Bayesian spectral density estimation using P-splines with quantile-based knot placement

Response letter to Statistical Papers

Patricio Maturana-Russel, Renate Meyer

October 31, 2020

Referee #1

We would like to thank the referee for a thorough review of our paper. Below, you find a point-by-point response to your points raised. We have now carefully revised the paper taking all of your – and the other referee's – comments into account and hope that you will now find the paper acceptable for publication. For your convenience, all changes made are marked in red in the manuscript.

Major comment:

p.8, Quantile-based knot placement
I could not understand the meaning of "calculate its distribution function $F_z = P(z \leq z_i)$" in the second item of the procedure. The values of the transformed periodogram z_i are regarded as the values of the density $\tilde{f}$, which is non-monotonic, taken at equally-spaced frequencies. According to item 4, the distribution function F_z, and hence the density function $\tilde{f}$, should be defined on the frequency domain.
I had a look at the R codes of the package psplinePsd. I think that the density function $\tilde{f}$ is a linear spline that interpolates as

$$\tilde{f}\left(\frac{\lambda_k}{2\pi}\right) = \frac{z_k}{\sum_{i=1}^{\nu} z_i}, \quad k = 1, \dots, \nu.$$

That's correct! The $\widetilde{f}$ is actually a probability mass function (not a probability density function), which is interpolated. However, we have simplified a bit our knot allocation algorithm avoiding that step. The new version of the algorithm is described step by step in our manuscript. This change leads to small changes in our application section, which has been consequently updated. We have also updated the `psplinePsd` package.

Minor comments:

1. *p.4, line 10*
   A spline of degree $r+1 \implies$ A spline of degree r (or, of order $r+1$)

   Fixed.

2. *p.5 line 13*
   In the definition of $\tau, f(\tau\omega) \implies f(\pi\omega)$

   That is correct. Fixed.

Referee #2

We would like to thank the referee for a thorough review of our paper. Below, you find a point-by-point response to your points raised. We have now carefully revised the paper taking all of your – and the other referee's – comments into account and hope that you will now find the paper acceptable for publication. For your convenience changes are marked in red in the manuscript.

Specific comments:

1. *Page 5, how to choose penalty or prior* ϕ*?*

   In a Bayesian setting, the roughness penalty is translated into the prior distribution for the spline parameters via a normal distribution. See for instance Bremhorst and Lambert [2016]. In order to make the posterior less sensitive to the prior, we suggest to follow their proposal and use their robust specifications for the hierarchical priors on ϕ, which is also described in our paper.

2. *To select the number of B-spline densities, the authors employ a heuristic rule of thumb* K^* *in the application Section, it is of interest to compare it with other criteria such as AIC and BIC.*

   This could be a potential avenue for future research, however, we think it is beyond the scope of this paper. An analysis of those characteristics can be quite extensive and could constitute a paper in itself, see for instance Likhachev [2017]. We would like to highlight that the number of knots is considered non-important by many, since the penalty parameter plus the penalty order difference penalties specified in the penalty matrix $\mathbf{P}$ penalise the inclusion of B-spline densities, controlling the smoothness of the estimate. This means that we only need to select a minimum value, the penalty terms take care of the rest. To quote Eilers et al. [2015]: "With the number of B-splines in the basis we can tune the smoothness of a curve to the data at hand. A smaller number of splines gives a smoother result. However, this is not the only possibility. We can also use a large basis and additionally constrain the coefficients of the B-splines, to achieve as much smoothness as desired. A properly chosen penalty achieves this"; and also "The number of B-splines can be (much) larger than the number of observations. The penalty makes the fitting procedure well-conditioned. This should be taken literally: even a thousand splines will fit ten observations without problems. Such is the power of the penalty."

   Other minor comments:

3. *Page 2, line 17, what is* λ_i*?*

   λ_i denote the Fourier frequencies. These have now been defined just after the first equation on page 2.

4. *Page 4, lines 5-7, "In particular, the approximation error of the mixture of B-spline distributions can be made arbitrarily small by finding suitable knot locations*

and increasing the number of knots", what does "made arbitrarily small" mean? Do you mean the approach with the beta densities the error cannot be made arbitrarily small?

Arbitrarily small means that the approximation error converges to zero with increasing number of knots. We have cited the paper by Perron and Mengersen [2001] that make this statement precise. But we do not mean to imply that this is not the case for the mixture of beta densities. Of course, according to the well-known Weierstrass approximation theorem, any continuous function can be approximated in the sup-norm by a linear combination of Bernstein polynomials. If the continuous function is a density, the approximant becomes a mixture of beta distributions. By increasing the number of beta distributions, the approximation error can be made arbitrarily small. But as shown in Perron and Mengersen (2001) the rate of approximation is faster for Bsplines than for Bernstein polynomials. We have changed this paragraph in the manuscript accordingly.

5. *Page 4, line 15, "A spline of degree r + 1 is a function", I think degree" should be order", we use degree for polynomials.*

   Fixed.

6. *Page 5, line 20, "If* $\tau = \int_0^1 f(\tau\omega)d\omega$ *is the normalizing constant", it implies* s_r *is a density, is it correct?*

   Yes, s_r is a pdf. Note that it is actually a mixture of B-spline densities.

Further Changes

Here some extra changes:

1. We have defined K^* as the number of internal knots (B-splines and BspDP prior section).
2. We have specified the prior on τ (P-spline prior subsection of section 2).
3. We have added $s_r(\omega)$ in Eq(3), short notation used in the posterior distribution of τ (Posterior computation section).
4. We have slightly simplified our knot location approach. This has led to minor variations in the results shown in the application section, which has been updated. This change has not impacted the conclusions. This version of the knot location algorithm is described step by step in the Quantile-based knot placement section.

References

Vincent Bremhorst and Philippe Lambert. Flexible estimation in cure survival models using Bayesian p-splines. *Computational Statistics & Data Analysis*, 93:270 – 284, 2016. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2014.05.009. URL `http://www.sciencedirect.com/science/article/pii/S0167947314001492`.

Paul H. C. Eilers, Brian D. Marx, and María Durbán. Twenty years of p-splines. *SORT: statistics and operations research transactions*, 39(2), 2015. ISSN 1696-2281. URL `http://hdl.handle.net/2117/88526`.

D.V. Likhachev. Selecting the right number of knots for b-spline parameterization of the dielectric functions in spectroscopic ellipsometry data analysis. *Thin Solid Films*, 636:

519 – 526, 2017. ISSN 0040-6090. doi: https://doi.org/10.1016/j.tsf.2017.06.056. URL http://www.sciencedirect.com/science/article/pii/S0040609017304911.

F. Perron and K. Mengersen. Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics*, 57(2):518–528, 2001. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/3068361.