Parameter Sensitivity Analysis in Mixtures of Polynomials

Ángel T. Sáez-Ruiz ANGELSAEZ@UAL.ES
Department of Mathematics
University of Almeria
Almeria, Spain

Rafael Rumí RRUMI@UAL.ES
Department of Mathematics and Center for the Development and Transfer of Mathematical Research to Industry (CDTIME)
University of Almeria
Almeria, Spain

Ana D. Maldonado ANA.D.MALDONADO@UAL.ES
Department of Mathematics and Center for the Development and Transfer of Mathematical Research to Industry (CDTIME)
University of Almeria
Almeria, Spain

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Parameter sensitivity analysis examines how a model is affected when some of its parameters are changed. In this study, we consider Mixtures of Polynomial (MoP) models to represent the probability distributions of continuous variables in Bayesian networks (BN). The objective of this study is to analyse the difference between a MoP density function and its perturbed versions, and how these perturbations propagate to other variables in the BN. A sensitivity measure, which is based on L^1 distance, is defined to compare the original distribution and the obtained distribution after perturbation. The theoretical results indicate that the sensitivity measure of a marginal distribution is at most the same as the sensitivity measure of the perturbed density. Overall, perturbations in local distributions of MoP Bayesian networks propagate to marginal distributions with decreasing intensity —meaning that their effect on marginal distributions of descendant variables becomes progressively weaker—which suggests that sensitivity analysis should primarily focus on local components.

Keywords: Mixtures of Polynomials, Hybrid Bayesian Networks, Parameter Sensitivity Analysis.

1. Introduction

Probabilistic graphical models (PGMs) (Lauritzen, 1996; Bishop and Nasrabadi, 2006; Murphy, 2012; Ghahramani, 2015) represent the dependencies between the variables of a random vector using a graph. This representation allows to simplify the information, and the Chain Rule of Probability can be used to compute all probabilities in the model. Bayesian Networks (Pearl, 1988) (BNs) are a special case of PGM where the graph represents the conditional probability distributions among the variables. BNs have been applied in a wide range of fields (Aguilera et al., 2011; Bielza and Larrañaga, 2014; Phan et al., 2019; Machado et al., 2023; Zheng et al., 2025), which shows the versatility of these models.

© 2026 Á.T. Sáez-Ruiz, R. Rumí & A.D. Maldonado.

Originally, BNs were introduced for discrete random vectors. As a result, when continuous variables are involved, a common practice is to discretise them and treat them as discrete variables. Gaussian BNs (Shachter and Kenley, 1989) provide an opportunity to consider continuous random variables, but they assume that the joint distribution of continuous variables is Gaussian. Nevertheless, there are many situations where the Gaussian assumption is unrealistic. This limitation has motivated the development of more flexible representations for continuous Bayesian networks, such as kernel density estimation, Mixtures of Truncated Exponentials (Moral et al., 2001), Mixtures of Polynomials (Shenoy and West, 2011) (MoPs), and, more generally, Mixtures of Truncated Basis Functions (Langseth et al., 2012) (MoTBFs); and semiparametric BNs (Atienza et al., 2022).

Sensitivity analysis in discrete BNs has been widely studied in the literature (Jensen and Nielsen, 2007; Rohmer, 2020; Ballester-Ripoll and Leonelli, 2025). In discrete BNs, the parameters are directly associated with the evidence and values of the variables. In this way, data incompleteness, partial knowledge of the domains, and expert knowledge are factors that could affect the accuracy of the BN parameters (Druzdzel and Van Der Gaag, 1995). The objective of sensitivity analysis is to study how variations in the parameters affect the output probability of interest. The simplest case is when only a few parameters in the BN are modified while the rest are fixed. When sensitivity analysis is carried out by varying only a single parameter, this procedure is called *one-way sensitivity analysis*. In general, if n parameters are changed at the same time, we refer to it as a *n-way sensitivity analysis*.

For continuous random variables, the parameters are not directly linked to a single state of the variable, and a change in them affects the probability of all outcomes. In the case of conditional linear Gaussian (CLG) models, the parameters represent the mean and the variance of the variable. One-way sensitivity analysis in CLG networks has been studied in Castillo and Kjærulff (2003); Gómez-Villegas et al. (2007); Gómez-Villegas et al. (2008); Zhu et al. (2017). In these studies, the objective was to analyse the differences in a target distribution using different perturbations of a single parameter of the BN. To the best of our knowledge, no analogous parameter sensitivity analyses have been reported for semiparametric BNs or MoP BNs in the literature. In MoP BNs, the conditional probability distributions (CPDs) are expressed as polynomial densities. The changes in the polynomial terms imply that the resulting expression may not be a density function , which constitutes an additional challenge when performing parameter sensitivity analysis in MoP BNs.

The objective of this study is to analyse, using a one-way sensitivity analysis, the differences between a MoP density function and its perturbed versions, and how these perturbations propagate to other variables in the BN. To do this, the paper is organised as follows. Section 2 introduces Bayesian networks based on MoP representations. Then, Section 3 develops a one-way sensitivity analysis for the case of univariate MoP densities. The extension of this analysis to BNs is presented in Section 4. An experimental analysis is carried out in section 5. Finally, the conclusions of this study are presented in Section 6. All proofs of the theoretical results are included in Appendix A. The code for generating the examples and the data results is freely available via https://github.com/asaez97/parameter_sensitivity_MoP_PGM.

2. Bayesian Networks using Mixtures of Polynomials

In this paper, the following notation is used. Uppercase letters denote random variables, while lowercase letters represent specific values of a random variable. Boldfaced characters indicate random vectors (i.e., multidimensional random variables). The set of all possible values of a random vector $\boldsymbol{X}$ (i.e., its *support*) is denoted by $\Omega_{\boldsymbol{X}}$. In particular, we use X to denote a generic random variable; the letters Y and Z represent discrete (multinomial) and continuous random variables, respectively. Finally, the support of a continuous random variable Z is given by $\Omega_Z = [\alpha, \beta]$.

A simple example of a BN representing the joint distribution of seven variables, Y_1, Y_2, Z_3, Z_4, Z_5, Y_6, and Z_7, is illustrated in Figure 1. This BN encodes the following factorisation:

$$p(Y_1, Y_2, Z_3, Z_4, Z_5, Y_6, Z_7) = p(Y_1)p(Y_2|Y_1)p(Z_3|Y_1)p(Z_4)p(Z_5|Y_2, Z_3)p(Y_6|Z_3)p(Z_7|Z_3, Z_4).$$

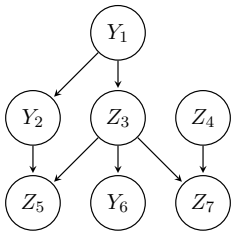


Figure 1: Example of a hybrid Bayesian network with seven variables. Y denotes discrete random variables and Z denotes continuous random variables.

In this paper, we consider hybrid BNs where continuous random variables follow a Mixture of Polynomials (MoP) (Langseth et al., 2012).

Definition 1 *Let $\boldsymbol{X}$ be a mixed n-dimensional random vector. Let $\boldsymbol{Y} = \{Y_1, \ldots, Y_d\}$ and $\boldsymbol{Z} = \{Z_1, \ldots, Z_c\}$ be the discrete and continuous parts of $\boldsymbol{X}$, respectively, with $d + c = n$. We say that a function $\hat{f} : \Omega_{\boldsymbol{X}} \rightarrow \mathbb{R}_0^+$ is a Mixture of Polynomials potential to degree r if one of the following two conditions holds:*

(i) $\hat{f}$ can be written as

$$\hat{f}(\boldsymbol{x}) = \hat{f}(\boldsymbol{y}, \boldsymbol{z}) = \sum_{i=0}^{r} \prod_{j=1}^{c} a_{i,\boldsymbol{y}}^{(j)} z_j^i, \tag{1}$$

where $a_{i,\boldsymbol{y}}^{(j)}$ are real numbers.

(ii) There is a partition $\mathcal{I}_1, \ldots, \mathcal{I}_m$ of $\Omega_{\boldsymbol{X}}$ for which the domain of the continuous variables, $\Omega_{\boldsymbol{Z}}$, is divided into hyper-cubes and such that $\hat{f}$ is defined as

$$\hat{f}(\boldsymbol{x}) = \hat{f}_l(\boldsymbol{x}) \quad \text{if} \quad \boldsymbol{x} \in \mathcal{I}_l,$$

where each $\hat{f}_l$, $l = 1, \ldots, m$, can be written in the form of (1).

An MoP potential is said to be a density if

$$\sum_{\boldsymbol{y}\in\Omega_{\boldsymbol{Y}}}\int_{\Omega_{\boldsymbol{Z}}}\hat{f}(\boldsymbol{y},\boldsymbol{z})d\boldsymbol{z}=1.$$

In this study, we consider that the support of a continuous random variable Z is not partitioned in its CPD. Definitions 2, 3 and 4 expose the different cases of CPDs.

Definition 2 *Let Z be a continuous random variable with support $\Omega_Z=[\alpha,\beta]$. We say that Z follows a univariate MoP distribution of degree r if its density function $p(z)$ can be written as:*

$$p(z)=\begin{cases}a_0+a_1z+\cdots+a_rz^r & \text{if } \alpha\leq z\leq\beta,\\ 0 & \text{in another case.}\end{cases}$$

where a_i are real numbers.

Definition 3 *Let Z be a continuous random variable in a MoP BN, with parent set $\boldsymbol{Pa}(Z)$ and let $\Omega^1_{\boldsymbol{Pa}(Z)},\ldots,\Omega^m_{\boldsymbol{Pa}(Z)}$ be a partition in hyper-cubes of $\Omega_{\boldsymbol{Pa}(Z)}$. We say that Z follows a conditional MoP distribution if the conditional density of $Z|\boldsymbol{Pa}(Z)$ can be written as*

$$p(z|\boldsymbol{Pa}(Z))=\begin{cases}p_l(z) & \text{if } z\in\Omega_Z,\boldsymbol{Pa}(Z)\in\Omega^l_{\boldsymbol{Pa}(Z)},\\ 0 & \text{in another case.}\end{cases}$$

where $p_l(z)$, $l=1,\cdots,m$ is a MoP density following Definition 2. We refer to $p_l(z)$ as a component of the conditional MoP distribution.

Definition 4 *Let Y be a discrete random variable in a MoP BN, with parent set $\boldsymbol{Pa}(Y)$ and let $\Omega^1_{\boldsymbol{Pa}(Y)},\ldots,\Omega^m_{\boldsymbol{Pa}(Y)}$ be a partition in hyper-cubes of $\Omega_{\boldsymbol{Pa}(Y)}$. The CPD of Y is defined as*

$$p(y|\boldsymbol{Pa}(Y))=\begin{cases}a_y^{(l)} & \text{if } y\in\Omega_Y,\boldsymbol{Pa}(Y)\in\Omega^l_{\boldsymbol{Pa}(Y)},\\ 0 & \text{in another case.}\end{cases}$$

where $0\leq a_y^{(l)}\leq 1$ and $\sum_{y\in\Omega_Y}a_y^{(l)}=1$ for each $l=1,\ldots,m$.

Example 1 shows some examples of the CPDs of the BN illustrated in Figure 1.

Example 1 *Example of the CPDs associated with nodes Y_1, Y_2 and Z_3 of the MoP BN considered in Figure 1.*

$$p(y_1)=\begin{cases}0.65 & \text{if } y_1=0,\\ 0.35 & \text{if } y_1=1,\\ 0 & \text{in another case.}\end{cases}\qquad p(y_2|y_1)=\begin{cases}0.45 & \text{if } y_1=0,y_2=0,\\ 0.55 & \text{if } y_1=0,y_2=1,\\ 0.3 & \text{if } y_1=1,y_2=0,\\ 0.7 & \text{if } y_1=1,y_2=1,\\ 0 & \text{in another case.}\end{cases}$$

$$p(z_3|y_1)=\begin{cases}0.5936-0.7441z_3-0.4069z_3^2+0.9779z_3^3+0.2102z_3^4 & \text{if } y_1=0,\ -1<z_3<1,\\ 0.5224-0.0511z_3+0.2487z_3^2+0.0740z_3^3-0.5266z_3^4 & \text{if } y_1=1,\ -1<z_3<1,\\ 0 & \text{in another case.}\end{cases}$$

3. Parameter Sensitivity Analysis in MoP Densities

Parameter sensitivity analysis consists of evaluating how a target distribution changes when some parameters of the BN are perturbed. In a MoP BN, the interpretation of the parameters differs depending on whether the variable is discrete or continuous. When the variable is discrete, the parameters have the same interpretation as in a discrete BN, i.e., the parameters of a node represent the probability of each specific value, given a configuration of the parent variables. In contrast, when a continuous random variable is considered, the parameters correspond to the polynomial coefficients.

In this study, we considered a *one-way sensitivity analysis*, i.e., a sensitivity analysis in which a single parameter of the CPD is perturbed at a time. In this section, we define how the perturbation of a single parameter is carried out in the case of a MoP density. These perturbations aim to evaluate the impact of numerical errors during parameter learning introduced by Langseth et al. (2014). Furthermore, they help identify which MoP parameters have more influence on inference. Example 2 shows that, regardless of the perturbation applied, the resulting function is no longer a valid density, since it may become negative and fail to integrate to 1.

Example 2 *Let $p(z)$ be the following MoP density:*

$$p(z) = \begin{cases} \frac{4}{9} - \frac{2}{9}z - \frac{2}{9}z^2 & \text{if } -2 \leq z \leq 1, \\ 0 & \text{in another case.} \end{cases}$$

We perturb the coefficient a_2, obtaining $\tilde{a}_2 = a_2 - \frac{2}{9} = \frac{-4}{9}$, and define the new function q as

$$q(z) = \begin{cases} \frac{4}{9} - \frac{2}{9}z - \frac{4}{9}z^2 & \text{if } -2 \leq z \leq 1, \\ 0 & \text{in another case.} \end{cases}$$

Function q is not a valid density since it does not integrate to 1 and may become negative. For instance, $q(1) = -\frac{2}{9} < 0$. We now illustrate that modifying other coefficients cannot in general restore validity. For example, if we consider modifying a_1, we obtain

$$q(z; a_1) = \frac{4}{9} + a_1 z - \frac{4}{9}z^2.$$

By imposing non-negativity at the boundaries of the support, we obtain bounds for a_1 as

$$q(-2; a_1) = \frac{4}{9} + (-2)a_1 - \frac{4}{9}(-2)^2 = -\frac{12}{9} - 2a_1 \geq 0 \Rightarrow a_1 \leq -\frac{2}{3},$$
$$q(1; a_1) = \frac{4}{9} + a_1 - \frac{4}{9} = a_1 \geq 0.$$

These conditions are incompatible, and therefore there is no value a_1 that can make $q(z; a_1)$ a valid density function.

Definition 5 defines how to construct a valid perturbed density from a MoP density p when a parameter is modified by a perturbation $h \in \mathbb{R}$.

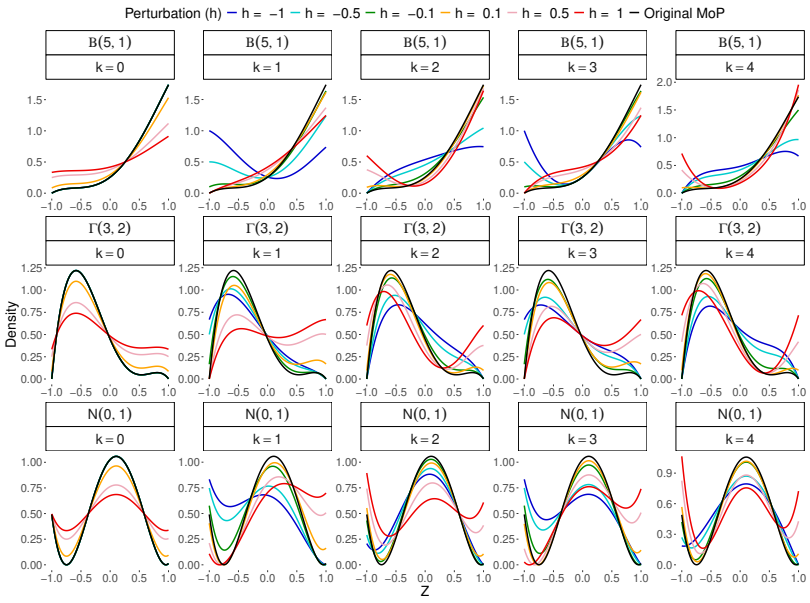


Figure 2: Example of perturbed MoP densities with $h = \pm 1, \pm 0.5, \pm 0.1$ and $k = 0, 1, 2, 3, 4$. All MoP densities are of degree 4, and the original (unperturbed) MoP densities are fitted from samples of 1000 observations drawn from $\mathcal{B}(5, 1)$, $\Gamma(3, 2)$ and $\mathcal{N}(0, 1)$ distributions. All samples are scaled to the [-1,1] interval.

Definition 5 *Let Z be a continuous random variable with a MoP density p as in Definition 2. Let $h \in \mathbb{R}$ be a perturbation factor and $k \in \{0, 1, \ldots, r\}$ the degree of the perturbed polynomial term. We define the perturbed density associated with the perturbation (k, h) as*

$$\tilde{p}(z; k, h) = \frac{p(z) + hz^k - a(p, k, h)}{A(p, k, h)},$$

where $a(p, k, h) = \min_{z \in \Omega_Z} \left(p(z) + hz^k, 0 \right)$ *and* $A(p, k, h) = 1 + \frac{h\left(\beta^{k+1} - \alpha^{k+1}\right)}{k+1} - a(p, k, h)\,(\beta - \alpha)$.

Note that the perturbation introduced in Definition 5 has an influence on all the coefficients of the resulting MoP density. Therefore, while the focus is on modifying a single component, the correction and normalization terms modify the entire density, which means that the resulting perturbation is global rather than a strict local change to a single coefficient. Proposition 6 shows that $\tilde{p}$ is a valid MoP density. The proof is given in the Appendix.

Proposition 6 *Let Z be a continuous random variable with a MoP density p of degree r as in Definition 2, and let $h \in \mathbb{R}$ and $k \in \{0, 1, \ldots, r\}$. Then, the transformation $\tilde{p}(z; k, h)$ introduced in Definition 5 is a MoP density.*

Figure 2 illustrates the effect of several values of h on three different MoP densities when perturbing polynomial terms of degree $k = 0, 1, 2, 3, 4$. These MoP densities have been fitted using samples of 1000 observations drawn from $\mathcal{B}(5, 1)$, $\Gamma(3, 2)$, and $\mathcal{N}(0, 1)$ distributions, rescaled to the interval [-1,1]. When the perturbation is applied to the constant term ($k = 0$), negative values of h may have no visible effect when the minimum of the original MoP density is zero ($\min_{z\in\Omega_Z} p(z) = 0$). This occurs because the correction term $a(p, 0, h)$ cancels the shift and the normalization factor equals one. As $|h|$ increases, the perturbed MoP density deviates increasingly from the original one. Therefore, the influence of the original MoP decreases as $|h|$ increases for each degree k.

To evaluate the effect of these perturbations, a sensitivity measure has to be considered. In this study, the *total variation distance* (Tsybakov, 2009, p. 83-84) is used to quantify the difference between the perturbed and the original MoP densities.

Definition 7 *Let Z be a continuous random variable with MoP density p of degree r as in Definition 2, and let $\tilde{p}(z)$ be a perturbed version of $p(z)$. Then, the sensitivity measure between $p(z)$ and $\tilde{p}(z)$ is defined as*

$$S\big(p(z), \tilde{p}(z)\big) = \frac{1}{2}\int_{\Omega_Z} \big|p(z) - \tilde{p}(z)\big| dz.$$

The sensitivity measure $S\big(p(z), \tilde{p}(z)\big)$ can be interpreted as the maximum difference in the probability of any event A between the perturbed $\tilde{p}(z)$ and the original $p(z)$ distributions. An important property of the sensitivity measure is that (Tsybakov, 2009). Example 3 shows the sensitivity measure for a MoP density when a perturbation of 0.05 is applied to the polynomial term of degree 2.

Example 3 *Let p be the MoP density with expression*

$$p(z) = \begin{cases} 0.2577 + 0.7121z + 0.8998z^2 + 0.1562z^3 - 0.2882z^4 & \text{if } -1 \leq z \leq 1, \\ 0 & \text{in another case.} \end{cases}$$

and consider a perturbation with $k = 2$ and $h = 0.05$.
Therefore, the perturbed MoP density $\tilde{p}(z; 2, 0.05)$ is given by

$$\tilde{p}(z; 2, 0.05) = \begin{cases} 0.2881 + 0.6689z + 0.7982z^2 + 0.1467z^3 - 0.2707z^4 & \text{if } -1 \leq z \leq 1, \\ 0 & \text{in another case.} \end{cases}$$

The sensitivity measure associated with the perturbation is $S\big(p(z), \tilde{p}(z; 2, 0.05)\big) = 0.0305$.

4. Parameter Sensitivity Analysis in MoP BNs

In the previous section, we analysed the sensitivity in a univariate MoP density. In this section, we extend the analysis to full MoP BNs. We focus on the case where a single parameter within a CPD of a MoP BN is perturbed, and its effect on the resulting marginal distributions.

On the one hand, in the case of continuous variables, perturbations are applied to a single parameter of one MoP component of its CPD. Theorem 8 shows how such a perturbation propagates to the marginal distributions in a MoP BN.

Theorem 8 *Let $\boldsymbol{X}$ be the random vector associated with a MoP BN, and let $Z \in \boldsymbol{X}$ be a continuous random variable whose CPD follows Definition 3. Let $l^* \in \{1, \ldots, m\}$ be the index of a component of the CPD of Z. Suppose that the l^*-th component of the CPD of Z is perturbed, and let $\tilde{p}_{l^*}(z)$ denote the perturbed MoP density of $Z \mid \boldsymbol{Pa}(Z) \in \Omega^{l^*}_{\boldsymbol{Pa}(Z)}$ (i. e., the conditional probability distribution of Z restricted to the hypercube $\Omega^{l^*}_{\boldsymbol{Pa}(Z)} \times \Omega_Z$). Let $X \in \boldsymbol{X} - \{Z\}$ be a random variable.*

(i) When X is not a descendant of Z, the marginal distribution of X is not perturbed.

(ii) The marginal distribution of Z, $p(z)$, is perturbed and is given by

$$\tilde{p}(z) = \sum_{l=1; l \neq l^*}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Z)}\right) p_l(z) + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \tilde{p}_{l^*}(z),$$

where $P\left(\Omega^{l}_{\boldsymbol{Pa}(Z)}\right)$ is the probability of the l-th parent configuration. The sensitivity measure between the perturbed and original marginal distributions is given by

$$S(p(z), \tilde{p}(z)) = P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big).$$

(iii) When X is a descendant of Z, the marginal distribution of X, $p(x)$, is perturbed and satisfies

$$S\big(p(x), \tilde{p}(x)\big) \leq P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big).$$

On the other hand, in the case of discrete random variables, the perturbation is analysed using the one–way sensitivity analysis from Jensen and Nielsen (2007, Chap.5). In these perturbations, the probability of a state y of a discrete random variable Y, given a configuration $\boldsymbol{pa}$ of its parents $\boldsymbol{Pa}(Y)$, is changed to h, with $h \in (0, 1)$. Formally,

$$P(Y = y | \boldsymbol{Pa}(Y) = \boldsymbol{pa}) = h, \quad P(Y \neq y | \boldsymbol{Pa}(Y) = \boldsymbol{pa}) = 1 - h, \quad h \in (0, 1).$$

Theorem 9 shows how a perturbation of a parameter in a discrete CPD propagates to the marginal distributions in a MoP BN.

Theorem 9 *Let $\boldsymbol{X}$ be the random vector associated with a MoP BN, and let $Y \in \boldsymbol{X}$ be a discrete random variable whose CPD follows Definition 4. Let $X \in \boldsymbol{X} - \{Y\}$ be a random variable, and let $\tilde{p}\left(y | \boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) = \tilde{a}_y^{(l^*)}$ be the perturbed conditional probability corresponding to the configuration $\Omega^{l^*}_{\boldsymbol{Pa}(Y)}$.*

(i) When X is not a descendant of Y, the marginal distribution of X is not perturbed.

(ii) The marginal distribution of Y, $p(y)$, is perturbed and is given by

$$\tilde{p}(y) = \sum_{l=1; l \neq l^*}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right) a_y^{(l)} + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) \tilde{a}_y^{(l^*)},$$

where $P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right)$ is the probability of the l-th configuration of the parents. The sensitivity measure between perturbed and original marginal distributions is given by

$$S\big(p(y), \tilde{p}(y)\big) = P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) S\left(p\left(y | \boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right), \tilde{p}\left(y | \boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\right)$$

(iii) When X is a descendant of Y, the marginal distribution of X, $p(x)$, is perturbed and

$$S\big(p(x),\tilde{p}(x)\big) \leq P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) S\left(p\left(y|\boldsymbol{Pa}(Y)\in\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right),\tilde{p}\left(y|\boldsymbol{Pa}(Y)\in\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\right)$$

Both Theorems 8 and 9 show that a perturbation in a MoP BN modifies the marginal of the perturbed node and propagates to descendant marginals, with the resulting sensitivity being bounded by the sensitivity of the perturbation scaled by the probability of the partition where the perturbation is applied.

5. Results and Discussion

To perform the experimental analysis, we used the MoTBFs R package (Pérez-Bernabé et al., 2020). We considered two different cases (i) a sensitivity analysis of univariate MoP densities; (ii) and a sensitivity analysis in the BN shown in Example 1. In all cases, the experiments were carried out by adding a perturbation proportional to the value of the parameter.

Sensitivity Analysis in univariate MoP densities

For the experimental analysis, we considered six probability distributions with distinct shapes, including symmetric, skewed, and heavy-tailed cases, and conducted a sensitivity analysis for each of them. The original MoP densities were learnt from samples of 1000 observations drawn from the $\beta(0.5,0.5)$, $\beta(5,1)$, $\chi^2(8)$, $\Gamma(3,2)$, $log-normal$, and standard Gaussian $\mathcal{N}(0,1)$ distributions, after standardising the data using z-score normalisation. The number of parameters in each MoP component is selected by maximising the Bayesian Information Criterion. We performed one-way sensitivity analysis by adding proportional noise to each parameter, following Definition 5. For each parameter a_k, perturbations are introduced as a proportion of the value of the parameter, so that $h=\pm\delta a_k$, where $\delta\in\{0.01,0.05,0.1,0.5,1\}$, corresponding to percentage perturbations of $\pm1\%,\pm5\%,\pm10\%,\pm50\%$ and $\pm100\%$, respectively.

Figure 3 summarises the sensitivity measures obtained for each distribution and perturbation. Each panel corresponds to one of the six distributions considered. The horizontal axis represents the perturbed polynomial term, whereas the vertical axis represents the percentage of perturbation. The resulting sensitivity measure is represented using a red colour scale, where darker tones indicate larger sensitivity values.

In the majority of cases, adding at least 5% of noise to a parameter means that the sensitivity measure will be greater than 0.2. This fact means that the difference between the probability of some event A can be greater than 0.2 when the original MoP distribution is perturbed. On the other hand, symmetric random variables tend to show lower sensitivity than asymmetric ones under the same proportional perturbation. Finally, lower degrees ($k=0,1,2$) generally yield lower sensitivity values than higher degrees ($k=6,7,8$). These results are as expected due to the collinearity in the polynomial regression, which induces high sensitivity in the parameter estimates.

To reduce collinearity in polynomial regression, Shacham and Brauner (1997) suggests transforming the data to the interval $[-1,1]$. We performed another experiment using the same aforementioned samples and applying this transformation instead of the z-score

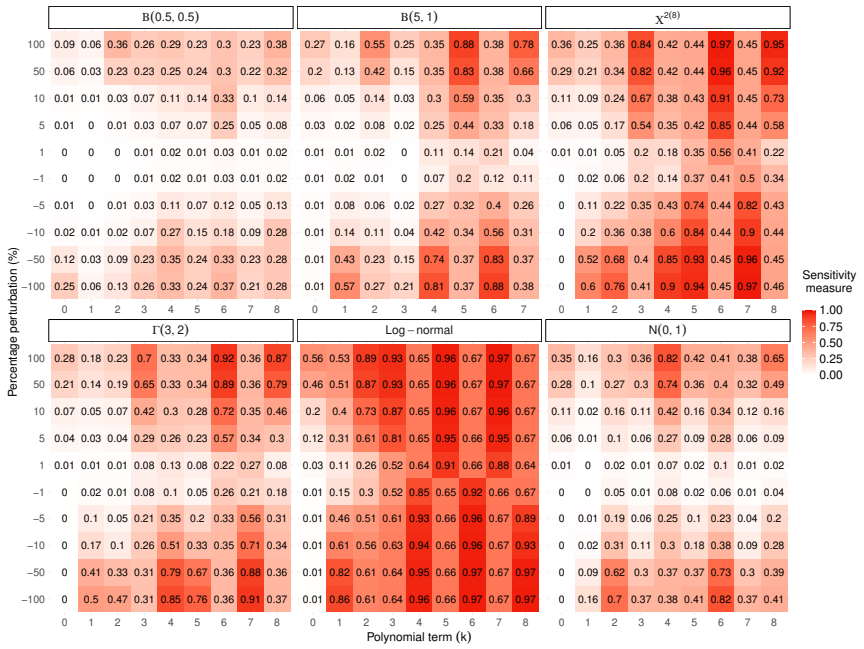


Figure 3: Sensitivity values of several MoP densities fitted to samples drawn from the $\beta(0.5, 0.5)$, $\beta(5, 1)$, $\chi^2(8)$, $\Gamma(3, 2)$, $log - normal$ and $\mathcal{N}(0, 1)$ distributions, after z-score standardization. The perturbations of each parameter are proportional to its value.

standardization. The results are summarised in Figure 4. Overall, the sensitivity of the parameters is clearly reduced when the training data follows an asymmetric distribution. In this case, perturbations of at most 5% result in sensitivity values lower than 0.2 in most cases. The highest sensitivity values are mainly obtained for large perturbations (50 and 100% of the parameter value). Finally, higher-degree parameters usually show higher sensitivity than the remaining parameters. To sum up, the numerical transformation to the interval $[-1, 1]$ tends to reduce the parameter sensitivity compared to the z-score standardization. In addition, heavy-tailed distributions present higher sensitivity in the parameters than symmetric ones.

Sensitivity Analysis in MoP BN

In this section, we analyse how perturbations propagate to the marginal distribution in a MoP BN. To do this, we consider the BN in Example 1. The continuous variable Z_3 is selected as the reference node, as it propagates perturbations to the largest number of descendant nodes in the network (Z_5, Y_6, and Z_7). Figure 5 shows the sensitivity values of the marginal distribution of each node after applying perturbations to each parameter of each MoP component of the Z_3 CPD.



Figure 4: Sensitivity values of several MoP densities fitted to samples drawn from the $\beta(0.5, 0.5)$, $\beta(5, 1)$, $\chi^2(8)$, $\Gamma(3, 2)$, $log - normal$ and $\mathcal{N}(0, 1)$ distributions, after normalizing to the interval $[-1, 1]$. The perturbations of each parameter are proportional to its value.

The marginal distribution of Z_3 exhibits the highest sensitivity value for each perturbation. The effects of the perturbations vary depending on the polynomial term. When a parameter of $Z_3|Y_1$ is perturbed, the largest effects on the other variables are observed in the terms a_1 and a_3. These differences are related to the partition structure of Z_3 in the CPDs of its children. When the probability of the partitions is not changed, then the marginal distribution of the descendants does not change, which is aligned with Theorem 8.

6. Conclusions

In this paper, the behaviour of the sensitivity of the MoP models introduced by Langseth et al. (2012) has been studied. To do this, we have introduced a sensitivity measure between the original and the perturbed MoP densities based on total variation distance. The main conclusion of this study is that perturbations within a component of a MoP distribution propagate to marginal distributions in the BN with decreasing intensity as conditional relationships are encoded through the discretization of continuous random variables. Consequently, parameter sensitivity in MoP BNs aligns with that observed in discrete BNs (Laskey, 1995; Chan and Darwiche, 2002; Ballester-Ripoll and Leonelli, 2022). Consequently, sensitivity analysis in MoP BNs should primarily focus on these distributions.

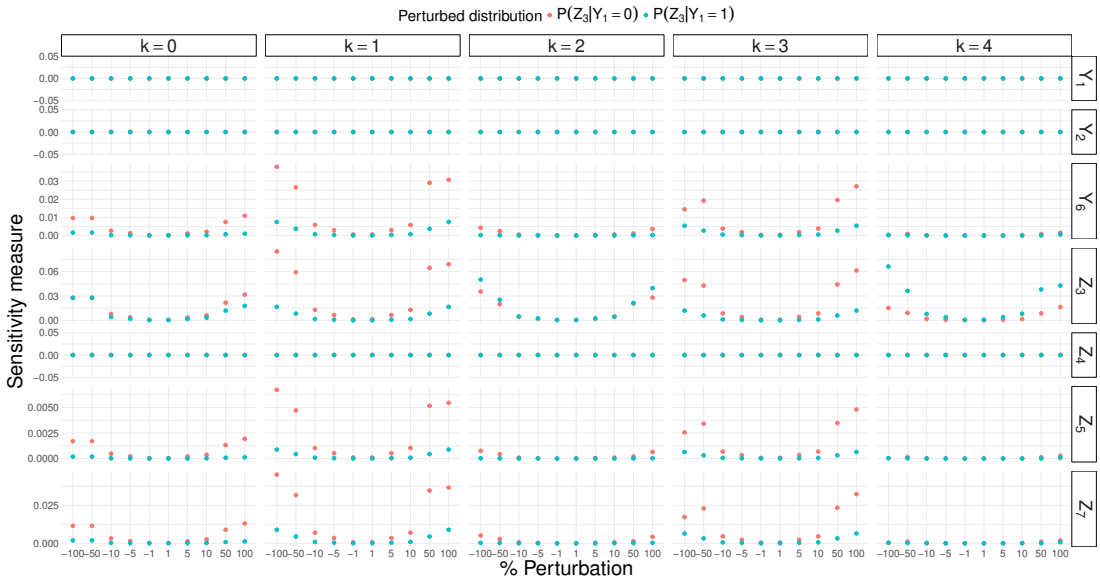


Figure 5: Sensitivity measure of each marginal distribution of the BN of Example 1 when the parameters of $Z_3|Y_1$ distribution are perturbed. The perturbation considered are proportional to parameter value, i. e. same percentage of perturbation.

Firstly, the analytical results are as expected. The conditional dependence between variables is encoded through support partitions. Therefore, these partitions are relevant in parameter sensitivity, as shown in Theorems 8 and 9.

Secondly, the experimental results confirm that in the parameter sensitivity of MoP densities, collinearity is an important factor, similar to polynomial regression. The numerical transformations can reduce collinearity and sensitivity in MoP distributions. On the other hand, the MoP BN experiments show that the sensitivity measure between original and perturbed marginal distributions can be lower than that between the original and the perturbed MoP distribution. In addition, partitioning the support of continuous variables in the conditional representation can reduce the perturbation effect.

Finally, this study has raised some questions about parameter sensitivity in MoPs that will be addressed in future work. For instance, (i) how evidence affects the model when a parameter is perturbed; (ii) how important the degree of the polynomial is in a MoP density; (iii) how distributions are affected when the partition of parent variables changes.

Acknowledgments

The work of Á. T. Sáez-Ruiz, A. D. Maldonado, and R. Rumí is supported by grant PID2022-139293NB-C31 and PID2025-172960NB-C33 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU "A way of Making Europe", by the research group FQM-244, grant PPIT-UAL, Junta de Andalucía-ERDF 2021-2027, Objective RSO1.1. Programme: 54.A.

References

Pedro A. Aguilera, Antonio Fernández, Rosa Fernández, Rafael Rumí, and A. Salmerón. Bayesian networks in environmental modelling. *Environmental Modelling & Software*, 26 (12):1376–1388, 2011.

David Atienza, Concha Bielza, and Pedro Larrañaga. Semiparametric Bayesian networks. *Information Sciences*, 584:564–582, 2022.

Rafael Ballester-Ripoll and Manuele Leonelli. You only derive once (YODO): automatic differentiation for efficient sensitivity analysis in Bayesian networks. In *International Conference on Probabilistic Graphical Models*, pages 169–180. PMLR, 2022.

Rafael Ballester-Ripoll and Manuele Leonelli. Global sensitivity analysis of uncertain parameters in Bayesian networks. *International Journal of Approximate Reasoning*, 180: 109368, 2025.

Concha Bielza and Pedro Larrañaga. Bayesian networks in neuroscience: a survey. *Frontiers in Computational Neuroscience*, 8, 2014. ISSN 1662-5188.

Christopher M. Bishop and Nasser M. Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.

Enrique Castillo and Uffe Kjærulff. Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique. *Reliability Engineering & System Safety*, 79(2): 139–148, 2003.

Hei Chan and Adnan Darwiche. When do numbers really matter? *Journal of artificial intelligence research*, 17:265–287, 2002.

Marek J. Druzdzel and Linda C. Van Der Gaag. Elicitation of probabilities for belief networks: combining qualitative and quantitative information. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, page 141–148. Morgan Kaufmann Publishers Inc., 1995. ISBN 1558603859.

Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521 (7553):452–459, 2015.

Miguel A. Gómez-Villegas, Paloma Maín, and Rosario Susi. Extreme inaccuracies in Gaussian Bayesian networks. *Journal of Multivariate Analysis*, 99(9):1929–1940, 2008.

Miguel A. Gómez-Villegas, Paloma Maín, and Rosario Susi. Sensitivity analysis in Gaussian Bayesian networks using a divergence measure. *Communications in Statistics - Theory and Methods*, 36(3):523–539, 2007.

Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer, 2007.

Helge Langseth, Thomas D. Nielsen, Rafael Rumí, and Antonio Salmerón. Mixtures of truncated basis functions. *International Journal of Approximate Reasoning*, 53(2):212–227, 2012.

Helge Langseth, Thomas D. Nielsen, Inmaculada Pérez-Bernabé, and Antonio Salmerón. Learning mixtures of truncated basis functions from data. *International Journal of Approximate Reasoning*, 55(4):940–956, 2014.

Kathryn B. Laskey. Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(6):901–909, 1995.

Steffen L. Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

Pedro G. Machado, Celma O. Ribeiro, and Claudio A. O. do Nascimento. Risk analysis in energy projects using Bayesian networks: A systematic review. *Energy Strategy Reviews*, 47:101097, 2023.

Serafín Moral, Rafael Rumí, and Antonio Salmerón. Mixtures of truncated exponentials in hybrid Bayesian networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 156–167. Springer, 2001.

Kevin P Murphy. *Machine Learning: a Probabilistic Perspective*. MIT press, 2012.

Judea Pearl. Probabilistic Reasoning in Intelligent Systems. *Morgan Kaufmann*, 1988.

Inmaculada Pérez-Bernabé, Ana D. Maldonado, Antonio Salmerón, Thomas D Nielsen, et al. MoTBFs: An R package for learning hybrid Bayesian networks using mixtures of truncated basis functions. *The R Foundation*, 2020.

Thuc D. Phan, James C. R. Smart, Ben Stewart-Koster, Oz. Sahin, Wade L. Hadwen, Lien T. Dinh, Iman Tahmasbian, and Samantha J. Capon. Applications of Bayesian networks as decision support tools for water resource management under climate change and socio-economic stressors: A critical popraisal. *Water*, 11(12), 2019.

Jeremy Rohmer. Uncertainties in conditional probability tables of discrete Bayesian belief networks: A comprehensive review. *Engineering Applications of Artificial Intelligence*, 88:103384, 2020.

Modechai Shacham and Neima Brauner. Minimizing the effects of collinearity in polynomial regression. *Industrial & Engineering Chemistry Tesearch*, 36(10):4405–4412, 1997.

Ross D. Shachter and C. Robert Kenley. Gaussian influence diagrams. *Management Science*, 35(5):527–550, 1989.

Prakash P. Shenoy and James C. West. Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52(5):641–657, 2011.

Alexandre B. Tsybakov. *Lower bounds on the minimax risk*, pages 77–135. Springer, 2009.

Junhua Zheng, Yue Zhuo, Xiaoyu Jiang, Lingquan Zeng, and Zhiqiang Ge. Advances in Bayesian networks for industrial process analytics: Bridging data and mechanisms. *Expert systems with applications*, 271, 2025.

Mingmin Zhu, Sanyang Liu, and Jiewei Jiang. A novel divergence for sensitivity analysis in Gaussian Bayesian networks. *International Journal of Approximate Reasoning*, 90:37–55, 2017.

Appendix A. Proofs

In this section, the proofs omitted in the paper are included.

Proposition 7 *Let Z be a continuous random variable with a MoP density p of degree r as in Definition 2, and let $h \in \mathbb{R}$ and $k \in \{0, 1, \ldots, r\}$. Then, the transformation $\tilde{p}(z; k, h)$ introduced in Definition 5 is a MoP density.*

Proof If we consider $q(z; k, h) = p(z) + hz^k - a(p, k, h)$, we have

$$\begin{aligned} q(z; k, h) = p(z) + hz^k - a(p, k, h) &\geq \min_{\alpha \leq z \leq \beta} \left(p(z) + hz^k \right) - a(p, k, h) \\ &\geq \min_{\alpha \leq z \leq \beta} \left(p(z) + hz^k, 0 \right) - a(p, k, h) = a(p, k, h) - a(p, k, h) = 0. \end{aligned}$$

And,

$$\begin{aligned} \int_{\Omega_Z} q(z; k, h) dz &= \int_{\alpha}^{\beta} \left(p(z) + hz^k - a(p, k, h) \right) dz = \int_{\alpha}^{\beta} p(z) dz + \int_{\alpha}^{\beta} hz^k dz - \int_{\alpha}^{\beta} a(p, k, h) dz \\ &= 1 + \frac{h\left(\beta^{k+1} - \alpha^{k+1}\right)}{k+1} - a(p, k, h)\left(\beta - \alpha\right) = A(p, k, h). \end{aligned}$$

Then, $q(z; k, h)$ is a MoP potential and $\tilde{p}(z; k, h) = \frac{q(z;k,h)}{A(p,k,h)}$ is a MoP density. ■

Theorem 9 *Let $\boldsymbol{X}$ be the random vector associated with a MoP BN, and let $Z \in \boldsymbol{X}$ be a continuous random variable whose CPD follows Definition 3. Let $l^* \in \{1, \ldots, m\}$ be the index of a component of the CPD of Z. Suppose that the l^*-th component of the CPD of Z is perturbed, and let $\tilde{p}_{l^*}(z)$ denote the perturbed MoP density of $Z \mid \boldsymbol{Pa}(Z) \in \Omega^{l^*}_{\boldsymbol{Pa}(Z)}$. Let $X \in \boldsymbol{X} - \{Z\}$ be a random variable.*

(i) When X is not a descendant of Z, the marginal distribution of X is not perturbed.

(ii) The marginal distribution of Z, $p(z)$, is perturbed and is given by

$$\tilde{p}(z) = \sum_{l=1; l \neq l^*}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Z)}\right) p_l(z) + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \tilde{p}_{l^*}(z),$$

where $P\left(\Omega^{l}_{\boldsymbol{Pa}(Z)}\right)$ is the probability of the l-th parent configuration. The sensitivity measure between the perturbed and original marginal distributions is given by

$$S(p(z), \tilde{p}(z)) = P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big).$$

(iii) When X is a descendant of Z, the marginal distribution of X, $p(x)$, is perturbed and satisfies

$$S\big(p(x), \tilde{p}(x)\big) \leq P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big).$$

Proof

(i) Trivially. If there is no evidence, Z and X are $d-$separated.

(ii) Using variable elimination, we only have to considered the BN formed by Z and its parents, $\boldsymbol{Pa}(Z)$. Let $q()$ be the MoP density of the joint distribution of $\boldsymbol{Pa}(Z)$. Remaining the CPD of Z, Definition 3 we have that

$$p(z, \boldsymbol{Pa}(Z)) = q(\boldsymbol{Pa}(Z))p(Z|\boldsymbol{Pa}(Z)) = \begin{cases} q(\boldsymbol{Pa}(Z))p_l(z) & \text{if } z \in \Omega_Z, \boldsymbol{Pa}(Z) \in \Omega^l_{\boldsymbol{Pa}(Z)}, \\ 0 & \text{in another case.} \end{cases}$$

with $l = 1, \ldots, m$. Marginalizing, we have

$$p(z) = \sum_{l=1}^{m} \int_{\Omega^l_{\boldsymbol{Pa}(Z)}} q(\boldsymbol{Pa}(Z))p_l(z)d\boldsymbol{Pa}(Z) = \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z),$$

where $P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right)$ denotes the probability of $\Omega^l_{\boldsymbol{Pa}(Z)}$ for each $l = 1, \ldots, m$. Therefore, when p_{l^*} is perturbed

$$\tilde{p}(z) = \sum_{l=1; l\neq l^*}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \tilde{p}_{l^*}(z).$$

Finally, the sensitivity measure in the case of $\tilde{p}$ is computed as

$$\begin{aligned} &S\big(p(z), \tilde{p}(z)\big) = \frac{1}{2}\int_{\Omega_Z} |p(z) - \tilde{p}(z)|dz \\ &= \frac{1}{2}\int_{\Omega_Z} \left| \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) - \left(\sum_{l=1; l\neq l^*}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \tilde{p}_{l^*}(z) \right) \right| dz \\ &= \frac{1}{2}\int_{\Omega_Z} \left| P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \left(p_{l^*}(z) - \tilde{p}_{l^*}(z)\right) \right| dz = \frac{1}{2} P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \int_{\Omega_Z} \left|\left(p_{l^*}(z) - \tilde{p}_{l^*}(z)\right)\right| dz \\ &= P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big). \end{aligned}$$

(iii) Firstly, we use variable elimination to obtain the BN with only Z and X as variables. We have to consider two cases: X is a continuous variable, or X is a discrete variable. When $X = Z_1$ is a continuous variable, the CPD of $Z_1|Z$ has the expression

$$p(z_1|z) = \begin{cases} p_s(z_1) & \text{if } z_1 \in \Omega_Z, z \in \Omega^s_Z, \\ 0 & \text{in another case.} \end{cases}$$

where Ω^s_Z, with $s = 1, \ldots, t$ is a partition of Ω_Z. Using the expression of marginal distribution of Z computed in the proof of (ii), we have

$$\begin{aligned} p(z_1) &= \int_{\Omega_Z} \sum_{s=1}^{t} p_s(z_1)p(z)dz = \int_{\Omega_Z} \sum_{s=1}^{t} p_s(z_1) \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z)dz \\ &= \sum_{s=1}^{t} \int_{\Omega^s_Z} p_s(z_1) \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z)dz = \sum_{s=1}^{t} p_s(z_1) \int_{\Omega^s_Z} \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z)dz. \end{aligned}$$

Then, if the perturbed $\tilde{p}(z)$ is considered,

$$\tilde{p}(z_1) = \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} \left(\sum_{l=1;l\neq l^*}^{m} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l}\right) p_l(z) + P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \tilde{p}_{l^*}(z) \right) dz.$$

Then,

$$\begin{aligned}
|p(z_1) - \tilde{p}(z_1)| &= \left| \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} \sum_{l=1}^{m} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l}\right) p_l(z) dz \right. \\
&\quad \left. - \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} \left(\sum_{l=1;l\neq l^*}^{m} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l}\right) p_l(z) + P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \tilde{p}_{l^*}(z) \right) dz \right| \\
&= \left| \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \left(p_{l^*}(z) - \tilde{p}_{l^*}(z)\right) dz \right| \\
&\leq \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz \\
&= P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz.
\end{aligned}$$

In this way, if we apply that $p_s(z_1)$ is a density in Ω_{Z_1} for each $s = 1, \ldots, t$ and $\Omega_Z^1, \ldots, \Omega_Z^t$ a partition of Ω_Z, the sensitivity measure is upper bounded by

$$\begin{aligned}
S(p(z_1), \tilde{p}(z_1)) &= \frac{1}{2} \int_{\Omega_{Z_1}} |p(z_1) - \tilde{p}(z_1)| dz_1 \\
&\leq \frac{1}{2} \int_{\Omega_{Z_1}} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \sum_{s=1}^{t} p_s(z_1) \int_{\Omega_Z^s} \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz dz_1 \\
&= \frac{1}{2} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \sum_{s=1}^{t} \int_{\Omega_{Z_1}} p_s(z_1) \int_{\Omega_Z^s} \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz dz_1 \\
&= \frac{1}{2} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \sum_{s=1}^{t} \int_{\Omega_{Z_1}} p_s(z_1) dz_1 \int_{\Omega_Z^s} \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz \\
&= \frac{1}{2} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \sum_{s=1}^{t} \int_{\Omega_Z^s} \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz \\
&= \frac{1}{2} P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) \int_{\Omega_Z} \left|p_{l^*}(z) - \tilde{p}_{l^*}(z)\right| dz \\
&= P\left(\Omega_{\boldsymbol{Pa}(Z)}^{l^*}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big).
\end{aligned}$$

Finally, if $X = Y$ is a discrete random variable, the obtained CPD of $Y|Z$ has the expression

$$p(y|z) = \begin{cases} a_y^{(s)} & \text{if } y \in \Omega_Y, z \in \Omega_Z^s, \\ 0 & \text{in another case.} \end{cases}$$

where Ω_Z^s, with $s = 1, \ldots, t$ is a partition of Ω_Z. In an analogous way to the previous case,

$$
\begin{aligned}
p(y) &= \int_{\Omega_Z} \sum_{s=1}^{t} a_y^{(s)} p(z) dz = \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) dz, \\
\tilde{p}(y) &= \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} \left(\sum_{l=1; l\neq l^*}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \tilde{p}_{l^*}(z) \right) dz. \\
|p(y) - \tilde{p}(y)| &= \left| \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} \sum_{l=1}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) dz \right. \\
&\quad \left. - \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} \left(\sum_{l=1; l\neq l^*}^{m} P\left(\Omega^l_{\boldsymbol{Pa}(Z)}\right) p_l(z) + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \tilde{p}_{l^*}(z) \right) dz \right| \\
&\leq P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} |p_{l^*}(z) - \tilde{p}_{l^*}(z)| \, dz.
\end{aligned}
$$

Therefore, changing integral by sum in the definition 7, the sensitivity measure can be computed as

$$
\begin{aligned}
S(p(y), \tilde{p}(y)) &= \frac{1}{2} \sum_{y \in \Omega_Y} |p(y) - \tilde{p}(y)| \\
&\leq \frac{1}{2} \sum_{y \in \Omega_Y} P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} |p_{l^*}(z) - \tilde{p}_{l^*}(z)| \, dz \\
&= \frac{1}{2} \left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \sum_{y \in \Omega_Y} \sum_{s=1}^{t} a_y^{(s)} \int_{\Omega_Z^s} |p_{l^*}(z) - \tilde{p}_{l^*}(z)| \, dz \\
&= \frac{1}{2} \left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \sum_{s=1}^{t} \int_{\Omega_Z^s} |p_{l^*}(z) - \tilde{p}_{l^*}(z)| \, dz \sum_{y \in \Omega_Y} a_y^{(s)}.
\end{aligned}
$$

Note that $\sum_{y \in \Omega_Y} a_y^{(s)} = 1$ for each $s = 1, \ldots, t$. Therefore, the sensitivity measure between $p(y)$ and $\tilde{p}(y)$ is bounded by

$$
\begin{aligned}
S(p(y), \tilde{p}(y)) &\leq \frac{1}{2} \left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \sum_{s=1}^{t} \int_{\Omega_Z^s} |p_{l^*}(z) - \tilde{p}_{l^*}(z)| \, dz \\
&= \frac{1}{2} \left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) \int_{\Omega_Z} |p_{l^*}(z) - \tilde{p}_{l^*}(z)| \, dz \\
&= P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Z)}\right) S\big(p_{l^*}(z), \tilde{p}_{l^*}(z)\big).
\end{aligned}
$$

■

Theorem 10 *Let $\boldsymbol{X}$ be the random vector associated with a MoP BN, and let $Y \in \boldsymbol{X}$ be a discrete random variable whose CPD follows Definition 4. Let $X \in \boldsymbol{X} - \{Y\}$ be a random variable, and let $\tilde{p}\left(y|\boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) = \tilde{a}_y^{(l^*)}$ be the perturbed conditional probability corresponding to the configuration $\Omega^{l^*}_{\boldsymbol{Pa}(Y)}$.*

(i) When X is not a descendant of Y, the marginal distribution of X is not perturbed.

(ii) The marginal distribution of Y, $p(y)$, is perturbed and is given by

$$\tilde{p}(y) = \sum_{l=1; l\neq l^*}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right) a_y^{(l)} + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) \tilde{a}_y^{(l^*)},$$

where $P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right)$ is the probability of the l-th configuration of the parents. The sensitivity measure between perturbed and original marginal distributions is given by

$$S\big(p(y), \tilde{p}(y)\big) = P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) S\left(p\left(y|\boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right), \tilde{p}\left(y|\boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\right)$$

(iii) When X is a descendant of Y, the marginal distribution of X, $p(x)$, is perturbed and

$$S\big(p(x), \tilde{p}(x)\big) \leq P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) S\left(p\left(y|\boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right), \tilde{p}\left(y|\boldsymbol{Pa}(Y) \in \Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\right)$$

Proof

(i) Trivially. If there is no evidence, Y and X are X $d-$separated

(ii) Using variable elimination, we have to consider only the BN formed by Y and its parents $\boldsymbol{Pa}(Y)$. Lets $q(\boldsymbol{Pa}(Y))$ the MoP density of the joint distribution of $\boldsymbol{Pa}(Y)$. Remaining the definition 4, we have that

$$p(y, \boldsymbol{Pa}(Y)) = q(\boldsymbol{Pa}(Y))p(y|\boldsymbol{Pa}(Y)) = \begin{cases} a_y^{(l)} q(\boldsymbol{Pa}(Y)) & \text{if } y \in \Omega_Y, \boldsymbol{Pa}(Y) \in \Omega^{l}_{\boldsymbol{Pa}(Y)}, \\ 0 & \text{in another case.} \end{cases}$$

where $0 \leq a_y^{(l)} \leq 1$ and $\sum_{y \in \Omega_Y} a_y^{(l)} = 1$ for each $y \in \Omega_Y$ and $l = 1, \ldots, m$. Marginalizing,

$$p(y) = \sum_{l=1}^{m} \int_{\Omega^{l}_{\boldsymbol{Pa}(Y)}} a_y^{(l)} q(\boldsymbol{Pa}(Y)) d\boldsymbol{Pa}(Y) = \sum_{l=1}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right) a_y^{(l)}.$$

Therefore, the perturbed marginal distribution of Y is

$$\tilde{p}(y) = \sum_{l=1; l\neq l^*}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right) a_y^{(l)} + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right) \tilde{a}_y^{(l^*)}.$$

Finally, the perturbation measure between the true and the perturbed marginal distribution of Y is

$$\begin{aligned}
S(p(y),\tilde{p}(y)) &= \frac{1}{2}\sum_{y\in\Omega_Y}|p(y)-\tilde{p}(y)| = \frac{1}{2}\sum_{y\in\Omega_Y}\left|\sum_{l=1}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right)a_y^{(l)}\right. \\
&\quad \left. - \sum_{l=1;l\neq l^*}^{m} P\left(\Omega^{l}_{\boldsymbol{Pa}(Y)}\right)a_y^{(l)} + P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\tilde{a}_y^{(l^*)}\right| \\
&= \frac{1}{2}\sum_{y\in\Omega_Y}\left|P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)a_y^{(l)} - P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\tilde{a}_y^{(l^*)}\right| \\
&= \frac{1}{2}P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\sum_{y\in\Omega_Y}\left|a_y^{(l)} - \tilde{a}_y^{(l^*)}\right| \\
&= P\left(\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)S\left(p\left(y|\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right),\tilde{p}\left(y|\Omega^{l^*}_{\boldsymbol{Pa}(Y)}\right)\right)
\end{aligned}$$

(iii) As well as the proof of Theorem 8(iii), using variable elimination, we only have to prove the BN with only Y and X as variables. We have to consider two cases, X is a continuous or a discrete random variable. Firstly, we consider $X = Z$. Using the marginal distribution obtained in the previous case, we have

$$p(z,y) = \begin{cases} p(y)p_y(z) & \text{if } z\in\Omega_Z, y\in\Omega_Y, \\ 0 & \text{in another case.}\end{cases}$$

Marginalizing the previous expression,

$$p(z) = \sum_{y\in\Omega_Y} p(y)p_y(z), \quad \tilde{p}(z) = \sum_{y\in\Omega_Y}\tilde{p}(y)p_y(z)$$

Then,

$$\begin{aligned}
&S(p(z),\tilde{p}(z)) = \frac{1}{2}\int_{\Omega_Z}|p(z)-\tilde{p}(z)|\,dz = \frac{1}{2}\int_{\Omega_Z}\left|\sum_{y\in\Omega_Y}p(y)p_y(z) - \sum_{y\in\Omega_Y}\tilde{p}(y)p_y(z)\right|dz \\
&= \frac{1}{2}\int_{\Omega_Z}\left|\sum_{y\in\Omega_Y}p(y)p_y(z) - \sum_{y\in\Omega_Y}\tilde{p}(y)p_y(z)\right|dz = \frac{1}{2}\int_{\Omega_Z}\left|\sum_{y\in\Omega_Y}\left(p(y)-\tilde{p(y)}\right)p_y(z)\right|dz \\
&\leq \frac{1}{2}\int_{\Omega_Z}\sum_{y\in\Omega_Y}|p(y)-\tilde{p}(y)|\,p_y(z)dz = \frac{1}{2}\sum_{y\in\Omega_Y}|p(y)-\tilde{p}(y)|\int_{\Omega_Z}p_y(z)dz \\
&= \frac{1}{2}\sum_{y\in\Omega_Y}|p(y)-\tilde{p}(y)| = S(p(y),\tilde{p}(y)).
\end{aligned}$$

On the other hand, if $X = Y_1$, discrete random variable, the joint distribution of Y, Y_1 has the expression

$$p(y,y_1) = \begin{cases} p(y)a_{y_1}^{(y)} & \text{if } y_1\in\Omega_{Y_1}, y\in\Omega_Y, \\ 0 & \text{in another case.}\end{cases}$$

Then, marginalizing, we have

$$p(z) = \sum_{y \in \Omega_Y} p(y) a_{y_1}^{(y)}, \quad \tilde{p}(z) = \sum_{y \in \Omega_Y} \tilde{p}(y) a_{y_1}^{(y)}.$$

Therefore,

$$\begin{aligned}
S(p(y_1), \tilde{p}(y_1)) &= \frac{1}{2} \sum_{y_1 \in \Omega_{Y_1}} |p(y_1) - \tilde{p}(y_1)| = \frac{1}{2} \sum_{y_1 \in \Omega_{Y_1}} \left| \sum_{y \in \Omega_Y} p(y) a_{y_1}^{(y)} - \sum_{y \in \Omega_Y} \tilde{p}(y) a_{y_1}^{(y)} \right| \\
&= \frac{1}{2} \sum_{y_1 \in \Omega_{Y_1}} \left| \sum_{y \in \Omega_Y} p(y) a_{y_1}^{(y)} - \sum_{y \in \Omega_Y} \tilde{p}(y) a_{y_1}^{(y)} \right| = \frac{1}{2} \sum_{y_1 \in \Omega_{Y_1}} \left| \sum_{y \in \Omega_Y} \left(p(y) - \tilde{p(y)} \right) a_{y_1}^{(y)} \right| \\
&\leq \frac{1}{2} \sum_{y_1 \in \Omega_{Y_1}} \sum_{y \in \Omega_Y} |p(y) - \tilde{p}(y)| \, a_{y_1}^{(y)} = \frac{1}{2} \sum_{y \in \Omega_Y} |p(y) - \tilde{p}(y)| \sum_{y_1 \in \Omega_{Y_1}} a_{y_1}^{(y)} \\
&= \frac{1}{2} \sum_{y \in \Omega_Y} |p(y) - \tilde{p}(y)| = S(p(y), \tilde{p}(y)).
\end{aligned}$$

■