



# Copula based factorization in Bayesian multivariate infinite mixture models<sup>☆</sup>



Martin Burda<sup>a,b,\*</sup>, Artem Prokhorov<sup>c,d,e</sup>

<sup>a</sup> Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada

<sup>b</sup> IES, Charles University, Prague, Czech Republic

<sup>c</sup> Business School, University of Sydney, NSW, Australia

<sup>d</sup> CIREQ, 1455 de Maisonneuve Blvd West, Montreal, QC H3G 1M8, Canada

<sup>e</sup> Faculty of Economics, St. Petersburg State University, St. Petersburg, Russia

## ARTICLE INFO

### Article history:

Received 4 February 2013

Available online 1 March 2014

### AMS 2010 subject classifications:

62C10

62G07

65C40

62H99

### Keywords:

Nonparametric copula

Nonparametric consistency

Mixture modeling

## ABSTRACT

Bayesian nonparametric models based on infinite mixtures of density kernels have been recently gaining in popularity due to their flexibility and feasibility of implementation even in complicated modeling scenarios. However, these models have been rarely applied in more than one dimension. Indeed, implementation in the multivariate case is inherently difficult due to the rapidly increasing number of parameters needed to characterize the joint dependence structure accurately. In this paper, we propose a factorization scheme of multivariate dependence structures based on the copula modeling framework, whereby each marginal dimension in the mixing parameter space is modeled separately and the marginals are then linked by a nonparametric random copula function. Specifically, we consider nonparametric univariate Gaussian mixtures for the marginals and a multivariate random Bernstein polynomial copula for the link function, under the Dirichlet process prior. We show that in a multivariate setting this scheme leads to an improvement in the precision of a density estimate relative to the commonly used multivariate Gaussian mixture. We derive weak posterior consistency of the copula-based mixing scheme for general kernel types under high-level conditions, and strong posterior consistency for the specific Bernstein–Gaussian mixture model.

© 2014 Elsevier Inc. All rights reserved.

## 1. Bayesian nonparametric copula kernel mixture

Bayesian infinite mixture models are useful both as nonparametric estimation methods and as a way of uncovering latent class structure that can explain the dependencies among the model variables. Such models express a distribution as a mixture of simpler distributions without a priori restricting the number of mixture components which is stochastic and

<sup>☆</sup> We would like to thank the participants of the 6th Annual Bayesian Econometric Workshop of the Rimini Center for Economic Analysis, Toronto, 2012, the Canadian Economics Association Conference meetings, HEC Montreal, 2013, the Canadian Econometrics Study Group meetings, Kitchener, ON, 2013, the Midwest Econometrics Group meetings, Bloomington, IN 2013, the 19th International Panel Data Conference, London, UK, 2013 and the seminar audience at the University of New South Wales for their insightful comments and suggestions. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: [www.sharcnet.ca](http://www.sharcnet.ca)) and it was supported by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC: [www.sshrc-crsh.gc.ca](http://www.sshrc-crsh.gc.ca)).

\* Corresponding author at: Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada.

E-mail addresses: [martin.burda@utoronto.ca](mailto:martin.burda@utoronto.ca) (M. Burda), [artem.prokhorov@concordia.ca](mailto:artem.prokhorov@concordia.ca) (A. Prokhorov).

data-driven. In many contexts, a countably infinite mixture is also a more realistic model than a mixture with a small fixed number of components.

Even though their theoretical foundations were developed early [18,2,39], infinite mixture models have only recently become computationally feasible for practical implementation on larger data sets with the development of Markov chain Monte Carlo (MCMC) methods [16,42]. Bayesian infinite mixture models are becoming increasingly popular. Among the many areas of applications are treatment effects [11], autoregressive panel data [30], finance [31], latent heterogeneity in discrete choice models [33,7], contingent valuation models [19], and instrumental variables [12]. There is a growing number of applications in pattern recognition and other fields of machine learning (see, e.g., [17]), in biology (see, e.g., [37]), in network traffic (see, e.g., [1]), in DNA profiling (see, e.g., [65]), and other fields.

Most of these applications are univariate, or structured as conditionally independent copies of the univariate case, albeit in some fields, such as machine learning, infinite mixture models have been used in fairly high dimensions. Fundamentally, the computational complexity associated with algorithms based on popular nonparametric priors, such as the Dirichlet process, is not directly related to the dimensionality of the problem. However, the rapid increase in the number of mixture components required to represent a nonparametric dependence structure accurately in high dimensions poses a practical problem at the implementation level.

From a practical point of view, Dirichlet process priors can facilitate the feasibility of implementation by selecting relatively few latent classes in each MCMC step. However, this sparsity is traded off with the accuracy of estimation. The analyst does have the option to generate a large number of latent class proposals by tightening the prior of the concentration parameter in the Dirichlet process mixture. Nonetheless, many of the latent classes proposed in this way are likely to be specified over regions of the parameter space that are only weakly supported by the data. Hence, they will either not be accepted or quickly discarded during the MCMC run, leading to a noisy estimate. The severity of the trade-off is further exacerbated with higher dimensions: few mixing components can provide a very inaccurate representation of the data generating process, but strengthening the prior to increase their number will yield a higher rejection rate.

From our practical experience, we can alleviate some of these problems by relaxing the joint parametric specification of the mixing kernel, such as in the case of the typically used multivariate Gaussian kernel. If the joint parametric mixing kernel can be decomposed into flexible building blocks, each of which can be parsimoniously determined with high probability strongly supported by the data then we should expect to obtain a more accurate representation. However, decompositions by conditional expansions such as the Cholesky factorization of the covariance matrix of the multivariate Gaussian kernel still preserve the joint parametric dependence structure of the kernel. What is needed for our purpose is a decomposition of the dependence structure itself.

In this paper, we propose a copula-based factorization scheme for Bayesian nonparametric mixture models whereby each marginal dimension in the mixing parameter space is modeled as a separate mixture and these marginal models are then joined by a nonparametric copula function based on random Bernstein polynomials. In the implementation, only a few latent classes are required for each of the marginals, regardless of the overall number of dimensions. We show that this scheme leads to an improvement in the precision of a density estimate in finite samples relative to mixtures of Gaussian kernels, providing a suitable tool for applications requiring joint dependence modeling in the multivariate setting. Bearing in mind Freedman's [20] result concerning a topologically wide class of priors leading to inconsistent posteriors in Bayesian nonparametric models, we specify the conditions under which our approach yields posterior consistency; both for weak topologies for a general class of kernels and strong topologies for the specific case of random Bernstein polynomial copula and Gaussian mixture marginals (Bernstein–Gaussian mixture).

In a related literature Chen et al. [10] consider a copula sieve maximum likelihood estimator with a parametric copula and nonparametric marginals, while Panchenko and Prokhorov [43] analyze the converse problem, with parametric marginals and a nonparametric copula. In contrast, our procedure is based on both nonparametric copula and marginals. Moreover, in our case the number of mixture components is stochastic and automatically selected during the MCMC run, without the need for model selection optimization required for approaches based on maximum likelihood.

Nonparametric copula-based mixture models have been analyzed in several specific contexts distinct from ours. Silva and Gramacy [55] present various MCMC proposals for copula mixtures. Fuentes et al. [21] analyze a spatial Dirichlet process (DP) copula model based on the stick-breaking DP representation. Rey and Roth [49] introduce a copula mixture model to perform dependency-seeking clustering when co-occurring samples from different data sources are available. Their model features nonparametric marginals and a Gaussian copula with block-diagonal correlation matrix. Rodriguez et al. [50] construct a stochastic process where observations at different locations are dependent, but have a common marginal distribution. Dependence across locations is introduced by using a latent Gaussian copula model, resulting in a latent stick-breaking process. Parametric Bayesian copula models and their mixtures have also been analyzed in [48,56,3,14,38,5,29,64], among others.

Posterior consistency<sup>1</sup> can fail in infinite-dimensional spaces for quite well-behaved models even for seemingly natural priors [20,13,32]. In particular, the condition of assigning positive prior probabilities in “usual” neighborhoods of the

<sup>1</sup> Although consistency is intrinsically a frequentist property, it implies an eventual agreement among Bayesians with different priors. For a subjective Bayesian who dispenses of the notion of a true parameter, consistency has an important connection with the stability of predictive distributions of future observations — a consistent posterior will tend to agree with calculations of other Bayesians using a different prior distribution in the sense of weak topology. For an objective Bayesian who assumes the existence of an unknown true model, consistency can be thought of as a validation of the Bayesian method as approaching the mechanism used to generate the data [28]. As pointed out by a referee, our treatment of posterior consistency is concerned with the behavior of the posterior with respect to draws from a fixed sampling distribution, and so can be viewed as frequentist style asymptotics.

Download English Version:

<https://daneshyari.com/en/article/1145582>

Download Persian Version:

<https://daneshyari.com/article/1145582>

[Daneshyari.com](https://daneshyari.com)