Journal of Machine Learning Research 21 (2020) 1-32 Submitted 9/18; Revised 4/20; Published 5/20

Model-Preserving Sensitivity Analysis for Families of Gaussian Distributions

Christiane Görgen CHRISTIANE.GOERGEN@MIS.MPG.DE
Max Planck Institute for Mathematics in the Sciences
Inselstraße 22,
04103 Leipzig, Germany

Manuele Leonelli MANUELE.LEONELLI@IE.EDU
School of Human Sciences and Technology
IE University
Madrid, Spain

Editor: Peter Spirtes

Abstract

The accuracy of probability distributions inferred using machine-learning algorithms heavily depends on data availability and quality. In practical applications it is therefore fundamental to investigate the robustness of a statistical model to misspecification of some of its underlying probabilities. In the context of graphical models, investigations of robustness fall under the notion of sensitivity analyses. These analyses consist in varying some of the model's probabilities or parameters and then assessing how far apart the original and the varied distributions are. However, for Gaussian graphical models, such variations usually make the original graph an incoherent representation of the model's conditional independence structure. Here we develop an approach to sensitivity analysis which guarantees the original graph remains valid after any probability variation and we quantify the effect of such variations using different measures. To achieve this we take advantage of algebraic techniques to both concisely represent conditional independence and to provide a straightforward way of checking the validity of such relationships. Our methods are demonstrated to be robust and comparable to standard ones, which can break the conditional independence structure of the model, using an artificial example and a medical real-world application.

Keywords: Conditional independence, Gaussian models, Graphical models, Kullback-Leibler divergence, Sensitivity analysis

1. Introduction

The validation of both machine-learnt and expert-elicited statistical models is one of the most critical phases of any applied analysis. Broadly speaking, this validation phase consists of checking that a model produces outputs that are in line with current understanding, following a defensible and expected mechanism (French, 2003; Pitchforth and Mengersen, 2013). A critical aspect of such a validation is the investigation of the effects of variations in the model's inputs to outputs of interest. These types of investigations are usually referred to as *sensitivity analyses* (Borgonovo and Plischke, 2016).

Various sensitivity methods are now in place for generic statistical models (Rohmer, 2020; Saltelli et al., 2000). A large proportion of these have focused on graphical models (Lauritzen, 1996) and, in particular, on Bayesian network (BN) models (Pearl, 1988). A BN is a graphical representation of a statistical model defined via a set of conditional independence statements (Dawid, 1979).

Sensitivity analysis in BNs usually consists of two main steps. First *local* changes on outputs of interest are investigated via sensitivity functions: so probabilities are studied as functions of the input parameters as these vary in some appropriate interval. Once possible input parameter changes have been identified, the *global* effects that these would have on the overall distribution of the network are studied. These global effects are usually quantified by some divergence or distance between the original and the varied distributions, for instance using the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951).

Sensitivity methods in BNs usually focus on either models consisting of discrete random variables or, in the continuous case, of multivariate Gaussian distributions. The properties of sensitivity functions in the discrete case have been studied extensively (Castillo et al., 1997; Coupé and van der Gaag, 2002). Here, when some probabilities of interest are varied, then some others, namely those associated to the same conditional probability tables, need to *covary* in order to respect the sum-to-one condition of probabilities. The gold-standard in this context is *proportional covariation* which assigns to the covarying parameters the same proportion of the remaining probability mass as they originally had (Laskey, 1995; Renooij, 2014). The use of proportional covariation is justified by a variety of optimality criteria because it often minimizes the distance between the original and varied distribution among all possible covariations (Chan and Darwiche, 2002; Leonelli et al., 2017; Leonelli and Riccomagno, 2018). In the continuous case, the properties of sensitivity functions for Gaussian BNs (GBNs) have been known for quite some time (Castillo et al., 1997). They are rational functions of both the mean parameters and the entries of the covariance matrix of the Gaussian distribution associated to the network. Following these early developments, methods to quantify the distance between the original distribution and the one obtained from perturbations of the mean vector and covariance matrix were introduced (Gómez-Villegas et al., 2007, 2008, 2013), entailing the computation of the KL divergence between the two distributions.

One of the major drawbacks of the established Gaussian sensitivity methods is that in most cases perturbations of the covariance matrix make the graph of the original BN a non-faithful representation of the new distribution. This is because entries of the covariance matrix relate directly to conditional independence relationships between the depicted variables. In the discrete case this issue does not arise since a perturbation is applied directly to the conditional probability distributions associated to the BN rather than to the covariance structure of the model so that any new distribution automatically respects all the conditional independences of the model.

In practice, however, GBN users may want to apply a perturbation to some parameters whilst retaining the original graphical structure of their model and all of its entailed conditional independences. To tackle this issue we introduce a new class of perturbations of Gaussian vectors, called *model-preserving*, which have the property that the graphical representation of the original distribution remains valid after the perturbation. Whilst standard sensitivity methods act additively over the entries of the covariance matrix of the underlying Gaussian distribution, our model-preserving approach acts multiplicatively as formalized below. Furthermore, and conversely to standard sensitivity methods which only vary the entries of interest of the covariance matrix, in model-preserving perturbations additional parameters need to covary so that all conditional independences of the model are retained. In particular, this covariation ensures that the matrix under perturbation remains a covariance matrix of the original Gaussian model. This can be thought of as the continuous analogue of covariation techniques in the discrete case in the sense that it ensures that the varied object remains inside its original class.

Below we introduce various ways to select the parameters that need to covary for a given perturbation and we quantify the distance between the original and varied distributions using a variety of measures. We achieve this by adopting an algebraic approach which characterizes conditional independences as specific vanishing minors of a covariance matrix. Algebraic methods have been already used extensively in machine learning problems (see for instance Rusakov and Geiger, 2005; Zwiernik, 2011) but, to the best of our knowledge, we provide here their first application to sensitivity studies.

An implementation of the methods developed in this paper in the open-source R software (R Core Team, 2019) is given in the package `bnmonitor` and available at `https://github.com/manueleleonelli/bnmonitor`.

2. Conditional Independence and Gaussian Graphical Models

We start by reviewing the theory of Gaussian conditional independence models. We then focus on two graphical representations of specific sets of conditional independences, namely undirected and directed graphical models.

2.1. Gaussian Conditional Independence Models

Let Y be a n-dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}_{\text{psd}}$, where $\mathbb{R}^{n\times n}_{\text{psd}} \subset \mathbb{R}^{n\times n}$ denotes the cone of symmetric, positive semidefinite $n \times n$ matrices. Let $f_{\mu,\Sigma}$ be the density of a Gaussian distribution parametrized by μ and Σ. For index sets $A, B \subseteq [n] = \{1, \dots, n\}$, let $\mu_A = (\mu_i)_{i\in A}$ be the subvector of the mean with entries indexed by A and $\Sigma_{A,B}$ be the submatrix of Σ with rows indexed by A and columns indexed by B. Both marginal and conditional distributions of Gaussian vectors are Gaussian. In particular, for any two disjoint sets $A, B \subset [n]$, the random vector $Y_A = (Y_i)_{i\in A}$ has density $f_{\mu_A,\Sigma_{A,A}}$ and $Y_A|Y_B = y_B$ has density $f_{\mu^{A|B},\Sigma^{A|B}}$ where

$$\mu^{A|B} = \mu_A + \Sigma_{A,B}\Sigma^{-1}_{B,B}(y_B - \mu_B) \text{ and } \Sigma^{A|B} = \Sigma_{A,A} - \Sigma_{A,B}\Sigma^{-1}_{B,B}\Sigma_{B,A}. \qquad (1)$$

In this paper we consider Gaussian models defined by sets of conditional independence statements. The random vector Y_A is henceforth said to be *conditionally independent of* Y_B *given* Y_C for disjoint subsets $A, B, C \subseteq [n]$ if and only if the density factorizes as

$$f_{\mu^{A\cup B|C},\Sigma^{A\cup B|C}} = f_{\mu^{A|C},\Sigma^{A|C}} f_{\mu^{B|C},\Sigma^{B|C}}.$$

We sometimes abbreviate this statement to $A \perp\!\!\!\perp B \mid C$. The following lemma from Drton et al. (2008) demonstrates that conditional independence relationships in multivariate Gaussian models can be characterized in a straightforward algebraic way.

Lemma 1 (Proposition 3.1.13 of Drton et al. (2008)) *For a n-dimensional Gaussian random vector Y with density $f_{\mu,\Sigma}$ and disjoint $A, B, C \subset [n]$, the conditional independence statement $A \perp\!\!\!\perp B \mid C$ is true if and only if all $(\#C + 1) \times (\#C + 1)$ minors of the matrix $\Sigma_{A\cup C, B\cup C}$ are equal to zero. Here, $\#C$ denotes the cardinality of the set C.*

This duality between conditional independence and the vanishing of a set of equations provides the key insight on which we build our new algebraic sensitivity methods. In particular, in the subsequent sections we easily establish techniques for Gaussian graphical models which ensure that if a set of equations vanished before a perturbation of one or multiple entries of the covariance matrix then it will continue to vanish after an appropriate covariation of some of the other entries of that matrix.

Let henceforth $\mathrm{CI} = \{A_1 \perp\!\!\!\perp B_1 \mid C_1, \dots, A_r \perp\!\!\!\perp B_r \mid C_r\}$ denote a set of conditional independence statements for disjoint index sets $A_i, B_i, C_i \subset [n]$ and $i \in [r]$, with $r \in \mathbb{N}$. A *Gaussian conditional independence model* $\mathcal{M}_{\mathrm{CI}}$ for a n-dimensional random vector Y for which all conditional independence statements are true is a special subset of all possible Gaussian densities $f_{\mu,\Sigma}$:

$$\mathcal{M}_{\mathrm{CI}} \subseteq \{f_{\mu,\Sigma} \mid \mu \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n\times n}_{\mathrm{psd}}\}.$$

By Lemma 1, the parameter space of $\mathcal{M}_{\mathrm{CI}}$ is equal to the algebraic set

$$\begin{aligned}\mathcal{A}_{\mathrm{CI}} = \{\mu \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n\times n}_{\mathrm{psd}} \mid g(\Sigma) = 0 \text{ for all polynomials } g \text{ which are}\\ (\#C_i + 1) \times (\#C_i + 1) \text{ minors of } \Sigma_{A_i\cup C_i, B_i\cup C_i}, i \in [r]\}.\end{aligned} \tag{2}$$

Thus, every Gaussian conditional independence model is the image of a bijective parametrization map $(\mu, \Sigma) \mapsto f_{\mu,\Sigma}$ whose domain is given by equation (2).

Example 1 *Let Y_1, Y_2 and Y_3 be jointly Gaussian and suppose $Y_3 \perp\!\!\!\perp Y_1 \mid Y_2$. Then by Lemma 1, the 2×2 minors of the submatrix*

$$\Sigma_{\{2,3\},\{1,2\}} = \begin{pmatrix} \sigma_{21} & \sigma_{22} \\ \sigma_{31} & \sigma_{32} \end{pmatrix}$$

need to vanish. Here the only vanishing minor simply corresponds to the determinant. So $g = \sigma_{21}\sigma_{32} - \sigma_{31}\sigma_{22}$ is a polynomial which is zero in equation (2).

Example 2 *For a Gaussian random vector $Y = (Y_i)_{i\in[4]}$ together with the conditional independence $Y_2 \perp\!\!\!\perp \{Y_1, Y_3\} \mid Y_4$, the 2×2 minors of the submatrix*

$$\Sigma_{\{2,4\},\{1,3,4\}} = \begin{pmatrix} \sigma_{21} & \sigma_{23} & \sigma_{24} \\ \sigma_{41} & \sigma_{43} & \sigma_{44} \end{pmatrix}$$

need to vanish. Explicitly, $\sigma_{21}\sigma_{43} - \sigma_{41}\sigma_{23} = 0$, $\sigma_{21}\sigma_{44} - \sigma_{41}\sigma_{24} = 0$ and $\sigma_{23}\sigma_{44} - \sigma_{43}\sigma_{24} = 0$.

The following two sections review some basic results on directed and undirected graphical models. In particular, here we recall a second duality: the one between graphs and conditional independence relationships. In conjunction with Lemma 1, these form the basis for algebraic sensitivity methods which ensure that after covariation a graph remains a faithful representation of the model.

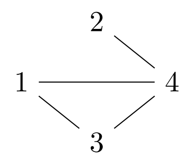

Figure 1: An undirected graph for the conditional independence model $Y_2 \perp\!\!\!\perp \{Y_1, Y_3\} \mid Y_4$ in Examples 2 and 3.

2.2. Undirected Gaussian Graphical Models

For Gaussian random vectors, many sets of conditional independences can be represented visually by a graph. We start by defining families of Gaussians supported by undirected graphs.

Definition 2 *A* Gaussian undirected graphical model *for a random vector $Y = (Y_i)_{i\in[n]}$ is defined by an undirected graph $\mathcal{G} = (V, E)$ with vertex set $V = [n]$ and a family of densities $f_{\mu,\Sigma}$ whose covariance matrix Σ is such that $(\Sigma^{-1})_{ij} = 0$ if and only if $(i, j) \notin E$.*

Thus the zero entries in the inverse of the covariance matrix of a Gaussian undirected graphical model correspond to conditional independence statements. This is usually called the *pairwise Markov* property (Lauritzen, 1996). In particular, if $(\Sigma^{-1})_{ij} = 0$ then $Y_i \perp\!\!\!\perp Y_j \mid Y_{[n]\setminus\{i,j\}}$: so the absence of an edge between two random variables implies that these are conditionally independent given all the others.

By Lemma 1 and Definition 2, the fact that an entry in the inverse of the covariance matrix is equal to zero exactly corresponds to the vanishing of the minors of an appropriate submatrix of Σ.

Example 3 (Example 2 continued) *The statement $Y_2 \perp\!\!\!\perp \{Y_1, Y_3\} \mid Y_4$ can be represented by the undirected graph in Figure 1 where the edges $(1, 2)$ and $(2, 3)$ are not present.*

2.3. Gaussian Bayesian Networks

For directed graphical models, conditional independence relationships cannot be explicitly represented by zeros in the inverse of the covariance matrix. GBNs can however be constructed from the definition of a conditional univariate Gaussian distribution at each of its vertices (Richardson and Spirtes, 2002).

Definition 3 *A* GBN *for a random vector $Y = (Y_i)_{i\in[n]}$ is defined by a directed acyclic graph $\mathcal{G} = (V, E)$ with vertex set $V = [n]$ such that to each $i \in [n]$ is associated a conditional Gaussian density f_{m_i,v_i} with mean $m_i = \beta_{0i} + \sum_{j\in \mathrm{pa}(i)} \beta_{ji} y_j$ and variance $v_i \in \mathbb{R}_+$. Here, $\mathrm{pa}(i) \subseteq [i-1]$ denotes the parent set of the vertex i in $\mathcal{G}$, $\beta_{0i} \in \mathbb{R}$ and $\beta_{ji} \in \mathbb{R}$ for all $j \in \mathrm{pa}(i)$.*

The Gaussian densities f_{m_i,v_i} in a GBN are associated to conditional independence statements of the form $Y_i \perp\!\!\!\perp Y_{[i-1]\setminus \mathrm{pa}(i)} \mid Y_{\mathrm{pa}(i)}$. Definition 3 then assigns a multivariate

$$1 \longrightarrow 2 \longrightarrow 3$$

Figure 2: A directed acyclic graph for the conditional independence model $Y_3 \perp\!\!\!\perp Y_1 \mid Y_2$ in Examples 1 and 4.

Gaussian distribution to the full vector Y as follows. Let $\beta_0 = (\beta_{0i})_{i\in[n]}$, B be the strictly upper triangular matrix with entries $B_{ji} = \beta_{ji}$ if $j \in \text{pa}(i)$ and zero otherwise, and $V = \text{diag}(v_1, \dots, v_n)$ be the diagonal matrix of the conditional variances. Then Y has Gaussian density $f_{\mu,\Sigma}$ with mean $\mu = (I-B)^{-\top}\beta_0$ and covariance matrix $\Sigma = (I-B)^{-\top}V(I-B)^{-1}$ where I denotes the identity. A matrix Σ constructed in this way naturally respects the conditional independences associated to a directed acyclic graph $\mathcal{G}$. In general, whether or not a covariance matrix can be associated to a specific directed acyclic graph can be checked by the vanishing of the minors of appropriate submatrices of Σ, as formalized in Lemma 1.

Example 4 (Example 1 continued) *The conditional independence model* $Y_3 \perp\!\!\!\perp Y_1 \mid Y_2$ *on three Gaussian random variables can be associated to the directed acyclic graph reported in Figure 2 where the directed edge* $(1,3)$ *is not present.*

3. Sensitivity Analysis

We now briefly review standard sensitivity methods for GBNs before we introduce our new model-preserving formalism.

3.1. Standard Methods for Gaussian Bayesian Networks

Sensitivity methods for GBNs have been extensively studied (Gómez-Villegas et al., 2007, 2008, 2013). For a generic Gaussian random vector Y with density $f_{\mu,\Sigma}$, robustness is usually studied by perturbing the mean vector μ and the covariance matrix Σ. Such a perturbation is carried out by defining a perturbation vector $d \in \mathbb{R}^n$ and a matrix $D \in \mathbb{R}^{n\times n}$ which act additively on the original mean and variance, giving rise to a vector $\tilde{Y}$ with a new distribution$f_{\mu+d,\Sigma+D}$. The dissimilarity between these two vectors is then usually quantified via the KL divergence.

For any two n-dimensional Gaussian vectors Y and Y' with distributions $f_{\mu,\Sigma}$ and $f_{\mu',\Sigma'}$ respectively, the *KL divergence* between Y and Y' is given by

$$\text{KL}(Y'||Y) = \frac{1}{2}\left(\text{tr}(\Sigma^{-1}\Sigma') + (\mu-\mu')^\top\Sigma^{-1}(\mu-\mu') - n + \ln\left(\frac{\det(\Sigma)}{\det(\Sigma')}\right)\right). \quad (3)$$

The KL divergence is not a distance and in particular violates the symmetry requirement, so in general $\text{KL}(Y'||Y) \neq \text{KL}(Y||Y')$. Symmetric extensions of KL divergences have been recently considered in sensitivity studies for GBNs (Zhu et al., 2017) but a comprehensive review of these goes beyond the scope of this paper.

From equation (3), the KL divergence between a perturbation $\tilde{Y}$ and the original Y is

$$\text{KL}(\tilde{Y}||Y) = \frac{1}{2}\left(\text{tr}(\Sigma^{-1}D) + d^\top\Sigma^{-1}d + \ln\left(\frac{\det(\Sigma)}{\det(\Sigma+D)}\right)\right).$$

In sensitivity analyses, the Gaussian vector Y is often partitioned into two subvectors Y_E and Y_O such that $E \cup O = [n]$, including the evidential and output variables, respectively. Evidential variables are those for which a value y_E is observed, whilst the output variables are those of interest to the user. Then also the perturbation mean vector and covariance matrix can be partitioned as

$$d = \begin{pmatrix} d_O \\ d_E \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} D_{O,O} & D_{O,E} \\ D_{E,O} & D_{E,E} \end{pmatrix}$$

where $D_{O,E} = D_{E,O}^\top$. This formalism enables the user to study the dissimilarity between the Gaussian vector $Y_O|Y_E = y_E$ and perturbed output variables $\tilde{Y}_O|Y_E = y_E$ with distributions $f_{\mu^{O|E},\Sigma^{O|E}}$ and $f_{\tilde{\mu}^{O|E},\tilde{\Sigma}^{O|E}}$, respectively, where $\mu^{O|E}$ and $\Sigma^{O|E}$ are as in equation (1) and

$$\begin{aligned}\tilde{\mu}^{O|E} &= \mu_O + d_O + (\Sigma_{O,E} + D_{O,E})(\Sigma_{E,E} + D_{E,E})^{-1}(y_E - \mu_E - d_E),\\ \tilde{\Sigma}^{O|E} &= \Sigma_{O,O} + D_{O,O} - (\Sigma_{O,E} + D_{O,E})(\Sigma_{E,E} + D_{E,E})^{-1}(\Sigma_{E,O} + D_{E,O}).\end{aligned}$$

Efficient algorithms to propagate evidence $Y_E = y_E$ and to speedily compute these conditional distributions are available for GBNs (Castillo and Kjærulff, 2003; Malioutov et al., 2006). The form of the KL divergence between these two conditional distributions depends on the block of parameters that are perturbed (see Gómez-Villegas et al., 2013, for more details). Although in the following we do not consider distributions updated via evidence propagation, notice that our approach would equally apply to the updated probabilities $Y_O|Y_E = y_E$.

This standard approach has the critical drawback that if the Gaussian distribution is associated to a specific conditional independence model, for instance represented by a directed graph, then a perturbation may break its conditional independences. We illustrate this point below.

Example 5 (Example 4 continued) *Suppose in the GBN of Example 4 that the covariance matrix Σ is perturbed by a 3×3 matrix D of all zeros except for a $d \in \mathbb{R}$ in some of the positions of the minor that vanishes such that $\Sigma + D \in \mathbb{R}^{3\times 3}_{\text{psd}}$. The directed graph in Figure 2 is a faithful representation of this new Gaussian distribution if and only if the 2×2 corresponding minor $\sigma_{21}\sigma_{32} - \sigma_{31}\sigma_{22}$ is still equal to zero (this is the only vanishing minor of the model). But this is the case if and only if $d = 0$: so if there is no perturbation.*

If alternatively the only non-zero entry of D were in some of the positions outside the minor that vanishes, such that $\Sigma + D \in \mathbb{R}^{3\times 3}_{psd}$, then no matter what the value of $d \in \mathbb{R}$ the representation in Figure 2 would be a faithful description of the underlying conditional independence structure, obviously.

A possible approach to overcome the breaking of conditional independences in the case of GBNs is to vary the parameters of the univariate conditional Gaussian distributions in Definition 3. The perturbation of the matrix Φ of conditional variances can then affect the covariance matrix of the overall Gaussian distribution (see Section 6 of Gómez-Villegas et al., 2013, for an example). Another possibility is to perturb the matrix B including the regression parameters and observe the effects of this perturbation on both the overall mean μ and covariance Σ. This second approach has been used to quantify the effect of adding or deleting edges in GBNs (Gómez-Villegas et al., 2011). However and critically, both these approaches lose the intuitiveness of acting directly on the unconditional mean and covariance of the Gaussian distribution.

3.2. Model-Preserving Sensitivity Analysis

To overcome the difficulties arising in classical sensitivity analyses, we now introduce a novel approach which extends sensitivity methods usually applied exclusively to GBNs to more general Gaussian conditional independence models, including undirected Gaussian graphical models. In particular, we establish specific conditions under which a perturbed covariance matrix is within the original algebraic parameter set of the model at hand, so that all conditional independence relationships of the original model continue to be valid. We show below that this can be easily achieved by considering covariation schemes which act multiplicatively rather than additively on these matrices.

Henceforth, we think of a Gaussian model $\mathcal{M}_{\text{CI}}$ for a random vector $Y = (Y_i)_{i\in[n]}$ together with conditional independence assumptions $\text{CI} = \{A_k \perp\!\!\!\perp B_k \mid C_k \text{ for } k \in [r]\}$ as being represented by a collection of vanishing minors of its covariance matrix $\Sigma \in \mathbb{R}^{n\times n}_{\text{psd}}$ as introduced in Lemma 1. Because this rationale concerns only the covariance matrix, we can assume without loss that Y has zero mean, $\mu = 0_n$. For ease of notation, we thus write f_Σ rather than $f_{0_n,\Sigma}$ for its associated density.

We denote by the circle $\circ$ the *Schur product* of two matrices Δ and Σ of the same dimension, so $\Delta \circ \Sigma = (\delta_{ij}\sigma_{ij})_{i,j\in[n]}$ is the componentwise product of their entries. Let

$$\Phi_\Delta : \Sigma \mapsto \Delta \circ \Sigma$$

denote the map which sends a covariance matrix to its Schur product with a matrix Δ. We call the map Φ_Δ *model-preserving* if under this operation the algebraic parameter set in equation (2) is mapped onto itself, $\Phi_\Delta(\mathcal{A}_{\text{CI}}) \subseteq \mathcal{A}_{\text{CI}}$.

In the following sections, we always decompose the perturbation of a covariance matrix Σ into two steps, and hence two Schur products, as follows. In the first step, the original covariance Σ is mapped to its Schur product with a symmetric *variation* matrix $\Delta \in \mathbb{R}^{n\times n}_{\neq 0}$. Hereby, usually only some of the original covariances σ_{ij} are assigned a new value $\sigma_{ij} \mapsto \delta_{ij}\sigma_{ij}$ at selected positions (i, j) while the remaining parameters are untouched. This is achieved by having all non-(i, j) entries of Δ equal to one. In demanding that all entries δ_{ij} are non-zero, we automatically avoid setting a non-zero covariance $\sigma_{ij} \neq 0$ to zero via multiplication by an entry of Δ. This type of perturbation would force the corresponding variables to be independent, $X_i \perp\!\!\!\perp X_j$, in the perturbed model, which would clearly violate the assumptions in the original model $\mathcal{M}_{\text{CI}}$. The second Schur product is then calculated between $\Delta \circ \Sigma$ and a symmetric *covariation* matrix $\tilde{\Delta} \in \mathbb{R}^{n\times n}_{\neq 0}$. This matrix $\tilde{\Delta}$ has ones in the positions (i, j) selected previously, whilst the values of the remaining entries need to be set according to some agreed procedure which ensures that for every vanishing minor of Σ, the appropriate minor of $\tilde{\Delta} \circ \Delta \circ \Sigma$ vanishes as well. In this process, in order to guarantee symmetry, whenever an entry (i, j) is changed in one of the matrices, its corresponding entry (j, i) needs to be changed in the exact same fashion. Explicitly, the composition of Schur products is of the following form:

$$\tilde{\Delta} \circ \Delta \circ \Sigma = \begin{pmatrix} \star & \cdots & \cdots & \star \\ \vdots & \ddots & 1 & \vdots \\ \vdots & 1 & \ddots & \vdots \\ \star & \cdots & \cdots & \star \end{pmatrix} \circ \begin{pmatrix} 1 & \cdots & \cdots & 1 \\ \vdots & \ddots & \delta_{ij} & \vdots \\ \vdots & \delta_{ji} & \ddots & \vdots \\ 1 & \cdots & \cdots & 1 \end{pmatrix} \circ \begin{pmatrix} \sigma_{11} & \cdots & \cdots & \sigma_{1n} \\ \vdots & \ddots & \sigma_{ij} & \vdots \\ \vdots & \sigma_{ji} & \ddots & \vdots \\ \sigma_{n1} & \cdots & \cdots & \sigma_{nn} \end{pmatrix}$$

Here, the stars indicate entries in $\tilde{\Delta}$ which need to be specified. Thus, for a given covariance matrix $\Sigma \in \mathcal{A}_{\mathrm{CI}}$ and a given variation $\Delta \in \mathbb{R}^{n\times n}_{\neq 0}$ of that matrix, we develop methods to find some covariation matrices $\tilde{\Delta}$ such that $\tilde{\Delta} \circ \Delta \circ \Sigma \in \mathcal{A}_{\mathrm{CI}}$. Then the map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving.

Example 6 (Example 5 continued) *Suppose in the Gaussian model $Y_3 \perp\!\!\!\perp Y_1 \mid Y_2$ we perform a perturbation to the parameter σ_{21} of the covariance matrix Σ. Then, under our formalism, the matrix Δ is defined as*

$$\Delta = \begin{pmatrix} 1 & \delta & 1 \\ \delta & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

and the only vanishing minor of $\Delta \circ \Sigma$ takes the form $\delta\sigma_{12}\sigma_{32} - \sigma_{31}\sigma_{22}$. This polynomial is equal to zero in either of three cases: when σ_{22} is covaried by δ; when σ_{31} and σ_{13} are covaried by δ; or when σ_{22}, σ_{31}, σ_{13}, σ_{32} and σ_{23} are covaried by δ. The associated covariation matrices $\tilde{\Delta}$ should equal, respectively,

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & \delta & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & \delta \\ 1 & 1 & 1 \\ \delta & 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & \delta \\ 1 & \delta & \delta \\ \delta & \delta & 1 \end{pmatrix}. \tag{4}$$

For each of these choices of $\tilde{\Delta}$, we have that $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving.

The structure associated to these perturbations is much clearer if we only consider the submatrix $\tilde{\Delta}_{\{2,3\},\{1,2\}} \circ \Delta_{\{2,3\},\{1,2\}}$, for each of the covariation matrices, whose determinant is the relevant minor in Lemma 1. For the three cases above, this submatrix is equal to, respectively,

$$\begin{pmatrix} \delta & \delta \\ 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} \delta & 1 \\ \delta & 1 \end{pmatrix}, \qquad \begin{pmatrix} \delta & \delta \\ \delta & \delta \end{pmatrix}. \tag{5}$$

So here the perturbation is applied either to a full row, a full column or the full matrix. We demonstrate below that this feature is in general associated to model-preservation.

The formalism we set up in this section enables us to interpret a model-preserving map as a homomorphism between polynomial rings in the indeterminates given by entries of the covariance and variation/covariation matrices. This observation together with Lemma 1 enables us to employ the powerful language of real algebraic geometry to study Gaussian conditional independence models. Over the next few sections, we make a first important step in using these notions for sensitivity analyses.

4. One-Way Model-Preserving Sensitivity Analysis

Throughout this section, we study single-parameter variations. Here, a user identifies precisely one entry σ_{ij} of a covariance matrix at a fixed position (i,j) that she intends to adjust to $\delta \cdot \sigma_{ij}$ for some $\delta \neq 0, 1$. The corresponding variation matrix $\Delta = (\delta_{ij})_{i,j} \in \mathbb{R}^{n\times n}_{\neq 0}$ is a symmetric matrix which has either one entry $\delta_{ij} = \delta$, if $i = j$, or two entries $\delta_{ij}, \delta_{ji} = \delta$, if $i \neq j$, not equal to one and entries $\delta_{kl} = 1$ for all $(k,l) \neq (i,j)$. We assume the user believes the conditional independence relationships of the model to be valid and that these should remain valid after the perturbation. Before setting up an appropriate covariation scheme for this setting, we fix some notation. This will be used in the two cases of our analysis presented below: first for models defined by a single conditional independence relationship and then for models defined by a collection of conditional independence statements.

4.1. Covariation Matrices

For any symmetric matrix $D \in \mathbb{R}^{n\times n}$ and two index sets $A, B \subseteq [n]$, we henceforth denote with $\lfloor D_{A,B} \rfloor^1$ the symmetric, full dimension $n \times n$ matrix where:

- all positions indexed by A and B are equal to the corresponding entries in D;
- entries not indexed by A and B are set to ensure symmetry;
- all other entries are equal to one.

We also let $\mathbb{1}_{A,B}$ be the matrix with all entries equal to one and with rows indexed by A and columns indexed by B.

Example 7 *Let $D \in \mathbb{R}^{3\times 3}$ and suppose*

$$D_{\{1,2\},\{2,3\}} = \begin{pmatrix} 1 & \delta \\ 1 & \delta \end{pmatrix}.$$

Then

$$\lfloor D_{\{1,2\},\{2,3\}} \rfloor^1 = \begin{pmatrix} 1 & 1 & \delta \\ 1 & 1 & \delta \\ \delta & \delta & 1 \end{pmatrix}.$$

We now define different types of covariation matrices which are motivated by those studied in Example 6.

Definition 4 *For a single-parameter variation matrix $\Delta \in \mathbb{R}^{n\times n}$ with $\delta_{ij} = \delta_{ji} = \delta$, we say that the covariation matrix $\tilde{\Delta}$ is*

- total *if $\tilde{\Delta} \circ \Delta = \delta \mathbb{1}_{[n],[n]}$;*
- partial *if $\tilde{\Delta} \circ \Delta = \lfloor \delta \mathbb{1}_{A\cup C, B\cup C} \rfloor^1$.*
- row-based *if $\tilde{\Delta} \circ \Delta = \lfloor \delta \mathbb{1}_{E, B\cup C} \rfloor^1$ for a subset $E \subseteq A \cup C \subseteq [n]$;*
- column-based *if $\tilde{\Delta} \circ \Delta = \lfloor \delta \mathbb{1}_{A\cup C, F} \rfloor^1$ for a subset $F \subseteq B \cup C \subseteq [n]$.*

In words, the Schur product of a variation with a total covariation matrix is a matrix filled with δ, and the Schur product of a variation with a partial covariation matrix is a matrix which only has a symmetric sub-block filled with δ and entries equal to one otherwise. Row-based and column-based covariation matrices result in Schur products which have δ entries only in some specific subsets of the rows and columns. An illustration of row-based, column-based and partial covariation matrices was given in equations (4) and (5) for the setting of Example 6. For the same example, a total covariation matrix would take the form:

$$\begin{pmatrix} \delta & 1 & \delta \\ 1 & \delta & \delta \\ \delta & \delta & \delta \end{pmatrix}.$$

By construction total, partial, row- and column-based covariations ensure symmetry. Henceforth, we assume that $\tilde{\Delta} \circ \Delta \circ \Sigma \in \mathbb{R}^{n \times n}_{\text{psd}}$. We discuss in Section 7 methods to assess the positive semidefiniteness of the covariance matrix after a model-preserving perturbation.

4.2. One Conditional Independence Statement

We first consider the case where a Gaussian conditional independence model $\mathcal{M}_{\text{CI}}$ is specified by a single relationship $\text{CI} = \{A \perp\!\!\!\perp B \mid C\}$ for some index sets $A, B, C \subset [n]$. Throughout, $\Sigma \in \mathcal{A}_{\text{CI}}$ is a covariance matrix in this model and Δ is a single-parameter variation matrix with non-one entry $\delta_{ij} = \delta_{ji} = \delta$ at a fixed position (i, j) and (j, i). We can now specify covariation matrices for this setup which result in model-preserving perturbations.

Proposition 5 *If $(i, j), (j, i) \notin (A \cup C, B \cup C)$ then the map $\Phi_{\tilde{\Delta} \circ \Delta}$ is model-preserving for a covariation $\tilde{\Delta} = \mathbb{1}_{[n],[n]}$.*

This result is a straightforward consequence of Lemma 1. Indeed, if both (i, j) and (j, i) are not entries of the submatrix $\Sigma_{A \cup C, B \cup C}$ whose vanishing minors specify the model, then no changes induced by multiplication with δ appear in the vanishing polynomials. This can be illustrated in Example 5 for the variation of the variance σ_{11}. Proposition 5 thus formalizes the cases when a perturbation has no effect on the underlying conditional independence structure and no covariation matrix is needed.

Proposition 6 *If $C = \emptyset$ then the map $\Phi_{\tilde{\Delta} \circ \Delta}$ is model-preserving for $\tilde{\Delta} = \mathbb{1}_{[n],[n]}$.*

This result easily follows by noting that standard independence statements $A \perp\!\!\!\perp B$ correspond to zeros in the covariance matrix. Multiplication of such zeros by δ still returns zeros which automatically results in a model-preserving map. Thus, if the model consists of one standard independence statement, any perturbation is model-preserving. For standard sensitivity methods which act additively on the covariance matrix, such a property does not in general hold unless one perturbs with the same value all the elements in the appropriate submatrix associated to the vanishing minor.

We next focus on the case where a perturbation makes some of the original vanishing polynomials non-equal to zero. Henceforth, unless otherwise stated, we thus assume that either (i, j) or (j, i) are an entry of $\Sigma_{A \cup C, B \cup C}$.

Theorem 7 *The map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving for total and partial covariation matrices $\tilde{\Delta}$.*

A proof of this result is given in the appendix.

Observe that by default, we need to enforce that $\delta > 0$ for total covariation matrices. This is because otherwise the entries of the diagonal of Σ, the variances of the model, become negative. For partial covariations this constraint may not have to be enforced, even though in reality it is controversial to investigate the effect of changing the sign of an entry in a covariance matrix. Furthermore, there has been a growing interest on covariance matrices with the property that all their entries are positive (Fallat et al., 2017; Slawski and Hein, 2015).

As a consequence of Proposition 5, perturbations by δ outside of the submatrix $\Sigma_{A\cup C,B\cup C}$ in total covariation matrices have no effect on the vanishing polynomials. Henceforth we thus consider only the submatrix $\tilde{\Delta}_{A\cup C,B\cup C}$ and identify the entries that need to have a δ so that the map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving. Intuitively, this approach fits a user who may want to change the least possible number of entries of a covariance matrix after a perturbation. In this direction, in Section 6 we demonstrate that if two covariation matrices are nested, meaning that the entries which need to covary for one matrix are a subset of those of the other, then the perturbed distribution associated to the simpler covariation matrix is closer to the original distribution.

Theorem 8 *The map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving in the following cases:*

- *if (i,j) or $(j,i) \in (A,B)$ for a row-based covariation $\tilde{\Delta}$ whenever $i \in E \subseteq A$, and for a column-based covariation $\tilde{\Delta}$ whenever $j \in F \subseteq B$;*
- *if (i,j) or $(j,i) \in (A,C)$ for a row-based covariation $\tilde{\Delta}$ whenever $i \in E \subseteq A$, and for a column-based covariation $\tilde{\Delta}$ whenever $F = C$;*
- *if (i,j) or $(j,i) \in (C,B)$ for a row-based covariation $\tilde{\Delta}$ whenever $E = C$, and for a column-based covariation $\tilde{\Delta}$ whenever $i \in F \subseteq B$;*
- *if (i,j) and $(j,i) \in (C,C)$ for a row-based covariation $\tilde{\Delta}$ whenever $E = C$, and for a column-based covariation $\tilde{\Delta}$ whenever $F = C$.*

A proof of this result is given in the appendix.

In words, whenever the perturbed entry (i,j) or (j,i) is not an element of the conditioning set (C,C), a row- or a column-based covariation consisting of one row or column only can give a model-preserving map $\Phi_{\tilde{\Delta}\circ\Delta}$. Conversely, if the entry $(i,j) \in (C,C)$ is in the conditioning set then row and column-based covariation matrices have δ entries over all rows and columns in C, respectively. This is because if for instance one row of $\tilde{\Delta}_{C,C}$ has δ elements then other entries of $\tilde{\Delta}_{C,C}$ need to be equal to δ to ensure symmetry. However this then implies that other full rows of $\tilde{\Delta}_{C,C}$ need to have all δ entries.

Example 8 *To illustrate Theorem 8, consider a Gaussian random vector $Y = (Y_i)_{i\in[4]}$ together with the conditional independence statement $Y_1 \perp\!\!\!\perp Y_4 \mid \{Y_2, Y_3\}$. In our notation $A = \{1\}$, $B = \{4\}$ and $C = \{2,3\}$. Notice that this model could be depicted by a BN* *with an edge (Y_i, Y_j) for all $i < j$, $i, j \in [4]$, except for the edge (Y_1, Y_4). By Lemma 1 the determinant of*

$$\Sigma_{A\cup C,B\cup C} = \begin{pmatrix} \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{32} & \sigma_{33} & \sigma_{34} \end{pmatrix}$$

must vanish. Theorem 8 specifies that:

- *from the first bullet, for a variation of σ_{14} or σ_{41}, then a row-based covariation of the first row of $\Sigma_{A\cup C,B\cup C}$ or a column-based covariation of the third column of $\Sigma_{A\cup C,B\cup C}$ gives a model-preserving map;*
- *from the second bullet, for a variation of σ_{12}, σ_{21}, σ_{13} or σ_{31}, then a row-based covariation of the first row of $\Sigma_{A\cup C,B\cup C}$ or a column-based covariation of the first and second columns of $\Sigma_{A\cup C,B\cup C}$ gives a model-preserving map;*
- *from the third bullet, for a variation of σ_{24}, σ_{42}, σ_{34} or σ_{43}, then a row-based covariation of the second or third rows of $\Sigma_{A\cup C,B\cup C}$ or a column-based covariation of the third column of $\Sigma_{A\cup C,B\cup C}$ gives a model-preserving map;*
- *from the fourth bullet, for a variation of σ_{22}, σ_{23}, σ_{32} or σ_{33}, then a row-based covariation of the second or third rows of $\Sigma_{A\cup C,B\cup C}$ or a column-based covariation of the first and second columns of $\Sigma_{A\cup C,B\cup C}$ gives a model-preserving map.*

Therefore for the first three cases there is at least one row- or column-based covariation consisting of simply one row or column.

4.3. Multiple Conditional Independence Statements

We now generalize the results of the previous section by considering models which are defined by a collection of multiple conditional independence relationships. Using the notation introduced in Section 2, in the following let $\text{CI} = \{A_1 \perp\!\!\!\perp B_1 \mid C_1, \ldots, A_r \perp\!\!\!\perp B_r \mid C_r\}$. Let also always $A = \cup_{k\in[r]} A_k$, $B = \cup_{k\in[r]} B_k$ and $C = \cup_{k\in[r]} C_k$.

First we introduce a result which simplifies the task of checking whether a covariation is model-preserving or not. Suppose hereby without loss of generality that the conditional independence relationships defining the model are ordered such that for all for $k \in [t]$ we have proper conditional independence statements $A_k \perp\!\!\!\perp B_k \mid C_k$ where $C_k \neq \emptyset$, whilst for all $l \in [r] \setminus [t]$ the conditioning set is empty, $C_l = \emptyset$, for some index for $t \leq r$. We then denote by $\text{CI}^* = \{A_1 \perp\!\!\!\perp B_1 \mid C_1, \ldots, A_t \perp\!\!\!\perp B_t \mid C_t\} \subseteq \text{CI}$ the set of statements with non-empty conditioning set.

Proposition 9 *If the map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving for $\mathcal{M}_{\text{CI}^*}$ then it is also model-preserving for $\mathcal{M}_{\text{CI}}$.*

This result is a straightforward consequence of Proposition 6, since zero entries in the covariance matrix are not affected by our model-preserving covariation. It is extremely useful since it allows us to check whether a map is model-preserving by using only a subset of all conditional independences of a Gaussian model. Henceforth, we can thus without loss assume that the set CI is such that $C_i \neq \emptyset$ for all $i \in [r]$.

The following example shows that in general it is not simply sufficient to create a $\tilde{\Delta}$ matrix for each conditional independence statement independently.

Example 9 *Consider a model for $Y = (Y_i)_{i\in[5]}$ defined by $Y_4 \perp\!\!\!\perp Y_{\{1,2\}} \mid Y_3$ and $Y_{\{2,4\}} \perp\!\!\!\perp Y_5|Y_3$. The submatrices associated to these independence statements are, respectively,*

$$\begin{pmatrix} \sigma_{31} & \sigma_{32} & \sigma_{33} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} \sigma_{23} & \sigma_{25} \\ \sigma_{33} & \sigma_{35} \\ \sigma_{43} & \sigma_{45} \end{pmatrix}.$$

Suppose the entry σ_{43} of Σ is varied by δ and that for both conditional independences the matrices $\tilde{\Delta}$ are column-based and consisting of one column only. If we compute the $\circ$ product between Δ and the two column-based $\tilde{\Delta}$ we have

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & \delta & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & \delta & 1 & 1 \\ 1 & \delta & \delta & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & \delta & 1 \\ 1 & 1 & \delta & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & \delta & 1 & 1 \\ 1 & \delta & \delta^2 & \delta & 1 \\ 1 & 1 & \delta & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

When the above matrix is then multiplied with Σ we have that, for instance, the minor $\delta\sigma_{31}\sigma_{43} - \delta^2\sigma_{41}\sigma_{33} \neq 0$ does not vanish and thus the resulting map is not model-preserving.

The problem here is that the entry σ_{33} appears in both submatrices whose minors need to vanish. Therefore if the submatrices $\Sigma_{A\cup C,B\cup C}$ associated to two different conditional independence statements do not have entries in common, then one could define a covariation matrix for each of them and then simply compute their Schur product to achieve a model-preserving map. We formally define this property in the appendix.

Consider a generic Gaussian conditional independence model. In direct analogy to Proposition 5 and Theorem 7, we find the following.

Proposition 10 *The map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving for*

- $\tilde{\Delta} = \mathbb{1}_{[n],[n]}$ *if* $(i,j),(j,i) \notin (A \cup C, B \cup C)$*;*
- $\tilde{\Delta} = \mathbb{1}_{[n],[n]}$ *if* $(i,j),(j,i) \notin (A_k \cup C_k, B_k \cup C_k)$ *for all* $k \in [r]$*;*
- *total and partial covariation matrices* $\tilde{\Delta}$*.*

The first two statements easily follow by noting that no changes appear in any of the vanishing polynomials and therefore no covariation is necessary. The third point is a straightforward consequence of Theorem 7.

Next we again look for covariation matrices which include a smaller number of elements than total and partial covariation matrices. This generalizes the concept of row-based and column-based covariations from models defined by single conditional independences to models defined by multiple relationships. Following the results of Section 4.2, it is reasonable to consider simplifications of partial covariation matrices where some of the rows/columns have entries equal to one. Thus we study covariation for the submatrix $\Sigma_{A\cup C,B\cup C}$. The following example illustrates some of the difficulties we might encounter.

Example 10 (Example 9 continued) *For the Gaussian model defined by* $Y_4 \perp\!\!\!\perp Y_{\{1,2\}} \mid Y_3$ *and* $Y_{\{2,4\}} \perp\!\!\!\perp Y_5 \mid Y_3$ *where we varied the covariance* σ_{43}*, we need to consider the submatrix*

$$\Sigma_{\{2,3,4\},\{1,2,3,5\}} = \begin{pmatrix} \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{45} \end{pmatrix}.$$

Simple row-based and column-based covariations are associated to the matrices $\tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}}$ *corresponding to*

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \delta & \delta & \delta & \delta \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} 1 & 1 & \delta & 1 \\ 1 & 1 & \delta & 1 \\ 1 & 1 & \delta & 1 \end{pmatrix}, \tag{6}$$

respectively, since σ_{43} *is the entry in position* $(3,3)$ *of* $\Sigma_{\{2,3,4\},\{1,2,3,5\}}$.

For the row-based covariation on the left of equation (6) we have

$$\tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}} = (\lfloor \tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}} \rfloor^1)_{\{2,3,4\},\{1,2,3,5\}}$$

since σ_{14}*,* σ_{24}*,* σ_{34} *and* σ_{54} *(those covariances that must be equal to entries in the last row of* $\Sigma_{\{2,3,4\},\{1,2,3,5\}}$ *to ensure symmetry) are not in* $\Sigma_{\{2,3,4\},\{1,2,3,5\}}$*. This means that no entries of* $(\tilde{\Delta} \circ \Delta)_{\{2,3,4\},\{1,2,3,5\}}$ *are altered during the creation of the* 5×5 *matrix* $\lfloor \tilde{\Delta} \circ \Delta \rfloor^1$*. It is straightforward to check that this covariation matrix gives rise to a model-preserving map. Conversely, consider the column-based covariation on the right of equation (6). In this case,* $\tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}} \neq (\lfloor \tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}} \rfloor^1)_{\{2,3,4\},\{1,2,3,5\}}$ *because*

$$(\lfloor \tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}} \rfloor^1)_{\{2,3,4\},\{1,2,3,5\}} = \begin{pmatrix} 1 & 1 & \delta & 1 \\ 1 & \delta & \delta & 1 \\ 1 & 1 & \delta & 1 \end{pmatrix}. \tag{7}$$

The map based on such a covariation is not model-preserving. This is because covariation matrices need to be filled with full-row or full-columns of δ*s in order to preserve a model's structure, as demonstrated in Theorem 8. To see this consider the independence statement* $Y_4 \perp\!\!\!\perp Y_{\{1,2\}} \mid Y_3$*. All minors of size 2 of the matrix*

$$\begin{pmatrix} \sigma_{31} & \sigma_{32} & \sigma_{33} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} \end{pmatrix}$$

must vanish. However, notice that the column-based covariation multiplies by δ *the entries* σ_{32}*,* σ_{33} *and* σ_{43}*, thus making some minors different from zero.*

We can fix this issue by simply filling up the second column of the matrix in equation (7) with δ *entries. Indeed,*

$$\tilde{\Delta}_{\{2,3,4\},\{1,2,3,5\}} \circ \Delta_{\{2,3,4\},\{1,2,3,5\}} = \begin{pmatrix} 1 & \delta & \delta & 1 \\ 1 & \delta & \delta & 1 \\ 1 & \delta & \delta & 1 \end{pmatrix}.$$

gives a model-preserving map because σ_{24} *(the only covariance that must be equal to entries in the second column of* $\Sigma_{\{2,3,4\},\{1,2,3,5\}}$ *to ensure symmetry) is not an entry of* $\Sigma_{\{2,3,4\},\{1,2,3,5\}}$.

The above example demonstrated that again row-based and column-based covariations can be associated to model-preserving maps. We can thus generalize Theorem 8 to the following result.

Proposition 11 *The map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving for a row-based or a column-based covariation matrix $\tilde{\Delta}$ if*

$$\tilde{\Delta}_{A\cup C,B\cup C} \circ \Delta_{A\cup C,B\cup C} = (\lfloor \tilde{\Delta}_{A\cup C,B\cup C} \circ \Delta_{A\cup C,B\cup C} \rceil^{1})_{A\cup C,B\cup C}. \tag{8}$$

This result easily follows by noting that under the condition in equation (8) the map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving for each $\mathcal{M}_{\mathrm{CI}^k}$ with $\mathrm{CI} = \{A_k \perp\!\!\!\perp B_k \mid C_k\}$, since by construction every submatrix $(\tilde{\Delta}\circ\Delta)_{A_k\cup C_k,B_k\cup C_k}$ is a row-based or column-based covariation matrix. In other words, model-preservation is guaranteed if by creating the full-dimensional covariation matrix no entries with indexes in $A\cup C$ and $B\cup C$ are affected to ensure symmetry of the resulting matrix.

5. Multi-Way Model-Preserving Sensitivity Analysis

We can now generalize the results of Section 4 by studying multi- rather than single-parameter variations in Gaussian conditional independence models. In particular, we show below that the characterization of a parameter set as sets of vanishing polynomial equations provide a powerful language to straightforwardly tackle this much more general case.

Theorem 12 *Compositions of model-preserving maps are model-preserving. In particular, for any two matrices Δ and Δ' we have $\Phi_{\Delta}(\Phi_{\Delta'}) = \Phi_{\Delta\circ\Delta'}$.*

A proof of this result is given in the appendix.

Theorem 12 immediately implies that if $\tilde{\Delta}\circ\Delta$ is a model-preserving covariation scheme then any further model-preserving covariation $\tilde{\Delta}'\circ\Delta'\circ\tilde{\Delta}\circ\Delta$ does not violate the conditional independences of the model. This implies that parameters can be varied sequentially.

In fact, we can write any symmetric multi-way variation matrix Δ as the Schur product of matrices of the form considered in Section 4, namely matrices Δ^k of ones with at most two entries $\delta^k_{ij} = \delta^k_{ji}$ different from one and not equal to zero. In this notation we use superscripts in order to avoid double indices. Explicitly, consider m single-parameter variations. We then have $\Delta = \Delta^1 \circ \Delta^2 \circ \cdots \circ \Delta^m$ where every Δ^k enforces a single-parameter variation. We can now covary every single-parameter variation Δ^k by a matrix $\tilde{\Delta}^k$ using for instance row-based and column-based covariation matrices as in Proposition 11. Because the Schur product is commutative, this induces a map

$$\Phi_{\tilde{\Delta}^1\circ\Delta^1\circ\tilde{\Delta}^2\circ\Delta^2\circ\cdots\circ\tilde{\Delta}^m\circ\Delta^m} = \Phi_{\tilde{\Delta}^1\circ\tilde{\Delta}^2\circ\cdots\circ\tilde{\Delta}^m\circ\Delta^1\circ\Delta^2\circ\cdots\circ\Delta^m} = \Phi_{\tilde{\Delta}\circ\Delta}$$

where $\tilde{\Delta} = \tilde{\Delta}^1 \circ \tilde{\Delta}^2 \circ \cdots \circ \tilde{\Delta}^m$ is the covariation matrix for Δ. By Theorem 12, this map is model-preserving.

Example 11 (Example 10 continued) *For the Gaussian model defined by $Y_4 \perp\!\!\!\perp Y_{\{1,2\}} \mid Y_3$ and $Y_{\{2,4\}} \perp\!\!\!\perp Y_5 \mid Y_3$ suppose that not only the covariance σ_{43} is varied by a quantity δ_1, but also the entry σ_{32} is varied by δ_2. From Example 9 we know that the row-based covariation* *matrix on the left hand side of equation (6) is model-preserving for the variation by δ_1. From Proposition 11 it easily deduced that*

$$(\tilde{\Delta}^2 \circ \Delta^2)_{\{2,3,4\},\{1,2,3,5\}} = \begin{pmatrix} 1 & \delta_2 & \delta_2 & 1 \\ 1 & \delta_2 & \delta_2 & 1 \\ 1 & \delta_2 & \delta_2 & 1 \end{pmatrix}$$

is associated to a model-preserving covariation matrix. Therefore, using Theorem 12 we can construct the matrix

$$\tilde{\Delta} \circ \Delta = \begin{pmatrix} 1 & 1 & 1 & \delta_1 & 1 \\ 1 & \delta_2 & \delta_2 & \delta_1\delta_2 & 1 \\ 1 & \delta_2 & \delta_2 & \delta_1\delta_2 & 1 \\ \delta_1 & \delta_1\delta_2 & \delta_1\delta_2 & 1 & \delta_1 \\ 1 & 1 & 1 & \delta_1 & 1 \end{pmatrix}$$

which is associated to a model-preserving map.

6. Divergence Quantification

The previous sections formalized how variations of the covariance matrix of a Gaussian model can be coherently performed without affecting its conditional independence structure. Next, as usual in sensitivity studies, we quantify the dissimilarity between the original and the new distribution. We start by considering the KL divergence.

Using notation from Section 3.1, let Y be a Gaussian vector with density f_Σ and let $\tilde{Y}$ be the vector resulting from a model-preserving variation and having density $f_{\tilde{\Delta}\circ\Delta\circ\Sigma}$. Thus both covariance matrices belong to the parameter set of the same model, that is $\Sigma, \tilde{\Delta}\circ\Delta\circ\Sigma \in \mathcal{A}_{\text{CI}}$. Here, Δ and $\tilde{\Delta}$ may be associated to either single- or multi-parameter variations. In the latter case, as formalized in Section 5, we again denote $\Delta = \Delta^1 \circ \cdots \circ \Delta^m$ as a Schur product of matrices Δ^k associated to a single-parameter variation and by $\tilde{\Delta}^k$ their model-preserving covariation matrix. Then $\tilde{\Delta} = \tilde{\Delta}^1 \circ \cdots \circ \tilde{\Delta}^m$. Let δ_k be the variation associated to the matrix Δ^k. The KL divergence between Y and $\tilde{Y}$ in model-preserving sensitivity analyses can be written as

$$\text{KL}(\tilde{Y}||Y) = \frac{1}{2}\left[\text{tr}(\Sigma^{-1}(\tilde{\Delta}\circ\Delta\circ\Sigma)) - n + \log\frac{\det(\Sigma)}{\det(\tilde{\Delta}\circ\Delta\circ\Sigma)}\right].$$

This result easily follows by substituting the definition of our variation and covariation matrices into equation (3).

Whilst for partial and row/column-based covariations KL does not entertain a closed form, for total covariation matrices KL divergence has the following simple closed-form formula:

$$\text{KL}(\tilde{Y}||Y) = \frac{1}{2}\big(n(\delta - \log(\delta) - 1)\big)$$

where $\delta = \prod_{i\in[n]} \delta_i$ for a multi-way variation.

Since the KL divergence is the one most often used in sensitivity studies in GBNs (Gómez-Villegas et al., 2007, 2008, 2013), in the following we focus on this measure of dissimilarity. However, from a theoretical point of view there is no difficulty in considering a symmetric divergence instead, as for instance Jeffrey's divergence (see e.g. Pardo, 2006).

Formally, Jeffrey's divergence between $\tilde{Y}$ and Y, $\mathrm{J}(Y,\tilde{Y})$, for a model-preserving sensitivity analysis can be written as

$$\mathrm{J}(Y,\tilde{Y}) = \frac{1}{2}\left(\mathrm{tr}(\Sigma^{-1}(\tilde{\Delta}\circ\Delta\circ\Sigma) + (\tilde{\Delta}\circ\Delta\circ\Sigma)^{-1}\Sigma) - 2n\right).$$

Therefore, Jeffreys divergence can be written in closed form and its computation can be straightforwardly implemented in software. For this reason, it is included in the `bnmonitor` package.

Our examples in the next section demonstrate that KL divergences often behave counter-intuitively and differently depending on the form of the covariance matrix analysed. Similar results were observed not only for KL divergences but also for other members of the class of ϕ-divergences, for instance the inverse KL divergence and the Hellinger distance (Ali and Silvey, 1966). For this reason we recommend using the KL divergence in conjunction with another measure which takes into account the number of entries that have been varied. One such measure is the Frobenius norm, defined below, which has been recently used in econometrics and finance to quantify the distance between two covariance matrices (Amendola and Storti, 2015; Laurent et al., 2012). Notice in particular that the Frobenius norm is symmetric, thus addressing one possible drawback of using the KL divergence. Although in the following we focus on the Frobenius norm, notice that any p-norm (see e.g. Ando, 1994) would be an appropriate symmetric measure of dissimilarity that takes into account the number of entries varied.

Definition 13 *Let Y and Y' be two Gaussian vectors with distribution f_Σ and $f_{\Sigma'}$, respectively. The* Frobenius norm *between Y and Y' is defined as*

$$\mathrm{F}(Y,Y') = ||\Sigma-\Sigma'||_{\mathrm{F}}^2 = \mathrm{tr}((\Sigma-\Sigma')^\top(\Sigma-\Sigma')).$$

In words, the Frobenius norm is defined as the sum of the element-wise squared differences of the two covariance matrices. For standard sensitivity analyses where a variation matrix D acts additively on Σ, the Frobenius norm is simply equal to $\mathrm{tr}(D^\top D)$, consisting of the sum of the squared variations. For our multiplicative covariation, we have the following result.

Proposition 14 *Let $\tilde{\Delta}\circ\Delta = (\delta_{ij})_{ij}$ be model-preserving. Then*

$$\mathrm{F}(Y,\tilde{Y}) = \sum_{i,j\in[n]}(1-\delta_{ij})^2\sigma_{ij}^2.$$

This result easily follows by substituting $\Sigma' = \tilde{\Delta}\circ\Delta\circ\Sigma$ into the formula given in Definition 13.

Proposition 14 enables us to deduce a useful ranking based on the Frobenius norm of the various model-preserving covariation schemes we introduced. Letting $\tilde{Y}_{\mathrm{total}}$, $\tilde{Y}_{\mathrm{partial}}$, $\tilde{Y}_{\mathrm{row}}$ and $\tilde{Y}_{\mathrm{column}}$ be the random vectors resulting from total, partial, row-based and column-based covariations, respectively, the following inequalities hold:

$$\mathrm{F}(Y,\tilde{Y}_{\mathrm{total}}) \geq \mathrm{F}(Y,\tilde{Y}_{\mathrm{partial}}), \quad \mathrm{F}(Y,\tilde{Y}_{\mathrm{partial}}) \geq \mathrm{F}(Y,\tilde{Y}_{\mathrm{row}}), \quad \mathrm{F}(Y,\tilde{Y}_{\mathrm{partial}}) \geq \mathrm{F}(Y,\tilde{Y}_{\mathrm{column}}).$$

This is true simply because, by definition, total covariations affect more entries of the covariance matrix than partial ones. Similarly, partial covariations affect more entries than row- and column-based covariations.

Since it is always possible to find a variation d_{ij} that acts additively on σ_{ij} such that $d_{ij} + \sigma_{ij} = \delta_{ij}\sigma_{ij}$, we can also deduce using the same reasoning that

$$\mathrm{F}(Y, \tilde{Y}_{\text{column}}) \geq \mathrm{F}(Y, \tilde{Y}_{\text{standard}}) \quad \text{and} \quad \mathrm{F}(Y, \tilde{Y}_{\text{row}}) \geq \mathrm{F}(Y, \tilde{Y}_{\text{standard}}),$$

where $\tilde{Y}_{\text{standard}}$ is the vector resulting from standard sensitivity methods which in general break the conditional independence structure of the model. Our examples in the following give an empirical illustration of the above inequalities.

7. Model-Preserving Maps and Interval Matrices

The observation that one can always relate an additive variation d_{ij} of a parameter σ_{ij} to a multiplicative one δ_{ij} by setting $d_{ij} = \sigma_{ij}(\delta_{ij} - 1)$ allowed us in Section 6 to identify a partial ranking of model-preserving variations according to the Frobenius norm. However, and more generally, the same observation can be used for two additional purposes: (1) determining if a perturbed covariance matrix is still positive semidefinite; (2) provide an upper bound on the KL divergence.

For a covariation matrix $\tilde{\Delta}$ and a variation matrix Δ define

$$D = \tilde{\Delta} \circ \Delta \circ \Sigma - \Sigma. \tag{9}$$

The matrix D performs exactly the same perturbation of $\tilde{\Delta} \circ \Delta$ but additively. Given this observation we have the following result.

Proposition 15 *For a model-preserving map $\Phi_{\tilde{\Delta}\circ\Delta}$, $\tilde{\Delta} \circ \Delta \circ \Sigma \in \mathbb{R}^{n\times n}_{\text{psd}}$ if*

$$\rho(D) \leq \lambda_{\min}(\Sigma),$$

where ρ is the spectral radius and $\lambda_{\min}$ is the smallest eigenvalue.

The result easily follows from the theory of interval matrices (Rohn, 1994; Horn and Johnson, 2012) and was already used in Gómez-Villegas et al. (2013) for standard sensitivity analyses.

Proposition 15 is relevant for partial, row- and column-based covariations. This is because covariance matrices resulting from total model-preserving covariations are always positive semidefinite. To see this, recall that a symmetric matrix $\Sigma \in \mathbb{R}^{n\times n}$ is positive semidefinite if and only if $x^\top \Sigma x \geq 0$ for all $x \in \mathbb{R}^n$. If Σ is multiplied by a positive constant δ, as in total model-preserving analyses, then by default $x^\top \delta\Sigma x \geq 0$ for all $x \in \mathbb{R}^n$.

Using again the notion of interval matrices, Gómez-Villegas et al. (2013) derived bounds for the KL divergence of standard sensitivity methods. Using the same reasoning we can derive the following result.

Proposition 16 *Let $\lambda_{\max}$ be the largest eigenvalue of a matrix and $f(x) = \ln(1+x) - x/(1+x)$. Let Y be a Gaussian vector and $\tilde{Y}$ be the vector resulting from a model-preserving variation. Then*

$$\mathrm{KL}(\tilde{Y}||Y) \leq 0.5n \max\left\{f(\lambda_{\max}(\tilde{\Delta}\circ\Delta)), f(\lambda_{\min}(\tilde{\Delta}\circ\Delta))\right\}$$

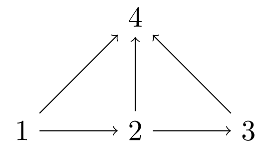

Figure 3: The directed acyclic graph representing the conditional independence model in Section 8.1.

The result follows from Proposition 4 of Gómez-Villegas et al. (2013) and equation (9). Therefore for model-preserving analyses these bounds only depend on the variation and covariation matrices. This is different to standard sensitivity analyses where the bounds are also a function of the original covariation matrix Σ. The bounds can be used to assess to which parameters' misspecifications the network is less robust before actually carrying out any sensitivity analysis. We give an illustration of this in the following section.

8. Illustrations

We now illustrate the results of the previous sections using two examples: one artificial and one based on a real-world data application. Furthermore, we investigate the efficiency of the model-preserving approach in comparison to the standard one of Gómez-Villegas et al. (2007) and Gómez-Villegas et al. (2013) using these two examples as well as others based on simulated data.

8.1. A First Example

Consider the BN model represented in Figure 3 and associated to the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 2 & 2 & 7 \\ 2 & 5 & 5 & 17 \\ 2 & 5 & 6 & 19 \\ 7 & 17 & 19 & 63 \end{pmatrix}. \tag{10}$$

This matrix was deduced using the formalism of Section 2.3 by setting $\beta_{0i} = 0$, $v_i = 1$, for $i \in [4]$, $\beta_{12} = 2$, $\beta_{13} = 0$, $\beta_{23} = 1$, $\beta_{14} = 1$, $\beta_{24} = 1$ and $\beta_{34} = 2$. Covariance matrices with a structure similar to the one in equation (10) are often encountered when the β_{ij} parameters are expert-elicited (see for instance Gómez-Villegas et al., 2011; Gómez-Villegas et al., 2013). Notice that this BN is defined by only one conditional independence statement, namely $Y_3 \perp\!\!\!\perp Y_1 \mid Y_2$. This is equivalent to the vanishing minor $\sigma_{12}\sigma_{23} - \sigma_{22}\sigma_{13} = 0$ by Lemma 1. Thus only variations of the parameters σ_{21}, σ_{22}, σ_{31} and σ_{32} may break the conditional independence structure of this model.

Figure 4 reports the KL divergence for one-way sensitivity analyses of each of the above parameters when entries are either increased or decreased by 25%. The plots show that the KL divergence is considerably smaller for total covariation matrices than for all the other covariations as well as for standard sensitivity methods. All other methods have similar KL divergences and we see that for most variations there is one model-preserving covariation

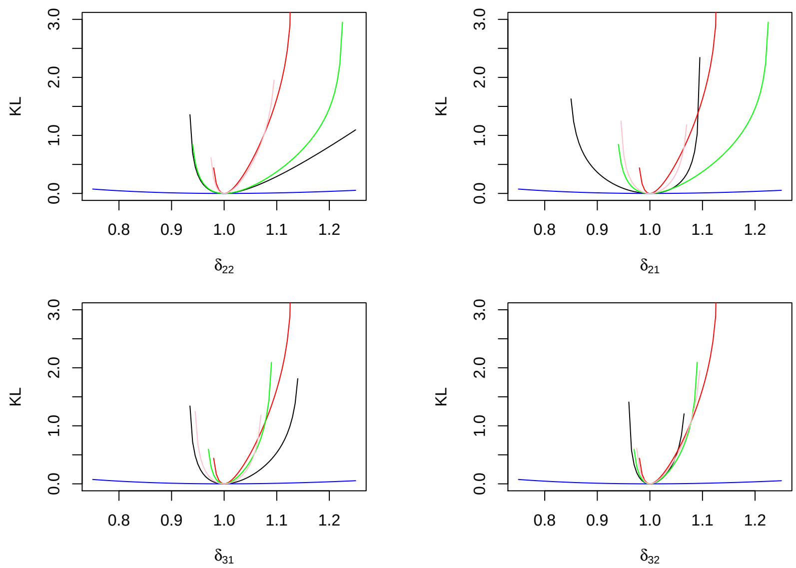

Figure 4: KL divergence for one-way variations $\sigma_{ij} \mapsto \delta_{ij}\sigma_{ij}$ of the parameters of the network in Figure 3. We use the color codes black = standard variation; blue = full; red = partial; green = row-based; pink = column-based.

Row	Column	Standard	Total	Partial	Row-based	Column-based
1	1	0.063	1.682	1.623	1.623	1.623
1	2	0.718	1.682	1.647	1.631	1.631
1	3	0.261	1.682	1.647	1.635	1.635
1	4	NA	1.682	NA	NA	NA
2	2	0.262	1.682	1.647	1.631	1.631
2	3	2.565	1.682	1.647	1.635	1.635
2	4	NA	1.682	NA	NA	NA
3	4	0.633	1.682	NA	NA	NA
4	4	1.328	1.682	1.623	1.623	1.623

Table 1: Bounds for the KL divergence for all parameters of the BN in Figure 3 a for all covariation schemes after a multiplicative perturbation by $\delta = 1.05$. The NA entries denote situations where the resulting covariance matrix was not positive semidefinite.

with KL divergence either smaller or comparable to the one of the standard method. The KL divergence takes similar values for all parameters varied and therefore none of these has a predominant effect on the robustness of the network.

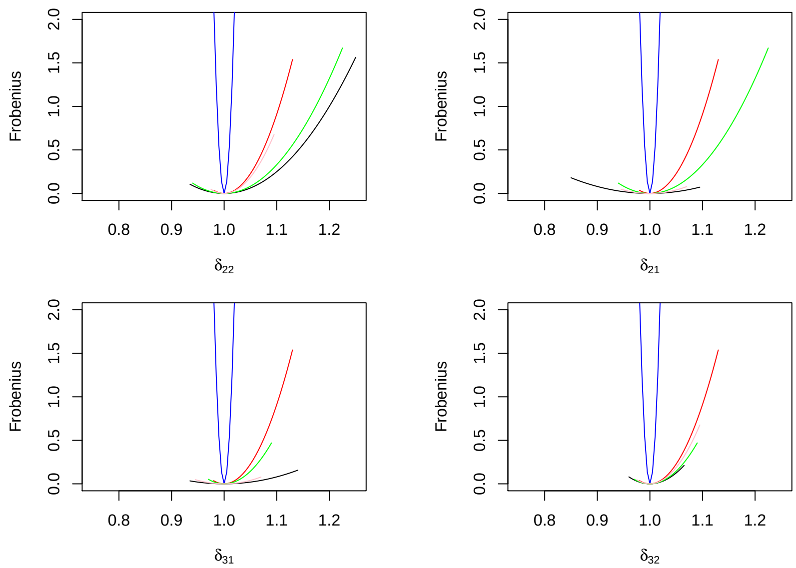

Figure 5: Frobenius norm for one-way variations $\sigma_{ij} \mapsto \delta_{ij}\sigma_{ij}$ of the parameters of the network in Figure 3. We use the color codes black = standard; blue = full; red = partial; green = row-based; pink = column-based.

Table 1 reports the KL bounds for variations of all parameters by a $\delta = 1.05$ (a 5% increase). We can notice that as expected the bounds for total variation are all equal, whilst the bounds for partial covariations are equal to two values depending as to whether the entry of the covariance matrix appears in the vanishing polynomial or not. Furthermore, the bounds for all model-preserving approaches are quite close to each other and do not exhibit much variability between different entries of the covariance matrix. Conversely, the bounds for the standard approach have greater variability: notably the bound for the entry in position $(2, 3)$ is the largest among all parameters and all approaches.

Figure 5 reports the Frobenius norms under the same settings as above, confirming the theoretical results of Section 6. In particular, standard sensitivity methods always have a smaller Frobenius norm than the others because in this case less parameters are varied. The plots also confirm that there is no fixed ranking between column-based and row-based covariations and demonstrate that full covariation has a considerably larger Frobenius norm than the other approaches. Again, the Frobenius norm appears to be comparable between all parameters varied and therefore none of these seems to be critical.

In Figures 4 and 5 the distance between the original and the varied distribution is not reported for all possible variations since for such variations the resulting covariance matrix is not positive semidefinite. This is even more evident in Figures 6 and 7 reporting the KL divergence and the Frobenius norm, respectively, for the multi-way sensitivity analysis of the parameters σ_{22} and σ_{33}. In these plots the white regions correspond to combinations of variations such that the resulting covariance matrix is not positive semidefinite: such a

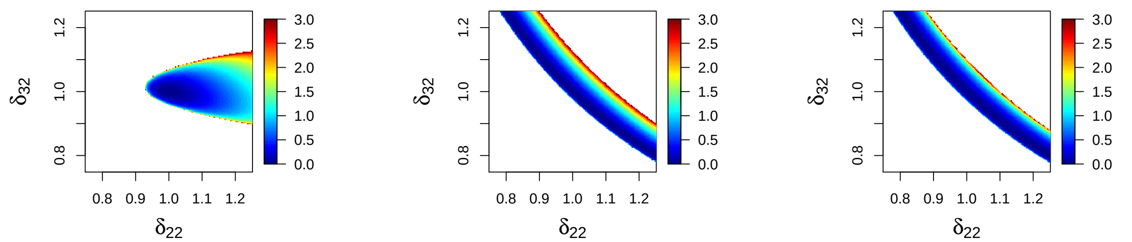

Figure 6: KL divergence for multi-way variation of the parametes σ_{22} and σ_{32} of the parameters of the network in Figure 3: standard (left); partial (central); column-based (right).

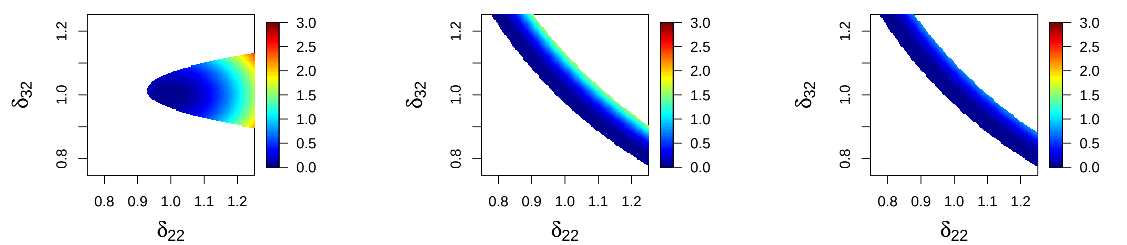

Figure 7: Frobenius norm for multi-way variation of the parametes σ_{22} and σ_{32} of the parameters of the network in Figure 3: standard (left); partial (central); column-based (right).

region is very different for the case of the standard sensitivity method (on the left) and the model-preserving ones (on the right and in the center).

Notice that Figures 6 and 7 do not report the divergence between the original and varied distribution in the case of total model-preserving covariations since their inclusion would have made the other plots not particularly informative: the KL divergence for full covariations can be shown to be way smaller than the others reported in Figure 6, whilst its Frobenius norm is considerably larger. Conversely, standard, partial and column-based (row-based is not included since for this example it would coincide with the partial one) have comparable divergences. However, as expected, the Frobenius norm is smaller for the standard method, although the difference does not appear to be very large.

8.2. A Real-World Application

In this section we study a subset of the data set of Eisner et al. (2011) including metabolomic information of 77 individuals: 47 of them suffering of cachexia, whilst the remaining do not. Cachexia is a metabolic syndrome characterized by loss of muscle with or without loss of fat mass. Although the study of Eisner et al. (2011) included 71 different metabolics which

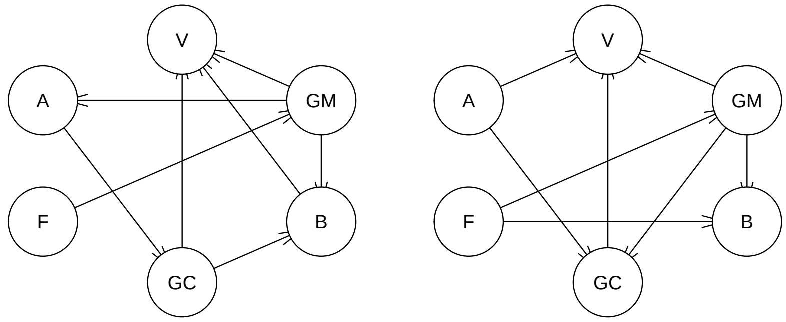

Figure 8: Learnt BN model of the metabolics for patients with Cachexia (left network) and for the control group (right network).

could possibly distinguish individuals who suffer of Cachexia from those who do not, for our illustrative purposes we focus on only six of them: Adipate (A), Betaine (B), Fumarate (F), Glucose (GC), Glutamine (GM) and Valine (V). Two GBN models were learnt for the two different populations (ill and not ill) using the `bnlearn` R package (Scutari, 2010) resulting in the networks in Figure 8. The order of the variables was kept fixed for the two populations for ease of comparison. The estimated covariance matrix for individuals suffering of Cachexia is

	F	GM	A	GC	B	V
F	304	3262	220	2963	414	208
GM	3262	98456	6637	89431	12489	6279
A	220	6637	3950	53223	1693	839
GC	2963	89431	53223	3050126	65012	31858
B	414	12489	1695	65012	7279	1791
V	208	6279	839	31858	1791	1124

whilst for the control group this is estimated as

	F	GM	A	GC	B	V
F	38647	1004	0	310	168	51
GM	1004	109	0	11923	10192	1974
A	0	0	41	376	0	77
GC	310	11923	376	8952	3144	1092
B	168	10192	0	3144	5171	520
V	51	1974	77	1092	520	192

After transforming the two above covariance matrices into correlations it was observed that only two covariances had a disagreement larger than 0.4 (in correlation scale) between the following Metabolics: GC/F and GC/GM. Therefore these are considered of interest. Furthermore there is interest in the covariance between A/F and A/GM since these pairs are

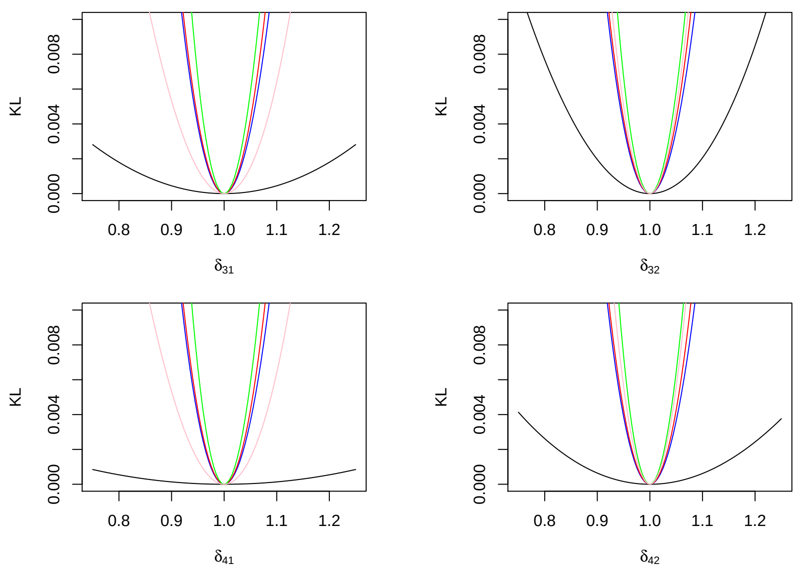

Figure 9: KL divergence for one-way variations $\sigma_{ij} \mapsto \delta_{ij}\sigma_{ij}$ of the parameters of the network for patients suffering of Cachexia. We use the color code black = standard variation; blue = full; red = partial; green = row-based; pink = column-based.

estimated independent in the control group network, whilst they are dependent for patients suffering of Cachexia. A sensitivity analysis over these parameters is carried out for the network learnt using the data of patients suffering of Cachexia to investigate its robustness. For ease of exposition, we report here a one-way sensitivity analysis over such parameters only, though multi-way analyses can be conducted as formalized in Section 5 and illustrated in Section 8.1.

Figure 9 reports the KL divergence for the chosen parameters of the network for patients suffering of Cachexia. We can notice that conversely to the sensitivity analysis carried out in Section 8.1, now standard methods have a much smaller KL divergence than model preserving ones. Furthermore, variations of different parameters lead to substantially different KL divergences under the traditional approach. For model-preserving variations we notice that the KL divergences are fairly similar for variations of different parameters. In addition, row-based model-preserving variations lead to significantly smaller KL divergences in two out of four cases. Thus, based on the result from both row-based and traditional methods, the covariances between A/GM and GC/GM appear to have a much stronger effect on the robustness of the network and therefore the validity of their estimated values needs to be carefully validated, for instance using expert information.

Similar conclusions can be drawn from Figure 10 reporting the logarithm of the Frobenius norm for the different parameter variations. For these plots, the goodness of the row-based model-preserving scheme is much more evident and especially for the covariance between GC/F, the Frobenius norm of such scheme is almost equal to the one of the traditional approach. Of course model-preserving covariations are expected to have a larger

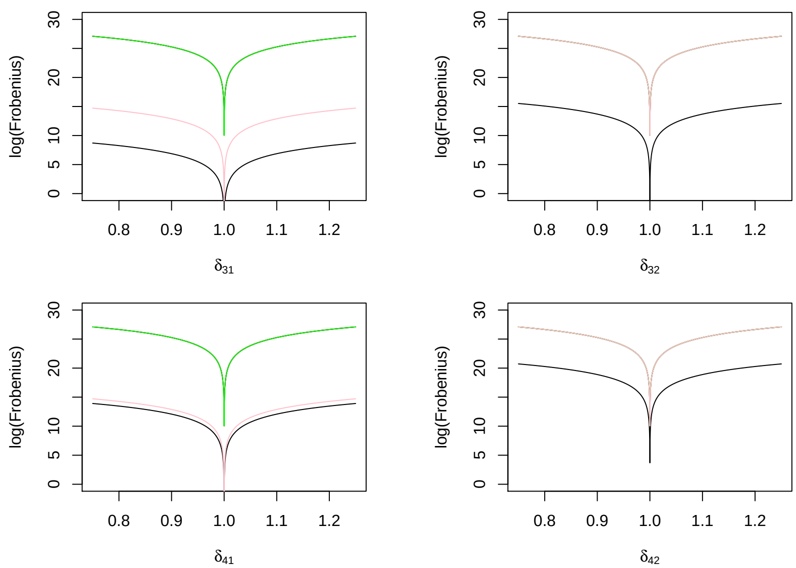

Figure 10: Frobenius norm for one-way variations $\sigma_{ij} \mapsto \delta_{ij}\sigma_{ij}$ of the parameters of the network in Figure 3. We use the color codes black = standard variation; blue = full; red = partial; green = row-based; pink = column-based.

Frobenius norm than in the standard approach, since in general more than one parameter needs to be perturbed to ensure the conditional independence structure is preserved. Notice that, because of the structure of the covariance matrix for this example, all variations considered were admissible and lead to a positive semidefinite matrix, irrespective of the approach used.

8.3. Efficiency of Model-Preserving Covariation

The `bnmonitor` package offers an implementation of model-preserving as well as standard covariation methods for Gaussian Bayesian networks. We next investigate the computational cost of model-preserving covariations in comparison to standard ones using our implementation in `bnmonitor`.

Our study compares the times required to compute the KL divergence for a vector of variations of length 20 under different covariation approaches over six BNs of increasing size. The first two networks correspond to the synthetic and real-world data analyzed above. For the remaining four networks, random data sets of uniform numbers were generated and BNs were learnt using the `bnlearn` R package. Details about these networks, specifically number of vertices and edges, are given in Table 2. The computing times include also a preprocessing step before the computation of the KL divergence to transform objects of class `bn.fit` (the output of the search method and maximum likelihood estimation of `bnlearn`) to objects including the mean vector and the covariance matrix (for standard sensitivity), and also all conditional independence statements of the BN (for model-preserving sensitivity).

	$\lvert V\rvert = 4$ $\lvert E\rvert = 5$	$\lvert V\rvert = 6$ $\lvert E\rvert = 8$	$\lvert V\rvert = 10$ $\lvert E\rvert = 38$	$\lvert V\rvert = 25$ $\lvert E\rvert = 205$	$\lvert V\rvert = 50$ $\lvert E\rvert = 758$	$\lvert V\rvert = 100$ $\lvert E\rvert = 2450$
Standard	0.02	0.03	0.04	0.17	1.61	26.78
Total	0.02	0.03	0.05	0.19	1.54	26.55
Partial	0.03	0.05	0.07	0.25	1.95	27.60
Row-based	0.03	0.03	0.05	0.31	2.54	28.86
Column-based	0.04	0.05	0.06	0.31	2.50	29.19

Table 2: Computation times for the KL divergence over a vector of variations of length 20, for different sensitivity methods and BNs.

Computations were carried out on a Intel Core I7 of 8th generation. The results in Table 2 suggest that, although the computational times for partial, row- and column-based covariation are slightly larger, all approaches require the same computational efforts. Furthermore, for networks of moderate size, up to 50 nodes, computations are completed almost instantaneously. For the larger BN with 100 nodes, the computational times increase notably for all approaches. This is actually mostly due to the pre-processing step which took 24.8 and 26.2 seconds for standard and model-preserving sensitivity methods respectively. Therefore, the actual computation of the model-preserving sensitivity matrices is still very quick even for larger BNs.

9. Discussion

Algebraic tools have proved to be extremely powerful to characterize conditional independence models and inferences based on such models. Here we have taken advantage of these tools to perform sensitivity analyses in GBNs which do not break the structure of the model. We demonstrated through various examples that our new methods are robust, meaning that the divergences computed under our paradigm are often comparable to those arising from standard methods, with the difference that in our approach the underlying network continues to be a coherent representation of the model.

The development of the `bnmonitor` R package provides an intuitive platform to implement the methods developed in this paper, as well as standard ones, in a variety of applications. Currently, the software implements only one-way sensitivity methods, but we plan to include multi-way methods in future releases. Given that currently almost no software allows for sensitivity studies, the continuous development of such a package is critical and could be of great benefit for the whole AI community.

Acknowledgements

Christiane Görgen and Manuele Leonelli were supported by the programme “Oberwolfach Leibniz Fellows” of the Mathematisches Forschungsinstitut Oberwolfach in 2017.

Appendix A. Proofs

A.1. Proof of Theorem 7

From Proposition 1, the result follows if all $(\#C+1)\times(\#C+1)$ minors of $(\tilde{\Delta}\circ\Delta\circ\Sigma)_{A\cup C,B\cup C}$ vanish. First recall that by Leibniz formula we can write any $(\#C+1)\times(\#C+1)$ minor of $\Sigma_{A\cup C,B\cup C}$ as a polynomial g

$$g(\Sigma_{A\cup C,B\cup C}) = \sum_{\tau\in S_{\#C+1}} \mathrm{sgn}(\tau)\prod_{i=1}^{\#C+1}\sigma_{i\tau(i)},$$

where $S_{\#C+1}$ denotes the symmetric group of permutations of the $\#C+1$ indices and $\mathrm{sgn}(\tau)$ is the signature of τ. Since for both total and partial covariation matrices all the entries of $(\tilde{\Delta}\circ\Delta)_{A\cup C,B\cup C}$ are equal to δ, we have that

$$g((\tilde{\Delta}\circ\Delta\circ\Sigma)_{A\cup C,B\cup C}) = \sum_{\tau\in S_{\#C+1}} \mathrm{sgn}(\tau)\prod_{i=1}^{\#C+1}\delta\sigma_{i\tau(i)} = \delta^{\#C+1}g(\Sigma_{A\cup C,B\cup C})$$

which is equal to zero since $g(\Sigma_{A\cup C,B\cup C})=0$ by Lemma 1.

A.2. Proof of Theorem 8

In analogy to the proof of Theorem 7, the result follows if all $(\#C+1)\times(\#C+1)$ minors of $(\tilde{\Delta}\circ\Delta\circ\Sigma)_{A\cup C,B\cup C}$ vanish. To prove the result in this case we use Laplace expansion formula which states that any $(\#C+1)\times(\#C+1)$ minor of $\Sigma_{A\cup C,B\cup C}$ can be written as a polynomial g

$$g(\Sigma_{A\cup C,B\cup C}) = \sum_{j=1}^{\#C+1}(-1)^{i+j}\sigma_{ij}\det\Sigma^{-ij}_{A\cup C,B\cup C} = \sum_{i=1}^{\#C+1}(-1)^{i+j}\sigma_{ij}\det\Sigma^{-ij}_{A\cup C,B\cup C},$$

where $\Sigma^{-ij}_{A\cup C,B\cup C}$ denotes the matrix $\Sigma_{A\cup C,B\cup C}$ without the i-th row and the j-th column.

Start considering the case (i,j) or $(j,i)\in(A,B)$ and suppose for a row-based covariation $E=\{i\}$. Then

$$\begin{aligned} g((\tilde{\Delta}\circ\Delta\circ\Sigma)_{A\cup C,B\cup C}) &= \sum_{k=1}^{\#C+1}(-1)^{i+k}\delta\sigma_{ik}\det(\tilde{\Delta}\circ\Delta\circ\Sigma)^{-ik}_{A\cup C,B\cup C} \\ &= \delta\sum_{k=1}^{\#C+1}(-1)^{i+k}\sigma_{ik}\det(\tilde{\Delta}\circ\Delta\circ\Sigma)^{-ik}_{A\cup C,B\cup C} \end{aligned} \tag{11}$$

where we use superscripts in matrices to denote rows and columns to be eliminated.

The result follows if $\det(\tilde{\Delta}\circ\Delta\circ\Sigma)^{-ik}_{A\cup C,B\cup C} = \det\Sigma^{-ik}_{A\cup C,B\cup C}$ for all $k=1,\dots\#C+1$. However this is true since, for $E=\{i\}$, δs are only in entries (i,k) and no entries (k,i), that need to be equal to δ for symmetry, are in $(A\cup C,B\cup C)$.

Consider next the case $E = \{i, l\}$ for a row-based covariation. Then in this case $\det(\tilde{\Delta} \circ \Delta \circ \Sigma)^{-ik}_{A\cup C, B\cup C} \neq \det \Sigma^{-ik}_{A\cup C, B\cup C}$. However, using again Laplace formula in equation (11), we have that

$$\begin{aligned} g((\tilde{\Delta} \circ \Delta \circ \Sigma)_{A\cup C, B\cup C}) &= \delta \sum_{k=1}^{\#C+1} (-1)^{i+k} \sigma_{ik} \sum_{r=1}^{\#C} (-1)^{l+r} \delta \sigma_{lr} \det(\tilde{\Delta} \circ \Delta \circ \Sigma)^{-\{i,l\}\{k,r\}}_{A\cup C, B\cup C} \\ &= \delta^2 \sum_{k=1}^{\#C+1} (-1)^{i+k} \sigma_{ik} \sum_{r=1}^{\#C} (-1)^{l+r} \sigma_{lr} \det(\tilde{\Delta} \circ \Delta \circ \Sigma)^{-\{i,l\}\{k,r\}}_{A\cup C, B\cup C}. \end{aligned}$$

The result follows again since $\det(\tilde{\Delta} \circ \Delta \circ \Sigma)^{-\{i,l\}\{k,r\}}_{A\cup C, B\cup C} = \det(\Sigma)^{-\{i,l\}\{k,r\}}_{A\cup C, B\cup C}$. Laplace expansion can now be used iteratively to demonstrate that all row-based covariation matrices with $E \subseteq A$ induce a model-preserving map when (i, j) or $(j, i) \in (A, B)$.

The result follows using the same reasoning for column-based covariation matrices when (i, j) or $(j, i) \in (A, B)$ by using the Laplace formula expansion over the rows of the matrix. The result is equally proven for row-based covariations when (i, j) or $(j, i) \in (A, C)$ and column-based covariations when (i, j) or $(j, i) \in (C, B)$.

The proof of the result needs to be slightly adapted when δs appear in the submatrix $(\tilde{\Delta} \circ \Delta)_{C,C}$: that is for column-based covariation if (i, j) or $(j, i) \in (A, C)$, for row-based covariation if (i, j) or $(j, i) \in (C, B)$ and for both covariations if (i, j) and $(j, i) \in (C, C)$ where for symmetric reasons both belong to the submatrix. In such cases, because the matrix $\lfloor (\tilde{\Delta} \circ \Delta \circ \Sigma)_{A\cup C, B\cup C} \rfloor^1$ needs to be symmetric, extra δs already appear within $\tilde{\Delta}_{A\cup C, B\cup C}$. So for instance, if all the entries in the i-th row of $(\tilde{\Delta} \circ \Delta)_{C,C}$ are δ, then also its i-th column must have δs for symmetry. But because of this then we only have that $\det(\tilde{\Delta} \circ \Delta \circ \Sigma)^{-CC}_{A\cup C, B\cup C} = \det(\Sigma)^{-CC}_{A\cup C, B\cup C}$, thus requiring us to apply the Laplace expansion over all rows or all columns with index in C.

A.3. Proof of Theorem 12

Let $\mathcal{A}$ be a parameter set. If D and D' are matrices such that Φ_D and $\Phi_{D'}$ map $\mathcal{A}$ to a subset of itself then also the composition of these two maps sends $\mathcal{A}$ to a subset of itself, $\Phi_D(\Phi_{D'}(\mathcal{A})) \subseteq \mathcal{A}$. Furthermore, $\Phi_D(\Phi_{D'}(\Sigma)) = \Phi_D(D' \circ \Sigma) = D \circ D' \circ \Sigma$ for any matrix Σ.

Appendix B. Separable Models

Definition 17 *We say that two relationships $A_k \perp\!\!\!\perp B_k \mid C_k$ and $A_l \perp\!\!\!\perp B_l \mid C_l$ in a Gaussian conditional independence model $\mathcal{M}_{\mathrm{CI}}$ are* separated *if for any entry σ_{kl} of $\Sigma_{A_k\cup C_k, B_k\cup C_k}$ neither σ_{kl} nor σ_{lk} are in $\Sigma_{A_l\cup C_l, B_l\cup C_l}$ and viceversa. A model $\mathcal{M}_{\mathrm{CI}}$ is called* separable *if all its pairs of conditional independence statements are separated.*

By separability of $\mathcal{M}_{\mathrm{CI}}$, the following result holds.

Proposition 18 *Let $\mathcal{M}_{\mathrm{CI}}$ be separable with parameter space $\mathcal{A}_{\mathrm{CI}}$. Given a covariance matrix $\Sigma \in \mathcal{A}_{\mathrm{CI}}$, the map $\Phi_{\tilde{\Delta}\circ\Delta}$ is model-preserving for $\tilde{\Delta} = \circ_{k\in[r]}\tilde{\Delta}_k$, where $\tilde{\Delta}_k$ is a $\mathcal{M}_{\mathrm{CI}_k}$-preserving covariation matrix for $\mathrm{CI}_k = \{A_k \perp\!\!\!\perp B_k \mid C_k\}$.*

References

S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society Series B*, 1(28):131–142, 1966.

A. Amendola and G. Storti. Model uncertainty and forecast combination in high-dimensional multivariate volatility prediction. *Journal of Forecasting*, 34(2):83–91, 2015.

T. Ando. Majorizations and inequalities in matrix theory. *Linear algebra and its Applications*, 199:17–67, 1994.

E. Borgonovo and E. Plischke. Sensitivity analysis: a review of recent advances. *European Journal of Operational Research*, 248(3):869–887, 2016.

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

E. Castillo, J. M. Gutiérrez, and A. S Hadi. Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(4):412–423, 1997.

H. Chan and A. Darwiche. When do numbers really matter? *Journal of Artificial Intelligence Research*, 17:265–287, 2002.

V. M. H. Coupé and L. C. van der Gaag. Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36(4):323–356, 2002.

A. P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B*, 41:1–31, 1979.

M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on Algebraic Statistics.* Oberwolfach Seminars, vol. 39, Springer, Berlin, 2008.

R. Eisner, C. Stretch, T. Eastman, J. Xia, D. Hau, Sa. Damaraju, R. Greiner, D. S. Wishart, and V. E. Baracos. Learning to predict cancer-associated skeletal muscle wasting from 1h-nmr profiles of urinary metabolites. *Metabolomics*, 7(1):25–34, 2011.

S. Fallat, S. Lauritzen, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik. Total positivity in Markov structures. *The Annals of Statistics*, 45(3):1152–1184, 2017.

S. French. Modelling, making inferences and making decisions: the roles of sensitivity analysis. *Top*, 11:229–251, 2003.

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

M. A. Gómez-Villegas, P. Main, and R. Susi. Extreme inaccuracies in Gaussian Bayesian networks. *Journal of Multivariate Analysis*, 99(9):1929–1940, 2008.

M. A. Gómez-Villegas, P. Main, H. Navarro, and R. Susi. Evaluating the difference between graph structures in Gaussian Bayesian networks. *Expert Systems with Applications*, 38 (10):12409–12414, 2011.

M. A. Gómez-Villegas, P. Main, and R. Susi. The effect of block parameter perturbations in Gaussian Bayesian networks: sensitivity and robustness. *Information Sciences*, 222: 439–458, 2013.

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 2012.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

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

S. Laurent, J. V. K. Rombouts, and F. Violante. On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics*, 27(6):934–955, 2012.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.

M. Leonelli and E. Riccomagno. A geometric characterisation of sensitivity analysis in monomial models. *arXiv:1901.02058*, 2018.

M. Leonelli, C. Görgen, and J. Q. Smith. Sensitivity analysis in multilinear probabilistic models. *Information Sciences*, 411:84–97, 2017.

D. V. Malioutov, J. K. Johnson, and A. S. Willsky. Walk-sums and belief propagation in Gaussian graphical models. *Journal of Machine Learning Research*, 7:2031–2064, 2006.

L. Pardo. *Statistical Inference Based on Divergence Measures*. Taylor & Francis, 2006.

J. Pearl. *Probabilistic reasoning in intelligent systems*. Morgan-Kauffman, San Francisco, 1988.

J. Pitchforth and K. Mengersen. A proposed validation framework for expert elicited Bayesian networks. *Expert Systems with Applications*, 40:162–167, 2013.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL `https://www.R-project.org/`.

S. Renooij. Co-variation for sensitivity analysis in Bayesian networks: properties, consequences and alternatives. *International Journal of Approximate Reasoning*, 55:1022–1042, 2014.

T. Richardson and P. Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

J. Rohmer. Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: A comprehensive review. *Engineering Applications of Artificial Intelligence*, 88: 103384, 2020.

J. Rohn. Positive definiteness and stability of interval matrices. *SIAM Journal on Matrix Analysis and Applications*, 15(1):175–184, 1994.

D. Rusakov and D. Geiger. Asymptotic model selection for naive Bayesian networks. *Journal of Machine Learning Research*, 6:1–35, 2005.

A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity analysis.* Wiley, New York, 2000.

M. Scutari. Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software*, 35:1–22, 2010.

M. Slawski and M. Hein. Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra and its Applications*, 473: 145–179, 2015.

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

P. Zwiernik. An asymptotic behaviour of the marginal likelihood for general Markov models. *Journal of Machine Learning Research*, 12:3283–3310, 2011.