Computational Statistics and Data Analysis 103 (2016) 229–241

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Functional regression approximate Bayesian computation for Gaussian process density estimation☆

G.S. Rodrigues [a,b,*], David J. Nott [c], S.A. Sisson [a]

[a] School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia

[b] CAPES Foundation, Ministry of Education of Brazil, Brasília - DF 70040-020, Brazil

[c] Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form 11 May 2016
Accepted 13 May 2016
Available online 26 May 2016

Keywords:
Approximate Bayesian computation
Nonparametric density estimation
Gaussian process prior
Hierarchical models

ABSTRACT

A novel Bayesian nonparametric method is proposed for hierarchical modelling on a set of related density functions, where grouped data in the form of samples from each density function are available. Borrowing strength across the groups is a major challenge in this context. To address this problem, a hierarchically structured prior, defined over a set of univariate density functions using convenient transformations of Gaussian processes, is introduced. Inference is performed through approximate Bayesian computation (ABC) via a novel functional regression adjustment. The performance of the proposed method is illustrated via simulation studies and an analysis of rural high school exam performance in Brazil.

1. Introduction

We introduce a new statistical procedure for hierarchical modelling on a set of related densities $f_i(x)$ for $i = 1, \ldots, g$, based on random samples $\{X_{ij}\}$ such that $X_{ij} \sim f_i(x)$ for $j = 1, \ldots, n_i$. The benefits of hierarchical modelling are well known in the parametric context, where the same model is fitted to different but related datasets or groups, and where model parameters are allowed to vary across groups (Gelman et al., 2004). A hierarchical prior may be defined on the group varying parameters, and the data used to determine how much pooling of information to perform across groups. The problem considered here is the nonparametric equivalent of this. Here, the nonparametric density functions $f_i(x)$ are thought to be related and we aim to share information hierarchically to improve estimation of each density, especially for those densities $f_i(x)$ for which the corresponding sample size n_i is small.

Bayesian nonparametric methods have made enormous advances over the last two decades. The Dirichlet process (DP) (Ferguson, 1973) has played a central role in this development. Methods based on Dirichlet process mixture (DPM) models, where a mixing distribution is given a Dirichlet process prior, are a standard approach to flexible Bayesian density estimation (Lo, 1984; West et al., 1994; Escobar and West, 1995). These methods also have extensions to grouped data and the estimation of a set of related density functions, the problem considered here. Important methods for grouped data include the analysis of densities model of Tomlinson and Escobar (1999) which uses a DPM model for each density in which the base measure for the mixing densities is the same and given a DPM prior; the hierarchical Dirichlet process (Teh et al., 2006) where mixing distributions are given DP priors with a common base measure which is given a DP prior; Dirichlet

☆ The *R* code used in Section 4 is available for download in the article's website (see Appendix A).

* Corresponding author at: School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia.
E-mail address: g.rodrigues@unsw.edu.au (G.S. Rodrigues).

http://dx.doi.org/10.1016/j.csda.2016.05.009

process mixture of ANOVA models (De Iorio et al., 2004) where the atoms in a Dirichlet process are modelled as dependent on a covariate following an ANOVA type dependence structure; and the hierarchical model of Muller et al. (2004) in which distributions are modelled as a mixture of a common and group specific component, with these components being given DP mixture priors.

One alternative to DP based methods in Bayesian nonparametrics is the use of Gaussian processes (Leonard, 1978; Thorburn, 1986; Lenk, 1988, 1991; Tokdar, 2007; Tokdar and Ghosh, 2007). However, to the best of our knowledge hierarchical versions of Gaussian process priors for grouped data situations have not been developed in the literature. Gaussian process methods can be attractive because the parameters in the resulting priors allow very easy expression of relevant prior information, such as smoothness of the densities and, in the hierarchical setting, the extent of sharing information between groups. For the DP mixture based methods, on the other hand, generally prior information must be expressed through a prior on a mixing distribution, through which it is difficult to adequately express similar prior beliefs. One reason that Gaussian process density estimation methods are not more popular is perhaps the computational difficulty. Such difficulties are even more acute in the hierarchical setting involving grouped data.

Here we introduce a hierarchical Gaussian process (HGP) prior, formulated as a univariate hierarchical extension of the multivariate prior proposed by Adams et al. (2009), and discuss tractable methods for computation with this prior. Our construction, besides being able to handle an arbitrary number of hierarchy levels, provides a convenient way of expressing prior beliefs regarding both the degree of similarity between the densities and the nature of their characterising features, such as smoothness and support. We also establish a remarkably different approach for making inference. Instead of relying on Markov chain Monte Carlo (MCMC) methods to draw samples from the posterior distribution, as is commonly implemented in other approaches to Gaussian process density estimation (Adams et al., 2009) which can suffer from poor performance, we alternatively introduce an approximate Bayesian computation (ABC) (Beaumont et al., 2002) functional regression-adjustment to draw approximate samples from the posterior. The use of ABC to estimate functional objects in itself represents an important contribution to the ABC literature. Moreover, this approach provides a great practical advantage in terms of flexibility, as, for most cases, changes to the prior specification are operationally straightforward to accommodate and do not require the derivation and coding of a new MCMC sampler.

In Section 2, we introduce the hierarchical Gaussian process prior and present an algorithm for sampling data from this prior. Section 3 describes the inferential strategy for estimating the functional parameters. Performance of the proposed density estimator is investigated in the simulation studies in Sections 4 and 5. Finally, in Section 6, we use the proposed estimator to compare rural high school exam performance across states in Brazil.

2. The hierarchical Gaussian process prior

We specify a hierarchical Gaussian process prior on the set of densities $f_i(x),\ i = 1, \ldots, g$, as follows:

$$f_i(x) = \frac{L(Z_i(x))b(x|\phi)}{c_i(\phi, Z_i)} \tag{1a}$$

$$Z_i(x) \sim \mathcal{GP}(\mu(x), k(x, x'|\theta_Z)) \tag{1b}$$

$$\mu(x) \sim \mathcal{GP}(m(x), k(x, x'|\theta_\mu)) \tag{1c}$$

$$(\theta_Z, \theta_\mu) \sim \pi(\cdot). \tag{1d}$$

Here $L(z) = 1/(1 + \exp(-z))$ denotes the logistic function and $\mathcal{GP}(\mu(x), k(x, x'|\theta))$ represents a Gaussian process with mean function $\mu(x)$ and covariance function $k(x, x'|\theta)$ with parameters θ. $b(x|\phi)$ is an arbitrarily chosen parametric base density with hyperparameters ϕ. It can be regarded as the modeller's 'best guess' of the densities $f_i(x)$ and the 'centre' of the prior distribution—the term $L(Z_i(x))$ acts by 'deforming' $b(x)$ to create new densities. The function $m(x)$ is a conveniently chosen function discussed further below, $\pi(\cdot)$ is a hyperprior for the parameters of each covariance function, and $c_i(\phi, Z_i) = \int L(Z_i(x))b(x|\phi)dx$ is the normalising constant of $f_i(x)$.

Gaussian processes are a common tool for modelling functional data and can be understood as an infinite-dimensional generalisation of the Gaussian distribution. A comprehensive review of Gaussian processes is found in Rasmussen and Williams (2006). See also Shi and Choi (2011) for a book length treatment of Gaussian processes in functional data analysis.

Eq. (1a) defines a deterministic map from auxiliary functions $Z_i(x)$ to the normalised densities $f_i(x)$. Operating on the transformed space avoids difficulties otherwise implied by the intrinsic properties of density functions—namely, that they be non-negative and integrate to one. Under this prior, all marginal densities are considered potentially related, and bound by a common Gaussian process mean function $\mu(x)$ (Eq. (1b)), which is itself unknown. The above formulation, though simple, will be shown to be highly adaptable, and capable of adequately describing a wide range of possible prior beliefs. For ease of presentation only two levels of hierarchy are described above, but it is straightforward to incorporate multiple hierarchical levels, as is illustrated in the analysis of Brazilian school exam performance in Section 6.

For this article, we adopt as covariance function the squared exponential kernel $k(x, x') = \sigma^2 \exp\left(-(x - x')^2/(2\ell^2)\right)$, although other options are easily accommodated. The hyperparameters $(\theta_Z, \theta_\mu) = (\sigma_Z, \ell_Z, \sigma_\mu, \ell_\mu)$ play a crucial role in the behaviour of $f_i(x)$. Fig. 1 shows $g = 5$ densities drawn from the proposed prior under various parameter settings for $(\sigma_Z, \ell_Z, \sigma_\mu, \ell_\mu)$. The standard uniform distribution was chosen as the base density $b(x|\phi)$ and we set $m(x) = -10$ (the motivation for this choice will be made clear in Section 3).