



Comments on: Hybrid semiparametric Bayesian networks

Antonio Salmerón^{1,2} 

Received: 27 April 2022 / Accepted: 6 May 2022 / Published online: 13 June 2022
© The Author(s) 2022

The authors present an interesting work that extends their previous contribution on semiparametric Bayesian networks to a more general class of models, namely hybrid Bayesian networks, in which discrete (or categorical) and continuous variables coexist. The proposal consists of extending conditional linear Gaussian (CLG) networks by allowing some of the conditional densities in the model to be represented by a conditional kernel where some of the conditioning variables can be discrete or categorical. This is achieved by considering a different conditional kernel density for each possible value of the discrete/categorical variables. In practice, this is equivalent to treating the discrete variables as categorical, in the sense that their possible values do not explicitly appear in the closed-form formula of the corresponding kernel density. This is also the case of other formalisms for representing hybrid Bayesian networks, like the above-mentioned CLG networks and mixtures of truncated basis functions (MoTBFs) (Langseth et al. 2012).

The proposal in the paper is able to determine which densities are better represented by a CLG or a conditional kernel density, and therefore, the resulting model inherits the good properties of CLG networks in what concerns the estimation of both parameters and network structure from data. Regarding the estimation of the network structure, the semiparametric hybrid Bayesian networks also inherit the limitation of CLGs, more precisely, the restriction that conditional densities of discrete/categorical can only be defined given other discrete/categorical variables, but not other continuous variables. It means that discrete variables are not allowed to have continuous parents in the network. Whether or not this restriction is important strongly depends on what the model is meant to be used for. If the semiparametric hybrid Bayesian network is assumed to be a classifier (Pérez et al. 2006), it is not problematic. However, if the links in the network are expected to have a causal meaning, then it can be problematic, as some

This comment refers to the invited paper available at: <https://doi.org/10.1007/s11749-022-00812-3>.

✉ Antonio Salmerón
antonio.salmeron@ual.es

¹ Department of Mathematics, University of Almería, Ctra. Sacramento s/n, 04120 Almería, Spain

² Center for the Development and Transfer of Mathematical Research to Industry (CDTIME), University of Almería, Ctra. Sacramento s/n, 04120 Almería, Spain

causal relationships could not be represented (namely those in which a continuous variable is the cause of a discrete effect).

Perhaps the proposal could be enriched by allowing the possibility of having discrete variables conditional on continuous ones, using logistic regression or softmax functions to define the conditional model (Lerner et al. 2001). However, this would result in difficulties when trying to solve some typical tasks like probabilistic inference, as we will discuss later.

The authors propose to estimate the network structure by optimizing a penalized likelihood score. I find the discussion about the optimization process useful and interesting, especially the consideration about the risk of overestimating the goodness of fit of a model when using kernel densities if the training data is also used for evaluating the score, since the elements in the training sample would cause some of the $K_{\mathbf{H}}$ functions to be evaluated at their maximum. In order to avoid this, the authors proposed a cross-validated computation of the score.

An important issue when using kernel densities is the complexity of the resulting model. This is especially important within the context of Bayesian networks, which are commonly employed in high dimensional settings and/or in large sample problems, including streaming data. The case of streaming data is particularly problematic for kernel densities, since the sample size is continuously growing and thus the size of the required kernel densities would quickly become unmanageable. Typically, the fact that the sample incorporates new items thus requiring the estimation of the model parameters to be updated motivates that data streams are handled using distributions within the exponential family (Masegosa et al. 2020) or at least models whose components belong to the exponential family (Ramos-López et al. 2018).

Even in problems with limited sample size, the complexity of the kernel densities can render them inefficient even for simple evaluations, as was already pointed out by two of the authors in a previous work (López-Cruz et al. 2014) within the context of multivariate densities, where it is shown that the evaluation time is dramatically higher than the time required by MoP densities, while MoPs turned out to be as accurate as kernel densities. A similar finding would probably hold for conditional densities as well, and therefore, a comparison between semiparametric hybrid Bayesian networks and MoPs (or MoTBFs in general) seems to be a relevant subject for future research.

A typical task carried out over Bayesian networks is *probabilistic inference*, also known as *belief update*. Assume a Bayesian network over variables $\mathbf{X} = \{X_1, \dots, X_n\}$. The goal of probabilistic inference is to compute the density of some target variable $X_i \in \mathbf{X}$ given that some other variables $\mathbf{X}_E \in \mathbf{X}$ take on value $\mathbf{x}_E \in \Omega_{\mathbf{X}_E}$. In the case of a hybrid Bayesian network with unobserved discrete variables $\mathbf{X}_D \in \mathbf{X}$ and unobserved continuous variables $\mathbf{X}_C \in \mathbf{X}$, it amounts to computing

$$f(x_i | \mathbf{x}_E) = \frac{\sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} \int_{\Omega_{\mathbf{X}_C}} f(x_i, \mathbf{x}_C, \mathbf{x}_D, \mathbf{x}_E) d\mathbf{x}_C}{\sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} \int_{\Omega_{\mathbf{X}_i}} \int_{\Omega_{\mathbf{X}_C}} f(x_i, \mathbf{x}_C, \mathbf{x}_D, \mathbf{x}_E) d\mathbf{x}_C dx_i}. \quad (1)$$

Computing the joint density $f(x_i, \mathbf{x}_C, \mathbf{x}_D, \mathbf{x}_E)$ in Eq. (1) can be done efficiently taking advantage of the factorization induced by the Bayesian network structure. However, it can still be a difficult task if conditional kernel densities are used. As an example, consider a network with three continuous variables X, Y and Z and structure $X \rightarrow Y \rightarrow Z$. Assume we want to compute $f(z)$ (i.e., $\mathbf{X}_E = \emptyset$). This is achieved by calculating

$$f(z) = \int_{\Omega_X} \int_{\Omega_Y} f(x, y, z) dy dx = \int_{\Omega_X} \int_{\Omega_Y} f(z|y) f(y|x) f(x) dy dx. \quad (2)$$

If, for instance, the three conditional densities in Eq. (2) are conditional kernel densities, estimated from a sample of size $N = 1000$, the product would be, in the worst case, a density with 10^9 terms. This complexity can be somewhat sidestepped by using an approximate alternative. In that direction, the authors define a procedure for sampling from the conditional kernel densities, so that probabilistic inference can be carried out (even though in an approximate way) using Monte Carlo methods.

Besides the model complexity, the fact that two types of densities coexist (CLGs and conditional kernels) also represents a difficult problem from the point of view of probabilistic inference, as the result of multiplying both types of densities would belong to a different class of distributions, i.e., the marginal in Eq. (2) would not be a Gaussian nor a kernel. It would even be more problematic if logistic regression or softmax models were adopted in order not to restrict the possible network structures. Altogether, these considerations suggest that future research on semiparametric hybrid Bayesian networks might have good perspectives from the point of view of probabilistic inference.

I congratulate the authors for their paper, since it provides useful insight on a difficult subject where reaching a trade-off between accuracy and model complexity is difficult to find. Finally, I would like to thank the editors of TEST for giving me the opportunity to comment on this paper.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Langseth H, Nielsen TD, Rumí R, Salmerón A (2012) Mixtures of truncated basis functions. *Int J Approx Reason* 53:212–227
- Lerner U, Segal E, Koller D (2001) Exact inference in networks with discrete children of continuous parents. In: Proceedings of the 17th conference on Uncertainty in Artificial Intelligence (UAI-01), pp 319–332

- López-Cruz PL, Bielza C, Larrañaga P (2014) Learning mixtures of polynomials of multidimensional probability densities from data using B-spline interpolation. *Int J Approx Reason* 55:989–1010
- Masegosa AR, Ramos-López D, Salmerón A, Langseth H, Nielsen TD (2020) Variational inference over nonstationary data streams for exponential family models. *Mathematics* 8:1942
- Pérez A, Larrañaga P, Inza I (2006) Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *Int J Approx Reason* 43:1–25
- Ramos-López D, Masegosa AR, Salmerón A, Rumí R, Langseth H, Nielsen TD, Madsen AL (2018) Scalable importance sampling estimation of Gaussian mixture posteriors in Bayesian networks. *Int J Approx Reason* 100:115–134

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.