

Mixtures of Polynomial With Tails as Probability Distributions: Theoretical and Experimental Evaluation

Juan C. Luengo¹, Rafael Rumí^{1,2} , Ana D. Maldonado^{1,2} , Darío Ramos-López^{1,2}  *

¹ Department of Mathematics, University of Almería (Spain)

² Center for Development and Transfer of Mathematical Research to Industry, University of Almería (Spain)

* Corresponding author: jll249@inlumine.ual.es (J. C. Luengo), rrumi@ual.es (R. Rumí), ana.d.maldonado@ual.es (A. D. Maldonado), dramoslopez@ual.es (D. Ramos-López).

Received 29 April 2025 | Accepted 13 May 2026 | Published 28 August 2026



ABSTRACT

Bayesian networks based on mixtures of polynomials (MoPs) are used to model multivariate hybrid or continuous probability distributions. MoPs provide a flexible yet simple model to deal with probabilistic reasoning, approximating complex or empirical probability distributions. We review the main models based on MoPs in the literature, introducing an alternative MoP approach: the mixtures of polynomials with tails (tMoPs). Also, we present procedures for learning tMoP marginal and conditional densities from data. This learning algorithms are tested with many well-known probability distributions. In these experiments, the tMoP models yield satisfactory results in comparison to other techniques, including other available MoPs alternatives. We also propose a meta-model that can directly, using interpolation, provide a tMoP expression for a specific density within a given family. This idea is tested with several probability distributions of one and two parameters, giving promising results.

KEYWORDS

Approximation, Density Estimation, Hybrid Bayesian Networks, Mixtures of Polynomials.

DOI: 10.9781/ijimai.2026.2244
eLocator: e2244

I. INTRODUCTION

BAYESIAN networks (BNs) are flexible and powerful probabilistic graphical models used for probabilistic reasoning under uncertainty and decision-making problems [1], [2]. BNs have been employed in a large number of fields and real-world applications in many different domains [3]–[6]. Moreover, BNs are more explainable and understandable models than other state-of-the-art techniques.

Formally, a BN consists of an acyclic directed graph, that represents the statistical dependencies among variables, and a set of probability distributions of each variable conditioned on its parents in the graph. Thus, the joint probability factorizes as

$$p(\mathbf{X}) = p(X_1, \dots, X_s) = \prod_{i=1}^s p(X_i | \text{Pa}(X_i)), \quad X_i \in \Omega_i. \quad (1)$$

where p stands for a probability distribution, $\text{Pa}(X_i)$ represents the set of parents of variable X_i and Ω_i is the support of variable X_i .

The most interesting aspects of BNs are revealed when both continuous and discrete variables coexist within the same model (hybrid networks). In this scenario, it is crucial to have a uniform and

coherent representation of the probability distributions used, so that operations on them result in a distribution within the same family; otherwise, standard inference algorithms cannot be applied.

An alternative approach involves the use of discrete Bayesian Networks, which requires prior discretization of the relevant continuous variables. This discretization process can be implemented through relatively simple methods, such as partitioning the data into equal-length intervals, or by employing more sophisticated techniques like the Classification and Regression Trees (CART) method [7]. Another well-established option is the Conditional Linear Gaussian (CLG) model, which, despite not being strictly closed (since its posterior distributions are mixtures of Gaussian distributions), provides a foundation for integrating linear regression models between continuous variables [8], but assumes continuous variables follow a Gaussian distribution. Additionally, Kernel Density Estimation (KDE) techniques have been employed to model generic continuous variables in a non-parametric manner [9], [10], although these approaches lead to a substantial increase in both the number of parameters that must be estimated and the computational complexity of the inference process. In addition to the former ones, few alternatives exist that satisfy the closed-family requirement while minimizing the drawbacks of the

Please cite this article as:

J. C. Luengo, R. Rumí, A. D. Maldonado, D. Ramos-López. Mixtures of Polynomial With Tails as Probability Distributions: Theoretical and Experimental Evaluation, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 10, no. 1, e2244, 2026, <http://doi.org/10.9781/ijimai.2026.2244>

former approaches. One possibility is the use of Mixtures of Truncated Exponentials (MTEs; [11], [12]). Another option is BNs based on mixtures of polynomials [13], in which we focus on this paper. MTEs and mixtures of polynomials (MoPs) schemes are sometimes unified in the literature under the name *Mixture of Truncated Basis Functions* (MoTBFs; [14], [15]). The main advantage of MoPs over MTEs are the interpretability, since polynomials are simple to understand, and the operational ease, since the multiplication and integration of polynomials are trivially closed. However, this methodology is only partially developed, and most well-known BN techniques (such as learning and inference) have not yet been adapted to the MoPs case, with results on MoPs being scattered.

In this paper, which is an extension of the conference paper [16], we present mixtures of polynomials with tails (tMoPs) and compare them to other MoP approaches in the literature. Also, we relate the MoPs distribution to some other common continuous probability distributions. In [16], we presented a basic theoretical study about the MoPs distribution. In this paper, we focus on the study of the learning procedure for the probability distributions, leaving aside their usage within BNs. Both marginal (univariate) distributions and conditional distributions coexist in BNs, so learning procedures for both types have to be available. The main problem in hybrid BNs lies in accurately learning marginal distributions, since for conditional ones to continuous parents, the domain of the parent variables is split into subdomains, and a marginal density is then learned for each.

II. STATE OF THE ART

BNs based on MoPs [13], [17], [18] are used to model multivariate continuous probability distributions. Like MTEs [11], MoPs try to provide a flexible yet simple model to deal with probabilistic reasoning and approximate complex or empirical probability distributions.

In a nutshell, MoPs are defined as piecewise functions that are polynomials in each subdomain. In this section we define the MoP model and review the previous work on this topic, to build on them by presenting an approach that we will refer to as the tMoPs model in the next section.

MoPs were originally introduced in [13] with the next definition: A one-dimensional real function $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be a mixture of polynomials (MoP) if it is a non-negative piecewise function of the form

$$f(x) = \begin{cases} a_{0i} + a_{1i}x + a_{2i}x^2 + \dots + a_{ni}x^n & \text{for } x \in A_i, \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \dots, k \quad (2)$$

where $A_1, \dots, A_k$ (subdomains) are disjoint intervals in $\mathbb{R}$ that do not depend on x , and $a_{0i}, a_{1i}, a_{2i}, \dots, a_{ni}$ (coefficients) are real constants for all i .

Such a function is said to be a k -piece, n -degree MoPs function or potential. In addition, if it integrates up to one, it is said to be a MoPs density.

The main motivation for working with models based on MoPs is the ease of working with them both analytically and computationally. Furthermore, they are closed under addition, differentiation, multiplication, marginalization and integration. Additionally, learning the parameters is easier than in MTEs, as there are many ways to obtain them in the literature (Taylor expansion, splines, regression...).

However, if the amount k of subdomains or the degree n are large, these general MoPs may become too complex models and prone to overfitting.

A. Shenoy-West MoPs Approximation

Shenoy and West present in [19] a method based on the Taylor series expansion, so it requires the analytical expression of the density function to be approximated. The non-negativity of the Taylor approximation over the entire domain is not guaranteed. In case it presents some roots, two possibilities are provided: either modify the degree of the polynomial, or increase the number of subintervals, thereby increasing considerably not only the accuracy but also the complexity of the fit.

A standardized method to fit any one-dimensional distribution is not presented, but the characteristics such as extremes or inflection points are studied beforehand to decide the degree of the polynomials and the number of intervals into which the domain is divided.

A restricted version of Shenoy's method [19] has been used in this paper. The Taylor expansion series, which we will call $t(x)$, has been computed with respect to the midpoint of the interval Ω_X :

$$t(x) = \begin{cases} \sum_{i=0}^n \frac{f^{(i)}(x_m)}{i!} (x - x_m)^i & \text{for } x \in \Omega_X, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $f(x)$ is the density function of X and x_m is the midpoint of the interval Ω_X .

Using the maximum allowed degree, and since the non-negativity of the polynomial in that region is not guaranteed, whenever $t(x) < 0$ for some $x \in \Omega_X$, the degree n is reduced. If that situation is repeated, ultimately a uniform distribution may be employed, i.e., with $n = 0$.

B. B-Splines MoPs Approximation

In [18], the authors present a new method for learning MoPs based on uniform B-splines. The domain Ω_X is split into $L_X + 1$ equi-spaced subintervals with knots $\{a_0, \dots, a_{L_X}\}$, defining $M_X = L_X + r_X - 1$ B-splines with order r_X as follows:

$$B_j(x) = (a_j - a_{j-r_X}) \cdot H(x - a_{j-r_X}) \quad (4)$$

$$\sum_{i=0}^{r_X} \frac{(a_{j-r_X+i} - x)^{r_X-1} \cdot H(a_{j-r_X+i} - x)}{w'(a_{j-r_X+i})}, \quad x \in \Omega_X \quad (5)$$

where $j = 1, \dots, M_X$ and $w'_{j-r_X}(x)$ is the first derivative of $w_{j-r_X}(x) = \prod_{u=0}^{r_X} (x - a_{j-r_X+u})$. Also $H(x)$ is the Heaviside function.

Starting from a sample $\{x_1 \dots x_N\} \in \Omega_X = [a, b]$, the MoPs $\varphi(x; \alpha)$ can be defined as a linear combination of the previous M_X B-splines:

$$\varphi(x; \alpha) = \sum_{j=1}^{M_X} \alpha_j \cdot B_j(x), \quad x \in \Omega \quad (6)$$

where $\alpha_1 \dots \alpha_{M_X}$ are the mixing coefficients calculated using the recursive formula:

$$\hat{\alpha}_j^{(q)} = \frac{1}{N \cdot c_j} \sum_{i=1}^N \frac{\hat{\alpha}_j^{(q-1)} \cdot B_j(X_i)}{\varphi(X_i; \hat{\alpha}^{(q-1)})} \quad (7)$$

where q is the iteration number in the process and $c_j = \frac{a_j - a_{j-r_X}}{r_X}$.

C. MoTBF Model

MoTBFs, introduced in [20] is a general model based on the abstract notion of real-valued basis functions $\psi(\cdot)$, so that an MoTBF can be defined as a function $g_k : \Omega \rightarrow \mathbb{R}_0^+$ that:

$$g_k = \sum_{i=0}^k a_i \psi_i(x) \quad (8)$$

where $a_i \in \mathbb{R}$. It can be noted that particular cases of this type of functions are MoPs and MTEs.

To fit a tMoP it starts with a sample $\{x_1 \dots x_N\} \in \Omega_X = [a, b]$ from which the empirical distribution function is obtained as:

$$G_N(x) = \frac{1}{N} \sum_{i=1}^N 1\{x_i \leq x\}, \quad x \in \mathbb{R} \quad (9)$$

Lets denote $\eta_i(x)$ to the anti-derivative of ψ_i , so the next minimization problem with restrictions must be solved to fit the cumulative distribution function (CDF):

$$\begin{aligned} \min_{c_0 \dots c_k} \quad & \sum_{i=1}^N \left(G_N(x_i) - \sum_{j=0}^k c_j \eta_j(x_i) \right)^2 \\ \text{subject to} \quad & \sum_{j=1}^{k-1} c_j \psi_j(x) \geq 0, \quad \forall x \in \Omega_X \\ & \sum_{j=0}^k c_j \eta_j(a) = 0 \quad \text{and} \quad \sum_{j=0}^k c_j \eta_j(b) = 1 \end{aligned} \quad (10)$$

III. METHODS

In this section, we define the mixtures of polynomials with tails (tMoPs) and we propose learning algorithms for univariate and conditional densities. Also, we propose a meta-model to obtain a tMoP that fits some probability distribution with certain parameters based on interpolation of previously learned distributions.

A. tMoPs Model

This method is a particular case of the original MoPs definition made in [13] where only a maximum of three pieces are taken: a polynomial function in the central one and two constant tails in the other two subintervals (that appear only if they are necessary).

Definition of a Mixture of Polynomials with tails function (tMoP):

A one-dimensional function $f: \mathbb{R} \rightarrow \mathbb{R}$ is said to be a tailed mixture of polynomials function if it is a piecewise function of the form

$$f(x) = \begin{cases} b_1 & \text{for } x \in [l_0, l_1], \\ \sum_{i=0}^n a_i x^i & \text{for } x \in [l_1, l_2], \\ b_2 & \text{for } x \in [l_2, l_3], \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $l_0 \leq l_1 < l_2 \leq l_3$ and the coefficients $a_0, a_1, a_2 \dots a_n$ are constants. The values b_1 and b_2 are called the MoPs tails and they are relevant only if $l_0 < l_1$ or $l_2 < l_3$ (i.e., when $[l_0, l_1]$, $[l_2, l_3]$ or both are proper intervals).

When we try to fit a density using a polynomial it is common that there are areas where it is negative, especially those where the probability is close to 0. Achieving non-negativity in a polynomial fit is not an easy task since it generally causes a loss of accuracy, especially in the areas near the extremes of the domain. To avoid this mismatch, the polynomial fit is restricted to the area where it is positive (this process is explained in detail in Section III.B.) using a 0-degree polynomial, i.e. a constant, to fit the tails. In that case we can get the advantages similar to discretization in the tails, and a better fit in the center of the distribution.

The main properties of tMoPs as probability distributions, such as the moments (expectation, variance, ...), cumulative distribution function, or the moment generating function can be consulted in [16].

B. Learning tMoP Parameters for Marginal Densities

Considering a sample of size N of X in its domain $\Omega_X = [X_{min}, X_{max}]$, we can note $\{X_{min} = X_1 \leq X_2 \leq \dots \leq X_{max} = X_N\}$, our goal is to estimate a univariate tMoP density as defined in Section III.A.

We will split this process into two steps: firstly, we will calculate the set of kernel points; and secondly, we will calculate the parameters for the polynomial fit and the tails, if necessary. A brief scheme of

the whole process can be seen in Algorithm 1 and its corresponding flowchart in Fig. 1.

Algorithm 1. Algorithm to learn an univariate tMoP from a sample.
New tMoP($X, d_{max}, X_{min}, X_{max}$)

Input: Sample $X = \{x_1, \dots, x_N\}$ of variable X . Maximum degree d_{max} . Minimum and maximum possible values of variable X .

Output: Normalized polynomial, roots and probability of tails.

```

1: begin
2:   Calculate balanced kernel points  $X' = \{(x, y)\}$  from  $X$ .
3:   Split  $(X_{min}, X_{max})$  into 10 subintervals
4:   Filter the kernel points of the subintervals with less than 10%
     of the points
5:   Let  $d = d_{max}$ .
6:   while  $d \geq 0$  and BIC increases do
7:     Fit  $p(x)$  of degree  $d$ .
8:     Find roots  $r_1, \dots, r_k$  of  $p(x)$  in  $(X_{min}, X_{max})$ .
9:     if  $k = 0$  then
10:      No tails ( $t_1 = t_2 = 0$ ).
11:       $(r_i, r_{i+1}) = (X_{min}, X_{max})$ .
12:      Normalize  $p(x)$  in  $(X_{min}, X_{max})$ .
13:     else
14:      Find the most suitable interval  $R_i = [r_i, r_{i+1}]$ 
15:      Compute the tails  $t_1$  and  $t_2$  keeping the same
        probability in that interval.
        Normalize  $p(x)$  in  $R_i$ .
16:     end
17:     Compute BIC.
18:      $d \leftarrow d - 1$ .
19:   end
20:   return  $p(x), (r_i, r_{i+1}), (t_1, t_2)$ .
21: end

```

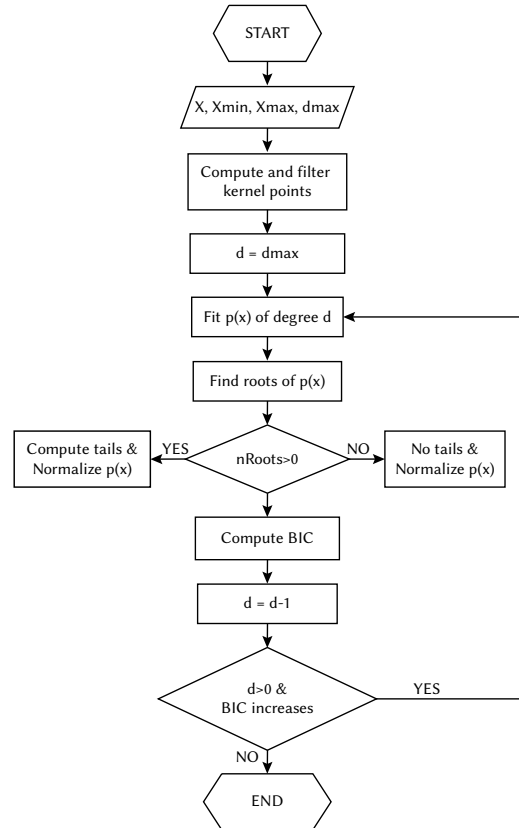


Fig. 1. Flowchart of the algorithm to learn an univariate tMoP from a sample.

1. Obtaining Balanced Kernel Points

We start estimating an empirical density from the sample through the Kernel Smoothing technique [23]. We consider a set of kernel points $\{(x, y)\}$, with y being the kernel density value at the point x in the domain of X . Initially, we take 500 equispaced points in Ω_X as follows.

In the way the kernel points are obtained, the density in the areas near the extremes is lower. In order to avoid this distortion, the sample points are mirrored to the right taking as symmetry point X_{max} (likewise to the left) forming a sample of size $3N$ as follows: $\{X_N^{(l)}, \dots, X_1^{(l)}, X_1, \dots, X_N, X_N^{(r)}, \dots, X_1^{(r)}\}$. From that sample we compute the kernel points (x, y) which are subsequently restricted to the domain Ω_X having kernel points (x', y') such that $x' \in \Omega_X$.

The next step is to match the kernel point density to the distribution of the sample along Ω_X . The set of kernel points is equispaced; however, it is common in samples to have areas with higher kernel point density and areas with lower kernel point density. This is corrected by splitting the domain into 10 disjoint subintervals $I_1, \dots, I_{10}$ and calculating the proportion of points from the original sample in each of these intervals as follows:

$$p_i = \frac{\#\{X_j : X_j \in I_i\}}{N} \quad (12)$$

Since the kernel points are equispaced, in each interval I_i there will be 10% of the points; however, in those intervals where $p_i < 0.1$ we randomly eliminate kernel points until their proportion is p_i .

2. Building a tMoP Density From a Polynomial Fit

Now, a polynomial fit of the above kernel points must be performed using least squares regression.

The process begins by establishing a maximum degree for the polynomial, which we will call n , and then fitting a polynomial $p(x)$ of that degree. Several scenarios are possible in such a case.

If the polynomial has no roots within the domain, then $p(x) > 0, \forall x \in \Omega_X$ and the tMoP will consist of a single truncated polynomial, without the need to set tails, which will eventually be normalized to ensure that the tMoP is a density.

If there are roots, it is necessary to select the interval in which the polynomial will be defined, leaving the rest as one or two tails. Let $R = \{X_{min} = r_0, r_1, \dots, r_k, X_{max} = r_{k+1}\}$ be the set of roots in Ω_X , including the two endpoints, and let $R_i = [r_i, r_{i+1}]$ denote each of the subregions.

Among the regions R_i where $p(x) > 0$ we calculate the one with the highest density in terms of kernel points, R_A , and also the one that covers the largest range of the domain, R_B . If they are different, two tMoPs, $f_A(x)$ and $f_B(x)$, are created with their central parts in R_A and R_B , respectively, and one or two tails are created in the side regions (if necessary) by dividing the proportion of kernel points by the length of the interval. Between these two models, the one that minimizes the mean square error with respect to the kernel points is selected.

Then the degree is reduced and the above scheme is followed. This process continues as the Bayesian information criterion (BIC) [24] increases, which may lead to reaching degree 0 and getting a uniform density. The process concludes by normalizing the polynomial along with the tail or tails so that it becomes a density function. The formula used to calculate the BIC is specified in Section III.E.

C. Learning tMoP Parameters for Conditional Densities

Since Bayesian networks contain conditional distributions in their topology, it is necessary to derive a learning algorithm to fit a conditional density. In the next sections, we derive the basics of such algorithm, beginning with the simplest case (a single parent) and

showing the way in which this can be extended to multiple parents. Details are specified in the case of two parents, but the methodology can be extended to a general case.

In all cases, the domain of the continuous parent variables is split and, for each subdomain, a marginal density is fitted. The problem is to find the partition of the parent domain so the the global fit of the conditional density is optimal. This methodology fits also for the hybrid case, when the parent set contains discrete variables. In that case the procedure is standard, fit a only-continuous distribution for each combination of the states of the discrete parents. Since that solution is obvious, we will focus in the continuous case, i.e., when all variables are continuous.

1. One Continuous Parent

A partition of the domain of the parent Y , Ω_Y , is carried out, so that a marginal fit for X is computed in each subinterval of the partition. An iterative process begins by dividing Ω_Y into halves and fitting the corresponding marginals in each subinterval. Then, each subinterval is split again, the corresponding BIC is computed for each possible division, and the best one in terms of BIC is selected. In this partition process, one interval can be split at most three times. Therefore, if an interval covers $\frac{1}{8}$ of Ω_Y , it is not split again.

The maximum number of subintervals is 8 when it is split uniformly along the domain, however it may be that some parts of the domain are split repeatedly while others keep unchanged. In both cases it depends on the distribution of the observations. Divisions in which $\frac{\#\{X : Y \in \Omega_Y\}}{N} < 0.05$ for any of the new intervals are dismissed.

Details can be found in [23].

2. Two Continuous Parents

A process of dynamic partitioning the domains of both parents Y and Z is carried out. They are simultaneously split by choosing the best fit according to the BIC.

Let us call Ω the two-dimensional domain of Y and Z . Following the same idea as before, the domain of each parent is splitted into two halves, obtaining 4 sub-regions. An initial fit of $f(X|Y, Z)$ is obtained learning a marginal distribution for X in each sub-region of Ω .

Then, an iterative process begins in which each subregion dimension is split in two halves, obtaining 8 new partitions of Ω , and fitting $f(X|Y, Z)$ for each one. The best partition is selected in terms of the BIC, obtaining now a 5 sub-regions partition.

This process is repeated whenever the following three conditions are satisfied: the BIC increases, the number of observations of the child within each subinterval exceeds 5% of the total observations; i.e., $\frac{\#\{X : P \in \Omega_i\}}{n} \geq 0.05$ and each subinterval covers at least one eighth of the total domain of each parent.

This process, schematized in Algorithm 2, leads to a partition of the domain Ω into rectangles of different shapes.

D. tMoP Meta-model for Common Probability Distributions

Since known probability distributions can model properly specific real-world situations it is usual that, when modeling real problems, the distribution of some of the variables involved is known except for some of their parameters. This section shows a new method to learn a distribution of any parameter from a finite set of previously fitted distributions, in order to avoid the computational complexity of fitting new distributions from scratch.

In the case of the Gaussian (normal) distributions, to obtain a MoP approximation to a general $N(\mu, \sigma)$ density we only need to transform the MoP approximation to the $N(0, 1)$. Let F be the cumulative distribution function of $X \rightarrow N(\mu, \sigma)$, and G the cumulative distribution function of $Z \rightarrow N(0, 1)$. Then it is known that

Algorithm 2. Estimation of a tMoP density with two continuous parents.

Function Fit2ContParents($\mathbf{X}, \mathbf{Y}, \mathbf{Z}$)

Input: Observations $\mathbf{X} = \{X_1, \dots, X_N\}$. Its parents

$\mathbf{Y} = \{Y_1, \dots, Y_N\}$ and $\mathbf{Z} = \{Z_1, \dots, Z_N\}$.

Output: Conditional density. Partition matrix of $\mathbf{Y}$ and $\mathbf{Z}$.

```

1: begin
2:   Split domain in 4 rectangles  $\Omega_1, \dots, \Omega_4$  by mid points of  $\mathbf{Y}, \mathbf{Z}$ .
3:   Fit  $f(\mathbf{X}|\mathbf{Y}, \mathbf{Z})$ .
4:   Compute BIC.
5:    $nRegions \leftarrow 4$ .
6:   while BIC increases do
7:     for  $i \leftarrow 1$  to  $nRegions$  do
8:       Split  $\Omega_i$  by the mid point of  $\mathbf{Y}$ .
9:       Fit  $f(\mathbf{X}|\mathbf{Y}, \mathbf{Z})$  in each region and calculate BIC.
10:      Split  $\Omega_i$  by the mid point of  $\mathbf{Z}$ .
11:      Fit  $f(\mathbf{X}|\mathbf{Y}, \mathbf{Z})$  in each region and calculate BIC.
12:    end
13:    Select highest BIC partition.
14:     $nRegions \leftarrow nRegions + 1$ .
15:  end
16:  return  $f(\mathbf{X}|\mathbf{Y}, \mathbf{Z})$ , partition matrix.
17: end
    
```

$$F(x) = P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = G\left(\frac{x - \mu}{\sigma}\right). \quad (13)$$

To obtain the corresponding density functions of $\mathbf{X}$, f and of $\mathbf{Z}$, g , we derive and get

$$f(x) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right) \quad (14)$$

So, to obtain the corresponding MoP approximation to a general $N(\mu, \sigma)$ we need to apply the former transformation to the MoP approximation to the standardized Normal distribution.

To complete the definition of tMoP as expressed in Section III.A, we need the roots and the height of the tails, if any.

If the roots obtained in the standard Normal distribution are $\{r_{min}, r_{max}\}$, the new roots will be $\{r'_{min}, r'_{max}\} = \{r_{min} \cdot \sigma + \mu, r_{max} \cdot \sigma + \mu\}$.

The tails must also be rescaled according to the length of the domain in which they are defined. If the tails obtained in the standard Normal distribution are $\{t_{min}, t_{max}\}$, defined in (x_{min}, r_{min}) and (r_{max}, x_{max}) respectively, the new tails will be:

$$\{t'_{min}, t'_{max}\} = \left\{ t_{min} \cdot \frac{r_{min} - x_{min}}{r'_{min} - x'_{min}}, t_{max} \cdot \frac{x_{max} - r_{max}}{x'_{max} - r'_{max}} \right\} \quad (15)$$

where (x'_{min}, x'_{max}) is the domain of definition of the tMoP of $\mathbf{X}$ that, in order to maintain the same structure of the original one, it should contain both roots, if they exist.

The remaining known distributions do not support such transformation. However, thanks to the ease with which MoPs can be combined, it is also possible to combine two tMoPs efficiently without the need to learn a new model from a sample. During the process of learning a BN, it is necessary to learn a large number of distributions. As described in Section III.C, fitting a conditional density with a tMoP involves an iterative process of fitting marginal distributions. Therefore, having a previously learned database is especially efficient to reduce the computational cost of the process.

We will now develop the interpolation process for a distribution with two parameters, $\alpha \in \Omega_\alpha$ and $\beta \in \Omega_\beta$. The fitting process for distributions with one parameter is a particular case.

The previous step is to learn a set of distributions within those parameters intervals. Let $\alpha_0, \dots, \alpha_n$ be $n+1$ equispaced points in $\Omega_\alpha = [\alpha_0, \alpha_n]$ and $\beta_0, \dots, \beta_m$ be $m+1$ equispaced points in $\Omega_\beta = [\beta_0, \beta_m]$.

In total, $n \cdot m$ subintervals have been created in the bidimensional domain $\Omega_\alpha \times \Omega_\beta$. The grid is made of $(n+1) \cdot (m+1)$ points. A tMoP is fitted in each point of the grid.

When we want to fit a distribution $B(\alpha', \beta')$ we search the four points between which it is located, i.e. $\alpha' \in [\alpha_p, \alpha_{p+1}]$ and $\beta' \in [\beta_q, \beta_{q+1}]$. The four selected points would be $A = (\alpha_p, \beta_q)$, $B = (\alpha_p, \beta_{q+1})$, $C = (\alpha_{p+1}, \beta_q)$ and $D = (\alpha_{p+1}, \beta_{q+1})$. In distributions with one parameter, the previous step can be omitted since the two closest points are obtained directly.

Using the Euclidean distance between the parameters, we select the two closest ones. Without loss of generality, let be $A = (\alpha_p, \beta_p)$ and $B = (\alpha_p, \beta_{q+1})$ whose distances are d_A and d_B . An scheme of this procedure can be seen in Fig. 2.

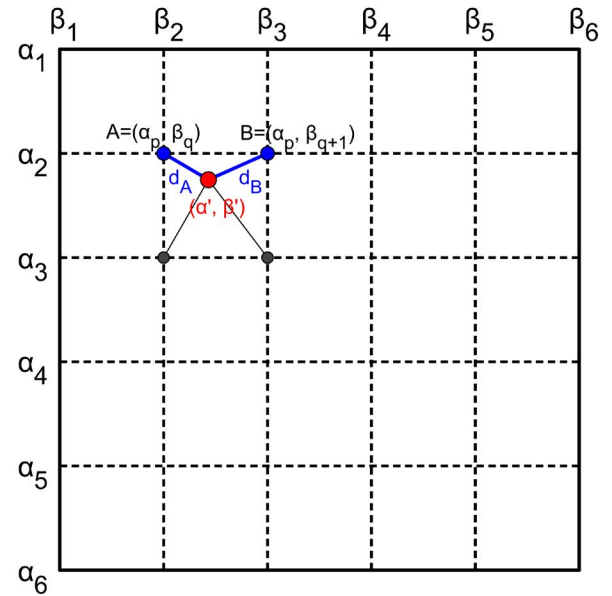


Fig. 2. Simplified diagram of neighboring points in a grid.

The new polynomial will be a weighted mean of the polynomials of the previous distributions taking the weights inversely proportional to d_A and d_B . Regarding the tails, if any, the same procedure is performed.

Suppose that the roots of the two learned distributions are $r_{Amin} < r_{Amax}$ and $r_{Bmin} < r_{Bmax}$, in that case the roots of the new model will be:

$$(\max\{r_{Amin}, r_{Bmin}\}, \min\{r_{Amax}, r_{Bmax}\}), \quad (16)$$

so the non-negativity of the new polynomial is guaranteed. In the case one distribution has zero or one root, the maximum or minimum will be used instead.

As the final step, the tMoP thus obtained is normalized.

This process is schematized in Algorithm 3.

E. Metrics for Goodness-of-fit Assessment

To assess the goodness-of-fit of an approximate univariate density $\tilde{f}(x)$, given by any MoP alternative, to a known probability density function $f(x)$, we use the Kullback-Leibler (KL) divergence, defined as:

Algorithm 3. Algorithm to interpolate a two parameter univariate distribution.

tMoPs interpolation($\Omega_\alpha, \Omega_\beta, n, m, \alpha', \beta'$)

Input: Intervals Ω_α and Ω_β , domain of definition of α and β . Number of subdivisions n and m of Ω_α and Ω_β , respectively. Parameters α' and β' of the target distribution.

Output: Normalized polynomial, roots and probability of tails.

```

1: begin
2:   Create  $\alpha_0, \dots, \alpha_n$  and  $\beta_0, \dots, \beta_m$  equispaced points in  $\Omega_\alpha$  and  $\Omega_\beta$ 
3:   Fit a tMoP for each point  $(\alpha_i, \beta_j)$  of the grid
4:   Find 4 points of the grid surrounding  $(\alpha', \beta')$ 
5:   Get the 2 closest points of the grid using the euclidean distance:  $P_A$  and  $P_B$ 
6:   Let  $d_A$  and  $d_B$  be the distance from  $(\alpha', \beta')$  to  $P_A$  and  $P_B$ 
7:   Compute the weights  $w_A = \frac{f}{d_A}$  and  $w_B = \frac{f}{d_B}$  where  $f = \frac{1}{\frac{1}{d_A} + \frac{1}{d_B}}$ 
8:   Calculate the polynomial as  $p'(x) = w_A \cdot p_A(x) + w_B \cdot p_B(x)$ 
9:   Let  $r_A$  and  $r_B$  the left roots and  $t_A$  and  $t_B$  the left tails (same to right)
10:  if  $t_A > 0$  or  $t_B > 0$  then
11:    Compute new left tail:  $t' = w_A \cdot t_A + w_B \cdot t_B$ 
12:    Compute the left root:  $r' = \max(r_A, r_B)$ 
13:  Normalize the tMoP
14: end
15: return  $\{p'(x), r', t'\}$ .
```

$$D_{KL}(f(x), \bar{f}(x)) = \int_{\Omega} f(x) \log \frac{f(x)}{\bar{f}(x)} dx \quad (17)$$

In the case of a conditional density, the KL divergence is computed as in (17) but replacing the marginal densities by the corresponding conditional densities. As an example, in the case of only one parent, the KL divergence between the approximate density $\bar{f}(y|x)$ and the true conditional density $f(y|x)$ is computed as:

$$\begin{aligned} D_{KL}(f(y|x), \bar{f}(y|x)) &= \mathbb{E}_{f(y|x)} \left[\frac{f(y|x)}{\bar{f}(y|x)} \right] = \\ &= \int_X \int_Y f(x, y) \cdot \log \frac{f(y|x)}{\bar{f}(y|x)} dy dx = \\ &= \int_X \int_Y f(x) \cdot f(y|x) \cdot \log \frac{f(y|x)}{\bar{f}(y|x)} dy dx. \end{aligned} \quad (18)$$

Finally, to measure the goodness-of-fit for a distribution learned from real data, for which the true distribution is unknown, we use the Bayesian Information Criterion (BIC), given by:

$$BIC(f) = \sum_{i=1}^N \log f(x_i) - \frac{k}{2} \log N \quad (19)$$

where f can be a univariate or conditional probability distribution of certain variables $\mathbf{X}$, $\{x_i\}$ is a N -size sample for the variables $\mathbf{X}$ that appear in f , and k is the number of parameters in the probability density function f .

IV. RESULTS

In this section, we conduct a series of experiments to validate the methods for tMoPs presented in Section III, compare them with the other MoP alternatives and assess their performance in various scenarios. For MoPs based on B-splines, the method is not directly available in any published piece of software and therefore we have not included it in the comparison.

A. Approximation to Common Univariate Probability Distributions

MoPs can be used to approximate any probability distribution, and we test their goodness of fit with six well-known distributions with different parameters. For distributions whose domain is unbounded, it has been truncated in such a way that the probability of the resulting finite subdomain is at least 0.99. The distributions used are:

- Exponential with three different parameters: $Exp(1)$ on the interval $[0, 6]$, $Exp(\frac{1}{3})$ on the interval $[0, 20]$ and $Exp(3)$ on the interval $[0, 3]$.
- Standard normal $N(0, 1)$ on the interval $[-3, 3]$.
- Gamma with three parameter combinations: $\Gamma(6, 1)$ on the interval $[0, 15]$, $\Gamma(8, 1)$ on the interval $[0, 17]$ and $\Gamma(11, 1)$ on the interval $[0, 20]$.
- Beta with five parameter combinations: $\beta(0.4, 0.6)$, $\beta(1.1, 3)$, $\beta(2.5, 2.5)$, $\beta(2.7, 1.3)$ and $\beta(1.3, 2.7)$.
- Lognormal with three parameter combinations: $LN(0, 0.25)$ on the interval $[0, 4]$, $LN(0, 0.5)$ on the interval $[0, 6]$ and $LN(0, 1)$ on the interval $[0, 11]$.
- Chi-square with five different degrees of freedom: $\chi(10)$ on the interval $[0, 27]$, $\chi(8)$ on the interval $[0, 25]$, $\chi(6)$ on the interval $[0, 22]$, $\chi(4)$ on the interval $[0, 17]$ and $\chi(2)$ on the interval $[0, 13]$.
- Student's t distribution with three different degrees of freedom: $t(1)$ in the interval $[-30, 30]$, $t(2)$ in the interval $[-8, 8]$ and $t(4)$ in the interval $[-5, 5]$.
- Weibull distribution with five parameter combinations: $W(1, 1)$ in the interval $[0, 5]$ (equivalent to $Exp(1)$), $W(1.5, 1)$ in the interval $[0, 3]$, $W(2, 2)$ in the interval $[0, 5]$, $W(6, 3)$ in the interval $[0, 4]$ and $W(8, 4)$ in the interval $[0, 5]$.
- Bimodal as three different linear combination of two Normal distributions: $0.25 \cdot N(0, 1) + 0.75 \cdot N(6, 2)$ on the interval $[-3, 12]$, $0.4 \cdot N(1, 1) + 0.6 \cdot N(11, 3)$ on the interval $[-2, 20]$ and $0.5 \cdot N(1, 1) + 0.5 \cdot N(7, 1)$ on the interval $[-2, 10]$.

We compare the proposed tMoP approach (Section III.A) with the previous methods of Shenoy and MoTBFs (Section II.A).

For Shenoy's method, two parameters must be fixed: the number of equal-length subintervals into which the domain is divided, and the maximum degree of the polynomial in each interval. We conducted experiments using between 1 and 4 subintervals, with maximum polynomial degrees of 7, 4 and 3.

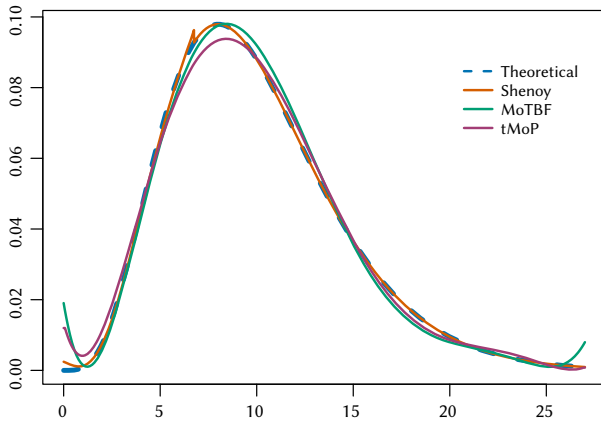
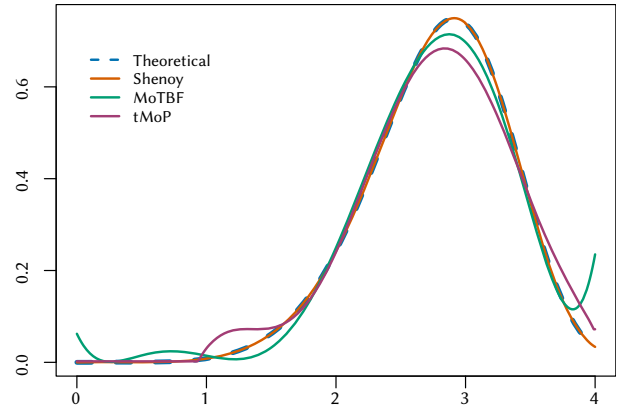
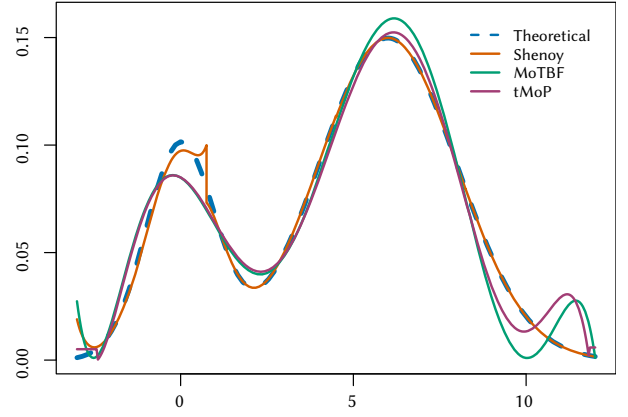
For MoTBF and tMoP approaches, unlike Shenoy's method, it is not necessary to know the analytic expression of the distribution, but only a sample. Thus, random samples of sizes 50, 100 and 1000 have been drawn from the true distribution. For each size, the fitting process was repeated 10 times taking different samples, and the results were averaged.

We assess the quality of fit of each alternative using the KL divergence (see Section III.E). In total, 12 combinations of parameters were used for Shenoy's method and 9 for MoTBFs and tMoPs, for each univariate distribution tested. The best result obtained for each method is shown in Table I. In general, the best results were obtained, as expected, for the largest number of subintervals, maximum degree or sample size.

Some examples of the resulting fits can be found in Fig. 3, Fig. 4 and Fig. 5. The accuracy provided by Shenoy's method with 4 subintervals is clearly visible. On the other hand, the difference between the MoTBF and tMoP fit are usually found at the extremes of the domain. These errors typical of the polynomial fit can be mitigated by using the tails (as in Fig. 4 where the left tail avoids big errors).

TABLE I. BEST-CASE KULLBACK-LEIBLER DIVERGENCE FOR COMMON PROBABILITY DISTRIBUTIONS FIT USING MoPs

PDF	Shenoy	MoTBF	tMoPs
$Exp(1)$	1.603e-13	3.274e-03	7.715e-03
$Exp(1/3)$	8.834e-13	2.498e-03	6.309e-03
$Exp(3)$	1.169e-10	4.504e-03	9.311e-03
$N(0,1)$	7.482e-09	9.547e-03	1.034e-02
$\Gamma(6,1)$	2.174e-03	1.471e-02	8.922e-03
$\Gamma(8,1)$	3.045e-04	1.254e-02	1.067e-02
$\Gamma(11,1)$	5.134e-04	1.867e-02	1.243e-02
$\beta(0.4,0.6)$	1.221e-01	1.282e-01	1.293e-01
$\beta(1.1,3)$	1.967e-04	4.000e-03	4.933e-03
$\beta(2.5,2.5)$	6.691e-02	2.575e-03	8.424e-03
$\beta(2.7,1.3)$	1.311e-02	4.357e-03	5.727e-03
$\beta(1.3,2.7)$	1.311e-02	4.347e-03	4.920e-03
$LN(0,0.25)$	6.563e-02	1.435e-01	7.049e-01
$LN(0,0.5)$	1.339e-01	9.417e-02	1.193e-01
$LN(0,1)$	5.099e-02	7.062e-02	5.233e-02
$\chi^2(2)$	5.865e-13	6.171e-03	6.337e-03
$\chi^2(4)$	3.986e-08	6.003e-03	9.604e-03
$\chi^2(6)$	2.557e-02	8.944e-03	1.435e-02
$\chi^2(8)$	6.115e-04	6.723e-03	1.475e-02
$\chi^2(10)$	1.199e-03	1.177e-02	1.312e-02
$t(1)$	9.893e-02	8.690e-01	4.781e-01
$t(2)$	6.921e-03	1.888e-01	7.467e-02
$t(4)$	2.096e-04	7.259e-02	4.647e-02
$W(1,1)$	8.034e-15	3.341e-03	5.602e-03
$W(1.5,1)$	5.480e-04	4.056e-03	5.934e-04
$W(2,2)$	2.488e-07	2.227e-04	8.488e-03
$W(6,3)$	3.286e-06	9.325e-03	8.122e-03
$W(8,4)$	1.547e-04	4.701e-02	1.136e-02
$BiMod(0,1)(6,2)$	2.656e-03	3.018e-02	2.380e-02
$BiMod(1,1)(11,3)$	2.330e-01	1.066e-01	6.782e-02
$BiMod(1,1)(7,1)$	2.348e-05	4.481e-02	8.573e-02

Fig. 3. Fit of a $\chi^2(10)$ using the three MoP alternatives.Fig. 4. Fit of a $W(6,3)$ using the three MoP approaches.Fig. 5. Fit of a bimodal distribution made from a $N(0,1)$ and a $N(6,2)$ using the three MoP alternatives.

In terms of efficiency, we consider it is appropriate to refer to both the execution time and the number of parameters to be estimated by the model. The example in Fig. 5 has been designed to carry out a small experiment comparing the three methods mentioned before (Shenoy, MoTBF, and tMoP). The results are shown in Table II.

TABLE II. EXECUTION TIME (IN SECONDS) AND NUMBER OF PARAMETERS OF EACH PRESENTED METHOD TO FIT THE BIMODAL DISTRIBUTION COMPOSED BY $N(0,1)$ AND $N(6,2)$ USED IN FIG. 5

Method	Execution time	No. of parameters
Shenoy	0.30	34
MoTBF	0.14	7
tMoP	0.07	9

Hereafter, we investigate the effect of the maximum degree and the sample size in the tMoP fit. Also, we compare the influence of the type of MoP used in the goodness-of-fit. To analyze these effects, we perform hypothesis tests over the KL divergence results to detect significant differences between each alternative. We use non-parametric tests: the Friedman test when comparing more than two groups, and the Wilcoxon signed-rank test when comparing only two groups. For all tests, the significance level was set at 0.05 and the null hypothesis stated that there was no difference between the groups.

1. Effect of the Maximum Degree on the Univariate tMoP Fit

For sample sizes of 50 and 100, the choice between the three maximum polynomial degrees (3, 4 and 7) did not lead to statistically different results, according to the Friedman test. With a sample size of 50, the p-value was 0.205 while for size 100, the p-value was 0.088.

On the other hand, for a sample size of 1000, the Friedman test rejects the null hypothesis with a p-value of 8.591e-09. In this case, paired Wilcoxon tests have been performed, whose results are displayed in Table III. Significant differences are detected in all cases, in favor of the highest degree.

Thus, the maximum degree in the fit of tMoPs is significant only when the sample size is large enough.

TABLE III. P-VALUES OF THE WILCOXON TEST FOR THE MAXIMUM DEGREE IN THE tMoP FIT WITH A SAMPLE SIZE OF 1000

Max. degree	p-value
3 vs 4	2.658e-06
3 vs 7	1.105e-06
4 vs 7	5.949e-06

2. Effect of the Sample Size on the Univariate tMoP Fit

According to the Friedman test applied to assess the effect of sample size (50, 100 and 1000) on the tMoP fit, the null hypothesis was rejected for the three polynomial degrees tested, with a p-value of 1.499e-12. To analyze differences between the groups, paired Wilcoxon tests have been performed for the effect of the sample size for each degree. The p-values obtained are displayed in Table IV.

Thus, the sample size has a significant effect on the fit of tMoPs independently of the maximum degree fixed.

TABLE IV. P-VALUES OF THE WILCOXON TEST FOR THE SAMPLE SIZE IN THE tMoP FIT FOR EACH MAXIMUM DEGREE

Degree	Sample size	p-value
3	50 vs 100	9.313e-10
	50 vs 1000	4.656e-10
	100 vs 1000	4.581e-06
4	50 vs 100	4.656e-10
	50 vs 1000	9.313e-10
	100 vs 1000	1.494e-06
7	50 vs 100	4.656e-10
	50 vs 1000	4.977e-07
	100 vs 1000	1.105e-06

3. Comparison of Univariate tMoPs to the Other Univariate MoP Approaches

Finally, we compared the tMoP approach to two other MoP alternatives: Shenoy's method and MoTBFs. For MoTBFs and tMoPs, all comparisons were conducted with a sample size of 1000. As before, the KL divergences of tMoPs and the alternatives were analyzed using the Wilcoxon sign rank test.

The results of the comparison between tMoPs and Shenoy's method is detailed in Table V. The results are mainly affected by the number of subintervals used to fit in Shenoy's method (nPieces). Regarding the case of 1 or 2 intervals, the results obtained with tMoP are better than those obtained with Shenoy (all p-values are less than 0.05). When the number of intervals is 3, the results are balanced, and with higher degrees the null hypothesis cannot be rejected. On the other hand, Shenoy's method outperforms tMoPs when 4 intervals are used.

In Table VI, the resulting p-values of the comparison between MoTBFs and tMoPs with different degrees are reported. When the degree is low, the Wilcoxon test indicates that the hypothesis of equality is rejected, and the adjustments made with tMoPs are more accurate. On the contrary, when the degree is 7 the null hypothesis cannot be rejected.

TABLE V. P-VALUES OF THE WILCOXON TEST COMPARING tMoPs AND SHENOY'S METHOD FOR VARIOUS COMBINATIONS OF PARAMETERS

nPieces	Degree	Best method	p-value
1	3	tMoPs	6.519e-09
	4	tMoPs	9.313e-10
	7	tMoPs	1.253e-05
2	3	tMoPs	2.720e-03
	4	tMoPs	2.720e-03
	7	tMoPs	1.871e-02
3	3	Shenoy	4.907e-04
	4	No sign. diff.	0.178
	7	No sign. diff.	0.173
4	3	Shenoy	8.696e-06
	4	Shenoy	3.493e-05
	7	Shenoy	4.868e-03

TABLE VI. P-VALUES OF THE WILCOXON TEST COMPARING tMoPs AND MoTBFs FOR VARIOUS DEGREES

Degree	Best method	p-values
3	tMoPs	1.412e-05
4	tMoPs	9.767e-05
7	No sign. diff.	0.382

B. Approximation to Univariate Real-World Data

Shenoy's method cannot be used to fit probability distributions to real data, since it is based in the Taylor series of the true density, which is unknown. Thus, here we only test MoTBFs and tMoPs. Also, when fitting real-world data, the sample size is not a free parameter. Thus, the comparisons are more limited than in the case of fitting known densities. Still, we can analyze the influence of the maximum degree and compare the two MoP approaches.

For this purpose, we used some datasets from the UCI repository [25]:

- *Appendicitis*: 7 continuous variables and 106 observations
- *Cloud*: 6 continuous variables and 108 observations
- *Concrete*: 8 continuous variables and 1030 observations
- *Housing*: 12 continuous variables and 506 observations
- *Hungarian*: 4 continuous variables and 261 observations

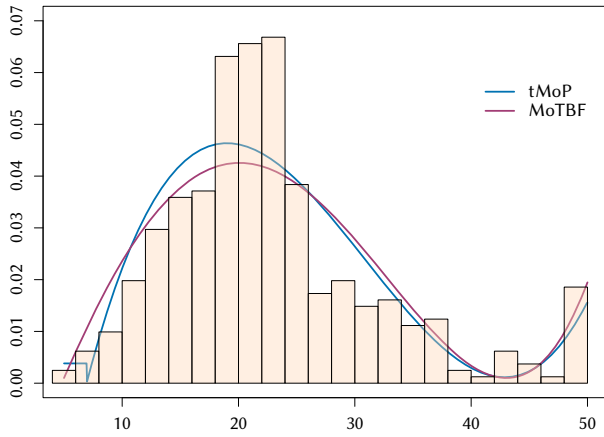
From each dataset, 3 different variables have been fitted with both MoP alternatives, using as maximum degrees 3, 4 and 7. In this case, the BIC was used to measure the goodness-of-fit (see Section III.E). In all cases, 80% of the data were randomly selected for training while the remaining 20% were used for testing. The BIC results in the test dataset for the best case of each method can be seen in the Table VII.

The two methods were compared using the pairwise Wilcoxon test. For degrees 3 and 4, the null hypothesis cannot be rejected (the p-values were 0.359 and 0.210, respectively). However, the hypothesis is rejected for degree 7, with a p-value of 3.108e-02. In that case, the tMoP fits were more accurate than those with MoTBFs.

As an example, the fit with tMoPs and MoTBFs with degree 4 for the variable *V14* of dataset *Housing* is depicted in Fig. 6.

TABLE VII. BIC OF UNIVARIATE DATA FROM REAL-WORLD DATASETS FITTED WITH tMoTBFs AND tMoPs

Dataset	PDF	tMoPs	MoTBFs
<i>Appendicitis</i>	$f(x_{v_2})$	2.536e-00	1-146e-00
	$f(x_{v_3})$	5.716e-01	2.678e-01
	$f(x_{v_4})$	1.637e-00	2.198e-00
<i>Cloud</i>	$f(x_{v_4})$	-1.459e+01	-1.521e+01
	$f(x_{v_6})$	-1.577e+01	-1.590e+01
	$f(x_{v_8})$	-1.825e+01	-1.829e+01
<i>Concrete</i>	$f(x_{v_1})$	-5.069e+02	-0.64e+02
	$f(x_{v_3})$	-3.891e+02	-3.884e+02
	$f(x_{v_6})$	-5.060e+02	-5.054e+02
<i>Housing</i>	$f(x_{v_5})$	3.506e+01	3.624e+01
	$f(x_{v_7})$	-1.892e+02	-1.913e+02
	$f(x_{v_{14}})$	-1.587e+02	-1.590e+02
<i>Hungarian</i>	$f(x_{v_1})$	-8.050e+01	-8.101e+01
	$f(x_{v_5})$	-1.155e+02	-1.158e+01
	$f(x_{v_8})$	-9.912e+01	-9.830e+01

Fig. 6. Fit of V_{14} of the dataset *Housing* using tMoPs and MoTBFs and histogram of the variable.

C. tMoP Meta-Model for Common Univariate Probability Distributions

To test the meta-model proposed in Section III.D, we have used three well-known probability distributions: one with a single parameter and two with two parameters:

- Exponential distribution: A tMoP model is fitted at different values of the exponential distribution parameter, which is considered within the interval $[1,6]$. A step size of 0.05 is used, resulting in a total of 101 fitted distributions. The variable domain used for them was $[0,5]$.
- Lognormal distribution: the parameter μ is estimated in the interval $[0,1]$ and the parameter σ , in the interval $[0.2,1.2]$. One tMoP is fitted every 0.1 in each parameter, so in total we have previously needed 121 distributions. The variable domain used for them was $[0,15]$.
- Beta distribution: both parameters are estimated in the interval $[0,2]$. One tMoP is fitted every 0.1 in each parameter, so in total we have previously needed 121 distributions for the domain $[0,1] \times [0,1]$, another 121 distributions for the domain $[0,1] \times [1,2]$ and 121 distributions for the domain $[1,2] \times [1,2]$. It is unnecessary to fit distributions in $[1,2] \times [0,1]$ due to the symmetry of the Beta distribution with respect its parameters. The variable domain was always $[0,1]$.

In all cases, we compare the direct fit by tMoPs and the tMoP interpolation using the pre-learned distributions. Table VIII shows the

results obtained for the exponential distribution. We experimented with 15 different parameters within the established interval. This table shows the KL divergence for the fit by interpolation and direct fit by tMoPs using a sample of size 1000. According to the paired Wilcoxon test, the hypothesis of equality between both methods has been rejected with a p-value of $3.051e-05$.

TABLE VIII. KL DIVERGENCE OF INTERPOLATION AND DIRECT tMoP FITTING OF THE EXPONENTIAL DISTRIBUTION

PDF	Direct fit	Interpolation
<i>Exp</i> (1.33)	5.716e-03	6.272e-02
<i>Exp</i> (1.41)	6.586e-03	6.256e-02
<i>Exp</i> (1.67)	8.377e-03	4.897e-02
<i>Exp</i> (2.12)	7.291e-03	3.237e-02
<i>Exp</i> (2.44)	6.819e-03	3.034e-02
<i>Exp</i> (2.93)	6.822e-03	3.921e-02
<i>Exp</i> (3.08)	6.927e-03	3.705e-02
<i>Exp</i> (3.51)	7.341e-03	4.831e-02
<i>Exp</i> (3.72)	7.485e-03	2.439e-02
<i>Exp</i> (4.11)	8.192e-03	1.950e-02
<i>Exp</i> (4.52)	8.556e-03	3.487e-02
<i>Exp</i> (4.96)	8.933e-03	3.867e-02
<i>Exp</i> (5.28)	9.193e-03	2.882e-02
<i>Exp</i> (5.47)	9.341e-03	2.575e-02
<i>Exp</i> (5.81)	9.595e-03	2.012e-02

The results obtained for the beta distribution are shown in Table IX and for the lognormal distribution are shown in Table X. Both tables contain the comparison between the approximation using the meta-model and the direct fit using a tMoP model from a sample of 1000 observations. In both cases we experimented using 15 different combinations of both parameters.

According to the Wilcoxon test, with a 0.05 significance level, the hypothesis of equality between both methods has been rejected in the lognormal distribution with a p-value of $3.051e-05$; however it has not been rejected in the beta distribution with a p-value of 0.085.

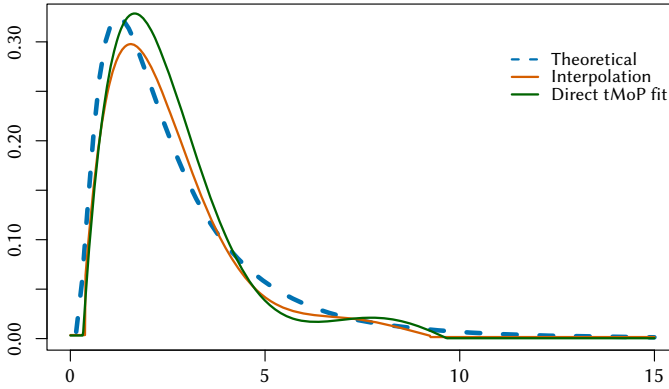
An example of this method to estimate the $LN(0.77, 0.57)$ can be seen in Fig. 7. It shows that the difference between the tMoPs calculated from the previously learned database and the tMoPs calculated directly from a sample of the theoretical distribution are very similar, whereas the computational cost has been reduced considerably.

TABLE IX. KL DIVERGENCE OF INTERPOLATION AND DIRECT tMoP FITTING OF THE LOGNORMAL DISTRIBUTION

PDF	Direct fit	Interpolation
<i>LN</i> (0.08,1.14)	7.470e-02	9.989e-02
<i>LN</i> (0.11,0.76)	1.412e-01	3.978e-01
<i>LN</i> (0.16,0.25)	7.457e-01	1.563e-00
<i>LN</i> (0.27,0.69)	6.492e-02	1.245e-01
<i>LN</i> (0.35,0.87)	8.683e-02	1.290e-01
<i>LN</i> (0.39,1.04)	6.551e-02	8.765e-02
<i>LN</i> (0.44,0.62)	4.548e-02	1.901e-01
<i>LN</i> (0.53,1.12)	5.875e-02	9.568e-02
<i>LN</i> (0.57,0.46)	2.108e-01	2.821e-01
<i>LN</i> (0.64,0.88)	7.011e-02	1.033e-01
<i>LN</i> (0.69,1.07)	5.725e-02	1.108e-01
<i>LN</i> (0.77,0.57)	6.304e-02	1.190e-01
<i>LN</i> (0.81,1.15)	5.721e-02	1.292e-01
<i>LN</i> (0.91,0.93)	5.339e-02	9.698e-02
<i>LN</i> (0.94,0.36)	3.378e-01	5.680e-01

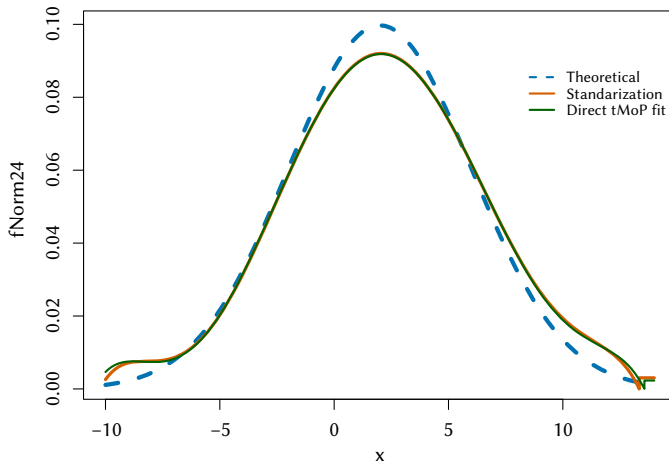
TABLE X. KL DIVERGENCE OF INTERPOLATION AND DIRECT tMoP FITTING OF THE BETA DISTRIBUTION

PDF	Direct fit	Interpolation
$\beta(0.33, 0.85)$	2.674e-01	2.642e-01
$\beta(0.45, 0.44)$	1.258e-01	1.198e-01
$\beta(0.56, 0.77)$	3.593e-02	3.478e-02
$\beta(0.62, 0.94)$	2.223e-02	1.942e-02
$\beta(0.82, 0.58)$	2.989e-02	2.971e-02
$\beta(0.27, 1.42)$	6.456e-01	2.528e-00
$\beta(0.35, 1.93)$	0.350e-01	1.770e-00
$\beta(0.52, 1.14)$	5.207e-02	5.162e-02
$\beta(0.85, 1.36)$	2.740e-03	2.782e-03
$\beta(0.92, 1.65)$	2.849e-03	2.066e-03
$\beta(1.18, 1.76)$	1.791e-03	2.922e-00
$\beta(1.33, 1.55)$	3.817e-03	2.055e-00
$\beta(1.52, 1.91)$	5.615e-03	1.141e-00
$\beta(1.61, 1.25)$	3.126e-03	3.709e-01
$\beta(1.84, 1.48)$	4.271e-03	3.291e-01

Fig. 7. Fit of a $LN(0.77, 0.57)$ PDF using a grid of equispaced tMoPs with a maximum degree of 7 in the domain of the parameters.

1. Standardization of a Univariate Normal Distribution

We will illustrate the method of approximation of the Gaussian distribution described in Section III.D. For this purpose, we started from the tMoP fitted to a standard normal distribution from a sample size of 1000 with a maximum degree of 7 in the interval $[-3, 3]$. Based on it, we fitted the distribution $X \sim N(2, 4)$ in the interval $[-10, 14]$, which can be seen in Fig. 8.

Fig. 8. Fit of a $N(2, 4)$ PDF rescaling the 7 degree tMoP of a $N(0, 1)$ and fitting a direct tMoP with a maximum degree of 7.

D. Approximation to Multivariate Probability Distributions

Now, we analyze the accuracy of learning tMoPs for continuous variables with one or two continuous parents, using the algorithms from Section III.C.

1. Conditional Distributions With One Parent

We tested the tMoPs approach for fitting a conditional density with one parent $f(y|x)$ with common well-known probability distributions, as well as with real data. For the first case, we use the bivariate normal and exponential distributions with different parameters (as before, unbounded domains were trimmed to finite subdomains of probability ≥ 0.95):

- $X \sim N(0, 1)$ and $Y \sim N(0, 1)$ where $\rho = \frac{1}{2}$. Therefore, $Y|X = x \sim N\left(\frac{x}{2}, \frac{\sqrt{3}}{2}\right)$ defined in the interval $[-3, 3] \times [-3, 3]$
- $X \sim N(0, 1)$ and $Y \sim N(1, 3)$ where $\rho = \frac{3}{4}$. Therefore, $Y|X = x \sim N\left(1 + \frac{9x}{4}, \frac{3\sqrt{7}}{4}\right)$ defined in the interval $[-3, 3] \times [-9, 11]$
- $X \sim N(1, 3)$ and $Y \sim N(0, 5)$ where $\rho = \frac{2}{3}$. Therefore, $Y|X = x \sim N\left(\frac{10(x-1)}{9}, \frac{5\sqrt{5}}{3}\right)$ defined in the interval $[-5, 7] \times [-12, 12]$
- $X \sim \text{Exp}\left(\frac{1}{2}\right)$ and $Y \sim \text{Exp}(1)$ where $\rho = \frac{1}{2}$. Therefore, the density function of Y , given $X = x$, is:

$$f(y|x) = \left(1 + \frac{x}{2}\right) \cdot \exp\left(-\left(1 + \frac{x}{2}\right) \cdot y\right) \cdot \left(1 + y - \frac{2}{2+x}\right), \quad (20)$$

defined in the interval $[0, 3] \times [0, 3]$

The marginal and conditional density functions have been normalized in the specified interval. In all cases, the tMoP fit has been performed using a sample of size 1000 randomly selected within the corresponding domain.

The density fit with tMoPs has been compared with discretization, with two alternatives (3 and 5 subintervals per variable), yielding a two-dimensional domain made up of 9 and 25 subdomains respectively. To illustrate this, the first bivariate normal distribution and its three approximations are displayed in Fig. 9.

To measure the accuracy, the Kullback-Leibler divergence has been computed for the three alternatives (see Section III.E). The results are reported in Table XI.

TABLE XI. KL DIVERGENCE OF CONDITIONAL DENSITIES WITH ONE PARENT WITH tMoPs AND DISCRETIZATION

PDF	tMoP	Discr-3	Discr-5
$N(0, 1), N(0, 1), \rho = 1/2$	6.8637e-02	2.0329e-01	9.8460e-02
$N(0, 1), N(1, 3), \rho = 3/4$	1.1754e-01	9.9874e-01	2.2599e-01
$N(1, 3), N(0, 5), \rho = 3/4$	5.0116e-02	1.9275e-01	8.6249e-02
$\text{Exp}(1/2), \text{Exp}(1), \rho = 1/2$	1.1691e-02	4.0153e-02	2.2261e-02

Similarly, experiments have been carried out with variables obtained from real data. These continuous variables are taken from the datasets:

- *Forestfires*: 7 continuous variables and 507 observations
- *Istanbul*: 9 continuous variables and 536 observations
- *Forestfires*: 5 continuous variables and 625 observations

An example of the approximation of real data can also be seen at Fig. 10.

For measuring the goodness-of-fit of real data, the BIC (see Section III.E) is calculated. The results of the experiments with real data can be seen in Table XII. In all cases, 80% of the data were randomly selected for training, while the remaining 20% were used as test. In all the cases, the tMoP produces a more accurate fit than the discretization.

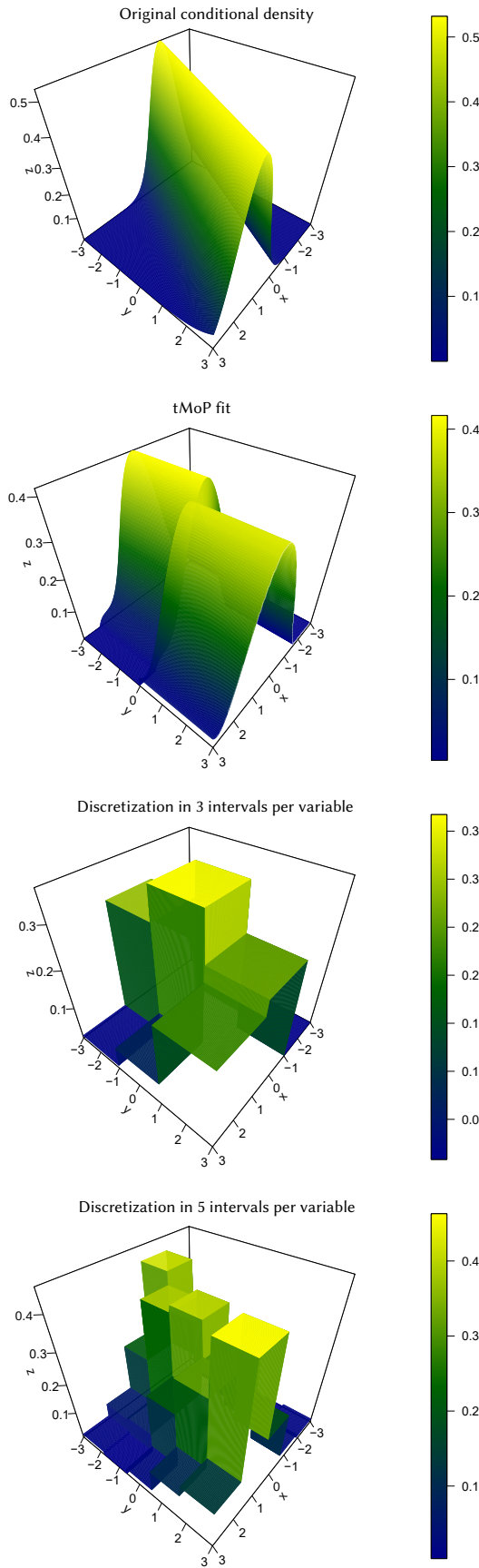


Fig. 9. Bivariate normal distribution where $X \sim N(0, 1)$, $Y \sim N(0, 1)$ and $\rho = \frac{1}{2}$ (top), and its approximations using 7-degree tMoPs (second row) and discretization with 3 and 5 intervals per variable (two bottom rows).

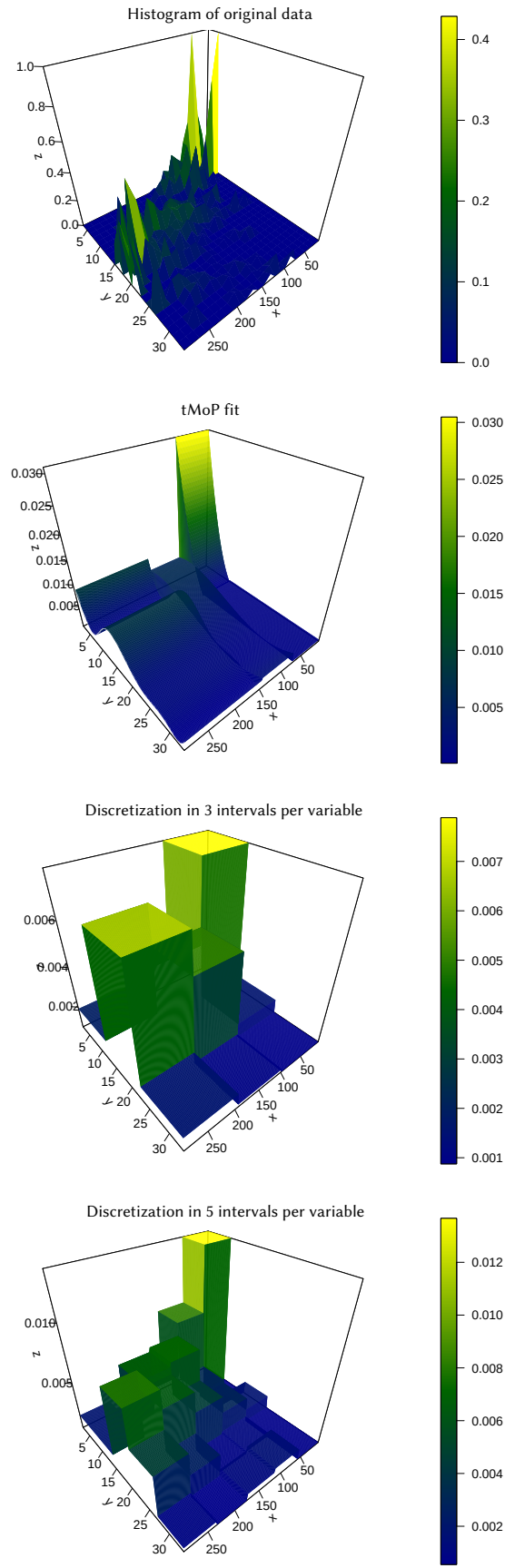


Fig. 10. Two-dimensional distribution of variables V_6 and V_9 of the dataset *Forestfires*. Approximate histogram (top), fit with 7-degree tMoPs (second row) and fit with discretization with 3 and 5 intervals per variable (two bottom rows).

TABLE XII. BIC OF CONDITIONAL DENSITIES WITH ONE PARENT FOR REAL DATA WITH tMoPs AND DISCRETIZATION

Dataset	PDF	tMoPs	Discr-3	Discr-5
Forestfires	$f(x_{v5} x_{v13})$	-124.6620	-158.0394	-146.8105
	$f(x_{v6} x_{v9})$	-48.8966	-247.1916	-253.7301
Istanbul	$f(x_{v1} x_{v8})$	148.5237	138.6659	138.2831
	$f(x_{v6} x_{v9})$	122.3097	119.1601	112.9496
Strikes	$f(x_{v4} x_{v7})$	-127.0834	-129.0789	-134.4338
	$f(x_{v5} x_{v6})$	-46.7847	-166.0334	-172.1099

2. Conditional Distributions With Two Parents

With the same considerations as in the previous section, we now fit a conditional distribution of a continuous variable depending on two parents, $f(z|x, y)$. In particular, we have tested the behavior of tMoP with several trivariate normal distributions of different parameters and also with some real-world datasets. The normal distributions employed were:

- ($distr_1$): $X \sim N(0, 1)$, $Y \sim N(0, 1)$ and $Z \sim N(0, 1)$ where X and Y are the parents of Z . The correlations are: $\rho_{xy} = 0$, $\rho_{xz} = \frac{1}{3}$ and $\rho_{yz} = \frac{2}{3}$. The vector of means and covariance matrix are:

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 1 \end{pmatrix}$$

Therefore the density function of $Z|X = x, Y = y$ is:

$$f(z|x, y) = \frac{1}{\frac{\pi}{2}\sqrt{2\pi}} \exp \left(-\frac{1}{8}x^2 - \frac{1}{2}y^2 - \frac{9}{8}z^2 - \frac{1}{2}xy + \frac{3}{4}xz + \frac{3}{2}yz \right) \quad (21)$$

defined in the interval $[-3, 3] \times [-3, 3] \times [-4, 4]$

- ($distr_2$): $X \sim N(0, 1)$, $Y \sim N(0, 3)$ and $Z \sim N(0, 2)$ where X and Y are the parents of Z . The correlations are: $\rho_{xy} = 0$, $\rho_{xz} = \frac{1}{2}$ and $\rho_{yz} = \frac{1}{2}$. The vector of means and covariance matrix are:

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 9 & 3 \\ 1 & 3 & 4 \end{pmatrix}$$

Therefore the density function of $Z|X = x, Y = y$ is:

$$f(z|x, y) = \frac{1}{2\sqrt{\pi}} \exp \left(-\frac{1}{4}x^2 - \frac{1}{36}y^2 - \frac{1}{4}z^2 - \frac{1}{6}xy + \frac{1}{2}xz + \frac{1}{6}yz \right) \quad (22)$$

defined in the interval $[-3, 3] \times [-6, 6] \times [-7, 7]$

- ($distr_3$): $X \sim N(0, 2)$, $Y \sim N(0, 1)$ and $Z \sim N(0, 2)$ where X is the parent of Y and X and Y are the parents of Z . The correlations are: $\rho_{xy} = \frac{1}{2}$, $\rho_{xz} = \frac{1}{4}$ and $\rho_{yz} = \frac{1}{2}$. The vector of means and covariance matrix are:

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 4 \end{pmatrix}$$

Therefore the density function of $Z|X = x, Y = y$ is:

$$f(z|x, y) = \frac{1}{\sqrt{6\pi}} \exp \left(-\frac{1}{6}y^2 - \frac{1}{6}z^2 + \frac{1}{3}yz \right) \quad (23)$$

defined in the interval $[-5, 5] \times [-3, 3] \times [-6, 6]$

Similarly to the previous experiments, the tMoP was compared

to the discretization of each variable into 3 and 5 intervals. Table XIII shows the KL divergence of each type of fit. Again, the results obtained by applying the tMoP are, in all the cases, better than the discretization.

TABLE XIII. KL DIVERGENCE OF CONDITIONAL DENSITIES WITH TWO PARENTS WITH tMoPs AND DISCRETIZATION

PDF	tMoPs	Discr-3	Discr-5
$distr_1$	1.7100e-01	5.0285e-01	3.7603e-01
$distr_2$	2.2098e-01	3.6682e-01	2.6279e-01
$distr_3$	6.5675e-02	2.2050e-01	1.3620e-01

TABLE XIV. BIC OF CONDITIONAL DENSITIES WITH TWO PARENTS FOR REAL DATA WITH tMoPs AND DISCRETIZATION

Dataset	PDF	tMoPs	Discr-3	Discr-5
Forestfires	$f(x_{v6} x_{v7}, x_{v8})$	-255.1025	-266.9436	-345.6007
	$f(x_{v9} x_{v10}, x_{v13})$	-204.1668	-184.0390	-266.6199
Istanbul	$f(x_{v1} x_{v2}, x_{v3})$	95.8300	119.4869	50.8221
	$f(x_{v4} x_{v5}, x_{v6})$	48.2965	101.8016	42.5188
Strikes	$f(x_{v4} x_{v6}, x_{v7})$	-167.1525	-149.3543	-208.713
	$f(x_{v6} x_{v5}, x_{v7})$	-295.4872	-245.9434	-305.8199

Table XIV lists the results for the BIC in six particular fits of continuous variables with two continuous parents from the datasets *Forestfires*, *Istanbul* and *Strikes*. In all of them, the tMoP fit outperforms the 5-interval discretization. Regarding the discretization with 3 intervals, tMoP gets more accurate results in one case. This is probably due to the number of parameters required, it is penalizing the tMoP and 5-interval discretization models.

V. DISCUSSION

In Section IV, the tMoP have been evaluated in various scenarios: fitting theoretical distributions and real data, fitting univariate and conditional distributions, and being used as a meta-model for common probability distributions.

In Section IV.A, tMoPs were compared to the MoPs variants using Shenoy's method and MoTBF in the univariate case for common probability distributions.

For Shenoy's method, as expected, the best results (minimum KL divergence) are obtained for the highest degree and greatest number of subintervals. However, the complexity of the resulting model also increases with these parameters.

An example of this complexity can be found in Table II where both the execution time and the number of parameters to be estimated in Shenoy's method are much greater than tMoPs. Comparing MoTBF with tMoP, it can be noticed that the number of parameters is similar. However, in the tMoP method the execution time has been reduced by half.

The fit with MoTBFs and tMoPs was assessed varying the maximum polynomial degree and the sample size. Again, the fit is more accurate as the sample size and maximum degree increase.

However, the results displayed in Table III show that when only small samples are available, the degree is not that relevant. In contrast, when the sample size is higher and the amount of information about the distribution is higher, the choice of maximum degree becomes relevant. In these cases, increasing the degree can substantially improve the accuracy of the fit.

Comparing between Shenoy's method to tMoPs, according to Table V, it is reasonable to assume that when the degree used in the fit is low, a larger number of pieces offers a great advantage and flexibility, but with a greater complexity.

Therefore, when the number of subintervals (and, therefore, of parameters) is comparable (1, 2 or 3), the tMoP is more accurate in 8 out of 9 options tested. This is especially relevant if we take into consideration that the tMoP does not need the analytic expression of the distribution, but only a sample.

Comparing tMoPs to MoTBFs for several degrees (see Table VI), the former outperforms the latter when the maximum degree is lower, and they are comparable in the highest tested degree.

Thus, the requirement of the theoretical distribution and its greater complexity are the main drawbacks of Shenoy's method. The comparison between MoTBFs and tMoPs gives similar or favorable results to the tMoPs adjustment. In general, tMoPs have the higher flexibility offered by the tails. This advantage is more evident with lower degrees, thus allowing to work with simpler polynomials and to reduce the number of parameters and the functional complexity.

In Section IV.B, tMoPs were used to fit real-world data. We used 3 continuous variables in 5 different datasets and compared the results obtained with tMoP and MoTBF. Shenoy's method is not applicable when the theoretical distribution is not available. The results are shown in Table VIII. When a small degree is used, such as 3 or 4, no significant differences between the two methods can be observed. On the other hand, with a higher degree (7), there are significant differences. In this case the results obtained with tMoP are better than those obtained with MoTBF. This shows that in real data the use of tails can be very helpful to avoid the mismatches that usually occur at the extremes of the intervals, while maintaining an optimal fit in the areas of highest probability density in the central part of the interval.

The use of distributions with tails presents some potential theoretical advantages. A tMoP, which consists of a polynomial distribution with uniform tails, combines the expressive power of the central polynomial part with the robustness of tails near the boundaries. The presence of tails can mitigate the influence of outliers and the problem of negative values in the adjusted density, which must be absolutely avoided. With the tMoP approach, the polynomial degree can be lowered with respect to a regular MoP while keeping the accuracy of the fit in high-density regions. Controlling the polynomial degree is generally a good numerical strategy, since polynomials of high degree are prone to large oscillations and to present areas with negative values due to sign changes. Therefore, tMoPs offer a good balance between flexibility, accuracy, and computational simplicity. These potential advantages have been reflected in the experimental results, justifying the better performance in general of the tMoP models.

In Section IV.C, a new method has been developed to fit well-known distributions more efficiently, which we have called tMoP meta-model. This method is based on pre-learning a certain set of tMoPs for some specific values of the parameters of a distribution, and then interpolating when other values of the parameter are required. The results compared to the direct tMoP fit can be seen in Table VIII, Table IX and Table X. The meta-models are satisfactory as they allow to improve the fit efficiency while maintaining a similar accuracy than the direct fit. As these three tables show, the results for a direct tMoP fit are slightly better than the results with the meta-model, but with a similar accuracy.

In the particular case of the Gaussian (normal) distribution, due to the way it is defined, it is possible to rescale the tMoP of the standard normal distribution. Therefore, in the normal distribution it is possible to obtain the fitting tMoP of any combination of parameters without loss of accuracy. An example can be seen in the Fig. 8.

Finally, in Section IV.D, tMoPs were employed to fit multivariate conditional probability distributions, for some common multivariate PDFs (normal and exponential distributions) and real-world data from 3 different datasets. In this case the tMoP model has been compared to

discretization process of all the variables (in 3 and 5 intervals).

In the case of a distribution conditioned to a continuous parent (see Table XI and Table XII), the tMoP fit provides considerably better results than both discretization approaches. For distributions conditioned to two continuous parents (see Table XIII and Table XIV), the results are similar. The tMoP model gets a more accurate fit for all the experimented distributions.

The tMoP method presented here allows to fit continuous distributions with any type of parents and, furthermore, it does not require to know the theoretical distribution. In addition, it allows to efficiently carry out any fit. This improvement over discretization makes the tMoP a versatile tool in parametric learning of probability distributions for hybrid BNs.

VI. CONCLUSIONS

In this work we have analyzed the mixtures of polynomials with tails (tMoPs) for fitting probability distributions in hybrid Bayesian networks. We reviewed the different MoPs alternatives in the literature, highlighting their main features. With respect to the work in [16], we have focused on the learning procedures and applications of tMoPs and their comparison to other techniques, increasing their knowledge in various ways.

We have provided learning procedures for both marginal (univariate) densities and conditional densities with tMoPs, that provide the basis of parametric learning for a Bayesian network with them.

Most of these ideas were originally introduced in [21] with a focus on classification and regression problems using a random sample, but some of them have been substantially improved or extended since then. On the other hand, we have tested this learning algorithm with several standard probability distributions (Exponential, Normal, Gamma, Beta, log-Normal, Chi-square, Student's t, Weibull and a bimodal (mixture of Gaussians)). The learning algorithms have been also analyzed for several real-world datasets. In addition, various experiments analyze the accuracy of tMoP fitting for multivariate distributions or data, which is a significant contribution with respect to [16]. In general, the experiments show the tMoPs models yield satisfactory results in comparison to other techniques, even with respect to other MoP available alternatives.

However, at the moment these learning procedures are constrained to up to two conditioning variables. We plan to go deeper into the learning procedures, especially to include conditional density functions in a general hybrid scheme with any number of conditioning variables.

Finally, we have extended in detail and broadened the experimental analysis of the meta-model introduced in [16]. This meta-model can provide an approximate tMoP expression for a specific density among a given family, from previously learned distributions for specific parameter values. This idea is tested with the well-known Exponential (one parameter), Lognormal (two parameters) and Beta (two parameters) distributions, giving promising results. The meta-models, thanks to its efficiency, are a very useful tool when it is necessary to make many adjustments for different values of the parameters, as can happen in the case of conditional variables. As a particular case, we also dealt with the case of fitting the normal distribution by tMoPs, which is simpler to analyze as it is based on a rescaling.

Globally, all these results show that tMoPs models are a valuable tool for probabilistic graphical models with hybrid variables. Also, both theoretical and practical knowledge about MoPs and tMoPs has been increased by providing new results and insights for them.

CREDiT AUTHORSHIP CONTRIBUTION STATEMENT

Juan C. Luengo: conceptualization, software, data curation, investigation, visualization, methodology, writing (original draft)

Rafael Rumí: conceptualization, formal analysis, supervision, funding acquisition, investigation, visualization, methodology, writing (review & editing).

Ana D. Maldonado: supervision, validation, investigation, visualization, methodology, writing (review & editing).

Darío Ramos-López: conceptualization, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing (review & editing).

DATA STATEMENT

The datasets used in the experiments can be found in the UCI repository [25].

The code necessary to reproduce the experiments can be found in the following GitHub repository: <https://github.com/jcluengolopez/IJMAI-bayesian-network-mops>.

DECLARATION OF CONFLICTS OF INTEREST

We have no conflict of interest to declare.

ACKNOWLEDGEMENTS

The last three authors (RR, AM and DRL) are supported by the research projects PID2022-139293NB-C31 and PID2025-172960NB-C33 funded by MICIU/AEI/10.13039/501100011033 and by "ERDF A way of making Europe" funds.

The last author (DRL) is also supported by the research projects PID2021-124472NB-I00 and PID2025-170285NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

The authors thank the support from the Center for Development and Transfer of Mathematical Research to Industry CDTIME (University of Almería), and from the Junta de Andalucía research groups FQM-229 and FQM-244, grant PPIT-UAL, Junta de Andalucía-ERDF 2021-2027, Objective RSO1.1. Programme: 54.A

REFERENCES

- [1] J. Pearl, *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA.: Morgan-Kaufmann. San Mateo, 1988.
- [2] K.B. Korb, E. A. Nicholson, *Bayesian artificial intelligence*. CRC Press/Chapman & Hall Book, 2010.
- [3] G. M. Castro-Luna, A. Martínez-Finkelshtein, D. Ramos-López, "Robust keratoconus detection with Bayesian network classifier for Placido-based corneal indices," *Contact Lens and Anterior Eye*, vol. 43, no. 4, pp. 366–372, 2020, <http://doi.org/10.1016/j.clae.2019.12.006>
- [4] R. Cabañas, A. M. Martínez, A. R. Masegosa, D. Ramos-López, A. Samerón, T. D. Nielsen, H. Langseth, A. L. Madsen, "Financial Data Analysis with PGMs Using AMIDST," in *IEEE 16th International Conference on Data Mining (ICDM 2016) Workshops*, 2016, pp. 1284–1287.
- [5] B. Cai, L. Huang, M. Xie, "Bayesian Networks in Fault Diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2227–2240, 2017, <http://doi.org/10.1109/TII.2017.2695583>
- [6] A. D. Maldonado, A. Valdivielso, A. Rescia, P. A. Aguilera, "Probabilistic graphical models for species richness prediction: Are current protected areas effective to face climate emergency," *Global Ecology and Conservation*, vol. 23, p. e01162, 2020.
- [7] T. Talvitie, R. Eggeling, M. Koivisto, "Learning Bayesian networks with local structure, mixed variables and exact algorithms," *International Journal of Approximate Reasoning*, vol. 115, pp. 69–95, 2019.

- [8] M. Grzegorzczak, "Being Bayesian about learning Bayesian networks from hybrid data," *International Journal of Approximate Reasoning*, vol. 187, p. 109549, 2025.
- [9] D. Atienza, P. Larrañaga, C. Bielza, "Hybrid semiparametric Bayesian networks," *TEST*, vol. 31, pp. 299–327, 2022.
- [10] C. Wei, C. Li, Y. Liu, S. Chen, Z. Zuo, P. Wang, Z. Ye, "Causal Discovery and Reasoning for Continuous Variables with an Improved Bayesian Network Constructed by Locality Sensitive Hashing and Kernel Density Estimation," *Entropy*, vol. 27, no. 2, p. 123, 2025.
- [11] S. Moral, R. Rumí, A. Salmerón, "Mixtures of truncated exponentials in hybrid Bayesian networks," in *ECSQARU'2001. Lecture Notes in Artificial Intelligence*, 2001, pp. 156–167.
- [12] H. Langseth, T. D. Nielsen, R. Rumí, A. Salmerón, "Mixtures of Truncated Basis Functions," *International Journal of Approximate Reasoning*, vol. 53, pp. 212–227, 2012.
- [13] P. P. Shenoy, J. C. West, "Mixtures of polynomials in hybrid Bayesian networks with deterministic variables," in *Proceedings of the 8th Workshop on Uncertainty Processing (WUPES'09)*, 2009, pp. 202–212.
- [14] H. Langseth, T. D. Nielsen, I. Pérez-Bernabé, A. Salmerón, "Learning mixtures of truncated basis functions from data," *International Journal of Approximate Reasoning*, vol. 55, pp. 940–956, 2014.
- [15] I. Pérez-Bernabé, A. D. Maldonado, A. Salmerón, T. D. Nielsen, "MoTBFS: An R package for learning hybrid bayesian networks using mixtures of truncated basis functions," *The R Journal*, vol. 12, no. 2, pp. 321–341, 2021, <http://doi.org/10.32614/RJ-2021-019>
- [16] J. C. Luengo, D. Ramos-López, R. Rumí, A. D. Maldonado, "Contributions on mixtures of polynomials for hybrid Bayesian networks," in *Intelligent Data Engineering and Automated Learning – IDEAL 2024*, 2025, pp. 460–471, Springer Nature Switzerland.
- [17] P. P. Shenoy, "Two issues in using mixtures of polynomials for inference in hybrid Bayesian networks," *International Journal of Approximate Reasoning*, vol. 53, pp. 847–866, 2012.
- [18] P. L. López-Cruz, T. D. Nielsen, C. Bielza, P. Larrañaga, "Learning Mixtures of Polynomials of Conditional Densities from Data," in *CAEPIA'13. Lecture Notes in Artificial Intelligence*, 2013, pp. 363–372.
- [19] P. P. Shenoy, J. C. West, "Inference in hybrid Bayesian networks using mixtures of polynomials," *International Journal of Approximate Reasoning*, vol. 52, pp. 641–657, 2011.
- [20] I. Pérez-Bernabé, A. Fernández, R. Rumí, A. Salmerón, "Parameter learning in hybrid Bayesian networks using prior knowledge," *Data Mining and Knowledge Discovery*, vol. 30, pp. 576–604, 2016.
- [21] J. C. Luengo, R. Rumí, "Naive Bayes classifier with mixture of polynomials," in *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM-2015)*, 2015, pp. 14–24.
- [22] J. C. Luengo, R. Rumí, "Mixtures of polynomials for regression problems," in *Proceedings of the 12th International Conference on Data Mining (ICDM'2016)*, 2016, pp. 208–214.
- [23] M. C. Wand, M. C. Jones, *Kernel Smoothing*. Boca Raton, FL.: Chapman & Hall, 1995.
- [24] M. Stone, "Comments on Model Selection Criteria of Akaike and Schwartz," *Journal of the Royal Statistical Society B*, no. 41, pp. 276–278, 1979.
- [25] M. Kelly, R. Longjohn, K. Nottingham, "UCI Machine Learning Repository." [Online]. Available: <http://archive.ics.uci.edu/ml>

Juan C. Luengo



Juan C. Luengo is a Postdoctoral Student in Mathematics at the University of Almería in Spain. He holds a Bachelor's degree in Mathematics and a Master's degree in Mathematic Engineering. His research interests include numerical approximation techniques, probabilistic modeling, data analysis and computational mathematical methods. He is a Secondary School Teacher specialized in Mathematics with 15 years of experience. He has developed his academic activity in different countries such as Spain, Hungary and Slovakia.



Rafael Rumí

Rafael Rumí is a Professor of Statistics at the University of Almería, Spain. Member of the Data Analysis Research Group and of the Centre for the Development and Transfer of Mathematical Research to Industry (CDTIME). He holds a Bachelor degree in Mathematics and a PhD in Mathematics. His research interests are, probabilistic modeling, machine learning, data analysis, supervised and unsupervised classification and ecological modelling. He has participated in more than 20 research projects, and supervised 4 National Research Projects, and is currently involved in the project TOPAI (Towards Probabilistic Trustworthy Artificial Intelligence). He has published more than 40 papers in journals and international conferences.



Ana D. Maldonado

Ana D. Maldonado is an Associate Professor of Statistics at the University of Almería, Spain. She is a member of the Data Analysis Research Group and the Centre for the Development and Transfer of Mathematical Research to Industry (CDTIME). Her research interests are probabilistic modeling, machine learning, and applications to the environment and ecology. Her multidisciplinary profile stems from her academic background in environmental sciences and her PhD training in data analysis, machine learning, and Bayesian networks. She has published more than 20 papers in recognized journals and conferences and has an h-index of 11. She is currently involved in the research project TOPAI (Towards Probabilistic Trustworthy Artificial Intelligence).



Darío Ramos-López

Darío Ramos-López is an Associate Professor of Applied Mathematics at the University of Almería, Spain. He is a member of the Approximation Theory and Orthogonal Polynomials Research Group and of the Centre for the Development and Transfer of Mathematical Research to Industry (CDTIME). He holds two Bachelor's degrees, in Mathematics and in Technical Engineering in Computer Systems, two Master's degrees, and a PhD in Mathematics. His research interests span computational mathematical methods, numerical approximation techniques, probabilistic modeling, and efficient data analysis. He published more than 30 papers in recognized journals and conferences, with an h-index of 15. He is currently involved in the research projects NETCATA (New Trends in Constructive Approximation) and TOPAI (Towards Probabilistic Trustworthy Artificial Intelligence).