Center for Analytical Finance
University of California, Santa Cruz

Working Paper No. 17

Bayesian mixture modelling for spectral density estimation

Annalisa Cadonna, Athanasios Kottas and Raquel Prado*

October 2015

Abstract

We develop a Bayesian modelling approach for spectral densities, built from a local Gaussian mixture approximation to the Whittle log-likelihood. The implied model for the log-spectral density is a mixture of linear functions with frequency-dependent logistic weights, which allows for general shapes for smooth spectral densities. The proposed approach facilitates efficient posterior simulation as it casts the spectral density estimation problem in a mixture modelling framework for density estimation. It also sets the stage for hierarchical extensions for spectral analysis of multiple time series. The methodology is illustrated with synthetic and real data sets.

Key words: Logistic mixture weights; Markov chain Monte Carlo; Normal mixtures; Whittle likelihood.

About CAFIN

The Center for Analytical Finance (CAFIN) includes a global network of researchers whose aim is to produce cutting edge research with practical applications in the area of finance and financial markets. CAFIN focuses primarily on three critical areas:

- Market Design
- Systemic Risk
- Financial Access

Seed funding for CAFIN has been provided by Dean Sheldon Kamieniecki of the Division of Social Sciences at the University of California, Santa Cruz.

---

* Annalisa Cadonna (acadonna@soe.ucsc.edu) is Ph.D. student, and Athanasios Kottas (thanos@soe.ucsc.edu) and Raquel Prado (raquel@soe.ucsc.edu) are Professors in the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA. This research was supported in part by the National Science Foundation under award DMS 1407838.

Bayesian mixture modelling for spectral density estimation

Annalisa Cadonna, Athanasios Kottas and Raquel Prado *

Abstract

We develop a Bayesian modelling approach for spectral densities, built from a local Gaussian mixture approximation to the Whittle log-likelihood. The implied model for the log-spectral density is a mixture of linear functions with frequency-dependent logistic weights, which allows for general shapes for smooth spectral densities. The proposed approach facilitates efficient posterior simulation as it casts the spectral density estimation problem in a mixture modelling framework for density estimation. It also sets the stage for hierarchical extensions for spectral analysis of multiple time series. The methodology is illustrated with synthetic and real data sets.

Key words: Logistic mixture weights; Markov chain Monte Carlo; Normal mixtures; Whittle likelihood.

1 Introduction

Spectral density estimation from multiple observed time series is important in fields where information about frequency behavior is relevant, as in neuroscience, econometrics and geoscience. For example, electrical signals measuring brain activity, such as electroencephalograms, are typically recorded at multiple locations over the scalp of a given subject in neuroscience studies. In addition, such studies often involve several subjects and various treatments or experimental conditions. Therefore, providing flexible methods for spectral analysis of multiple time series is key, particularly since the methods that are currently available in the literature cannot be easily extended to deal with these cases.

*Annalisa Cadonna (acadonna@soe.ucsc.edu) is Ph.D. student, and Athanasios Kottas (thanos@soe.ucsc.edu) and Raquel Prado (raquel@soe.ucsc.edu) are Professors in the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA. This research was supported in part by the National Science Foundation under award DMS 1407838.

We develop a mixture modelling approach for a single spectral density, which is amenable to extensions for hierarchical modelling of multiple spectral densities. Throughout, we assume n realizations $x_1, \ldots, x_n$ from a zero-mean stationary time series $\{X_t : t = 1, 2, ...\}$, with absolutely summable autocovariance function $\gamma(\cdot)$. The spectral density function is defined as

$$f(\omega) = \sum_{k=-\infty}^{+\infty} \gamma(k) \exp(-ik\omega), \quad \text{for} - \pi \leq \omega \leq \pi,$$

where $\gamma(k) = \mathrm{E}(X_{t+k}X_t)$ denotes the autocovariance function. An estimator of the spectral density is the periodogram, defined as $I_n(\omega) = \left|\sum_{t=1}^{n} x_t \exp(-it\omega)\right|^2 / n$. Although $I_n(\omega)$ is defined for all $\omega \in [-\pi, \pi]$, it is computed at the Fourier frequencies $\omega_j = 2\pi j/n$, for $j = 0, \ldots, \lfloor n/2 \rfloor$, where $\lfloor n/2 \rfloor$ is the largest integer not greater than $n/2$. Because of the symmetry of the periodogram, there are only $\lfloor n/2 \rfloor + 1$ effective observations. In fact, following common practice, we exclude the observations at $\omega_j = 0, \pi$, and therefore the sample size is $N = \lfloor n/2 \rfloor - 1$. It is well known that the periodogram is not a consistent estimator of the spectral density. Consistent estimators have been obtained by smoothing the periodogram or the log-periodogram through windowing methods, as in, for instance, Parzen (1962).

Model-based approaches to spectral density estimation are typically built from the Whittle likelihood approximation to the periodogram (Whittle, 1957). In fact, for relatively large sample sizes, the periodogram realizations at the Fourier frequencies, $I_n(\omega_j)$, can be considered independent. In addition, for large n and for zero-mean Gaussian time series, the $I_n(\omega_j)$, for $\omega_j \neq 0, \pi$, can be considered independent exponentially distributed with mean $f(\omega)$. The main advantage of the Whittle likelihood with respect to the true likelihood is that the spectral density appears explicitly and not through the autocovariance function. Moreover, under the Whittle likelihood, the estimation problem can be cast in a regression framework with observations given by the log-periodogram ordinates and regression function defined by the log-spectral density: $\log(I_n(\omega_j)) = \log(f(\omega_j)) + \epsilon_j$, for $j = 1, \ldots, N$. Here, the ϵ_j follow a log-exponential distribution with scale parameter 1. In this context, frequentist estimation approaches include approximating the distribution of the ϵ_j with a normal distribution and fitting a smoothing spline to the log-periodogram (Wahba, 1980) and maximizing the Whittle likelihood with a roughness penalty term (Pawitan & O'Sullivan, 1994). Bayesian approaches have also been developed. Carter & Kohn (1997) approximate the distribution of the ϵ_j with a mixture of normal distributions and assign a smoothing prior to $\log(f(\omega))$. Rosen & Stoffer (2007) express the log-spectral density as

$\log(f(\omega)) = \alpha_0 + \alpha_1\omega + h(\omega)$, with a Gaussian process prior on $h(\omega)$. A different approach based on Bernstein polynomial priors (Petrone, 1999) was considered by Choudhuri et al. (2004), with emphasis on posterior consistency results. Pensky et al. (2007) propose Bayesian wavelet-based smoothing of the log-periodogram.

In this paper we present a new Bayesian approach to spectral density estimation. Building from theoretical results in Jiang & Tanner (1999a) and Norets (2010), we approximate the Whittle log-likelihood with a mixture of normal distributions, with frequency-dependent means and logistic weights. The implied prior model for $\log(f(\omega))$ enables a wide range of shapes for smooth spectral densities. The mixture representation of the log-spectral density is advantageous in that it facilitates extensions to hierarchical modelling in the multiple time series setting. The model described here is a starting point for such generalizations.

The outline of the paper is as follows. In Section 2, we describe the mixture modelling approach for spectral density estimation. In Section 3, we present results from synthetic and real datasets, and Section 4 concludes with a summary.

2 Mixture model for the spectral density

Norets (2010) presents approximation properties of finite local mixtures of normal regressions as flexible models for conditional densities. The work in Norets (2010) focuses on the joint distribution of the responses and covariates, showing that, under certain conditions, the joint distribution can be approximated in Kullbach-Leibler divergence by different specifications of local finite mixtures of normals in which means, variances, and weights can depend on the covariates. In this paper we consider instead fixed covariates, specifically the Fourier frequencies.

We propose to approximate the Whittle log-likelihood with a local mixture of normal distributions, with frequency-dependent means and weights. We assume that $\log(f(\omega))$ and its first and second derivatives are continuous and bounded. In practice, this requirement translates into a smoothness assumption on the log-spectral density.

At the Fourier frequencies, for $j = 1, .., N$, we have $\mathrm{E}[\log(I_n(\omega_j))] = \log(f(\omega_j)) - \gamma$, where γ is the Euler-Mascheroni constant. Let $y_j = \log(I_n(\omega_j)) + \gamma$. Under the Whittle likelihood approximation

– that is, the exponential distribution for the $I_n(\omega_j)$ with mean $f(\omega)$ – the y_j are independent with the following distribution:

$$f_Y(y) = \exp\{y - \gamma - \log(f(\omega)) - \exp(y - \gamma - \log(f(\omega)))\}, \quad y \in \mathbb{R}. \tag{1}$$

Notice that the distribution in (1) is in the exponential family, and $-y_j$ are Gumbel distributed with scale parameter 1 and location parameter defined additively through $\log(f(\omega))$ and γ, such that the mean is $-\log(f(\omega))$. Although the Whittle approximate likelihood for either the $I_n(\omega_j)$ or the y_j is based on standard distributions, the spectral density enters the likelihood in a non-standard fashion through the mean parameter. Our proposal is to apply a further approximation replacing the distribution in (1) with a structured mixture of normal distributions, which induces a flexible model for the log-spectral density. The key advantage is that for inference, we can draw from well established methods for mixtures. This is practically relevant even when modelling a single spectral density, but it also sets the stage for hierarchical extensions to modelling and inference for multiple related spectral densities.

More specifically, we approximate the distribution of y_j with a mixture of normal distributions, with means that depend linearly on ω_j and frequency-dependent logistic weights:

$$y_j \mid \Theta \overset{ind}{\sim} \sum_{k=1}^{K} g_k(\omega_j; \lambda, \zeta, \phi)\mathrm{N}(y_j \mid \alpha_k + \beta_k\omega_j, \sigma^2), \qquad j = 1, \ldots, N \tag{2}$$

where $g_k(\omega_j; \lambda, \zeta, \phi) = \exp\{(\zeta_k + \phi_k\omega_j)/\lambda\}/\sum_{i=1}^{K}\exp\{(\zeta_i + \phi_i\omega_j)/\lambda\}$. Here, Θ collects all model parameters, in particular, it includes the parameters for the logistic weights, $\zeta = \{\zeta_k : k = 1, ..., K\}$ and $\phi = \{\phi_k : k = 1, ..., K\}$, the intercept and slope parameters for the means of the normal mixture components, $\alpha = \{\alpha_k : k = 1, ..., K\}$ and $\beta = \{\beta_k : k = 1, ..., K\}$, as well as parameters λ and σ^2. The logistic weights partition the support with soft boundaries. The parameter λ controls the smoothness of the transition between the subsets of $(0, \pi)$ induced by the logistic weights. The larger the value of λ, the smoother is the corresponding estimate of the spectral density.

Under the approximation in (2), the model for the log-spectral density is given by

$$\log(f(\omega)) = \sum_{k=1}^{K} g_k(\omega; \lambda, \zeta, \phi)(\alpha_k + \beta_k\omega). \tag{3}$$

Hence, the induced prior model for the log-spectral density admits a representation as a mixture of linear functions with component specific intercept and slope parameters, and with frequency-dependent weights that allow for local adjustment, and thus flexible spectral density shapes.

In addition to the appealing interpretation of the implied spectral density model, further theoretical justification for the approximation in (2) can be provided by means of results in the L_p norm for the spectral density; see Appendix A for details. In particular, the L_p distance between the true log-spectral density and the proposed mixture is bounded by a constant proportional to K_0/K^2, where K_0 is related to the smoothness of the true spectral density. Hence, the distance decreases quadratically with the number of components. If we have prior knowledge on the smoothness of the log-spectral density, we can use it to fix K. For practical purposes, confirmed from empirical investigation with several data sets including the ones of Section 3, we have observed that in general a relatively small number of mixture components suffices to capture different spectral density shapes, with inference results being robust to the choice of K.

To complete the full Bayesian model, we assume prior independence among and between the parameters of each mixture component. Specifically, we use a normal prior distribution with mean μ_α and variance σ^2_α for the α_k, and a normal prior with mean μ_β and variance σ^2_β for the β_k. For the common variance parameter, σ^2, we use an inverse-gamma prior, and for the smoothness parameter, λ, a gamma prior. We place standard normal priors on the ζ_k and ϕ_k. This choice is motivated by the fact that the variances of ζ_k and ϕ_k and λ cannot be estimated simultaneously. The prior of λ expresses our belief on the degree of smoothness of the spectral density. As demonstrated with the data examples of Section 3, the smoothness parameter can be learned from the data. The prior on the intercept parameters α_k summarizes information about the spectral density value near $\omega = 0$. Moreover, the prior on the slope parameters β_k can be used to express beliefs about the shape of the spectral density. For instance, for multimodal spectral densities, we expect some β_k to be positive and some negative, whereas for unimodal spectral densities, we expect all β_k to have the same sign.

The model in (2) can be expanded in hierarchical form by introducing configuration variables $(r_1, \ldots, r_N)$, where each r_j, $j = 1, \ldots, N$, has a discrete distribution with values in $\{1, \ldots, K\}$:

$$
\begin{aligned}
y_j \mid r_j, \alpha, \beta, \sigma^2 &\overset{ind}{\sim} \mathrm{N}(y_j \mid \alpha_{r_j} + \beta_{r_j}\omega_j, \sigma^2), \qquad j = 1, \ldots, N \\
r_j \mid \zeta, \phi, \lambda &\overset{ind}{\sim} \sum_{k=1}^{K} g_k(\omega_j; \lambda, \zeta, \phi)\delta_k(r_j), \qquad j = 1, \ldots, N
\end{aligned}
$$

where $\delta_k(\cdot)$ denotes a point mass at k. We develop a Markov chain Monte Carlo algorithm to simulate from the joint posterior distribution of the parameters, which is based almost exclusively on Gibbs

sampling steps. The full conditional distribution for each r_j is a discrete distribution on $\{1, \ldots, K\}$ with updated probabilities. We have conjugate full conditional distributions for the (α_k, β_k), and for σ^2. Updating the parameters for the logistic weights is more challenging, and we use a data augmentation step based on auxiliary Pólya-Gamma variables (Polson et al., 2013). Sampling λ requires a Metropolis-Hastings step. Details of the posterior simulation algorithm are provided in Appendix B.

Note that the prior variance of the spectral density increases with ω. To minimize this effect, when fitting the model we normalize the support $(0, \pi)$ of the spectral density to the unit interval; the results can be reported on the original scale through straightforward transformation.

3 Numerical illustration

3.1 Synthetic datasets

We evaluate the performance of our method through four synthetic data examples. In all cases, for the α_k and β_k, we used zero-mean normal priors with variances $\sigma_\alpha^2 = \sigma_\beta^2 = 100$. An inverse-gamma prior with shape parameter 3 and mean 3 was placed on σ^2. For λ, we used a gamma prior with shape parameter 2 and mean 0.2, which places almost all its mass in $(0, 1)$. The number of mixture components was fixed at $K = 10$; results were essentially the same with larger K.

Focusing first on monotonic spectral densities, we simulated data from two autoregressive processes of order one. Specifically, we considered $x_{1,t} = 0.7x_{1,t-1} + \epsilon_{1,t}$, with $\epsilon_{1,t} \sim \text{N}(0, 1)$ and $x_{2,t} = -0.9x_{2,t-1} + \epsilon_{2,t}$, with $\epsilon_{2,t} \sim \text{N}(0, 1)$, and simulated 400 observations from each of these processes. The estimated log-spectral densities and the corresponding 95% posterior intervals are reported in Figure 1, together with the true log-spectral densities and the periodogram observations. The spectral densities for these processes are respectively monotonically decreasing and increasing. Notice that our model successfully captures their monotonic trends. The posterior density for λ for each case is plotted in Figure 3.

To evaluate model performance for more complex spectral densities, we simulated data from the sum of two autoregressive processes and a white noise term. Specifically, let $x_{3,t} = 0.9x_{3,t-1} + \epsilon_{3,t}$, with $\epsilon_{3,t} \sim \text{N}(0, 1)$, $x_{4,t} = 0.9x_{4,t-1} - 0.9x_{4,t-2} + \epsilon_{4,t}$, with $\epsilon_{4,t} \sim \text{N}(0, 1)$, and $x_{5,t} = -0.8x_{5,t-1} - 0.8x_{5,t-2} + \epsilon_{5,t}$, with $\epsilon_{5,t} \sim \text{N}(0, 1)$. We construct $z_{1,t} = x_{3,t} + x_{5,t} + \nu_{1,t}$, where $\nu_{1,t} \sim \text{N}(0, 1)$, and

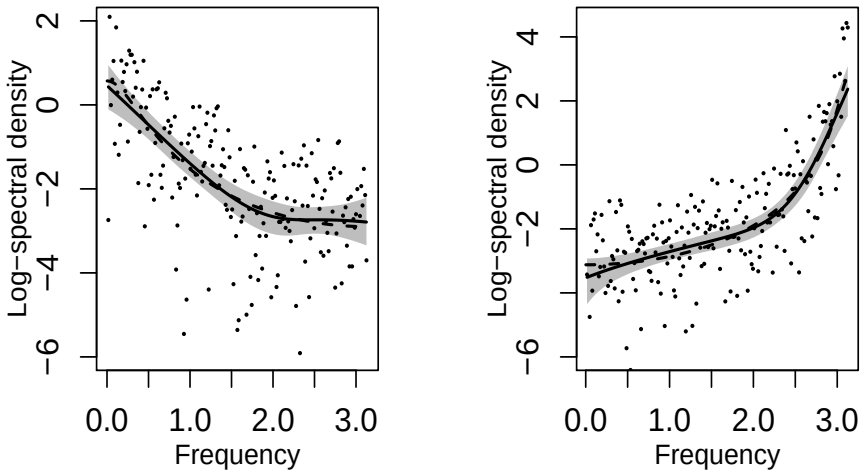

Figure 1: Synthetic data. Posterior mean (solid dark line), 95% credible interval (shaded area), true log-spectral density (dashed line), and log-periodogram (dots) for data simulated from autoregressive processes of order one, with parameters 0.7 (left panel) and -0.9 (right panel).

$z_{2,t} = x_{4,t} + x_{5,t} + \nu_{2,t}$, where $\nu_{2,t} \sim \mathrm{N}(0, 1)$. We simulated 400 observations from each of these two processes. The spectral density of $z_{1,t}$ is decreasing for low frequencies and has a peak around $\omega = 1.1$; the spectral density of $z_{2,t}$ is bimodal, with two peaks around $\omega = 1.1$ and $\omega = 2$. The estimated log-spectral densities for the two processes along with 95% posterior intervals are reported in Figure 2. Our model does well in capturing the shape of the log-spectral densities, and it successfully identifies the peaks. The corresponding posterior densities for λ are plotted in Figure 3. In both cases, the posterior distribution of λ is supported by smaller values than for the autoregressive processes of order one, in agreement with the fact that the spectral densities of $x_{1,t}$ and $x_{2,t}$ are smoother than those of $z_{1,t}$ and $z_{2,t}$.

3.2 Electroencephalogram data

We consider four times series that correspond to portions of electroencephalograms taken from a larger dataset. The original time series were recorded at 19 locations over the scalp of a patient who received electroconvulsive therapy. Further details and data analysis can be found in Krystal et al. (1999). The

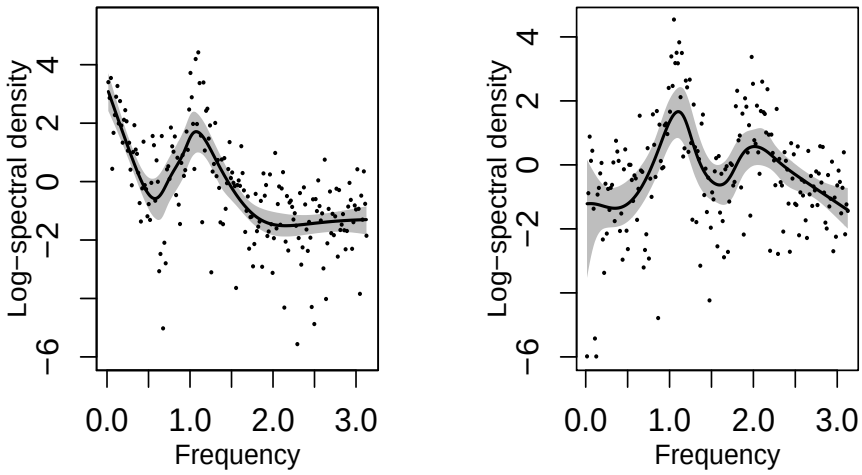

Figure 2: Synthetic data. Posterior mean (solid dark line), 95% credible interval (shaded area), and log-periodogram (dots) for the sum of autoregressive processes and white noise.

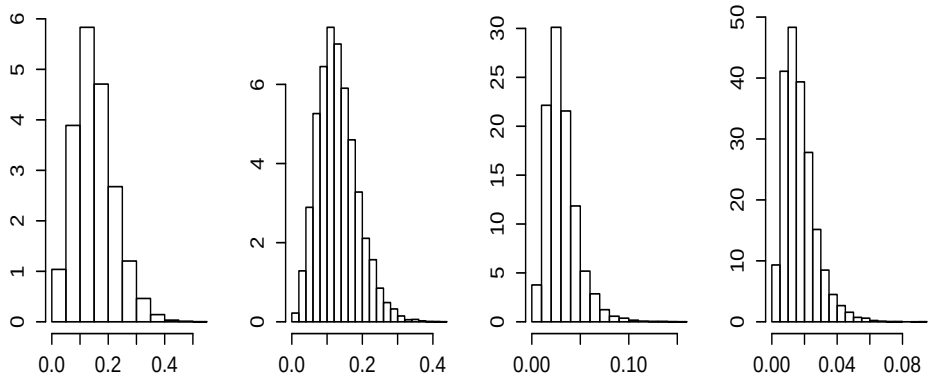

Figure 3: Synthetic data. Posterior density for λ for data simulated from (from the left) autoregressive processes of order one, with parameter 0.7 and -0.9, and for the sum of autoregressive processes and white noise.

time series were recorded in four left channels, two of which are in the frontal region of the scalp (F7 and F3), one is in the temporal region (T5), and one is in the parietal region (P3). For each time series, we have 299 observations, obtained by subsampling the electroencephalogram signal every sixth observation

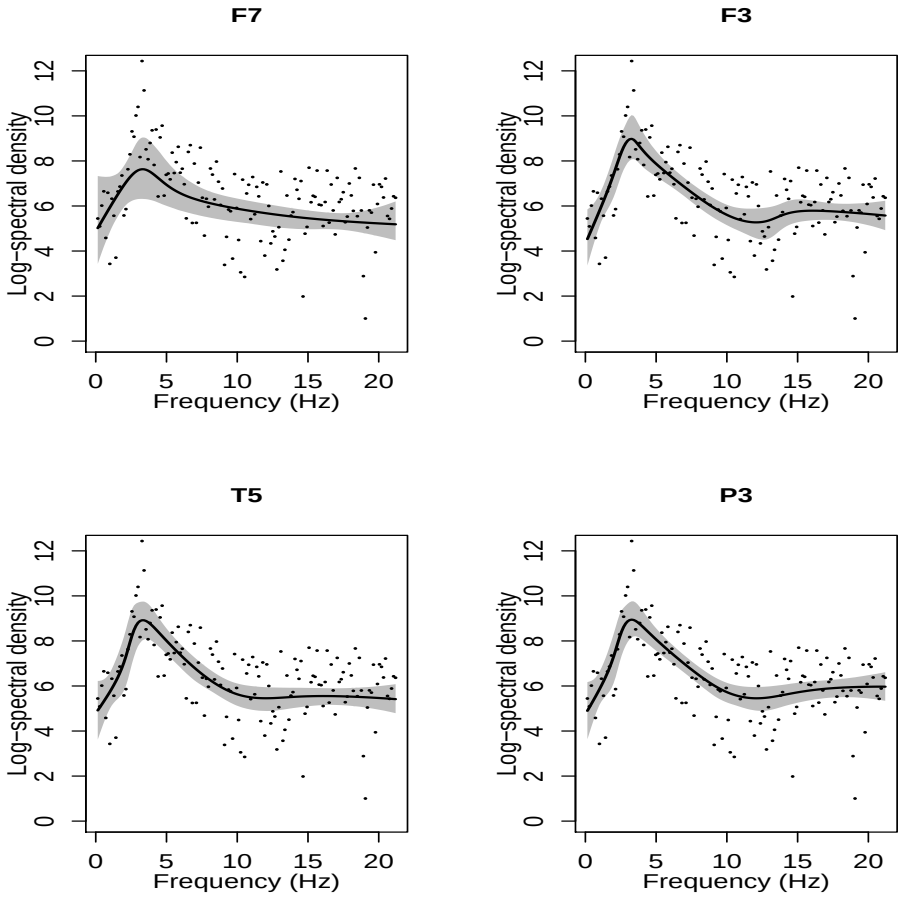

Figure 4: Posterior mean (solid dark line), 95% credible interval (shaded area), and log-periodogram (dots) for the electroencephalogram datasets recorded in four channels.

from a mid-seizure section. The original sampling rate was 256 Hz. The priors were the same with Section 3.1, save for the prior for λ that here is given by a gamma distribution with shape parameter 3 and mean 0.1, which places almost all its mass in $(0, 0.5)$. We choose this prior because we expect at least one pronounced peak in the spectral density, reflecting brain activity in at least one frequency band, and we thus want to avoid oversmoothing. The number of components is fixed to $K = 10$; also in this case, the results were robust with respect to the number of components.

Figure 4 shows the posterior mean estimates and 95% posterior credible intervals for the spectral densities together with the logged spectral periodogram. All the channels show a peak around 3.4 Hz. which is slightly shifted to the left in T5 and P3. These results are consistent with previous analyses

which indicate that the observed quasi-periodicity is dominated by activity in the delta frequency range, that is, in the range from 1 to 5 Hz. We also note that although there are important similarities across the spectral densities, each density has its own features with channels F3, T5 and P3 sharing more similarities and F7 being different from the rest.

4 Summary

We have presented a Bayesian approach to spectral density estimation that builds from an approximation to the Whittle log-likelihood through a mixture of normal distributions with linear means and frequency-dependent logistic weights. We have used simulated data examples to demonstrate the capacity of the model to uncover both monotonic and multimodal shapes for the underlying spectral density, working with a fixed number of mixture components. The modelling approach can be generalized with random K, albeit at the expense of more computationally challenging posterior simulation. In the electroencephalogram data example, we have shown that the spectral densities corresponding to different channels from the same subject can present different characteristics. The fact that the spectral density estimation problem has been cast in a mixture modelling framework is key for extensions to hierarchically dependent spectral densities. Future research will focus on building nonparametric hierarchical models for multiple spectral densities.

References

CARTER, C. K. & KOHN, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society, Series B* **59**, 255–268.

CHOUDHURI, N., GHOSAL, S. & ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association* **99**, 1050–1059.

JIANG, W. & TANNER, M. A. (1999a). Hierarchical mixtures-of-experts for exponential family regression models: Approximation and maximum likelihood estimation. *The Annals of Statistics* **27**, 987–1011.

JIANG, W. & TANNER, M. A. (1999b). On the approximation rate of hierarchical mixtures-of-experts for generalized linear models. *Neural computation* **11**, 1183–1198.

KRYSTAL, A., PRADO, R. & WEST, M. (1999). New methods of time series analysis of non-stationary EEG data: eigenstructure decompositions of time-varying autoregressions. *Clinical Neurophysiology* **110**, 2197–2206.

NORETS, A. (2010). Approximation of conditional densities by smooth mixtures of regressions. *The Annals of Statistics* **38**, 1733–1766.

PARZEN, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* **33**, 1065–1076.

PAWITAN, Y. & O'SULLIVAN, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association* **89**, 600–610.

PENSKY, M., VIDAKOVIC, B. & DECANDITIIS, D. (2007). Bayesian decision theoretic scale-adaptive estimation of a log-spectral density. *Statistica Sinica* **17**, 635–666.

PETRONE, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics* **26**, 373–393.

POLSON, N. G., SCOTT, J. G. & WINDLE, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association* **108**, 1339–1349.

ROSEN, O. & STOFFER, D. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika* **94**, 335–345.

WAHBA, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association* **75**, 122–132.

WHITTLE, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society, Series B* **19**, 38–63.

Appendix A: Theoretical results

Here, we show that a smooth function $h(\cdot)$ on $[0, \pi]$ can be approximated by a smooth mixture of linear functions, $h_K(\omega) = \sum_{k=1}^{K} g_k(\omega; \lambda, \zeta, \phi)(\alpha_k + \beta_k \omega)$, with logistic weights, $g_k(\omega; \lambda, \zeta, \phi) = \exp\{(\zeta_k + \phi_k \omega)/\lambda\}(\sum_{m=1}^{K} \exp\{(\zeta_m + \phi_m \omega)/\lambda\})^{-1}$. Let $\|f(\cdot)\|_p = \left(\int_0^\pi |f(\omega)|^p \mathrm{d}P(\omega)\right)^{\frac{1}{p}}$ denote the L_p norm, where P is a probability measure on $\Omega = [0, \pi]$ absolutely continuous with respect to Lebesgue measure. Moreover, denote by χ_B the indicator function for $B \subseteq \Omega$, and define the partition $\{Q_1^K, ..., Q_K^K\}$ of Ω, where $Q_k^K = [(k-1)\pi/K, k\pi/K)$, for $k = 1, \ldots, K-1$, and $Q_K^K = [(K-1)\pi/K, \pi]$. The following lemma from Jiang & Tanner (1999a) is used to obtain the main result.

LEMMA *For logistic weights, $g_k(\omega; \lambda, \zeta, \phi)$, $k = 1, \ldots, K$, as defined above, we have that, for all K and for each $\epsilon > 0$, there exist $\zeta^{(\epsilon)}, \phi^{(\epsilon)}, \lambda^{(\epsilon)}$ such that* $\sup_{1 \leq k \leq K} ||g_k(\cdot; \lambda^{(\epsilon)}, \zeta^{(\epsilon)}, \phi^{(\epsilon)}) - \chi_{Q_k^K}(\cdot)||_p < \epsilon$.

Based on the lemma, the logistic mixture weights arbitrarily approximate the set of indicator functions on the partition $\{Q_1^K, ..., Q_K^K\}$, for any fixed K. The following result establishes that the distance in the L_p norm between the target log-spectral density, h, and the proposed mixture model, h_K, is bounded by a constant that is inversely proportional to the square of the number of mixture components K.

THEOREM *Let $h \in W_{2,K_0}^{\infty}$, that is, the Sobolev space of continuous functions bounded by K_0, with the first two derivatives continuous and bounded by K_0. Then,* $\inf_{h_K} \|h_K(\cdot) - h(\cdot)\|_p \leq \pi^2 K_0/(2K^2)$.

PROOF The proof is along the lines of the one in Jiang & Tanner (1999b). We start by proving that, for fixed K, any $h \in W_{2,K_0}^{\infty}$ can be approximated by a piecewise linear function on the partition $\{Q_1^K, ..., Q_K^K\}$, with the L_p distance bounded by a constant that depends on K. For each interval Q_k^K, consider a point $\omega_k^* \in Q_k^K$ and the linear approximation based on the first-order Taylor series expansion: $\hat{h}_k(\omega) = \hat{\alpha}_k + \hat{\beta}_k \omega$, for $\omega \in Q_k^K$, where $\hat{\alpha}_k = h(\omega_k^*) - \omega_k^* h'(\omega_k^*)$ and $\hat{\beta}_k = h'(\omega_k^*)$; here, $h'(\omega_k^*)$ denotes the first derivative of $h(\omega)$ evaluated at ω_k^*, with similar notation used below for the second derivative. We have $\left\| \left\{ \sum_{k=1}^{K} \chi_{Q_k^K}(\cdot) \hat{h}_k(\cdot) \right\} - h(\cdot) \right\|_p = \left\| \sum_{k=1}^{K} \chi_{Q_k^K}(\cdot) \left\{ \hat{h}_k(\cdot) - h(\cdot) \right\} \right\|_p \leq \sup_{1 \leq k \leq K} \left\| \hat{h}_k(\cdot) - h(\cdot) \right\|_\infty$, where $\|\cdot\|_\infty$ denotes the L_∞ norm. Now, for each interval Q_k^K, we consider the second-order expansion of $h(\omega)$ around the same $\omega_k^* \in Q_k^K$. Note that the partition $\{Q_1^K, ..., Q_K^K\}$ satisfies the property that, for any k, and for any ω_1 and ω_2 in Q_k^K, $|\omega_1 - \omega_2| \leq \pi/K$. Using this

property and the fact that the second derivative of h is bounded by K_0, we obtain $|\hat{h}_k(\omega) - h(\omega)| \leq |0.5(\omega - \omega_k^*)^2 h''(\omega_k^*)| \leq \pi^2 K_0/(2K^2)$. Therefore, $\left\| \left\{ \sum_{k=1}^K \chi_{Q_k^K}(\cdot)\hat{h}_k(\cdot) \right\} - h(\cdot) \right\|_p \leq \pi^2 K_0/(2K^2)$.

Using the triangular inequality, we can write

$$\left\| \left\{ \sum_{k=1}^K g_k(\cdot;\lambda,\zeta,\phi)\hat{h}_k(\cdot) \right\} - h(\cdot) \right\|_p \leq \left\| \sum_{k=1}^K \left\{ g_k(\cdot;\lambda,\zeta,\phi) - \chi_{Q_k^K}(\cdot) \right\} \hat{h}_k(\cdot) \right\|_p + \left\| \left\{ \sum_{k=1}^K \chi_{Q_k^K}(\cdot)\hat{h}_k(\cdot) \right\} - h(\cdot) \right\|_p .$$

Based on the previous result, the second term is bounded by $\pi^2 K_0/(2K^2)$. Regarding the first term, $\left\| \sum_{k=1}^K \left\{ g_k(\cdot;\lambda,\zeta,\phi) - \chi_{Q_k^K}(\cdot) \right\} \hat{h}_k(\cdot) \right\|_p \leq \sum_{k=1}^K \left\| g_k(\cdot;\lambda,\zeta,\phi) - \chi_{Q_k^K}(\cdot) \right\|_p \left\| \hat{h}_k(\cdot) \right\|_\infty \leq K\epsilon(1+\pi)K_0$. For the last inequality, we have used the lemma and the fact that $|\hat{h}_k(\omega)| \leq |h(\omega_k^*)| + |h'(\omega_k^*)(\omega - \omega_k^*)| \leq K_0 + \pi K_0$. Finally, $\left\| \left\{ \sum_{k=1}^K g_k(\cdot;\lambda,\zeta,\phi)\hat{h}_k(\cdot) \right\} - h(\cdot) \right\|_p \leq K\epsilon(1+\pi)K_0 + \{\pi^2 K_0/(2K^2)\}$, and letting ϵ tend to zero, we obtain the result.

Appendix B: Markov Chain Monte Carlo algorithm

Posterior simulation for the model is based on a Gibbs sampler with data augmentation and Metropolis-Hastings steps, as outlined below.

The full conditional for each configuration variable r_j, $j = 1, \ldots, N$, is a discrete distribution on $\{1, \ldots, K\}$ with probabilities proportional to $g_k(\omega_j;\lambda,\zeta,\phi)\mathrm{N}(y_j \mid \alpha_k + \beta_k\omega_j, \sigma^2)$, for $k = 1, ..., K$.

Let $\mu_0 = (\mu_\alpha, \mu_\beta)'$, and Σ_0 the diagonal covariance matrix, with diagonal elements σ_α^2 and σ_β^2. Then, the posterior full conditional for each (α_k, β_k), $k = 1, \ldots, K$, is bivariate normal with covariance matrix $\Sigma^* = (\sigma^{-2}\sum_{\{j:r_j=k\}} z_j z_j' + \Sigma_0^{-1})^{-1}$ and mean $\Sigma^*(\Sigma_0^{-1}\mu_0 + \sum_{\{j:r_j=k\}} y_j z_j)$, where $z_j = (1, \omega_j)'$.

The posterior full conditional for σ^2 is inverse-gamma with parameters $n_0 + N/2$ and $d_0 + 0.5\sum_{j=1}^N \{y_j - (\alpha_{r_j} + \beta_{r_j}\omega_j)\}^2$, where n_0 and d_0 are the parameters of the inverse-gamma prior.

The posterior full conditional distribution for each (ζ_k, ϕ_k) is proportional to

$$\pi(\zeta_k, \phi_k) \prod_{\{j:r_j=k\}} \frac{\exp\{(\zeta_k + \phi_k\omega_j)/\lambda\}}{\sum_{i=1}^K \exp\{(\zeta_i + \phi_i\omega_j)/\lambda\}} \prod_{\{j:r_j\neq k\}} \frac{1}{\sum_{i=1}^K \exp\{(\zeta_i + \phi_i\omega_j)/\lambda\}}.$$

To sample (ζ_k, ϕ_k), for $k = 1, \ldots, K$, we use data augmentation introducing, for each k, N latent Pólya-Gamma random variables. Let $\psi_{kj} = (\zeta_k + \phi_k\omega_j)/\lambda$, and $c_{kj} = \log(\sum_{\{i\neq k\}} \exp\{(\zeta_i + \phi_i\omega_j)/\lambda\})$. Then, using Theorem 1 from Polson et al. (2013), we can write

$$\frac{\exp\{(\zeta_k + \phi_k\omega_j)/\lambda\}}{\sum_{i=1}^K \exp\{(\zeta_i + \phi_i\omega_j)/\lambda\}} = \frac{\exp(\psi_{kj} - c_{kj})}{1 + \exp(\psi_{kj} - c_{kj})} = 2^{-1}\exp\{(\psi_{kj} - c_{kj})/2\}\mathrm{E}[\exp\{-q_j(\psi_{kj} - c_{kj})^2/2\}]$$

and

$$\frac{1}{\sum_{i=1}^{K}\exp\{(\zeta_i+\phi_i\omega_j)/\lambda\}}=\frac{\exp(-c_{kj})}{1+\exp(\psi_{kj}-c_{kj})}\propto 2^{-1}\exp\{-(\psi_{kj}-c_{kj})/2\}\mathrm{E}[\exp\{-q_j(\psi_{kj}-c_{kj})^2/2\}],$$

where the q_j, $j=1,...,N$, are Pólya-Gamma random variables with parameters 1 and 0, and the expectation is taken with respect to the distribution of q_j. Now, the augmented full conditional distribution for (ζ_k,ϕ_k), given the q_j, is normal with covariance matrix $\Sigma_k^*=(\mathbb{I}+\sum_{j=1}^{N}z_jz_j'q_j/\lambda^2)^{-1}$ and mean $\Sigma_k^*(\sum_{\{j:r_j=k\}}z_jc_{kj}q_j/\lambda+\sum_{\{j:r_j=k\}}z_j/2\lambda-\sum_{\{j:r_j\neq k\}}z_j/2\lambda)$, where $\mathbb{I}$ is the identity matrix. Moreover, the full conditional for q_j, $j=1,\ldots,N$, is Pólya-Gamma with parameters 1 and $\psi_{kj}-c_{kj}$.

Finally, λ is updated using a random walk Metropolis step on $\log(\lambda)$.