + h(\varphi).$$

We first show that for each $\varphi \in \Theta$, $l_n(\varphi) \xrightarrow{P} 0$ as $n \to \infty$, and then strengthen it to uniform convergence.

Since for any $\varphi \in \Theta$, φ and φ_{par} are both lower bounded away from 0 and upper bounded, the spectral norm of $C_n(\varphi)\Gamma_{\text{par}}^{-1}C_n(\varphi)$ is upper bounded uniformly over $n \in \mathbb{N}$ (Gray, 2006, Theorem 2.1 and Lemma 4.1). Therefore, by Corollary 1 of Yaskov (2018), we have

$$\left| \frac{1}{n} \left[(X_1^n)' C_n(\varphi) \Gamma_{\text{par}}^{-1} C_n(\varphi) X_1^n \right] - \frac{1}{n} E \left[(X_1^n)' C_n(\varphi) \Gamma_{\text{par}}^{-1} C_n(\varphi) X_1^n \right] \right| \xrightarrow{P} 0.$$

In view of equations (5.1) and (5.2), the above result yields $|l_n(\varphi)| \xrightarrow{P} 0$.

It is straightforward to show that $\sup_{\varphi \in \Theta} |l_n(\varphi, \omega)|$ is $\sigma(X_1^n)$-measurable. Since $R_n(\varphi, \omega)$ is a Carathéodory function and $h(\varphi)$ is $\mathcal{T}$-measurable, $|l_n(\varphi, \omega)|$ is also a Carathéodory function. According to the measurable maximum theorem (Aliprantis and Border, 2006, Theorem 18.19), $\sup_{\varphi \in \Theta} |l_n(\varphi, \omega)|$ is $\sigma(X_1^n)$-measurable.

We now show that

$$\sup_{\varphi \in \Theta} |l_n(\varphi)| \xrightarrow{P} 0, \ n \to \infty.$$

By Proposition 3.1, we know that Θ is compact with respect to d_∞. According to Davidson (1994, Theorem 21.9), we can strengthen the pointwise convergence to uniform convergence by showing that $l_n(\varphi)$ is asymptotically uniformly stochastically equicontinuous. In order to prove equicontinuity, it

suffices to show that $l_n(\varphi)$ satisfies a Lipschitz-type condition on Θ (David-son, 1994, Theorem 21.10). That is, there exists $S \in \mathbb{N}$ such that for any $\phi, \psi \in \Theta$ and $n \geqslant S$,

$$|l_n(\phi) - l_n(\psi)| \leqslant B_n \|\phi - \psi\|_\infty, \qquad (5.3)$$

where $B_n = O_p(1)$. Since the Lipschitz continuity of $h(\varphi)$ has been verified in the preceding subsection, we now prove that $\frac{1}{n} \ln R_n(\theta, \omega)$ satisfies (5.3). In light of equation (5.1), we have

$$\frac{1}{n} \ln R_n(\theta, \omega) = -\frac{\ln(2\pi)}{2} - \frac{1}{2} \frac{\ln \det \left(\Gamma_{\text{par}} C_n^{-2}(\varphi) \right)}{n} - \frac{1}{2} \frac{(X_1^n)' C_n(\varphi) \Gamma_{\text{par}}^{-1} C_n(\varphi) X_1^n}{n}$$

$$=: -\frac{\ln(2\pi)}{2} - \frac{1}{2} l_n^{(1)}(\varphi) - \frac{1}{2} l_n^{(2)}(\varphi, \omega).$$

It suffices to prove that $l_n^{(i)}$, $i = 1, 2$ satisfy the Lipschitz condition on Θ.

For all $x, y > 0$, as per the mean value theorem, there exists $z \in (\min\{x, y\}, \max\{x, y\})$ such that

$$|\ln x - \ln y| = \frac{1}{z} |x - y| \leqslant \frac{1}{\min\{x, y\}} |x - y|.$$

Consequently, for any $\phi, \xi \in \Theta$, $n \in \mathbb{N}$,

$$\left| l_n^{(1)}(\phi) - l_n^{(1)}(\xi) \right| = \frac{1}{n} \left| \ln \det C_n^{-2}(\phi) - \ln \det C_n^{-2}(\xi) \right| = \frac{1}{n} \left| \sum_{j=1}^{N} (\ln \phi(y_j) - \ln \xi(y_j)) \right|$$

$$\leqslant \frac{1}{n} \sum_{j=1}^{N} \frac{1}{\min\{\phi(y_j), \xi(y_j)\}} |\phi(y_j) - \xi(y_j)| \leqslant \frac{1}{2m} \max_{j=1,2,\cdots,N} |\phi(y_j) - \xi(y_j)|$$

$$\leqslant \frac{1}{2m} \|\phi - \xi\|_\infty.$$

Furthermore,

$$\left| l_n^{(2)}(\phi) - l_n^{(2)}(\xi) \right| = \frac{1}{n} \left| (X_1^n)' \left(C_n(\phi) \Gamma_{\text{par}}^{-1} C_n(\phi) - C_n(\xi) \Gamma_{\text{par}}^{-1} C_n(\xi) \right) (X_1^n) \right|$$

$$= \frac{1}{n} \left\| (X_1^n)' \left(C_n(\phi) \Gamma_{\text{par}}^{-1} C_n(\phi) - C_n(\xi) \Gamma_{\text{par}}^{-1} C_n(\xi) \right) (X_1^n) \right\|_s$$

$$\leqslant \|C_n(\phi) \Gamma_{\text{par}}^{-1} C_n(\phi) - C_n(\xi) \Gamma_{\text{par}}^{-1} C_n(\xi)\|_s \frac{\sum_{k=1}^{n} X_k^2}{n},$$

where $\| \cdot \|_s$ denotes the matrix spectral norm and the last inequality is due to the submultiplicativity of the spectral norm (see e.g. Gloube and van Loan 1996, formula (2.3.3)). By the triangle inequality, we have

$$\|C_n(\phi) \Gamma_{\text{par}}^{-1} C_n(\phi) - C_n(\xi) \Gamma_{\text{par}}^{-1} C_n(\xi)\|_s$$

$$\leqslant \|(C_n(\phi) - C_n(\xi)) \Gamma_{\text{par}}^{-1} C_n(\phi)\|_s + \|C_n(\xi) \Gamma_{\text{par}}^{-1} (C_n(\phi) - C_n(\xi))\|_s.$$

For the first term of the right hand side, again by the submultiplicativity of the spectral norm, we get

$$\|(C_n(\phi) - C_n(\xi))\Gamma_{\text{par}}^{-1}C_n(\phi)\|_s \leqslant \|C_n(\phi) - C_n(\xi)\|_s \|\Gamma_{\text{par}}^{-1}\|_s \|C_n(\phi)\|_s.$$

According to Gray (2006, Theorem 2.1 and Lemma 4.1), $\|\Gamma_{\text{par}}^{-1}\|_s \leqslant \frac{1}{\beta_{\text{par}}}$. By direct calculation, we have

$$\|C_n(\phi) - C_n(\xi)\|_s = \max_{j=1,\cdots,N} \left| \varphi_{\text{par}}^{\frac{1}{2}}(y_j) \left(\phi^{-\frac{1}{2}}(y_j) - \xi^{-\frac{1}{2}}(y_j) \right) \right|$$

$$= \max_{j=1,\cdots,N} \frac{\varphi_{\text{par}}^{\frac{1}{2}}(y_j)}{\phi^{\frac{1}{2}}(y_j)\xi^{\frac{1}{2}}(y_j) \left(\xi^{\frac{1}{2}}(y_j) + \phi^{\frac{1}{2}}(y_j) \right)} |\xi(y_j) - \phi(y_j)|$$

$$\leqslant \frac{\|\varphi_{\text{par}}^{\frac{1}{2}}\|_\infty}{2m^{\frac{3}{2}}} \|\xi - \phi\|_\infty$$

and

$$\|C_n(\phi)\|_s = \max_{j=1,\cdots,N} \frac{\varphi_{\text{par}}^{\frac{1}{2}}(y_j)}{\phi^{\frac{1}{2}}(y_j)} \leqslant \frac{\|\varphi_{\text{par}}^{\frac{1}{2}}\|_\infty}{m}.$$

Therefore,

$$\|(C_n(\phi) - C_n(\xi))\Gamma_{\text{par}}^{-1}C_n(\phi)\|_s \leqslant \frac{\|\varphi_{\text{par}}\|_\infty}{2\beta_{\text{par}}m^{\frac{5}{2}}} \|\xi - \phi\|_\infty.$$

By the same token,

$$\|C_n(\xi)\Gamma_{\text{par}}^{-1}(C_n(\phi) - C_n(\xi))\|_s \leqslant \frac{\|\varphi_{\text{par}}\|_\infty}{2\beta_{\text{par}}m^{\frac{5}{2}}} \|\xi - \phi\|_\infty.$$

Putting things together, we have

$$\left| l_n^{(2)}(\phi) - l_n^{(2)}(\xi) \right| \leqslant \frac{\sum_{k=1}^n X_k^2}{n} \frac{\|\varphi_{\text{par}}\|_\infty}{\beta_{\text{par}}m^{\frac{5}{2}}} \|\xi - \phi\|_\infty = O_p(1)\|\xi - \phi\|_\infty,$$

where the last equation is due to Corollary 1 of Yaskov (2018).

Since $l_n^{(i)}$, $i = 1, 2$ satisfy the Lipschitz condition, we have shown that $l_n = -\frac{\ln(2\pi)}{2} - \frac{1}{2}l_n^{(1)} - \frac{1}{2}l_n^{(2)} + h$ does indeed satisfy the Lipschitz-type condition (5.3). Therefore, according to Davidson (1994, Theorem 21.10), l_n is asymptotically uniformly stochastically equicontinuous in probability over Θ.

Proof of posterior consistency

Our goal is to show that for any $r > 0$, $\Pi_n(B^c(\varphi_0, r)) \to 0$ in P-probability. If $\Pi_0(B^c(\varphi_0, r)) = 0$, then clearly we have $\Pi_n(B^c(\varphi_0, r)) = 0$ for any $n \in \mathbb{N}$. If $\Pi_0(B^c(\varphi_0, r)) > 0$, then according to Theorem 2.1, we only need to show

$$h\left(B^c(\varphi_0, r)\right) = \operatorname*{ess\,inf}_{\varphi \in B^c(\varphi_0, r)} h(\varphi) > h(\Theta) = \operatorname*{ess\,inf}_{\varphi \in \Theta} h(\varphi),$$

where here and throughout the essential infimum is with respect to the prior measure Π_0.

First, notice that the continuity of h on Θ has been proved when verifying Assumption 2.2. We will now focus on proving that h has a unique minimum point at φ_0. Since for any $y > 0$, it holds $y - \ln y - 1 \geqslant 0$ with equality if and only if $y = 1$, and Θ is a collection of continuous functions on $[0, 1]$, we have for any $\varphi \in \Theta$,

$$\frac{\varphi_0}{\varphi} - \ln \frac{\varphi_0}{\varphi} - 1$$

is a non-negative continuous function on $[0, 1]$. Hence,

$$\int_0^1 \left(\frac{\varphi_0(x)}{\varphi(x)} - \ln \frac{\varphi_0(x)}{\varphi(x)} - 1 \right) dx \geqslant 0,$$

where the equality holds if and only if $\varphi = \varphi_0$. By virtue of equation (5.2), we get

$$\varphi_0 = \operatorname*{argmin}_{\varphi \in \Theta} h(\varphi).$$

Now we will prove that $h(\Theta) = h(\varphi_0)$. For any $z > h(\varphi_0)$, by the continuity of h, there exists $\delta > 0$ such that $h(\varphi_0) \leqslant h(\varphi) < z$, for all $\varphi \in B(\varphi_0, \delta)$. Consequently, $\Pi_0(\{\varphi \in \Theta : h(\varphi) < z\}) \geqslant \Pi_0(B(\varphi_0, \delta)) > 0$ by Proposition 3.2 showing that indeed $h(\Theta) = h(\varphi_0)$.

Finally, we will show that $h\left(B^c(\varphi_0, r)\right) > h(\Theta)$ for all $r > 0$. It is clear that $B^c(\varphi_0, r)$, as a closed subset of a compact set Θ is also compact. Since h is continuous on $B^c(\varphi_0, r)$, it can attain its local minimum on this set. i.e.

$$h\left(B^c(\varphi_0, r)\right) \geqslant \min_{\varphi \in B^c(\varphi_0, r)} h(\varphi) > h(\varphi_0) = h(\Theta).$$

Theorem 3.1 now follows from Theorem 2.1.

6   Discussion and outlook

The general consistency theorem introduced in this paper does not rely on the KL formulation which is otherwise popular in the literature. It has the capability to deal with situations where the proxy model does not approximate the data generating process sufficiently well with respect to the KL

divergence. Examples of such time series with singular marginal distributions are presented in subsection 4.1. Since the proxy model in this case corresponds to a non-degenerate Gaussian probability measure, the Radon-Nikodym derivative and consequently the KL divergence between these two measures is not defined. On the other hand, these time series meet the assumptions of this paper so that the proposed methodology is applicable. Moreover, the proxy model is not required to be tied to a probability distribution, it has the potential to deal with generalized Bayesian updating such as Gibbs posterior distributions.

The main objective in this paper was to make as weak assumptions on the time series as possible. Indeed posterior consistency is proven without invoking stronger assumptions such as the existence of higher order cumulants and certain complicated dependence structures which are commonly seen in the literature on nonparametric Bayesian (and frequentist) spectral density estimation (e.g. Choudhuri et al. 2004; Pawitan and O'sullivan 1994; Kakizawa 2006).

Mild assumptions on the time series structure, however, require stronger assumptions elsewhere. Similar to (3.2) of Meier et al. (2020) and (15) of Gao and Zhou (2016), the parameter space in this paper is chosen to be compact under a pre-specified topology which is a common assumption. This restriction may be relaxed by using a sieve defined on the parameter space of positive continuous functions on [0,1] endowed with the sup-norm. Nevertheless, this comes at the cost of stronger assumptions on the time series. To elaborate, suppose we use $\Theta = \{f \in C([0,1]) : f > 0\}$ as the parameter space and a suitable sequence of sets $\Theta_n = \{f \in C([0,1]) : f \geqslant m_n, \|f\|_{\text{Lip}} \leqslant M_n\}$ that approximate Θ well for growing n where the covering number of Θ_n needs to be controlled. The latter often relies on some sort of concentration inequality which in turn requires the existence of higher order moments and certain dependence structures (e.g. linear process, mixing condition, etc.). These are exactly the kind of assumptions which we aim to avoid. Such concentration inequalities are also typically required to obtain posterior contraction rates.

It is of future interest to establish a posterior contraction rate under minimal constraints on the time series structure. Another direction of future research, which may be of practical interest, is to generalize the results to priors other than the Bernstein-Dirichlet process prior used in this paper. As the proof relies mainly on the uniform approximation property of the Bernstein polynomials, we conjecture that it can be generalized as long as the basis functions have the uniform approximation property. Some examples are priors based on B-spline (Edwards et al. 2019 and Belitser and Serra 2014) and wavelet basis (Giné and Nickl, 2011) functions. Furthermore, our future research will also aim to extend the consistency proof to the estimation of the Hermitian positive definite spectral density matrix of a non-Gaussian multivariate time series using a matrix-Gamma process prior

(Meier, 2018).

A   Proof of Theorem 2.1

The proofs in this section are closely related to the proofs for Lemmas 1, 2, 3, 6 and Theorem 1 in Shalizi (2009).

Lemma A.1. *Let $Q^n \subseteq \Theta \times \Omega$, $n = 1, 2, \cdots$ be a sequence of jointly measurable sets, with sections $Q_{\theta,\bullet}^n = \{\omega \in \Omega : (\theta, \omega) \in Q^n\}$ and $Q_{\bullet,\omega}^n = \{\theta \in \Theta : (\theta, \omega) \in Q^n\}$.*

If, for some probability measure P on Ω,

$$\lim_{n \to \infty} P\left(Q_{\theta,\bullet}^n\right) = 1, \ \forall \theta \in \Theta.$$

Then for any probability measure Π on Θ,

$$\Pi\left(Q_{\bullet,\omega}^n\right) \xrightarrow{P} 1.$$

Proof. As in the proof of Lemma 1 in Shalizi (2009), we obtain by Tonelli's theorem

$$\int_\Theta P(Q_{\theta,\bullet}^n)\Pi(d\theta)) = \mathrm{E}\,\Pi(Q_{\bullet,\omega}^n),$$

such that by the dominated convergence theorem it holds

$$\lim_{n \to \infty} \mathrm{E}\,|\Pi(Q_{\bullet,\omega}^n) - 1| = 1 - \int_\Theta \lim_{n \to \infty} P(Q_{\theta,\bullet}^n)\Pi(d\theta) = 0,$$

which implies

$$\Pi\left(Q_{\bullet,\omega}^n\right) \xrightarrow{P} 1.$$

$\square$

Lemma A.2. *Under Assumptions 2.1, 2.2 and 2.3, for any $\eta > 0$ it holds $\Pi_0\left(Q_{\bullet,\omega}^n(\eta)\right) \xrightarrow{P} 1$, where $Q_{\bullet,\omega}^n(\eta)$ is the ω-section of*

$$Q^n(\eta) := \left\{(\theta, \omega) \in \Theta \times \Omega : \frac{1}{n} \ln R_n(\theta, \omega) \geqslant -h(\theta) - \eta\right\}.$$

Proof. For any $\eta > 0$ and $n \in \mathbb{N}$, $Q^n(\eta)$ is jointly measurable according to Assumptions 2.1 and 2.2. Furthermore, Assumption 2.3 implies that for all $\theta \in \Theta$

$$\lim_{n \to \infty} P\left(Q_{\theta,\bullet}^n(\eta)\right) = 1,$$

such that by Lemma A.1, we have

$$\Pi_0\left(Q_{\bullet,\omega}^n(\eta)\right) \xrightarrow{P} 1.$$

$\square$

Lemma A.3. *Under Assumptions 2.1, 2.2 and 2.3, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : \Pi_0\left(R_n\right) > \exp\{-n\left(h\left(\Theta\right) + \epsilon\right)\}\}\right) = 1,$$

where $\Pi_0\left(R_n\right) := \int_\Theta R_n(\theta, \omega)\Pi_0(d\theta)$.

Proof. For any $\omega \in \Omega$, we have

$$\Pi_0(R_n) > \exp\{-n\left(h\left(\Theta\right) + \epsilon\right)\} \Leftrightarrow \int_\Theta \exp\left\{n\left(h\left(\Theta\right) + \epsilon + \frac{\ln R_n(\theta)}{n}\right)\right\}\Pi_0(d\theta) > 1.$$

Consequently, it is sufficient to prove

$$\lim_{n\to\infty} P\left(\left\{\omega \in \Omega : \int_\Theta \exp\left\{n\left(h\left(\Theta\right) + \epsilon + \frac{\ln R_n(\theta)}{n}\right)\right\}\Pi_0(d\theta) > 1\right\}\right) = 1.$$
(A.1)

For $N_{\frac{\epsilon}{2}} = \{\theta : h(\theta) \leqslant h(\Theta) + \frac{\epsilon}{2}\}$ it holds for any $\theta \in N_{\frac{\epsilon}{2}} \cap Q^n_{\bullet,\omega}\left(\frac{\epsilon}{3}\right)$ by definition of these sets

$$\frac{\ln R_n(\theta, \omega)}{n} \geqslant -h(\theta) - \frac{\epsilon}{3} \geqslant -h(\Theta) - \frac{5}{6}\epsilon.$$

Consequently,

$$\int_\Theta \exp\left\{n\left(h\left(\Theta\right) + \epsilon + \frac{\ln R_n(\theta)}{n}\right)\right\}\Pi_0(d\theta) \geqslant \Pi_0\left(N_{\frac{\epsilon}{2}} \cap Q^n_{\bullet,\omega}\left(\frac{\epsilon}{3}\right)\right)\exp\left\{\frac{\epsilon}{6}n\right\}$$
(A.2)

According to Lemma A.2,

$$\Pi_0\left(\Theta \setminus Q^n_{\bullet,\omega}\left(\frac{\epsilon}{3}\right)\right) \xrightarrow{P} 0$$

and by definition that $\Pi_0\left(N_{\frac{\epsilon}{2}}\right) > 0$, such that

$$\Pi_0\left(N_{\frac{\epsilon}{2}} \cap Q^n_{\bullet,\omega}\left(\frac{\epsilon}{3}\right)\right) \geqslant \Pi_0\left(N_{\frac{\epsilon}{2}}\right) - \Pi_0\left(\Theta \setminus Q^n_{\bullet,\omega}\left(\frac{\epsilon}{3}\right)\right) \xrightarrow{P} \Pi_0\left(N_{\frac{\epsilon}{2}}\right) > 0.$$

This concludes the proof in combination with (A.1) and (A.2). ☐

Lemma A.4. *Let Assumptions 2.1, 2.2 and 2.3 hold, and take any set $G \in \mathcal{T}$ such that $\Pi_0(G) > 0$. Then for any $\epsilon > 0$,*

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : \Pi_0\left(GR_n\right) \leqslant \exp\{n\left(-h(G) + \epsilon\right)\}\}\right) = 1,$$

where $\Pi_0\left(GR_n\right) = \int_G R_n(\theta, \omega)\Pi_0(d\theta)$.

Proof. For any $\omega \in \Omega$ and sufficiently large n such that

$$\sup_{\theta \in G} \left| \frac{1}{n} \ln \left(R_n(\theta, \omega) \right) + h(\theta) \right| \leqslant \epsilon,$$

we have

$$\Pi_0 \left(GR_n \right) = \int_G \exp\{\ln(R_n(\theta, \omega))\} \Pi_0(d\theta) \leqslant \int_G \exp\left\{ n \left(-h(\theta) + \epsilon \right) \right\} \Pi_0(d\theta)$$
$$\leqslant \exp\left\{ n \left(-h(G) + \epsilon \right) \right\} \Pi_0(G) \leqslant \exp\left\{ n \left(-h(G) + \epsilon \right) \right\}.$$

$\square$

With these lemmas, we can proceed to proving Theorem 2.1.

Proof of Theorem 2.1. By the assumption that $h(A) > h(\Theta)$, we can choose ϵ such that $h(\Theta) - h(A) + 2\epsilon < 0$. Then, on the intersection of the asymptotic 1-sets of Lemmas A.3 and A.4 it holds

$$\Pi_n(A) = \frac{\Pi_0(AR_n)}{\Pi_0(R_n)} \leqslant \exp\left\{ n \left(h(\Theta) - h(A) + 2\epsilon \right) \right\} \to 0,$$

such that indeed $\Pi_n(A) \overset{P}{\to} 0$ for $n \to \infty$.

$\square$

B   Parameter space, Lipschitz functions and Bernstein polynomials

In this section, we will first prove Proposition 3.1, and then give some results regarding Lipschitz functions and Bernstein polynomials.

Proof of Proposition 3.1. According to the Arzelà-Ascoli theorem, the compactness of Θ is guaranteed if it is bounded, closed, and equicontinuous. The boundedness of Θ with respect to $\|\cdot\|_\infty$ follows directly from the definition of Θ. The equicontinuity follows immediately from the uniform boundedness of the Lipschitz constants over Θ.

To prove that Θ is closed, consider an arbitrary sequence $\{\varphi_n\}_{n=1}^\infty$ of functions in Θ which converge uniformly to $\varphi \in C([0,1])$. It is then sufficient to show that $\varphi \in \Theta$. By the triangle inequality it follows from $\|\varphi_n - \varphi\|_\infty \to 0$, that $\|\varphi_n\|_\infty \to \|\varphi\|_\infty$. Thus, for any $x \in [0,1]$,

$$\varphi(x) = \lim_{n \to \infty} \varphi_n(x) \geqslant m.$$

Furthermore,

$$\|\varphi\| + L(\varphi) = \lim_{n \to \infty} \|\varphi_n\| + \sup_{0 \leqslant x \neq y \leqslant 1} \lim_{n \to \infty} \frac{|\varphi_n(x) - \varphi_n(y)|}{|x - y|}$$
$$= \sup_{0 \leqslant x \neq y \leqslant 1} \lim_{n \to \infty} \left(\|\varphi_n\| + \frac{|\varphi_n(x) - \varphi_n(y)|}{|x - y|} \right) \leqslant \limsup_{n \to \infty} \|\varphi_n\|_{\mathrm{Lip}} \leqslant M.$$

$\square$

Results regarding Lipschitz norm and the Bernstein polynomials

Let
$$\mathrm{Lip}([0,1]) \coloneqq \{\varphi \in C([0,1]) : \|\varphi\|_{\mathrm{Lip}} < \infty\}.$$

The following inequalities regarding the Lipschitz norm will be useful in the proof of positivity of prior probabilities in Appendix C.

Proposition B.1.

(a) If $f \in \mathrm{Lip}([0,1])$ and $f > 0$, then we have $\frac{1}{f}, f^\delta \in \mathrm{Lip}([0,1])$ for any $0 \leqslant \delta \leqslant 1$.

(b) For any $f, g \in \mathrm{Lip}([0,1])$, we have $fg \in \mathrm{Lip}([0,1])$. Furthermore, $\|fg\|_{\mathrm{Lip}} \leqslant \|f\|_{\mathrm{Lip}} \|g\|_{\mathrm{Lip}}$.

Proof. The proof of (a) is a simple consequence of Theorems 12.5 and 12.6 of Eriksson et al. (2004). The proof of (b) can be found in Dudley (2004, Proposition 11.2.1) □

The general n-th order Bernstein polynomial is of the form
$$b(x; n, w) = \sum_{i=1}^{n} w_i \beta(x; i, n - i + 1), \ x \in [0, 1],$$

where $w = (w_1, \cdots, w_n) \in \mathbb{R}^n$ and $\beta(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$, $x \in [0, 1]$, $a, b \geqslant 0$.

Proposition B.2. *For any $w = (w_1, \cdots, w_n) \in \mathbb{R}^n$, we have*

(a) $\|b(\cdot; n, w)\|_\infty \leqslant n \max_{i=1,2,\cdots,n} |w_i|$.

(b) $L(b(\cdot; n, w)) \leqslant 2n^2 \max_{i=1,2,\cdots,n} |w_i|$.

Proof. The first inequality is proved in Lemma 2.2 of Kruijer (2008, p. 21). We will focus on proving the second inequality. By direct calculation, we obtain
$$b'(x; n, w) = n \sum_{i=1}^{n-1} (w_{i+1} - w_i) \beta(x; i, n - i)$$

with the convention $\sum_{i=1}^{0} = 0$. By noticing that $\sum_{i=1}^{n-1} \beta(x; i, n - i) = n - 1$, we have

$$|b'(x; k, w)| \leqslant n \sum_{i=1}^{n-1} |w_{i+1,n} - w_{i,n}| \beta(x; i, n - i)$$

$$\leqslant n \max_{i=1,\cdots,n-1} \{|w_{i+1} - w_i|\} \sum_{i=1}^{n-1} \beta(x; i, n - i) = n(n-1) \max_{i=1,\cdots,n-1} \{|w_{i+1} - w_i|\}$$

26

for all $0 \leqslant x \leqslant 1$. Therefore,

$$L(b(\cdot; n, w)) = \left\| b'(\cdot; n, w) \right\|_\infty \leqslant n(n-1) \max_{i=1,\cdots,n-1} \{|w_{i+1} - w_i|\} \leqslant 2n^2 \max_{i=1,2,\cdots,n} |w_i|.$$

$\square$

Let $f \in \mathrm{Lip}([0,1])$ be a probability density function with corresponding cumulative distribution function F, then the n-th order (Kantorovich-type) Bernstein polynomial associated with f is defined as

$$b(x; n, w_n(F)) = \sum_{i=1}^{n} w_{i,n}(F)\beta(x; i, n-i+1), \ x \in [0,1],$$

where $w_{i,n}(F) = F\left(\frac{i}{n}\right) - F\left(\frac{i-1}{n}\right)$, $i = 1, 2, \cdots, n$, and $w_n(F) = (w_{1,n}(F), w_{2,n}(F), \cdots, w_{n,n}(F))$.

Proposition B.3. *We have*

(a) $\| f - b\left(\cdot; n, w_n(F)\right) \|_\infty \leqslant \frac{3L(f)}{\sqrt{n}}$.

(b) If f is continuously differentiable, then $\| f - b\left(\cdot; n, w_n(F)\right) \|_{\mathrm{Lip}} \to 0$.

(c) $\| b\left(\cdot; n, w_n(F)\right) \|_{\mathrm{Lip}} \leqslant \| f \|_{\mathrm{Lip}}$.

Proof. (a) is proved in Lemma 2.1 of Kruijer (2008, p. 20). (b) is a direct consequence of Theorem 2.4 of Özarslan and Duman (2016). As for (c), we have by definition

$$w_{i,n}(F) = F\left(\frac{i}{n}\right) - F\left(\frac{i-1}{n}\right) = \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(x)dx.$$

Thus, we get $w_{i,n}(F) \leqslant \frac{1}{n}\|f\|_\infty$, $i = 1, 2, \cdots, n$, which implies $\|b(\cdot; k, w_n(F))\|_\infty \leqslant \|f\|_\infty$ in light of the fact that $\sum_{i=1}^{n} \beta(x; i, n-i+1) = n$.

As for the bound of Lipschitz constant, because for any $i = 1, 2, \cdots, n-1$, we have

$$|w_{i+1,n}(F) - w_{i,n}(F)| \leqslant \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left| f\left(x + \frac{1}{n}\right) - f(x) \right| dx \leqslant \frac{L(f)}{n^2}.$$

Therefore, for all $x \in [0,1]$,

$$|b'(x; k, w_n(F))| = n \left| \sum_{i=1}^{n-1} (w_{i+1,n}(F) - w_{i,n}(F))\beta(x; i, n-i) \right| \leqslant L(f),$$

which implies $L(b(\cdot; k, w_n(F))) \leqslant L(f)$. Together this yields (c). $\square$

C Proof of Proposition 3.2

Let

$$\tau_0 = \int_0^1 \frac{\varphi_0(x)}{\varphi_{\text{par}}^\delta(x)} dx,$$

and $q_0 = \frac{1}{\tau_0} \frac{\varphi_0}{\varphi_{\text{par}}^\delta}$. Then q_0 is a probability density function on $[0,1]$. Denote by Q_0 the cumulative distribution function corresponding to q_0. Our goal is to show that $\Pi_C\left(B(\varphi_0, r)\right) > 0$ for all $r > 0$. To this end, consider

$$B_{M_1,M_2,M_3} = \Big\{ \tau b\left(\cdot; k, w_k(G)\right) \varphi_{\text{par}}^\delta :$$

$$k = M_1, |\tau - \tau_0| \leqslant M_2, \max_{i=1,2,\cdots,k} |w_{i,k}(G) - w_{i,k}(Q_0)| \leqslant M_3 \Big\}$$

We will show that (a) $\Pi_C(B_{M_1,M_2,M_3}) > 0$ for all $M_1 = 1, 2, \ldots$ and $M_2, M_3 > 0$ and (b) that there exist such M_1, M_2, M_3 that $B_{M_1,M_2,M_3} \subset B(\varphi_0, r)$ for any $r > 0$.

Recall that ρ, which is the probability mass function of K, has full support. Clearly, we have $\rho(M_1) > 0$. Since both the distribution of τ and the Dirichlet process have full support (see e.g. Choudhuri et al. 2004, Appendix B.1), the following two sets:

$$\{\tau \in (0,\infty) : |\tau - \tau_0| \leqslant M_2\}$$

and

$$\Big\{ G : \max_{i=1,2,\cdots,M_1} |w_{i,M_1}(G) - w_{i,M_1}(Q_0)| \leqslant M_3 \Big\}$$

also have positive probabilities. Therefore, $\Pi_C(B_{M_1,M_2,M_2}) > 0$, showing (a).

For (b), first note that by the triangle inequality, Propositions B.2(a) and B.3(a),

$$\|\tau b(\cdot; k, w_k(G))\varphi_{\text{par}}^\delta - \tau_0 q_0 \varphi_{\text{par}}^\delta\|_\infty$$

$$\leqslant \tau \|b(\cdot; k, w_k(G) - w_k(Q_0))\|_\infty \|\varphi_{\text{par}}\|_\infty^\delta + |\tau - \tau_0| \|b(\cdot; k, w_k(Q_0))\|_\infty \|\varphi_{\text{par}}\|_\infty^\delta$$

$$+ \tau_0 \|b(\cdot; k, w_k(Q_0)) - q_0\|_\infty \|\varphi_{\text{par}}\|_\infty^\delta$$

$$\leq (\tau_0 + M_2) M_1 M_3 \|\varphi_{\text{par}}\|_\infty^\delta + M_2 \|q_0\|_\infty \|\varphi_{\text{par}}\|_\infty^\delta + \tau_0 \frac{L(q_0)}{\sqrt{M_1}},$$

which by Propositions B.1(a) can be made arbitrarily small by first choosing an M_1 large enough for the last term to become small as well as an M_2 small enough for the second term to become small and finally (dependent on M_1 and M_2) an M_3 small enough for the first term to become small.

This shows in particular that for any $\epsilon > 0$ there exists M_1 big and M_2, M_3 small enough that $\|\varphi - \varphi_0\|_\infty < \epsilon$ for any $\varphi \in B_{M_1,M_2,M_3}$ as well

as $\varphi(x) \geq \varphi_0(x) - \|\varphi - \varphi_0\|_\infty > m$ for ϵ small enough by Assumption 3.1. In particular, for any $r > 0$ it holds $\|\varphi - \varphi_0\|_\infty < r$ and φ fulfills the lower bound in (3.4). We will now complete the proof that for sufficiently large M_1 and small M_2, M_3 it holds that $\|\varphi\|_{\text{Lip}}$ fulfills the upper bound. Indeed, we get by the triangle inequality, Propositions B.1(b) and B.2

$$\|\tau b(\cdot; k, w_k(G))\varphi_{\text{par}}^\delta\|_{\text{Lip}}$$
$$\leqslant \|\tau b(\cdot; k, w_k(Q_0))\varphi_{\text{par}}^\delta\|_{\text{Lip}} + \|\tau b(\cdot; k, w_k(G) - w_k(Q_0))\varphi_{\text{par}}^\delta\|_{\text{Lip}}$$
$$\leqslant \left(1 + \frac{M_2}{\tau_0}\right) \left\|\tau_0 b(\cdot; k, w_k(Q_0))\varphi_{\text{par}}^\delta\right\|_{\text{Lip}} + 3(\tau_0 + M_2)M_1^2 M_3\|\varphi_{\text{par}}^\delta\|_{\text{Lip}}.$$

The second part becomes again arbitrarily small for appropriately chosen M_j, $j = 1, 2, 3$, such that the upper bound of (3.4) can be met as soon as $\|\tau_0 b(\cdot; k, w_k(Q_0))\varphi_{\text{par}}^\delta\|_{\text{Lip}} < M$.

Indeed, by Propositions B.1(b) and B.3(c), we have

$$\|\tau_0 b(\cdot; k, w_k(Q_0))\varphi_{\text{par}}^\delta\|_{\text{Lip}} \leq \tau_0 \|q_0\|_{\text{Lip}} \|\varphi_{\text{par}}^\delta\|_{\text{Lip}} = \left\|\frac{\varphi_0}{\varphi_{\text{par}}^\delta}\right\| \|\varphi_{\text{par}}^\delta\|_{\text{Lip}}, \quad \text{(C.1)}$$

such that the assertion follows from Assumption 3.1.

Proof of Remark 3.4. In this case, we can prove (C.1) without Assumption 3.1, since by the triangle inequality and Proposition B.1(b),

$$\|\tau_0 b(\cdot; k, w_k(Q_0))\varphi_{\text{par}}^\delta\|_{\text{Lip}} \leqslant \tau_0 \|b(\cdot; k, w_k(Q_0)) - q_0\|_{\text{Lip}} \|\varphi_{\text{par}}^\delta\|_{\text{Lip}} + \|\varphi_0\|_{\text{Lip}},$$

where the first term becomes arbitrarily small for $M_1 = k$ large enough due to Proposition B.3(b) and the second one is strictly smaller than M. $\square$

Acknowledgements

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag, 3rd edition.

Andrews, D. W. K. (1984). "Non-Strong Mixing Autoregressive Processes." *Journal of Applied Probability*, 21(4): 930–934.

Belitser, E. and Serra, P. (2014). "Adaptive priors based on splines with random knots." *Bayesian Analysis*, 9(4): 859–882.

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). "A general framework for updating belief distributions." *Journal of the Royal Statistical Society - Series B*, 78(5): 1103–1130.

Bradley, R. C. (2005). "Basic properties of strong mixing conditions. A survey and some open questions." *Probability Surveys*, 2: 107–144.

— (2007). *Introduction to strong mixing conditions*, volume I. Kendrick press.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, 2nd edition.

Cadonna, A., Kottas, A., and Prado, R. (2017). "Bayesian mixture modeling for spectral density estimation." *Statistics & Probability Letters*, 125: 189–195.

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

Chopin, N., Rousseau, J., and Liseo, B. (2013). "Computational aspects of Bayesian spectral density estimation." *Journal of Computational and Graphical Statistics*, 22(3): 533–557.

Choudhuri, N., Ghosal, S., and Roy, A. (2004). "Bayesian estimation of the spectral density of a time series." *Journal of the American Statistical Association*, 99(468): 1050–1059.

Contereras-Cristán, A., Gutiérrez-Peña, E., and Walker, S. G. (2006). "A note on Whittle's likelihood." *Communications in Statistics - Simulation and Computation*, 35(4): 857–875.

Dahlhaus, R., Janas, D., et al. (1996). "A frequency domain bootstrap for ratio statistics in time series analysis." *Annals of Statistics*, 24(5): 1934–1963.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. New York: Oxford University press.

Dovgoshey, O., Martio, O., Ryazanov, V., and Vuorinen, M. (2006). "The Cantor function." *Expositiones Mathematicae*, 24: 1–37.

Dudley, R. M. (2004). *Real Analysis and Probability*. Cambridge University Press.

Edwards, M., Meyer, R., and Christensen, N. (2019). "Bayesian nonparametric spectral density estimation using B-spline priors." *Statistics and Computing*, 29: 67–78.

Eriksson, K., Estep, D., and Johnson, C. (2004). *Applied Mathematics: Body and Soul Volume 1: Derivatives and Geometry in IR3*. Springer-Verlag Berlin Heidelberg, 1 edition.

Farouki, R. T. (2012). "The Bernstein polynomial basis: A centennial retrospective." *Computer Aided Geometric Design*, 29: 379–419.

Francq, C. and Zakoian, J. (2019). *GARCH Models structure, statistical inference and financial applications*. John Wiley & Sons, 2nd edition.

Gangopadhyay, A., Mallick, B., and Denison, D. (1999). "Estimation of spectral density of a stationary time series via an asymptotic representation of the periodogram." *Journal of statistical planning and inference*, 75(2): 281–290.

Gao, C. and Zhou, H. (2016). "Rate exact Bayesian adaptation with modified block priors." *The Annals of Statistics*, 44(1): 318–345.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*, volume 44. Cambridge University Press.

Giné, E. and Nickl, R. (2011). "Rates of contraction for posterior distributions in L_r-metrics, $1 \leqslant r \leqslant \infty$." *The Annals of Statistics*, 39(6): 2883–2911.

Gloube, G. H. and van Loan, C. F. (1996). *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 3rd edition.

Gray, R. M. (2006). *Toeplitz and circulant matrices: A review*. Now Publishers Inc.

Hannan, E. J. (1973). "The Asymptotic Theory of Linear Time-Series Models." *Journal of Applied Probability*, 10(1): 130–145.
URL http://www.jstor.org/stable/3212501

Hermansen, G. H. (2008). "Bayesian nonparametric modelling of covariance functions, with application to time series and spatial statistics." Ph.D. thesis, Universitetet i Oslo.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Kakizawa, Y. (2006). "Bernstein polynomial estimation of a spectral density." *Journal of Time Series Analysis*, 27(2): 253–287.

Katznelson, Y. (2004). *An Introduction to Harmonic Analysis*. Cambridge University Press, 3rd edition.

Kirch, C., Edwards, M. C., Meier, A., and Meyer, R. (2019). "Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis." *Bayesian Analysis*, 14(4): 1037–1073.

Kleijn, B. J. K. and van der Vaart, A. W. (2006). "Misspecification in infinite-dimensional Bayesian statistics." *The Annals of Statistics*, 34(2): 837–877.

Kreiss, J.-P., Paparoditis, E., et al. (2003). "Autoregressive-aided periodogram bootstrap for time series." *The Annals of Statistics*, 31(6): 1923–1955.

Kruijer, W. (2008). "Convergence Rates in Nonparametric Bayesian Density Estimation." Ph.D. thesis, Vrije University of Amsterdam.

Liseo, B., Marinucci, D., and Petrella, L. (2001). "Bayesian semiparametric inference on long-range dependence." *Biometrika*, 88(4): 1089–1104.

Meier, A. (2018). "A matrix Gamma process and applications to Bayesian analysis of multivariate time series." Ph.D. thesis, Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik.
URL http://dx.doi.org/10.25673/13407

Meier, A., Kirch, C., Edwards, M. C., and Meyer, R. (2017). *beyondWhittle: Bayesian Spectral Inference for Stationary Time Series*. R package.

Meier, A., Kirch, C., and Meyer, R. (2020). "Bayesian Nonparametric Analysis of Multivariate Time Series: A Matrix Gamma Process Approach." *Journal of Multivariate Analysis*, 175.

Özarslan, M. A. and Duman, O. (2016). "Smoothness Properties of Modified Bernstein-Kantorovich Operators." *Numerical Functional Analysis and Optimization*, 37(1): 92–105.

Pawitan, Y. and O'sullivan, F. (1994). "Nonparametric spectral density estimation using penalized Whittle likelihood." *Journal of the American Statistical Association*, 89: 600–610.

Peligrad, M. and Wu, W. B. (2010). "Central limit theorem for Fourier transforms of stationary processes." *The Annals of Probability*, 38(5): 2009–2022.

Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory*. John Wiley & Sons.

Rao, S. S. and Yang, J. (2020). "Reconciling the Gaussian and Whittle Likelihood with an application to estimation in the frequency domain."

Schmidt, K. D. (1991). "The Cantor set in probability theory." Technical report, Universität Mannheim, Fakultät für Mathematik und Informatik.

Schwartz, L. (1965). "On Bayes procedures." *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 4: 10–26.

Searcòid, M. O. (2006). *Metric Space*. Springer London.

Serov, V. (2017). *Fourier Series, Fourier Transform and Their Applications to Mathematical Physics*. Springer, 1st edition.

Sethuraman, J. (1994). "A CONSTRUCTIVE DEFINITION OF DIRICH-LET PRIORS." *Statistica Sinica*, 4(2): 639–650.

Shalizi, C. R. (2009). "Dynamics of Bayesian updating with dependent data and misspecified models." *Electronic Journal of Statistics*, 3: 1039–1074.

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013). "Posterior consistency of Bayesian quantile regression based on the misspecifed asymmetric Laplace density." *Bayesian Analysis*, 8(2): 479–504.

Sykulski, A. M., Olhede, S. C., Lilly, J. M., Guillaumin, A. P., and Early, J. J. (2019). "The de-biased Whittle likelihood." *Biometrika*, 106: 251–266.

Syring, N. (2017). "Gibbs Posterior Distributions: New Theory and Applications." Ph.D. thesis, University of Illinois at Chicago.

Syring, N., Hong, L., and Martin, R. (2019). "Gibbs posterior inference on value-at-risk." *Scandinavian Actuarial Journal*, 2019(7): 548–557.

Tamaki, K. (2008). "The Bernstein-von Mises theorem for stationary processes." *Journal of the Japan Statistical Society*, 38(2): 311–323.

Whittle, P. (1957). "Curve and periodogram smoothing." *JRSSB*, 38–63.

Yaskov, P. (2018). "LLN for quadratic forms of long memory time series and its applications in random matrix theory." *Journal of Theoretical Probability*, 31: 2032–2055.