Additive Bayesian Networks for Multivariate Data: Parameter Learning, Model Fitting and Applications in Veterinary Epidemiology

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Marta Pittavino

aus

Italien

Promotionskomitee

Prof. Dr. Reinhard Furrer (Vorsitz)
Prof. Dr. Paul Torgerson
Prof. Dr. Leonhard Held
Dr. Fraser Lewis

Zürich, 2016

'Above all, don't fear difficult moments.
The best comes from that.'
Rita Levi Montalcini

To Andrea,
for being always beside me.

To Marilisa,
for our special relationship and her outstanding Foundation.

To my family: Claudio, Delfina, Luca & **Lucia**,
for their constant support, also if far away.

Acknowledgements

'We often take for granted the very things that most deserve our gratitude.'
Cynthia Ozick

When I think at the path taken to arrive at this point, it comes to my mind a mountain landscape. The aim is reaching the highest summit, however it never arrives.. when a peak is achieved one realizes it is not the top and better can always be done. This journey has been demanding and hard, but looking back it has been worth to complete the 'hike'. It simply would not have been possible without the help of many people, along the way, that have added their own unique form of guidance, support, and motivation.

In first place, I want to thank a lot my supervisor Reinhard Furrer for having given me the chance to continue my PhD under his excellent guidance. I am very grateful for his endless patience, constant encouragement, and full support. It has been an honor working with an experienced statistician like him. He taught me a lot about modern statistics techniques and visualization. Moreover, he gave me also the opportunity to be his first female doctoral student and having Tukey as an academic great grandfather. I am also thankful for having given me the opportunity to attend international conferences and to carry on my studies in Switzerland, one of the most enriching experiences of my life. Thank you very much!

A special gratitude goes to Paul Torgerson and Fraser Lewis for offering me the opportunity to do a PhD at the University of Zurich. They had a very important influence, contributing with precious and right advice to enhance the quality and content of the thesis.

I likewise thank Leonhard Held for his useful comments during the committee meetings, which helped to improve and look at further developments of the research.

I had the great honor of working with Anou Dreyfus and Sonja Hartnack: not only are they outstanding epidemiologists, but also brilliant and wonderful women! Their talent and wisdom have been an excellent example to follow. Anou always managed to bring me happiness and make me smile, also when everything was completely 'cloudy': Grazie Cara! Sonja taught me how to be more determinate and stronger: Danke Vielmals! I have benefited a lot from our collaboration and I am glad they chose me as collaborator.

My PhD fellows were invaluable during the work on my thesis. Especially, many thanks to Belen and Fabien, I have no idea how I could survive my first months in Zurich without you. Belen, thank you to take care of me, as I was your little sister, and to be a great reference point to me. Fabien, thank you for your patience in the office, and to help us understanding and discovering, a bit more, the Swiss rules. Overall, thank you very much to all the members of the Veterinary Epidemiology group of the University of Zurich!

A significant acknowledgement is for the Applied Statistics group of the University of Zurich: many many thanks to Clément, David, Julia, Florian and Mattia! I want to thank you a lot for the interesting statistical, and not only, discussions during lunch at the Mensa. Each of you

helped and contributed to enhance my statistical skills and clarify my programming doubts. I would like also to thanks a lot Steven Geinitz and Wei Wei for technical support with the thesis template and printing.

I want to thank Eva Furrer and Milo Puhan for the excellent PhD program offered. The seminars series was extremely interesting and I profited a lot to gain novel and further epidemiological knowledge. Thanks to Eva's coordination and organization, I managed to keep track of the activity done and still to do, in order to complete the PhD: Thank you!

An enormous gratitude goes to Franziska Robmann and her daughter Niamh. They helped me very much with all the administration issues, but not only. They took always care of my possible needs and contributed to make me feeling at home in Zurich. I feel deeply indebted for their heartfelt support in any occasion.

Many many thanks to Carsten Rose and Rafael Ostertag for helping me in every moment with technological issues and to provide a wonderful service with the server at the I-Math.

I would probably stop this hard journey earlier, if it would not be for the constant encouragements of a very good mathematician, Alessandra Cipriani. She believed in me much earlier than when I started. We still have to write a book together, I hope we will find the time soon.

Thinking about another great mathematician, I want to thank a lot Martina Dal Borgo. She was patient with me since the first time we met in the office. Although she could have asked to change desk after our acquaintance, she was enough brave to keep her assignment and she allowed an important friendship to start.

Many many thanks to Emma Hovhannisyan and Wei Xue, to stand me in the office and always help and support me when needed, especially during the last hectic months.

I want also to thank Ashkan Nikeghbali to provide useful suggestions for the life in France.

Special thanks are for Annalisa Massaccesi and Camillo De Lellis, to organize the SOLA in 2014. Another big thank is for Beniamino Guerra and the BioStats Epi Students, to organize the SOLA in 2015. Thank you very much to all the runners, to participate to SOLA with me! Beside the running, also the tennis and dancing time was important to achieve this goal, Chiara Boccato and Paola Bacigaluppi, thank very much to take part to this activity with me!

Overall, a huge thank you to all the I-Math members and BioStats Epi students, especially to Giacomo and Gosia, that contributed to create memorable moments during the thesis writing.

A special mention must be made to an extraordinary woman who had a profound influence, thank you very much Marilisa! I would have never thought that our long phone calls would have helped to build such a special relationship. Your wisdom advice and frequent encouragement were a precious motivation for me to keep going, especially after hard times.

Heartfelt thanks to my wonderful parents, to my mum Lucia and my dad Claudio because they love me even when we do not agree. Without all their education and encouragement, I could never arrived where I managed to: 'Grazie Mamma e Grazie Papà!'

An exceptional thank to my brother Luca and my grandma Delfina, for being simply among the few people on which I can always count for, also at last minute.

Last, but not least, my beloved Andrea. There are not enough words to thank you for all your attentions and care during these years. Your presence has been fundamental for me. Thanks for being always beside me, during positive and, especially, negative situations.

For those I might have forgotten to thank here, all my apologizes and acknowledgments for their contribution to this dissertation.

Marta Pittavino

Zurich, March 2016

Zusammenfassung

Veterinärepidemiologie, eine der facettenreichen Anwendungen der Statistik, zielt darauf vermutete Zusammenhänge zwischen Kovariaten oder Prädiktoren und einer oder mehr Zielvariablen zu untersuchen. Häufig sind die zugrundeliegenden biologischen Prozesse komplex und resultieren in multiplen Abhängigkeiten innerhalb der Prädiktoren und der Zielvariablen. Standardverfahren der Epidemiologie und Statistik sind nur begrenzt geeignet, um multiple Abhängigkeiten multivariater Daten zu beschreiben. Die hier vorgestellte Arbeit verwendet und entwickelt eine Methodik weiter, die sich dieser Herausforderung stellt: Additive Bayesianische Netze (ABN). ABN ist ein graphisches Modell, dass durch die Darstellung der gemeinsamen Wahrscheinlichkeitsverteilung die üblichen generalisierten linearen Modelle (GLM) ausweitet auf multiple abhängige Variablen.

Die PhD Arbeit besteht aus vier Teilen. In den ersten beiden Teilen wird die praktische Anwendung von ABN anhand von zwei veterinärepidemiologischen Fallstudien dargestellt. Hierbei wird der zusätzliche Nutzen durch ABN im Vergleich zu klassischen Verfahren deutlich. Die ausgewerteten multivariaten Daten weisen hauptsächlich binäre, aber auch kontinuierliche und Poisson Datenformate auf. Ziel der ersten Studie war es, vergleichend ausgewertet mit ABN und GLM, Risikofaktoren für eine Infektion mit Leptospira interrogans sv pomona zu bestimmen. Dass persönliche Schutzausrüstung die Odds einer Infektion erhöht, also nicht schützt, wurde nur in der Auswertung mit ABN deutlich. Grund hierfür ist die Möglichkeit in ABN die Abhängigkeiten zwischen allen Variablen zu berücksichtigen. Die zweite Fallstudie beschäftigt sich mit der Einstellung von Tierärzten gegenüber der Euthanasie in der Kleintierpraxis und Prädiktoren wie zum Beispiel Alter und Geschlecht. Mit klassischen Verfahren ist es schwierig, die Effekte von Alter und Geschlecht in Beobachtungsstudien getrennt zu schätzen, da die jüngeren Tierärzte mehrheitlich weiblich und die älteren männlich sind. Auch hier erwies sich die Anwendung von ABN, aufgrund der Möglichkeit komplexe Abhängigkeiten zwischen verschiedenen Variablen darzustellen, als vorteilhaft. Ebenfalls nur durch ABN wurde die Bedeutung der Arbeit in einem Team deutlich: diese Variable wies die höchste Anzahl an Verknüpfungen zu allen anderen Variablen auf und unterstreicht die unterstützende Rolle eines Teams in stressvollen Situationen. Die Zuverlässigkeit der ABN Modelle wurde durch ein parametrisches Bootstrapverfahren mittels Markov Chain Monte Carlo (MCMC) mithilfe der Software JAGS überprüft. Der dritte Teil der PhD Arbeit beinhaltet Anpassung und Verbesserung einer Software zum Lernen und Anpassen von ABN Modellen: dem R Paket (ABN). Dies beinhaltete die Modifikation von Funktionen für die graphische Darstellung und die entsprechende Dokumentation. Der Höhepunkt dieser PhD Arbeit liegt im Erkenntnisgewinn der ABN zugrundeliegenden Theorie. Hierbei sind zwei Herausforderungen im Zusammenhang mit der Bayesianischen Modellauswahl herauszustreichen: Die Spezifikation der Parameterprior und die Berechnung der resultierenden Posteriorwahrscheinlichkeiten mittels marginalem Likelihood. Ein geeigneter konjugierter Parameterprior für ABN, der die Dirichlet-Dichte für additive Parameter generalisiert, wird vorgestellt. Dieser Prior erfüllt die erwünschte Eigenschaft der Unabhängigkeit der Bayesianischen Netze und überwindet das Problem der kompletten Datenseparation, die mit anderen ausgewählten Priors vorkommen kann. Weiterhin wurde eine analytische Lösung der marginalen Likelihood gefunden, die ohne Laplace Approximation oder MCMC Methoden angewendet werden kann. Nachgewiesen wurde ebenfalls die Score Äquivalenzeigenschaft, dass äquivalente Netze die gleiche Scorefunktion erlangen. Durch die praktische Anwendung in zwei veterinärepidemiologischen Studien, die Anpassung einer ABN-Software und einer vereinfachten Berechnungsmöglichkeit der marginalen Likelihood trägt diese PhD Arbeit zu einer Weiterentwicklung der ABN-Methodik bei.

Abstract

Veterinary epidemiology, one of the multifaceted applications of statistics, primarily aims to investigate hypothesized relationships between covariates or predictors of interest and one, or more, outcome variables. Commonly, the biological processes, which generated the data, are extremely complex, resulting in multiple dependencies between explanatory and response variables. Standard epidemiological and statistical approaches have shown a limited ability to sufficiently describe such inter-dependent multivariate connections. The following work extends and improves a methodology that addresses these issues: additive Bayesian networks (ABNs). ABNs are types of graphical model that extend the usual Generalized Linear Model (GLM) to multiple dependent variables through the representation of their joint probability.

The PhD thesis consists of four parts. The work begins with the presentation of the commonly 'used' ABN methodology in veterinary epidemiology. Two relevant case studies are presented, giving evidence that ABN models offer added value compared to existing standard statistical and epidemiological methods, i.e., GLM. The multivariate data analyzed are mainly binary, but also continuous and count data. The objective of the first case study was to identify factors associated with Leptospira interrogans sv Pomona infection by exploring the advantages and disadvantages of the two methodologies. Thanks to ABN's capacity to model the relationships between all the variables, the results prove that personal protective gears increased the odds of infection, hence they are in fact not protective. This information was not obtained when the data were analyzed only with GLM. The second case study examines the attitudes of Austrian veterinarians towards euthanasia of small animals. Association between gender and age with views on euthanasia have been found. ABN methodology helped to disentangle the role of gender in relation with age, mainly young females working in small animal practices were influencing the outcome. These features were revealed by ABN due to its ability to capture the natural complexity of data more effectively. Evidence on the importance of the number of veterinarians working together was demonstrated considering the highest number of links, in ABN models, to others variables. This highlights the supporting role of a team in stressful situations. To ensure robustness and reliability of ABN models a parametric bootstrapping approach was implemented, using a Markov Chain Monte Carlo (MCMC) technique in the software JAGS. The third part consists of the update and improvement of a software for fitting and learning ABN models: the R package abn. Modifications of functions, more related to the model graphical representation, were implemented and the documentations related to the R package entirely restructured and rewritten. The final part of this work relies on an improvement related to the underlying theory for ABN models. Two main challenges posed by Bayesian model selection have been addressed: the specification of parameter priors and the computation of the resulting posterior model probabilities via the marginal likelihood. A suitable conjugate prior for ABN which generalizes the Dirichlet density for additive parameters has been introduced. This prior satisfies the desirable independence assumptions for Bayesian networks and overcomes the issue of complete data separation occurring with previous prior choices. Furthermore, an analytic expression for the marginal likelihood was found, which avoids using the Laplace Approximation or MCMC method. Then, the score equivalence property, i.e., equivalent networks get the same score function, has been shown. This work contributes to a better promotion of ABN methodology by illustrating their practical application to veterinary epidemiology, by improving software useful to deal with these models and by gaining better knowledge of the posterior density and an easier computation of the marginal likelihood.

Thesis outline

Notation

Introduction

Thesis Summary

References

Paper I: **Comparison between Generalized Linear Modelling and Additive Bayesian Network. Identification of Factors associated with the Incidence of Antibodies against Leptospira interrogans sv Pomona in Meat Workers in New Zealand**

Marta Pittavino & Anou Dreyfus*, Cord Heuer, Jackie Benschop, Peter Wilson, Julie Collins-Emerson, Paul Torgerson, Reinhard Furrer*

*: joint first authorship.

Paper submitted to *BMC Veterinary Research*.

Paper II: **Attitudes of Austrian veterinarians towards euthanasia in small animal practice: impacts of age and gender on views on euthanasia**

Sonja Hartnack, Svenja Springer, Marta Pittavino & Herwig Grimm

Paper published on *BMC Veterinary Research*.

Paper III: **abn: an R package for modelling multivariate data using additive Bayesian networks**

Marta Pittavino, Fraser Lewis & Reinhard Furrer

Manual for the R-package *abn* published on the Comprehensive R Archive Network (CRAN).

Paper IV: **Conjugate Priors for Additive Bayesian Networks**

Marta Pittavino & Reinhard Furrer

Paper under revision for submission to *Bayesian Analysis*.

Curriculum Vitae

Notation

	Random variables
$\boldsymbol{X}$	Random vector, set of random variables or corresponding sets of nodes
$X_j \in \boldsymbol{X};\ \ X_1, \ldots, X_n$	Random variables or their corresponding sets of nodes
$X_j = x_j^s$	The variable X_j is in the state s
$\mathcal{D} = \{x_{1\cdot}, \ldots, x_{m\cdot}\}$	A data set, a set of observations i, where $x_{i\cdot}$ is a complete assignment to the variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$
P	The joint probability distribution of all the variables in $\boldsymbol{X}$
$P(x_j \mid x_k)$	The probability that $X_j = x_j$ given $X_k = x_k$, where $k \neq j$ (also used to describe a probability density, probability distribution)
$\mathbb{E}$	The expected value of a random variable X_j

	Bayesian networks
$\mathcal{S} = (V, E)$	A Bayesian network structure (a directed acyclic graph)
$\mathcal{B} = (\mathcal{S}, \boldsymbol{\theta}_{\mathcal{B}})$	A Bayesian network model
V	$= \{1, \ldots, n\}$ node (vertex) set
$j \in V$	node (vertex)
E	directed edge set
$e \in E$	directed edge
$op\,(\mathcal{S}, E)$	The single edge operation E on the structure $\mathcal{S}$
$\Delta\left(X_i \rightarrow X_j\right)$	Difference from an edge i to an edge j.
$P(\mathcal{D} \mid \mathcal{S})$	The total marginal likelihood of the structure given the data
$\mathbf{Pa}_j$	The variables or nodes corresponding to the parents of node X_j in a Bayesian network structure, $\dim(\mathbf{Pa}_j) = P_j$
$\mathbf{pa}_j$	A configuration of the variables $\mathbf{Pa}_j$, $\dim(\mathbf{pa}_j) = C_j$
S_j	The number of states of discrete variable X_j
C_j	The number of configurations of $\mathbf{Pa}_j$
θ_{jcs}	The binary parameter corresponding to the probability $P(X_j = x_j^s \mid \mathbf{Pa}_j = \mathbf{pa}_j^c) \equiv P(X_j = s \mid \mathbf{Pa}_j = c)$
$\boldsymbol{\theta}_{\mathcal{B}}$	$= \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n\}$
$\boldsymbol{\theta}_j$	$= \{\boldsymbol{\theta}_{j1}, \ldots, \boldsymbol{\theta}_{jC_j}\}$
$\boldsymbol{\theta}_{jc}$	$= \{\theta_{jc1}, \ldots, \theta_{jcS_j}\}$
δ	An equivalent sample size
δ_{jcs}	The Dirichlet hyperparamer corresponding to θ_{jcs}
δ_{jc}	$= \sum_{s=1}^{S_j} \delta_{jcs}$
$\mathcal{C}_{jcs}$	The number of cases in data set $\mathcal{D}$ where $X_j = x_j^s$ and $\mathbf{Pa}_j = \mathbf{pa}_j^c$
$\mathcal{C}_{jc}$	$= \sum_{s=1}^{S_j} \mathcal{C}_{jcs}$

	Additive Bayesian networks
$\mathcal{A} = (\mathcal{S}, \boldsymbol{\beta}_{\mathcal{A}})$	Additive Bayesian network model
$\boldsymbol{\beta}_{\mathcal{A}}$	$= \{\beta_1, \ldots, \beta_n\}$
$\boldsymbol{\beta}_j$	$= \{\beta_{j1}, \ldots, \beta_{jC_j}\}$
$\boldsymbol{\beta}_{jc}$	$= \{\beta_{jc1}, \ldots, \beta_{jcS_j}\}$
β_{jcs}	The logit transformed multinomial parameter corresponding to the probability $P(X_j = s \mid \mathbf{Pa}_j = c)$
β_{jcs}	$= \sum_{p=1}^{c} \left[\phi_{j,p-1} + \sum_{p=2,\ p/X_{p-1} \in \mathbf{Pa}_j}^{\min(c,P_j+1)} \phi_{j,p-1} X_{p-1} + \sum_{p=P_j+2}^{c} \phi_{j,p-1} \right]$
$\phi_{j,p-1}$	Logistic regression coefficients
$x_j[i]$	ith case of the response variable X_j
$\boldsymbol{Z}_{ij}$	$(S_j - 1) \times (S_j - 1)C_j$ design matrix, for node j and case i, constructed from $\mathbf{Pa}_j$ and from $\boldsymbol{z}_{ij}^T$
$\boldsymbol{z}_{ij}^T$	$= [z_{ij,1}, \ldots, z_{ij,C_j}]$ for each case i, element of $\boldsymbol{Z}_{ij}$
$z_{ij,c}$	indicator vector with $z_{ij,c} = 1$ if configuration c observed; $z_{ij,c} = 0$ otherwise
α_j	$= \{\alpha_{j,1}, \ldots, \alpha_{j,C_j}\}$, the numerator coefficients for integrand of $P(\mathcal{D}_j \mid \mathcal{S})$, node j, as a function of β_{jcs}
γ_j	$= \{\gamma_{j,1}, \ldots, \gamma_{j,C_j}\}$, the denominator coefficients for integrand of $P(\mathcal{D}_j \mid \mathcal{S})$, node j, as a function of β_{jcs}
λ_j	$= \{\lambda_{j,0}, \ldots, \lambda_{j,C_j-1}\}$, the numerator coefficients for integrand of $P(\mathcal{D}_j \mid \mathcal{S})$, node j, as a function of $\phi_{j,c-1}$
ω_j	$= \{\omega_{j,0}, \ldots, \omega_{j,C_j-1}\}$, the numerator coefficients for integrand of $P(\mathcal{D}_j \mid \mathcal{S})$, node j, as a function of $\phi_{j,c-1}$
Pa_{jc+}	Sum of all parents configuration c for node j
X_{jc+}	Sum of all $X_j = 1$ that corresponds to the parents configuration c
$\mathcal{E}$	Exponential family distribution type
E	Approximation error

	Likelihood theory
$\pi(.)$	The parameter prior
$\pi(.\mid\mathcal{D})$	The posterior distribution (or density)
$L(.)$	The likelihood function
$l(.)$	The log-likelihood function
$h(\boldsymbol{\beta}_j)$	$= -\frac{l(\boldsymbol{\beta}_j) + \log \pi(\boldsymbol{\beta}_j)}{m}$ a convex and twice differentiable function
$\boldsymbol{\beta}^*$	The posterior mode or maximum for the function $h(\boldsymbol{\beta}_j)$
$\underset{\beta}{\operatorname{argmax}} f(\beta)$	The argument β which maximizes the function f
max	The maximum of a given function

Introduction

"Statistics starts with a problem, proceeds with the collection of data, continues with the data analysis and finishes with conclusions."
Faraway

Biostatistics, the application of statistics to a wide range of topics in biology and epidemiology, as an independent field of study has been recognized slightly less than a century ago. Biostatistical reasoning and modeling were of critical importance to the foundation theories of modern biology and medicine. In particular, since the 1930s, thanks to the book written by Fisher (1930), statisticians and models built on statistical reasoning have helped to give birth to this fairly new discipline. Statistical models illustrate, in a mathematical way, how deterministic and stochastic components generate observables events. The models are defined by their structure and have model parameters that endow them with flexibility. A large part of statistical models can be associated with regression techniques, a method which has been of paramount importance for much of the scientist throughout the 20th century. Regression methods are largely used in epidemiology, literally from Greek origin $\epsilon\pi\iota$ =on, $\delta\eta\mu o\varsigma$ =population, $\lambda o\gamma o\varsigma$ = study: 'study on the population', one of the main application areas of biometry. An essential part of an epidemiological study is the identification of variables that affects animal and human health. It is a biomedical discipline that deals with the spread of disease and the patterns, causes, and effects of health conditions in defined populations, in a given space and at a given time, in order to identify and analyze the risk factors, variables associated with an increased risk of disease. Specifically, we focus on studies which primarily aims to investigate hypothesized relationships between covariates of interest and one, or more, outcome variables. Commonly, the biological processes, which generated the data, are extremely complex, resulting in multiple dependencies between explanatory and response variables. Standard epidemiological and statistical approaches, such as regression methods, have shown a limited ability to sufficiently describe such inter-dependent multivariate connections. In this thesis I develop and improve a methodology that addresses this challenge: additive Bayesian networks (ABNs).

ABN is a form of graphical model that extends the generalized linear model (GLM) to multiple dependent variables, through the representation of their joint probability. The key aspect which distinguish ABN from standard methods, i.e., GLM, is that ABN attempt to model all dependencies between all variables, and so they can potentially separate out variables which are merely correlated with each other, from those which are actually directly dependent. It is extremely important to identify the variables that are directly associated with health status, because they are the natural targets for interventions stategies. While this feature is highly desirable in any statistical modelling methodology, we do not present ABN as a replacement for existing standard approaches, this would be absurd as no single methodology is a universal panacea, but rather it should be viewed as an additional analytical tool, which offer new insights into both existing and new data from complex epidemiological systems.

In Section 1, regression methods are introduced and their link with ABNs is outlined. Bayesian network (BN) and ABN models are described in Section 2. Then, BN's origins and a comparison with Markov networks is presented in Section 3. In Section 4 the process of learning a BN is clarified. Finally, in Section 5, conclusive remarks on the concept of Laplace approximation methods and Markov Chain Monte Carlo simulations are given, followed by a brief outlook on further research developments in Section 6.

1. Regression models

Regression methods are one of the most widely used statistical techniques in epidemiological analyses. These models have been introduced by Fisher in the 1920*s*, bringing an innovation in the statistical thinking for the beginning of the century that it has become standard for statistician nowdays.

Linear regression

The classical regression models assumes that the response variables y_i $(i = 1, \ldots, m)$, which represents the outcomes of some experiments or study, are independent and normally distributed conditional on the covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$, representing the available data, with expectation $\eta_i = \beta_0 + \mathbf{x_i}^T\boldsymbol{\beta}$ and identical variance σ^2. This assumption can be written as:

$$y_i \overset{ind}{\sim} \mathcal{N}(\eta_i, \sigma^2),$$
$$\eta_i = \beta_0 + \mathbf{x_i}^T\boldsymbol{\beta} + \epsilon_i.$$

The intercept term is represented by β_0, while $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ indicates the regression coefficients vector. Let ϵ_i represents the error term and let us assume that $\mathbb{E}(\epsilon_i) = 0$, $\text{Var}(\epsilon_i) = \sigma^2$, $\text{Cov}(\epsilon_j, \epsilon_k) = 0 \ \forall j \neq k$.
The goals of using this type of analysis are mainly the prediction of new observations, the determination of the strength and the identification of a quantitative law (how does y change when x is changed). The predictors can be continuous or categorical or a mixture of both, see Faraway (2005).

Generalized linear regression

In an epidemiology context, the commonly regression methods used are the *generalized linear models* (GLM), see Clayton and Hills (1993); McCullagh (1989), where the normal distribuiton of the response variable is replaced by a member of the exponential family. This includes many important ditribution, such as the Bernoulli, Poisson, binomial, negative-binomial and exponential distributions. GLM can thus be applied to data with binary and count response as well to data with stricly positive continuous response. The response function (or inverse link function) h transforms the linear predictor η_i to the mean $\mathbb{E}(y_i) = \mu_i = h(\eta_i)$, which in turn is mapped to the canonical parameter $\theta_i = (db/d\theta)^{-1}(\mu_i)$ of the distribution. Often the canonical response function $h = db/d\theta$ is used where $\theta_i = \eta_i$. Here $h = db/d\theta$ is the first derivative of the function b as defined in the likelihood contribution

$$p(y_i|\beta_0, \boldsymbol{\beta}_i) \propto \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i}\right\}$$

of the ith observation. The dispersion $\phi_i = \phi/w_i$ can incorporate weights w_i. The variance $Var(y_i) = \phi_i d^2b(\theta_i)/d\theta^2(\theta_i)$ is expressed through the variance function
$v(\mu_i) = d^2b/d\theta^2((db/d\theta)^{-1}(\mu_i))$ as $\text{Var}(y_i) = \phi_i v(\mu_i)$, as in McCullagh (1989).
These two types of regression methods, with one dependent variable and multiple independent variables, are known as *multivariable regression*.

Multivariate regression models

Multivariable regression, for one dependent variable, can be extended to the scenario that more than one response is measured on each sample unit, *multivariate regression*. The regression model can be extended to the situation where we have measured n responses $y_1, y_2, \ldots, y_n$ and the same set of p predictors $x_1, x_2, \ldots, x_p$ on each sample unit, see Alexopoulos (2010).

Each response follows its own regression model:

$$\begin{aligned} y_1 &= \beta_{01} + \beta_{11}x_1 + \cdots + \beta_{p1}x_p + \epsilon_1, \\ y_2 &= \beta_{02} + \beta_{12}x_1 + \cdots + \beta_{p2}x_p + \epsilon_2, \\ &\vdots = \vdots \\ y_n &= \beta_{0n} + \beta_{1n}x_1 + \cdots + \beta_{pn}x_p + \epsilon_n. \end{aligned}$$

The error vector term $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$ has expectation $\mathbf{0}$ and variance matrix $\Sigma_{p\times p}$. The errors associated with different responses on the same sample unit may have different variances and may be correlated.

Suppose we have a sample of size m and the design matrix is denoted with Z has dimension $m \times (p+1)$. Moreover the n m-dimensional vectors of responses and errors, $i = 1, \ldots, n$, are arranged in an $m \times n$ matrix, the multivariate multiple regression model can now be formulated as:

$$\begin{aligned} &Y_{m\times n} = Z_{m\times(p+1)}\boldsymbol{\beta}_{(p+1)\times m} + \boldsymbol{\epsilon}_{m\times n}, \\ &\mathbb{E}(\epsilon_{(i)}) = 0,\ \mathrm{Cov}(\epsilon_{(i)}, \epsilon_{(k)}) = \sigma_{ik}I,\ i,k = 1,2,\ldots,n. \end{aligned}$$

The *multivariate regression* model can be incorporated into an *additive Bayesian network* model, that consists of a directed acyclic graph where each variable (possible response) in the graph consists of a generalized linear model, combined together in a multivariate framework as a result of the factorization of the joint probability distribution. All the details of ABN models, with the corresponding notation, are explained in the Section 2, while in the next paragraph follows a brief explanation of the advantages of multivariate techniques over multivariable one, with some reasons why is worst applying them in Veterinary Epidemiology.

Additive Bayesian networks as a tool for developing veterinary systems epidemiology

Multivariable regression methods have been widely used during the 20th century in veterinary epidemiology, see Dhand *et al.* (2007); Johnson *et al.* (2010); Phillips *et al.* (2010); Schemann *et al.* (2012); Wagner *et al.* (2003); Zeileis *et al.* (2008), even though there is a risk that approaches which do not take into account relationships between covariates, and more outcome variables, may give unreliable results, as the *Yule-Simpson paradox* (Yule 1900; Simpson 1951) shows by Hand *et al.* (1997); Tu *et al.* (2008): "An apparent relationship between variables may disappear or even be reversed when others are taken into account. We would like to be able to analyse the effects of possible confounding factors and hence avoid the incorrect simplifications which result from collapsing across factors and studying marginal distributions". In order to better address these challenging situations, it may be preferable to use a generalization of usual multivariable regression where all random variables are mutually statistically dependent. In other words, a full *multivariate regression* modelling approach which it coincides with the additive Bayesian network method.

ABN modelling is of particular relevance to veterinary systems epidemiology. The term *'systems epidemiology'* is taken into account in several different scenary (Roux, 2007; Lusis *et al.*,

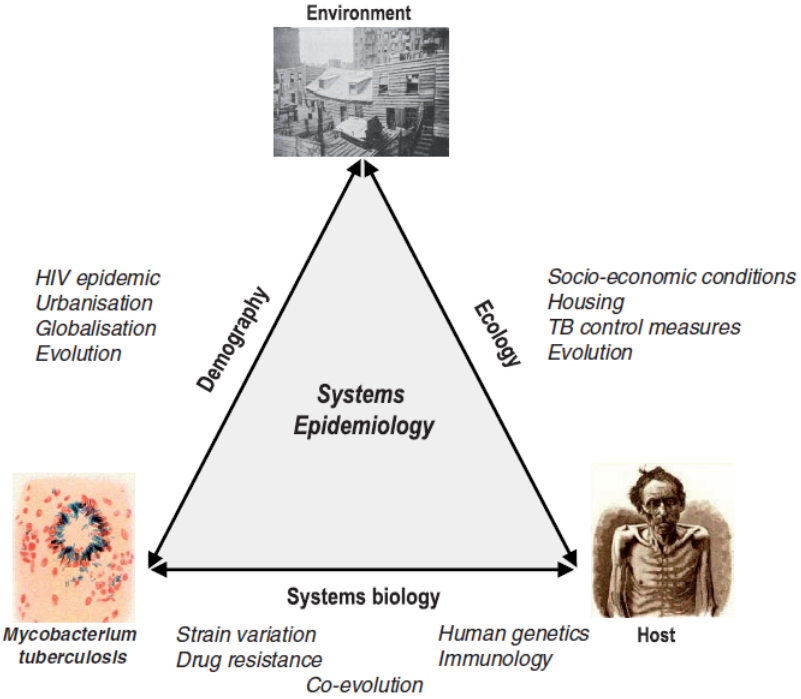


Figure 1.: A 'systems epidemiology' approach to tuberculosis, which integrates demography, ecology and systems biology. Picture reported in Fenner *et al.* (2009) and drawing from Koch R. 'Die Aetiologie der Tuberkulose'. Berliner Klinische Wochenschrift, 1882; Dens of Death.

2008; Fenner *et al.*, 2009; Galea *et al.*, 2010). It indicates a collection of mutually interdependent variables some or all of which can forecast or influence the health outcomes of interest, as shown in Figure 1. This terminology is used with the purpose of highlighting a holistic view of epidemiology is more appropriate for reflecting the true complexity of disease occurence. BN is a natural methodology for developing veterinary systems epidemiology in an analogous way as, in the last decade, it has already happened in biomedical science/systems biology (Jansen *et al.*, 2003; Dojer *et al.*, 2006; Poon *et al.*, 2007a,b; Needham *et al.*, 2007; Poon *et al.*, 2008; Djebbari and Quackenbush, 2008; Lycett *et al.*, 2009; Hodges *et al.*, 2010; Aminian *et al.*, 2010; Kuschner *et al.*, 2010; Milns *et al.*, 2010). Since in the standard epidemiological and statistical approaches only one dependent variable is taken into account, this doesn't allowed to have a general and detailed vision of the presence of disease as part of a complex system and cannot adequately describe such inter-dependent multi-factorial relationships. However, only in the last few years, ABN models have been applied to veterinary epidemiology. To date, we are aware of a general introduction to BN modelling in Lewis *et al.* (2011) and of further applications of ABN, see (Lewis and McCormick, 2012; Lewis, 2012; Sanchez-Vazquez *et al.*, 2012; Firestone *et al.*, 2013; Lewis and Ward, 2013; Ludwig *et al.*, 2013; McCormick *et al.*, 2013; Schemann *et al.*, 2013; Ward and Lewis, 2013; Wilson *et al.*, 2013; Firestone *et al.*, 2014), resulting in dozens of publications. Hence, knowledge of this methodology within the veterinary epidemiology field is extremely rare previous sentence "dozens", possibly because it comes from the machine learning and artificial intelligence literature (Buntine, 1991; Cooper and Herskovits, 1992; Chickering, 1996; Heckerman *et al.*, 1995), opposed to the most familiar statistical literature. This is an important reason why, in *Paper I* and *Paper II*, we present this novel method in comparison with the standard regression approaches, within a veterinary epidemiologicy context.

2. Bayesian networks

Bayesian networks are probabilistic graphical models, see Lauritzen (1996), developed in the late 80's by Lauritzen and Spiegelhalter (1988); Pearl (1988) with an easy and detailed introduction in Jensen (2001). They represent a convergence between statistical methodology and data mining, machine learning, see Cooper and Herskovits (1992); Heckerman *et al.* (1995); Friedman *et al.* (1997); Jensen (2001); Friedman and Koller (2003); Boettcher (2004); Scutari (2007). The *joint, multidimensional* aspect of a BN makes this methodology so attractive for analysis of complex data. These structures result extremely interesting for their ability of expressing in a simple way a set of complex relationship, for that they represent an ideal tool to deal with problems of uncertainty and complexity. A quite recent overview about their different applications is available in Lauritzen (2003).

A BN is a form of graphical model that derives a directed acyclic graph from empirical data, describing the dependency structure between random variables. It provides a compact representation of the joint probability distribution using a combination of graph (the qualitative part) and probability (the quantitative part) theory.
More precisely, a BN model $\mathcal{B}$ for a set of random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ consists of:

- A *directed acyclic graph* (DAG) structure $\mathcal{S} = (V, E)$, where V is a finite set of vertices or nodes and E is a finite set of directed edges between the vertices. A DAG is *acyclic*; hence, the edges in E do not form directed cycles. A random variable X_j corresponds to each node $j \in V = \{1, \ldots, n\}$ in the graph. We do not distinguish between a variable X_j and the corresponding node j.
- A set of parents for a node j is denoted by $\mathbf{Pa}_j$. A vertex j is said to be a *parent* of a node k if the edge set E contains an edge from j to k. P_j indicates the total number of parents for a node j : $\dim(\mathbf{Pa}_j) = P_j$.
- A set of local probability distributions for all variables in the network called $\boldsymbol{\theta}_{\mathcal{B}}$. Each node j, with parent set $\mathbf{Pa}_j$, is parametrized by a local probability distribution: $P(X_j|\mathbf{Pa}_j)$.

Edges represent both *marginal* and *conditional dependencies*. The main role of the network structure is to express the conditional independence relationships among the variables in the model through graphical separation, thus specifying the factorization of the global probability distribution:

$$P(\boldsymbol{X}) = \prod_{j=1}^{n} P(X_j|\mathbf{Pa}_j).$$

We denote a BN model, $\mathcal{B}$, for a set of random variables, $\boldsymbol{X}$, by a pair $\mathcal{B} = (\mathcal{S}, \boldsymbol{\theta}_{\mathcal{B}})$. The DAG defines the *structure* $\mathcal{S}$, and $\boldsymbol{\theta}_{\mathcal{B}}$ the *parametrization* of the model $\mathcal{B}$. In order to specify a $\mathcal{B}$ for $\boldsymbol{X}$, we must therefore specify a DAG structure and a set of local probability distributions.
Figure 2 shows an example of $\mathcal{B}$ for five random variables; the joint probability distribution can be factorized into five factors, one for each random variable conditioned on its parents: $P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2)P(X_3|\mathbf{Pa}_3 = \{X_1, X_2\})P(X_4|\mathbf{Pa}_4 = \{X_3\})P(X_5|\mathbf{Pa}_5 = \{X_3\})$.

$P(X_1, X_2, X_3, X_4, X_5) =$
$P(X_1)P(X_2)$
$P(X_3|\mathbf{Pa}_3 = \{X_1, X_2\})$
$P(X_4|\mathbf{Pa}_4 = \{X_3\})P(X_5|\mathbf{Pa}_5 = \{X_3\})$

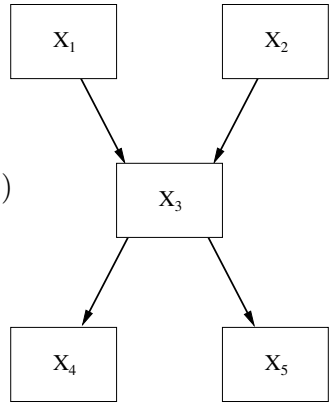


Figure 2.: A Bayesian network model $\mathcal{B}$ for five random variables.

2.1. Additive Bayesian networks

In order to introduce an additive Bayesian Network (ABN) model $\mathcal{A}$, some further notation is needed.

Let S_j be the number of states of the variable X_j, and $s = \{1, \ldots, S_j\}$ the corresponding set of indexes. Let $C_j = \prod_{p:\, X_p \in \mathbf{Pa}_j} S_p$ be the number of configurations of $\mathbf{Pa}_j$ and $c = \{1, \ldots, C_j\}$ indicates the corresponding set of indexes for the different parents configurations of $\mathbf{Pa}_j$. Let $X_j = s$ indicate the possible observations for X_j. Hence, let $P(X_j = s|\mathbf{Pa}_j = c)$ be the probability that $X_j = s$, given the c-th parent configuration of $\mathbf{Pa}_j$, denoted by the multinomial parameter θ_{jcs}. Therefore, the following notation is used:

$$\boldsymbol{\theta}_{jc} = \bigcup_{s=1}^{S_j} \{\theta_{jcs}\},$$
$$\boldsymbol{\theta}_j = \bigcup_{c=1}^{C_j} \{\boldsymbol{\theta}_{jc}\},$$
$$\boldsymbol{\theta}_{\mathcal{B}} = \bigcup_{j=1}^{n} \{\boldsymbol{\theta}_j\}.$$

This means that $\boldsymbol{\theta}_{jc}$ denotes the set of local probability distributions associated with a node j, its parent configuration c and node states s. $\boldsymbol{\theta}_j$ denotes the set of all parameters associated with a node j and its parent configuration c. $\boldsymbol{\theta}_{\mathcal{B}}$ denotes the set of local probability distributions for all variables in the Bayesian network $\mathcal{B}$. All the local probability distributions are unrestricted, discrete distributions with $P(X_j = s|\mathbf{Pa}_j = c) \geq 0 \ \forall j \in V$. Then $\sum_{s=1}^{S_j} \theta_{jcs} = 1$ and $0 \leq \theta_{jcs} \leq 1$. Using this parametrization, the joint probability distribution factorizes into:

$$P(\mathbf{X}|\boldsymbol{\theta}_{\mathcal{B}}, \mathcal{S}) = \prod_{j=1}^{n} P(X_j|\mathbf{Pa}_j = c, \boldsymbol{\theta}_{jc}).$$

It is now possible to introduce an ABN model $\mathcal{A}$. In this work, we are going to refer to additive Bayesian networks with the abbreviation ABN and the notation $\mathcal{A}$ used interchangeably.

An additive Bayesian network $\mathcal{A}$ consists of a Bayesian network $\mathcal{B}$ that generalizes the multi-

nomial logistic regression model $\mathcal{M}$, introduced by Rijmen (2008). The multinomial logistic regression model $\mathcal{M}$ can be integrated into a Bayesian network $\mathcal{B}$ by modelling each conditional probability table $P(X_j|\mathbf{Pa}_j) = \theta_{jcs}$ of a particular Bayesian network $\mathcal{B}$ via a multinomial logistic regression model, where X_j is progressively the outcome variable, and the design matrix $\boldsymbol{Z}_{ij}$ is constructed from $\mathbf{Pa}_j$.

An additive Bayesian network model $\mathcal{A}$ without restrictions on the conditional probability tables is obtained by constructing $\boldsymbol{Z}_{ij}$ from $\mathbf{Pa}_j$ as follows. For each possible configuration c on $\mathbf{Pa}_j$, $c = 1, \ldots, C_j$, a dummy variable is defined. For each observation i, the covariate vector $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijC_j})^T$ is defined as an indicator vector with $z_{ijc} = 1$ if the configuration c is observed, and $z_{ijc} = 0$ if not. The $(S_j - 1) \times (S_j - 1)C_j$ design matrix $\boldsymbol{Z}_{ij}$ is constructed from $\boldsymbol{z}_{ij}$ and $\boldsymbol{\beta}_j = (\beta_{j11}, \ldots, \beta_{jcs}, \ldots, \beta_{jC_j(S_j-1)})^T$, of dimension $(S_j - 1)C_j \times 1$, is the coefficients vector for the *additive parameters*. Then, the expression for the linear predictor for each observation i is instantiated by:

$$\boldsymbol{\eta}_{ij} = \begin{pmatrix} \eta_{ij1} \\ \vdots \\ \eta_{ij(S_j-1)} \end{pmatrix} = \boldsymbol{Z}_{ij}\boldsymbol{\beta}_j = \begin{bmatrix} \boldsymbol{z}_{ij}^T & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{z}_{ij}^T \end{bmatrix} \begin{bmatrix} \beta_{j11} \\ \vdots \\ \beta_{jC_j(S_j-1)} \end{bmatrix}.$$

The corresponding conditional probabilities are obtained by applying the inverse of the link function to the linear predictor.

Therefore, we denote an additive Bayesian network model $\mathcal{A}$ for a set of random variables $\boldsymbol{X}$ by a pair $\mathcal{A} = (\mathcal{S}, \boldsymbol{\beta}_\mathcal{A}) = (\mathcal{S}, h(\boldsymbol{\theta}_\mathcal{B}))$, where $h(\boldsymbol{\theta}_\mathcal{B}) = \text{logit}(\boldsymbol{\theta}_\mathcal{B}) = \boldsymbol{\beta}_\mathcal{A}$. The main difference between a $\mathcal{B}$ and a $\mathcal{A}$ is the re-parametrization of the $\boldsymbol{\theta}_\mathcal{B}$ parameters, seen as a function of the *additive parameters* $\boldsymbol{\beta}_\mathcal{A}$. From the definition of the additive Bayesian network model $\mathcal{A}$, a 'transformed' notation can be used to indicate the parameters in an $\mathcal{A}$ resulting from the *logit link transformation function*, with a similar meaning to that of a $\mathcal{B}$ model:

$$\begin{aligned} \boldsymbol{\beta}_\mathcal{A} &= \text{logit}(\boldsymbol{\theta}_\mathcal{B}) = h(\boldsymbol{\theta}_\mathcal{B}) = \bigcup_{j=1}^{n} \{\boldsymbol{\beta}_j\}, \\ \boldsymbol{\beta}_j &= \bigcup_{c=1}^{C_j} \{\boldsymbol{\beta}_{jc}\}, \\ \boldsymbol{\beta}_{jc} &= \bigcup_{s=1}^{S_j} \{\beta_{jcs}\}. \end{aligned}$$

In this thesis, specifically in *Paper IV*, we are interested in specifying networks for random variables $\boldsymbol{X}$ that follow a Bernoulli distribution (with only two states: *binary* variables), as specified in Dai *et al.* (2013). Hence, a special case of the multinomial logistic regression model is treated, namely the binary logistic regression model. Therefore, we are going to describe this specific discrete Bernoulli case, where we have $S_j = 2$ and $C_j = \prod_{p:\, X_p \in \mathbf{Pa}_j} S_p = 2^{P_j}$ as the number of configurations of $\mathbf{Pa}_j$. In particular, each conditional probability table $P(X_j = 1|\mathbf{Pa}_j) = \theta_{jc1}$ from a $\mathcal{B}$ is modelled via a binary logistic regression model. Thus, we get:

$$\theta_{jc1} = \frac{e^{\boldsymbol{Z}_{ij}\boldsymbol{\beta}_j}}{1 + e^{\boldsymbol{Z}_{ij}\boldsymbol{\beta}_j}} = \frac{e^{\beta_{jc1}}}{1 + e^{\beta_{jc1}}} \Rightarrow \beta_{jc1} = h(\theta_{jc1}) = \text{logit}(\theta_{jc1}) = \log\left(\frac{\theta_{jc1}}{1 - \theta_{jc1}}\right).$$

Moreover, in order to better understand this parametrization, a specific example is described. The structure shown in Figure 2 is considered; the related additive Bayesian network model is called $\mathcal{A}_I$ and leads to the additive parameters:

$$\begin{aligned}
&\boldsymbol{\beta}_{\mathcal{A}_I} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\beta}_5, \},\\
&\text{For } j = \{1,2\},\ \boldsymbol{\beta}_j = \{\boldsymbol{\beta}_{j1}\} = \{\beta_{j11} \cup \beta_{j12}\},\\
&\boldsymbol{\beta}_3 = \{\boldsymbol{\beta}_{31}, \boldsymbol{\beta}_{32}, \boldsymbol{\beta}_{33}, \boldsymbol{\beta}_{34}\},\\
&\boldsymbol{\beta}_{31} = \{\beta_{311} \cup \beta_{312}\}\ ,\ \boldsymbol{\beta}_{32} = \{\beta_{321} \cup \beta_{322}\}, \boldsymbol{\beta}_{33} = \{\beta_{331} \cup \beta_{332}\}\ ,\ \boldsymbol{\beta}_{34} = \{\beta_{341} \cup \beta_{342}\},\\
&\text{For } j = \{4,5\},\ \boldsymbol{\beta}_j = \{\boldsymbol{\beta}_{j1}, \boldsymbol{\beta}_{j2}\} = \{\beta_{j11} \cup \beta_{j12}, \beta_{j21} \cup \beta_{j22}\}.
\end{aligned}$$

Specifically, the following reparametrization occurs, proceeding with a description of the node that follows an increase in parent's orders:

$$\begin{aligned}
&\text{For } j = \{1,2\}, \beta_{j11} = h(\theta_{111}) = \log\left(\frac{\theta_{111}}{1-\theta_{111}}\right),\ \theta_{111} = p(X_1 = 1 | \mathbf{Pa}_1 = 1 = \varnothing, \boldsymbol{\theta}_{11}),\\
&j = \{4,5\},\ \beta_{j11} = h(\theta_{j11}) = \log\left(\frac{\theta_{j11}}{1-\theta_{j11}}\right),\ \theta_{j11} = p(X_j = 1 | \mathbf{Pa}_j = 1 = \{X_3 = 0\}, \boldsymbol{\theta}_{j1}),\\
&j = \{4,5\},\ \beta_{j21} = h(\theta_{j21}) = \log\left(\frac{\theta_{j21}}{1-\theta_{j21}}\right),\ \theta_{j21} = p(X_j = 1 | \mathbf{Pa}_j = 2 = \{X_3 = 1\}, \boldsymbol{\theta}_{j2}).
\end{aligned}$$

For node $j = 3$ further explanation is provided because two parents are involved, implying more parents configuration that require carefulness:

$$\begin{aligned}
&\beta_{311} = h(\theta_{311}) = \log\left(\frac{\theta_{311}}{1-\theta_{311}}\right),\ \theta_{311} = p(X_3 = 1 | \mathbf{Pa}_3 = 1 = \{X_1 = 0, X_2 = 0\}, \boldsymbol{\theta}_{31}),\\
&\beta_{321} = h(\theta_{321}) = \log\left(\frac{\theta_{321}}{1-\theta_{321}}\right),\ \theta_{321} = p(X_3 = 1 | \mathbf{Pa}_3 = 2 = \{X_1 = 1, X_2 = 0\}, \boldsymbol{\theta}_{32}),\\
&\beta_{331} = h(\theta_{331}) = \log\left(\frac{\theta_{331}}{1-\theta_{331}}\right),\ \theta_{331} = p(X_3 = 1 | \mathbf{Pa}_3 = 3 = \{X_1 = 0, X_2 = 1\}, \boldsymbol{\theta}_{33}),\\
&\beta_{341} = h(\theta_{341}) = \log\left(\frac{\theta_{341}}{1-\theta_{341}}\right),\ \theta_{341} = p(X_3 = 1 | \mathbf{Pa}_3 = 4 = \{X_1 = 1, X_2 = 1\}, \boldsymbol{\theta}_{34}).
\end{aligned}$$

The main novelty for an additive Bayesian network is the change of focus from a parametrization expressed in terms of θ_{jcs} to a corresponding one represented in terms of β_{jcs}. It is a one-to-one transformation from the $\boldsymbol{\theta}_{\mathcal{B}}$ to the $\boldsymbol{\beta}_{\mathcal{A}}$ parameters.

3. Causal networks

The definition of Bayesian networks does not refer to causality, and there is no requirement that the links represent causal impact. However the origin of BNs lies in the *causal networks* framework and they can be defined as causal networks with the strength of the causal links represented as conditional probabilities, under particular assumptions and interventions, see Jensen (2001). Hence, in this section we describe the main features characterizing *causal network* models.

Specifically, a causal network consists then of a set of *variables* and a set of *directed links* (also called *arcs*) between variables; structure mathematically called *directed graph*. In a causal network, a variable represents a set of possible states of affairs, where for state of affair we mean

to know that the variable is in a particular state. Casual networks can be used to check how a change of certainty in one variable could have consequences about certainty of other variables.
Below we list the three possible types of connections (serial, diverging and converging) typical of causal networks and rules for reasoning about relevance in causal networks are given. We then explain how they are linked to the concept of *d-separation* and *Markov blanket*.
In the following subsections, we say that a variable is *instantiated* when its state is known. Moreover, *evidence* about a variable is a statement of the certainties of its states. If the variable is instantiated, it is called *hard evidence*; otherwise, it is called *soft*.

3.1. Serial Connections

We start considering the model in Figure 3. Here we have three variables: A has an influence on B, which then has an influence on C. Obviously, if we have an evidence (i.e., an information known from the beginning) about A, this will influence the certainty of B, which in turn influences the certainty of C. Similarly, making the argument to the contrary, an evidence about C will influence, through B, the certainty of A. On the other hand, if the state of B is known, then the link is blocked, and A and C become independent; mathematically we say that A and C are d-separated given B. From this first example, we note that evidence may be

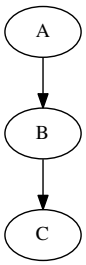


Figure 3.: Serial connection. When B is known, it blocks information between A and C.

transmitted through a serial connection unless the state of the variable in the middle of the connection is istantiated.

3.2. Diverging Connections

We continue with the situation in Figure 4, called a *diverging* connection. In this case information can pass to all the children of A unless the state of A is known. In other words if we know the state of A, the passage of the information through its children is blocked and mathematically we say that B,C,D are d-separated given A. Finally, we can conclude that evidence may pass through a diverging connection unless it is instantiated.

3.3. Converging Connections

A description of the situation in Figure 5, converging connection (known also as *v-structure*), requires more attention. If nothing is known about A, except what may be deduced from knowledge of its parents, then we say that parents are independent: evidence about one of

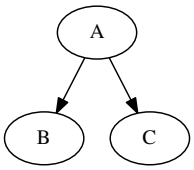


Figure 4.: Diverging connection. If A is instantiated, it blocks communication between its children.

them can not influence the certainties of the others through A. Knowledge of one possible cause of an event does not tell us anything about the other possible causes of that event. Nevertheless, if anything is known about the consequences, then information on one possible

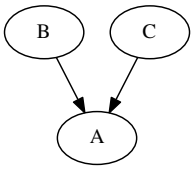


Figure 5.: Converging connection. If A change certainty, communications between its parents are open.

cause may tell us something about the other causes (*explaining away* effect).

Therefore, we conclude that evidence may pass through a converging connection only if either the variable in the connection or one of its descendants has received evidence.

3.4. d-Separation

The three preceding connections cover all ways in which evidence may be transmitted through variables, and following the rules it is possible in a causal network to decide, for any pair of variables, whether they are independent given the evidence entered into the network. The rules can be summarized in the **d-separation** concept, formulated below.

Definition 1. *In a causal network, two distinct variables A and B are **d-separated** ("d" is for directed graph) if for all paths between A and B, there is an intermediate variable V (different from A and B) such that either:*

- *the connection is serial or diverging and V is instantiated or,*
- *the connection is converging, and neither V nor any of V's descendants have received evidence.*

*We say that A and B are **d-connected**, if they are not d-separated.*

One way to check the property of d-separation is by the *Markov blanket*, defined below:

Definition 2. *The **Markov blanket** of a variable A is the set consisting of the parents and the children of A and the variables sharing a child with A.*

When the Markov blanket of A is instantiated, we have a property: A is d-separated from the rest of the network. In Figure 6 we can see an example of a Markov blanket.

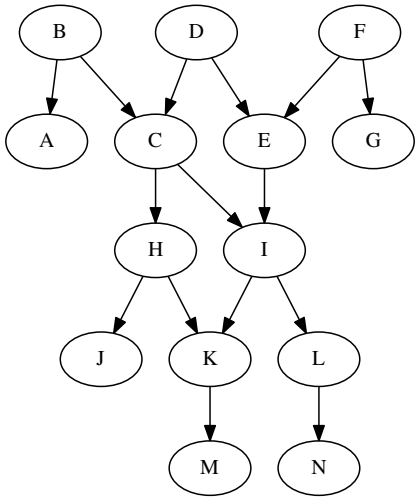


Figure 6.: The Markov blanket for I is $\{C, E, H, K, L\}$. Note that J is not d-separated from I if only I's neighbours are instantiated.

From the definition of d-separation we see that in order to analyze whether two variables, say A and B, are d-separated given hard evidence on a set of variables $\mathcal{C}$ it is necessary to control whether all paths connecting A and B are d-separating paths. An easier way of performing this control, without having to consider the different types of connections, is as follows (see Figure 7).

- Firstly, the so-called *ancestral graph* consisting of A, B and $\mathcal{C}$ is built together with all nodes from which exist a directed path to either A, B and $\mathcal{C}$.
- After, undirected links between each pair of nodes with a common child are inserted and then all links become undirected. The resulting graph is called the *moral graph*.
- The moral graph can be used to control whether A and B are d-separated given $\mathcal{C}$: if all paths connecting A and B intersect $\mathcal{C}$, then A and B are d-separated given $\mathcal{C}$.

3.5. Bayesian and Markov networks

Markov networks (also known as Markov random fields) are, together with Bayesian networks, the subjects of most past and current literature on graphical models. These two classes of graphical models share many common traits. Markov networks differs from Bayesian networks from the underlying structure $\mathcal{S} = (V, E)$, is an undirected graph. All the arcs of $\mathcal{S}$, which are usually called edges in this setting, are undirected; the relationship between the two nodes linked by an edge is symmetric, without the distinction between parents and children that characterizes Bayesian networks.

In other word, this imply that situations like the converging connections can not happen for Markov networks. Hence, in this case instead of having the so-called parents and children, we have the terminology 'neighborhoods'. Furthermore, only situation of conditional depen-

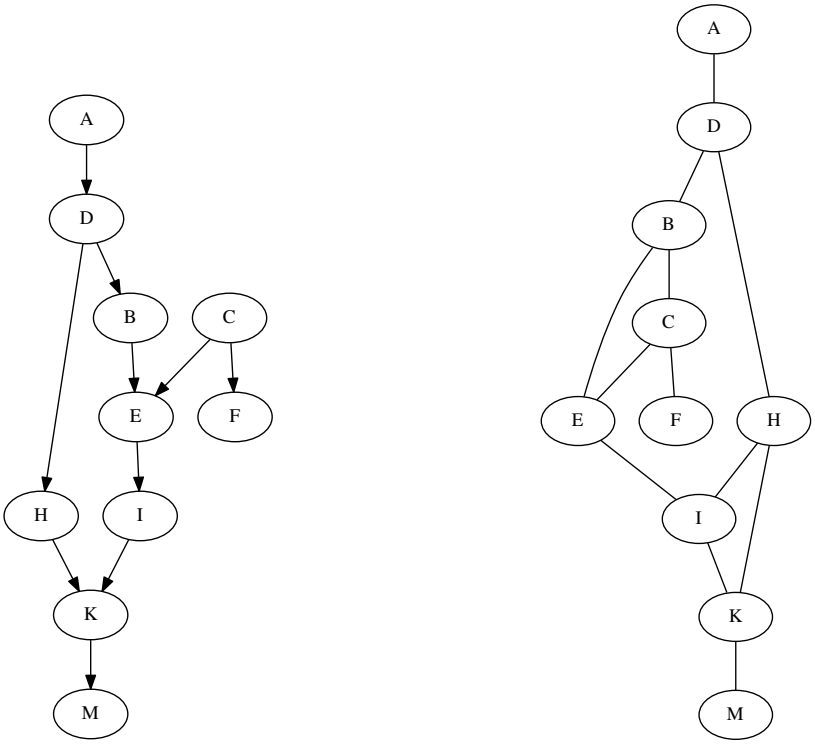


Figure 7.: To test whether A is d-separated from F given evidence on B and M, we first construct the ancestral graph for $\{A, B, F, M\}$ (DAG on the left). Next we add undirected links between pairs of nodes with common child and then the direction is ignored on all links (graph on the right). In the resulting graph we have that the path A-D-H-K-I-E-C-F does not intersect B and M, hence A and F are d-connected given B and M.

dency may arise and not of marginal dependence, like in the v-structure case for Bayesian networks.

4. Learning an additive Bayesian network model

In the Bayesian network literature, see Buntine (1991); Heckerman *et al.* (1995); Heckerman (1998); Friedman and Koller (2003); Boettcher (2004), the parameter estimation and the model selection process are known as *learning*: 1) *parameter learning*: specifying the local probability distributions (model parameters $\beta_{\mathcal{A}}$); and 2) *structure learning*: specifying the DAG structure $\mathcal{S}$. Hence, when constructing an additive Bayesian network model $\mathcal{A}$, two steps need to be considered. Being in a Bayesian framework, given a data set $\mathcal{D}$, we have:

$$P(\mathcal{A}|\mathcal{D}) = \underbrace{P(\beta_{\mathcal{A}}, \mathcal{S}|\mathcal{D})}_{\text{model learning}} = \underbrace{P(\beta_{\mathcal{A}}|\mathcal{S}, \mathcal{D})}_{\text{parameter learning}} \cdot \underbrace{P(\mathcal{S}|\mathcal{D})}_{\text{structure learning}} .$$

Both the *learning* procedures are relevant and necessary in order to understand the final model. They are interconnected and dependent on each other, as explained in Jensen (2001).

Both the two procedures played an important role for the work done in this thesis. The structure learning approach is more linked with the software part, explained in *Paper III*, while the parameter learning is mainly is connected to the choice of the prior and its properties, developed in *Paper IV*. Hence, for this reason, both the procedures will be clarified.

First, the *structure learning* process is explained. The principal score functions and related searching strategies to look for the best model, are presented. The *parameter learning* process is then presented via a list of key assumptions that helps to simplify the most demanding computations.

4.1. Learning the Structure

In this section, the main score functions and searching procedures, used to learn the structure of an ABN network, are presented.

Learning the Structure of Bayesian networks

Consider having a data set $\mathcal{D}$ from an ABN $\mathcal{A}_1$ over the set of variables $\boldsymbol{X}$. The task is now to find a Bayesian network $\mathcal{A}_2$ from the data set $\mathcal{D}$ that is close to $\mathcal{A}_1$. In theory, this can be done by performing parameter learning for all possible structures, and then selecting as candidates those models for that the distribution of $\mathcal{A}_2$ is close to the sample distribution. Unfortunately, by following this simplified approach, three fundamental problems for learning Bayesian networks can arise:

1. The space of all Bayesian network structures is extremely large. It has been shown that the number of different structures (Sloane, 2013), $f(n)$, grows more than exponentially in the number n of nodes, as represented by:

$$f(n) = \sum_{j=1}^{n} (-1)^{j+1} \frac{n!}{(n-j)!n!} 2^{j(n-j)} f(n-1).$$

2. When searching through the network structures, it is likely that the result will be several equally good candidate structures. Since a Bayesian network can represent any distribution across the set of variables over a complete graph, several candidates may appear; this implies that a Bayesian network over a complete graph cannot be the correct answer.
3. There is also the problem of overfitting: a complete graph can represent the sample distribution exactly, but $\mathcal{D}$ could have been sampled from a sparse network. Alternatively, the selected model may be so close to the sample distribution that it also covers the smallest deviances in the distribution of the original model $\mathcal{A}_1$.

Moreover, in Chickering (1996); Chickering *et al.* (2004) are presented one of the latest in a series of results that show that the task of learning Bayesian network structures is NP-hard. Therefore, another searching strategy needs to be followed. The first method for the automated learning of a BN was the method that learned tree-structured models (Chow and Liu, 1968). At present, there are two different types of methods for learning the structure of a BN: *constraint-based* and *score-based*. The first establishes a set of conditional independence statements holding for the data, and uses this set to build a network with graphical separation properties corresponding to the conditional independence properties determined. The second creates some candidate BNs, calculates a score for each candidate, and returns the network with the highest score. Cowell (2001), has shown that, according to often quoted assumptions, constraint-based learning and score-based learning are equivalent.

In the next subsection score-based methods and an exact structure algorithm are presented, because of their link with the developed work in the thesis.

Score-based learning methods

When performing structural learning, the aim is to look for a BN structure that can represent the data set sufficiently well without being overly complex.

Score-based methods assign a number (a *score*) to each BN structure. The score reflects the 'usefulness' of the structure, in other words, how likely it is that the structure could have been used to generate the data set at hand. The task of score-based learning can then be considered to be a search problem: looking for the model structure with the highest score.

Therefore, a BN can be learned from a data set by performing a search of all the DAGs and selecting the one with the highest score. Hence, in order to specify a score-based learning algorithm entirely, two components are needed: a score function and a search procedure.

Bayesian score functions

A good score function should, at least have the following two properties: (a) a balance between the accuracy and the complexity of the structure; and (b) it should be computationally tractable to evaluate.

Moreover, a desirable property for a score function is the *decomposability*, that occurs if it can be expressed as a sum of local scores, one for each node in the data $\mathcal{D}$:

$$score\left(\mathcal{D},\mathcal{S}\right)=\sum_{j=1}^{n}score\left(X_j,\mathbf{Pa}_j,\mathcal{D}\right).$$

An example of a good Bayesian score function, that contains both a term measuring how well the data fits the model and a term that controls model complexity, is the *Bayesian Information Criterion* (BIC) (Bernardo and Smith, 2000).

The *marginal likelihood* is the classical Bayesian approach for measuring the fitness of a candidate BN structure, $\mathcal{S}$. Specifically, we have:

$$P\left(\mathcal{S}|\mathcal{D}\right)=\frac{P\left(\mathcal{S}\right)P\left(\mathcal{D}|\mathcal{S}\right)}{P\left(\mathcal{D}\right)}, \tag{1}$$

where $P\left(\mathcal{D}\right)^{-1}$ is the normalization constant, and is considered a constant because it does not depend on $\mathcal{S}$. From (1), it is easy to see that, in order to score a structure based on its posterior probability given the data, we need two terms, namely the prior probability for the structures $P\left(\mathcal{S}\right)$ and the marginal likelihood of the structure given the data $P\left(\mathcal{D}|\mathcal{S}\right)$. Generally, the prior probability distribution for the structures is chosen in order to be relatively easy to calculate (it is usually assumed that all structures are equally supported, leading to an uninformative structure prior) or with appropriate studies (Scutari, 2013). Therefore, *the main computational problem is the calculation of the marginal likelihood* which is needed to deal with the parameters of the model $\boldsymbol{\beta}_{\mathcal{A}}$:

$$P\left(\mathcal{D}|\mathcal{S}\right)=\int_{\boldsymbol{\beta}_{\mathcal{A}}}P\left(\mathcal{D}|\mathcal{S},\boldsymbol{\beta}_{\mathcal{A}}\right)\pi\left(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}\right)d\boldsymbol{\beta}_{\mathcal{A}}, \tag{2}$$

where $\pi\left(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}\right)$ is the prior probability distribution over the parameters, conditioned on $\mathcal{S}$. The integral in the above equation is over all the parameters and over all the possible Bayesian

networks with the same structure, but with different conditional probability distributions. Intuitively, the marginal likelihood can therefore be interpreted as the probability that the data $\mathcal{D}$ could be generated if the parameters for $\mathcal{S}$ were selected randomly according to the parameter prior $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S})$.

As specified above, the difficult part in the calculation of $P(\mathcal{D}|\mathcal{S})$ is the evaluation of the integral in (2). Fortunately, it has been shown by Cooper and Herskovits (1992); Heckerman *et al.* (1995) that, for a standard Bayesian network model $\mathcal{B}$, the evaluation of this integral can be reduced to a simple counting problem, that can be executed in polynomial time based on three crucial assumptions for the data set $\mathcal{D}$ (A1 to A3) and five regarding the parameters (A4 to A8), that are later clarified in the *Learning the Parameters* section. The first 3 assumptions are important to guarantee that we are working with data that are fully representative of a BN, because their completeness and independence facilitate the factorization of each entry. In particular, we have:

A1. The data set $\mathcal{D}$ is a faithful sample of a Bayesian network.

A2. Observations in the data set $\mathcal{D}$ are independent, given the Bayesian network model.

A3. The data set $\mathcal{D}$ is complete.

In this thesis, we work with a fully observed data set $\mathcal{D} = \{x_{1\cdot}, \ldots, x_{m\cdot}\}$, where each $x_{i\cdot}$ is a set of simultaneous values of the set of variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$. Hence, we consider a data set $\mathcal{D}$ that fulfils assumptions A2 and A3.

In the literature, it has been shown that the BIC score of a model is an asymptotic approximation of the marginal likelihood of that model, and it is equivalent to the minimum description length proposed by Rissanen (1987), and adopted as a decomposable consistent score for Bayesian networks by Lam and Bacchus (1994) and Friedman and Goldszmidt (1998). A Bayesian metric for scoring models was proposed by Cooper and Herskovits (1992), where a search algorithm that performs a greedy search conditioned on a linear ordering of the variables (known as the K2 algorithm) was also suggested. Finally, Friedman and Koller (2003) provided a method for calculating the posterior probability of the absence or presence of individual arcs in the generating net given the data.

The score function for ABN

The *marginal likelihood* has been introduced previously. In this subsection, we describe how it is adapted for ABN models. As a result of the decomposability property of the score function, the total network score, the *marginal likelihood* for an ABN model, can be written as $P(\mathcal{D}|\mathcal{S}) = \prod_{j=1}^{n} P(\mathcal{D}_j|\mathcal{S})$. In a binary logistic additive Bayesian network model $\mathcal{A}$, the network score for node j is given by:

$$P(\mathcal{D}_j|\mathcal{S}) = \int_{\boldsymbol{\beta}_j} \prod_{i=1}^{m} \left(\frac{e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{x_{ij}} \left(\frac{1}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{1 - x_{ij}} \pi(\boldsymbol{\beta}_j|\mathcal{S}) d\boldsymbol{\beta}_j, \tag{3}$$

where $\mathcal{D}_j$ are the observed data at node j, and consist of tuples of $[x_{ij}, \boldsymbol{z}_{ij}^T]$. The parameter vector at node j is represented by $\boldsymbol{\beta}_j$, and has the same length as the possible parent configuration: $\dim(\boldsymbol{\beta}_j) = C_j$. The prior at node j is indicated by $\pi(\boldsymbol{\beta}_j|\mathcal{S})$, and is the unknown quantity that we characterize.

The main difficulty in moving towards an additive model is the computation of the *marginal likelihood*. In fact, additive Bayesian network models $\mathcal{A}$ require considerably more computational time than do standard Bayesian network models $\mathcal{B}$ because, thus far, no work that aims

to simplify the integral has been developed (3) in a similar framework to that of Cooper and Herskovits (1992); Geiger and Heckerman (1994); Heckerman *et al.* (1995); Boettcher (2004).

In *Paper IV*, we show that a simplified expression for (3), based on assumptions A1 to A3 and A4 to A8, listed below, can be also obtained for an ABN model. In order to achieve this goal, a crucial role is played by the prior $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S})$, that is chosen properly.

Search Procedures

Given a score function, the task is to find the highest-scoring Bayesian network structure in the set of all possible network structures. In other words, the task of structural learning is reduced to a searching problem.

Researchers have developed heuristic search strategies that move around in the search space by iteratively performing small changes to the current structure. Specifically, these search methods usually work directly in the space of the Bayesian network structures; hence, each point in this *search space* corresponds to a particular DAG structure, and *search operators* need to be defined. Search operators are used to move from one structure to another, and to determine the neighbourhood of a DAG, namely those DAGs that can be reached in one step from the current DAG. The operators consist of

- *arc addition*: insert a single arc between two nonadjacent nodes;
- *arc deletion*: remove a single arc between two nodes;
- *arc reversal*: reverse the direction of a single arc.

The notation $op(\mathcal{S}, E)$ represents the result of performing the edge operation E on the structure $\mathcal{S}$; in other words $op(\mathcal{S}, E)$ is a DAG that differs from S in terms of one edge only. One important property of these operators is that they only result in local changes to the current structure, i.e., if an arc between X_i and X_j is inserted or deleted, then only the family of X_j is changed, while if an arc between X_i and X_j is reversed, the families of both X_i and X_j are changed. This property is tightly connected to the decomposability of a score function.

If we insert an edge from X_i into X_j, only the local score for X_j will change; thus, when evaluating whether such a move is beneficial, we need only to compute the score difference (or gain) $\Delta(X_i \rightarrow X_j) = score(X_j, \mathbf{Pa}_j \cup \{X_i\}, \mathcal{D}) - score(X_j, \mathbf{Pa}_j, \mathcal{D})$. Of all the possible searching procedures, a simple heuristic approach is the *greedy search*. In the next subsection, the principal steps of a greedy search algorithm will be explained.

Greedy search

The *greedy search* algorithm chooses some initial structures (usually an empty structure, a randomly chosen structure or a prior structure specified by the user), and then calculates the gain for each legal arc operation; by legal, it is meant that the resulting graph must be acyclic. Specifically, a *greedy search algorithm* consists of the following steps:

1. Let S be an initial structure.
2. Repeat
   (a) Calculate $\Delta(E)$ for each legal operation E
      - Let $\Delta^* = \max_E \Delta(E)$ and $E^* = \text{argmax}_E \Delta(E)$.
   (b) If $\Delta^* > 0$, then
      - Set $S = op(S, E^*)$.
3. Until $\Delta^* \leq 0$.

Note that, in the previous algorithm, if the parents of two nodes do not change from one iteration to another, the gain $\Delta\left(X_i \rightarrow X_j\right)$ of any edge operation involving these two nodes will remain unchanged. This gain can therefore be cached for subsequent iterations, so that the calculations can be reused. These properties are important consequences of the decomposition property of the score function.

Using the R package *abn*, it is possible to conduct an heuristic search, through the command *search.hillclimber*, whose functionalities are explained in details in *Paper III*.

However, the limitation of heuristic search algorithms is that it does not guarantee the finding of an optimal global structure, but only of a local optimal structure. Various solutions have been proposed, and an example is the greedy search algorithm with multiple restarts; in other words, after a local maximum is found, the search is reinitialized with a random structure. After this first attempt, the reinitialization is repeated for a fixed number of iterations, and the best structure found, via the entire process is selected. Conversely, a valid alternative solution is to use an exact algorithm (Koivisto and Sood, 2004; Lewis *et al.*, 2014), explained later, in which an optimal global structure is found via a reduction in the number of variables that need to be considered.

Equivalence class search and score equivalence

It can sometimes be advantageous to define the search space using a more abstract representation than that of DAGs. An example is a procedure called the *greedy equivalence search*. The search is based on the observation that data alone cannot be used to discriminate among network structures that represent the same assertions of conditional independence. We start to define this particular form of structures, which are also of particular importance for the work developed in *Paper IV*.

Definition 3. *Two DAG network structures, $\mathcal{S}_1$ and $\mathcal{S}_2$ are **equivalent** if they represent the same independence constraints.*

To better understand the previous definition, an example is provided. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ indicate two DAG network structures, as represented in Figure 8. They are equivalent DAGs because the same independence relations are represented: $P(X_1, X_2) = P(X_1)P(X_2|X_1) = P(X_2)P(X_1|X_2)$. This equivalence is referred to in Heckerman *et al.* (1995) as *Likelihood Equivalence*, which implies that, if $\mathcal{S}_1$ and $\mathcal{S}_2$ are independent equivalent networks that are related to two BN models ($\mathcal{B}_1$ and $\mathcal{B}_2$), they have the same joint likelihood $P(\boldsymbol{X}|\boldsymbol{\theta}_{\mathcal{B}_1}, \mathcal{S}_1) = P(\boldsymbol{X}|\boldsymbol{\theta}_{\mathcal{B}_2}, \mathcal{S}_2)$. The equivalence relation is reflexive, symmetric, and transitive; hence, the relationship defines a collection of *equivalence classes*.

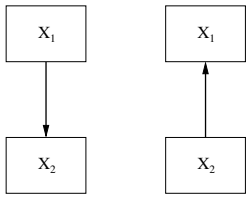


Figure 8.: $\mathcal{S}_1$ and $\mathcal{S}_2$, two equivalent DAG structures.

Definition 4. *A score function that assigns the same score to equivalent structures is said to be **score equivalent**.*

This definition means that it is not possible to distinguish between different DAGs in an *equivalence class* from observations alone. Hence, fitting each DAG to the same dataset should give the same likelihood of observing the data in each model: *equivalent score function*.

BIC represents an example of a score equivalent function. Heckerman *et al.* (1995), considered the specification of prior information, such as that of equivalent network structures (Chickering, 1995), is given the same score. This was the first example of an *equivalent Bayesian score function*. In this work, a parameter prior that leads to *Score Equivalence* for ABN models is introduced.

This property means that, if we have found a particular structure using a score equivalent function, we could just as well select any other structure that is equivalent to the one identified.

In order to move around in the space of equivalence classes, where each point in the search space corresponds to an equivalence class, it is possible to identify some search operators, that are a bit more complex than the ones used in DAG spaces, due to the nature of the search space. These operators define the *neighbourhood* for an equivalence class, which is the set of structures reachable by a single change to the current structure or to one of its equivalents. An *upper neighbourhood* is one consisting of equivalence classes with fewer dependence statements, and a *lower neighbourhood* is one with more dependence statements. All the neighbourhoods are based on the definition of equivalence classes in terms of independence statements.

The two neighbourhoods are defined as the equivalence classes that can be obtained by either adding or deleting a single arc from a DAG in the current equivalence class.

Based on this specification of the search space, the **Greedy Equivalence Search Algorithm** consists of two steps:

1. Start with the equivalence class without dependencies among the variables, and perform a greedy search upwards until a local maximum is reached.
2. Starting from the equivalence class just identified, perform a greedy search downwards until a local maximum is reached.

If the database is sufficiently large, the resulting equivalence class is guaranteed to include the Bayesian network from which the data were generated.

In the context of equivalent structures, greedy search procedures have been proposed by Chickering (2002); Chickering and Meek (2002), and are guaranteed to identify the correct structure when the amount of data becomes large. On the other hand, in the work of Heckerman *et al.* (1995); Geiger and Heckerman (1994); Boettcher (2004) it is shown for the discrete, the Gaussian and the conditional Gaussian case, respectively, that when a specific choice of the parameter priors is made, the marginal likelihood is the same for equivalent network structures, leading to an *Equivalent Network Score* scenario: $P(\mathcal{D}|\mathcal{S}_1) = P(\mathcal{D}|\mathcal{S}_2)$.

Finally, it should be emphasized that, even though another specification of the search space has been made, the general complexity problem present in DAG spaces has unfortunately not been solved: the number of equivalence classes also grows super exponentially in line with the number of variables.

Order based searches

As mentioned before, it is practically impossible to search all possibles DAGs to find the globally best model except in very small problems with only a handful of variables. This is particularly true when dealing with additive models. A fairly tight upper bound on the number of unique DAGs is $n!2^{\binom{n}{2}}$. One way to avoid this problem is not to search across DAGs but search across orders, introduced by Friedman and Koller (2003).

Specifically, a *node ordering* is simply a list of the nodes, say as indexes 1 through n. A given DAG structure is consistent with an ordering if and only if the parents of each node precede their child node in this list. Orderings can be thought of as groups of DAG structures, those structures which are consistent with that particular ordering.

The crucial idea is that by *searching across orderings* the model search space can be reduced from $n!2^{\binom{n}{2}}$ down to only $n!$, which it represents a massive decrease! While this search space is much smaller it is still quickly computationally infeasible for larger problems. However, the drawback for searching across the much smaller space of orders rather than DAGs, it is the consistency of any given DAG with numerous different orders. For example the independence DAG, with no arcs, will be consistent with every possible order. This means that when searching across orders, DAGs which are consistent with more orders will be favoured!

Results from searching across DAGs and searching across orders may, and most likely will, differ. It could or not be seen as a possible problem. A purist might be unhappy with any sort of bias given that searching across DAGs is obviously the "gold standard". On the other hand, the bias is towards simpler models and so that might be considered an acceptable trade-off. Of course the actual magnitude of the bias will be problem specific and dependent on things like sample size.

Often "heuristic" approaches are not fully appreciated, hence it leads to a pragramatic advantage of exact-order based methods. However, the reality is that model selection, which is absolutely essential, cannot generally be done in any other way without resorting to exhaustive searches which are generally not feasible. MCMC is also a heuristic. Even so, providing results using an exact, exhaustive search, using an order based approach can help both with confirming results and also addressing such issues.

In the literature, two different approaches to order based searching have been proposed. The first method proposed for order based searching was through the use of MCMC simulation (Friedman and Koller, 2003). The basic idea is that a search algorithm is constructed which randomly samples across the landscape of orders and it collects information about the degree of statistical support for structural features of interest. The results are the posterior probability for each arc, say. Conceptually, a nice approach.

Later an exact alternative was proposed which visits every order - using the Fast Möbius Transform in order to vastly reduce the computation (Koivisto and Sood, 2004). This method is implemented in the R package *abn* and it can be used through the function *mostprobable*.

4.2. Learning the Parameters

In this section, we assume that the structure of a BN model over the variables $\boldsymbol{X}$ is known, but that the estimates for the conditional probabilities are not known. Hence, the specification of the parameters in the distributions is considered, and the aim is to estimate the parameters of the model: the conditional probabilities.

The first assumption is related to the parameter distribution, called the *Multinomial sample*:

A4. The parameters define a *Multinomial distribution* for each variable X_j and for each configuration of the parents.

When working in a multivariate framework with more than one variable involved, we look at the relationship of the parameters for all the variables in the network. In order to ensure that the parameters can be learned independently we will satisfy two independence properties in order to ensure that all the mathematical properties that enable the computation of the

integral (3) are met. These properties were introduced by Spiegelhalter and Lauritzen (1990), and later expanded by Heckerman *et al.* (1995). They are denoted by *global* and *local parameter independence*. The former means that the parameters for the various variables are independent that, in practice, means that it is possible to modify the tables for the variables independently. The latter means that the parameters are independent for each configuration of the discrete parents. In practice, this means that by having two different configurations, $\mathbf{Pa}_j^1$ and $\mathbf{Pa}_j^2$, the uncertainties on $P\left(X_j|\mathbf{Pa}_j^1\right)$ and on $P\left(X_j|\mathbf{Pa}_j^2\right)$ are independent, and it is possible to modify the parameters for the two distributions independently. If the parameters satisfy the aforementioned independence property, then we have:

A5. *Global parameter independence* : $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}) = \prod_{j=1}^{n} \pi(\boldsymbol{\beta}_j|\mathcal{S})$.

A6. *Local parameter independence* : $\pi(\boldsymbol{\beta}_j|\mathcal{S}) = \prod_{c=1}^{C_j} \pi(\boldsymbol{\beta}_{jc}|\mathcal{S})$, $j = 1, \ldots, n$.

The next assumption is related to the choice of the prior. Specific distributions guarantee a close form expression for the posterior, helping with the computation of (3):

A7. The prior distribution of the parameters is a Dirichlet distribution.

Another important assumption is the *parameter modularity*:

A8. If a node X_j has the same parents in two structures $\mathcal{S}_1$ and $\mathcal{S}_2$ ($\mathbf{Pa}_j^{\mathcal{S}_1} = \mathbf{Pa}_j^{\mathcal{S}_2}$), then $P(\boldsymbol{\beta}_{jc}|\mathcal{S}_1) = P(\boldsymbol{\beta}_{jc}|\mathcal{S}_2)$, $c = 1, \ldots, C_j$.

This means that each discrete distribution has the property whereby, if the joint probability distribution $P(\boldsymbol{X})$ can be factorized according to a structure $\mathcal{S}$, it can also be factorized according to all other structures that represent the same set of conditional independencies as $\mathcal{S}$. The parameters are learned using the *principle of maximum likelihood* (Jensen, 2001; Held and Sabanés Bové, 2014). These five assumptions (A4 to A8), together with A1 to A3, complete the 8 points that allow reducing the computation of the integral (3) to a counting problem, that it is tackled in *Paper IV*.

5. Numerical techniques for Bayesian inference

A potential problem in the application of Bayesian inference to more complex models is the integration necessary to compute the normalizing constant of the posterior distribution in Bayes' theorem, which it is also linked to the computation of the score function seen in the previous section. The calculation of certain characteristics of the posterior distribution such as the posterior mean may require additional numerical integration. In this section we will discuss numerical techniques to perform such integrations. We start describing classical as the Laplace approximation. We then move on to so-called Markov chain Monte Carlo (MCMC) methods, which enable us to avoid explicit integration by simulating from the posterior distribution (Held and Sabanés Bové, 2014).

5.1. Laplace method

The Laplace approximation (LA) is used to calculate characteristics of the posterior distribution of an unknown scalar parameter θ. Application of the LA to this task involves optimization, rather than integration, which is typically much easier. Analytical results are available to study the approximation error in detail.

Let $p(\theta|x)$ denote a posterior distribution. Assume we are interested in

$$\mathbb{E}(g(\theta)|x) = \int g(\theta) \cdot p(\theta|x) d\theta, \tag{4}$$

for a certain positive function $g(\theta)$ on the parameter space. For example, if the parameter θ is positive and $g(\theta) = \theta$ and we obtain the posterior mean $\mathbb{E}(\theta|x)$. To calculate (4) we write

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)},$$

to obtain

$$\mathbb{E}(g(\theta)|x) = \frac{\int g(\theta)p(x|\theta)p(\theta)d\theta}{\int p(x|\theta)p(\theta)d\theta}, \tag{5}$$

i.e., a ratio of two integrals. Suppose now that the data $x = (x_1, \ldots, x_n)$ is a realisation of a random sample $X = (X_1, \ldots, X_n)$, then (5) can be written as

$$\mathbb{E}(g(\theta)|x) = \frac{\int \exp(-nh_g(\theta))d\theta}{\int \exp(-nh(\theta))d\theta}, \tag{6}$$

where

$$\begin{aligned} -nh(\theta) &= \log p(x|\theta) + \log p(\theta) \text{ and} \\ -nh_g(\theta) &= \log g(\theta) + \log p(x|\theta) + \log p(\theta). \end{aligned}$$

Now let $\hat{\theta}$ and $\hat{\theta}_g$ denote the locations of the minima of $h(\theta)$ and $h_g(\theta)$, respectively, i.e., the values where the terms $-nh(\theta)$ and $-nh_g(\theta)$ are maximal. $h(\theta)$ is a convex and twice differentiable function with a maximum at $\theta = \hat{\theta}$. Further let

$$\hat{k} = \frac{d^2h(\hat{\theta})}{d\theta^2}\bigg|_{\theta=\hat{\theta}} \quad \text{and} \quad \hat{k}_g = \frac{d^2h(\hat{\theta})}{d\theta^2}\bigg|_{\theta=\hat{\theta}_g}$$

denote the curvature of $h(\theta)$ and $h_g(\theta)$, respectively, at the corresponding minimum. Separate application of the Laplace approximation (a second-order Taylor expansion of h around $\hat{\theta}$) to both the numerator and denominator gives:

$$\mathbb{E}(g(\theta)|x) = \sqrt{\frac{\hat{k}}{\hat{k}_g}} \exp\{-n(h_g(\hat{\theta}_g) - h(\theta))\}. \tag{7}$$

Tierney and Kadane (1986), have studied the approximation error of (7) in detail. They showed that, although the approximation error of the Laplace approximation in the two integrals in (6) is of order $\mathcal{O}(n^{-1})$, the leading terms in the two errors cancel in (7). As a result, the error of the Laplace approximation of the posterior mean is only of order $\mathcal{O}(n^{-2})$, so smaller. Similar results can be obtained for the posterior variance. The regularity condition required is that the likelihood times the prior be unimodal.

5.2. Monte Carlo integration

In order to calculate characteristics of the posterior distribution we often need to integrate certain functions. We have discussed the Laplace approximation as potentially useful approach to integration when analytic computation is not possible. A second approach are so-called *Monte Carlo methods*.

Assume first, that it is easy to generate independent samples $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}$ from the posterior distribution $p(\boldsymbol{\theta}|x)$ of interest. A Monte Carlo estimate of the posterior mean

$$\mathbb{E}(\boldsymbol{\theta}|x) = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|x) d\boldsymbol{\theta} \tag{8}$$

is then given by

$$\hat{\mathbb{E}}(\boldsymbol{\theta}|x) = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\theta}^{(m)}.$$

The law of large numbers ensures, that this estimate is simulation-consistent, i.e., the estimate converges (almost sure) to the true posterior mean for $M \to \infty$. This approach is called *Monte Carlo integration* and avoids the integration in (8).

5.3. Markov chain Monte Carlo

Application of ordinary Monte Carlo methods is difficult if the unknown parameter is of high dimension. However, Markov chain Monte Carlo (MCMC) methods will then be a useful alternative. The idea is to simulate a Markov chain $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(m)}, \ldots$, which is designed in a way such that it converges to the posterior distribution $p(\boldsymbol{\theta}|x)$. After convergence, one obtains random samples from the target distribution, which can be used to estimate posterior characteristics. However, these samples will typically be dependent, an inherent feature of Markov chains. Similar to rejection sampling, there is great liberty in the actual choice of the underlying random samples of MCMC algorithms. These random samples typically depend on the current state $\boldsymbol{\theta}^{(m)}$ of the Markov chain, here m denotes the current iteration index. The new random variates are now generated from some arbitrary proposal distribution with density $h(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$, say. The so-called Metropolis-Hastings algorithm, the most general MCMC approach, accepts the proposed random number $\boldsymbol{\theta}^*$ from $h(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ as the new state of the Markov chain with probability

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|x)}{p(\boldsymbol{\theta}^{(m)}|x)} \cdot \frac{h(\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta}^*)}{h(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(m)})}\right\},$$

i.e., $\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^*$, otherwise $\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)}$, i.e., $\boldsymbol{\theta}^*$ is rejected. The random number $\boldsymbol{\theta}^*$ is often simply called proposal. The term $\frac{p(\boldsymbol{\theta}^*|x)}{p(\boldsymbol{\theta}^{(m)}|x)}$ is often called *posterior ratio* while $\frac{h(\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta}^*)}{h(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(m)})}$ is the *proposal ratio*. So the acceptance probability is the product of the posterior ratio and the proposal ratio, suitably truncated to the unit interval. Under some regularity conditions one can show that this algorithm converges to the target distribution $p(\boldsymbol{\theta}|x)$, regardless of the specific choice of $h(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$. However, the speed of convergence and the dependence between successive samples will depend heavily on the choice of the proposal distribution. Some special cases of this general algorithm have special names. If the proposal density is symmetric around the current value, i.e., $h(\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta}^*) = h(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(m)})$, one obtains the Metropolis

algorithm with acceptance probability: $\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|x)}{p(\boldsymbol{\theta}^{(m)}|x)}\right\}$.

A special case of this is the so-called random walk proposal , which is defined as the current value $\boldsymbol{\theta}^{(m)}$ plus a random number variate of a zero-centred symmetric distribution. If the proposal density does not depend on $\boldsymbol{\theta}^{(m)}$, the proposal is called independence proposal . Another special case occurs if the acceptance probability α always equals unity. This is clearly the case if $h(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(m)}) = p(\boldsymbol{\theta}^*|x)$, i. e. if the proposal density is equal to the posterior density, the target density. At first sight this appears to be of limited value, as we implicitly assumed that direct sampling from the target density is unavailable. However, α will also equal unity, if a specific component θ_j of $\boldsymbol{\theta}$ is updated by a sample from its full conditional distribution $p(\theta_j|x, \boldsymbol{\theta}_{-j})$, here $\boldsymbol{\theta}_{-j}$ denotes the vector $\boldsymbol{\theta}$ without the component θ_j. Because $p(\theta_j|x, \boldsymbol{\theta}_{-j}) \propto p(\boldsymbol{\theta}|x), j = 1, \ldots, p$, the acceptance probability is still one in this case. Iteratively updating all components of $\boldsymbol{\theta}$ with samples from their corresponding full conditionals is called *Gibbs sampling*. The approach can be adopted to updating multidimensional blocks (not just scalars) of $\boldsymbol{\theta}$ from their respective full conditional distributions. The *Gibbs sampling* has been implemented in the software JAGS, which it has been used both in *Paper I* and *II* to perform MCMC approach for model validation.

The efficiency of the Metropolis-Hastings algorithm depends crucially on the acceptance rate, i.e., the relative frequency of acceptance (typically assessed after convergence of the Markov chain). However, an acceptance rate close to one is not always good. For example, for random walk proposals an acceptance rate too large implies that the proposal density is too close around the current value, so the algorithm needs many small steps to explore the target distribution sufficiently. On the other hand, if the acceptance rate of a random walk proposal is too small, large moves are often proposed, but rarely accepted. In some cases, the algorithm may even get stuck at a specific value and subsequent proposals will get rejected for a large number of iterations. For random walk proposals acceptance rates between 30 and 50% are typically recommended, which can be easily achieved through appropriate choice of the variance of the proposal distribution. Things are different for independence proposals, where a high acceptance rate, which means that the proposal density is close to the target density, is desired.

6. Outlook

This dissertation focuses on additive Bayesian network models. Applications, implementations and methodological improvements of these models are developed and tackled.

This thesis can be considered as an attempt to bring Bayesian networks modelling in the Veterinary Epidemiology literature, with a first description of the technical details and mathematical notation underlying these complex models. Full description of the important statistical features can be found in *Paper IV* with an emphasis of an initial improvement for choice of the prior, the model computation and fitting.

Further developments of ABN models can be performed through the implementation of the prior in *Paper IV* in the R package *abn*. To date, only the multivariate Bernoulli case has been considered, for methodological improvements. However, other multivariate distributions (Gaussian, Poisson), always belonging to the exponential family, can be examined using similar procedures and suggestions as in *Paper IV*. Another interesting, also if very challenging question, is related to the characterization and description of ABN models for repeated measurements, hence where instead of modelling each variable in the graph with a generalized linear model, a generalized linear mixed model (GLMM) is considered.

Thesis Summary

This thesis consists of four chapters, presented in a temporal order reflecting the chronology of the work done. Their content and contribution are briefly summarized below.

Paper I

Comparison between Generalized Linear Modelling and Additive Bayesian Network. Identification of Factors associated with the Incidence of Antibodies against Leptospira interrogans sv Pomona in Meat Workers in New Zealand by Marta Pittavino* & Anou Dreyfus*, Cord Heuer, Jackie Benschop, Peter Wilson, Julie Collins-Emerson, Paul Torgerson, Reinhard Furrer.
*: joint first authorship.

In this paper, we present a comparison between the results from Generalized Linear Modelling (GLM) with those from Additive Bayesian Network (ABN) analysis, used to identify factors associated with Leptospira interrogans sv Pomona (Pomona) infection. The advantages and disadvantages of these two methodologies are explored, to corroborate inferences informing health and safety measures at abattoirs in New Zealand (NZ). In a cohort study in four sheep slaughtering abattoirs in NZ, sera were collected twice a year from 384 meat workers and tested by Microscopic Agglutination with a 91% sensitivity and 94% specificity for Pomona. The study primarily addressed the effect of work position, personal protective equipment (PPE) and non-work related exposures such as hunting on a new infection with Pomona. Directly, significantly associated with Pomona were "Work position" and "Abattoir" (GLM), and "Work position" (ABN). The odds of Pomona infection (OR, [95% CI]) was highest at stunning and hide removal (ABN 41.0, [6.9-1044.2]; GLM 56.9, [6.5-496.6]), followed by removal of intestines, bladder, and kidneys (ABN 30.7, [4.9-788.4]; GLM 28.8, [3.3-252.4]). In ABN analysis of indirectly linked variables, the odds of Pomona infection was 3.0 [1.9-4.8] times higher when wearing a facemask compared to not wearing a facemask and 2.2 [1.3-3.7] times higher when wearing glasses compared to not wearing glasses for workers in the offal and/or pet food area, once adjusted for the effect of work position.

The idea for this paper came from Paul Torgerson and Reinhard Furrer, who thought to exploit the potential of ABN methodology with the Leptospira incidence data in order to further assess the effect of the work position, PPE and related covariates, on a new infection with Pomona. Anou Dreyfus conceived and designed the cohort study, helped by Cord Heuer, Jackie Benschop, Julie Collins-Emerson and Peter Wilson, and collected the data as part of her PhD project. Anou Dreyfus and I analyzed the data, she was responsible for the GLM method part and I conducted all the analysis related to the ABN approach. I wrote a primary summary of the analysis and results achieved, I created all the figures reported in the manuscript. Anou and I drafted the manuscript together. All authors read and approved the final manuscript.

The main contribution of this paper is to highlight the multivariate properties of the ABN methodology, finding new results in the usage of PPE. Hence, facemasks and safety glasses did not show any indication of being protective in GLM, on the contrary, such PPE increased the odds of infection in ABN, but this requires verification using other research methodology. In ABN, all relationships between variables are modeled; hence it has an advantage over GLM due to its capacity to capture the natural complexity of data more effectively.

Paper II

Attitudes of Austrian veterinarians towards euthanasia in small animal practice: impacts of age and gender on views on euthanasia by Sonja Hartnack, Svenja Springer, Marta Pittavino & Herwig Grimm.

Veterinarians commit more often suicide than other members of the general population, but little is known about the contributing factors. It has been hypothesized that occupational stressors including long working hours, heavy workload, poor work-life balance, difficult relations with owners and performing euthanasia act as contributing factors. In this paper we aim to gain insight into the attitudes of Austrian veterinarians towards euthanasia of small animals using both standard regression techniques and additive Bayesian network models. This implied verifying their agreement with euthanasia in specific case scenarios, potentially explained by demographic variables (e.g. gender, age, working in small animal practice, employment, working in a team, numbers of performed euthanasia) and also describing the veterinarians' agreement with a number of different normative and descriptive statements, including also presumed coping strategies. Euthanasia of pets has been described by veterinarians as "the best and the worst" of the profession. The most commonly mentioned ethical dilemmas veterinarians face in small animal practice are: limited treatment options due to financial constraints, euthanizing of healthy animals and owners wishing to continue treatment of terminally ill animals. A questionnaire with 9 euthanasia scenarios, 26 normative and descriptive statements, and demographic data were sent to all members of the Austrian Chamber of Veterinary Surgeons (n=2478). In total, 486 veterinarians fully answered to enable analyses. Initially, responses were explored descriptively before being analyzed using both linear regression and additive Bayesian networks (ABN). The main purpose of using ABN models was to identify joint relationships between demographic variables, statements and each of the 9 euthanasia scenarios. Mutual dependencies between the demographic variables were found, i.e., female compared to male veterinarians worked mostly in small animal practice, and working mostly in small animal practice was linked to performing more euthanasia per month.

The idea for this paper came from Sonja Hartnack, who designed and coordinated the study. Svenja Springer, with the supervision of Herwig Grimm, collected the data in order to include them as a part of her doctoral thesis. Sonja, Svenja and I analyzed and interpreted the data. Svenja conducted the descriptive analysis part, Sonja and I analyzed and interpreted the data, both using regression techniques and ABN models. In particular, I supervised Sonja for the ABN models part, helping with my expertise with the R package 'abn'. I created all the 9 figures, one for each of the 9 euthanasia scenario, present in the paper. All authors helped to draft the manuscript, read and approved the final manuscript.

The main contribution of this paper is the association found between gender and age with views on euthanasia: female veterinarians and veterinarians having worked for less years were more likely to disagree with euthanasia in at least some of the convenience euthanasia scenarios. The paper provide evidence on the importance of the number of veterinarians working together, which was found to be the variable with the highest number of links to other variables, demographic as well as ethical statements. This highlights the role of a team potentially providing support in stressful situations. The results are useful for a better understanding of coping strategies for veterinarians with moral stress due to euthanasia of small animals.

Paper III

abn: an R package for modelling multivariate data using additive Bayesian networks by Marta Pittavino, Fraser Lewis & Reinhard Furrer.

This manual describes the *R* package *abn* which provides functionality for identifying statistical dependencies in complex data using additive Bayesian network models. This methodology is ideally suited for both univariate - one response variable, and multiple explanatory variables - and multivariate analysis, where in both cases all statistical dependencies between all variables in the data are sought. These models comprise of directed acyclic graphs (DAGs) where each node in the graph comprises a generalized linear model, where model search algorithms re used to identify those DAG structures most supported by the data. Currently implemented are models for data comprising of categorical, continuous and/or count variables. Further relevant information about *abn* can be found at: *http://www.r-bayesian-networks.org*.

Fraser Lewis designed, created and coded the *R* package *abn*, together with an initial draft of the manual. Following ideas from Reinhard Furrer, I amended and extended the package. Graphical representation of the models and their related functions have been edited. I entirely restructured and rewritten the manual. All authors read and approved the final version of the vignette, before uploading to CRAN.

The main contribution of this manual is to provide an accessible guide for the users of the *R* package *abn*. Description of the different functionalities of the package are present, including the explanation of a selected complete case study to show all the steps necessary to perform when analyzing data using additive Bayesian networks.

Paper IV

Conjugate Priors for Additive Bayesian Networks by Marta Pittavino & Reinhard Furrer

This paper addresses the parameter learning process of an additive Bayesian network (ABN) model for binary data. When an additive parametrization for Bayesian networks is used, the marginal likelihood (ABN network score) computation is the major objective. In this paper, we introduce a novel conjugate prior distribution for ABN that belongs to a flexible family of conjugate priors called the Diaconis−Ylvisaker conjugate priors. We show that the suggested prior is a generalization of the Dirichlet prior. Moreover, we prove that this prior satisfies the desirable independence assumptions for a parameter prior in DAG models. Hence, it helps to address the goodness of fit calculation. The resulting ABN network score is equal to the Gaussian ordinary hypergeometric function. However, it can be approximated using the Laplace method. We then present a method for selecting the hyperparameter priors in order to have the score equivalence property satisfied. Finally, the priors, the derived methods and the usefulness are illustrated by means of an example of a binary variable network.

The idea for this paper came from Reinhard Furrer, who proposed to address two of the main challenges posed by ABN models linked to the specification of the prior and the computation of the resulting marginal likelihood. Moreover, another aim was to satisfy the score equivalence property, i.e., equivalent networks get the same score function, with the new introduced prior. I did all the computations and worked on properties and theorems reported in the manuscript, always fully supported and adviced by Reinhard Furrer, who constantly contributed with good ideas and new thoughts to the work. I drafted the manuscript. Both authors read and approved the final version of the manuscript.

The main contribution of this paper is the introduction of a prior which is conjugate with respect to ABN models and satisfy all their desirable assumptions, overcoming the issue of complete data separation occurring with the previous suggested prior. Moreover, an analytic expression for the marginal likelihood is provided, which has been compared with the usual Laplace approximation method. The error between the analytic expression and the Laplace approximation is negligible, also for small sample size. The results are useful from the model fitting and parameter estimation perspective of ABN models. The analytic form expression leads to an easier computation of the marginal likelihood, while the conjugacy property helps to have a better knowledge of the posterior density. Moreover, as a novel result a proof of the score equivalence, for ABN models, is provided.

References

Alexopoulos, E. (2010). Introduction to multivariate regression analysis, *Hippokratia* **14**: 23–28.

Aminian, M., Shabbeer, A. and Bennett, K. P. (2010). A conformal Bayesian network for classification of Mycobacterium tuberculosis complex lineages, *BMC Bioinformatics* **11**: S4.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian theory*, Wiley Series In Probability, 389 edn, John Wiley & Sons, Inc., Chichester.

Boettcher, S. G. (2004). *Learning Bayesian networks with mixed variables*, PhD thesis, Aalborg University - Department of Mathematical Sciences.

Buntine, W. (1991). Theory refinement on Bayesian networks, *Uncertainty in artificial intelligence: proceedings of the seventh conference*, pp. 52–60. 7th Conference on Uncertainty in Artificial Intelligence, 1991.

Chickering, D. M. (1995). A transformational characterization of Bayesian networks, *in* I. P. Besnard and S. Hanks (eds), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pp. 87–98.

Chickering, D. M. (1996). Learning Bayesian networks is NP-complete, *in* D. Fisher and H.-J. Lenz (eds), *Learning from Data*, Vol. 112 of *Lecture Notes in Statistics*, Springer New York, pp. 121–130.

Chickering, D. M. (2002). Optimal structure identification with greedy search, *Journal of Machine Learning Research* **3**: 507–554.

Chickering, D. M. and Meek, C. (2002). Finding optimal Bayesian networks, *in* I. A. Darwiche and N. Friedman (eds), *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pp. 94–102.

Chickering, D. M., Heckerman, D. and Meek, C. (2004). Large-sample learning of Bayesian networks is NP-hard, *Journal of Machine Learning Research* **5**: 1287–1330.

Chow, C. and Liu, C. (1968). Approximating discrete probability distributions with dependence trees, *IEEE Transactions On Information Theory* **14**(3): 462–467.

Clayton, D. and Hills, M. (1993). *Statistical models in epidemiology*, Oxford University Press.

Cooper, G. F. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* **9**(4): 309–347.

Cowell, R. G. (2001). Conditions under which conditional independence and scoring methods lead to identical selection of Bayesian network models., *in* I. J. Breese and D. Koller (eds), *Proceedings of the Seventeenth International Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pp. 91–97.

Dai, B., Ding, S. and Wahba, G. (2013). Multivariate Bernoulli distribution, *Bernoulli* **19**(4): 1465–1483.

Dhand, N. K., Eppleston, J., Whittington, R. J. and Toribio, J.-A. L. (2007). Risk factors for ovine johne's disease in infected sheep flocks in australia, *Preventive Veterinary Medicine* **82**(1–2): 51–71.

Djebbari, A. and Quackenbush, J. (2008). Seeded Bayesian networks: constructing genetic networks from microarray data, *BMC Systems Biology* **2**: 57.

Dojer, N., Gambin, A., Mizera, A., Wilczynski, B. and Tiuryn, J. (2006). Applying dynamic Bayesian networks to perturbed gene expression data, *BMC Bioinformatics* **7**: 249.

Faraway, J. J. (2005). *Linear models with R*, Chapman & Hall/Crc Texts in Statistical Science Series.

Fenner, L., Egger, M. and Gagneux, S. (2009). Annie Darwin's death, the evolution of tuberculosis and the need for systems epidemiology, *International Journal of Epidemiology* **38**(6): 1425–1428.

Firestone, S. M., Lewis, F. I., Schemann, K., Ward, M. P., Toribio, J. A. and Dhand, N. K. (2013). Understanding the associations between on-farm biosecurity practice and equine influenza infection during the 2007 outbreak in Australia, *Preventive Veterinary Medicine* **110**(1): 28–36.

Firestone, S. M., Lewis, F. I., Schemann, K., Ward, M. P., Toribio, J. A., Taylor, M. R. and Dhand, N. K. (2014). Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a subsequent outbreak, *Preventive Veterinary Medicine* **116**(3): 243–251.

Fisher, R. A. (1930). *The genetical theory of natural selection*, Oxford University Press, Oxford.

Friedman, N. and Goldszmidt, M. (1998). Learning Bayesian networks with local structure, *Learning in Graphical Models* pp. 421–459.

Friedman, N. and Koller, D. (2003). Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks, *Machine Learning* **50**(1-2): 95–125.

Friedman, N., Geiger, D., Goldszmidt, M., Provan, G., Langley, P. and Smyth, P. (1997). Bayesian network classifiers, *Machine Learning*, pp. 131–163.

Galea, S., Riddle, M. and Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology, *International Journal of Epidemiology* **39**(1): 97–106.

Geiger, D. and Heckerman, D. (1994). Learning Gaussian networks, *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, UAI'94, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 235–243.

Hand, D. J., McConway, K. J. and Stanghellini, E. (1997). Graphical models of applicants for credit, *IMA Journal of Management Mathematics* **8**(2): 143–155.

Heckerman, D. (1998). A tutorial on learning with Bayesian networks, *in* M. Jordan (ed.), *Learning in graphical models*, Vol. 89, pp. 301–354.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data, *Machine Learning* **20**(3): 197–243.

Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference*, Springer, Heidelberg. Likelihood and Bayes.

Hodges, A. P., Dai, D. J., Xiang, Z. S., Woolf, P., Xi, C. W. and He, Y. Q. (2010). Bayesian network expansion identifies new ros and biofilm regulators, *PLOS One* **5**(3): e9513.

Jansen, R., Yu, H. Y., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S. B., Emili, A., Snyder, M., Greenblatt, J. F. and Gerstein, M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data, *Science* **302**(5644): 449–453.

Jensen, F. V. (2001). *Bayesian network and decision graphs*, Springer-Verlag, New York.

Johnson, H. L., Liu, L., Fischer-Walker, C. and Black, R. E. (2010). Estimating the distribution of causes of death among children age 1-59 months in high-mortality countries with incomplete death certification, *International Journal of Epidemiology* **39**(4): 1103–1114.

Koivisto, M. and Sood, K. (2004). Exact Bayesian structure discovery in Bayesian networks, *Journal of Machine Learning Research* **5**: 549–573.

Kuschner, K. W., Malyarenko, D. I., Cooke, W. E., Cazares, L. H., Semmes, O. J. and Tracy, E. R. (2010). A Bayesian network approach to feature selection in mass spectrometry data, *BMC Bioinformatics* **11**: 177.

Lam, W. and Bacchus, F. (1994). Learning Bayesian belief networks. an approach based on the mdl principle, *Computational Intelligence* **10**: 269–293.

Lauritzen, S. L. (1996). *Graphical models*, Oxford Univ Press, New York.

Lauritzen, S. L. (2003). *Some modern applications of graphical models*, Highly Structured Stochastic Systems, Oxford University Press, Oxford, pp. 13–32.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society Series B-methodological* **50**(2): 194–224.

Lewis, F. and Ward, M. (2013). Improving epidemiologic data analyses through multivariate regression modelling, *Emerging Themes in Epidemiology* **10**(1): 4.

Lewis, F. I. (2012). Bayesian networks as a tool for epidemiological systems analysis, *in* S. Sivasundaram (ed.), *9th International Conference On Mathematical Problems In Engineering, Aerospace and Sciences (icnpaa 2012), Amer Inst Physics*, Vol. 1493 of *AIP Conference Proceedings*, pp. 610–617.

Lewis, F. I. and McCormick, B. J. J. (2012). Revealing the complexity of health determinants in resource-poor settings, *American Journal of Epidemiology* **176**(11): 1051–1059.

Lewis, F. I., Brulisauer, F. and Gunn, G. J. (2011). Structure discovery in Bayesian networks: an analytical tool for analysing complex animal health data, *Preventive Veterinary Medicine* **100**(2): 109–115.

Lewis, F. I., Pittavino, M. and Furrer, R. (2014). *abn: data modelling with additive Bayesian networks*. R package version 0.85.

Ludwig, A., Berthiaume, P., Boerlin, P., Gow, S., Léger, D. and Lewis, F. I. (2013). Identifying associations in Escherichia coli antimicrobial resistance patterns using additive Bayesian networks, *Preventive Veterinary Medicine* **110**(1): 64–75.

Lusis, A. J., Attie, A. D. and Reue, K. (2008). Metabolic syndrome: from epidemiology to systems biology, *Nature Reviews Genetics* **9**(11): 819–830.

Lycett, S. J., Ward, M. J., Lewis, F. I., Poon, A. F. Y., Pond, S. L. K. and Brown, A. J. L. (2009). Detection of mammalian virulence determinants in highly pathogenic avian influenza h5n1 viruses: multivariate analysis of published data, *Journal of Virology* **83**(19): 9901–9910.

McCormick, B., Sanchez-Vazquez, M. and Lewis, F. (2013). Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs, *Preventive Veterinary Medicine* **110**(1): 54–63.

McCullagh, Peter; Nelder, J. (1989). *Generalized linear models*, Chapman and Hall.

Milns, I., Beale, C. M. and Smith, V. A. (2010). Revealing ecological networks using Bayesian network inference algorithms, *Ecology* **91**(7): 1892–1899.

Needham, C. J., Bradford, J. R., Bulpitt, A. J. and Westhead, D. R. (2007). A primer on learning in Bayesian networks for computational biology, *PLOS Computational Biology* **3**(8): e129.

Pearl, J. (1988). *Probabilistic reasoning in intelligence systems: networks of plausible inference*, Morgan Kaufmann, Los Altos, CA.

Phillips, G., Tam, C. C., Rodrigues, L. C. and Lopman, B. (2010). Prevalence and characteristics of asymptomatic norovirus infection in the community in England, *Epidemiology and Infection* **138**(10): 1454–1458.

Poon, A. F. Y., Lewis, F. I., Frost, S. D. W. and Pond, S. L. K. (2008). Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models, *Bioinformatics* **24**(17): 1949–1950.

Poon, A. F. Y., Lewis, F. I., Pond, S. L. K. and Frost, S. D. W. (2007a). Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope, *PLOS Computational Biology* **3**(1): e11.

Poon, A. F. Y., Lewis, F. I., Pond, S. L. K. and Frost, S. D. W. (2007b). An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope, *PLOS Computational Biology* **3**(11): e231.

Rijmen, F. (2008). Bayesian networks with a logistic regression model for the conditional probabilities, *International Journal of Approximate Reasoning* **48**(2): 659–666.

Rissanen, J. (1987). Stochastic complexity, *Journal of the Royal Statistical Society, Series B* **49**(3): 223–229. With discussions.

Roux, A. V. D. (2007). From the american college of epidemiology annual meeting 2006 - integrating social and biologic factors in health research: A systems view, *Annals of Epidemiology* **17**(7): 569–574.

Sanchez-Vazquez, M., Nielen, M., Edwards, S., Gunn, G. and Lewis, F. (2012). Identifying associations between pig pathologies using a multi-dimensional machine learning methodology, *BMC Veterinary Research* **8**(1): 151.

Schemann, K., Firestone, S., Taylor, M., Toribio, J.-A., Ward, M. and Dhand, N. (2012). Horse owners'/managers' perceptions about effectiveness of biosecurity measures based on their experiences during the 2007 equine influenza outbreak in Australia, *Preventive Veterinary Medicine* **106**(2): 97–107.

Schemann, K., Lewis, F. I., Firestone, S. M., Ward, M. P., Toribio, J.-A. L. M. L., Taylor, M. R. and Dhand, N. K. (2013). Untangling the complex inter-relationships between horse managers' perceptions of effectiveness of biosecurity practices using Bayesian graphical modelling, *Preventive Veterinary Medicine* **110**(1, SI): 37–44.

Scutari, M. (2007). *Network Bayesiani: un approccio non parametrico basato sull'entropia per la selezione del modello.*, PhD thesis, Università degli Studi di Padova.

Scutari, M. (2013). On the prior and posterior distributions used in graphical modelling, *Bayesian Analysis* **8**(3): 505 – 532.

Sloane, N. J. A. (2013). The on-line encyclopedia of integer sequences, *Annales Mathematicae et Informaticae. International Journal for Mathematics and Computer Science* **41**: 219–234.

Spiegelhalter, D. J. and Lauritzen, S. L. (1990). Sequential updating of conditional probabilities on directed graphical structures, *Networks* **20**(5): 579–605.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**(393): 82–86.

Tu, Y.-K., Gunnell, D. and Gilthorpe, M. S. (2008). Simpson's paradox, Lord's paradox, and suppression effects are the same phenomenon: the reversal paradox, *Emerging themes in epidemiology*.

Wagner, B. A., Salman, M. D., Dargatz, D. A., Morley, P. S., Wittum, T. E. and Keefe, T. J. (2003). Factor analysis of minimum-inhibitory concentrations for Escherichia coli isolated from feedlot cattle to model relationships among antimicrobial-resistance outcomes, *Preventive Veterinary Medicine* **57**(3): 127–139.

Ward, M. P. and Lewis, F. I. (2013). Bayesian graphical modelling: applications in veterinary epidemiology, *Preventive Veterinary Medicine* **110**(1): 1–3.

Wilson, A. J., Ribeiro, R. and Boinas, F. (2013). Use of a Bayesian network model to identify factors associated with the presence of the tick ornithodoros erraticus on pig farms in southern portugal., *Preventive Veterinary Medicine* **110**(1, SI): 45–53.

Zeileis, A., Kleiber, C. and Jackman, S. (2008). Regression models for count data in R, *Journal of Statistical Software* **27**(8): 8.

PAPER I

Comparison between Generalized Linear Modelling and Additive Bayesian Network. Identification of Factors associated with the Incidence of Antibodies against Leptospira interrogans sv Pomona in Meat Workers in New Zealand

Pittavino Marta & Dreyfus Anou, Heuer Cord, Benschop Jackie, Wilson Peter, Collins-Emerson Julie, Torgerson Paul, Furrer Reinhard

Paper submitted to *BMC Veterinary Research*.

Comparison between Generalized Linear Modelling and Additive Bayesian Network; Identification of Factors associated with the Incidence of Antibodies against *Leptospira interrogans* sv *Pomona* in Meat Workers in New Zealand

M. PITTAVINO[1#] and A. DREYFUS[2#], C. HEUER[3], J.BENSCHOP[3], P. WILSON[3], J. COLLINS-EMERSON[3], P. R. TORGERSON[2], R. FURRER[1]*

(1) Institute of Mathematics, University of Zurich, Zurich, Switzerland
(2) Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland
(3) Institute of Veterinary Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

***** *Author to whom correspondence should be addressed: Marta Pittavino, Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, Office +41 44 63 55897, E-Mail: marta.pittavino@math.uzh.ch*

[#] *Joint first authors*

Received xxx 2016; Final revision xxx 2016; Accepted xxx 2016

ABSTRACT

Background: Additive Bayesian Network (ABN) is a graphical model which extends Generalized Linear Modelling (GLM) to multiple dependent variables. The present study compares results from GLM with those from ABN analysis used to identify factors associated with *Leptospira interrogans* sv Pomona (Pomona) infection by exploring the advantages and disadvantages of these two methodologies, to corroborate inferences informing health and safety measures at abattoirs in New Zealand (NZ).

Methodology and Principal Findings: in a cohort study in four sheep slaughtering abattoirs in NZ, sera were collected twice a year from 384 meat workers and tested by Microscopic Agglutination with a 91% sensitivity and 94% specificity for Pomona.

The study primarily addressed the effect of work position, personal protective equipment (PPE) and non-work related exposures such as hunting on a new infection with Pomona. Directly, significantly associated with Pomona were "Work position" and "Abattoir" (GLM), and "Work position" (ABN). The odds of Pomona infection (OR, [95% CI]) was highest at stunning and hide removal (ABN 41.0, [6.9-1044.2]; GLM 56.9, [6.5-496.6]), followed by removal of intestines, bladder, and kidneys (ABN 30.7, [4.9-788.4]; GLM 28.8, [3.3-252.4]). In ABN analysis of indirectly linked variables, the odds of Pomona infection was 3.0 (1.9-4.8) times higher when wearing a facemask compared to not wearing a facemask and 2.2 [1.3-3.7] times higher when wearing glasses compared to not wearing glasses for workers in the offal and/or pet food area, once adjusted for the effect of work position.

Conclusions/Significance: Facemasks and safety glasses did not show any indication of being protective in GLM, on the contrary, such PPE increased the odds of infection in ABN, but this requires verification using other research methodology. In ABN, all relationships between variables are modelled; hence it has an advantage over GLM due to its capacity to capture the natural complexity of data more effectively.

Key words: ABN, GLM, Pomona, Leptospirosis, MCMC, R, JAGS, abattoir workers, risk factors, protective equipments.

AUTHOR SUMMARY

A bacterial zoonotic disease, called “Leptospirosis” burdens New Zealand's (NZ) rural communities with most cases occurring in farmers and meat workers, due to transmission from livestock. A study in four sheep abattoirs had been conducted to see how many meat workers tested positive against one *Leptospira* type, called “Pomona”. Our aim was to identify factors associated with *Leptospira interrogans sv* Pomona infection in meat workers, comparing two different statistical methodologies called Generalized Linear Modelling (GLM) and Additive Bayesian Network (ABN) to inform public health policy and get new insights in statistical modelling.

This research showed that the odds of a Pomona infection were highest at stunning and hide removal in both methodologies. While, facemasks and safety glasses did not show any indication of being protective in GLM, such protective equipment increased the odds of infection in ABN. The latter finding requires further discussion, verifying it with other research methodology.

The strength of this new analytical tool (ABN) is given by the potential to provide far greater insights into both existing and new complex data across all areas of epidemiological research.

INTRODUCTION

The present study compares results from Generalized Linear Modelling (GLM) with those from Additive Bayesian Network (ABN) analysis by exploring the advantages and disadvantages of these two analytical methods while analysing risk factors for occupational leptospirosis in New Zealand (NZ).

A primary objective of many epidemiological studies is to investigate hypothesized relationships between covariates of interest, and one or more outcome variables. To date, a large variety of statistical models is available to analyse epidemiological data (i.e. cross validation criteria, ANOVA), and one of the most popular are GLM (1). Typically, the biological and epidemiological processes, which generated this data, are highly complex, resulting in multiple correlations/dependencies between covariates and also between outcome variables. Unfortunately, standard epidemiological and statistical approaches have a limited ability to adequately describe such inter-dependent multi-factorial relationships. ABN is a form of probabilistic graphical model that extends the usual GLM to multiple dependent variables, through the representation of the joint probability distribution of random variables. It is a statistical model that allows the analysis of complex data and derives a directed acyclic graph (DAG) from empirical data, describing the dependency structure between random variables as opposed to fixed variables in GLM (2, 3). ABN models comprise two reciprocally dependent parts: a DAG and a set of parameters. A DAG is a graphical representation of the joint probability distribution of all random variables in the data. Each node in the DAG is the equivalent to the dependent variable in a GLM regression model. In a graphical statistical model there is no distinction between covariates and an outcome variable. Hence, while a standard GLM focuses on the association between covariates and a single dependent or outcome variable, an ABN is a multivariate (conditional) regression model, analysing the associations between all covariates with all variables being potentially dependent (4). Therefore, in a multifactorial complex disease system, interdependencies between risk factors may be revealed in ABN, that may or may not be discovered in GLM, as the latter imposes a linear relationship between covariates and the outcome (4). By comparing ABN with GLM using identical data, we explore the likely impact of such an analytical difference on the inferences from this study.

The ABN models described here, also if consisting of a DAG, are only related with statistical dependency, and arcs present in such models do not imply any causal relationship. While the identification of a statistical dependency is often a step towards the conclusion of causal mechanisms, it is, however, more demanding to further claim that the given dependency exists within a particular causal web.

In the last decades, Bayesian Network (BN) modelling has been widely used in biomedical science/systems biology (5-13) to analyse multi-dimensional data. However, only in the last few years, it has been applied in the veterinary epidemiology field. A general introduction to BN modelling in veterinary epidemiology is provided by Lewis et al. (14). Further applications of BN to veterinary studies were described by Ward et al., Wilson et al. and Sanchez-Vazquez et al. (15-17). Graphical modelling techniques used to analyse epidemiological data were used by Firestone et al., Schemann et al., Lewis et al., Ludwig et al. and McCormick et al. (2, 18-23). Some of these do not compare results from ABN and GLM (18-20), whereas others do (2, 21-23). In the literature, a detailed comparison of these two methodologies can be found in Lewis et al. (23). However, the aforementioned study was based on simulated (artificial) epidemiological data and differences of results were mainly discussed with graphical outputs (qualitatively), whereas this analysis also compares ORs of parameters directly and indirectly linked to the outcome, focusing on the contrast as well on a quantitative point of view.

Leptospirosis is a zoonotic disease occurring in many mammals and is caused by a bacterium of the genus *Leptospira* spp. Transmission occurs from exposure to urine or aborted tissues of infected animals, either directly or via contact with contaminated water or soil (24). Pathogenic leptospires enter the body through mucous membranes or skin abrasions. In humans, infection with *Leptospira* spp. varies from being sub-clinical (asymptomatic), through a mild to a severe acute disease. A mild form with fever and "influenza-like" symptoms appears to be more common in New Zealand (25). The acute disease is characterized by jaundice, renal failure, hepatic failure, myocarditis, uveitis and/or pulmonary haemorrhage (26, 27).

Among temperate developed countries New Zealand (NZ) has a relatively high incidence of notified human leptospirosis cases of an average annual incidence risk of 2-3 cases per 100,000 population (28, 29). However, under-ascertainment is common and estimated to be 15-65 fold in sheep abattoir workers (25). The three most common serovars in humans are *Leptospira interrogans* sv Pomona (Pomona) and *Leptospira borgpetersenii* sv Hardjo (Hardjo) a*nd Leptospira interrogans* sv Ballum (Ballum) (29). The serovar Pomona is highly prevalent in cattle, deer and sheep in NZ (30-32). Therefore, livestock are a frequent source of human leptospirosis in farmers and meat workers (28) who are most at risk with less than 10% of deer mobs, sheep flocks or beef herds currently vaccinated against leptospirosis (33, 34). Dreyfus et al. (25) found that in 2011 the annual cumulative Pomona incidence risk (%) in sheep abattoir workers was on average 11.9% (95% CI 8.5-14.8%) with a range for four different abattoirs of 8.4-16.4%. The annual risk of confirmed

clinical leptospirosis was 0.78% (3/384, 95% CI 0.20-2.46%) and new infections with Pomona increased the risk of illness with 'influenza-like' symptoms 2.1-fold (= relative risk) (95% CI 1.5-3.0) (25).

This study used the data of the study described above (25) with the following two aims: the first aim was to identify factors associated with Pomona infection in sheep abattoir workers in NZ, with two different methodologies GLM and ABN, in order to untangle the web of causality of human infection with Pomona with a real data set. Specifically, we aimed to test the hypothesis of work position being the main effect variable (most important risk factor), to evaluate the role of personal protective equipment (PPE) and non-work related exposures, such as hunting, home slaughtering and farming. If PPE had a protective effect, it would be a good measure to protect workers. If workers were mainly exposed in their work place and not while hunting or home slaughtering, then it becomes clear where the emphasis on their protection should be. The second and equally important aim was to compare the results between GLM and ABN and discuss advantages and disadvantages of the two statistical analyses.

MATERIALS AND METHODS

Case study

A prospective cohort study amongst voluntarily participating meat workers from four purposively selected sheep abattoirs in the North Island of NZ was conducted. Study methods were described in detail by Dreyfus et al. (25). Participants were blood sampled by certified phlebotomists or nurses and interviewed at the same time by trained researchers using a questionnaire (Supplementary Material). Serum antibodies against Pomona were analysed by the microscopic agglutination test (MAT) at doubling dilutions from 1:24 to 1:1536 as described previously (35). Blood samples and data were collected twice at intervals ranging from 50 – 61 weeks in order to estimate the incidence of new infections with Pomona. Study participants of "Abattoir 1" were sampled the first time between February and April 2008 and the second time in April 2009. All other abattoirs were sampled initially in November 2009 - March 2010, and again in November 2010 - May 2011. Hence, one abattoir ("Abattoir 1") was studied twice in two consecutive years and three abattoirs were studied in the second year once. New infection occurred where a worker sero-converted (a sero-negative worker had a MAT titre increase to equal or higher than 1:48) or had an anamnestic response (a sero-positive worker had a MAT titre increase by two or more dilutions) (25).

Data structure

Serological test results and questionnaire information were entered into an Access© database. The data resulting from the serological test results and interviews comprised of 384 observations across 13 variables, including the outcome variable (Table 1). There were no missing data. To estimate the odds of infection in different work positions this was categorised as follows: the reference category 0 included workers with no or presumed low exposure to organs of the urinary tract or to urine and they worked in the office, "boning" room (where the carcass is cut into pieces), "chillers", "freezers" or "blood processing"; category 1 were workers from areas where organs were handled, such as the "offal"/ "casing"/ "pet food", hide processing positions, also including cleaners, renderers or engineers; category 2 included workers at the middle and end of the slaughter board, where animals were opened, organs removed and carcasses were inspected; category 3 were workers in the yards, where animals were washed and waiting for slaughter and at the beginning of the slaughter board, where animals were stunned, bled and hides were removed.

Workers were asked about the PPE worn for every task in the abattoir. PPE variables were "Facemasks" (mask with movable transparent protective shield covering the whole face), "Safety (= goggles) or normal glasses" and "Gloves on two hands" (made out of latex, or similar material or plastic). They were further asked about the frequency PPE was worn. Frequency category 1 was "always or often" and frequency category 0 "sometimes or never". Further variables of interest were number of months worked during the study and in the three preceding seasons, years worked in an abattoir, whether workers went hunting, were farming, home slaughtering in the study year and the previous three years and personal data such as age and gender. Variable names and their description are presented in Table 1.

In order to analyse the data with ABN methodology, the multicategorical variables had to be split into binary ones. Therefore, the variables "Work position" and "Abattoir" with four and five categories respectively, were split into three ("Work1", "Work2", "Work3") and four ("Plant1", "Plant2", "Plant3" and "Plant4") binary variables. Hence, the variable "Work1" corresponds to "category one" of the variable "Work position", where workers remove the offal or work in the pet food area (Table 1). To address over parametrization and collinearity "Work0" and "Plant0" were omitted from the model.

75 **Table 1: Potential determinants of new infection with *Leptospira interrogans* sv Pomona in sheep abattoir meat**
76 **workers (n=384) in New Zealand. Multinomial variables in generalized linear modelling (GLM) were**
77 **transformed into binary variables in Additive Bayesian Network (ABN) modelling.**

Variables and Categories in		**Positive cases**[4]	**DAG**[3] **label**
GLM	**ABN**		
Work position			
0 Boning, chillers, office	Omitted[2]		
1 Offal removal, pet food	**Work position 1**		**Work1**
	0 Not working in offal removal, pet food	340	
	1 Working in offal removal, pet food	44	
2 Intestines or kidney removal, meat inspection	**Work position 2**		**Work2**
	0 Not removing intestines or kidneys, not inspecting meat	296	
	1 Intestines or kidney removal, meat inspection	88	
3 Yards, stunning, pelting	**Work position 3**		**Work3**
	0 Not working in yards, not stunning or pelting	274	
	1 Working in yards, stunning or pelting	110	
Abattoir			
0 Working in Abattoir 1 (A1)[1] (n=82)	Omitted[2]		
1 Working in Abattoir 1 (A2) (n=135)	**Abattoir 1**		**Plant1**
	0 Not working in Abattoir 1 (A2)	302	
	1 Working in Abattoir 1 (A2)	82	
2 Working in Abattoir 2 (n=68)	**Abattoir 2**		**Plant2**
	0 Not working in Abattoir 2	316	
	1 Working in Abattoir 2	68	
3 Working in Abattoir 3 (n=21)	**Abattoir 3**		**Plant3**
	0 Not working in Abattoir 3	363	
	1 Working in Abattoir 3	21	
4 Working in Abattoir 4 (n=78)	**Abattoir 4**		**Plant4**
	0 Not working in Abattoir 4	306	
	1 Working in Abattoir 4	78	
Gender	**Gender**		**Gender**
0 Female	0 Female	128	
1 Male	1 Male	256	
Hunter of goats, pigs & or deer	**Hunter of goats, pigs & or deer**		**Hunt**
0 No	0 No	355	
1Yes	1Yes	29	
Slaughter of sheep, goats, pigs, beef & or deer at home	**Slaughter of sheep, goats, pigs, beef & or deer at home**		**Kill**
0 No	0 No	320	
1 Yes	1 Yes	64	
Owning a farm with pigs, goats, sheep, beef cattle, alpaca & or deer	**Owning a farm with pigs, goats, sheep, beef cattle, alpaca & or deer**		**Farm**
0 No	0 No	322	
1Yes	1 Yes	62	
Wearing normal or safety glasses	**Wearing normal or safety glasses**		**Glass**
0 Sometimes/ never	0 Sometimes/ never	166	
1 Always/ often	1 Always/ often	218	
Wearing gloves on both hands	**Wearing gloves on both hands**		**Gloves**
0 Sometimes/ never	0 Sometimes/ never	134	
1 Always/ often	1 Always/ often	250	
Wearing a facemask	**Wearing a facemask**		**Mask**
0 Sometimes/ never	0 Sometimes/ never	320	
1 Always/ often	1 Always/ often	64	
Months worked in the meat industry	**Months worked in the meat industry**		**Time**
Continuous	Continuous	[4]217; 9-636	
Age	**Age**		**Age**
Continuous	Continuous	[4]48;19-73	

[1]Sheep Abattoir 1 (A1) is the same as sheep Abattoir 1 (A2), but sampled in subsequent years with 57 of 135 participants being resampled; [2]Omitted due to over parametrization; [3]Directed acyclic graph. [4]For continuous variables the mean and range are given.

Exploratory data analysis (EDA) with correlation matrices, captured using the Spearman's correlation coefficients, and parallel coordinate plots evaluation was conducted on the raw data to test the correlation between predictor variables.

Analysis with GLM

Data were analysed using the software R (36). We used multivariable logistic regression (MLR) to test the hypotheses that work position, hunting, slaughtering at home and farming were risk factors and PPE was a protective factor for new infection with Pomona. We evaluated risk factors and confounding variables by a manual forward stepwise selection in the MLR model, starting with a null model with only an intercept included and then adding one risk factor at a time. A variable was retained if the Likelihood Ratio Test (LRT) was statistically significant at a p-value ≤ 0.05 (= risk factor) or if its presence changed the OR of work position (main effect variable) in the model by more than 15% (= confounder) (37). Interaction between risk factors was tested by LRT. If the LRT was statistically significant ($p \leq 0.05$) the interaction term was retained in the model. The tested interaction terms were "Gender*Wearing gloves", "Work position*Wearing gloves", "Work position*Wearing safety/normal glasses", "Work position*Gender", "Wearing safety/normal glasses*Abattoir" and "Wearing gloves*Abattoir".

Since 57 people from Abattoir 1 participated twice in the study, robust standard errors (SE) were calculated using generalised estimating equations (GEE) as adjustment for clustering due to repeated measurements (38). Results of GEE were compared to the ones without and adjustment kept if SE increased by $\geq 5\%$.

The Hosmer-Lemeshow statistic was used to test the distributional assumption and the Pseudo R-square was used to evaluate the overall model fit. Influential covariate patterns and leverage were examined using described methods (39).

Analysis with ABN

All analyses were conducted using the software R (36) and specifically the R package abn (40) which is maintained by one of the two principal authors and is available from CRAN (cran.r-project.org) with additional documentation and case studies at "http://www.r-bayesian-networks.org/". The resulting networks were manually created with the programme Xfig.

Multicategorical variables were transformed to binary and hence, not presenting real associations, arcs (links, edges) between these were not allowed to be part of the model. Prior distributions were defined: all DAG structures were equally supported by a uniform prior in the absence of any data. A uniform prior was used to guarantee that no structure was preferred over the others, in order to allow a fully data-driven approach. While, uninformative Gaussian priors were applied for the parameters at each node: specifically independent Gaussian priors with mean zero and variance 1000 for the additive terms, equivalent to beta coefficients in a conventional logistic regression, and a diffuse Gamma distribution with shape and scale of 0.001 for the precision, i.e., the inverse of the variance parameter in the Gaussian nodes.

A three step procedure was utilized to determine a robust model for the case study data and to estimate the parameters. The first step was to find an optimal model, represented by the DAG, which is a graphical representation of the joint probability distribution of all the random variables where no cycles exist. The best goodness of fit to the available data was computed using the marginal likelihood method, which is the standard goodness of fit metric in Bayesian modelling and includes an implicit penalty for model complexity. This was estimated using the Laplace approximation at each node (41). The process of identifying an optimal DAG is referred to in the literature as structure learning (42, 43). This was found with an order based exact search method (44), which determines a DAG with goodness of fit being equal to the best possible goodness of fit of any DAG. In order to find the best DAG, the maximum number of parents allowed per node (= number of covariates in each regression model at each node) was increased until the goodness of fit remained constant and thereby identified the same globally optimal DAG. The model selection procedure started from three possible parents per node and then the parent limit increased gradually until ten possible parents per node (Figure 1). A best fitting ABN was identified at the end of this first step, with a maximum number of possible parents per node.

In the second step, the model was adjusted by checking it for over-fitting (45) using Markov chain Monte Carlo (MCMC) simulation implemented in JAGS (‘just another Gibbs sampler’) (45, 46). Simulated

datasets were generated with MCMC as iterations of an identical size as the original one, from the optimal model found in step one. An identical exact search for an optimal model structure was then performed exactly as in the first step, but applied to the bootstrapped data rather than original data. It was repeated 2560 times, a large enough number to get robust results, using the same parent limit per node as the one found in the initial search. Arcs present in less than 50% of the globally optimal DAGs – estimated from the bootstrapped data – were considered not to be robust and removed from the DAG generated in the first step. A threshold of 50% structural support is the usual cut-off in ABN analysis (4). For sensitivity analysis, the arcs coverage after 640 and 1280 simulations were compared. A most robust ABN model fully adjusted for over-fitting was identified at the end of this second step, equivalent to a multivariate GLM. The R package coda (47) was used to evaluate the mixing of MCMC chain. Both visual and statistical techniques have been used with the Gelman and Geweke diagnostics (48).

In the third step of ABN analysis, the marginal posterior log odds ratio and 95% credible intervals were estimated for each parameter from the posterior distribution, expressed by the DAG identified at the second step. Being in a Bayesian statistics framework, the parameters were the maximum likelihood estimates (MLE) based on the joint posterior distribution. With ABN methodology it is possible to evaluate both "direct" and "indirect" relations present in the data. An arc between two variables in the final DAG model is referred to as a "direct" relationship, whereas an "indirect" relationship is defined as two arcs connecting two variables with an intermediate variable. For example, Figure 2 shows variables "Pomona" (= Pomona infection) and "Mask" (= wearing a facemask) being "indirectly" linked through the presence of work position ("Work1", "Work2" and "Work3") variables that are all "directly" linked to "Pomona".

In order to estimate the parameters of "directly" linked variables, a specific function (fitabn) of the R package 'abn' was used. With the latter, it is possible to compute the odds ratio at each node, connected with an arc in the final model. In order to estimate the parameters for associations between "indirectly" dependent variables and new infection with Pomona, a long MCMC (roughly 400000 independent iterations) was performed using JAGS, to simulate values from the joint posterior distribution across all variables (i.e. from the final DAG). Then the empirical odds ratios of interest were computed. Because of the joint nature of the model, these were marginal odds ratios across all variables in the data, hence implicitly included other variables effects. Fixing the values of some specific nodes across the network also allowed the estimation of adjusted odds ratios for previously estimated variables, resulting from the effect of fixing a selected outcome

of a variable and re-calculating the parameter of interest. This step described dependencies between PPE and new infection with Pomona given that a person was working in a specific work position.

At the end of this third step, the marginal posterior odds ratio of the main variables in the analysis and their 95% credibility intervals were obtained. Data and R codes are available as Supplementary Material.

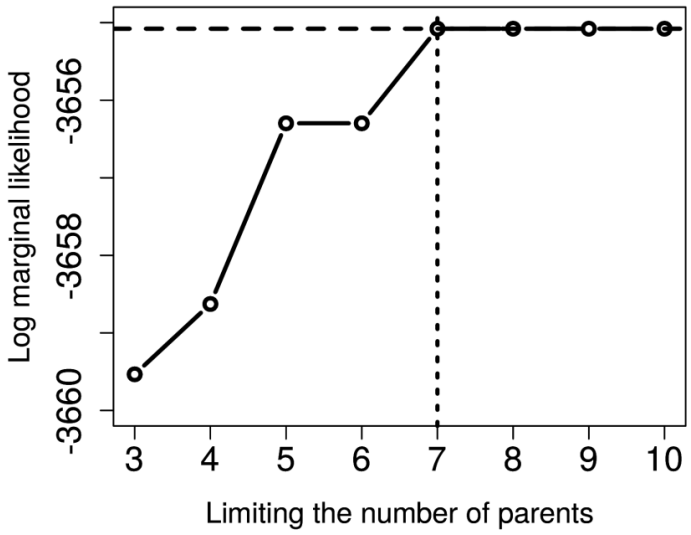


Figure 1: Comparison of goodness of fits (log marginal likelihood) for different parent limits (number of covariates in each regression model at each node), resulting from the first step of model selection in Additive Bayesian Network (ABN) methodology, for data of new infection with *Leptospira interrogans* sv Pomona in abattoir workers processing sheep in New Zealand.

RESULTS

The number of participating workers by abattoir ranged from 21-135 with a total of 384 workers (Table 1). The exploratory data analysis revealed the strongest correlation between the continuous variables "Age" and "Time" (c=0.61), the variables "Work3" and "Gender" (c=0.38) and "Work2" and "Mask" (c=0.33). The variable "Pomona", with 36 positive cases, was mainly linked with variable "Work3" (c=0.2), for all the other variables there was a correlation coefficient < 0.15. A similar pattern in the data was reflected in the results from ABN analysis.

Risk factors for new infection with Pomona analysed by GLM

Statistically significant risk factors in the final GLM model were "Work position" and working in a specific abattoir (Table 2). Compared with the workers in the office, boning room or chillers (reference group), workers in the offal room had 20 times (95% CI 2.6-423.9, p=0.009), workers removing the intestines and kidneys, and meat inspectors had 29 times (95% CI 4.8-562.1; p=0.003), and workers stunning, pelting and working in the yards had 57 times (95% CI 9.5-1109.6; p<0.001) the odds of infection with Pomona. These associations were independent of working in a specific abattoir. Compared with "Abattoir 1/A1", persons resampled at this abattoir ("Abattoir 1/A2") had four times the odds of infection (95% CI 1.3-11.8; p=0.013) compared to the previous year irrespective of work position.

Even though the variables "Home slaughter", "Gender" and "Wearing glasses" were not statistically significant , they were left in the model as potential confounders, as they changed the work position OR by ≥15% (Table 2). None of the other potential risk factors, confounders or interactions was significantly associated with new infection in GLM and did not improve the model fit.

The SE changed by less than 5% with GEE and therefore adjustment for repeated sampling of the same worker in two subsequent years was not required.

Model diagnostics indicated that the data fitted the logit-normal distribution. One outlier was identified, but its removal and collapsing work position categories zero and one did not change any of the significant model coefficients by more than 8% and hence, did not impact on the inferences.

Association between variables analysed by ABN

The resulting best fitting ABN comprised 30 arcs and a maximum number of seven parents, for the variable

"Gender". The MCMC revealed a good mixing of the chain, with no evidence of non-convergence toward the stationary distribution resulting from the Gelman and Geweke diagnostics. After the bootstrap analysis, four of the arcs in the globally optimal DAG were only weakly supported. Therefore the number of arcs was reduced from 30 to 26. Identical results were obtained in the sensitivity analysis, where we started with 2560 bootstraps, which was a large enough number to generate robust results, and then performed half (1280) and a quarter (640) of the bootstrap analyses, suggesting a robust conclusion. The final globally optimal additive Bayesian network model after adjustment for over-fitting is shown in Figure 2. The ABN models considered here are concerned only with statistical dependency, and arc direction in such networks has no epidemiological interpretation. Therefore, the graphical models are presented without arc direction.

In the final ABN model shown in Figure 2, the only variables directly linked to Pomona infection were work positions. More specifically, people working in stunning, pelting and yards had the highest odds of infection, compared to those who were not working in this particular category, as the odds of infection with Pomona was 41.0 (95% CI 6.9-1044.1) times higher than in workers not working in these positions. Workers removing the intestines and kidneys and meat inspectors had 30.7 (95% CI 4.9 -788.4) times the odds of infection with Pomona compared to workers not working in these positions. Workers removing offal and pet food had 18.3 (95% CI 2.2 – 506.7) times the odds of infection with Pomona compared to workers not working in these positions (Table 2). As illustrated in the final DAG, work positions were strongly inter-dependent with PPE. The odds of Pomona infection was 7.0 (95% CI 3.5 – 14.0) times higher when wearing a facemask compared with not wearing a facemask, without adjusting for the effect of work position. The odds of Pomona infection was 3.03 (95% CI 1.9 – 4.8) times higher when wearing a facemask compared to not wearing a facemask for workers in offal and/or pet food, once adjusted for the effect of work position. Similar ORs were found for wearing a facemask and working in work position 2 and 3, as seen in Table 3 and Figure 3. The odds of Pomona infection were 4.6 (95% CI 2.2 – 9.8) times as high when wearing normal or safety glasses compared with not wearing them. However, this value was halved when adjusting for the effect of work positions (Table 3). The odds ratios for the effect of wearing gloves were not statistically significant, as the confidence intervals included "one".

Table 2: Odds ratios (OR) and confidence intervals (CI) of significant covariates and confounders in generalized linear modelling (GLM) (left) and DIRECTLY[1] dependent covariates in Additive Bayesian Network analysis (ABN) (right) for new infection with *Leptospira interrogans* sv Pomona in abattoir workers processing sheep (*n*=384) in New Zealand.

GLM[2] covariates with categories			**ABN covariates** with categories		
	OR	**95% CI**		**OR**	**95% CI[7]**
Work position					
0 Boning, chillers, office	ref		**Omitted[6]**		
1 Offal removal, pet food	20.2	2.6-423.9	**Work position 1**		
			0 Not working in offal removal, pet food		
			1 Working in offal removal, pet food	18.3[5]	2.2-506.7
2 Intestines & kidney removal, meat inspection	28.8	4.8-562.1	**Work position 2**		
			0 Not removing intestines or kidneys, not inspecting meat		
			1 Intestines or kidney removal, meat inspection	30.7	4.9-788.4
3 Yards, stunning, pelting	56.9	9.5-1109.6	**Work position 3**		
			0 Not working in yards, not stunning or pelting		
			1 Working in yards, stunning or pelting	41.0	6.9-1044.2
Abattoir			Indirectly dependent		
1 (A1)[3]	ref				
1 (A2)	3.8	1.3-11.8			
2	1.1	0.1-6.2			
3	2.5	0.8-7.9			
4	0.4	0.1-1.4			
Wearing normal or safety glasses			Indirectly dependent[8]		
No	ref				
Yes	1.3[4]	0.5-3.6			
Gender			Indirectly dependent		
Female	ref				
Male	0.5[4]	0.2-1.4			
Slaughter of sheep, goats, pigs, beef & or deer at home			Indirectly dependent		
No	ref				
Yes	0.5[4]	0.1-1.3			

[1]One arc between variables
[2]The Log likelihood of the nested GLM model was -95.327 compared to the empty model with -119.474;
[3]Sheep Abattoir 1 (A1) is the same as sheep Abattoir 1 (A2), but sampled in a different year with 57 of 135 participants being resampled;
[4]Statistically not significant, but kept in the model due to a confounding effect on the work position variable;
[5] The odds ratios are marginal, i.e., for work position 1, persons working in the offal removal or pet food area are 18.3 times as likely to get infected, than everyone else not working in these positions, taking all the other variables into account;
[6]Omitted due to over parametrization;
[7]ABN methodology does not generate p-values because of the joint mathematical formulation;
[8]Odds ratio in Table 3

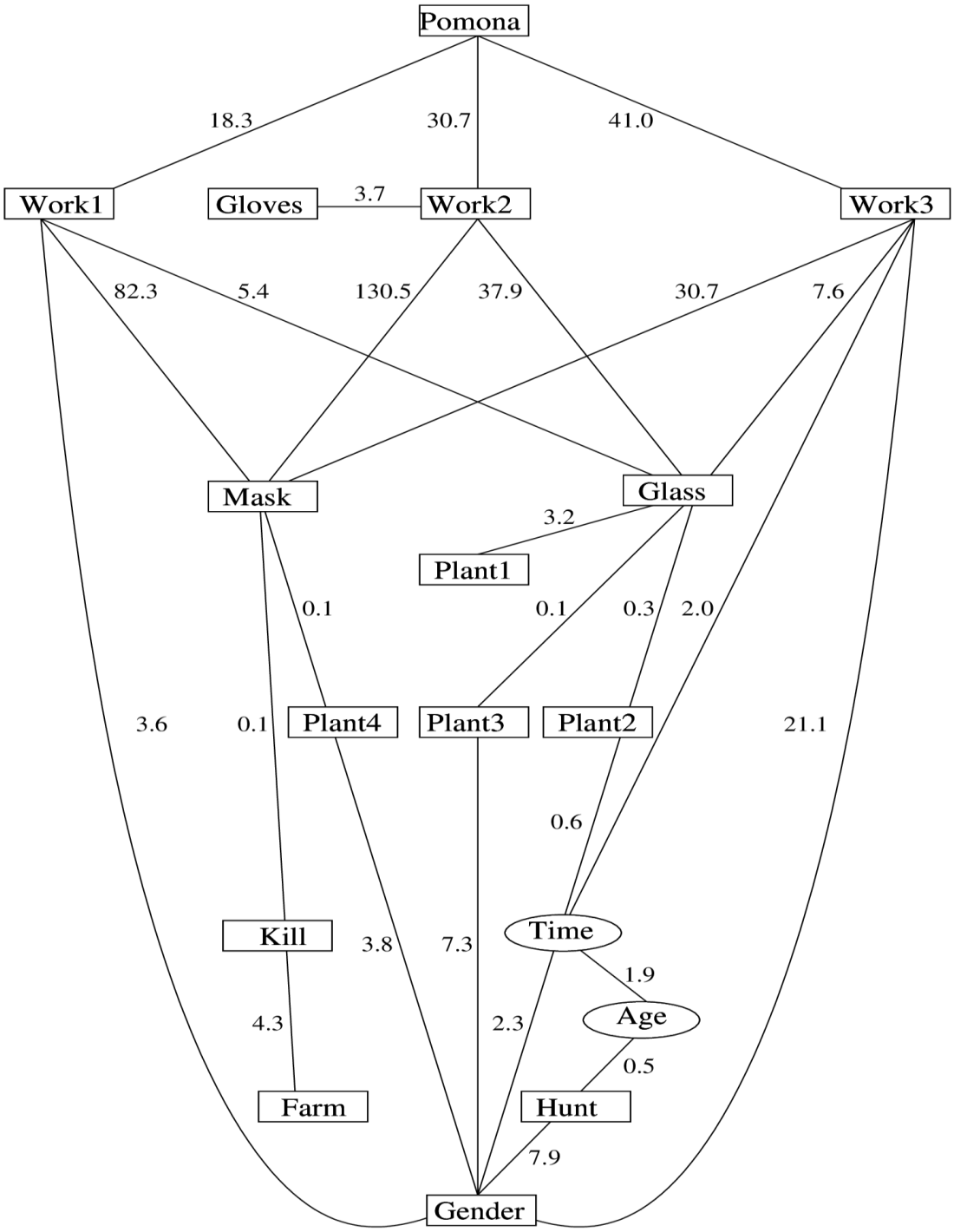


Figure 2: Final globally optimal additive Bayesian Network (ABN) model, after adjustment for over-fitting, evaluating factors linked with the odds of new infection with *Leptospira interrogans* sv Pomona ("Pomona") in sheep abattoir workers (n=384) in New Zealand. Directly dependent variables were various work positions ("Work1", "Work2", and "Work3"). Binary variables are shown as squares and continuous variables as ovals. Numbers represent odds ratios of significant directly dependent variables in ABN model, as reported in Table 2. Arc direction is omitted to not create confusion with the usual epidemiological DAGs, which imply causality and possible variables intervention, absent in ABN model where only statistical dependency are relevant.

Table 3: Associations between INDIRECTLY[1] linked variables of new infection with *Leptospira interrogans* sv Pomona (Pomona) in abattoir workers processing sheep (n=384) in New Zealand and odds ratios (OR) and 95% credible intervals (95% CI) analysed by Additive Bayesian Network (ABN)

Indirectly associated variables	OR	95% CI
Odds of Pomona infection when wearing a facemask in general and when working in various work positions		
Pomona|Mask[2]	7.0	3.5-14.0
Pomona|Mask,Work1[3]	3.0	1.9-4.8
Pomona|Mask,Work2	2.9	1.9-4.5
Pomona|Mask,Work3	2.3	1.5-3.6
Odds of Pomona infection when wearing safety or normal glasses in general and when working in various work positions		
Pomona|Glass	4.6	2.2-9.8
Pomona|Glass, Work1	2.2	1.3-3.7
Pomona|Glass, Work2	2.0[4]	0.9-4.7
Pomona|Glass, Work3	1.4[4]	0.9-2.4
Odds of Pomona infection when wearing gloves in general and when working in various work positions		
Pomona|Gloves	0.8[4]	0.4-1.6
Pomona|Gloves, Work1	0.7[4]	0.6-1.3
Pomona|Gloves, Work2	1.0[4]	0.6-1.7
Pomona|Gloves, Work3	0.9[4]	0.6-1.4

[1] More than one arc between the variables

[2]Interpretation: the probability of Pomona infection is seven times as likely when wearing a facemask than not wearing a facemask (=crude odds ratio);

[3]Interpretation: the probability of Pomona infection is three times as likely when wearing a facemask and working in work position 1 (offal room or pet food) than not wearing a facemask (=adjusted odds ratio);

[4]Statistically not significant; variable names are explained in Table 1.

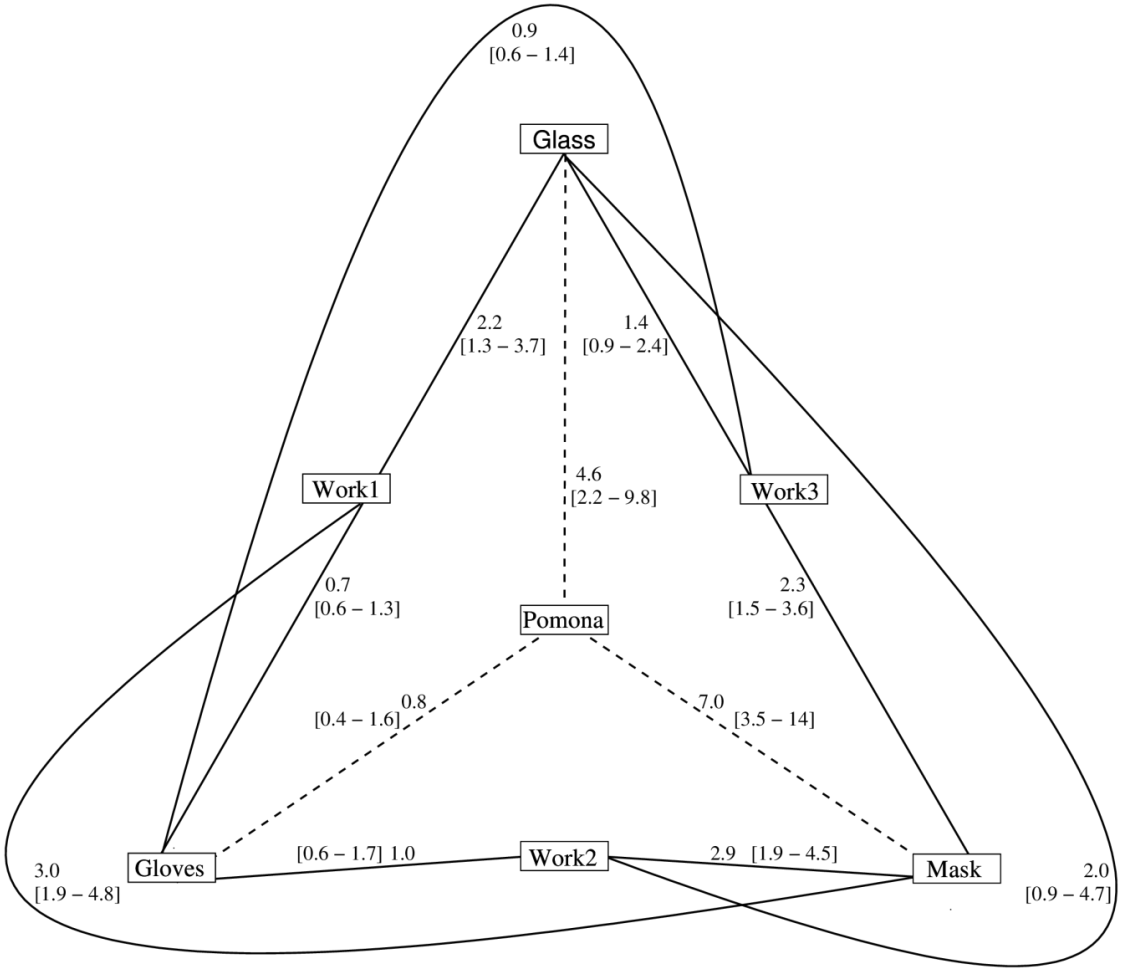


Figure 3: Visual representation of the associations between the INDIRECTLY linked variables "Wearing gloves", "Wearing a facemask" and "Wearing normal or safety glasses" (=PPE) and new infection wit*h Leptospira interrogans* sv Pomona (Pomona) in abattoir workers processing sheep (n=384) in New Zealand as reported in Table 3. Dashed lines represent crude associations between PPE and Pomona infection and solid lines represent for work position adjusted associations between PPE and Pomona infection. The numbers between two variables are the odds ratios and the numbers in square brackets the [95% credible intervals].

DISCUSSION

In the last four decades, four cross-sectional studies investigated *Leptospira* sero-prevalence in meat workers in NZ (49-52) estimating sero-prevalences against Pomona, Hardjo, and/or *Leptospira borgpetersenii* sv Tarassovi of being between 4.1% and 31%. One longitudinal study investigating risk factors for *Leptospira* incidence risk in abattoirs has been conducted recently (53). However, in the latter study, the *Leptospira* serovars contributing to "new infection" were the two serovars Hardjo and Pomona as a combined outcome. Since a study found that risk factors for Hardjo and Pomona infection in livestock varied substantially (30), and since Pomona was associated with the majority of new infections in workers in all abattoirs and with more signs of flu-like illness as opposed to Hardjo, we omitted Hardjo infection from our analysis outcome in this analysis. The analysis of risk factors for new infection with Pomona is therefore novel and has not been done in the former study (53).

Had serovar Hardjo been associated with more than 13 new infections, we would have incorporated it in the analysis as a variable, as ABN could have demonstrated the dependencies between Hardjo, Pomona and all other variables, differentiating the roles of these two serovars in the risk factor scenario at sheep abattoirs for leptospirosis. However, given the few sero-positive cases, one third with respect to Pomona new infection, it would have resulted in a poorer model fit. Further, in GLM it would have been nonsensical to include Hardjo infection as a risk factor for Pomona infection in the outcome. Hence, an inclusion of Hardjo was not possible for comparing the two methods.

The objective of the presented analysis was to identify risk factors for new infection with Pomona in sheep abattoir workers and to compare results from GLM with those from ABN. GLM and ABN confirmed the hypothesis that work position was the strongest risk factor for new infection with Pomona in sheep abattoir workers (Table 2). Hence, both methods appeared to be appropriate for identifying strong associations. ABN models are multidimensional multivariate regression models and analyse associations between all variables at the same time (4). Therefore, ABN and GLM are likely to identify the same risk factors when associations are strong and highly significant.

While work position was a categorical variable with four levels in GLM, it consisted of three binary variables in ABN due to splitting the multicategorical variables into binary ones, resulting in different baseline data. The ORs for the association between Pomona infection and the work position variable(s) did not only vary because of a different analytical method in GLM and ABN, but also varied due to different

baseline categories. In GLM the baseline category was working in boning, chillers, office, whereas in ABN the reference group was not being in the respective category (see Table 1). For example, in ABN working in yards, stunning or pelting versus working elsewhere resulted in an OR of 41, whereas in GLM working in yards, stunning or pelting versus working in the boning room and office gave an OR of 57 (Table 2). Since in GLM a highly exposed group was compared to with one with hardly any exposure to Pomona, the OR was higher than in ABN, where the baseline category included persons from various exposure levels. Despite the different baselines, the results are still comparable, being ABN a generalization of the usual GLM regression and as the aim of GLM and ABN was to identify the presence of statistical dependency and factors associated with Pomona infection. The work position variables in ABN ("Work1", "Work2", "Work3" in Figure 2) were all significantly related to Pomona infection ("Pomona"), as were the working position categories in GLM. The multivariate ABN can be viewed as a collection of multivariable models (GLM) along the arcs of the DAG, hence the parameter estimates are expected to be identical given the same explanatory variables. However, compared with the GLM, the ABN model is regarded as more flexible since each level of a categorical variable can potentially have different sets of dependencies to other variables, whereas GLM only associates independent variables with a single outcome. This flexibility was apparent with the variable 'wearing gloves' ("Gloves"), which was only connected to the removal of intestines/kidneys/meat inspection ("Work2"), suggesting that the risk attribution to wearing the PPE depended on work position.

As already discussed by Dreyfus et al. (53), the highest odds of infection in workers at the beginning of the slaughter board may be explained by contact with contaminated droplets due to frequent urination of stunned sheep. The relatively high odds of infection during removal of kidneys and at meat inspection may be attributable to direct exposure of workers to Pomona residing in the genital-urinary system. Kidneys pass through the offal room, possibly explaining the odds of infection in that working area. Working in the boning room, chillers or in the office was associated with little or no exposure to urine with much lower odds of infection with Pomona.

Wearing PPE at the work place (gloves, facemasks and glasses) were not statistically significantly associated with Pomona in GLM analysis. However, ABN suggested that such PPEs increased the odds of infection. This may be biologically plausible because workers wearing safety goggles or facemasks reported they sweat and presumably wipe their eyes with potentially contaminated hands more often than workers not

wearing them. Wearing glasses and/or a facemask increased the odds of infection at the slaughter board approximately two-fold. This is an important new finding by ABN, hence we recommend research to clarify whether this is actually true (e.g. by detecting *Leptospira* DNA in facemasks and glasses by PCR). Our findings about PPE should be interpreted with some caution, as there is a possibility of differential misclassification bias. When responding to questions about wearing PPE, participants may have overstated the use of PPE and not admitted non-compliance to the employment policy enforcing the use of PPE, despite a clear statement that interviews were confidential. This may have led to an overstatement of wearing PPE by meat workers in exposed work positions, reducing the chance of determining a protective effect of PPE in the analysis. Nevertheless, we believe that such bias were small because workers handling kidneys were, contrary to belief, less likely to sero-convert than workers at stunning/pelting.

Hunting, farming or slaughter of animals at home were not associated with Pomona infection in the GLM and only indirectly linked to Pomona infection through three to four arcs in the ABN model. This is an indication that in this study population exposure to Pomona was more likely occurring in the abattoir than through contact with livestock at times off-work. This finding underlines the role of leptospirosis as an occupational hazard in sheep abattoirs in NZ. These findings were confirmed in the study on sero-prevalence/incidence and risk factors by Dreyfus et al. (52, 53), but contrast with the findings of Heuer et al. (54), where home slaughter was found to be a risk factor for sero-prevalence of Hardjo or Pomona, where Hardjo titres were 5-fold more frequent than Pomona titres among workers of one abattoir (A1).

An advantage of ABN is the illustration of the dependencies between all variables by the graphical model (Figure 2), compared with GLM which only shows the dependencies between risk factors and outcome (Pomona infection). Hence, the GLM only identified variables that were directly associated with Pomona infection, and was restricted to a limited model space ignoring indirect relationships. Conversely, ABN considered all variables jointly allowing arcs to be present between any variables. For example, ORs for the effect of wearing PPE on Pomona infection at certain work positions could be estimated (Table 3, Figure 3). The DAG illustrates that farming was dependent on slaughter of animals at home, which was associated with working at the yards, stunning and pelting ("Work1"). Hence, persons working in these positions were more likely to slaughter at home and farm. Hunting was associated with the variables "Age" and "Time", meaning that older, long time workers were more likely to go hunting. The estimation of ORs indirectly linked to the potential outcome variable has, to the authors' knowledge, not been presented before

in the literature.

The Yule-Simpson paradox (55) states that taking a narrow univariate (single dependent variables/multivariable regression) approach to risk factor analysis will, in general, not give the same result as a joint and truly multivariate approach (4). In this study, ABN and GLM methodology did not produce exactly the same results: whereas work position was the only directly dependent variable upon new infection with Pomona in both ABN and GLM, the GLM found workers in Abattoir A2 to be at higher odds of infection than workers at A1. As shown in Figure 2, ABN suggested that the odds of infection in Abattoir A2 ("Plant1") was indirectly linked to the outcome "Pomona infection" through wearing normal or safety glasses and all three work positions. The results from the ABN method suggest that in Abattoir 1 (A2), Pomona infection occurred more often when using PPE, hence was associated with policy compliance at specific work positions. GLM had not detected these associations. Hence, while GLM only established a direct association between one abattoir and Pomona infection, ABN identified a network of inter-dependent factors linked to the outcome. Here ABN was more informative about potential causal pathways in the disease system than GLM (4). This potential advantage of the ABN method would specifically be useful for observational studies with large number of variables, where causal and time relationships are often unknown.

The technical foundations of ABN modelling lie within the machine learning and data mining literature (42, 43, 56-58). The main obstacle of using this methodology in practice is that it can be computationally rather demanding: determining the best model for a given data set has been shown to be NP-hard (59), the most difficult class of computational problem. This means that finding an optimal model must be done using heuristic search algorithms (42, 43, 56-58), rather than brute force computation. Hence, another limitation of ABN modelling is the restriction of the number of variables. To date, exact structure discovery with bootstrapping is only feasible for around 20 variables (44). Heuristic searches (43) and inexact order-based searches (58) for globally optimal DAGs offer an alternative, however they are less ideal approaches, because they are local techniques and not exact methods. In the future, advances in either technology or statistical methods will make larger computationally intensive analysis more feasible. An example could be that presented by (60). Moreover, another drawback of the current ABN methodology is the unfeasibility to take into account possible interactions. However, we checked possible effect modifications for the GLM model, as clarified in the previous sections, but they revealed to be not significant

and they did not improve the model fit. Therefore, although if this possible limitation is present, it does not harm our analysis and the two methodologies are still comparable due to absence of effect modifications.

The credible intervals were wider in the ABN than in the GLM results. This is due to the model nature, where all the variables are taken into account due to the joint mathematical model formulation, despite the use of GLM techniques in the estimation process. In addition, the estimation of the ORs between undirected variables (e.g. Work|Mask) introduces more uncertainty.

The stepwise algorithm used to select the best GLM model is not always recognized as the standard procedure for model selection (23). Nevertheless, in this context, the stepwise approach was appropriate, as the choice of the frequency of variables put into the model was based on hypotheses, formulated with knowledge from former studies (51, 53), with knowledge of infection pathways and the epidemiology of leptospirosis.

One abattoir ("Abattoir 1") was studied twice in two consecutive years and three abattoirs were studied in the second year once. Since only 14.8 % of workers were sampled repeatedly in "Abattoir 1" and since sero-conversion and anamnestic responses (= Pomona incidence) were measured, and not sero-prevalence, clustering was expected to be at very low level. This was confirmed when we extended the GLM model with generalised estimating equations (GEEs) and the SE changed by less than 5%.

Although the abattoirs were not fully representative of the whole of NZ, solely located in the west and east of the North Island, the animals originated from all areas of the North Island. Furthermore, as demonstrated in (52), the study population was recruited from almost 20% of the total sheep abattoir worker population.

Since participation was voluntary, it was likely a sampling bias had been introduced. A parallel analysis revealed that workers from high exposed work positions were more likely to participate (61). But this did not affect the results from the multivariable logistic regression analysis where working area was included as a covariate.

The MAT titre cut-off of 1:48 is appropriate to determine exposure to leptospires in humans, but is generally not recommended as a cut-off for diagnosing clinical disease (35, 62). Hence, a two-fold increase can determine new infection due to exposure to leptospires, but is not necessarily appropriate to diagnose clinical disease. The latter requires, by WHO definition, a single MAT antibody titre $\geq$800 or a four-fold increase in the convalescent blood sample. However, this definition has been challenged recently (30, 63), as

infection with certain serovars seem to lead to clinical disease with lower antibody responses. Seroconverting meat workers in this study had a two-fold risk of influenza-like symptoms compared to workers not seroconverting (25). Fang et al. (64) modelled the association between *Leptospira* sero-positivity and risk factors in meat workers of one sheep abattoir for different MAT cut-offs. While the percentage of sero-positive meat workers reduced by approximately 40%, when choosing a MAT titre cut-off of 1:96 rather than 1:48, the conclusions on risk factors did not change. In many countries a wide range of serologically related serovars is prevalent introducing the problem of cross reactivity in the MAT. However, the prevalence of six endemic serovars in NZ, which belong to different serogroups, should reduce the problem of cross reactivity (65). The MAT in the NZ context is therefore very specific and false positives should not represent a problem in this study context. In a study evaluating the MAT sensitivity and specificity of acute (MAT cut-off 1:100) and convalescent (MAT cut-off not mentioned) sera in an urban setting in Brazil (66), the MAT testing of convalescent sera had a sensitivity of 91% to 100% and specificity of 94% to 100%. If we assumed that the MAT in our study had a 91% sensitivity and 94% specificity, the tested incidence in meat plants was likely under-estimated. However, since we used a MAT titre cut-off of 1:48 and tested for the serovars Hardjobovis and Pomona, which are less likely to be encountered in a urban setting, where serovar Copenhageni is predominant (66), it is possible that the sensitivity and specificity of the MAT in NZ are not the same as in Brazil.

In conclusion, this study demonstrated that workers were at highest odds of new infection when working at the beginning of the slaughter (stunning and hide removal), followed by those removing intestines, bladder and kidneys, and workers in the offal/pet food area. PPEs like facemasks and safety glasses did not show any indication of being protective (GLM). On the contrary, the ABN model suggested that such PPE increased the odds of infection, but this requires verification using other research methodology. Further, other means of protection might be considered, like vaccination of farmed livestock or slaughter procedure changes. ABN has an advantage over GLM due to its capacity to capture the natural complexity of data more effectively. In ABN, all relationships between variables are modelled, which appears to be more explanatory in view of the inter-dependencies between study variables in complex disease systems. This is due to its ability to estimate relationships between undirected variables, and more associations between variables can be explored, for example the association between wearing PPE and Pomona infection.

ACKNOWLEDGEMENTS

The authors are indebted and grateful to study participants, managers and health and safety workers of the participating abattoirs, nurses and phlebotomists, without whom the study would have been impossible.

The ABN methodology is part of the PhD project of Marta Pittavino. She thanks Fraser Ian Lewis to have introduced her to this new methodology. Marta Pittavino is a PhD candidate in the Epidemiology and Biostatistics PhD Program of Life Science Zurich Graduate School and gratefully acknowledges its support.

Marta Pittavino expresses her deep gratitude to the Foundation ‘‘Franco and Marilisa Caligara per l’Alta Formazione Interdisciplinare’’ for its support before and during the PhD studies.

We thank Sarah Moore for tireless support of sampling and data management, occupational health physicians John Reekie & John Kerr for advice and communication with abattoirs, Heather Duckett for helping to organize sampling, Christine Cunningham and Wendy Maharey for administrative support, Brian O`Leary, Masood Sujau and Simon Verschaffelt for help developing the database, Fang Fang, Prakriti Bhattarai, Rayon Gregory, Claire Cayol and Emilie Vallee for interviewing, Neville Haack and Rae Pearson for MAT testing Roger Lentle for advice for the Massey University Human Ethics Committee application and Lesley Stringer and Sarah Rosanowski for analytical and software support. Further, we thank and the Department of Labour NZ for support.

FINANCIAL SUPPORT

We gratefully acknowledge funding by Rural Woman New Zealand, and commissioned by the Tertiary Education Commission (TEC) via the Institute of Veterinary, Animal and Biomedical Sciences, Massey University (TEC #RM12703 (2008)), and the Swiss National Science Foundation (PBBEBS-124186, SNF138562 and SNF144973).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ETHICAL STANDARDS

All procedures were approved by the Massey University Human Ethics Committee in 2008 and 2009 (HEC: Southern A, Application 05/123 and 09/08, Slaughter carcasses as a source for human infection with Leptospira serotypes Hardjo, Pomona and Ballum at abattoirs in New Zealand).

SUPPORTING INFORMATION

Files containing original data and codes

Study questionnaire

REFERENCES

1. McCulloch CE, Searle S. R., Neuhaus JM. Generalized, Linear, and Mixed Models: Wiley; 2008.
2. Lewis FI, editor Bayesian Networks as a tool for Epidemiological Systems Analysis. 9th international conference on mathematical problems in engineering, aerospace and sciences (ICNPAA 2012); 2012; Melville, NY 11747-4501 USA: American Institute of Physics.
3. Rijmen F. Bayesian networks with a logistic regression model for the conditional probabilities. International Journal of Approximate Reasoning. 2008 June;48(2):659-66.
4. Lewis FI, McCormick BJJ. Revealing the complexity of health determinants in resource-poor settings. Am J Epidemiol. 2012 2012 Dec 1 (Epub 2012 Nov;176(11):1051-9. PubMed PMID: MEDLINE:23139247.
5. Lycett SJ, Ward MJ, Lewis FI, Poon AFY, Pond SLK, Brown AJL. Detection of Mammalian Virulence Determinants in Highly Pathogenic Avian Influenza H5N1 Viruses: Multivariate Analysis of Published Data. Journal of Virology. 2009 Oct;83(19):9901-10. PubMed PMID: WOS:000269614300025.
6. Poon AFY, Lewis FI, Frost SDW, Pond SLK. Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. Bioinformatics. 2008 September;24(17):1949-50.
7. Poon AFY, Lewis FI, Pond SLK, Frost SDW. Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope. Plos Computational Biology. 2007 January;3(1):e11.
8. Poon AFY, Lewis FI, Pond SLK, Frost SDW. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. Plos Computational Biology. 2007 November;3(11):e231.
9. Dojer N, Gambin A, Mizera A, Wilczynski B, Tiuryn J. Applying dynamic Bayesian networks to perturbed gene expression data. Bmc Bioinformatics. 2006 May;7:249.
10. Hodges AP, Dai DJ, Xiang ZS, Woolf P, Xi CW, He YQ. Bayesian Network Expansion Identifies New ROS and Biofilm Regulators. Plos One. 2010 March;5(3):e9513.
11. Jansen R, Yu HY, Greenbaum D, Kluger Y, Krogan NJ, Chung SB, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science. 2003 October;302(5644):449-53.
12. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. A primer on learning in Bayesian networks for computational biology. Plos Computational Biology. 2007 August;3(8):e129.
13. Djebbari A, Quackenbush J. Seeded Bayesian Networks: Constructing genetic networks from microarray data. Bmc Systems Biology. 2008 July;2:57.
14. Lewis FI, Brulisauer F, Gunn GJ. Structure discovery in Bayesian networks: An analytical tool for analysing complex animal health data. Prev Vet Med. 2011 June;100(2):109-15.
15. Ward MP, Lewis FI. Bayesian Graphical modelling: Applications in veterinary epidemiology. Prev Vet Med. 2013 15 May 2013;110(1):1-3.
16. Wilson AJ, Ribeiro R, Boinas F. Use of a Bayesian network model to identify factors associated with the presence of the tick Ornithodoros erraticus on pig farms in southern Portugal. Prev Vet Med. 2013 MAY 15;110(1, SI):45-53.
17. Sanchez-Vazquez M, Nielen M, Edwards S, Gunn G, Lewis F. Identifying associations between pig pathologies using a multi-dimensional machine learning methodology. BMC Veterinary Research. 2012;8(1):151.
18. Firestone SM, Lewis FI, Schemann K, Ward MP, Toribio J-ALML, Dhand NK. Understanding the associations between on-farm biosecurity practice and equine influenza infection during the 2007 outbreak in Australia. Prev Vet Med. 2013 MAY 15;110(1, SI):28-36.
19. Firestone SM, Lewis FI, Schemann K, Ward MP, Toribio J-ALML, Taylor MR, et al. Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a

subsequent outbreak. Prev Vet Med. 2014 OCT 1;116(3, SI):243-51.
20. Schemann K, Lewis FI, Firestone SM, Ward MP, Toribio J-ALML, Taylor MR, et al. Untangling the complex inter-relationships between horse managers' perceptions of effectiveness of biosecurity practices using Bayesian graphical modelling. Prev Vet Med. 2013 MAY 15;110(1, SI):37-44.
21. Ludwig A, Berthiaume P, Boerlin P, Gow S, Léger D, Lewis FI. Identifying associations in Escherichia coli antimicrobial resistance patterns using additive Bayesian networks. Prev Vet Med. 2013 May;110(1):64-75.
22. McCormick BJJ, Sanchez-Vazquez MJ, Lewis FI. Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs. Prev Vet Med. 2013 May;110(1):54-63.
23. Lewis F, Ward MJ. Improving epidemiologic data analyses through multivariate regression modelling Emerging Themes in Epidemiology. 2013;10:4.
24. Hartskeerl RA, Collares-Pereira M, Ellis WA. Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. Clinical Microbiology and Infection. 2011 April;17(4):494-501. PubMed PMID: WOS:000288501700005.
25. Dreyfus A, Heuer C, Wilson P, Collins-Emerson J, Baker MG, Benschop J. Risk of infection and associated influenza-like disease among abattoir workers due to two Leptospira species. Epidemiol Infect. 2014 Sep 30:1-11. PubMed PMID: 25266854. Epub 2014/10/01. Eng.
26. Adler B, de la Pena Moctezuma A. Leptospira and leptospirosis. Veterinary Microbiology. 2010 Jan 27;140(3-4):287-96. PubMed PMID: WOS:000274986400012.
27. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, Lovett MA, et al. Leptospirosis: a zoonotic disease of global importance. Lancet Infectious Diseases. 2003 Dec;3(12):757-71. PubMed PMID: ISI:000186910800018. English.
28. Thornley CN, Baker MG, Weinstein P, Maas EW. Changing epidemiology of human leptospirosis in New Zealand. Epidemiol Infect. 2002 Feb;128(1):29-36. PubMed PMID: ISI:000174910400005. English.
29. Institute of Environmental Science and Research (ESR). Annual surveillance summary 2006-2010 New Zealand2006-2010. Available from: https://surv.esr.cri.nz/surveillance/annual_surveillance.php.
30. Dreyfus A. Leptospirosis in humans and pastoral livestock in New Zealand. Palmerston North, New Zealand: Massey University; 2013.
31. Marshall RB, Manktelow BW. Fifty years of leptospirosis research in New Zealand: a perspective. New Zealand Veterinary Journal. 2002;50(3):61-3. PubMed PMID: ISI:000176378400014. English.
32. Ayanegui-Alcerreca M, Wilson PR, Mackintosh CG, Collins-Emerson JM, Heuer C, Midwinter AC, et al. Regional seroprevalence of leptospirosis on deer farms in New Zealand. New Zealand Veterinary Journal. 2010 Aug;58(4):184-9. English.
33. Wilson P, Glossop JC, van der Kroef JW, Heuer C, Stringer L. Disease and deer farm productivity and profitability. Proceedings of the Deer Branch of the New Zealand Veterinary Association. Annual Conference 2008 - Volume 25. Palmerston North, New Zealand: The Deer Branch New Zealand Veterinary Association; 2008. p. 22-9.
34. Keenan B. Leptospirosis: reducing the impact on New Zealand workplaces. Wellington, New Zealand: Department of Labour, 2007.
35. Faine S, Adler B, Bolin C, Perolat P. *Leptospira* and Leptospirosis. Melbourne, Australia: MediSci; 1999. 272 p.
36. R Development Core Team. R: A Language and Environment for StatisticalComputing. Vienna, Austria2015.
37. Dohoo I, Wayne M, Stryhn H. Veterinary Epidemiologic Research. Veterinary Epidemiologic Research. Charlottetown, Canada: VER inc.; 2010.
38. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986 Apr;73(1):13-22. PubMed PMID: WOS:A1986A734100002. English.

39. Hosmer DW, Lemeshow S. Assessing the fit of the model. Applied Logistic Regression Second ed. New York: John Wiley & Sons Inc; 2000. p. 143-67.
40. Lewis F, Pittavino M, Furrer R. abn: Data Modelling with Additive Bayesian Networks. 2014.
41. Tierney L, Kadane JB. Accurate Approximations For Posterior Moments and Marginal Densities. Journal of the American Statistical Association. 1986 March;81(393):82-6.
42. Friedman N, Goldszmidt M, Wyner A. Data analysis with Bayesian networks: A Bootstrap approach. 1999. p. 196-205.
43. Heckerman D, Geiger D, Chickering DM. Learning Bayesian Networks - The Combination of Knowledge And Statistical Data. Machine Learning. 1995 September;20(3)(3):197-243.
44. Koivisto M, Sood K. Exact Bayesian structure discovery in Bayesian networks. Journal of Machine Learning Research. 2004 May;5:549-73.
45. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Vol, editor: Psychosomatic Medicine; 2004.
46. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Hornik, K, Leisch, F, Zeileis, A (Eds), Proc 3rd Int Work Dist Stat Comp (DSC 2003) Vienna, Austria, pp 20–22. 2003.
47. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006;6(1):7-11.
48. Cowles M, Carlin B. Markov chain Monte Carlo convergence diagnostics: A comparative review. Journal of the American Statistical Association. 1996 JUN;91(434):883-904.
49. Blackmore DK, Bell L, Schollum L. Leptospirosis in Meat Inspectors - Preliminary-Results of a Serological Survey. New Zealand Medical Journal. 1979;90(648):415-8. PubMed PMID: ISI:A1979HW18400001. English.
50. Blackmore DK, Schollum L. The Occupational Hazards of Leptospirosis in the Meat Industry. New Zealand Medical Journal. 1982;95(712):494-7. PubMed PMID: ISI:A1982PL92900003. English.
51. Benschop J, Heuer C, Jaros P, Collins-Emerson J, Midwinter A, Wilson P. Sero-prevalence of leptospirosis in workers at a New Zealand slaughterhouse. The New Zealand Medical Journal. 2009 2009 Dec;122(1307):39-47. PubMed PMID: MEDLINE:20148043.
52. Dreyfus A, Benschop J, Collins-Emerson J, Wilson P, Baker M, Heuer C. Sero-Prevalence and Risk Factors for Leptospirosis in Abattoir Workers in New Zealand. International Journal of Environmental Research and Public Health. 2014;11(2):1756-75. PubMed PMID: doi:10.3390/ijerph110201756.
53. Dreyfus A, Wilson P, Collins-Emerson J, Benschop J, Moore S, Heuer C. Risk factors for new infection with Leptospira in meat workers in New Zealand. Occupational and environmental medicine. 2014 Dec 17. PubMed PMID: 25520373. Epub 2014/12/19. Eng.
54. Heuer C, Dreyfus A, Wilson PR, Benschop J, Subharat S, Ayanegui-Alcerreca AM, et al. Epidemiology and control of leptospirosis in New Zealand. In: Alban L, Kelly LA, editors. Society for Veterinary Epidemiology and Preventive Medicine Proceedings, Nantes, France, 24-26 March, 20102010. p. 174-85.
55. Hand DJ, McConway KJ, Stanghellini E. Graphical models of applicants for credit. IMA Journal of Management Mathematics. 1997 March;8(2):143-55. Pubmed Central PMCID: 10.1093/imaman/8.2.143.
56. Buntine W. Theory refinement on Bayesian networks. In Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence, Los Angeles, CA, USA, pages 52-60: Morgan Kaufmann; 1991.
57. Cooper GF, Herskovits E. A Bayesian Method For the Induction of Probabilistic Networks From Data. Machine Learning. 1992 October;9(4):309-47.
58. Friedman N, Koller D. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. Machine Learning. 2003 January;50(1-2):95-125.
59. Chickering DM, Heckerman D, Meek C. Large-sample learning of Bayesian networks is

NP-hard. Journal of Machine Learning Research. 2004 October;5:1287-330.
60. Parviainen P, Koivisto M, editors. Exact structure discovery in Bayesian networks with less space. Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence; 2009; Arlington, Virginia, United States: AUAI Press.
61. Dreyfus A, Benschop J, Collins-Emerson J, Wilson P, Moore S, Heuer C. Adjusting the leptospirosis sero-prevalence of New Zealand abattoir workers for sampling bias [presentation]. Goldcoast, Australia2010.
62. Shivakumar S, Krishnakumar B. Diagnosis of leptospirosis--role of MAT. The Journal of the Association of Physicians of India. 2006 2006-Apr;54:338-9. PubMed PMID: MEDLINE:16944624.
63. Goris. M. Diagnostic tests for human leptospirosis. Second meeting of the European Leptospirosis Society on Leptospirosis and other rodent borne hemorrhagic fevers; Royal Tropical Institute, Amsterdam, The Netherlands: European Leptospirosis Society 2015.
64. Fang F, Heuer C, Collins-Emerson J, Wilson P, Benschop J, editors. Effect of antibody titer cut points on inferences from a cross-sectional sero-prevalence study of leptospirosis in meat workers. Leptocon 2009, the 6th Annual Scientific Meeting of International Leptospirosis Society; 2009; Cochin, India.
65. Hathaway SC. Leptospirosis in New Zealand: an ecological view. New Zealand Veterinary Journal. 1981 Jul;29(7):109-12. English.
66. McBride AJA, Santos BL, Queiroz A, Santos AC, Hartskeerl RA, Reis MG, et al. Evaluation of four whole-cell Leptospira-based serological tests for diagnosis of urban leptospirosis. Clinical and Vaccine Immunology. 2007 Sep;14(9):1245-8. PubMed PMID: ISI:000249488400028.

PAPER II

Attitudes of Austrian veterinarians towards euthanasia in small animal practice: impacts of age and gender on views on euthanasia

Sonja Hartnack, Svenja Springer, Marta Pittavino & Herwig Grimm

Paper published on *BMC Veterinary Research*.

Hartnack et al. BMC Veterinary Research (2016) 12:26
DOI 10.1186/s12917-016-0649-0

BMC Veterinary Research

RESEARCH ARTICLE Open Access

Attitudes of Austrian veterinarians towards euthanasia in small animal practice: impacts of age and gender on views on euthanasia

Sonja Hartnack[1*], Svenja Springer[2], Marta Pittavino[3] and Herwig Grimm[2]

Abstract

Background: Euthanasia of pets has been described by veterinarians as "the best and the worst" of the profession. The most commonly mentioned ethical dilemmas veterinarians face in small animal practice are: limited treatment options due to financial constraints, euthanizing of healthy animals and owners wishing to continue treatment of terminally ill animals. The aim of the study was to gain insight into the attitudes of Austrian veterinarians towards euthanasia of small animals. This included assessing their agreement with euthanasia in exemplified case scenarios, potentially predicted by demographic variables (e.g. gender, age, working in small animal practice, employment, working in a team, numbers of performed euthanasia). Further describing the veterinarians' agreement with a number of different normative and descriptive statements, including coping strategies. A questionnaire with nine euthanasia scenarios, 26 normative and descriptive statements, and demographic data were sent to all members of the Austrian Chamber of Veterinary Surgeons ($n = 2478$).

Results: In total, 486 veterinarians answered sufficiently completely to enable analyses. Responses were first explored descriptively before being formally analysed using linear regression and additive Bayesian networks – a multivariate regression methodology – in order to identify joint relationships between the demographic variables, the statements and each of the nine euthanasia scenarios. Mutual dependencies between the demographic variables were found, i.e. female compared to male veterinarians worked mostly in small animal practice, and working mostly in small animal practice was linked to performing more euthanasia per month.

Conclusions: Gender and age were found to be associated with views on euthanasia: female veterinarians and veterinarians having worked for less years were more likely to disagree with euthanasia in at least some of the convenience euthanasia scenarios. The number of veterinarians working together was found to be the variable with the highest number of links to other variables, demographic as well as ethical statements. This highlights the role of a team potentially providing support in stressful situations. The results are useful for a better understanding of coping strategies for veterinarians with moral stress due to euthanasia of small animals.

Keywords: Euthanasia, Human-animal bond, Multivariate additive Bayesian networks modelling, Small animal practice, Veterinary medical ethics

* Correspondence: Sonja.Hartnack@access.uzh.ch
[1]Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Winterthurerstr. 270, 8057 Zurich, Switzerland
Full list of author information is available at the end of the article

Background

Euthanasia of pets has been described by veterinarians as "the best and the worst" of the profession [1]. Although euthanasia presumably accounts for only less than 1 % of all veterinary services in a typical small animal practice [2], veterinarians face ethical dilemmas in this context regularly and consider them stressful [3]. Performing euthanasia has been described as an occupational stressor and related to suicidal behaviour in veterinarians [4] and systematically reviewed [5]. Studies indicate that suicidal thoughts seem to be higher among young, female veterinarians working in small animal practices [6, 7].

The most commonly mentioned ethical dilemmas in small animal practice are: limited treatment options due to financial constraints, euthanizing of healthy animals and owners wishing to continue treatment of terminally ill animals. Here, the principle to protect animals' lives on one hand and to reduce pain [8] on the other can conflict in a strict sense. Moreover, having responsibilities towards animal patients and pet owners at the same time, raises further fundamental questions in veterinary medical ethics [9, 10] or in other words: moral stress [11, 12].

Although ethics is included step-by-step in undergraduate veterinary curricula at least in European countries, and specific euthanasia guidelines such as the AVMA guidelines exist [13] it has been stated that there is no such thing as a common professional ethic within the veterinarian profession [14]. An approach such as the Principles of Biomedical Ethics [15] in human medicine, which integrates important ethical viewpoints is not in sight or applied in veterinary medicine. Looking at the legal requirements in the German speaking countries the situation becomes even more complex. The Austrian (https://www.globalanimallaw.org/database/national/austria/, accessed 28 October 2015) and German (https://www.globalanimallaw.org/database/national/germany/, accessed 28 October 2015) animal protection laws refer to the responsibility for the animal as a fellow creature or the concept of animal's dignity in Switzerland (https://www.globalanimallaw.org/database/national/switzerland/, accessed 28 October 2015). According to the Austrian and German animal protection law it is prohibited to kill an animal without a "good" – understood as a justifying – reason.

Thus veterinarians are faced with the challenge to clarify their ethos with regard to moral, legal, and societal responsibilities. This process takes place in a society where divergent views on animals and their standing are present and attitudes towards animals has significantly changed in recent decades [16].

The aim of this study was to assess if demographic variables such as work experience, gender or working time spent in small animal practice influence the veterinarians' attitudes towards euthanasia. To get a clearer picture of these attitudes, the level of agreement with euthanasia in a number of different case scenarios was utilised. The scenarios described situations with conflicting views between owners and veterinarians: either the owners requested euthanasia ("convenience euthanasia") or refused it in cases where euthanasia seemed to be the appropriate measure from a veterinary perspective. An additional aim was to assess the level of agreement with a number of different normative and descriptive statements in the context of small animal euthanasia and their potential links to demographic variables. The overall objective of this study was to establish a body of empirical knowledge describing normative and descriptive beliefs as well as underlying values of Austrian veterinarians regarding euthanasia in small animal practice. This included also insights into self-reported coping strategies concerning euthanasia related stress.

Methods

Questionnaire

The questionnaire utilised for the analysis comprised the following three sections: A: 9 scenarios, B: ethical and / or technical statements with 26 questions, C: demographic data. Seven scenarios described situations in which the person (animal owner) bringing an animal to the veterinary practice either requested the animal to be euthanized ($n = 5$) or refused euthanasia ($n = 2$). One scenario asked about the necessity to inform the official veterinarian in case of a terminally ill animal. One scenario asked about the willingness of the veterinarian to take the decision for or against euthanasia instead of the owners. For each scenario and statement the respondent was asked to rank agreement from 1 (rejection) to 9 (complete agreement). Statements and scenarios are presented in Tables 2 and 3. The questionnaire was prepared in German. A prototype version of the questionnaire was developed and pre-tested by veterinarians. Their comments on clarity and content were incorporated into a revised form. An additional file shows the original questionnaire [see Additional file 1].

The study population comprised all members ($n = 2478$) of the Austrian Chamber of Veterinary Surgeons defined using their e-mail distribution list. An electronic invitation to participate was sent outlining the aims of the study and assuring anonymity to the respondents. The questionnaire was implemented with the software LimeSurvey version 2.0 [17] and a reminder was sent via email within a month (November 2012). Since the study dealt with information that was regarded critical, ethical approval was explicitly asked for. According to the Ethics Commission of the Medical University Vienna no formal ethical approval was needed.

Data analysis

a) Linear regression models

Multivariable regression models were utilized to identify significant associations between the outcome "agreement with euthanasia" (in each of the different scenarios separately) and the demographic variables. The outcome variable ranged from total rejection to full agreement on a 9-point scale. The demographic predictor variables included: percentage of working time spent with small animals (dichotomized into ≤ 60 % and > 60 %) (*Small animals %*), working employed or self-employed (*Employment*), number of other veterinarians working in the same practice (*Nb vets*), number of euthanasia per month performed by the respondent (*Nb eutha*), number of times per year the respondent is asked to perform euthanasia of a healthy animal (*Request healthy eutha*), years working as a vet (*Years*) and gender (*Gender*). Stepwise model selection (backward and forward) by Akaike's information criterion (AIC) was performed using the MASS package [18] in the software R [19]. Only complete questionnaires with no missing values for the chosen variables were utilised for the multivariable analysis. We assumed that the response variable was continuous as this facilitates much clearer and more straightforward analyses. Given the large sample size and that the categories in the questionnaire are points on an underlying continuous scale from 1.0 (complete disagreement) to 9.0 (complete agreement) this is a reasonable approach. For completeness the data was also analysed in an analogous fashion using ordinal categories (proportional odds logistic models) and these results can be found in [Additional file 2].

b) Additive Bayesian networks (ABN)

A Bayesian network approach was used to analyze the results of the questionnaires with the software package abn [20]. In addition to the variables chosen for the linear regression 11 statements were included. The main reason to include only a subset of the 26 statements was technical, allowing for an exact search which is only possible for up to 20 variables. The 11 statements were chosen to represent all important ethical aspects. ABN is a well-established methodology for exploring complex observational data [21–23]. Bayesian network models, and specifically additive Bayesian networks (ABN), which we utilize here, are simply multi*variate* extensions of usual multi*variable* regression models, e.g. linear or generalized linear models (GLM). In contrast to a GLM, an ABN model does not require that we designate one variable in the study as a single response variable with the remaining variables all as predictors. Rather, an ABN allows all variables to be potentially mutually dependent, which is appropriate here as we have multiple response variables (i.e. scenarios and statements) we wish to consider, in all other respects it is a typical regression model. If the data are sufficiently simple then the ABN results will collapse to those using GLMs and so we lose nothing using this extended approach, but may gain additional insight into relationships which exist between all the different variables in the questionnaire. The results of our ABN analyses are presented as a graph, which is the usual presentation (see Figures 1 to 9), and which shows how the various different questionnaire responses are statistically related. Determining an optimal ABN model for a given data set is somewhat technical and full details are given in the additional material [see Additional file 3].

Results

a) Descriptive and linear regression analysis

Out of the 2478 contacted veterinarians, 764 returned the questionnaire, 486 were fully or sufficiently completed to enable analysis. In Table 1, the demographic variables are summarized, separately for male and female veterinarians. Based on median and the 25^{th} and 75^{th} percentiles, the level of agreement with each of the 26 statements was grouped into high or moderate agreement, ambivalent and disagreement or strong disagreement (Table 2). The level of agreement with euthanasia in terms of medians and the 25^{th} and 75^{th} percentiles (IQR) is shown in Table 3. Bar plots of the agreement with statements and with euthanasia are presented in the additional material [see Additional files 4 and 5].

For each of the nine different scenarios, linear regression models were utilised to assess if the predictor variables (gender, years having worked as a veterinarian, working mostly in small animals practice, type of employment, number of other veterinarians working in the same practice, number of performed euthanasia per month and number of requests per year to euthanize a healthy animal) were significantly associated with the response variable "agreement with euthanasia or else" in each of the nine scenarios. The detailed results including univariable models for all available data and the complete questionnaire as well as the final multivariable models with 95 % confidence intervals for the corresponding effects sizes are presented in the additional material [see Additional file 6].

In summary, for the scenarios describing "convenience euthanasia" gender was significantly associated with level of agreement in three out of five scenarios. Female veterinarians were more likely to disagree with euthanasia in the scenarios of the aggressive dog which had bitten a

Table 1 Summary statistics of the demographic data ($n = 486$) presented separately for male and female veterinarians

Variables	Female		Male		Missing values
	$n = 251$		$n = 167$		$n = 68$
	mean	95 % CI[b]	mean	95 % CI	
"Years" Number of years having worked as a veterinarian	12.4	[0;26.8]	22	[3.9;40.2]	$n = 65$
Age in years[a]	40.6	[24.9;56.2]	50.5	[32.9;68.1]	$n = 68$
	proportion in %	95 % CI[c]	proportion in %	95 % CI	
"Small animals %"Working >60 % in small animal practice	80	[74;84]	51	[43;59]	$n = 18$
"Employment"Being self-employed	72	[65;77]	95	[90;98]	$n = 63$
	median	[10th,90th] percentile	median	[10th,90th] percentile	
"Nb vets" Number of other veterinarians working in the same practice	1	[0;5]	1	[0;3]	$n = 74$
"Nb eutha" Number of euthanasia per month performed by respondent	3	[1;7]	3	[1;10]	$n = 92$
"Request healthy eutha" Number of times per year \respondent is asked to perform euthanasia of a healthy animal	2	[1;8]	2	[0;10]	$n = 81$

[a]Age was not considered in the statistical models
[b]95 % confidence interval corresponding to mean ± 1.96 standard deviation
[c]95 % Wald confidence interval

child even after specific training and therapy (F1), the young dog which would need a costly and time-demanding therapy (F3), and the rabbit owner who prefers to buy a new animal instead of spending money on his sick rabbit (F4). In the two remaining "convenience euthanasia" scenarios, the number of years having worked as a veterinarian was significantly associated with the level of agreement with euthanasia. Here more experienced veterinarians were more likely to agree with euthanasia. These two scenarios comprised situations in which a rabbit owner asks for euthanasia because of the wrong coat colour not meeting breeding standards (F2), and an old dog no longer fitting the living conditions of the owner (F5). The percentage of time spent in small animal practice was significantly associated with agreement to euthanize in the scenarios of the young dog in need of a costly and time-demanding therapy (F3), the rabbit owner who prefers to buy a new animal (F4) and the old dog no longer fitting to the owner's living conditions (F5). In all three scenarios, veterinarians working at least 60 % in small animal practice were more likely to disagree with euthanasia. The number of times being asked to euthanize a healthy animal was significantly associated with agreement in the following three scenarios: the aggressive dog (F1), the young dog in need of a costly and time-demanding therapy (F3) and the old dog no longer fitting his owner's living conditions (F5). Being asked more often to euthanize a healthy animal, veterinarians were more likely to disagree with euthanasia. With a higher number of monthly performed euthanasia, veterinarians were more likely to agree with euthanasia in the scenario of the old dog no longer fitting his owner's living conditions (F5).

Two other scenarios are related to situations in which the owner or the person in charge of the animal refuses humane euthanasia. In one scenario it is explicitly stated, that from a veterinary perspective euthanasia is to be recommended, whereas in the other scenario the presumed presence of lung metastases might suggest euthanasia. In both scenarios, the number of years having worked as a veterinarian was significantly associated with agreement to euthanize. In the scenario of the owner of a severely ill Persian cat, having a very close relationship with his cat and thus refusing euthanasia (F6), older veterinarians were more likely to disagree with euthanasia. In contrast, in the scenario of a dog sitter refusing to take the decision of euthanasia of an old dog with breathing problems and a history of malignancy when the owner cannot be reached (F7), older veterinarians were more likely to agree with euthanasia. In the scenario F8, a guinea pig owner refuses euthanasia of his animal with a tumour and wants to take it home instead. The question is raised if the official veterinarian has to be informed. Only gender was found to be significant with female veterinarians being more likely to notify the official veterinarian. In a scenario in which – on veterinary reasoning – no clear recommendation in favour or against euthanasia was possible (F9), the number of veterinarians working in the same practice, gender, and number of years having worked as a veterinarian, were found to be significantly associated with refusing

Table 2 Veterinarian's agreement with 26 normative and descriptive statements in the context of euthanasia in small animal practice

Name		Median (IQR)
High agreement		
S14[a]	It would be difficult for me to euthanize an animal against my conviction.	9 (9;9)
S17[a]	Treating the owners in an understanding way is a central part of euthanasia.	9 (9;9)
S16	Treating the dead animal in a respectful way is an important part of euthanasia.	9 (8;9)
S11[a]	Effective analgesia makes it easier for me to deal with the animal's suffering.	9 (8;9)
S10	It is easier for me to deal with euthanasia if the procedure is carried out according to the best technical standards.	9 (8;9)
S1	It is easier for me to deal with the animal's suffering if I know that I have done my best for its well-being.	9 (7;9)
S26[a]	I see reflected euthanasia as a central part of my practice as a vet.	9 (7;9)
S5[a]	Knowing that all veterinary medical, social and economic options have been considered makes it easier for me to deal with euthanasia.	9 (7;9)
S3	It is easier for me to deal with euthanasia if the owner has been well informed.	9 (7;9)
Moderate Agreement		
S21	It is easier for me to deal with euthanasia if I know that I have done my best for the animal's well-being.	8 (7;9)
S12	It is easier for me to deal with the animal's suffering if the owner has been well informed.	8 (5.75;9)
S13[a]	It is easier for me to deal with euthanasia if the animal has lived a rich live until its death.	8 (5;9)
S9[a]	Careful planning and the right moment make it easier for me to deal with euthanasia.	7 (5;9)
S2[a]	It is easier for me to deal with euthanasia if I know that the animal would only have lived on for a short time.	7 (5;9)
S24[a]	The animal's advanced (high) age makes it easier for me to deal with euthanasia.	7 (5;8)
S22	I see euthanasia as an unavoidable evil in my responsibility.	7 (4;9)
Ambivalent		
S4	It is easier for me to deal with euthanasia if the owner is satisfied about the way his animal has been euthanized.	5 (2;9)
S8[a]	I am still not used to euthanizing animals.	5 (2;8)
S18	Retrospectively, it becomes easier for me to deal with euthanasia.	5 (2;7)
Disagreement		
S15	It mostly causes more problems if the owners are present.	3 (1;6)
S7	It is easier for me to deal with euthanasia if the owners are present during the procedure.	3 (1;5)
S23	Knowing that my influence on the owner's decision is limited makes it easier for me to deal with euthanasia.	3 (1;5)
Strong Disagreement		
S20	Although I would reject euthanasia, I euthanize the animal because I am afraid that the owner will kill it himself.	2 (1;6)
S19	It is more difficult for me to euthanize an animal that does not have an owner (if all the other conditions are the same).	2 (1;5)
S6[a]	It is easier for me to euthanize an animal if I see that the owner does not have a close relationship to his animal.	2 (1;5)
S25	Although I would reject euthanasia, I euthanize the animal because I am afraid that the owner will see another vet.	1 (1;1)

Medians and interquartile ranges (IQR) of the agreement (1 = "I do not agree at all" to 9="I completely agree") given by the responding veterinarians to normative and descriptive statements in the context of euthanasia in small animal practice. Based on the results, the statements have been grouped arbitrarily into five different levels of (dis-)agreement. The names correspond to the designations given in the plotted graphs

[a]These statements have also been considered in the multivariate additive Bayesian networks modelling

to take the decision in the place of the owners. Veterinarians working in a team and being female were more likely to decline to take the decision. In contrast, older veterinarians were more likely willing to take the decision in favour or against euthanasia if the owners are not willing to decide.

b) Multivariate regression with additive Bayesian networks

Demographic variables, statements and agreement with scenarios

We now consider a more in-depth multivariate analysis where we examine each of the above scenarios in turn and additionally include eleven selected normative and descriptive statements (see Table 2) in the context of euthanasia in small animal practice, in addition to the demographic variables. Our objective here is to identify how and whether agreement with

Table 3 Veterinarians agreement with euthanasia or else in nine different euthanasia scenarios in small animal practice

Scenarios		Median (IQR)
"Convenience euthanasia"		
F1	Aggressive dogA dog has twice bitten persons. It has attended training courses and animal psychologists have tried to educate it. However, 2 days ago it severely injured a child that is now in hospital.	9[7;9]
F2	Rabbit breederA rabbit breeder wants to have some of her young animals euthanized because their coat colour does not meet the breeding standards and she will not be successful at exhibitions with those animals.	1[1;1]
F3	Young dog costly therapyAn animal owner comes to your office with a young dog. This dog is severely ill, but therapy is possible. This therapy would be time-consuming, but there are chances of success. The owner rejects the therapy because he has neither enough time nor enough money. He wants you to euthanize the dog.	3[1;5]
F4	Rabbit costly therapyA rabbit owner comes to your office. The animal suffers from a treatable disease, but the therapy would require some time and cost about 150 €. The owner does not want to spend the money on a therapy, but asks you to euthanize the rabbit. He wants to buy a new rabbit for 40 €.	1[1;3]
F5	Dog not fitting living conditionsA dog owner comes to your office and wants you to euthanize her dog. She argues that the 15 year old dog does not fit to her living conditions anymore because she will travel with her family for some time and does not want to bring a dog at this age to the animal shelter.	1[1;3]
"Owner's refusal to euthanize"		
F6	Persian catAn animal owner comes to your office with a severely ill Persian cat. You know that he has a very close relationship to his cat and does not want to part with it. In your opinion, euthanasia would be reasonable, but the owner does not agree. You reject any further treatment apart from analgesia.	7[5;9]
F7	Old sick dog without ownerA dog sitter comes to your office with a 17 year old dog that suffers from breathing problems. The owners have left for a trekking tour 3 days ago and cannot be reached. You removed a malign tumour in this dog 6 months ago and you are afraid that it has developed lung metastases. The dog sitter refuses to take a decision regarding euthanasia and cannot tell you what the owners might want.	6[2;8]
"Notification"		
F8	Guinea pig veterinary officerA guinea pig owner comes to your office because the guinea pig does not eat. You find a tumour of nut size in the region of the abdomen. As the animal's general condition is weak, you think that the prognosis is in Faust and recommend euthanasia. The owner thinks that the animal's condition is unproblematic and wants to take his pet home instead of having it euthanized. You are obliged to inform the veterinary officer.	2[1;7]
"Responsibility"		
F9	Dog veterinarian decisionA couple comes to your office with a dog of advanced age and asks: "What would you do if it was your animal?" You think that it is a 50/50 situation and that the couple will follow your advice. Would you refuse to make a clear recommendation and take the decision yourself?	7[5;9]

Medians and interquartile ranges (IQR) of the agreement for the different scenarios. For the scenarios F1 to F7, the veterinarians were asked to gauge their agreement with euthanasia in this case from 1="I reject euthanasia" to 9="I fully agree with euthanasia". In scenario F8 the question was about the necessity to notify an official veterinarian with the answer options ranging from 1="rejection" to 9="agreement". The answer options for scenario F9, asking about the willingness to take a decision concerning euthanasia in the place of the owners, ranged from 1="I would for sure make no recommendation" to 9="I would surely make a recommendation"

euthanasia in each of the scenarios is jointly related to these statements and also demographics. The results are presented as graphs, and the corresponding effects sizes are presented in the additional material [see Additional file 7]. In each graph an arc connecting two variables means that these are directly (statistically) related, a variable with no connecting arcs is statistical independent from all other variables.

For F1 (Fig. 1), the aggressive dog, gender was no longer found to be associated with agreement of euthanasia in the scenario of the biting dog. In contrast, gender was found to be linked to percentage of time spent in small animal practice and the number of other vets in the same practice. Compared to males, females worked more in small animal practice and with fewer colleagues. Older veterinarians were more often self-employed and worked with a lower number of colleagues. Being self-employed or employed was also linked with the number of other veterinarians working in the same practice. Percentage of time spent in small animal practice was linked directly to the number of performed euthanasia and indirectly via this variable also with number of times being asked to perform euthanasia of a healthy animal and with the number of other veterinarians working in the same practice. Thus spending more working time in small animal practice and being part of a larger team is associated with a higher number of performed euthanasia and more requests to euthanize a healthy animal, but not with agreeing with euthanasia in this scenario.

For F2 (Fig. 2), the rabbit breeder, the demographic variables are linked to each other in the same way and none of the variables was linked to F2.

In F3 (Fig. 3), costly therapy for a young dog, still prevailing a similar linking of the demographic variables, gender was found to be associated with agreement of euthanasia with females being more likely to disagree with

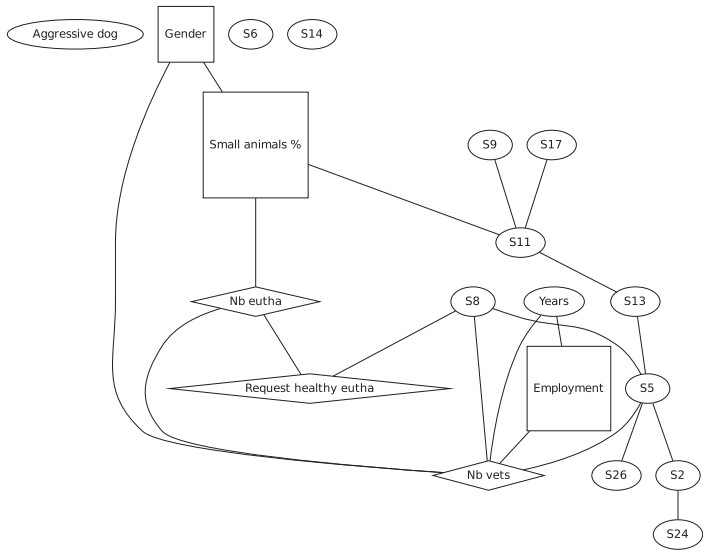


Fig. 1 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F1 (aggressive dog), seven demographic and eleven statements ($n = 301$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

euthanasia. Indirectly, either via gender or a statement, the number of performed euthanasia, the request for euthanizing a healthy animal as well as the number of veterinarians per practice are linked with the agreement of euthanasia in this scenario. Whereas agreeing with the statements S5 "Knowing that all veterinary medical, social and economic options have been considered makes it easier for me to deal with euthanasia"and S6 "It is easier for me to euthanize an animal if I see that the owner does not have a close relationship to his animal" were associated with a higher agreement of euthanasia, agreement with the statement S8 "I am still not used to euthanizing animals" was linked with disagreement of euthanasia.

In F4 (Fig. 4), costly therapy of a rabbit, a similar pattern of linking between the demographic variables was found. Gender was found to be associated with agreement of euthanasia in this scenario, with females being less likely to agree with euthanasia.

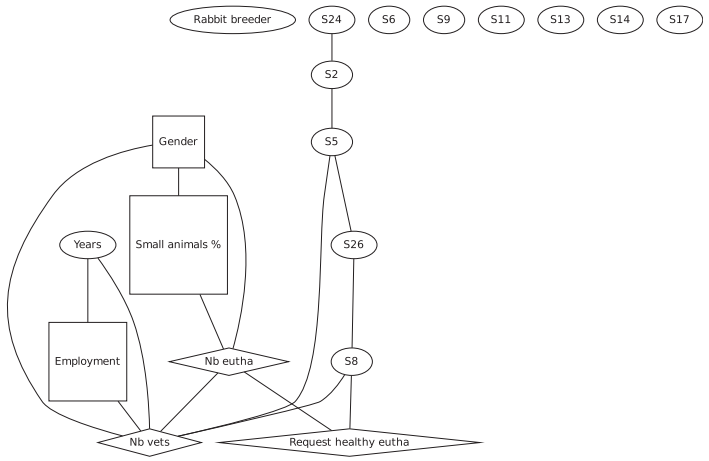


Fig. 2 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F2 (rabbit breeder), seven demographic and eleven statements ($n = 307$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

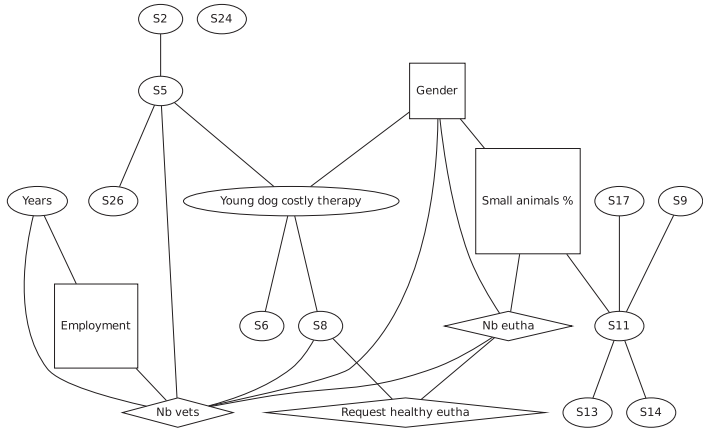


Fig. 3 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F3 (young dog costly therapy), seven demographic and eleven statements ($n = 303$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

In F5 (Fig. 5), a dog no longer fitting the owner's living conditions, next to the linking between the demographic variables similar to the other scenarios, the level of agreement with euthanasia was found to be directly linked to an increasing number of years having worked as a veterinarian, and to the number of monthly performed euthanasia. It was indirectly linked to the demographic variables being employed or self-employed, the number of other veterinarians in the same practice, the number of times per year being asked to euthanize a healthy animal and the percentage of time spent in small animal practice.

For F6 (Fig. 6), the Persian cat, a similar pattern of the linking between the demographic variables gender, working time spent in small animal practice, number of other vets, and being employed or self-employed, is seen. The number of performed euthanasia is linked to the number of times being asked to euthanize a healthy animal. The later variable is also linked indirectly with the number of other veterinarians working in the same practice.

In F7 (Fig. 7), an old dog with an absent owner, besides a linking of the demographic variables similar to the other scenarios, agreement of euthanasia was found to be linked to the number of other veterinarians in the same practice, directly to the number of monthly performed euthanasia and via this variable also with the number of times being asked to euthanize a healthy animal and working mostly in small animal practice. Here more veterinarians working in the same practice and a higher number of performed euthanasia were associated with an increased level of agreement. Indirectly agreement with euthanasia was linked via the number of veterinarians in the same practice with employment, professional years and gender.

In F8 (Fig. 8), the guinea pig, similarly to the other scenarios, the demographic variables were linked to each other, but the agreement of informing the official veterinarian was not linked to any other variables.

In F9 (Fig. 9), asking about the willingness of the veterinarian to take the decision for euthanasia instead of

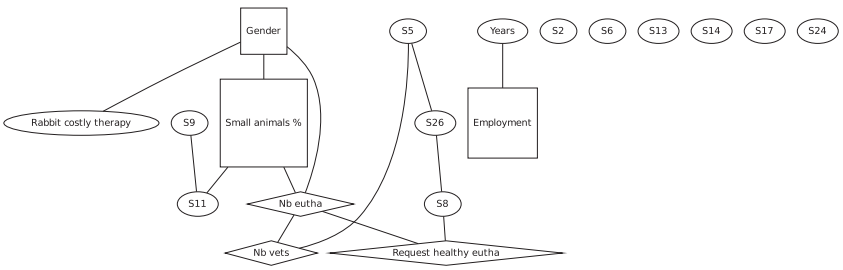


Fig. 4 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F4 (rabbit costly therapy), seven demographic and eleven statements ($n = 306$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

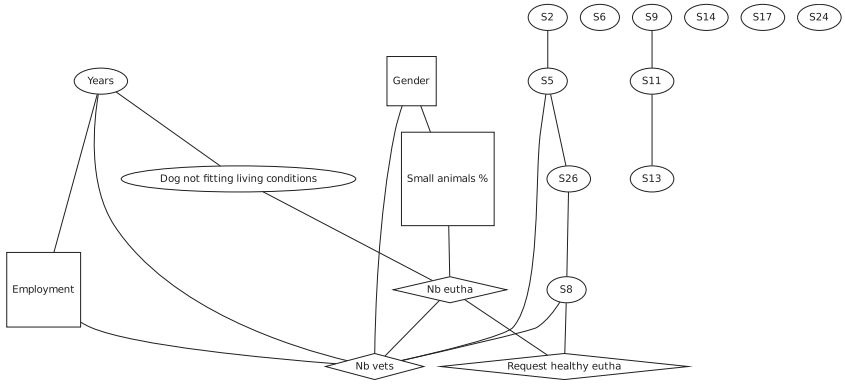


Fig. 5 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F5 (dog not fitting living conditions), seven demographic and eleven statements ($n = 303$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

the owner, the demographic variables being linked to each other in a similar way compared to the other scenarios. The willingness to take a decision for or against euthanasia was linked to the number of other veterinarians working in the same practice, with more veterinarians in the same practice being less likely to decide at the place of the owners. Indirectly agreement of euthanasia was linked via the size of the team with the status of being self-employed or employed, gender and the number of monthly performed euthanasia.

Agreement of statements with all other variables

In most of the graphs agreeing with the statement S8 "I am still not used to euthanizing animals" was associated with being asked to euthanize a healthy animal and the higher number of veterinarians working in the same practice. Additionally this statement was also found in close proximity to S26 "I see considerate euthanasia as a central part of my practice as a vet" and S5 "Knowing that all veterinary medical, social and economic options have been considered makes it easier for me to deal with euthanasia". The later statement was also found to be several times linked to the number of veterinarians working in the same practice and to S2 "It is easier for me to deal with euthanasia if I know that the animal would only have lived on for a short time". In some graphs, S2 was also found to be associated with S24 "The animal's advanced (high) age makes it easier for me to deal with euthanasia". In some graphs at least some of the four following statements were linked: S9 "Deliberate planning and the right moment make it easier for me to deal with euthanasia", S11 "Effective analgesia makes it easier for me to deal with the animal's suffering", S13 "It is easier for me to deal with euthanasia if the animal has lived a rich live until its death" and S17 "Treating the owners in an understanding way is a central part of euthanasia".

In conclusion, our multivariate results - given by the graphs - compared to those from our earlier linear multivariable analyses - have identified fewer associations between outcome (agreement with the scenario) and the demographic variables: for F1, F2, F6, and F8 no

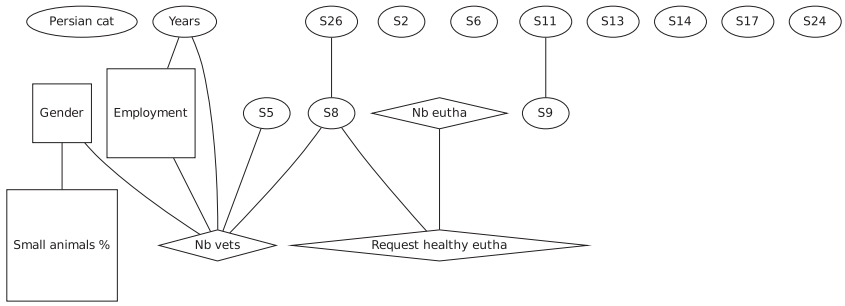


Fig. 6 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F6 (Persian cat), seven demographic and eleven statements ($n = 295$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

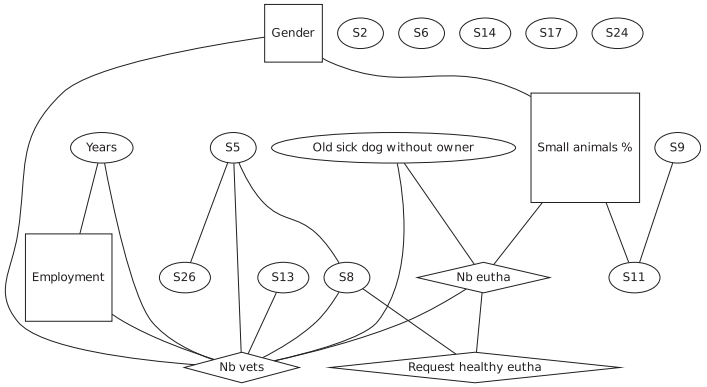


Fig. 7 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F7 (old sick dog without owner), seven demographic and eleven statements ($n = 288$). *Squares* denote variables which have been considered as *binary*, ovals as continuous and *diamond* shapes as Poisson

association was found. For F3, F4, F5, F7 and F9 either a direct association with gender and / or years spent as a veterinarian or an association via the number of veterinarians working in the same practice and the number of monthly performed euthanasia or the number of times per year being asked to euthanize a healthy animal was found. Some of the ethical and technical statements were found to be closely linked to each other and to some of the demographic variables. The number of other veterinarians working in the same practice was the variable with the highest number of links to other variables in all graphs. Younger veterinarians worked more often in a team and working in a team was associated with a higher agreement of the statements S5 "Knowing that all veterinary medical, social and economic options have been considered makes it easier for me to deal with euthanasia" and S8 "I am still not used to euthanizing animals".

Discussion

The aim of this study was to gather empirical knowledge regarding the normative and descriptive beliefs and underlying values of Austrian veterinarians regarding euthanasia in small animal practice. A questionnaire aiming at identifying agreement of veterinarians with ethical and technical statements, and/or euthanasia in case scenarios as well as significant associations with demographic variables was sent to all members of the Austrian Chamber of Veterinary Surgeons. Next to descriptive statistics, data analysis was performed using multivariable regression models and multivariate regression models (via Bayesian additive networks).

The importance of treating the owner of a euthanized animal in an understanding way was recognized by the veterinarians. This matches answers of bereaved owners highly appreciating veterinarians for their emotional support following pet death [24]. In former times, based on anecdotal evidence cited in [25], the owners were typically not present during euthanasia. Some decades ago, it was even questioned if the stay of an owner during euthanasia is beneficial or not [26]. In contrast, in recent publications [25, 27, 28] the majority of the owners is present. Accordingly, in our study the

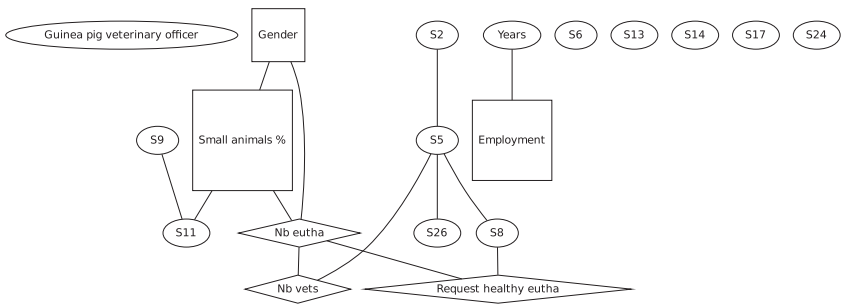


Fig. 8 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F8 (guinea pig veterinary officer), seven demographic and eleven statements ($n = 303$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

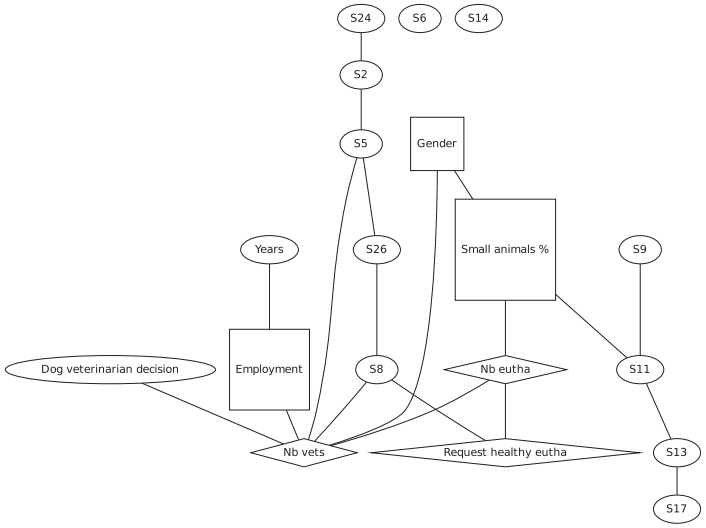


Fig. 9 Graph of the Bayesian model representing the globally optimal multivariate regression model (after bootstrapping) of scenario F9 (dog veterinarian decision), seven demographic and eleven statements ($n = 305$). *Squares* denote variables which have been considered as binary, *ovals* as continuous and *diamond* shapes as Poisson

presence of the owners during euthanasia was not perceived as a problem.

Several ethical statements, referring to an implicit idea of a "telos" or a completed life (S13, S24 and S2) [29] were ranked with a moderate agreement. The reason most often given by veterinarians in the context of wanting to refuse euthanasia, but not doing so was "Fear of what owners would otherwise do to the dog" [30]. In contrast, the responding veterinarians strongly disagreed with the statement S25 "Although I would reject euthanasia, I euthanize the animal because I am afraid that the owner will see another vet". A strong disagreement, albeit to a lesser extent was also found for the statement S20 "Although I would reject euthanasia, I euthanize the animal because I am afraid that the owner will kill it himself".

The statements were chosen and formulated based on the literature [11, 25, 29–33], but of course it is still possible that important moral aspects have not been covered in our analysis.

In the frequentist multivariable approach, the outcome of agreeing with euthanasia in all scenarios was found to be linked to at least one of the two predictors gender and years having worked as veterinarian. If significant, female veterinarians and younger veterinarians were always found to be more likely to disagree with euthanasia in the convenience euthanasia scenarios. In addition, working mostly in small animal practice, being asked more frequently to euthanize healthy animals and to a lesser extent the number of performed euthanasia were also found to be significant predictors for agreeing with euthanasia in the convenience euthanasia scenarios. Presumably, these variables are not independent, e.g. working mostly in small animal practice is likely to lead to a greater frequency of euthanasia performed on small animals. Female veterinarians might also have a preference for work in small animal practice, be on average younger than male veterinarians and being employed in a team. These mutual dependencies, e.g. confounding and collinearity might lead to biased results. Additionally the stepwise regression approach, although widely used, may introduce overfitting [34]. Thus the results of the frequentist regression models are presented, mainly for comparison with other studies, but should be interpreted with caution. In a multivariable approach it might be impossible to disentangle the "true" effect of any predictor on the outcome. Here generalizing multivariable regression to multivariate regression, allowing all variables being potentially statistically dependent, offers a richer modelling framework.

The multivariate approach gives also insights into the mutual dependencies between gender, working most of the time in small animal practice, being employed, working in a team, performing euthanasia and being asked to euthanize healthy animals. Albeit to a lesser extent compared to the frequentist multivariable approach, in the multivariate models gender and / or years were still found to be linked to the agreement of convenience euthanasia with female and younger veterinarians being more likely to disagree. Interestingly, in the graphs, the number of other vets working in the same practice was found to be the variable which was most frequently

associated with a number of other variables, including ethical statements. This highlights the role of a "team" to provide mutual support and was also suggested in a study focussing on beneficial services in addressing euthanasia-related stress in shelter workers [35, 36]. Amongst others, relationships with colleagues were important sources for job satisfaction [37].

The scenario with the highest number of links to demographic variables and statements was the scenario F3 describing the situation of a young dog which would need a costly therapy to recover. One reason for this finding could be that this scenario is closer to the situations that veterinarians face in their daily routine compared to scenarios like the rabbit breeder with the wrong coat colour or the guinea pig. It could also be that in the scenario of the rabbit breeder, no link to any other demographic or statement variable could be found because nearly every respondent disagreed with euthanasia here.

The scenarios have been based on literature suggesting the most common ethical dilemmas veterinarians are financial limitations restricting treatment options, euthanasia of healthy animals and clients wishing to continue treatment of terminally ill animals [3, 14, 29, 30, 33]. By including dogs, cats, rabbits and guinea pigs we aimed also to make the scenarios more realistic. Although some of the scenarios had been tested in another questionnaire addressing veterinarians and students of veterinary medicine, agriculture and law in Switzerland [38] and slightly modified for this questionnaire, it is possible that these situations are not seen or are badly worded, thus hampering the analysis and interpretation.

Amongst sources of ethical tension in veterinary medicine [33] describe that veterinarians may consider hamsters less morally relevant than dogs (or assume that the owner does). Based on the observation that the median of agreeing with euthanasia is lower for the rabbit scenario (F4) compared to the dog scenario F3, we conclude that there is no evidence that rabbits are considered less morally relevant than dogs (F3) in similar scenarios when high costs for therapy might be a reason for an owner to request euthanasia.

The analysis comprised a multivariate additive Bayesian networks modelling approach (ABN) which is a rather new technique. The classical ABN data formats considered datasets following a normal, binomial and Poisson distribution, whereas the scenarios and statements of the questionnaire are ordinal data.

The questionnaire comprised in total more than 50 questions. As it was time demanding, selection bias is possible with veterinarians being more sensitive about ethics being more willing to complete the questionnaire. In line with this, we cannot exclude, that the results might represent the attitudes of veterinarians being more sensitive about euthanasia than the general veterinary population.

Additionally, bias might have been introduced due to missing data. This might limit generalisability to the larger population of Austrian veterinarians. We are still confident that the results, especially the stated agreements with descriptive and normative statements are useful for a better understanding of coping strategies for veterinarians with moral stress due to euthanasia of small animals.

Conclusions

Agreement with euthanasia in specific case scenarios is not homogeneous among veterinarians. The variability in agreeing with convenience euthanasia can partly be explained by demographic factors such as gender, age and working mostly in small animal practice. Benefitting from the multivariate ABN framework, in contrast to classical multivariable models, it was possible to disentangle and assess separately the effects of different variables. Veterinarians which are female, which are younger and / or which work mostly in small animal practice are more likely to disagree with convenience euthanasia. This adds to previous findings that female, younger and veterinarians working in small or mixed practice are at a higher risk of work-related stress and suicidal thoughts demonstrating that differences due gender, age and working practice are already present in the attitudes towards euthanasia. The results of this study underlines that euthanasia is not just a professional task in order to avoid suffering on the animals' side. It rather implicates a complex situation in which veterinarians' attitude towards euthanasia is potentially affected by e.g. age, gender and working experience. The complexity of veterinarians' decisions to be taken in the context of euthanasia is further increases by the challenge to justify responsibilities. Moreover one important aspect seems to be the presence of colleagues at work - not only to discuss the medical point of view but also to provide a mutual support for several difficult experienced euthanasia cases.

Additional files

Additional file 1: Original questionnaire in German. (PDF 389 kb)

Additional file 2: Results of the proportional odds models. (DOCX 15 kb)

Additional file 3: Technical details on ABN modeling. (DOCX 14 kb)

Additional file 4: Bar plots of agreement with the 26 statements. (PDF 6 kb)

Additional file 5: Bar plots of agreement with euthanasia in the nine scenarios. (PDF 3 kb)

Additional file 6: Results of the linear regression models. (PDF 270 kb)

Additional file 7: Effect sizes of the ABN models (DAGS). (PDF 175 kb)

Abbreviations
ABN: additive Bayesian networks.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SH, SS and HG designed and coordinated the study. MP, SH and SS analysed and interpreted the data. All authors helped to draft the manuscript, read and approved the final manuscript.

Authors' information
The questionnaire is part of the master thesis of Svenja Springer. The ABN methodology is part of the PhD project of Marta Pittavino. She is a PhD candidate in the Epidemiology and Biostatistics PhD Program of Life Science Zurich Graduate School and gratefully acknowledges its support.

Acknowledgements
We are grateful to all veterinarians who completed the questionnaire. The authors thank the Österreichische Tierärztekammer (ÖTK) for supporting this study. We also gratefully acknowledge the useful comments from Dr. Regine Binder from the Institute of Animal Husbandry and Animal Welfare, University of Veterinary Medicine Vienna, when preparing the questionnaire. We would also like to thank Professor Ernst Singer (http://ethikkommission.meduniwien.ac.at/ethik-kommission/ueber-uns/) for his advice concerning the need of ethical approval. We gratefully acknowledge funding by the Fondazione Franco e Marilisa Caligara per l'Alta Formazione Interdisciplinare and by the Swiss National Science Foundation (PBBEBS-124186, SNF138562 and SNF144973).

Author details
[1]Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Winterthurerstr. 270, 8057 Zurich, Switzerland. [2]Unit of Ethics and Human-Animal-Studies, Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University Vienna, and University of Vienna, Veterinaerplatz 1, A-1210 Vienna, Austria. [3]Institute of Mathematics, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland.

Received: 7 September 2015 Accepted: 28 January 2016

References

1. Morris P. Blue juice : euthanasia in veterinary medicine. Philadelphia: Temple University Press; 2012.
2. Stauch S. Euthanasie in der Kleintierpraxis [The euthanasia of dogs and cats in a small animal veterinary practice]. Freie Universität Berlin; 2006. http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000002893.
3. Batchelor CE, McKeegan DE. Survey of the frequency and perceived stressfulness of ethical dilemmas encountered in UK veterinary practice. Vet Rec. 2012;170:19–U55. doi:10.1136/vr.100262.
4. Stark C, Dougall N. Effect of attitudes to euthanasia on vets' suicide risk. Vet Rec. 2012;171:172–3. doi:10.1136/vr.e5494.
5. Platt B, Hawton K, Simkin S, Mellanby RJ. Suicidal behaviour and psychosocial problems in veterinary surgeons: a systematic review. Soc Psychiatry Psychiatr Epidemiol. 2012;47:223–40. doi:10.1007/s00127-010-0328-6.
6. Gardner DH, Hini D. Work-related stress in the veterinary profession in New Zealand. NZVJ. 2006;54:119–24.
7. Bartram DJ, Baldwin DS. Veterinary surgeons and suicide: a structured review of possible influences on increased risk. Vet Rec. 2010;166:388–97. doi:10.1136/vr.b4794.
8. Passantino A, Quartarone V, Russo M. Informed consent in Italy: its ethical and legal viewpoints and its applications in veterinary medicine. ARBS Annu Rev Biomed Sci. 2012;14:16–26.
9. Tannenbaum J. Veterinary medical ethics: A focus of conflicting interests. J Soc Issues. 1993;49:143–56. doi:10.1111/j.1540-4560.1993.tb00914.x.
10. Passantino A, Fazio A, Quartarone V. Pain in veterinary medicine in the new millenium. Theor Biol Forum. 2012;105:77–85.
11. Rollin BE. Ethics and euthanasia. Can Vet J. 2009;50:1081–6.
12. Rollin BE. Euthanasia, moral stress, and chronic illness in veterinary medicine. Vet Clin North Am Small Anim Pract. 2011;41:651–9. doi:10.1016/j.cvsm.2011.03.005.
13. Nolen RS. AVMA board approves panel on euthanasia report: updated guidelines cover more species and methods. J Am Vet Med Assoc. 2011;239:1269.
14. Magalhães-Sant'Ana M. Ethics teaching in European veterinary schools: a qualitative case study. Vet Rec. 2014;175:592. doi:10.1136/vr.102553.
15. Beauchamp TL, Childress JF. Principles of biomedical ethics. 7th ed. New York: Oxford University Press; 2013.
16. Kalof L, Resl B, Boehrer B, Senior M, Kete K, Malamud R. A cultural history of animals. Oxford: Berg Publishers; 2011.
17. LimeSurvey: An Open Source survey tool: LimeSurvey Project Team; 2012. http://www.limesurvey.org. Accessed 29 Oct 2015
18. Venables WN, Ripley BD. MASS: Modern Applied Statistics with S. New York: Springer; 2002. Available from: http://www.stats.ox.ac.uk/pub/MASS4. Accessed 29 Oct 2015.
19. R Core Team. A language and environment for statistical: R Foundation for Statistical Computing; 2015. http://www.R-project.org/. Accessed 29 Oct 2015
20. Lewis FI. abn: Data Modelling with Additive Bayesian; 2014. https://cran.r-project.org/web/packages/abn/index.html.
21. Lewis FI, McCormick BJ. Revealing the complexity of health determinants in resource-poor settings. Am J Epidemiol. 2012;176:1051–9. doi:10.1093/aje/kws183.
22. Firestone SM, Lewis FI, Schemann K, Ward MP, Toribio JA, Taylor MR, et al. Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a subsequent outbreak. Prev Vet Med. 2014;116:243–51. doi:10.1016/j.prevetmed.2013.11.015.
23. Schemann K, Lewis FI, Firestone SM, Ward MP, Toribio JA, Taylor MR, et al. Untangling the complex inter-relationships between horse managers' perceptions of effectiveness of biosecurity practices using Bayesian graphical modelling. Prev Vet Med. 2013;110:37–44. doi:10.1016/j.prevetmed.2013.02.004.
24. Adams CL, Bonnett BN, Meek AH. Predictors of owner response to companion animal death in 177 clients from 14 practices in Ontario. J Am Vet Med Assoc. 2000;217:1303–9.
25. Morris P. Managing pet owners' guilt and grief in veterinary euthanasia encounters. J Contemp Ethnogr. 2012;41:337–65. doi:10.1177/0891241611435099.
26. Edney AT. Management of euthanasia in small animal practice. J Am Anim Hosp Assoc. 1979;15:645–9.
27. Hewson C. Grief for animal companions and an approach to supporting their bereaved owners. Bereave Care. 2014;33:103–10. doi:10.1080/02682621.2014.980985.
28. Dickinson GE, Roof PD, Roof KW. A survey of veterinarians in the US: Euthanasia and other end-of-life issues. Anthrozoös. 2011;24:167–74. doi:10.2752/175303711X12998632257666.
29. Fahrion A, Duerr S, Doherr MG, Hartnack S, Kunzmann P. Killing and dignity of animals: a problem for veterinarians? Schweiz Arch Tierheilkd. 2011;153:209–14. doi:10.1024/0036-7281/a000184.
30. Yeates JW, Main DCJ. Veterinary opinions on refusing euthanasia: justifications and philosophical frameworks. Vet Rec. 2011;168:263. doi:10.1136/vr.c6352.
31. Yeates J. Ethical aspects of euthanasia of owned animals. In Pract. 2010;32:70–3. doi:10.1136/inp.c516.
32. Main DCJ. Views on euthanasia. Vet Rec. 2007;161:144. doi:10.1136/vr.161.4.144.
33. \Morgan CA, McDonald M. Ethical dilemmas in veterinary medicine. Vet Clin North Am Small Anim Pract. 2007;37:165–79. doi:10.1016/j.cvsm.2006.09.008.
34. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med. 2004;66:411–21.
35. Rogelberg SG, Digiacomo N, Reeve CL, Spitzmueller C, Clark OL, Teeter L, et al. What shelters can do about euthanasia-related stress: An examination of recommendations from those on the front line. J Appl Anim Welf Sci. 2007;10:331–47.
36. Anderson KA, Brandt JC, Lord LK, Miles EA. Euthanasia in animal shelters: management's perspective on staff reactions and support programs. Anthrozoös. 2013;26:569–78. doi:10.2752/175303713X13795775536057.

37. Bartram DJ, Yadegarfar G, Baldwin DS. A cross-sectional study of mental health and well-being and their associations in the UK veterinary profession. Soc Psychiatry Psychiatr Epidemiol. 2009;44:1075–85. doi:10.1007/s00127-009-0030-8.
38. Duerr S, Fahrion A, Doherr MG, Grimm H, Hartnack S. Acceptance of killing of animals: Survey among veterinarians and other professions. Schweiz Arch Tierheilkd. 2011;153:215–22. doi:10.1024/0036-7281/a000185.

Supplementary materials outline:

Euthanasie in der Kleintierpraxis

Umfrage zur Euthanasie in der Kleintiermedizin

Liebe Kolleginnen und Kollegen aus der tierärztlichen Praxis,

das Thema „Tötung von Tieren“ spielt in der gesellschaftlichen Wahrnehmung eine zunehmend wichtige Rolle. Der Wandel der Mensch-Tier-Beziehung konfrontiert die praktischen Tierärztinnen und Tierärzte im Berufsleben und stellt sie vor neue Herausforderungen. Sie haben in Ihrem Berufsleben professionelle Erfahrungen gemacht, die für angehende Tierärztinnen und Tierärzte von großem Wert sein können. Wir möchten mit dem folgenden Fragebogen ein Bild der Ansichten zum Thema „Tötung von Tieren“ der österreichischen Tierärzteschaft erstellen: Was sind die prägenden Einstellungen? Was macht die Euthanasie zu (k)einem Problem? Welche Tipps und Strategien sind Ihnen wichtig?

Dieses interdisziplinäre Projekt wird gemeinsam im Rahmen einer Diplomarbeit von Svenja Springer unter der Leitung von Prof. Herwig Grimm (Messerli Forschungsinstitut) und Prof. Yves Moens (Veterinärmedizinische Universität Wien) in Kooperation mit Frau Dr. Sonja Hartnack (VetSuisse Zürich) durchgeführt.

Ziel ist es, die Ergebnisse auch dafür zu verwenden, um die angehenden Tierärztinnen und Tierärzte besser auf den Berufsalltag vorzubereiten. Die Ergebnisse der Befragung werden Ihnen über die Homepage des Messerli Forschungsinstitutes zugänglich gemacht und in geeigneter Form publiziert.

Das Ausfüllen des Fragebogens dauert ca. 15-25 Minuten.

Ihre Angaben werden vertraulich behandelt und selbstverständlich anonymisiert ausgewertet.

Diese Umfrage enthält 53 Fragen.

Fallbeispiele

Im folgenden Abschnitt geht es um Ihre Einstellung zur Euthanasie. Wir stellen Ihnen verschiedene Fallbeispiele vor und bitten Sie, uns Ihre persönliche Sicht mitzuteilen. Zur Beantwortung der einzelnen Fragen finden Sie eine Skala von 1 bis 9. Dabei bedeutet 1, dass Sie die Euthanasie sicherlich ablehnen, und 9, dass Sie der Euthanasie sicherlich zustimmen.

1 [F1]

Ein Hund hat bereits zweimal eine Person gebissen. In Erziehungskursen und bei Tierpsychologen wurde versucht, ihn zu erziehen. Vor zwei Tagen aber hat er ein Kind so stark verletzt, dass es seither im Spital liegt.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Zustimmung 9**
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

2 [F2]

Eine Kaninchenzüchterin möchte einige ihrer Jungtiere euthanasieren lassen, da die Fellfarbe nicht dem Zuchtstandard entspricht und sie folglich keinen Erfolg auf Ausstellungen mit den Tieren haben wird.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Zustimmung 9**
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

3 [F6]

Ein Tierbesitzer kommt mit seiner schwerkranken Perserkatze zu Ihnen in die Ordination. Sie wissen, dass dieser eine sehr enge Bindung zu seiner Katze hat und sich nicht von ihr trennen möchte. Aus Ihrer Sicht wäre eine Euthanasie angezeigt, aber der Tierbesitzer ist nicht einverstanden. Sie lehnen jede weitere Therapie, ausser Schmerzbehandlung ab.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Zustimmung 9**
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

4 [F3]

Ein Besitzer kommt mit einem jungen Hund in ihre Ordination. Dieser Hund ist schwerkrank, aber therapierbar. Diese Therapie würde viel Zeit in Anspruch nehmen, aber Erfolgschancen bestehen. Der Tierbesitzer verneint die Therapie, da er keine Zeit und finanziellen Mittel hat, und möchte, dass Sie das Tier euthanasieren.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Zustimmung 9**
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

5 [F4]

Ein Kaninchenbesitzer kommt zu Ihnen in die Ordination. Das Tier hat eine gut therapierbare Krankheit, die jedoch eine gewisse Zeit in Anspruch nimmt und mit einem Kostenfaktor von ungefähr 150€ verbunden ist. Der Besitzer möchte das Geld für die entsprechende Therapie nicht ausgeben, das kranke Kaninchen einschläfern und sich für 40€ ein neues Kaninchen kaufen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Zustimmung 9**
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

6 [F7]

In Ihre Ordination kommt eine Hundesitterin mit einem 17-jährigen Hund der an Atemproblemen leidet. Die Besitzer sind vor drei Tagen zu einer vierwöchigen Trekkingtour aufgebrochen und nicht erreichbar. Sie haben bei diesem Hund vor sechs Monaten ein malignen Tumor entfernt und befürchten nun, dass sich Lungenmetastasen entwickelt haben. Die Hundesitterin weigert sich, eine Entscheidung bezüglich Euthanasie zu treffen, und kann Ihnen auch keine Auskunft darüber geben, was die Besitzer vermutlich möchten.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Zustimmung 9**
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

7 [F5]

Eine Hundebesitzerin kommt mit dem Anliegen zu Ihnen in die Ordination, ihren Hund zu euthanasieren. Die Begründung lautet, dass ihr 15-jähriger Hund nicht mehr zu ihren Lebensumständen passt, da sie mit ihrer Familie für längere Zeit verreist und den Hund in diesem Alter nicht ins Tierheim abgegeben möchte.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	2	3	4	5	6	7	8	Zustimmung 9
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = Euthanasie ablehnen; 9 = Euthanasie befürworten).

8 [F8]

Ein Meerschweinchenbesitzer kommt zu Ihnen in die Ordination. Das Meerschweinchen frisst nicht. Sie stellen bei der Untersuchung einen walnussgroßen Tumor im Bereich des Abdomens fest. Aufgrund des schlechten Allgemeinzustandes des Tieres, sind Sie der Meinung, dass die Prognose infaust ist und raten dem Besitzer zu einer Euthanasie. Der Tierbesitzer hält den Zustand seines Tieres für unproblematisch und möchte das Tier wieder mit nach Hause nehmen und es nicht euthanasieren lassen. Der Amtstierarzt muss davon in Kenntnis gesetzt werden.

Hier möchten wir gerne von Ihnen wissen, ob Sie dem letzten Satz, dass "der Amtstierarzt in Kenntnis gesetzt werden muss", zustimmen oder nicht.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ablehnung 1	2	3	4	5	6	7	8	Zustimmung 9
Ihre Einstellung zur Euthanasie	○	○	○	○	○	○	○	○	○

Bitte kreuzen Sie die zutreffende Antwort auf der Skala an (1 = "Amtstierarzt in Kenntnis setzen" ablehnen; 9 = "Amtstierarzt in Kenntnis setzen" befürworten).

weitere Fallbeispiele

Bei den beiden folgenden Fallbeispielen möchten wir Sie bitten, dass Sie uns mithilfe der Antwortskala antworten.

9 [F9]

Ein Ehepaar kommt mit einem Hund mit fortgeschrittener Arthrose zu Ihnen und fragt Sie: „Was würden Sie tun, wenn es Ihr Tier wäre?" Sie klären die Besitzer gewissenhaft auf. Sie sind der Meinung, dass es sich um eine 50/50-Situation handelt und die Eheleute Ihrer Einschätzung folgen würden. Würden Sie sich weigern, eine eindeutige Empfehlung abzugeben und damit die Entscheidung zu übernehmen?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	sicher keine Empfehlung abgeben 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**sicher Empfehlung abgeben 9**
Empfehlung abgeben?	○	○	○	○	○	○	○	○	○

Bitte klicken Sie Ihre Antwort auf der Skala an (1 = ich würde keine Empfehlung abgeben, 9 = ich würde eine Empfehlung abgeben).

10 [F10]

Ein Unfallhund kommt in Seitenlage mit schwersten inneren Verletzungen in Ihre Ordination. Die Besitzer sind sehr aufgebracht und besorgt um Ihr Tier. Sie stellen während der Untersuchung fest, dass der Hund aufgrund der Unfallverletzungen nicht mehr am Leben zu halten ist und schlagen den Besitzern eine Euthanasie vor. Die Besitzer entscheiden sich für die Euthanasie. Würden Sie hier, trotz des komatösen Zustands des Tieres, noch einen Venenkatheter legen für eine iv-Applikation des Euthanasiepräparates?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

sicher Venenverweilkatheter legen 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**sicher keinen Venenverweilkatheter legen 9**
○	○	○	○	○	○	○	○	○

Bitte beantworten Sie diese Frage mithilfe der Antwortskala. Dabei bedeutet 1 = „Ich würde sicherlich einen Venenkatheter legen" und 9 = „Ich würde sicherlich keinen Venenkatheter

Aussagen und Statements I

In den folgenden drei Abschnitten finden Sie eine Reihe von Aussagen und Statements, die wir in der Literatur gefunden oder in Gesprächen mit Tierärzten und Tierärztinnen gehört haben. Einige der Aussagen beziehen sich ganz allgemein auf den Umgang mit der Euthanasie und Umstände, die den Umgang damit erleichtern oder erschweren. Andere Aussagen beziehen sich eher auf die praktische Gestaltung und Durchführung.

Hier ist der erste Abschnitt.

11 [S1]

Das Wissen, mich für das Wohl des Tieres eingesetzt zu haben, macht es mir leichter, mit dem Leiden des Tieres umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**stimme völlig zu 9**
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala. Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

12 [S2]

Das Wissen, dass das Tier nur noch eine kurze Lebensspanne vor sich hatte, macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**stimme völlig zu 9**
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala. Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

13 [S3]

Die sorgfältige Aufklärung des Tierbesitzers macht es mir leichter, mit der Euthanasie

umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala. Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

14 [S4]

Die Zufriedenheit des Kunden, bezüglich der Tötung seines Tieres, macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala. Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

15 [S5]

Das Wissen, dass alle veterinärmedizinischen, wie auch sozialen und ökonomischen Möglichkeiten bedacht wurden, macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**stimme völlig zu 9**
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

16 [S6]

Es fällt mir leichter, ein Tier zu euthanasieren, wenn ich sehe, dass die Tierbesitzer keine intensive Bindung zum Tier haben.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**stimme völlig zu 9**
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

17 [S7]

Wenn die Tierbesitzer bei der Euthanasie anwesend sind, erleichtert mir dies, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

18 [S8]

Ich habe mich noch immer nicht daran gewöhnt, Tiere zu euthanasieren.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

19 [S9]

Die sorgfältige Planung und die geschickte Wahl des Zeitpunktes machen es leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

20 [S10]

Die technisch einwandfreie Durchführung der Tötung des Tieres macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und

eine 9 = „Ich stimme völlig zu“.

Aussagen und Statements II

Hier beginnt der zweite Abschnitt.

21 [S11]

Eine wirksame Schmerztherapie macht es mir leichter, mit dem Leiden des Tieres umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

22 [S12]

Die sorgfältige Aufklärung des Patientenbesitzers macht es mir leichter, mit dem Leiden des Tieres umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

23 [S13]

Mein Wissen, dass das Tier bis zum Zeitpunkt der Euthanasie ein erfülltes Leben hatte, macht es mir leichter mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

24 [S14]

Belastend wäre es für mich, wenn ich gegen die eigene Überzeugung ein Tier euthanasiere.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

25 [S15]

Die Anwesenheit des Tierbesitzers verursacht tendenziell mehr Probleme.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

26 [S16]

Der respektvolle Umgang mit dem toten Tier ist ein wichtiger Teil der Euthanasie.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

27 [S17]

Der verständnisvolle Umgang mit den Tierbesitzern ist ein zentraler Bestandteil der Euthanasie.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

28 [S18]

Rückblickend fällt es mir zunehmend leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte klicken geben Sie Ihre Antwort mit durch Anklicken der entsprechenden Option auf der Skala an. Dabei bedeutet eine 1 = "ich stimme überhaupt nicht zu" und

eine 9 = "ich stimme völlig zu".

Aussagen und Statements III

Nun der dritte Abschnitt.

29 [S19]

Es fällt mir schwerer (unter sonst gleichen Bedingungen), ein besitzerloses Tier zu euthanasieren.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

30 [S20]

Obwohl ich eine Euthanasie eigentlich ablehnen würde, mache ich es dennoch, weil ich befürchte, dass der Tierbesitzer sein Tier eigenhändig töten würde.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

31 [S21]

Das Wissen, mich für das Wohl des Tieres eingesetzt zu haben, macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

32 [S22]

Ich erlebe Euthanasie als unvermeidliches Übel meiner Verantwortung.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

33 [S23]

Die Einsicht, dass die Möglichkeiten meiner Einflussnahme auf die Entscheidung des Besitzers begrenzt sind, macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

34 [S24]

Das fortgeschrittene (hohe) Alter des Tieres macht es mir leichter, mit der Euthanasie umzugehen.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

35 [S25]

Obwohl ich eine Euthanasie eigentlich ablehnen würde, mache ich es dennoch, weil ich befürchte, dass der Tierbesitzer in eine andere Ordination geht.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

36 [S26]

Ich sehe die wohlüberlegte Euthanasie als wesentlichen Bestandteil meiner tierärztlichen Tätigkeit.

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

stimme überhaupt nicht zu 1	2	3	4	5	6	7	8	stimme völlig zu 9
○	○	○	○	○	○	○	○	○

Bitte geben Sie Ihre Antwort durch Anklicken der entsprechenden Option auf der Skala.
Dabei bedeutet eine 1 = „Ich stimme überhaupt nicht zu“ und eine 9 = „Ich stimme völlig zu“.

Allgemeine und demografische Fragen

Im letzten Abschnitt haben wir noch ein paar demografische Fragen zu Ihnen, Ihrer beruflichen Tätigkeit und zur Euthanasie.

37 [D1]

Wie viel Arbeitszeit verbringen Sie mit der Behandlung von Kleintieren (Hund, Katze, Kaninchen und Meerschweinchen) in Ihrer Ordination?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

< 20 %	20 - 40 %	41 - 60 %	61 - 80 %	> 80 %
○	○	○	○	○

38 [D2]

Sind Sie selbstständig in Ihrem Beruf tätig oder arbeiten Sie im Angestelltenverhältnis in einer Ordination?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

selbständig	angestellt
○	○

39 [D3]

Wie viele Tierärzte und Tierärztinnen sind außer Ihnen in Ihrer Ordination beschäftigt?

Bitte geben Sie Ihre Antwort hier ein:

-

Bitte gegen Sie hier die Anzahl der Tierärzte/Tierärztinnen an, die ebenfalls in Ihrer Ordination tätig sind.

40 [D4]

Wie häufig werden in Ihrer Ordination Kleintiere durchschnittlich pro Monat euthanasiert?

Bitte geben Sie Ihre Antwort hier ein:

-

Bitte geben Sie die durchschnittliche monatliche Anzahl der euthanasierten Kleintiere an.

41 [D5]

Wie häufig pro Monat euthanasieren Sie selber?

Bitte geben Sie Ihre Antwort hier ein:

-

Bitte geben Sie hier an wie häufig Sie selber pro Monat euthanasieren.

42 [D6]

Wie häufig werden Sie jährlich schätzungsweise ersucht, ein (weitgehend) gesundes Tier zu euthanasieren?

Bitte geben Sie Ihre Antwort hier ein:

-

Bitte geben Sie hier die Häufigkeit pro Jahr an.

43 [D7]

Wie häufig werden Sie jährlich schätzungsweise ersucht, einen gesunden „Kampfhund“ zu euthanasieren?

Bitte geben Sie Ihre Antwort hier ein:

-

Bitte geben Sie hier die Häufigkeit pro Jahr an.

44 [D8]

Wenn Tierhalter ein (weitgehend) gesundes Tier euthanasieren lassen möchten, welche sind die drei häufigsten genannten Gründe?

Bitte geben Sie Ihre Antwort(en) hier ein:

- 1. häufigster Grund
- 2. häufigster Grund
- 3. häufigster Grund

45 [D9]

Welches Euthanasiepräparat verwenden Sie in Ihrer Ordination für die Euthanasie von Kleintieren (Hund, Katze, Kaninchen, Meerschweinchen)? (Mehrfachantwort möglich)

Bitte wählen Sie alle zutreffende Einträge aus und schreiben Sie einen Kommentar dazu:

- ☐ Kombinationspräparate mit Embutramid
- ☐ Derivate der Barbitursäure
- ☐ Inhalationsanästhetika
- ☐ Andere (bitte nennen)

46 [D10]

Wie lange sind Sie bereits tierärztlich tätig?

Bitte geben Sie Ihre Antwort hier ein:

•

Bitte geben Sie hier wieviele Jahre Sie bereits tierärztlich tätig sind.

47 [D11]

Bitte sagen Sie uns auch, wie alt Sie sind:

Bitte geben Sie Ihre Antwort hier ein:

•

Bitte geben Sie Ihr Alter in Jahren an.

48 [D12]

Bitte geben Sie hier Ihr Geschlecht an.

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ weiblich
- ○ männlich

49 [D13]

Was (oder wer) hat Sie am besten für Ihre beruflichen Aufgaben im Bereich der Euthanasie vorbereitet?

Bitte geben Sie Ihre Antwort hier ein:

50 [D14]

Was oder wer hilft Ihnen heute, um mit Euthanasien umzugehen?

Bitte geben Sie Ihre Antwort hier ein:

51 [D15]

Würden Sie sich mehr Unterstützung hinsichtlich der Thematik Euthanasie wünschen, und wenn ja von wem?

Bitte geben Sie Ihre Antwort hier ein:

52 [D18]

Würde Ihnen ein Kriterienkatalog zur Unterstützung in schwierigen Entscheidungssituationen bezüglich der Euthanasie helfen?

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Ja	**Unsicher**	**Nein**
	○	○	○

53 [D19]

Gibt es weitere Kommentare zum Thema Euthanasie oder zum Fragebogen, die Sie uns mitteilen möchten?

Bitte geben Sie Ihre Antwort hier ein:

Vielen Dank für das Ausfüllen des Fragebogens.

Mit freundlichen Grüßen

Svenja Springer, Sonja Hartnack und Herwig Grimm

Bitte übermitteln bis 11.12.2012 – 00:00

Übermittlung Ihres ausgefüllten Fragebogens:
Vielen Dank für die Beantwortung des Fragebogens.

Supplementary section on proportional odds logistic regression

Material and methods

Multivariable proportional odds logistic regressions were utilized to assess potential significant associations between the outcome "agreement with euthanasia" in each of the different scenarios and demographic data. The outcome variable ranged from total rejection to full agreement on a 9-point scale. The demographic predictor variables included: percentage of working time spent with small animals (dichotomized into ≤ 60 % and > 60 %) (*Small animals %*), working employed or self-employed (*Employment*), number of other veterinarians working in the same practice (*Nb vets*), number of euthanasia per month performed by the respondent (*Nb eutha*), number of times per year the respondent is asked to perform euthanasia of a healthy animal (*Request healthy eutha*), years working as a vet (*Years*) and gender (*Gender*). Stepwise model selection (backward and forward) by Akaike's information criterion (AIC) was performed using the MASS package [1] in R [2]. Only complete questionnaires were utilised for the analysis. The results of the multivariable proportional odds regression approach are presented as remaining predictors in the final models, p-values, proportional odds ratios and their corresponding confidence intervals.

For the multivariable proportional odds regression approach, the predictors in the final models, p-values, proportional odds ratios (pOR) and their corresponding confidence intervals are presented. The interpretation of a pOR of two for example would be that for a one unit increase in the predictor variable, the odds the highest category in the outcome variable (would be 9 equal to "agree"), the agreement of euthanasia in a specific scenario, versus the other eight lower combined categories are two times higher, given that all other variables are held constant in the model. Likewise, for a one unit increase in the predictor variable, the odds of the combined two highest categories 8 and 9 versus the other combined categories 1 to 7 are two times greater given that all other variables are held constant in the model and so on.

Results

Scenarios F1 to F5 could be titled “convenience euthanasia” meaning settings in which veterinarians are confronted with an owner requesting euthanasia which would be presumably against the veterinarian’s view.

F1 describes a scenario about a dog which had already bitten humans twice, attended training courses after the incident and visited animal psychologists and severely injured a child afterwards. The final model included only gender ($p<0.001$) with female veterinarians having a pOR of 0.41 [0.23;0.72] compared to males, thus being more likely to disagree with euthanasia in this scenario.

For F2, illustrates a scenario about a rabbit breeder asking for euthanasia of some of her young animals because of coat colour not meeting breeding standards thus excluding success at exhibitions. Here it was not possible to run proportional odds models, since almost all respondents rejected euthanasia in this scenario.

F3 describes a scenario in which an owner asks for euthanasia of a young dog which is severely ill, but could possibly be cured with an appropriate, albeit costly and time-consuming therapy. For F3, three variables remained in the final model: percentage of time spent in small animal practice, the number of times per year the respondents is asked by an owner to euthanize a healthy animal and gender. Spending more than 60 % of working time in small animal practice ($p=0.009$) led to a decrease in agreement with euthanasia with a pOR of 0.49 [0.28;0.83]. Being requested to euthanize a healthy animal with a $p =0.052$ and a pOR of 0.95 [0.91;1] was associated with a tendency to reject euthanasia with more requests. Females were found to be more likely to disagree with euthanasia ($p=0.005$) with a pOR of 0.47 [0.28;0.78].

In the scenario F4, a rabbit owner prefers to euthanize his animal, although the animal might be treated, and buy a new rabbit which would be cheaper. For F4, the final model included

three variables: the working time spent in small animal practice (p=0.12), being self-employed or employed (p=0.04) with employed veterinarians being more likely to reject euthanasia with a pOR of 0.49 [0.24;0.95], and the number of euthanasia performed monthly by the respondent (p=0.07).

F5 describes a scenario in which a dog owner wants her 15 year old dog to be euthanized because it no longer fits to her living conditions. She prefers to travel and does not want to bring her dog to the animal shelter at this age. In this scenario, six variables were present in the final model: percentage of time spent in small animal practice (p=0.002), number of euthanasia performed monthly by the respondent (p<0.001), number of times per year the respondent is asked to euthanize a healthy animal (p=0.03), years having worked as a veterinarian (p=0.001) and gender (p=0.13). Respondents spending more than 60 % of their working time in small animal practice were more likely to disagree with euthanasia with a pOR of 0.48 [0.25;0.9]. With an increasing number of self-performed euthanasia, respondents were more likely to agree with euthanasia with a pOR of 1.2 [1.1;1.3]. Being asked more often to euthanize a healthy animal, the respondents were more likely to disagree with euthanasia with a pOR of 0.93 [0.87;0.99].

Scenarios F6 and F7 describe situations in which euthanasia might be recommended on veterinary reasoning, but the owner or person in charge refuses it.

F6 describes a scenario in which an owner refuses euthanasia of a severely ill Persian cat, having a very close relationship with his cat. Here, the number of times per year the respondents is asked by an owner to euthanize an healthy animal was the sole variable based on AIC to be associated with the agreement of euthanasia, albeit a p-value of 0.12.

In F7, a dog sitter refuses to take the decision of euthanasia of an old dog with breathing problems and a history of malignancy, and the owner cannot be reached. In this scenario,

solely the years having worked as veterinarians remained in the model ($p<0.001$) with more professional experience being more likely to agree with euthanasia with a pOR of 1.05 [1.02;1.08].

In the scenario F8, a guinea pig owner refuses euthanasia of his animal with a tumour and wants to take it home instead. The question is raised if the official veterinarian has to be informed. In the final model remained two variables: being self-employed or employed ($p=0.016$) and the number of times per year the respondent is asked to perform euthanasia of a healthy animal ($p=0.13$). Being employed was found to be more likely to agree with the statement that the official veterinarian should be informed with a pOR of 2.06 [1.14;3.72].

F9 describes a scenario in which the owners urge the veterinarian to take the decision of euthanasia and the respondent is asked if she would reject the responsibility of taking a decision for or against euthanasia if on veterinary medical grounds both decisions could be justified. In the final model remained two variables: the number of other vets working in the same practice ($p=0.016$) and gender ($p<0.001$). Veterinarians working in a team and female veterinarians were less likely to take the decision at the owners' place with a pOR of 0.92 [0.87;0.98] and a pOR of 0.41 [0.24;0.67], respectively.

In summary, for the "convenience euthanasia" scenarios, in three out of four scenarios spending more or most of the working time in small animal practice was found to be significantly associated with disagreeing with euthanasia. Gender was found in two out of four scenarios to be significantly associated with disagreeing with euthanasia. Type of employment or more specifically being employed instead of working self-employed was found in one scenario (F4) to be significantly associated with disagreeing with euthanasia. In one scenario (F5) the number of euthanasias performed by the respondent was found to be significantly associated with the agreement with more animals euthanized being associated with a higher agreement. In contrast, in the same scenario, the number of times the respondent

had been asked to euthanize a healthy animal was associated with disagreeing with euthanasia. For the two scenarios in which euthanasia is refused, solely the number of professional years was associated with agreement with more experienced veterinarians agreeing more with euthanasia. With regard to the perceived need to inform an official veterinarian, type of employment was found to be significantly associated with employed veterinarians being more likely to inform the official veterinarian. In a scenario in which – on veterinary reasoning – no clear recommendation in favour or against euthanasia was possible, solely gender (female) and the number of veterinarians working in the same practice was found to be significantly associated with declining to take the decision at the place of the owners.

References

1. Venables WN, Ripley BD. MASS: Modern Applied Statistics with S. New York: Springer; 2002. Available from: URL: http://www.stats.ox.ac.uk/pub/MASS4. Accessed 9 Sept 2015.
2. R Core Team. A language and environment for statistical: R Foundation for Statistical Computing; 2015. Available from: URL: http://www.R-project.org/. Accessed 9 Sept 2015.

Supplementary material and methods section on additive Bayesian networks (ABN)

For the ABN analysis, first the optimal model in terms of the one with the highest marginal likelihood (model score) was determined by increasing subsequently the number of parents per node from one to nine. The marginal likelihood was considered as a goodness of fit metric including an implicit penalty for model complexity and was estimated using Laplace approximation at each node [1]. This process of identifying an optimal Bayesian graph is referred to in the literature as structure learning [2,3]. To allow for an exact search method approach [4] 19 variables (one scenario, seven demographic and eleven statements) were selected [5]. With respect to the structure a uniform prior was chosen assuming that all Bayesian graphical structures were equally plausible. Second in order to adjust for overfitting, bootstrapping with Markov chain Monte Carlo (MCMC) simulations in JAGS [6] were performed. Simulated datasets with an identical size as the original one were generated and an identical exact search was performed. At least 256 bootstraps were run for each of the nine scenarios with the demographic variables and eleven statements. Arcs or lines, representing associations between two nodes, present in less than 50% of the globally optimal Bayesian graph estimated from the bootstrap data were considered to be not robust enough and trimmed off from the Bayesian graph generated in the first step. A threshold of 50% structural support is the usual cut-off in Bayesian network analysis [7]. The analysis were performed using the software R [8] and the package abn [9]. The resulting networks or Bayesian graphs were visualized with GraphViz [10].

References

1. Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. American Statistical Association 1986; 81:82–6.
2. Friedman N, Koller D. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. Mach Learn 2003; 50:95–125.
3. Heckerman D, Geiger D, Chickering DM. Learning Bayesian Networks – the combination of knowledge and statistical-data. Mach Learn 1995; 20:197–243.
4. Koivisto M, Sood K. Exact Bayesian structure discovery in Bayesian networks. J Mach Learn Res 2004; 5:549–73.
5. Lewis FI, Ward MP. Improving epidemiologic data analyses through multivariate regression modelling. Emerg Themes Epidemiol 2013; 10:4.
6. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling; 2003. Available from: URL: http://mcmc-jags.sourceforge.net/. Accessed 9 Sept 2015.
7. Lewis FI, McCormick, Benjamin J. J. Revealing the complexity of health determinants in resource-poor settings. Am J Epidemiol 2012; 176:1051–9.
8. R Core Team. A language and environment for statistical: R Foundation for Statistical Computing; 2015. Available from: URL: http://www.R-project.org/. Accessed 9 Sept 2015.
9. Lewis FI. abn: Data Modelling with Additive Bayesian Networks; 2014. Available from: URL: http://CRAN.R-project.org/package=abn. Accessed 9 Sept 2015.
10. Junger M, Mutzel P, editors. Graph Drawing Software: Graphviz and dynagraph - static and dynamic graph drawing tools. Heidelberg: Springer-Verlag; 2003.

statement S1
statement S2
statement S3
statement S4
statement S5
statement S6
statement S7
statement S8
statement S9
statement S10
statement S11
statement S12
statement S13
statement S14
statement S15
statement S16
statement S17
statement S18
statement S19
statement S20
statement S21
statement S22
statement S23
statement S24
statement S25
statement S26
frequencies
0 100 300 500
agreement
1 2 3 4 5 6 7 8 9

scenario aggressive dog (F1)
scenario rabbit breeder (F2)
scenario young dog costly therapy (F3)
scenario rabbit costly therapy (F4)
scenario dog not fitting living conditions (F5)
scenario Persian cat (F6)
scenario old sick dog without owner (F7)
scenario guinea pig veterinary officer (F8)
scenario dog veterinarian decision (F9)
frequencies
agreement

"convenience euthanasia"

Scenario F1	**Small animals %**					
aggressive dog	(1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	
	univariable	456	0.094	336	0.060	
	multivariable			336	-	
	Employment					
	(1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	413	0.054	336	0.085	
	multivariable			336	-	
	Nb vets					
		n (all)	p-value	n (complete)	p-value	
	univariable	403	0.513	336	0.686	
	multivariable			336	-	
	Nb eutha					
		n (all)	p-value	n (complete)	p-value	
	univariable	386	0.240	336	0.311	
	multivariable			336	0.300	
	Request healthy eutha					
		n (all)	p-value	n (complete)	p-value	effect size
	univariable	396	0.058	336	0.180	
	multivariable			336	0.057	[-0.107;-0.003]
	Years					
		n (all)	p-value	n (complete)	p-value	
	univariable	412	0.012	336	0.042	
	multivariable			336	-	
	Gender					
	(1=M, 2=F)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	409	<0.001	336	<0.001	
	multivariable			336	<0.001	[-1.3;-0.41]

"convenience euthanasia"

Scenario F2	**Small animals %**					
rabbit breeder	(1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	
	univariable	467	0.307	342	0.773	
	multivariable			342	-	
	Employment					
	(1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	422	0.230	342	0.246	
	multivariable			342	-	
	Nb vets					
		n (all)	p-value	n (complete)	p-value	
	univariable	411	0.676	342	0.790	
	multivariable			342	-	
	Nb eutha					
		n (all)	p-value	n (complete)	p-value	
	univariable	393	0.333	342	0.316	
	multivariable			342	-	
	Request healthy eutha					
		n (all)	p-value	n (complete)	p-value	
	univariable	404	0.172	342	0.189	
	multivariable			342	-	
	Years					
		n (all)	p-value	n (complete)	p-value	effect size
	univariable	420	0.005	342	0.008	
	multivariable			342	0.008	[0.003;0.022]
	Gender					
	(1=M, 2=F)	n (all)	p-value	n (complete)	p-value	
	univariable	417	0.150	342	0.109	
	multivariable			342	-	

"convenience euthanasia"

Scenario F3 young dog costly therapy						
	Small animals % (1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	461	<0.001	338	<0.001	
	multivariable			338	<0.001	[-1.6;-0.39]
	Employment (1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	418	0.590	338	0.478	
	multivariable			338	-	
	Nb vets	n (all)	p-value	n (complete)	p-value	
	univariable	406	0.926	338	0.846	
	multivariable			338	-	
	Nb eutha	n (all)	p-value	n (complete)	p-value	
	univariable	389	0.672	338	0.676	
	multivariable			338	-	
	Request healthy eutha	n (all)	p-value	n (complete)	p-value	effect size
	univariable	399	0.094	338	0.242	
	multivariable			338	0.140	[-0.113;0.004]
	Years	n (all)	p-value	n (complete)	p-value	
	univariable	415	0.001	338	0.012	
	multivariable			338	-	
	Gender (1=M, 2=F)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	413	<0.001	338	<0.001	
	multivariable			338	<0.001	[-2.1;-0.95]

"convenience euthanasia"

Scenario F4 rabbit costly therapy						
	Small animals % (1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	462	<0.001	341	<0.001	
	multivariable			341	<0.001	[-1.1;-0.05]
	Employment (1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	417	0.011	341	0.030	
	multivariable			341	-	
	Nb vets	n (all)	p-value	n (complete)	p-value	
	univariable	407	0.190	341	0.182	
	multivariable			341	-	
	Nb eutha	n (all)	p-value	n (complete)	p-value	
	univariable	390	0.107	341	0.192	
	multivariable			341	-	
	Request healthy eutha	n (all)	p-value	n (complete)	p-value	
	univariable	401	0.551	341	0.804	
	multivariable			341	-	
	Years	n (all)	p-value	n (complete)	p-value	
	univariable	416	<0.001	341	<0.001	
	multivariable			341	-	
	Gender (1=M, 2=F)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	414	<0.001	341	<0.001	
	multivariable			341	<0.001	[-1.71;-0.73]

"convenience euthanasia"

Scenario F5 dog not fitting living conditions	**Small animals %** (1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	459	0.011	338	0.016	
	multivariable			338	0.011	[-1.29;-0.14]
	Employment (1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	418	0.186	338	0.121	
	multivariable			338	-	
	Nb vets	n (all)	p-value	n (complete)	p-value	
	univariable	406	0.303	338	0.291	
	multivariable			338	-	
	Nb eutha	n (all)	p-value	n (complete)	p-value	effect size
	univariable	389	0.001	338	0.002	
	multivariable			338	<0.001	[0.09;0.25]
	Request healthy eutha	n (all)	p-value	n (complete)	p-value	effect size
	univariable	400	0.532	338	0.785	
	multivariable			338	0.040	[-0.12;0]
	Years	n (all)	p-value	n (complete)	p-value	effect size
	univariable	414	<0.001	338	<0.001	
	multivariable			338	<0.001	[0.05;0.1]
	Gender (1=M, 2=F)	n (all)	p-value	n (complete)	p-value	
	univariable	411	<0.001	338	0.001	
	multivariable			338	-	

owner's refusal to euthanize

Scenario F6 Persian cat	**Small animals %** (1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	444	0.142	327	0.195	
	multivariable			327	0.193	[-1.12;0.09]
	Employment (1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	403	0.517	327	0.607	
	multivariable			327	-	
	Nb vets	n (all)	p-value	n (complete)	p-value	
	univariable	391	0.218	327	0.215	
	multivariable			327	-	
	Nb eutha	n (all)	p-value	n (complete)	p-value	
	univariable	375	0.371	327	0.278	
	multivariable			327	-	
	Request healthy eutha	n (all)	p-value	n (complete)	p-value	
	univariable	386	0.988	327	0.426	
	multivariable			327	-	
	Years	n (all)	p-value	n (complete)	p-value	effect size
	univariable	400	0.071	327	0.062	
	multivariable			327	0.032	[-0.064;-0.003]
	Gender (1=M, 2=F)	n (all)	p-value	n (complete)	p-value	
	univariable	398	0.790	327	0.461	
	multivariable			327	-	

"owner's refusal to euthanize"

Scenario F7 old sick dog without owner	Small animals % (1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	
	univariable	432	0.981	317	0.931	
	multivariable			317	-	
	Employment (1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	393	0.018	317	0.018	
	multivariable			317	-	
	Nb vets	n (all)	p-value	n (complete)	p-value	
	univariable	381	0.772	317	0.932	
	multivariable			317	-	
	Nb eutha	n (all)	p-value	n (complete)	p-value	effect size
	univariable	365	0.003	317	0.005	
	multivariable			317	0.005	[0.04;0.22]
	Request healthy eutha	n (all)	p-value	n (complete)	p-value	
	univariable	374	0.440	317	0.381	
	multivariable			317	-	
	Years	n (all)	p-value	n (complete)	p-value	effect size
	univariable	388	0.001	317	0.005	
	multivariable			317	0.004	[0.02;0.08]
	Gender (1=M, 2=F)	n (all)	p-value	n (complete)	p-value	
	univariable	385	0.410	317	0.479	
	multivariable			317	-	

"notification"

Scenario F8 guinea pig veterinary officer	Small animals % (1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	
	univariable	452	0.288	336	0.251	
	multivariable			336	-	
	Employment (1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	412	0.028	336	0.143	
	multivariable			336	-	
	Nb vets	n (all)	p-value	n (complete)	p-value	effect size
	univariable	399	0.036	336	0.053	
	multivariable			336	0.051	[-0.01;0.1]
	Nb eutha	n (all)	p-value	n (complete)	p-value	effect size
	univariable	384	0.285	336	0.121	
	multivariable			336	0.157	[-0.02;0.19]
	Request healthy eutha	n (all)	p-value	n (complete)	p-value	
	univariable	394	0.954	336	0.725	
	multivariable			336	-	
	Years	n (all)	p-value	n (complete)	p-value	
	univariable	406	0.036	336	0.039	
	multivariable			336	-	
	Gender (1=M, 2=F)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	403	0.022	336	0.019	
	multivariable			336	0.018	[0.15;1.58]

"responsability"

Scenario F9	**Small animals %**					
dog veterinarian decision	(1=<60%,2=60-100%)	n (all)	p-value	n (complete)	p-value	
	univariable	464	0.026	340	0.039	
	multivariable			340	-	
	Employment					
	(1=self,2=employed)	n (all)	p-value	n (complete)	p-value	
	univariable	420	0.001	340	0.009	
	multivariable			340	-	
	Nb vets					
		n (all)	p-value	n (complete)	p-value	effect size
	univariable	409	0.007	340	0.015	
	multivariable			340	0.013	[-0.08;0]
	Nb eutha					
		n (all)	p-value	n (complete)	p-value	
	univariable	391	0.635	340	0.373	
	multivariable			340	-	
	Request healthy eutha					
		n (all)	p-value	n (complete)	p-value	
	univariable	402	0.626	340	0.541	
	multivariable			340	-	
	Years					
		n (all)	p-value	n (complete)	p-value	effect size
	univariable	418	<0.001	340	<0.001	
	multivariable			340	<0.001	[0.003;0.075]
	Gender					
	(1=M, 2=F)	n (all)	p-value	n (complete)	p-value	effect size
	univariable	415	<0.001	340	<0.001	
	multivariable			340	0.020	[-1.46;-0.13]

Convenience euthanasia			
F1			
Aggressive dog			
	2.50%	50%	97.50%
small animal|gender	2.25	3.75	6.30
employment|years	0.06	0.12	0.21
nb vets|employment	5.86	7.26	9.02
nb vets|nb eutha	1.06	1.08	1.10
nb vets|years	0.71	0.79	0.89
nb vets|gender	0.48	0.59	0.72
nb vets|S5	1.22	1.35	1.50
nb vets|S8	1.12	1.21	1.30
nb eutha|small animal	1.25	1.43	1.64
Req healthy eutha|nb eutha	1.08	1.09	1.11
Req healthy eutha|S8	1.14	1.21	1.28
S2|S5	1.32	1.47	1.63
S5|S8	0.69	0.77	0.86
S5|S13	1.13	1.26	1.40
S11|small animal	1.27	1.59	1.99
S11|S9	1.17	1.30	1.45
S11|S17	1.12	1.24	1.38
S13|S11	1.15	1.28	1.43
S24|S2	1.36	1.51	1.67
S26|S5	1.29	1.43	1.60

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Convenience euthanasia			
F2			
Rabbit breeder			
	2.50%	50%	97.50%
small animal|gender	2.44	4.05	6.81
employment|years	0.06	0.12	0.21
nb vets|employment	6.12	7.55	9.35
nb vets|nb eutha	1.06	1.08	1.10
nb vets|years	0.71	0.80	0.90
nb vets|gender	0.47	0.58	0.71
nb vets|S5	1.22	1.35	1.50
nb vets|S8	1.11	1.20	1.29
nb eutha|small animal	1.41	1.63	1.88
nb eutha|gender	0.63	0.71	0.80
Req healthy eutha|nb eutha	1.08	1.10	1.11
Req healthy eutha|S8	1.12	1.19	1.27
S5|S2	1.33	1.48	1.64
S8|S26	0.70	0.78	0.87
S2|S24	1.35	1.50	1.66
S26|S5	1.29	1.43	1.59

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Convenience euthanasia			
F3			
Young dog costly therapy			
	2.50%	50%	97.50%
F3|gender	0.41	0.51	0.62
F3|S5	1.17	1.30	1.44
small animal|gender	2.29	3.81	6.42
employment|years	0.06	0.12	0.22
nb vets|employment	6.10	7.53	9.33
nb vets|nb eutha	1.06	1.08	1.10
nb vets|years	0.72	0.81	0.90
nb vets|gender	0.47	0.57	0.70
nb vets|S5	1.22	1.35	1.50
nb vets|S8	1.11	1.20	1.30
nb eutha|small animal	1.35	1.56	1.80
nb eutha|gender	0.64	0.72	0.81
Req healthy eutha|nb eutha	1.08	1.09	1.11
Req healthy eutha|S8	1.12	1.19	1.26
S5|S2	1.32	1.47	1.63
S6|F3	1.12	1.25	1.40
S8|F3	0.67	0.75	0.83
S11|small animal	1.30	1.63	2.03
S11|S9	1.18	1.31	1.45
S11|S17	1.12	1.24	1.38
S13|S11	1.15	1.29	1.44
S14|S11	1.10	1.23	1.37
S26|S5	1.29	1.43	1.60

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Convenience euthanasia			
F4			
Rabbit costly therapy			
	2.50%	50%	97.50%
F4|gender	0.45	0.56	0.70
small animal|gender	2.42	4.02	6.77
employment|years	0.06	0.12	0.21
nb vets|nb eutha	1.04	1.06	1.07
nb vets|S5	1.20	1.31	1.45
nb eutha|small animal	1.41	1.63	1.89
nb eutha|gender	0.63	0.71	0.80
Req healthy eutha|nb eutha	1.08	1.10	1.11
Req healthy eutha|S8	1.12	1.19	1.26
S8|S26	0.70	0.78	0.87
S11|small animal	1.25	1.57	1.98
S11|S9	1.19	1.32	1.47
S26|S5	1.29	1.43	1.59

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Convenience euthanasia			
F5			
Dog not fitting living conditions			
	2.50%	50%	97.50%
F5|years	1.22	1.36	1.51
small animal|gender	2.45	4.09	6.91
employment|years	0.06	0.11	0.20
nb vets|employment	5.82	7.21	8.95
nb vets|nb eutha	1.06	1.08	1.10
nb vets|years	0.69	0.78	0.87
nb vets|gender	0.46	0.57	0.69
nb vets|S5	1.24	1.38	1.54
nb vets|S8	1.10	1.19	1.29
nb eutha|F5	1.15	1.21	1.28
nb eutha|small animal	1.32	1.51	1.74
Req healthy eutha|nb eutha	1.08	1.09	1.11
Req healthy eutha|S8	1.13	1.20	1.27
S5|S2	1.33	1.47	1.64
S8|S26	0.69	0.77	0.86
S11|S9	1.20	1.34	1.49
S13|S11	1.16	1.29	1.44
S26|S5	1.29	1.43	1.59

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Owner's refusal to euthanize			
F6			
Persian cat			
	2.50%	50%	97.50%
small animal|gender	2.42	4.06	6.87
employment|years	0.07	0.13	0.22
nb vets|employment	5.76	7.08	8.71
nb vets|years	0.65	0.73	0.82
nb vets|gender	0.40	0.50	0.61
nb vets|S5	1.25	1.40	1.56
nb vets|S8	1.04	1.13	1.23
Req healthy eutha|nb eutha	1.08	1.09	1.10
Req healthy eutha|S8	1.10	1.17	1.25
S8|S26	0.68	0.76	0.85
S9|S11	1.17	1.30	1.45

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Owner's refusal to euthanize F7 Old sick dog without owner	2.50%	50%	97.50%
small animal|gender	2.22	3.72	6.31
employment|years	0.07	0.13	0.22
nb vets|F7	1.16	1.26	1.37
nb vets|employment	6.13	7.65	9.57
nb vets|nb eutha	1.06	1.08	1.10
nb vets|years	0.66	0.74	0.83
nb vets|gender	0.50	0.62	0.77
nb vets|S5	1.23	1.37	1.53
nb vets|S8	1.13	1.22	1.33
nb vets|S13	0.78	0.85	0.92
nb eutha|F7	1.12	1.19	1.27
Req healthy eutha|nb eutha	1.08	1.09	1.10
Req healthy eutha|S8	1.12	1.19	1.27
S8|S5	0.68	0.76	0.85
S11|small animal	1.28	1.61	2.04
S11|S9	1.21	1.34	1.50
S26|S5	1.28	1.42	1.59

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Notification			
F8			
Guinea pig veterinary officer			
	2.50%	50%	97.50%
small animal|gender	2.32	3.85	6.47
employment|years	0.06	0.12	0.21
nb vets|nb eutha	1.04	1.06	1.08
nb vets|S5	1.17	1.28	1.41
nb eutha|small animal	1.36	1.57	1.81
nb eutha|gender	0.63	0.71	0.81
Req healthy eutha|nb eutha	1.08	1.09	1.10
Req healthy eutha|S8	1.11	1.18	1.26
S5|S2	1.34	1.48	1.64
S8|S5	0.69	0.77	0.86
S11|small animal	1.29	1.62	2.03
S11|S9	1.19	1.33	1.48
S26|S5	1.29	1.43	1.60

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

Responsability			
F9			
Dog veterinarian decision			
	2.50%	50%	97.50%
small animal|gender	2.32	3.85	6.47
employment|years	0.06	0.12	0.21
nb vets|F9	0.75	0.81	0.87
nb vets|employment	7.38	8.85	10.62
nb vets|nb eutha	1.07	1.09	1.11
nb vets|gender	0.51	0.61	0.75
nb vets|S5	1.21	1.33	1.47
nb vets|S8	1.13	1.22	1.32
nb eutha|small animal	1.25	1.43	1.64
Req healthy eutha|nb eutha	1.08	1.09	1.11
Req healthy eutha|S8	1.13	1.20	1.28
S5|S2	1.30	1.44	1.60
S8|S26	0.70	0.78	0.87
S11|small animal	1.25	1.58	1.98
S11|S9	1.18	1.32	1.46
S13|S11	1.17	1.30	1.45
S17|S13	1.14	1.28	1.42
S2|S24	1.34	1.49	1.65
S26|S5	1.28	1.42	1.58

Small animal %	1= <60 %,	2= 60-100 %
Employment	1= self,	2= employed
Gender	1= male,	2= female

PAPER III

abn: an R package for modelling multivariate data using additive Bayesian networks

Marta Pittavino, Fraser Lewis & Reinhard Furrer

Manual for the R-package *abn* published on the Comprehensive R Archive Network (CRAN).

abn: an R package for modelling multivariate data using additive Bayesian networks

Marta Pittavino, Fraser Lewis, Reinhard Furrer

Abstract

This vignette describes the R package **abn** which provides functionality for identifying statistical dependencies in complex multivariate data using additive Bayesian network (ABN) models. This methodology is ideally suited for both univariate - one response variable, and multiple explanatory variables - and multivariate analysis, where in both cases all statistical dependencies between all variables in the data are sought. ABN models comprise of directed acyclic graphs (DAGs) where each node in the graph comprises a generalized linear model. Model search algorithms are used to identify those DAG structures most supported by the data. Currently implemented are models for data comprising of categorical, count and/or continuous variables. Further relevant information about **abn** can be found at: www.r-bayesian-networks.org.

Keywords: Graphical models, additive models, structure learning, exact order based search, parametric bootstrapping, JAGS, MCMC, parameter learning, heuristic search.

1. Introduction

Bayesian network (BN) modelling (Buntine 1991; Heckerman, Geiger, and Chickering 1995; Lauritzen 1996; Jensen 2001) is a form of graphical modeling which attempts to separate out indirect from direct association in complex multivariate data, a process typically referred to as structure discovery in Friedman and Koller (2003). In the last decades, BN modelling has been widely used in biomedical science/systems biology (Poon, Lewis, Pond, and Frost 2007a; Poon, Lewis, Frost, and Pond 2008; Poon, Lewis, Pond, and Frost 2007b; Needham, Bradford, Bulpitt, and Westhead 2007; Dojer, Gambin, Mizera, Wilczynski, and Tiuryn 2006; Jansen, Yu, Greenbaum, Kluger, Krogan, Chung, Emili, Snyder, Greenblatt, and Gerstein 2003; Djebbari and Quackenbush 2008; Hodges, Dai, Xiang, Woolf, Xi, and He 2010) to analyse multi-dimensional data. However, only in the last few years, ABN models have been applied to veterinary epidemiology field; thanks to their ability to generalize standard regression methodologies. Unlike other widely used multivariate approaches where dimensionality is reduced through exploiting linear combinations of random variables, such as in principal component analysis, graphical modeling does not involve any such dimension reduction.

BN is a generic and well established *data mining/machine learning* methodology, which has been demonstrated in other fields of study to be ideally suited to such analysis. They have been developed for analysing multinomial, multivariate Gaussian or conditionally Gaussian networks (a mix categorical and Gaussian variables), see Heckerman *et al.* (1995); Boettcher (2004); Geiger and Heckerman (1994). Additive Bayesian network (ABN) models are a special type of Bayesian network (BN) models, where each node in the graph comprises a generalized linear model. As the latter, they consist of statistical models which derive a directed acyclic graph (DAG) from empirical data, describing the

dependency structure between random variables. All types of BN models comprise of two reciprocally dependent parts: a qualitative (the structure) and a quantitative (the model parameters) part. The DAG is the graphical representation of the joint probability distribution of all random variables in the data. The model parameters are represented by a local probability distribution for all the variables in the network.

A number of libraries for fitting such BNs are available from CRAN. These types of BN have been constructed to ensure conjugacy, that is, enable posterior distributions for the model parameters and marginal likelihood to be calculated analytically. The purpose of **abn** is to provide a library of functions for more flexible BNs which can also not rely on conjugacy, which opens up an extremely rich modeling framework but at some considerable additional computational cost.

Currently **abn** includes functionality for fitting non-conjugate BN models which are multi-dimensional analogues of combinations of Binomial (logistic) and Gaussian regression. It includes also model with Poisson (log) distribution for count data and generalised linear models with random effects (with the previous distributions).

The objective in BN modeling structure discovery is to perform a model search on the data to identify an optimal model. Recall that BN models have a vast search space - super-exponential in the number of nodes - and it is generally impossible to determine a globally optimal model. How best to summarize a set of locally optimal networks with different structural features is an open question, and there are a number of widely used and intuitively reasonable possibilities. For example, one option is to conduct a series of heuristic searches and then simply select the best model found in Heckerman *et al.* (1995); alternatively, a single summary network can be constructed using results across many different searches (Hodges *et al.* 2010; Poon *et al.* 2007a). Otherwise, an exact search method, as the one presented in Koivisto and Sood (2004) which perform an exhaustive search, can be used. There are obvious pros and cons to either approaches and both are common in the literature and provide a good first exploration of the data.

For a general non-technical review of BN modeling applied in biology see Needham *et al.* (2007). While, a general introduction to BN modelling in veterinary epidemiology is provided by Lewis, Brulisauer, and Gunn (2011). Further applications of BN to veterinary studies were described by Ward and Lewis (2013); Wilson, Ribeiro, and Boinas (2013); Sanchez-Vazquez, Nielen, Edwards, Gunn, and Lewis (2012). Graphical modelling techniques used to analyse epidemiological data were used by Firestone, Lewis, Schemann, Ward, Toribio, and Dhand (2013); Firestone, Lewis, Schemann, Ward, Toribio, Taylor, and Dhand (2014); Lewis and McCormick (2012); Lewis (2012); Lewis and Ward (2013); Schemann, Lewis, Firestone, Ward, Toribio, Taylor, and Dhand (2013); Ludwig, Berthiaume, Boerlin, Gow, Léger, and Lewis (2013); McCormick, Sanchez-Vazquez, and Lewis (2013) resulting in dozens of publications, and references therein.

In this manual we first consider a series of examples illustrating how to fit different types of models and run different searches and summary analysis to a (synthetic) data set comprising of 250 observations from a joint distribution comprising of 17 categorical and 16 continuous variables which is included as part of the **abn** library. This data set is a single realization from a network of the same structure as that presented in Lewis *et al.* (2011), which is based on real data and sufficiently complex to provide a realistic example of data mining using Bayesian Network modeling. Then, a fully explained Case Study is provided, where all the important steps to conduct a data analysis using additive Bayesian networks are illustrated. Another detailed introduction and further relevant case studies about **abn** can be found at: www.r-bayesian-networks.org.

2. Pigs Case Study

We present data on disease occurrence in pigs provided by the industry body the 'British Pig Health Scheme' (BPHS). The main objective of BPHS is to improve the productivity of pig production in the UK, and reducing disease occurrence is a significant part of this process. The data we consider here comprise of a randomly chosen batch of 50 pigs from each of 500 randomly chosen pig producers in the UK. These are 'finishing pigs', animals about to enter the human food chain at an abattoir. Each animal is assessed for the presence of a range of different disease conditions by a specialist swine veterinarian. We consider here the following nine disease conditions: enzootic-pneumonia (EPcat); pleurisy (plbinary); milk spots (MS); hepatic scarring (HS); pericarditis (PC); peritonitis (PT); lung abscess (Abscess); tail damage (TAIL); and papular dermatitis (PDcat). The presence of any of these conditions results in an economic loss to the producer. Either directly due to the relevant infected part of the animal being removed from the food chain, or indirectly in cases such as enzooticpneumonia, which may potentially indicate poor herd health and efficiency losses on the farm. An additional loss, though not directly monetary, is the presence of tail damage which may be suggestive of welfare concerns, which may also be linked to sub-optimal production efficiency. Milk spots and hepatic scarring result from infestation with Ascaris suum which is particularly important as this is a zoonotic helminth parasite.

2.1. Deciding on a search method

As a very rough rule of thumb if there are less than 20 variables (and no random effects) then probably the most robust model search option is an exact search (as opposed to a heuristic) which will identify a globally best DAG. Followed then by parametric bootstrapping in order to assess whether the identified model contains excess structure (this is an adjustment for over-modelling). Although, the parametric bootstrapping might require access to a cluster computer to make this computationally feasible. This is arguably one of the most comprehensive and reliable statistical modelling approaches for identifying an empirically justified best guess (DAG) at 'nature's true model', the unknown mechanisms and processes which generated the study data.

Order Based Searches

It is generally not feasible to iterate over all possible DAG structures when dealing with more than a handful of variables, hence the reliance on heuristic searches. It is also extremely difficult to construct efficient Monte Carlo Markov chain samplers across BN structures. A solution to this was proposed in Friedman and Koller (2003) where rather than sample across DAGs, it was proposed to sample across node orderings. A node order is simply the set of integers 1 through n, where n is the number of variables in the data. A DAG is consistent with an ordering if for each node in the order its parents come before it. For example a DAG with only an arc from $1 \rightarrow 2$ is consistent with ordering $1, 2, 3, 4$ as the parent 1 comes before 2, but a DAG with an arc from $2 \rightarrow 1$ is not consistent with this ordering. In effect, each order is a collection of DAGs, and note that each DAG may be consistent with multiple orders, i.e. the empty DAG is consistent with every possible ordering. This introduces bias, in that averaging across orders need not give the same results as averaging across DAGs, if the latter were possible. This is relevant when estimating posterior probabilities of individual structural features, and is baised towards more parsimonious features as they are consistent with more orders. Note that this bias does not apply to maximising across orders, as in finding most probable structures (see later). The big advantage of searching across orders is that there are $n!$ different orders compared to a reasonably tight upper bound of $2^{\binom{n}{2}}$ for different DAGs.

There are (at least) two approaches for searching across orders. The first is to construct a Markov chain which samples from the posterior distribution of all orders, and is the approach presented in Friedman and Koller (2003). Alternatively, in Koivisto and Sood (2004) an exact method is proposed which rather than sample across orders, performs an exhaustive search. This has the advantage that it can also be used to find the globally optimal DAG of the data - the most probable structure - as well as posterior probabilities for structural features, such as individual arcs. The drawback is that this exact approach is only feasible with smaller number of variables e.g. up to 12 or 13 when dealing with additive models. For the code provided in **abn** this exact approach is readily feasible up to 20 variables using typical desktop computing, and potentially up to 25 variable with access to a shared memory cluster computer.

Most Probable Structure

Using the exact order based method due to Koivisto and Sood (2004) it is also possible to identify the DAG with globally best network score. Identification of a most probable structure is split into two parts. Firstly we calculate a cache of individual node scores, for example using `buildscorecache`. Next, an exhaustive order based structural search is carried out using the function `mostprobable` which relies on the information in the node cache.

As in the heuristic searches it is possible to ban or retain certain arcs, for example when splitting multinomial variables. This is done in the node cache computation step. There are two different structural priors implemented in this search, the first in the uniform prior where all structures (all parent combinations) are equally likely. This is the default `prior.choice=1` in `mostprobable` and the other functions. Also implemented is the prior used in Koivisto and Sood (2004) where all parent combinations of equal cardinality are equally weighted, this is `prior.choice=2`. The latter does give the same prior weight to a parent combination with no parents and a parent combination comprising off all possible parents (since there is only one choice of each, $n-1$ choose 0 and n-1 choose $n-1$). This may not be desirable but is included as a prior for completness. Note that the order based search is exact in the sense that it will identify a DAG who score is equal to the best possible score if it was possible to exhaustive search across all DAGs. For example, if using `prior.choice=1` then the network found should have a score greater than or equal to that found using the previously described heuristic searches. The structure found need not be unique in that others may exist with the same globally optimal score, the order based search is only guaranteed to find one such structure.

To calculate the most probable structure we again use `buildscorecache()` to calculate a cache of individual node scores. Next, the function `mostprobable()` does the actual exhaustive order based search, and works for both conjugate and additive models since as with calculating the posterior probabilities this step only involves structural searching and is not concerned with the precise parameterisation of each BN model.

2.2. Preparing the data

There are two main things which need to be checked in the data before it can be used with any of the abn model fitting functions.

- All records with missing variables must be either removed or else the values completed. There are a range of libraries available from CRAN for completing missing values using approaches such as multiple imputation. Ideally, marginalising over the missing values is preferable (as

opposed completing them as this then results in essentially dealing with models of models), but this is far from trivial here and not yet (and may never be) implemented in abn. To remove all records with one or more missing values then code similar to the following probably suffices:

```
> library( abn)     # Load the library
> mydat <- pigs.vienna[,-11]  # Get data, drop batch variable
> mydat[ complete.cases(mydat),]
```

N.b. this is not actually needed with pigs.vienna.

- All variables which are to be treated as binary must be coerced to factors. To coerce an existing variable into a factor then:

  ```
  > mydat[,1] <- as.factor(mydat[,1])
  ```

  coerces the first variable in data.frame pigs.vienna. The levels (labels of the factor) can be anything provided there are only two and a “success" here is take to be the second level. For example, the second value in the vector returned by:

  ```
  > levels( mydat[,1])
  ```

To include additional variables in the modeling, for example interaction terms or polynomials, then these must be created manually and included into the data.frame just like any other variable.

2.3. Initial searches for a optimal model

Below is some R code which will perform an exact search. We want to find the DAG with the best goodness of fit (network score - log marginal likelihood) and ideally we would search for this without any apriori complexity limit (max number of parents). However, this may be both not computationally feasible and also highly inefficient. For example, with 25000 observation is it really realistic to consider models with up to 9 covariates per node.

One approach is to start off with an apriori limit of one parent per node, find the best DAG, and then repeat an identical search process (again using functions buildscorecache and mostprobably) with a parent limit of 2. And so on, stopping when the maximal DAG found no longer changes when the complexity limit is relaxed (increased). The parent limits `max.par` are not inserted, in the code below, the `max.par` command can vary from one to an higher parent limits. These initial searches can be easily automatized, in the library subdirectory `system.file('bootstrapping_example', package='abn')` is possible to find a script file `initsearch.bash` that performs the initial search explained before in an automated way, increasing progressively the number of parents, from 1 to 5 (for this specific example). The run time for the mostprobable function increases very considerably with the number of parents.

```
> ban <- matrix( rep(0,dim(mydat)[2]^2),ncol=dim(mydat)[2])
```

The ban and retain matrix must have the names set:

```
> colnames( ban) <- rownames(ban) <- names(mydat)
> retain <- matrix( rep(0,dim(mydat)[2]^2),ncol=dim(mydat)[2])
> colnames( retain) <- rownames(retain) <- names(mydat)
```

Setup the distribution list for each node:

```
> mydists <- list( PC="binomial", PT="binomial", MS="binomial",
+                  HS="binomial", TAIL="binomial",
+                  Abscess="binomial", Pyaemia="binomial",
+                  EPcat="binomial", PDcat="binomial",
+                  plbinary="binomial")
```

Build a cache of all the local computations:

```
> mycache <- buildscorecache( data.df=mydat,
+            data.dists=mydists, dag.banned=ban,
+            dag.retained=retain, max.parents=max.par)
```

Run the actual exact search:

```
> mp.dag <- mostprobable( score.cache=mycache)
> fabn <- fitabn( dag.m=mp.dag,data.df=mydat,
+                 data.dists=mydists)
> datadir <- tempdir()
```

Save the results obtained:

```
> save( mycache, mp.dag, fabn, file=
+       paste(datadir,"mp",max.par,".RData",sep=""))
```

2.4. Results from the initial search

Searches across parent limits 1 through 5 were run and we now examine the results. What we are looking for is simply the model with the best score (the largest, the least negative, mlik value), checking that this does not improve when more parents are permitted. This then says we have found a DAG with maximal goodness of fit. What we find (below) is that the goodness of fit does not improve when we increase the parent limit beyond 3.

The code necessary to analyze the results from the script file and to visualize the resulting DAG, in a linux environment, is the following:

```
> load( "RData/Rout1.RData")
> ml1 <- fabn$mlik;
> tographviz( dag.m=mp.dag,data.df=mydat,data.dists=mydists,
+             outfile="DAG_cycle.dot")
> system( "dot -Tpdf -o DAG_cycle.pdf DAG_cycle.dot")
> system( "evince DAG_cycle.pdf&")
```

List of resulting marginal likelihood from the initial search:

```
> mp.mlik <- c( -44711.62,-44685.53,-44684.64,-44684.64,-44684.64)
```

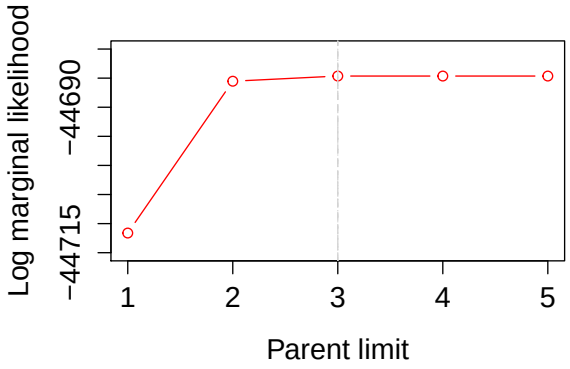


Figure 1: Comparison of goodness of fits for different parent limits

The actual DAG corresponding to the maximum marginal likelihood: `mp.mlik`$= -44684.64$ is in Fig. 2.

2.5. Adjustment for overfitting: parametric bootstrapping using MCMC

We have identified a DAG which has the best (maximum) possible goodness of fit according to the log marginal likelihood. This is the standard goodness of fit metric in Bayesian modelling (Mackay 1992), and includes an implicit penalty for model complexity. While it is sometimes not always apparent from the technical literature, the log marginal likelihood can easily (and sometimes vastly) overfit with smaller data sets. Of course the difficulty is identifying what constitutes 'small' here. In other words using the log marginal likelihood alone (or indeed any of the other usual metrics such as AIC or BIC) is likely to identify structural features, which, if the experiment/study was repeated many times, would likely only be recovered in a tiny faction of instances. Therefore, these features could not be considered robust. Overfitting is an ever present issue in model selection procedures, particular is common approaches such as stepwise regression searches, see Babyak (2004).

A well established approach for addressing overfitting is to use parametric bootstrapping (Friedman, Goldszmidt, and Wyner 1999). The basic idea is very simple. We take our chosen model and then simulate data sets from this, the same size as the original observed data, and see how often the different structural features are recovered. For example, is it reasonable for our data set of 434 observations to support a complexity of 29 arcs? Parametric bootstrapping is arguably one of the most defensible solutions for addressing overfitting, although it is likely the most computationally demanding, as for each simulated (bootstrap) data set we need to repeat the same exact model search as used with the original data. And we may need to repeat this analysis hundreds (or more) times to get robust results.

Performing parametric bootstrapping is easy enough to code up if done in small manageable chunks. Here we provide a step-by-step guide along with necessary sample code.

2.6. Software needed

We have a DAG model and MCMC software such as JAGS and WinBUGS are designed for simulating from exactly such models. So all we need to do is implement our DAG model, e.g. in Fig. 2, in the appropriate JAGS (or WinBUGS) syntax (which are very similar). Here I am going to use JAGS,

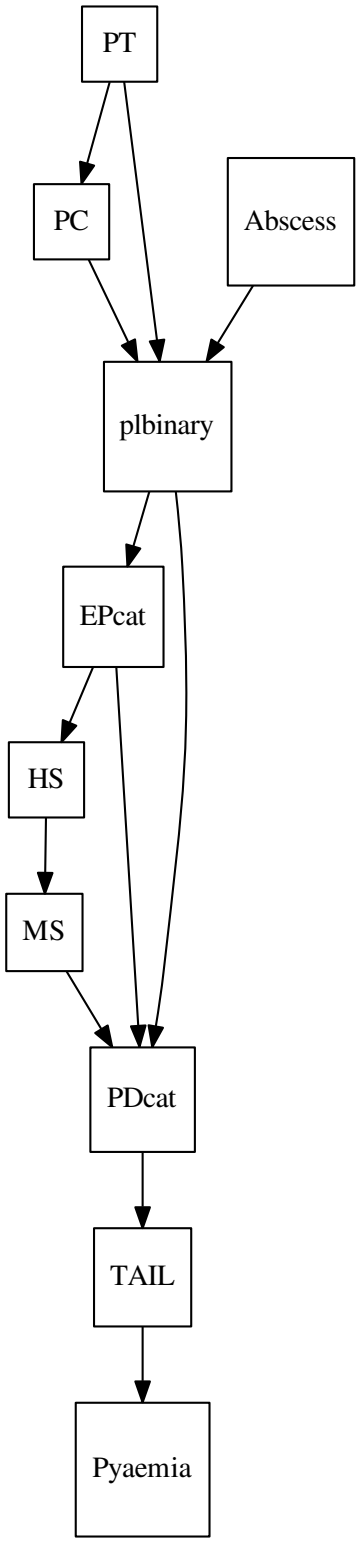


Figure 2: Globally optimal DAG, found with a maximum number of 3 or more parents.

developed by Plummer (2003), in preference to WinBUGS or OpenBUGS for no other reason than that is what I am most familiar with, and JAGS is relatively straightforward to install (compile from source) on a cluster. Binary versions of JAGS are also available for most common platforms (Linux, Windows, OS X). To implement our DAG in JAGS we need write a model definition file (a BUG file) which contains the structure of the dependencies in the model. We also need to provide in here the probability distributions for each and every parameter in the model. Note that in Bayesian modelling the parameter estimates will not generally conform to any standard probability distribution (e.g. Gaussian) unless we are in the very special case of having conjugate priors. The marginal parameter distributions required can be estimated using the `fitabn` function and then fed into the model definition. We next demonstrate one way of doing this which is to use empirical distributions, in effect we provide JAGS with a discrete distribution over a fine grid which approximates whatever shape of density we need to sample from.

2.7. Generating marginal densities

The function `fitabn` has functionality to estimate the marginal posterior density for each parameter in the model. The parameters can be estimated one at a time by manually giving a grid (e.g. the x values where we want to evaluate f(x)) or else all together. In the latter case a very simple algorithm will try and work out where to estimate the density. This can work better sometimes and others, although it seems to work fine here for most variables. In order to use these distributions with JAGS we must evaluate the density over an equally spaced grid as otherwise the approach used in JAGS will not sample correctly. The basic command needed here is:

```
> marg.f <- fitabn( dag.m=mydag,data.df=mydat,data.dists=mydists,
+                   compute.fixed=TRUE,n.grid=1000)
```

We should not simply assume that the marginals have been estimated accurately, and they should each be checked using some common sense. Generally speaking, estimating the goodness of fit (mlik) for a DAG comprising of GLM nodes is very reliable. This marginalises out all parameters in the model. Estimating marginal posterior densities for individual parameters, however, can run into trouble as this presupposes that the data contains sufficient information to accurately estimate the "shape" (density) for every individual parameter in the model. This is a stronger requirement than simply being able to estimate an overall goodness of fit metric. If a relatively large number of arcs have been chosen for a node with relatively few observations (i.e. “success” in a binary node) then this may not be possible, or at least the results are likely to be suspect. Exactly such issues - overfitting - are why we are performing the parametric bootstrapping in the first place but they can also pose some difficulties before getting to this stage.

It is essential to first visually check the marginal densities estimated from `fitabn`. Something like the following will create one pdf file where each page is a separate plot of a marginal posterior density.

```
> library( Cairo) # Available from CRAN
> CairoPDF( "MargPlots_PigsData.pdf")
> for( i in 1:length(marg.f$marginals)){
+   cat( "processing marginals for node:",
+   nom1 <- names( marg.f$marginals)[i],"\n")
+   cur.node <- marg.f$marginals[i]
+   # Get the marginal for current node, this is a matrix [x,f(x)]
```

```
+   cur.node <- cur.node[[1]]
+   # This is always [[1]] for models without random effects
+   for( j in 1:length(cur.node)){
+   cat( "processing parameter:",nom2<- names(cur.node)[j],"\n")
+   cur.param <- cur.node[[j]]
+   plot( cur.param,type="l",main=paste(nom1,":",nom2))}}
> dev.off()
```

These plots (available in `system.file('bootstrapping_example',package='abn')`) suggests that the all the marginal estimates looks fine. Moreover, we can now perform an additional common sense check on their reliability. A probability density must integrate to unity (the area under the curve is equal to one). The densities here are estimated numerically and so we would not expect to get exactly one (n.b. no internal standarization is done so we can check this), but if the numerical estimation has worked reliably then we would expect this to be close (e.g. $0.99, 1.01$) to on.

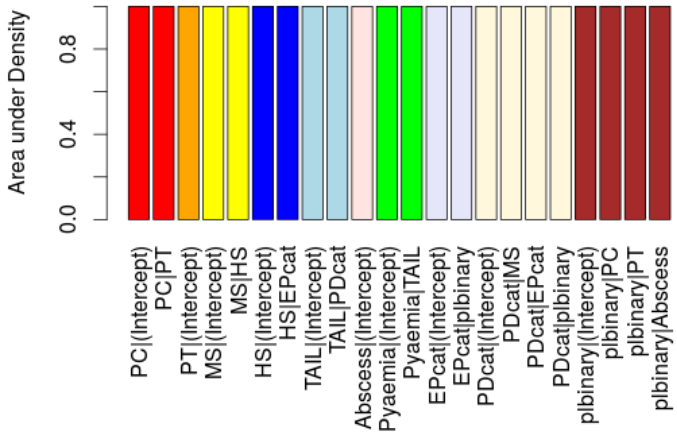


Figure 3: Approximated area under marginal densities

From Fig. 3 it is obvious that everything went fine in the marginal density estimation. To check the area we used the following code, which has as final output a file called `pigs_post_params.R` which contains all the information JAGS needs to sample from the marginal posterior distributions.

```
> marnew <- marg.f$marginals[[1]]
> for( i in 2: length(marg.f$marginals)){
+   marnew <- c(marnew, marg.f$marginals[[i]])}
```

A straightforward question is: how reliable are the marginals? If the numerical routines work well then the area under the density function should be close to unity.

```
> myarea <- rep( NA,length( marnew))
> names( myarea) <- names( marnew)
> for( i in 1:length(marnew)){
+   tmp <- spline(marnew[[i]])
+   # Spline just helps make the estimation smoother
+   myarea[i] <- sum(diff(tmp$x)*tmp$y[-1])}
```

```
>   # Just width x height of rectangles
> cbind( myarea)
```

Now visualise `myarea` as a plot, the colours need to be adjusted:

```
> library( Cairo)
> mycols <- rep("green",length(marnew))
> mycols[1:2] <- "red"; # PC
> mycols[3] <- "orange"; # PT
> mycols[4:5] <- "yellow"; # MS
> mycols[6:7] <- "blue"; # HS
> mycols[8:9] <- "lightblue"; # TAIL
> mycols[10] <- "mistyrose"; # Abscess
> mycols[11:12] <- "green"; # Pyaemia, lightcyan
> mycols[13:14] <- "lavender"; # EPcat
> mycols[15:18] <- "cornsilk"; # PDcat
> mycols[19:22] <- "brown" ; # plbinary
> CairoPNG( "Area_PigsData.png",pointsize=10,width=720,height=640)
> par( las=2, mar=c(8.1,4.1,4.1,2.1))
> barplot( myarea,ylab="Area under Density",ylim=c(0,2), col=mycols)
> dev.off()
```

Now we create the data in the right format ready for going to JAGS:

```
> print( names(marnew))
> # want to bind all the marginals the same nodes into a matrix
> m <- marnew;
> # PT -> PC, 1 parent, 2 params;
> PC.p <- cbind( m[["PC|(Intercept)"]],  m[["PC|PT"]])
> # PC --> NO PARENTS, 1 params;
> PT.p <- cbind( m[["PT|(Intercept)"]])
> # HS -> MS, 1 parent, 2 params;
> MS.p <- cbind( m[["MS|(Intercept)"]], m[["MS|HS"]])
> # EPcat -> HS, 1 parent, 2 params;
> HS.p <- cbind( m[["HS|(Intercept)"]],  m[["HS|EPcat"]])
> # PDcat -> TAIL, 1 parent, 2 params;
> TAIL.p <- cbind( m[["TAIL|(Intercept)"]],  m[["TAIL|PDcat"]])
> # Abscess --> NO PARENTS, 1 param;
> Abscess.p <- cbind( m[["Abscess|(Intercept)"]])
> # TAIL -> Pyaemia, 1 parent, 2 params;
> Pyaemia.p <- cbind(m[["Pyaemia|(Intercept)"]],m[["Pyaemia|TAIL"]])
> # plbinary -> EPcat, 1 parent, 2 params;
> EPcat.p <- cbind( m[["EPcat|(Intercept)"]], m[["EPcat|plbinary"]])
> # MS, EPcat, plbinary --> PDcat, 3 parents, 4 params;
> PDcat.p <- cbind( m[["PDcat|(Intercept)"]], m[["PDcat|MS"]],
+                   m[["PDcat|EPcat"]], m[["PDcat|plbinary"]])
> # PC, PT, Abscess --> plbinary, 3 parents, 4 params;
```

```
> plbinary.p <- cbind(m[["plbinary|(Intercept)"]],m[["plbinary|PC"]],
+                     m[["plbinary|PT"]], m[["plbinary|Abscess"]])
> dump( c( "PC.p","PT.p","MS.p","HS.p","TAIL.p","Abscess.p",
+          "Pyaemia.p","EPcat.p","PDcat.p","plbinary.p"),
+           file=paste('Jags/',"pigs_post_params.R",sep=""))
```

2.8. Building BUG model

Once we have the posterior distributions the next step is to actually create the DAG in JAGS. This involves creating a BUG file, a file which contains a definition of our DAG (from Fig. 2) in terms which can be understood by JAGS. This can easily be done by hand, if rather tedious, and should be checked carefully for errors (which JAGS will prompt about in any case). The BUG file is here. This file is fairly heavily commented (it might look complicated but most of it is just copy and paste with minor edit) - the syntax is similar to R - and should be fairly self explanatory. Note that unlike a more usual use of WinBUGS or JAGS we have no data here, we are simply providing JAGS with a set of marginal probability distributions and how they are inter-dependent, and we want it to then generate realisations from the appropriate joint probability distribution. The next step is to tell JAGS to perform this simulation, i.e. generate a single bootstrap data set of size $n = 25000$ observations based on the assumption that the DAG is Fig. 2 is our true model. The BUG file can be found in `system.file('bootstrapping_example',package='abn')` in the file `pigs_model.bug`.

2.9. A single bootstrap analysis

To run JAGS we use four separate files: i) the BUG model definition file (model.bug); ii) a file `pigs_post_params.R` which contains the marginal distributions referred to in the BUG file; iii) a script which runs the MCMC simulation `jags_pigs_script.R`; and finally iv) a file which sets of random number seed (`inits.R` - the value in this must be changed to use different streams of random numbers). These four files are contained in this tarball. To run this example extract all the files into one directory and then at the command line type `'jags jags_pigs_script.R'`. In terms of the MCMC component, a burn-in of 100000 is used but as there are no data here this takes no time to run and is likely of little use and could be shorted (it is just included to allow JAGS any internal adaptations or diagnostics that it might need to do). The actual MCMC is then run for 250000 iterations with a thin of 10, which gives 25000 observations for each variable - the same size as the original data. The number of MCMC steps and thin is a little arbitrary and this could be run for longer with a bigger thin, but for this data looking at autocorrelation plots for the Gaussian variables there appears no evidence of correlation at a thin of 10 and so this seems sufficient here.

The next step is to automate a single run, e.g. generate a single bootstrap sample and then perform an exact search on this, just as was done for the original data. This file performs a single such analysis in R - just drop this into the same directory as the four JAGS file above. For ease we also call the JAGS script from inside R which might require some messing about with the PATH variable on Windows (e.g. if you open up a cmd prompt then you should be able to type "jags" without needed to give a full path). The output from this is given below and "results" is a single matrix (DAG) which is saved in an R workspace called `boot1run.RData`.

Get the pigs data, drop batch variable

```
> orig.data <- pigs.vienna[,-11]
```

Now create a single bootstrap sample:

```
> system("jags jags_pigs_script.R")
```

Read in boot data and convert to data.frame in same format as the original data, e.g. coerce to factors, using the appropriate R package **coda**, described in Plummer, Best, Cowles, and Vines (2006):

```
> library( coda)
> boot.data <- read.coda("out1chain1.txt","out1index.txt")
> boot.data <- as.data.frame(boot.data)
> for( j in 1:dim(orig.data)[2]){if(is.factor(orig.data[,j]))
+         { boot.data[,j]<- as.factor(boot.data[,j])
+           levels(boot.data[,j])<- levels(orig.data[,j])}}
```

Now we have the boot.data in identical format as the original:

```
> ban <- matrix( rep(0,dim(orig.data)[2]^2),
+                ncol=dim(orig.data)[2])
> colnames( ban) <- rownames(ban) <- names(orig.data)
> retain <- matrix( rep(0,dim(orig.data)[2]^2),
+                   ncol=dim(orig.data)[2])
> colnames( retain) <- rownames(retain) <- names(orig.data)
> mydists <- list( PC="binomial",  PT="binomial", MS="binomial",
+                  HS="binomial",  TAIL="binomial",
+                  Abscess="binomial", Pyaemia="binomial",
+                  EPcat="binomial", PDcat="binomial",
+                  plbinary="binomial")
```

Set the parent limits to 3, equal to the original data:

```
> max.par <- 3;
```

Build a cache on bootstrap data:

```
> boot1.cache <- buildscorecache( data.df=boot.data,
+                 data.dists=mydists, max.parents=max.par,
+                 dag.banned=ban, dag.retained=retain)
```

Run `mostprobable` search on the bootstrap data:

```
> boot1.mp <- mostprobable( score.cache=boot1.cache)
> datadir <- tempdir()
> save( boot1.mp,file=paste( datadir,"boot1run.RData",sep=''))
```

Once this works on your local machine it is then a case of trying to automate this in the most efficient way, for example for use on a cluster. The crucial thing here is that the random number seed used each time (in the file `inits.R`) must be changed for each bootstrap simulation otherwise an identical bootstrap data set will be produced!

2.10. Ways to summarise results from parametric bootstrapping

The globally optimal DAG, Fig. 2, has 12 arcs. It way be that some of these are due to over-modelling which means they will be recovered in relatively few bootstrap analyses. 10240 bootstrap analyses were conducted (on a cluster) and all the R files, MPI wrapper, and also the actual results (in R workspaces) can be found in the folder `system.file('bootstrapping_example', package='abn')`.

The first step is to explore the bootstrap results. Of some interest is how many of the arcs were recovered during each of the bootstrap "simulations". This is given in Fig. 4. We can see right away that not all the arcs in the original DAG (Fig. 2 - with 12 arcs) were recovered - even just once. This provides overwhelming evidence that the original exact search has indeed overfitted to the original data. Not at all surprising given the relatively small sample size. We must therefore trim off some of the complexity - arcs - from the original DAG in order to justify that our chosen DAG is a robust estimate of the underlying system which generated the observed data.

A parametric bootstrapping approach was suggested in Friedman *et al.* (1999) which uses simulation to assess whether a chosen model comprises more complexity than could reasonably be justified given the size of the observed data set. Using Markov chain Monte Carlo simulation via JAGS (open source software), 10240 independent (assumed by inspecting autocorrelations from the MCMC output) data sets of the same size as the original data were generated from our chosen model in i). For each of these bootstrap data sets an identical exact order-based search as in i) was conducted.

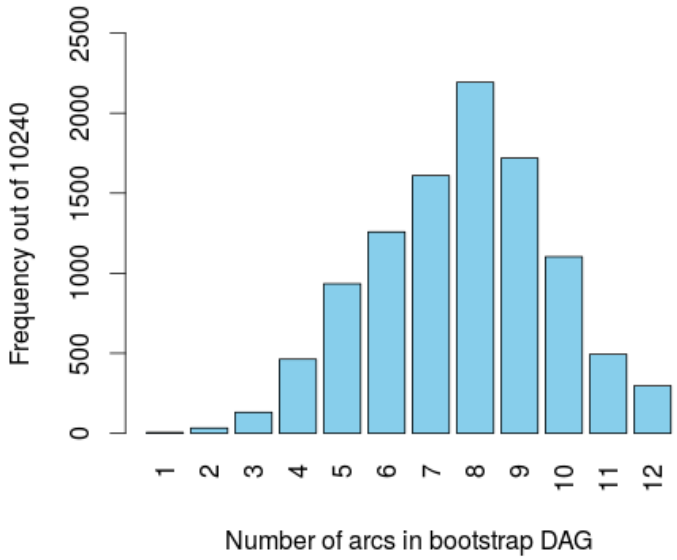


Figure 4: Number of arcs recovered in bootstrapping

There are a number of different options in terms of trimming/pruning arcs. One common option which apes the use of majority consensus trees in phylogenetics trees are just special cases of DAGs, is to remove all arcs which were not recovered in at least a majority (50%) of the bootstrap results. Fig. 5 shows the frequency at which each arc was recovered, the maximum value possible being 10240 $(100\%$ support). Collating results across these 10240 searches we find that only 14% of the globally optimal DAGs found comprised 12 or more arcs. Approximately 68% of DAGs had 11 or more arcs - therefore a robust model of the original data has no more than 11 arcs. Almost identical results were obtained using a random selection of 5012 searches suggesting that sufficient bootstrap samples had

been performed. The usual cut-off for structural support of features (arcs) is 50% in BN modeling (Poon *et al.* 2007a,b, 2008; Lycett, Ward, Lewis, Poon, Pond, and Brown 2009; Lewis *et al.* 2011), and is analogous to the widespread use of majority consensus trees in phylogenetics. We therefore conclude that our chosen model in i) with 11 arcs is robust. This is perhaps not surprising given we have a large data set of $25K$ observations.

	PC	PT	MS	HS	TAIL	Abscess	Pyaemia	EPcat	PDcat	plbinary
PC	0	6607	0	0	0	0	0	0	0	0
PT	0	0	0	0	0	0	0	0	0	0
MS	0	0	0	5243	0	0	0	0	0	0
HS	0	0	0	0	0	0	0	4402	0	0
TAIL	0	0	0	0	0	0	0	0	6190	0
Abscess	0	0	0	0	0	0	0	0	0	0
Pyaemia	0	0	0	0	3257	0	0	0	0	0
EPcat	0	0	0	0	0	0	0	0	0	9686
PDcat	0	0	5096	0	0	0	0	3703	0	9009
plbinary	9249	6606	0	0	0	9799	0	0	0	0

Figure 5: Frequencies at which each arc in the original DAG was recovered during bootstrapping

Note that this 50% is not in any way comparable with the usual 95% points used in parameter estimation as these are entirely different concepts (an explanation for this can be found here).

Another option, other than choosing a different level of support (which is entirely up to the researcher), is to consider an undirected network. That is, include all arcs if there support - considering both directions - exceeds 50%. This is justifiable due to likelihood equivalence which means that - generally speaking - the data cannot discriminate between different arc directions (see here for an explanation) and therefore considering arcs recovered in only one direction may be overly conservative. Again, this decision is likely problem specific. For example, from a purely statistical perspective being conservative is generally a good thing, but from the scientists point of view this may then remove most of the more interesting results from the study. Obviously a balance is required.

In this data it turns out that removing all arcs with have less than 50% support gives an identical pruned network as if we were to consider both arc directions jointly. In generally this need not be the case. Fig. 6 shows out optimal DAG after removing these arcs. This is our optimal model of the data.

All the code for analysing the results from the bootstrap analyses can be found here:

```
> mydata <- pigs.vienna[,-11]
> N <- 10240;
> # Write out manually, clearer than using rep()
> mydag <- matrix(c(
+ # PC  PT  MS  HS TAIL Abscess Pyaemia EPcat PDcat plbinary
+   0,  1,  0,  0,  0,  0,       0,      0,     0,      0,
+   0,  0,  0,  0,  0,  0,       0,      0,     0,      0,
+   0,  0,  0,  1,  0,  0,       0,      0,     0,      0,
+   0,  0,  0,  0,  0,  0,       0,      1,     0,      0,
+   0,  0,  0,  0,  0,  0,       0,      0,     1,      0,
+   0,  0,  0,  0,  0,  0,       0,      0,     0,      0,
+   0,  0,  0,  0,  1,  0,       0,      0,     0,      0,
+   0,  0,  0,  0,  0,  0,       0,      0,     0,      1,
```

```
+   0,  0,  1,  0,  0,  0,       0,     1,     0,     1,
+   1,  1,  0,  0,  0,  1,       0,     0,     0,     0),
+   byrow=TRUE,ncol=10)
> colnames(mydag) <- rownames(mydag) <- names(mydata)
> sum( mydag) # 12 arcs, as in original model, Figure 2
> mydists <- list( PC="binomial", PT="binomial", MS="binomial",
+                  HS="binomial", TAIL="binomial",
+                  Abscess="binomial", Pyaemia="binomial",
+                  EPcat="binomial", PDcat="binomial",
+                  plbinary="binomial")
> # Use fitabn to check mydag is correct (no typos mlik = -44684.64)
> print( fitabn(dag.m=mydag,data.df=mydata,data.dists=mydists)$mlik)
> bestdag <- mydag
```

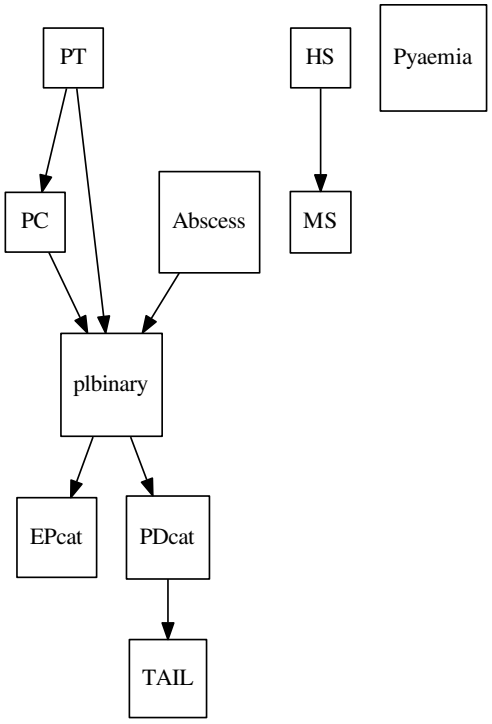


Figure 6: Optimal DAG after removal of arcs supported at less than 50% in bootstrapping. Contains 8 Arcs.

The next command reads all files with `mp[number].RData` and create a list of results:

```
> boot.dags <- list()
> these <- grep("mp10Kboot\\d+.RData", dir())
> num <- 1
> for( i in dir()[these]){# Load each file
+   load(i) # Provides dags, a list
+   tmp <- dags[which(unlist(lapply(dags,sum))>0)]
```

```
+   # Get valid entries in dags but as a list
+   for( j in 1:length(tmp)){
+   # For each entry copy into boot.dags, and increment counter
+     boot.dags[[num]]<- tmp[[j]]; num <- num+1 }
+   rm( dags)
+ }
```

It is always good practice, have a look at mlik values for the bootstraps viz a viz the original:

```
> if( FALSE){
+   scores <- rep(0,length(boot.dags))
+   for(i in 1:length(boot.dags)){
+     scores[i] <- fitabn(dag.m=boot.dags[[i]],data.df=mydata,
+
+   }
+   scores.b <- scores[-which(scores< -N)]
+   orig.score <- fitabn(dag.m=bestdag,data.df=mydata,
+
+   plot(density(scores.b,from=min(scores.b),to=max(scores.b)))
+   abline(v=orig.score,lwd=2,col="blue")
+ }
```

We now trim all arcs from the boot results which do not occur in the Master DAG - bestdag - since we know these are due to overfitting:

```
> boot.dags.trim <- boot.dags
> for( i in 1:length(boot.dags)){
+    boot.dags.trim[[i]] <- boot.dags.trim[[i]]*bestdag  }
> arc.freq <- lapply(boot.dags.trim,sum)
> arc.freq <- table(unlist(arc.freq))
> library( Cairo)
> CairoPNG("PigsFreqBootRes.png",pointsize=10,width=720,height=700)
> par(las=1, mar=c(6.1,6.1,4.1,2.1))
> barplot( arc.freq,ylab="",xlab="",col="skyblue",
+          names.arg=names(arc.freq), ylim=c(0,2500))
> par( las=1)
> mtext( "Number of arcs in bootstrap DAG",1,line=3,cex=1.5)
> par( las=3)
> mtext( "Frequency out of 10240",2,line=4,cex=1.5)
> dev.off()
> total.dag <- matrix(rep(0,dim(bestdag)[2]^2),ncol=dim(bestdag)[2])
> colnames(total.dag) <- rownames(total.dag)<- colnames(bestdag)
> # Get the support for each arc, total.dag:
> for( i in 1:length(boot.dags)){
+   if(sum(boot.dags[[i]])>0){total.dag <- total.dag+boot.dags[[i]]}}
> total.dag <- total.dag*bestdag # We only want arcs in the best DAG
> total.dag
```

The frequencies at which each arc in the original DAG was recovered during bootstrapping.
We get the majority consensus (directed DAG):

```
> f <- function(val,limit){ if(val<limit){return(0)}
+                           else {return(1)}}
> bestdag.trim <- apply( total.dag,c(1,2),FUN=f,limit=N/2)
```

We get the majority consensus (undirected DAG), but with arcs in the most supported direction:

```
> bestdag.trim.nodir <- bestdag
> bestdag.trim.nodir[,] <- 0 # Set zero
> child <- NULL; parent <- NULL
> for( i in 1:dim(total.dag)[1]){
+   for( j in 1:dim(total.dag)[2]){
+      if(i>j){ # Get most supported direction
+               if(total.dag[i,j]>total.dag[j,i]){
+                 m.i <- i; m.j <- j;}
+               else {m.i <- j; m.j <- i;}
+               # Does arc quality - exceed threshold of support
+               if(total.dag[i,j]+total.dag[j,i]>N/2){
+               # We want this as more than 5000
+               bestdag.trim.nodir[m.i,m.j] <- 1}}}}
> tographviz( dag.m=bestdag.trim,data.df=mydata,
+             data.dists=mydists, outfile="postbootpigs.dot")
> system( "dot -Tpdf -o postbootpigs.pdf postbootpigs.dot")
> system( "evince postbootpigs.pdf&")
> save( bestdag.trim,file=paste("bestdagpigs_trim.RData",sep=''))
```

2.11. Estimating marginals from the final DAG

Once we have identified our optimal DAG then it is usual to want to examine the parameters in this model, e.g. in Fig. 7. These are our results, the effects of the various variables in our study. This process is very similar to when estimating the marginals for the bootstrapping but should now be easier since we should have removed any difficulties due to over-fitting. The posterior density plots for the final DAG can be found in the pdf called `Pigs_PostBootPlots.pdf`. These all look fine. An outlook on the parameters can be done based on the variable plbinary.

The R code for creating the marginals and quantiles can be found below:

```
> mydata <- pigs.vienna[,-11]
> mydists <- list( PC="binomial", PT="binomial", MS="binomial",
+                  HS="binomial", TAIL="binomial",
+                  Abscess="binomial", Pyaemia="binomial",
+                  EPcat="binomial", PDcat="binomial",
+                  plbinary="binomial")
> marg.f <- fitabn(dag.m=bestdag.trim,data.df=mydata,
+                  data.dists=mydists,compute.fixed=TRUE,
+                  n.grid=1000)
```

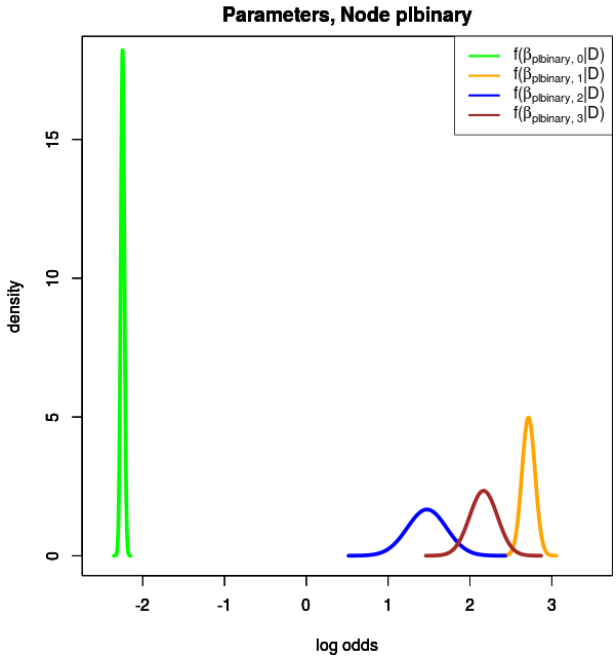


Figure 7: Marginal posterior density based in the node plbinary

```
> library( Cairo)
> CairoPDF("Pigs_PostBootPlots.pdf")
> for( i in 1:length(marg.f$marginals)){
+   cat( "processing marginals for node:",
+   nom1 <- names(marg.f$marginals)[i],"\n")
+   cur.node <- marg.f$marginals[i]
+   # Get marginal for current node, this is a matrix [x,f(x)]
+   cur.node <- cur.node[[1]]
+   # This is always [[1]] for models without random effects
+    for( j in 1:length(cur.node)){
+    cat("processing parameter:",
+    nom2 <- names(cur.node)[j],"\n")
+    cur.param <- cur.node[[j]]
+    plot( cur.param,type="l",main=paste(nom1,":",nom2))}}
> dev.off()
> marnew <- marg.f$marginals[[1]]
> for(i in 2: length(marg.f$marginals)){
+   marnew <- c(marnew, marg.f$marginals[[i]])}
> margs <- marnew;
> mymat <- matrix(rep(NA,length(margs)*3),ncol=3)
> rownames(mymat) <- names(margs)
> colnames(mymat) <- c("2.5%","50%","97.5%")
> ignore.me <- union(grep("\\(Int",names(margs)),
+                    grep("prec",names(margs)))
```

These are not effect parameters, but background constants and precisions:

		Odds Ratio	
	2.5%	**50%**	**97.5%**
PC|(Intercept)	0.03	0.03	0.03
PC|PT	17.71	26.44	39.21
PT|(Intercept)	0.00	0.00	0.00
MS|(Intercept)	0.06	0.06	0.07
MS|HS	0.20	0.30	0.44
HS|(Intercept)	0.06	0.06	0.06
TAIL|(Intercept)	0.00	0.00	0.00
TAIL|PDcat	2.61	4.32	6.78
Abscess|(Intercept)	0.00	0.01	0.01
Pyaemia|(Intercept)	0.00	0.00	0.00
EPcat|(Intercept)	0.36	0.37	0.38
EPcat|plbinary	1.97	2.14	2.31
PDCat|(Intercept)	0.04	0.04	0.04
PDcat|plbinary	1.78	2.08	2.41
plbinary|(Intercept)	0.10	0.11	0.11
PC|(Intercept)	12.92	15.12	17.71
PC|(Intercept)	2.72	4.36	6.97
PC|(Intercept)	6.23	8.71	12.17

Table 1: Marginal posterior quantiles for each parameter. The red lines indicate point estimates bigger than 1, corresponding to risk factor. While the blue line refers to estimates smaller than 1, indicating protective factor.

```
> comment <- rep("",length(margs))
> for(i in 1:length(margs)){
+   tmp <- margs[[i]]
+   tmp2 <- cumsum(tmp[,2])/sum(tmp[,2])
+   mymat[i,] <- c(tmp[which(tmp2>0.025)[1]-1,1],
+                  tmp[which(tmp2>0.5)[1],1],
+                  tmp[which(tmp2>0.975)[1],1])
+   myvec <- mymat[i,]
+   if( !(i%in%ignore.me) &&  (myvec[1]<0 && myvec[3]>0)){
+       comment[i] <- "not sig. at 5%"}
+   # Truncate for printing
+   mymat[i,] <- as.numeric(formatC(mymat[i,],digits=3,format="f"))}
> cbind( mymat)
```

The code above produce the table of percentiles that can be found in Table 1.

All of the effect parameters, that is, ignoring the intercept terms (which is just a background constant) and the precision parameters - have 95% confidence (or credible) intervals which do not cross the origin. Note this is not guaranteed to happen this criteria was not part of the model selection process, and all the parameters in the model have been justified using mlik and bootstrapping. But having such marginal intervals can make presenting the results possibly easier to a skeptical audience.

2.12. Pigs Case Study Conclusion

Based on the results presented in the previous section, it possible to draw a conclusion of the study. From Table 1, we can notice that, a part from HS, which has a protective effect on MS, all others parameters have a risk effect on possible disease outcome. Our final optimal ABN model of this disease system suggests that the presence of Pyaemia is independent from other conditions. Moreover, the remaining diseases split into two separate connected components.

Some interesting biological questions, resulting from ABN model, can be related to the similarity of causal pathways between these two groups of diseases and to the common causes shared within each of these groups.

The principal reasons because ABN models should be used are linked to their ability of generalization of the usual GLM to multiple dependent variables: fully multi-dimensional models. Specifically, results from ABN analyses can be used as a basis for developing new biological questions about factors potentially affecting disease's presence, and inform the design of future targeted studies.

3. An additional example

In the next subsections we illustrate how to fit an ABN model to different kinds of data, starting with the introdcution of the data we are going to deal with. The main purpose of BN structure discovery is to estimate the joint dependency structure of the random variables in the available data, and this is achieved by heuristically searching for optimal models and comparing their goodness of fit using Bayes factors. It is assumed that all structures are equally supported in the absence of any data - an uniformative prior on structures - and so comparing Bayes factors collapses to comparing the marginal likelihoods which is done on a log scale. The log marginal likelihood for a BN is typically referred to as the network score.

3.1. Simulated Data Example var33

Figure 1 shows the structure of the distribution which generated the data set `var33` included with **abn**. This diagram was created using the `tographviz()` function of **abn** (see later examples) which translates the matrix which defines a network - a directed acyclic graph - into a text file of suitable format for processing in Graphviz, where this processing was done outside of R. Graphviz is freely available and operates on most platforms and can be downloaded from www.graphviz.org, there is also an R package which interfaces to Graphviz available from the Bioconductor project (requires an installation of Graphviz).

3.2. Fitting an additive BN model to categorical data

An additive BN model for categorical data can be constructed by considering each individual variable as a logistic regression of the other variables in the data, and hence the network model comprises of many combinations of local logistic regressions (Rijmen 2008). The parameters in this model are the additive terms in a usual logistic regression and independent Gaussian priors are assumed for each covariate. Note that the variables here must all be binary, and so all multinomial variables need to be split into separate binary factors (and added to the original data.frame) in order to form the network model. This is analogous to forming the design matrix in a conventional additive model analysis. Similarly, interaction terms can be added by including appropriate additional columns in the data.frame. In these models the log marginal likelihood (network score) is estimated using Laplace

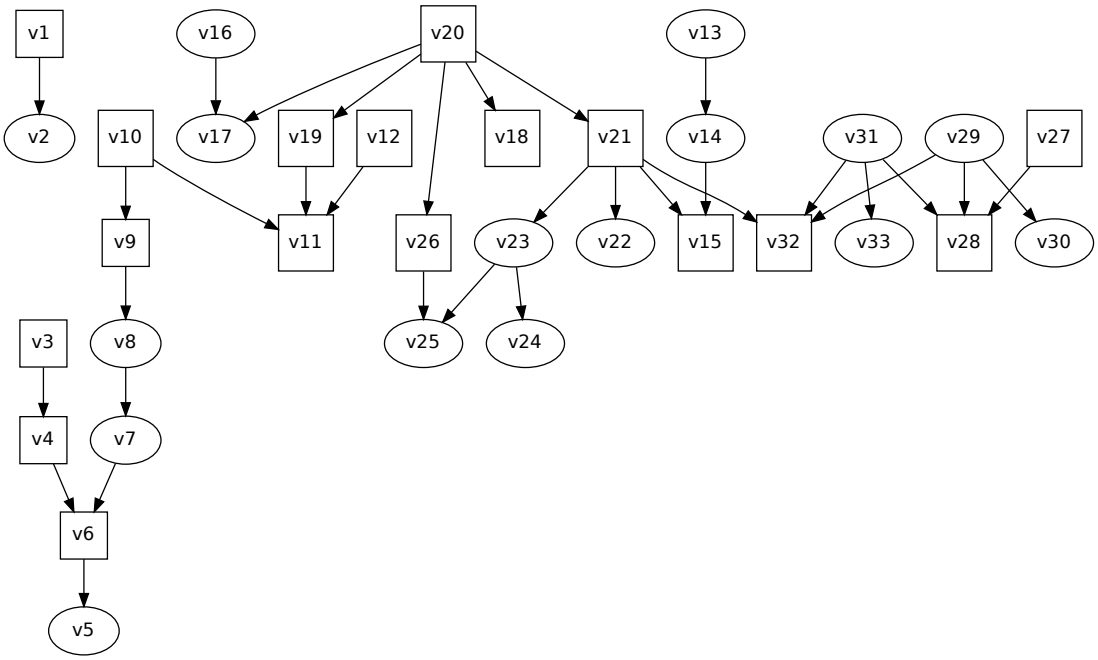


Figure 8: Directed acyclic graph representation of the joint probability distribution which generated data set `var33` which is included with **abn**. The square nodes are categorical (binary) and the oval nodes continuous variables.

approximations at each node. Hyperparameters for the means and variances in the Gaussian priors are fixed at zero and 1000 respectively, and other values can be given explicitly in the call to `fitabn` but this is not recommended without good reason.

To fit an additive model use `fitabn(data.df,dag.m,data.dists, ...)`. In the following code we fit first the independence model with no arcs and then the same dependence model as above. Turning on `verbose=TRUE` simply gives the individual log marginal likelihoods for each node (n.b. the numbering is that used internally and simply denotes the variables in the data.frame from left to right).

The following code fits a network to the subset of the variables from `var33` which are categorical. In this data these are all binary. Note that all categorical variables should be set as factors.

```
> library( abn)
> bin.nodes <- c( 1,3,4,6,9,10,11,12,15,18,19,20,21,26,27,28,32)
> var33.cat <- var33[,bin.nodes] #Categorical nodes only
> dag33 <- matrix( 0, 17, 17)
> colnames( dag33) <- rownames( dag33) <- names( var33.cat)#Set names
```

Move back to independence model:

```
> dag33["v11","v12"] <- 0; dag33["v11","v10"]<- 0; dag33["v4","v3"]<- 0;
```

Setup the distribution list for each categorical node:

```
> mydists.cat <- list( v1 ="binomial", v3 = "binomial",
+         v4 = "binomial",  v6 = "binomial",  v9 = "binomial",
+        v10 = "binomial", v11 = "binomial", v12 = "binomial",
+        v15 = "binomial", v18 = "binomial", v19 = "binomial",
+        v20 = "binomial", v21 = "binomial", v26 = "binomial",
+        v27 = "binomial", v28 = "binomial", v32 = "binomial")
> ind.mod.cat <- fitabn( data.df=var33.cat, dag.m=dag33,
+                       data.dists=mydists.cat, verbose=FALSE)
```

It is possible to change to `verbose=TRUE` if one want to check how change the score for each individual node.

The network score for a model with conditional independencies is:

```
> ind.mod.cat$mlik
```

```
[1] -2856.948
```

The structure of the network definition matrix is where each row is a "child" and each column is its "parents", where a `1` denotes a parent (or arc) is present. Now lets fit a model with some conditional dependencies, for example where `v11` is conditionally dependent upon `v12` and `v10`, and `v4` is conditionally dependent upon `v3`.

Now fit the model with some conditional dependencies:

```
> dag33["v11","v12"] <- 1;
> dag33["v11","v10"] <- 1;
> dag33["v4","v3"] <- 1;
> dep.mod.cat <- fitabn( data.df=var33.cat, dag.m=dag33,
+                        data.dists=mydists.cat, verbose=FALSE)
```

The network score for a model with conditional dependencies is:

```
> dep.mod.cat$mlik
```

```
[1] -2850.081
```

The network score is considerably improved and therefore suggests support for these new structural features. To produce a visual description of the model then we can export to graphviz as follows:

```
> tographviz( dag=dag33, data.df=var33.cat, data.dists=mydists.cat,
+             outfile="mydagcat.dot", directed=TRUE) #Create file
```

`mydagcat.dot` can then be processed with graphviz unix shell typing: "dot -Tpdf mydagcat.dot -o mydagcat.pdf" or using gedit if on Windows.

In `tographviz()` the `data.df` argument is used to determine whether the variable is a factor or not, where factors are displayed as squares and non-factors as ovals. To use the full range of visual

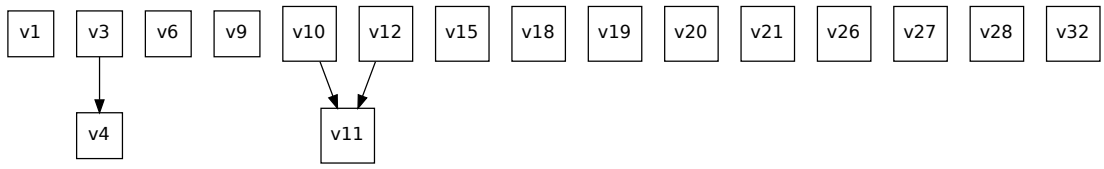


Figure 9: Directed acyclic graph `dag33` created using `tographviz()` and Graphviz

Graphviz options simply use the file created by `tographviz()` as a starting point and manually edit this in a text editor before running through `dot` or one of the other Graphviz layout processors.

3.3. Fitting an additive BN model to continuous data

We now consider analogous models to those in Section 3.2 but where the network comprises of Gaussian linear regressions rather than logistic regressions. The structure of these models again assumes independent Gaussian priors for each of the coefficients in the additive components for the mean response at each node (with hyper means = 0 and hyper variances = 1000). The Gaussian response distribution is parameterized in terms of precision $(1/\sigma^2)$, and independent Gamma priors are used with shape=0.001 and scale=1/0.001 (where these are as defined in the `rgamma` help page). By default, each variable in the data.frame is standardised to a mean of zero and standard deviation of one, this has no effect on the identification of dependencies between variables. We start droping the categorical nodes:

```
> var33.cts <- var33[,-bin.nodes]
> dag33 <- matrix( 0, 16, 16)
> colnames( dag33) <- rownames( dag33) <- names( var33.cts)
```

Setup the distribution list for each continuous node:

```
> mydists.cts <- list(       v2 = "gaussian",  v5 = "gaussian",
+           v7 = "gaussian",  v8 = "gaussian", v13 = "gaussian",
+         v14 = "gaussian", v16 = "gaussian", v17 = "gaussian",
+         v22 = "gaussian", v23 = "gaussian", v24 = "gaussian",
+         v25 = "gaussian", v29 = "gaussian", v30 = "gaussian",
+         v31 = "gaussian", v33 = "gaussian")
```

Now fit the model defined in dag33 - full independence

```
> ind.mod.cts <- fitabn( data.df=var33.cts, dag.m=dag33,
+                 data.dists=mydists.cts, verbose=FALSE)
```

We use as default priors, a Gaussian density with 0 mean and variance 1000. While for the precision, inverse of variance, we use a Gamma density of hyperparameters 0.001 and $1/0.001$. This is the network score (goodness of fit, log marginal likelihood):

```
> ind.mod.cts$mlik
```

```
[1] -5949.52
```

Now fit a model with conditional dependencies, for example let v33 depend on v31, and v24 depend on v23, and v14 depend on v13.

```
> dag33["v33","v31"] <- 1;
> dag33["v24","v23"] <- 1;
> dag33["v14","v13"] <- 1;
> dep.mod.cts <- fitabn( data.df=var33.cts, dag.m=dag33,
+                 data.dists=mydists.cts, verbose=FALSE)
```

The network score for a model with conditional independence is:

```
> dep.mod.cts$mlik
```

```
[1] -5704.547
```

```
> tographviz( dag=dag33, data.df=var33.cts, data.dists=mydists.cts,
+             outfile="mydagcts.dot", directed=TRUE) #Create file
```

mydagcat.dot can then be processed with graphviz unix shell typing: "dot -Tpdf mydagcat.dot -o mydagcat.pdf" or using gedit if on Windows.

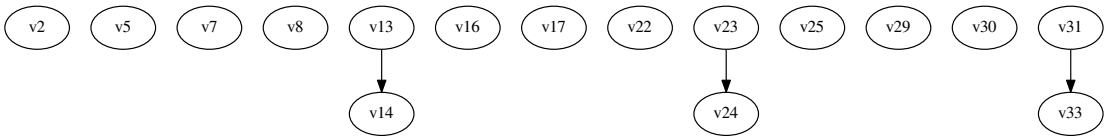


Figure 10: Directed acyclic graph dag33 for continuous variables only created using tographviz() and Graphviz

3.4. Fitting an additive BN model to mixed data

To conclude the fitting of a single pre-specified model to data, e.g. based on expert opinion, we consider an additive BN model which comprises both binary and Gaussian nodes and this comprises of a combination of Binomial (logistic) and Gaussian linear models. Again fitabn() is used and the code is almost identical to the previous examples.

```
> dag33 <- matrix( 0, 33, 33)
> colnames( dag33) <- rownames( dag33)<- names( var33)
```

Setup distribution list for each mixed node:

```
> mydists.mix <- list(      v1 = "binomial",  v2 = "gaussian",
+          v3 = "binomial",  v4 = "binomial",  v5 = "gaussian",
+          v6 = "binomial",  v7 = "gaussian",  v8 = "gaussian",
```

```
+          v9 = "binomial", v10 = "binomial", v11 = "binomial",
+         v12 = "binomial", v13 = "gaussian", v14 = "gaussian",
+         v15 = "binomial", v16 = "gaussian", v17 = "gaussian",
+         v18 = "binomial", v19 = "binomial", v20 = "binomial",
+         v21 = "binomial", v22 = "gaussian", v23 = "gaussian",
+         v24 = "gaussian", v25 = "gaussian", v26 = "binomial",
+         v27 = "binomial", v28 = "binomial", v29 = "gaussian",
+         v30 = "gaussian", v31 = "gaussian", v32 = "binomial",
+         v33 = "gaussian")
```

Now fit the model defined in dag33, full independence:

```
> ind.mod <- fitabn( data.df=var33, dag.m=dag33,
+            data.dists=mydists.mix, verbose=FALSE)
```

The network score with no conditional dependencies is:

```
> ind.mod$mlik
```

```
[1] -8806.468
```

We now fit a BN model which has the same structure as the joint distribution used to generate the data and later create a visual graph of this model. We then define a model with many independencies:

```
> dag33[2,1] <- 1;
> dag33[4,3] <- 1;
> dag33[6,4] <- 1; dag33[6,7] <- 1;
> dag33[5,6] <- 1;
> dag33[7,8] <- 1;
> dag33[8,9] <- 1;
> dag33[9,10] <- 1;
> dag33[11,10] <- 1; dag33[11,12] <- 1; dag33[11,19] <- 1;
> dag33[14,13] <- 1;
> dag33[17,16] <- 1;dag33[17,20] <- 1;
> dag33[15,14] <- 1; dag33[15,21] <- 1;
> dag33[18,20] <- 1;
> dag33[19,20] <- 1;
> dag33[21,20] <- 1;
> dag33[22,21] <- 1;
> dag33[23,21] <- 1;
> dag33[24,23] <- 1;
> dag33[25,23] <- 1; dag33[25,26] <- 1;
> dag33[26,20] <- 1;
> dag33[33,31] <- 1;
> dag33[33,31] <- 1;
> dag33[32,21] <- 1; dag33[32,31] <- 1; dag33[32,29] <- 1;
> dag33[30,29] <- 1;
```

```
> dag33[28,27] <- 1; dag33[28,29] <- 1; dag33[28,31] <- 1;
> dep.mod <- fitabn( data.df=var33, dag.m=dag33,
+            data.dists=mydists.mix, verbose=FALSE)
```

The network score for a model with conditional independence is:

```
> dep.mod$mlik
```

```
[1] -8019.887
```

```
> tographviz( dag=dag33, data.df=var33, data.dists=mydists.mix,
+             outfile="mydag_all.dot", directed=TRUE)#Create file
```

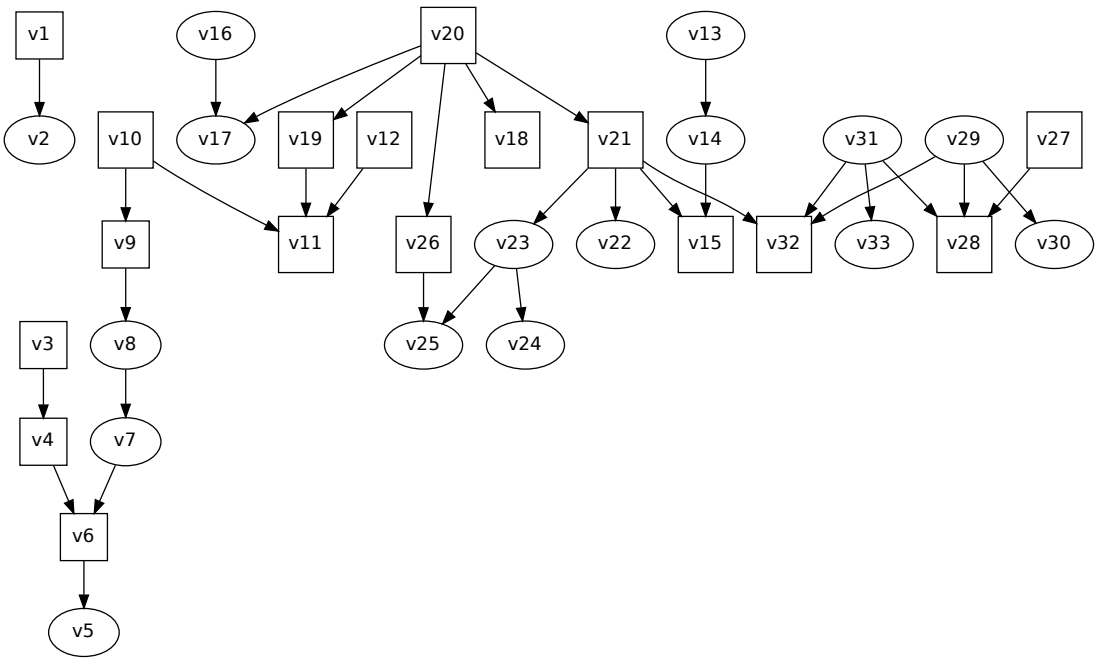


Figure 11: Directed acyclic graph `dag33` for mixed continuous and discrete variables

3.5. Model fitting validation

In order to validate the additive models for mixed binary and Gaussian models, estimates of the posterior distributions for the model parameters using Laplace approximations were compared with those estimated using Markov chain Monte Carlo. These were always in very close agreement for the range of models and data examined. This is an indirect validation of the Laplace estimate of the network score, e.g. if the posterior densities match closely then this implies that the denominator (the marginal likelihood - network score) must also be accurately estimated, as a "gold standard" estimate of the network score is generally unavailable for such non-conjugate models.

4. Further insights into searching strategy

The key objective of the **abn** library is to enable estimation of statistical dependencies in data comprising of multiple variables - that is, find a DAG which is robust and representative of the dependency structure of the (unknown) stochastic system which generated the observed data. The challenge here is that with such a vast model space it is impossible to enumerate over all possible DAGs, and there may be very many different DAGs with similar goodness of fit. In the next sections we first consider searching for additive (non-conjugate) models.

4.1. Single search for optimal additive BN model from categorical data

To run a single search heuristic use `search.hillclimber()`. This commences from a randomly created DAG which is constructed by randomly adding arcs to an empty network until all possible arcs have been tried. The function `search.hillclimber()` then searches stepwise from the initial random network for an improved structure, where three stepwise operations are possible: i) add an arc; ii) remove and arc; or iii) reverse and arc. The stepwise search is subject to a number of conditions, firstly only moves that do not generate a cycle are permitted, secondly, a parent limit is imposed which fixes the maximum number of parents which each child node can have (arcs go from parent to child), and thirdly it is possible to ban or retain arcs. If provided, `banned.m` is a matrix which defines arcs that are not allowed to be considered in the search process (or in the creation of the initial random network). Similarly, `retain.m` includes arcs which must always be included in any model. It is also possible to specific an explicit starting matrix, `start.m` and if using a retain matrix then `start.m` should contain at least all those arcs present in `retain.m`. Note that only very rudimentary checking is done to make sure that the ban, retain and start networks - if user supplied - are not contradictory.

To improve the computational performance of `search.hillclimber()` a cache of all possible goodness of fit must be built in advance, using the function `buildscorecache()`. Rather than re-calculate the score for each individual node in the network (the overall network score is the product of all the scores for the individual nodes) the score for each unique node found during the search is stored in the cache created by the function `buildscorecache()`.

```
> bin.nodes <- c( 1,3,4,6,9,10,11,12,15,18,19,20,21,26,27,28,32)
> var33.cat <- var33[,bin.nodes] #Categorical nodes only
> dag33 <- matrix( 0, 17, 17)
> colnames(dag33) <- rownames(dag33) <- names(var33.cat)#Set names
```

Create banned and retain empty DAGs:

```
> banned.cat <- matrix( 0, 17, 17)
> colnames(banned.cat) <- rownames(banned.cat) <- names(var33.cat)
> retain.cat <- matrix( 0, 17, 17)
> colnames(retain.cat) <- rownames(retain.cat) <- names(var33.cat)
```

Setup distribution list for each categorical node:

```
> mydists.cat <- list(     v1 = "binomial",  v3 = "binomial",
+         v4 = "binomial",  v6 = "binomial",  v9 = "binomial",
+        v10 = "binomial", v11 = "binomial", v12 = "binomial",
```

```
+        v15 = "binomial", v18 = "binomial", v19 = "binomial",
+        v20 = "binomial", v21 = "binomial", v26 = "binomial",
+        v27 = "binomial", v28 = "binomial", v32 = "binomial")
```

Build cache of all the local computations this information is needed later when running a model search:

```
> mycache.cat <- buildscorecache( data.df=var33.cat,
+                 data.dists=mydists.cat, dag.banned=banned.cat,
+                 dag.retained=retain.cat, max.parents=1)
```

Running a single search heuristic for an additive BN uses `search.hillclimber()`. It uses a parameter prior specifications (as detailed above). Several additional arguments are available which relate to the numerical routines used in the Laplace approximation to calculate the network score. The defaults appear to work reasonably well in practice and if it is not possible to calculate a robust value for this approximation in any model, for example due to a singular design matrix at one or more nodes, then the model is simply assigned a log network score of $-\infty$ which effectively removes it from the model search.

Run a single search heuristic for an additive BN:

```
> heur.res.cat <- search.hillclimber( score.cache=mycache.cat,
+                   num.searches=1, seed=0, verbose=FALSE,
+                   trace=FALSE, timing.on=FALSE)
```

Setting `trace=TRUE`, the majority consensus network is plotted as the searches progress.

4.2. Single search for optimal additive BN model for continuous data

As above but for a network of Gaussian nodes.

```
> var33.cts <- var33[,-bin.nodes] #Drop categorical nodes
> dag33 <- matrix( 0, 16, 16)
> colnames(dag33) <- rownames(dag33) <- names(var33.cts) #Set names
> banned.cts <- matrix( 0, 16, 16)
> colnames(banned.cts) <- rownames(banned.cts) <- names(var33.cts)
> retain.cts <- matrix( 0, 16, 16)
> colnames(retain.cts) <- rownames(retain.cts) <- names(var33.cts)
```

Setup distribution list for each continuous node:

```
> mydists.cts <- list(       v2 = "gaussian",  v5 = "gaussian",
+          v7 = "gaussian",  v8 = "gaussian", v13 = "gaussian",
+         v14 = "gaussian", v16 = "gaussian", v17 = "gaussian",
+         v22 = "gaussian", v23 = "gaussian", v24 = "gaussian",
+         v25 = "gaussian", v29 = "gaussian", v30 = "gaussian",
+         v31 = "gaussian", v33 = "gaussian")
```

Build cache of all local computations, information needed later when running a model search:

```
> mycache.cts<- buildscorecache( data.df=var33.cts,
+              data.dists=mydists.cts, dag.banned=banned.cts,
+              dag.retained=retain.cts, max.parents=1)
```

Run a single search heuristic for an additive BN:

```
> heur.res.cts<- search.hillclimber( score.cache=mycache.cts,
+               num.searches=1, seed=0, verbose=FALSE,
+               trace=FALSE, timing.on=FALSE)
```

Setting trace=TRUE, the majority consensus network is plotted as the searches progress.

4.3. Single search for optimal additive BN model for mixed data

Model searching for mixed data is again very similar to the previous examples. Note that in this example the parameter priors are specified explicitly (although those given are the same as the defaults). The `+1` in the hyperparameter specification is because a constant term is included in the additive formulation for each node.

```
> dag33 <- matrix( 0, 33, 33)
> colnames(dag33) <- rownames(dag33) <- names(var33)#Set names
```

Create empty DAGs:

```
> banned.mix <- matrix( 0, 33, 33)
> colnames(banned.mix) <- rownames(banned.mix) <- names(var33)
> retain.mix<- matrix( 0, 33, 33)
> colnames(retain.mix) <- rownames(retain.mix) <- names(var33)
```

Setup distribution list for mixed node:

```
> mydists.mix <- list(       v1 = "binomial",  v2 = "gaussian",
+          v3 = "binomial",  v4 = "binomial",  v5 = "gaussian",
+          v6 = "binomial",  v7 = "gaussian",  v8 = "gaussian",
+          v9 = "binomial", v10 = "binomial", v11 = "binomial",
+         v12 = "binomial", v13 = "gaussian", v14 = "gaussian",
+         v15 = "binomial", v16 = "gaussian", v17 = "gaussian",
+         v18 = "binomial", v19 = "binomial", v20 = "binomial",
+         v21 = "binomial", v22 = "gaussian", v23 = "gaussian",
+         v24 = "gaussian", v25 = "gaussian", v26 = "binomial",
+         v27 = "binomial", v28 = "binomial", v29 = "gaussian",
+         v30 = "gaussian", v31 = "gaussian", v32 = "binomial",
+         v33 = "gaussian")
```

Build cache of all local computations, information needed later when running a model search:

```
> mycache.mix <- buildscorecache( data.df=var33,
+              data.dists=mydists.mix, dag.banned=banned.mix,
+              dag.retained=retain.mix, max.parents=1)
```

Run a single search heuristic for an additive BN:

```
> heur.res.mix <- search.hillclimber( score.cache=mycache.mix,
+                num.searches=1, seed=0, verbose=FALSE,
+                trace=FALSE, timing.on=FALSE)
```

Setting `trace=TRUE`, the majority consensus network is plotted as the searches progress.

4.4. Multiple Search Strategies

To estimate a robust additive BN for a given dataset is it necessary to run many searches and then summarize the results of these searches. The function `search.hillclimber()` with `num.searches>1` run multiple searches. It is necessary to use a single joint node cache over all searches, using the function `buildscorecache`.

Conceptually it may seem more efficient to use one global node cache to allow node information to be shared between different searches, however, in practice as the search space is so vast for some problems this can result in extremely *slow* searches. As the cache becomes larger it can take much more time to search it (and it may need to be searched a very large number of times) than to simply perform the appropriate numerical computation. Profiling using the google performance tool google-pprof suggests that more than 80% of the computation time may be taken up by lookups. When starting searches from different random places in the model space the number of individual node structures in common between any two searches, relative to the total number of different node structures searched over can be very small meaning a common node cache is inefficient. This may not be the case when starting networks are relatively similar.

To help with performance monitoring it is possible to turn on timings using `timing.on=TRUE` which then outputs the number of seconds of CPU time each individual search takes (using standard libc functions declared in time.h).

```
> dag33 <- matrix( 0, 33, 33)
> colnames(dag33) <- rownames(dag33) <- names(var33) #Set names
```

Create empty DAGs:

```
> banned.mix <- matrix( 0, 33, 33)
> colnames(banned.mix)<- rownames(banned.mix)<- names(var33)
> retain.mix <- matrix( 0, 33, 33)
> colnames(retain.mix) <- rownames(retain.mix)<- names(var33)
```

Setup distribution list for mixed node:

```
> mydists.mix <- list(      v1 = "binomial",  v2 = "gaussian",
+          v3 = "binomial",  v4 = "binomial",  v5 = "gaussian",
+          v6 = "binomial" , v7 = "gaussian",  v8 = "gaussian",
+          v9 = "binomial", v10 = "binomial", v11 = "binomial",
+         v12 = "binomial", v13 = "gaussian", v14 = "gaussian",
+         v15 = "binomial", v16 = "gaussian", v17 = "gaussian",
+         v18 = "binomial", v19 = "binomial", v20 = "binomial",
```

```
+         v21 = "binomial", v22 = "gaussian", v23 = "gaussian",
+         v24 = "gaussian", v25 = "gaussian", v26 = "binomial",
+         v27 = "binomial", v28 = "binomial", v29 = "gaussian",
+         v30 = "gaussian", v31 = "gaussian", v32 = "binomial",
+         v33 = "gaussian")
> n.searches <-  10
```

The number 10 is an example only, it must be much larger in practice. Set the parent limits:

```
> max.par <- 1
```

We set only one parent, because the search with `buildscorecache()` take some minutes. Now we build the cache:

```
> mycache.mix <- buildscorecache( data.df=var33, data.dists=mydists.mix,
+ dag.banned=banned.mix, dag.retained=retain.mix, max.parents=max.par)
```

Repeat but this time have the majority consensus network plotted as the searches progress:

```
> myres.mlp <- search.hillclimber(score.cache=mycache.mix,
+               num.searches=n.searches, seed=0, verbose=FALSE,
+             trace=FALSE, timing.on=FALSE)
```

4.5. Creating a Summary Network: Majority Consensus

Having run many heuristic searches, then the next challenge is to summarise these results to allow for ready identification of the joint dependencies most supported by the data. One common, and very simple approach is to produce a single robust BN model of the data mimicing the approach used in phylogenetics to create majority consensus trees. A majority consensus DAG is constructed from all the arcs present in at least 50% of the locally optimal DAGs found in the search heuristics. This creates a single summary network. Combining results from different runs of `search.hillclimber()` or `search.hillclimber()` is straightforward, although note that it is necessary to check for duplicate random starting networks, as while highly unlikely this is theoretically possible. The following code provides a simple way to produce a majority consensus network and Figure 12 shows the resulting network - note that this is an example only and many thousands of searches may need to be conducted to achieve robust results. One simple ad-hoc method for assessing how many searches are needed is to run a number of searches and split the results into two (random) groups, and calculate the majority consensus network within each group. If these are the same then it suggests that sufficient searches have been run. To plot the majority consensus network use the result of the function `search.hillclimber`, see below for some example.

```
> tographviz( dag= myres.mlp$consensus, data.df=var33,
+ data.dists=mydists.mix, outfile="dagcon.dot") #Create file
```

`dagcon.dot` can then be processed with graphviz un a unix shell typing: "dot -Tpdf dagcon.dot -o dagcon.pdf" or using gedit if on Windows.

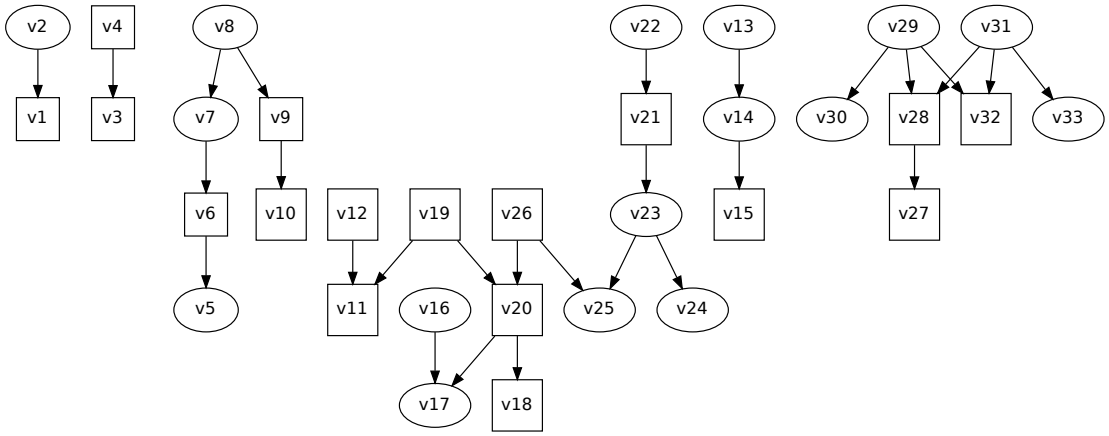


Figure 12: Example majority consensus network (from the results of only 10 searches)

4.6. Creating a Summary Network: Pruning

Rather than use the majority consensus network as the most appropriate model of the data, an alternative approach is to choose the single best model found during a large number of searches. To determine sufficient heuristic searches have been run to provide reasonable coverage of all the features of the model landscape, then again checking for a stable majority consensus network as in Section 4.5, seems a sensible approach. Once the best overall DAG has been identified then the next task is to check this model for over-fitting. Unlike with the majority consensus network, which effective "averages" over many different competing models and therefore should generally comprise only robust structural features, choosing the DAG from a single model search is far more likely to contain some spurious features. When dealing with smaller data sets, say, of several hundred observations then this is extremely likely, as can easily be demonstrated using simulated data. A simple assessment of overfitting can be made by comparing the number of arcs in the majority consensus network with the number of arcs in the best fitting model. We have found that in larger data sets the majority consensus and best fitting model can be almost identical, while in smaller data sets the best fitting models may have many more arcs - suggesting a degree of overfitting.

An advantage of choosing a DAG from an individual search is that unlike averaging over lots of different structures, as in the construction of a majority consensus network, the model chosen here has a structure which was actually found during a search across the model landscape. In contrast, the majority consensus network is a derived model which may never have been found chosen during even an exhaustive search, indeed it may even comprise of contradictory features as is a usual risk in averaging over different explanations (models) of data. In addition, a majority consensus network need also not be acyclic, although in practice this can be easily corrected by reversing one or more arcs to produce an appropriate DAG.

A simple compromise between the risk of over-fitting in choosing the single highest scoring DAG, and the risk of inappropriately averaging across different distinct data generating processes, is to prune the highest scoring DAG using the majority consensus model. In short, an element by element multiply of the highest scoring DAG and the majority consensus DAG, which gives a new DAG which only contains the structural features in *both* models.

5. Summary

The **abn** library provides a range of Bayesian network models to assist with identifying statistical dependencies in complex data, in particular models which are multidimensional analogues of generalised linear models. This process is typically referred to as structure learning, or structure discovery, and is computational extremely challenging. Heuristics are the only options for data comprising of larger numbers of variables. As with all model selection, over-modelling is an everpresent danger and using either: i) summary models comprising of structural features present in many locally optimal models or else; ii) using parametric bootstrapping to determine the robustness of the features in a single locally optimal model are likely essential to provide robust results. An alternative presented was exact order based searches, in particular finding the globally most probable structure. This approach is appealing as it is exact, but despite collapsing DAGs into orderings for larger scale problems it may not be feasible. For further in-depth analysis about **abn** refer to the website: www.r-bayesian-networks.org.

References

Babyak MA (2004). *What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models.* Psychosomatic Medicine.

Boettcher SG (2004). *Learning Bayesian networks with mixed variables*. Ph.D. thesis, Aalborg University - Department of Mathematical Sciences.

Buntine W (1991). "Theory refinement on Bayesian networks." In *Uncertainty in artificial intelligence: proceedings of the seventh conference*, pp. 52–60. Morgan Kaufmann Publishers Inc.

Djebbari A, Quackenbush J (2008). "Seeded Bayesian Networks: Constructing genetic networks from microarray data." *BMC Systems Biology*, **2**, 57.

Dojer N, Gambin A, Mizera A, Wilczynski B, Tiuryn J (2006). "Applying dynamic Bayesian networks to perturbed gene expression data." *BMC Bioinformatics*, **7**, 249.

Firestone SM, Lewis FI, Schemann K, Ward MP, Toribio JA, Dhand NK (2013). "Understanding the associations between on-farm biosecurity practice and equine influenza infection during the 2007 outbreak in Australia." *Preventive Veterinary Medicine*, **110**(1), 28–36. ISSN 0167-5877.

Firestone SM, Lewis FI, Schemann K, Ward MP, Toribio JA, Taylor MR, Dhand NK (2014). "Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a subsequent outbreak." *Preventive Veterinary Medicine*, **116**(3), 243–251. ISSN 0167-5877.

Friedman N, Goldszmidt M, Wyner A (1999). "Data analysis with Bayesian networks: A Bootstrap approach." In *Proc. Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99) (pp.206-215). San Francisco: Morgan Kaufmann*.

Friedman N, Koller D (2003). "Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks." *Machine Learning*, **50**(1-2), 95–125.

Geiger D, Heckerman D (1994). "Learning Gaussian networks." In *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, UAI'94, pp. 235–243. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 1-55860-332-8.

Heckerman D, Geiger D, Chickering DM (1995). "Learning Bayesian networks: the combination of knowledge and statistical data." *Machine Learning*, **20**(3), 197–243.

Hodges AP, Dai DJ, Xiang ZS, Woolf P, Xi CW, He YQ (2010). "Bayesian network expansion identifies new ros and biofilm regulators." *PLOS One*, **5**(3), e9513.

Jansen R, Yu HY, Greenbaum D, Kluger Y, Krogan NJ, Chung SB, Emili A, Snyder M, Greenblatt JF, Gerstein M (2003). "A Bayesian networks approach for predicting protein-protein interactions from genomic data." *Science*, **302**(5644), 449–453.

Jensen FV (2001). *Bayesian network and decision graphs*. Springer-Verlag, New York.

Koivisto M, Sood K (2004). "Exact Bayesian structure discovery in Bayesian networks." *Journal of Machine Learning Research*, **5**, 549–573.

Lauritzen SL (1996). *Graphical Models*. Oxford Univ Press, New York.

Lewis F, Ward M (2013). "Improving epidemiologic data analyses through multivariate regression modelling." *Emerging Themes in Epidemiology*, **10**(1), 4. ISSN 1742-7622.

Lewis FI (2012). "Bayesian networks as a tool for epidemiological systems analysis." In S Sivasundaram (ed.), *9th International Conference On Mathematical Problems In Engineering, Aerospace and Sciences (icnpaa 2012), Amer Inst Physics*, volume 1493 of *AIP Conference Proceedings*, pp. 610–617. ISBN 978-0-7354-1105-0. ISSN 0094-243X.

Lewis FI, Brulisauer F, Gunn GJ (2011). "Structure discovery in Bayesian networks: An analytical tool for analysing complex animal health data." *Preventive Veterinary Medicine*, **100**(2), 109–115.

Lewis FI, McCormick BJJ (2012). "Revealing the complexity of health determinants in resource-poor settings." *American Journal of Epidemiology*, **176**(11), 1051–1059.

Ludwig A, Berthiaume P, Boerlin P, Gow S, Léger D, Lewis FI (2013). "Identifying associations in Escherichia coli antimicrobial resistance patterns using additive Bayesian networks." *Preventive Veterinary Medicine*, **110**(1), 64–75. ISSN 0167-5877.

Lycett SJ, Ward MJ, Lewis FI, Poon AFY, Pond SLK, Brown AJL (2009). "Detection of mammalian virulence determinants in highly pathogenic avian influenza H5N1 viruses: multivariate analysis of published data." *Journal of Virology*, **83**(19), 9901–9910.

Mackay DJC (1992). "Bayesian Interpolation." *Neural Computation*, **4**(3), 415–447.

McCormick B, Sanchez-Vazquez M, Lewis F (2013). "Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs." *Preventive Veterinary Medicine*, **110**(1), 54–63. ISSN 0167-5877.

Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR (2007). "A primer on learning in Bayesian networks for computational biology." *PLOS Computational Biology*, **3**(8), e129.

Plummer M (2003). "JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling." *Proc 3rd Int Work Dist Stat Comp*.

Plummer M, Best N, Cowles K, Vines K (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News*, **6**(1), 7–11.

Poon AFY, Lewis FI, Frost SDW, Pond SLK (2008). "Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models." *Bioinformatics*, **24**(17), 1949–1950.

Poon AFY, Lewis FI, Pond SLK, Frost SDW (2007a). "Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope." *PLOS Computational Biology*, **3**(1), e11.

Poon AFY, Lewis FI, Pond SLK, Frost SDW (2007b). "An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope." *PLOS Computational Biology*, **3**(11), e231.

Rijmen F (2008). "Bayesian networks with a logistic regression model for the conditional probabilities." *International Journal of Approximate Reasoning*, **48**(2), 659–666.

Sanchez-Vazquez M, Nielen M, Edwards S, Gunn G, Lewis F (2012). "Identifying associations between pig pathologies using a multi-dimensional machine learning methodology." *BMC Veterinary Research*, **8**(1), 151. ISSN 1746-6148.

Schemann K, Lewis FI, Firestone SM, Ward MP, Toribio JALML, Taylor MR, Dhand NK (2013). "Untangling the complex inter-relationships between horse managers' perceptions of effectiveness of biosecurity practices using Bayesian graphical modelling." *Preventive Veterinary Medicine*, **110**(1, SI), 37–44. ISSN 0167-5877.

Ward MP, Lewis FI (2013). "Bayesian graphical modelling: applications in veterinary epidemiology." *Preventive Veterinary Medicine*, **110**(1), 1–3. ISSN 0167-5877.

Wilson AJ, Ribeiro R, Boinas F (2013). "Use of a Bayesian network model to identify factors associated with the presence of the tick Ornithodoros erraticus on pig farms in southern Portugal." *Preventive Veterinary Medicine*, **110**(1, SI), 45–53. ISSN 0167-5877.

Affiliation:

Marta Pittavino
Institute of Mathematics, University of Zurich
Winterthurerstrasse 190, Zurich 8057, Switzerland
E-mail: marta.pittavino@math.uzh.ch

Fraser Ian Lewis
Office of Health Economics
United Kingdom, London
E-mail: flewis@ohe.org

Reinhard Furrer
Institute of Mathematics, University of Zurich
Institute of Computational Science, University of Zurich
Winterthurerstrasse 190, Zurich 8057, Switzerland
E-mail: reinhard.furrer@math.uzh.ch

abn: an R package for modelling data using additive Bayesian networks

M Pittavino[1] **FI Lewis**[2] **R Furrer**[1]

[1] Institute of Mathematics, University of Zurich, Switzerland, Zurich
[2] Office of Health Economics, United Kingdom, London

marta.pittavino@math.uzh.ch, flewis@ohe.org, reinhard.furrer@math.uzh.ch

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION
SNF 138562, SNF 144973

Additive Bayesian network: ABN

Bayesian network analysis is a form of probabilistic modeling which derives from empirical data a directed acyclic graph (DAG) describing the dependency structure between random variables. An ABN consists of a DAG where each node in the graph comprises a generalized linear model (GLM).
ABNs can be applied in areas such as epidemiological and medical analysis.
Model search algorithms are used to identify those DAG structures most supported by the data. This process is typically referred to as structure learning and is computational extremely challenging. We propose new functions to do find a good ABN.

Main steps to perform data analysis using *abn*:

1. Load the data and prepare them for searching procedure
2a. Find the best fitting ABN structure using an exact search: `mostprobable`
2b. Find a fitting ABN structure using a heuristic search: `search.hillclimber`
3. Fit an ABN to data and estimate the posterior densities in ABN: `fitabn`
4. Parameter interpretation and conclusions

1. Load the data and prepare them for searching procedure

Load the data, included as part of `abn` R package, comprising of 10'000 observations from 10 variables:

```
> library( abn)
> mydat <- ex1.dag.data #see ?ex1.dag.data
```

Setup distribution list for each node, e.g. each variable, in the data set where:

- b: binomial,
- g: Gaussian,
- p: Poisson.

```
> mydists <- list( b1="binomial",p1="poisson",
+                  g1="gaussian",b2="binomial",
+                  p2="poisson",b3="binomial",
+                  g2="gaussian",b4="binomial",
+                  b5="binomial",g3="gaussian")
```

Set two matrices, `ban` and `ret`, refering to the arcs that can not be and which are forced to be present in the model.
Here we impose no constraints. Column and row names must be set:

```
> ban <- matrix( 0, nrow=10, ncol=10)
> colnames( ban) <- rownames( ban) <- names( mydat)
> ret <- matrix( 0, nrow=10, ncol=10)
> colnames( ret) <- rownames( ret) <- names( mydat)
```

Limited number of parents present in the model:

```
> max.par <- 4
```

The function `buildscorecache` builds a cache of goodness of fit, using default Gaussian uninformative parameter prior and uniform structural prior:

```
> mycache <- buildscorecache( data.df=mydat,
+                 dag.banned=ban, data.dists=mydists,
+                 dag.retained=ret, max.parents=max.par)
```

2a. Find the best fitting ABN structure using an exact search

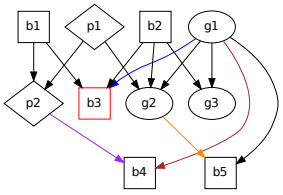
The function `mostprobable` finds the globally best DAG.
The exact approach is only feasible with a small number of variables, up to 20:

```
> best.dag <- mostprobable( score.cache=mycache)
```

The function `tographviz` creates a text file suitable for plotting with graphviz:

```
> tographviz( dag.m=best.dag,data.df=mydat,
+             data.dists=mydists,outfile="BestDag.dot")
```

Best DAG using exact search

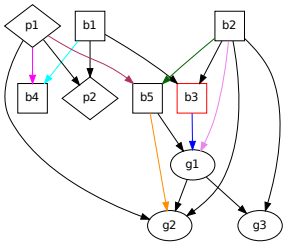


2b. Find a fitting ABN structure using a heuristic search

Heuristics are the only options for data set larger than 20 variables.
The function `search.hillclimber` repeats the search locally and plots the majority consensus network as the searches progress:

```
> heur.res <- search.hillclimber( score.cache=mycache,
+                  num.searches=1000, trace=TRUE,
+                  verbose=FALSE, timing.on=FALSE)
```

"Majority consensus" DAG from 1000 heuristic searches. Differences with best DAG are highlighted with colours



3. Fit an ABN to data and estimate the posterior densities

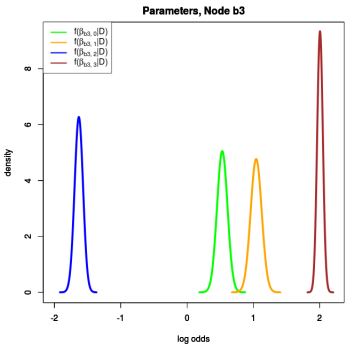
The function `fitabn` uses Laplace approximations to estimate the marginal posterior density for each parameter in the model.
We fit the model to calculate its goodness of fit, log marginal likelihood:

```
> myres.bestdag <- fitabn( dag.m=best.dag,data.df=mydat,
+                          data.dists=mydists)
```

The parameters can be estimated one at a time by manually giving a grid. We plot the marginal posterior densities for node b3 with three parents, the covariates b1, b2 and g1, as in the model:

$$\text{logit}\{P(\text{b3}=1)\} = \beta_{\text{b3},0} + \beta_{\text{b3},1} \cdot \text{b1} + \beta_{\text{b3},2} \cdot \text{b2} + \beta_{\text{b3},3} \cdot \text{g1}$$

Marginal posterior densities for intercept and covariates b1, b2 and g1 in node b3



4. Parameter interpretation and conclusions

The log odds ratio here is highly significantly different from 0. All of the effect parameters, have 95% credible intervals which do not cross the origin. This is not guaranteed to happen because no constraint was included in the model searching process, all parameters in the model have been justified using marginal likelihood.

The R package `abn` provides functionality for identifying statistical dependencies in complex data using ABN models, which are multidimensional analogues of GLMs. The key distinction between GLM techniques and ABN analysis is that the latter tries not only to identify statistically associated variables, but also to separate them into those directly and indirectly dependent. In conclusion, data analysis using ABNs have the potential to offer new insights into complex epidemiological systems.

References

www.r-bayesian-networks.org

PAPER IV

Conjugate Priors for Additive Bayesian Networks

Marta Pittavino & Reinhard Furrer

Paper under revision for submission to *Bayesian Analysis*.

Bayesian Analysis (0000) **00**, Number 0, pp. 1

Conjugate Priors for Additive Bayesian Networks

Marta Pittavino* and Reinhard Furrer*,†

Abstract. This paper addresses the parameter learning process of an additive Bayesian network (ABN) model for binary data. When an additive parametrization for Bayesian networks is used, the marginal likelihood (ABN network score) computation is the major objective. In this paper, we introduce a novel conjugate prior distribution for ABN that belongs to a flexible family of conjugate priors called the Diaconis−Ylvisaker conjugate priors. We show that the suggested prior is a generalization of the Dirichlet prior. Moreover, we prove that this prior satisfies the desirable independence assumptions for a parameter prior in DAG models. Hence, it helps to address the goodness of fit calculation. The resulting ABN network score is equal to the Gaussian ordinary hypergeometric function. However, it can be approximated using the Laplace method. We then present a method for selecting the hyperparameter priors in order to have the score equivalence property satisfied. Finally, the priors, the derived methods and the usefulness are illustrated by means of an example of a binary variable network.

MSC 2010 subject classifications: Primary 62C10, 62F15; secondary 62J12, 62P10, 62C12, 92C42.

Keywords: conjugacy, graphical models, marginal likelihood, hypergeometric function, Laplace approximation, score equivalence.

1 Introduction

Additive Bayesian network (ABN) models are types of graphical models that extend the usual multinomial logistic regression to multiple dependent variables through the representation of joint probability distribution. ABNs are a special type of Bayesian network (BN) models in that each node in the graph comprises a generalized linear model. With regard to the latter, they consist of statistical models that derive a directed acyclic graph (DAG) from empirical data, describing the dependency structure of random variables. All types of BN models consist of two reciprocally dependent parts, namely qualitative (the structure) and quantitative (the model parameters) parts. The DAG is the graphical representation of the joint probability distribution of all random variables in the data. The model parameters are represented by a local probability distribution for all the variables in the network. All the techinical details of BN and ABN models, together with the learning process, are presented in Sections 2 and 3.

In the last few decades, BN modelling has been widely used in biomedical science and in systems biology Poon et al. (2007a, 2008, 2007b); Needham et al. (2007); Dojer

*Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland marta.pittavino@math.uzh.ch reinhard.furrer@math.uzh.ch

†Institute of Computational Science, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland reinhard.furrer@math.uzh.ch

 DOI: 0000

et al. (2006); Jansen et al. (2003); Djebbari and Quackenbush (2008); Hodges et al. (2010) to analyse multi-dimensional data. However, it is only in the last few years that ABN models have been applied to the veterinary epidemiology field as a result of their ability to generalize standard regression methodologies. A general introduction to BN modelling in veterinary epidemiology is provided by Lewis et al. (2011). Further applications of BN to veterinary studies were described by Ward and Lewis (2013); Wilson et al. (2013); Sanchez-Vazquez et al. (2012). Graphical modelling techniques used to analyse epidemiological data were used by Firestone et al. (2013, 2014); Lewis and McCormick (2012); Lewis (2012); Lewis and Ward (2013); Schemann et al. (2013); Ludwig et al. (2013); McCormick et al. (2013), resulting in dozens of publications, and references therein.

The result of fitting an ABN to a data set is translated into optimizing the likelihood of observing the data under the studied model: the marginal likelihood (ABN network score). The marginal likelihood computation is the major objective when dealing with ABN models. The ABN network score is an integral over all the parameters that can be interpreted as the probability that we could generate the data set if we were to select the parameters for the structure randomly according to the parameter prior. To date, we are only aware of approximation and numerical techniques, such as that of Laplace (Tierney and Kadane, 1986) and the Newton method to compute the ABN network score (Lewis, 2012). The crucial part in this computation is the choice of the parameter prior. So far, only (uninformative) Gaussian priors, with a mean of zero and a large precision, have been taken into account. Unfortunately, this leads to very high costs in terms of computation time due to the difficulty of the integral evaluation.

In this paper, in order to simplify the marginal likelihood computation, in Section 4 we introduce a novel conjugate prior distribution for ABN that belongs to a flexible family of conjugate priors called the Diaconis−Ylvisaker conjugate prior (Diaconis and Ylvisaker, 1979). Moreover, this prior is also a special case of the classical Zellner's g-prior (Zellner, 1986), as shown in Sabanés Bové and Held (2011). We show that the suggested prior is a generalization of the Dirichlet density. We then ensure that the desirable independence assumptions are fulfilled (each node can be considered separately), as this is crucial in a parameter prior for DAG models, and in order to gain strong simplifications in the integral computation.

The advantages brought about by the conjugacy and the prior's independence properties are discussed and exploited in Section 5. The resulting ABN network score can be represented as the ${}_2F_1$ ordinary Gaussian hypergeometric functions. As is traditionally done in the Bayesian literature (e.g. Bernardo and Smith, 2000), it can be also approximated using the Laplace method. We illustrate the difference between the results found using the Laplace approximatio via some graphical examples. The posterior density, resulting from the choice of the prior, the derived methods and their usefulness are illustrated by the example of a network of binary variables.

Finally, we provide some concluding remarks about possible improvements and limitations of the developed methodology in Section 6. In particular, we mention a simplification of the ${}_2F_1$ hypergeometric function, represented by its link with the Beta Truncated function. We suggest some possible extensions of the additive models via

other distributions, that are always part of the exponential family and the generalized linear models. Proofs and technical results are presented in the appendix.

In the supplementary material, we prove that the introduced prior satisfies the *likelihood equivalence* property (data should not help to discriminate between network structures that represent the same assertions of conditional independence), and we present a method for selecting the hyperparameter priors in order to have the *score equivalence* property satisfied between two models that are *likelihood equivalent*. The '*equivalence*' part has not been placed in the appendix because it is a corollary of the results found, and to reduce the number of pages.

2 From Bayesian to additive Bayesian networks

A Bayesian network is a form of graphical model that derives a directed acyclic graph from empirical data, describing the dependency structure between random variables. It provides a compact representation of the joint probability distribution using a combination of graph (the qualitative part) and probability (the quantitative part) theory. The technical foundations of BN modelling lie within the machine learning and data mining literature (Cooper and Herskovits, 1992; Friedman et al., 1997; Friedman and Koller, 2003; Jensen, 2001; Boettcher, 2004; Heckerman et al., 1995).

More precisely, a BN model $\mathcal{B}$ for a set of random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ consists of:

- A *directed acyclic graph* (DAG) structure $\mathcal{S} = (V, E)$, where V is a finite set of vertices or nodes and E is a finite set of directed edges between the vertices. A DAG is *acyclic*; hence, the edges in E do not form directed cycles. A random variable X_j corresponds to each node $j \in V = \{1, \ldots, n\}$ in the graph. We do not distinguish between a variable X_j and the corresponding node j.
- A set of parents for a node j is denoted by $\mathbf{Pa}_j$. A vertex j is said to be a *parent* of a node k if the edge set E contains an edge from j to k. P_j indicates the total number of parents for a node j : $\dim(\mathbf{Pa}_j) = P_j$.
- A set of local probability distributions for all variables in the network called $\boldsymbol{\theta}_{\mathcal{B}}$. Each node j, with parent set $\mathbf{Pa}_j$, is parametrized by a local probability distribution: $P(X_j|\mathbf{Pa}_j)$.

Edges represent both *marginal* and *conditional dependencies*. The main role of the network structure is to express the conditional independence relationships among the variables in the model through graphical separation, thus specifying the factorization of the global probability distribution:

$$P(\boldsymbol{X}) = \prod_{j=1}^{n} P(X_j|\mathbf{Pa}_j). \tag{2.1}$$

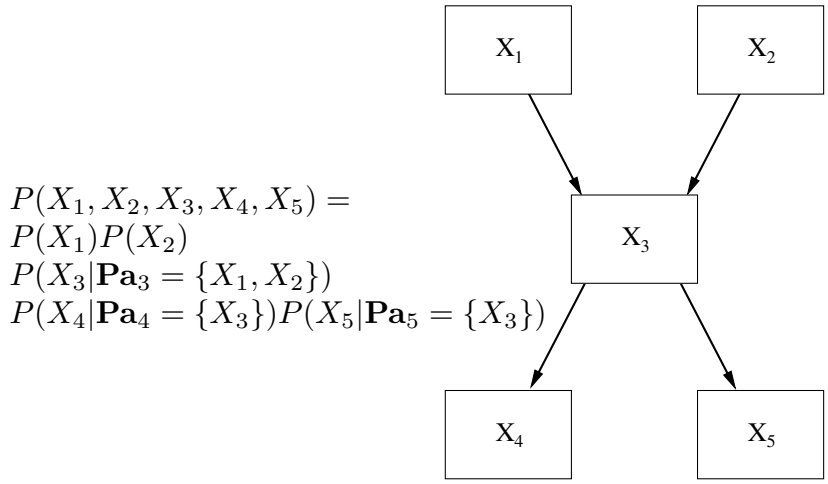


Figure 1: A Bayesian network model $\mathcal{B}$ for five random variables.

We denote a BN model, $\mathcal{B}$, for a set of random variables, $\boldsymbol{X}$, by a pair $\mathcal{B} = (\mathcal{S}, \boldsymbol{\theta}_{\mathcal{B}})$. The DAG defines the *structure* $\mathcal{S}$, and $\boldsymbol{\theta}_{\mathcal{B}}$ the *parametrization* of the model $\mathcal{B}$. In order to specify a $\mathcal{B}$ for $\boldsymbol{X}$, we must therefore specify a DAG structure and a set of local probability distributions.

Figure 1 shows an example of $\mathcal{B}$ for five random variables; the joint probability distribution can be factorized into five factors, one for each random variable conditioned on its parents: $P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2)P(X_3|\mathbf{Pa}_3 = \{X_1, X_2\})$ $P(X_4|\mathbf{Pa}_4 = \{X_3\})P(X_5|\mathbf{Pa}_5 = \{X_3\})$.

2.1 Additive Bayesian networks

In order to introduce an additive Bayesian Network (ABN) model $\mathcal{A}$, some further notation is needed.

Let S_j be the number of states of the variable X_j, and $s = \{1, \ldots, S_j\}$ the corresponding set of indexes. Let $C_j = \prod_{p:\, X_p \in \mathbf{Pa}_j} S_p$ be the number of configurations of $\mathbf{Pa}_j$ and $c = \{1, \ldots, C_j\}$ indicates the corresponding set of indexes for the different parents configurations of $\mathbf{Pa}_j$. Let $X_j = s$ indicate the possible observations for X_j. Hence, let $P(X_j = s|\mathbf{Pa}_j = c)$ be the probability that $X_j = s$, given the c-th parent configuration of $\mathbf{Pa}_j$, denoted by the multinomial parameter θ_{jcs}. Therefore, the following notation is used:

$$\boldsymbol{\theta}_{jc} = \bigcup_{s=1}^{S_j} \{\theta_{jcs}\},$$
$$\boldsymbol{\theta}_{j} = \bigcup_{c=1}^{C_j} \{\boldsymbol{\theta}_{jc}\},$$

$$\boldsymbol{\theta}_{\mathcal{B}} = \bigcup_{j=1}^{n} \{\boldsymbol{\theta}_j\}.$$

This means that $\boldsymbol{\theta}_{jc}$ denotes the set of local probability distributions associated with a node j, its parent configuration c and node states s. $\boldsymbol{\theta}_j$ denotes the set of all parameters associated with a node j and its parent configuration c. $\boldsymbol{\theta}_{\mathcal{B}}$ denotes the set of local probability distributions for all variables in the Bayesian network $\mathcal{B}$. All the local probability distributions are unrestricted, discrete distributions with $P(X_j = s|\mathbf{Pa}_j = c) \geq 0 \; \forall j \in V$. Then $\sum_{s=1}^{S_j} \theta_{jcs} = 1$ and $0 \leq \theta_{jcs} \leq 1$.

Using this parametrization, the joint probability distribution factorizes into:

$$P(\boldsymbol{X}|\boldsymbol{\theta}_{\mathcal{B}}, \mathcal{S}) = \prod_{j=1}^{n} P(X_j|\mathbf{Pa}_j = c, \boldsymbol{\theta}_{jc}).$$

It is now possible to introduce an ABN model $\mathcal{A}$. In this work, we are going to refer to additive Bayesian networks with the abbreviation ABN and the notation $\mathcal{A}$ used interchangeably.

An additive Bayesian network $\mathcal{A}$ consists of a Bayesian network $\mathcal{B}$ that generalizes the multinomial logistic regression model $\mathcal{M}$ (Rijmen, 2008). The multinomial logistic regression model $\mathcal{M}$ can be integrated into a Bayesian network $\mathcal{B}$ by modelling each conditional probability table $P(X_j|\mathbf{Pa}_j) = \theta_{jcs}$ of a particular Bayesian network $\mathcal{B}$ via a multinomial logistic regression model, where X_j is progressively the outcome variable, and the design matrix $\boldsymbol{Z}_{ij}$ is constructed from $\mathbf{Pa}_j$.

An additive Bayesian network model $\mathcal{A}$ without restrictions on the conditional probability tables is obtained by constructing $\boldsymbol{Z}_{ij}$ from $\mathbf{Pa}_j$ as follows. For each possible configuration c on $\mathbf{Pa}_j$, $c = 1, \ldots, C_j$, a dummy variable is defined. For each observation i, the covariate vector $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijC_j})^T$ is defined as an indicator vector with $z_{ijc} = 1$ if the configuration c is observed, and $z_{ijc} = 0$ if not. The $(S_j - 1) \times (S_j - 1)C_j$ design matrix $\boldsymbol{Z}_{ij}$ is constructed from $\boldsymbol{z}_{ij}$ and $\boldsymbol{\beta}_j = (\beta_{j11}, \ldots, \beta_{jcs}, \ldots, \beta_{jC_j(S_j-1)})^T$, of dimension $(S_j - 1)C_j \times 1$, is the coefficients vector for the *additive parameters*. Then, the expression for the linear predictor for each observation i is instantiated by:

$$\boldsymbol{\eta}_{ij} = \begin{pmatrix} \eta_{ij1} \\ \vdots \\ \eta_{ij(S_j-1)} \end{pmatrix} = \boldsymbol{Z}_{ij}\boldsymbol{\beta}_j = \begin{bmatrix} \boldsymbol{z}_{ij}^T & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{z}_{ij}^T \end{bmatrix} \begin{bmatrix} \beta_{j11} \\ \vdots \\ \beta_{jC_j(S_j-1)} \end{bmatrix}.$$

The corresponding conditional probabilities are obtained by applying the inverse of the link function to the linear predictor.

Therefore, we denote an additive Bayesian network model $\mathcal{A}$ for a set of random variables $\boldsymbol{X}$ by a pair $\mathcal{A} = (\mathcal{S}, \boldsymbol{\beta}_{\mathcal{A}}) = (\mathcal{S}, h(\boldsymbol{\theta}_{\mathcal{B}}))$, where $h(\boldsymbol{\theta}_{\mathcal{B}}) = \text{logit}(\boldsymbol{\theta}_{\mathcal{B}}) = \boldsymbol{\beta}_{\mathcal{A}}$. The main difference between a $\mathcal{B}$ and a $\mathcal{A}$ is the re-parametrization of the $\boldsymbol{\theta}_{\mathcal{B}}$ parameters, seen as a function of the *additive parameters* $\boldsymbol{\beta}_{\mathcal{A}}$. From the definition of the additive Bayesian network model $\mathcal{A}$, a 'transformed' notation can be used to indicate the parameters in

an $\mathcal{A}$ resulting from the *logit link transformation function*, with a similar meaning to that of a $\mathcal{B}$ model:

$$\boldsymbol{\beta}_{\mathcal{A}} = \text{logit}(\boldsymbol{\theta}_{\mathcal{B}}) = h(\boldsymbol{\theta}_{\mathcal{B}}) = \bigcup_{j=1}^{n} \{\boldsymbol{\beta}_j\},$$

$$\boldsymbol{\beta}_j = \bigcup_{c=1}^{C_j} \{\boldsymbol{\beta}_{jc}\},$$

$$\boldsymbol{\beta}_{jc} = \bigcup_{s=1}^{S_j} \{\beta_{jcs}\}.$$

In this work, we are interested in specifying networks for random variables $\boldsymbol{X}$ that follow a Bernoulli distribution, as specified in Dai et al. (2013). Therefore, we are going to work with networks containing variables with only two states: *binary* variables. Hence, a special case of the multinomial logistic regression model is treated, namely the binary logistic regression model. Therefore, in this specific discrete Bernoulli case, we have $S_j = 2$ and $C_j = \prod_{p:\, X_p \in \mathbf{Pa}_j} S_p = 2^{P_j}$ as the number of configurations of $\mathbf{Pa}_j$. In particular, each conditional probability table $P(X_j = 1|\mathbf{Pa}_j) = \theta_{jc1}$ from a $\mathcal{B}$ is modelled via a binary logistic regression model. Hence, we get:

$$\theta_{jc1} = \frac{e^{\boldsymbol{Z}_{ij}\boldsymbol{\beta}_j}}{1 + e^{\boldsymbol{Z}_{ij}\boldsymbol{\beta}_j}} = \frac{e^{\beta_{jc1}}}{1 + e^{\beta_{jc1}}} \Rightarrow \beta_{jc1} = h(\theta_{jc1}) = \text{logit}(\theta_{jc1}) = \log\left(\frac{\theta_{jc1}}{1 - \theta_{jc1}}\right).$$

The main novelty for an additive Bayesian network is the change of focus from a parametrization expressed in terms of θ_{jcs} to a corresponding one represented in terms of β_{jcs}. It is a one-to-one transformation from the $\boldsymbol{\theta}_{\mathcal{B}}$ to the $\boldsymbol{\beta}_{\mathcal{A}}$ parameters. From now on, we are going to work with the additive parameters and model $\mathcal{A} = (\mathcal{S}, \boldsymbol{\beta}_{\mathcal{A}})$ instead of using the standard Bayesian network notation $\mathcal{B} = (\mathcal{S}, \boldsymbol{\theta}_{\mathcal{B}})$. It will be clarified and used, only if necessary.

3 Learning an additive Bayesian network model

In the Bayesian network literature, Boettcher (2004); Buntine (1991); Friedman and Koller (2003); Heckerman (1998); Heckerman et al. (1995), the parameter estimation and the model selection process are known as *learning*: 1) *parameter learning*: specifying the local probability distributions (model parameters $\boldsymbol{\beta}_{\mathcal{A}}$); and 2) *structure learning*: specifying the DAG structure $\mathcal{S}$. Hence, when constructing an additive Bayesian network model $\mathcal{A}$, two steps need to be considered. Being in a Bayesian framework, given a data set $\mathcal{D}$, we have:

$$P(\mathcal{A}|\mathcal{D}) = \underbrace{P(\boldsymbol{\beta}_{\mathcal{A}}, \mathcal{S}|\mathcal{D})}_{\text{model learning}} = \underbrace{P(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}, \mathcal{D})}_{\text{parameter learning}} \cdot \underbrace{P(\mathcal{S}|\mathcal{D}).}_{\text{structure learning}}$$

Both the *learning* procedures are relevant and necessary in order to understand the final model. They are interconnected and dependent on each other. Even though, the

major aim in this work is the choice of the parameter prior and its properties, mainly linked to parameter learning, both the procedures will be clarified. First, the *structure learning* process is explained. The principal score functions, and related searching strategies to look for the best model, are presented. The *parameter learning* process is then presented via a list of key assumptions that helps to simplify the most demanding computations.

3.1 Learning the Structure

In this section, the main score functions, used for some searching methods to learn the structure of an ABN network, are presented. Finally, the specific ABN network score and its related difficulties in the computation are described.

Learning the Structure of Bayesian networks

The aim of this section is to introduce the process of learning the structure of an ABN, similar to that in Jensen (2001).

Consider having a data set $\mathcal{D}$ from an ABN $\mathcal{A}_1$ over the set of variables $\boldsymbol{X}$. The task is now to find a Bayesian network $\mathcal{A}_2$ from the data set $\mathcal{D}$ that is close to $\mathcal{A}_1$. In theory, this can be done by performing parameter learning for all possible structures, and then selecting as candidates those models for that the distribution of $\mathcal{A}_2$ is close to the sample distribution. Unfortunately, by following this simplified approach, computational problems and issues with feasibility arise. Moreover, in Chickering (1996); Chickering et al. (2004) are presented one of the latest in a series of results that show that the task of learning Bayesian network structures is NP-hard.

Therefore, another searching strategy needs to be followed. The first method for the automated learning of a BN was the method that learned tree-structured models (Chow and Liu, 1968). At present, there are two different types of methods for learning the structure of a BN: *constraint-based* and *score-based*. The first establishes a set of conditional independence statements holding for the data, and uses this set to build a network with graphical separation properties corresponding to the conditional independence properties determined. The second creates some candidate BNs, calculates a score for each candidate, and returns the network with the highest score. Cowell (2001) has shown that, according to often quoted assumptions, constraint-based learning and score-based learning are equivalent.

In the next section score-based methods are presented, because of their link with the score function, which is a crucial element in this work.

Score-based learning methods

When performing structural learning, the aim is to look for a BN structure that can represent the data set sufficiently well without being overly complex.

Score-based methods assign a number (a *score*) to each BN structure. The score

reflects the 'usefulness' of the structure, in other words, how likely it is that the structure could have been used to generate the data set at hand. The task of score-based learning can then be considered to be a search problem: looking for the model structure with the highest score.

Therefore, a BN can be learned from a data set by performing a search of all the DAGs and selecting the one with the highest score. Hence, in order to specify a score-based learning algorithm entirely, two components are needed: a score function and a search procedure.

The search procedure and the equivalent class search are presented in the supplementary material, while we introduce the Bayesian score functions below.

Bayesian score functions

A good score function should, at least have the following two properties: (a) a balance between the accuracy and the complexity of the structure; and (b) it should be computationally tractable to evaluate.

Moreover, a desirable property for a score function is the *decomposability*, that occurs if it can be expressed as a sum of local scores, one for each node in the data $\mathcal{D}$:

$$score\left(\mathcal{D},\mathcal{S}\right)=\sum_{j=1}^{n}score\left(X_j,\mathbf{Pa}_j,\mathcal{D}\right).$$

An example of a good Bayesian score function, that contains both a term measuring how well the data fits the model and a term that controls model complexity, is the *Bayesian Information Criterion* (BIC) (Bernardo and Smith, 2000).

The *marginal likelihood* is the classical Bayesian approach for measuring the fitness of a candidate BN structure, $\mathcal{S}$. Specifically, we have:

$$P\left(\mathcal{S}|\mathcal{D}\right)=\frac{P\left(\mathcal{S}\right)P\left(\mathcal{D}|\mathcal{S}\right)}{P\left(\mathcal{D}\right)}, \tag{3.1}$$

where $P\left(\mathcal{D}\right)^{-1}$ is the normalization constant, and is considered a constant because it does not depend on $\mathcal{S}$. From (3.1), it is easy to see that, in order to score a structure based on its posterior probability given the data, we need two terms, namely the prior probability for the structures $P\left(\mathcal{S}\right)$ and the marginal likelihood of the structure given the data $P\left(\mathcal{D}|\mathcal{S}\right)$. Generally, the prior probability distribution for the structures is chosen in order to be relatively easy to calculate (it is usually assumed that all structures are equally supported, leading to an uninformative structure prior) or with appropriate studies Scutari (2013). Therefore, *the main computational problem is the calculation of the marginal likelihood* which is needed to deal with the parameters of the model $\boldsymbol{\beta}_{\mathcal{A}}$:

$$P\left(\mathcal{D}|\mathcal{S}\right)=\int_{\boldsymbol{\beta}_{\mathcal{A}}}P\left(\mathcal{D}|\mathcal{S},\boldsymbol{\beta}_{\mathcal{A}}\right)\pi\left(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}\right)d\boldsymbol{\beta}_{\mathcal{A}}, \tag{3.2}$$

where $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S})$ is the prior probability distribution over the parameters, conditioned on $\mathcal{S}$. The integral in the above equation is over all the parameters and over all the possible Bayesian networks with the same structure, but with different conditional probability distributions. Intuitively, the marginal likelihood can therefore be interpreted as the probability that the data $\mathcal{D}$ could be generated if the parameters for $\mathcal{S}$ were selected randomly according to the parameter prior $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S})$.

As specified above, the difficult part in the calculation of $P(\mathcal{D}|\mathcal{S})$ is the evaluation of the integral in (3.2). Fortunately, it has been shown by Cooper and Herskovits (1992); Heckerman et al. (1995) that, for a standard Bayesian network model $\mathcal{B}$, the evaluation of this integral can be reduced to a simple counting problem, that can be executed in polynomial time based on three crucial assumptions for the data set $\mathcal{D}$ (A1 to A3) and five regarding the parameters (A4 to A8), that are later clarified in the *Learning the Parameters* section. The first 3 assumptions are important to guarantee that we are working with data that are fully representative of a BN, because their completeness and independence facilitate the factorization of each entry. In particular, we have:

A1. The data set $\mathcal{D}$ is a faithful sample of a Bayesian network.

A2. Observations in the data set $\mathcal{D}$ are independent, given the Bayesian network model.

A3. The data set $\mathcal{D}$ is complete.

We work with a fully observed data set $\mathcal{D} = \{x_{1\cdot}, \ldots, x_{m\cdot}\}$, where each $x_{i\cdot}$ is a set of simultaneous values of the set of variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$. The data set $\mathcal{D}$ is an $m \times n$ matrix, where the n columns are associated with the random variables in $\boldsymbol{X}$ and the m rows are the related realizations. Hence, we consider a data set $\mathcal{D}$ that fulfils assumptions A2 and A3.

In the literature, it has been shown that the BIC score of a model is an asymptotic approximation of the marginal likelihood of that model, and it is equivalent to the minimum description length proposed by Rissanen (1987), and adopted as a decomposable consistent score for Bayesian networks by Lam and Bacchus (1994) and Friedman and Goldszmidt (1998). A Bayesian metric for scoring models was proposed by Cooper and Herskovits (1992), where a search algorithm that performs a greedy search conditioned on a linear ordering of the variables (known as the K2 algorithm) was also suggested. Finally, Friedman and Koller (2003) provided a method for calculating the posterior probability of the absence or presence of individual arcs in the generating net given the data.

The score function for ABN

The *marginal likelihood* has been introduced previously. In this subsection, we describe how it is adapted for ABN models. As a result of the decomposability property of the score function, the total network score, the *marginal likelihood* for an ABN model, can be written as $P(\mathcal{D}|\mathcal{S}) = \prod_{j=1}^{n} P(\mathcal{D}_j|\mathcal{S})$. In a binary logistic additive Bayesian network

model $\mathcal{A}$, the network score for node j is given by:

$$P(\mathcal{D}_j|\mathcal{S}) = \int_{\boldsymbol{\beta}_j} \prod_{i=1}^{m} \left(\frac{e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{x_{ij}} \left(\frac{1}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{1-x_{ij}} \pi(\boldsymbol{\beta}_j|\mathcal{S}) d\boldsymbol{\beta}_j, \tag{3.3}$$

where $\mathcal{D}_j$ are the observed data at node j, and consist of tuples of $[x_{ij}, \boldsymbol{z}_{ij}^T]$. The parameter vector at node j is represented by $\boldsymbol{\beta}_j$, and has the same length as the possible parent configuration: $\dim(\boldsymbol{\beta}_j) = C_j$. The prior at node j is indicated by $\pi(\boldsymbol{\beta}_j|\mathcal{S})$, and is the unknown quantity that we characterize.

The main difficulty in moving towards an additive model is the computation of the *marginal likelihood*. In fact, additive Bayesian network models $\mathcal{A}$ require considerably more computational time than do standard Bayesian network models $\mathcal{B}$ because, thus far, no work that aims to simplify the integral has been developed (3.3) in a similar framework to that of Cooper and Herskovits (1992); Heckerman et al. (1995); Geiger and Heckerman (1994); Boettcher (2004).

We aim to show that a simplified expression for (3.3), based on assumptions A1 to A3 and A4 to A8, listed below, can be also obtained for an ABN model. In order to achieve this goal, a crucial role is played by the prior $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S})$, that has to be chosen properly and accurately.

We are going now to explain how the parameters estimation is conducted in the context of standard Bayesian networks and then describe and characterize our results in the two upcoming Sections 4 and 5.

3.2 Learning the Parameters

In this section, we assume that the structure of a BN model over the variables $\boldsymbol{X}$ is known, but that the estimates for the conditional probabilities are not known. Hence, the specification of the parameters in the distributions is considered, and the aim is to estimate the parameters of the model: the conditional probabilities.

The first assumption is related to the parameter distribution, called the *Multinomial sample*:

A4. The parameters define a *Multinomial distribution* for each variable X_j and for each configuration of the parents.

When working in a multivariate framework with more than one variable involved, we look at the relationship of the parameters for all the variables in the network. In order to ensure that the parameters can be learned independently we will satisfy two independence properties in order to ensure that all the mathematical properties that enable the computation of the integral (3.3) are met. These properties were introduced by Spiegelhalter and Lauritzen (1990), and later expanded by Heckerman et al. (1995). They are denoted by *global* and *local parameter independence*. The former means that the parameters for the various variables are independent that, in practice, means that

it is possible to modify the tables for the variables independently. The latter means that the parameters are independent for each configuration of the discrete parents. In practice, this means that by having two different configurations, $\mathbf{Pa}_j^1$ and $\mathbf{Pa}_j^2$, the uncertainties on $P\left(X_j|\mathbf{Pa}_j^1\right)$ and on $P\left(X_j|\mathbf{Pa}_j^2\right)$ are independent, and it is possible to modify the parameters for the two distributions independently. If the parameters satisfy the aforementioned independence property, then we have:

A5. *Global parameter independence* : $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}) = \prod_{j=1}^{n} \pi(\boldsymbol{\beta}_j|\mathcal{S})$.

A6. *Local parameter independence* : $\pi(\boldsymbol{\beta}_j|\mathcal{S}) = \prod_{c=1}^{C_j} \pi(\boldsymbol{\beta}_{jc}|\mathcal{S})$, $j = 1, \ldots, n$.

The next assumption is related to the choice of the prior. Specific distributions guarantee a close form expression for the posterior, helping with the computation of (3.3):

A7. The prior distribution of the parameters is a Dirichlet distribution.

Another important assumption is the *parameter modularity*:

A8. If a node X_j has the same parents in two structures $\mathcal{S}_1$ and $\mathcal{S}_2$ ($\mathbf{Pa}_j^{\mathcal{S}_1} = \mathbf{Pa}_j^{\mathcal{S}_2}$), then $P(\boldsymbol{\beta}_{jc}|\mathcal{S}_1) = P(\boldsymbol{\beta}_{jc}|\mathcal{S}_2)$, $c = 1, \ldots, C_j$.

This means that each discrete distribution has the property whereby, if the joint probability distribution $P(\boldsymbol{X})$ can be factorized according to a structure $\mathcal{S}$, it can also be factorized according to all other structures that represent the same set of conditional independencies as $\mathcal{S}$.

These five assumptions (A4 to A8), together with A1 to A3, complete the eight points that allow reducing the computation of the integral (3.3) to a counting problem.

The parameters are learned using the *principle of maximum likelihood*; see Held and Sabanés Bové (2014); Jensen (2001).

4 The suitable additive conjugate prior

In this section, we introduce a novel prior distribution for ABN, that belongs to a flexible family of conjugate priors called the Diaconis-Ylvisaker conjugate priors (Diaconis and Ylvisaker, 1979). We show that the suggested prior is a generalization of the Dirichlet distribution for the additive parameters in an ABN model, similar to that in Massam et al. (2009). Moreover, we prove that this prior satisfies the desirable independence assumptions for a parameter prior in DAG models. Likelihood equivalence, as shown in the supplementary material and important to obtain a score equivalent function, is also satisfied. All these properties help to address the goodness of fit calculation and the posterior parameter estimation.

In our work, we want to choose an appropriate prior for the models in object, in order to avoid the Lindley paradox (Lindley, 1957). Specifically, the principal idea we decided

to follow was to choose a *conjugate prior* for ABN models. Therefore, the computation of the network score should be easier, as a result of the *conjugacy property*, that simplifies the resulting product in equation (3.3). Moreover, the choice of a *conjugate* prior leads to an analytic form solution for the posterior mode β_1^* as shown later, without the necessity of a numerical solution approximation via the Newton method, as has been done to date.

Starting from the Diaconis-Ylvisaker flexible family of conjugate priors Diaconis and Ylvisaker (1979), Chen and Ibrahim (2003) generalized the result to conjugate priors for generalized linear models (GLMs). We started from the suggested prior in Chen and Ibrahim (2003), and adapted this finding to an ABN model. An ABN model consists of a DAG in which each node in the graph comprises a GLM; hence, the conjugacy with regard to a distribution belonging to the exponential family, and in particular to a GLM, was a good compromise.

So far, we are aware of work in which only an improper, informative Gaussian prior for the additive parameters has been considered, see, e.g., Lewis (2012). Therefore, we want to present an alternative to this kind of complex model in order to gain some theoretical results and to improve the computation of the *marginal likelihood*. The following proposition shows how it is possible to generalize the prior Dirichlet density, used in BN model $\mathcal{B}$, with a suitable prior useful for the additive parameters present in an ABN model $\mathcal{A}$. A model that does not impose any restriction on the conditional probabilities is called a *complete* model, which we work to address our aims.

Proposition 1. *For any complete Bayesian network structure $\mathcal{S}$ in a domain $\boldsymbol{X}$, for binary variables, if the prior density $\pi(\boldsymbol{\theta}_{\mathcal{B}})$ is Dirichlet:*

$$\pi(\boldsymbol{\theta}_{\mathcal{B}}) = c \prod_{j=1}^{n} \prod_{c=1}^{C_j} \prod_{s=1}^{2} [\theta_{jcs}]^{\delta_{jcs}-1}. \tag{4.1}$$

then the prior density $\pi(\boldsymbol{\beta}_{\mathcal{A}})$ for a complete additive Bayesian network model $\mathcal{A}$ is a result of the integration of the multinomial logistic regression model $\mathcal{M}$; hence the application of the 'logit transformation' to each parameter of the Bayesian network $\mathcal{B}$: $\theta_{jcs} = \dfrac{e^{\beta_{jcs}}}{1+e^{\beta_{jcs}}} = h^{-1}(\beta_{jcs})$ takes the following expression:

$$\pi(\boldsymbol{\beta}_{\mathcal{A}}) \propto \prod_{j=1}^{n} \prod_{c=1}^{C_j} \left(\frac{e^{\beta_{jc1}}}{1+e^{\beta_{jc1}}}\right)^{\delta_{jc1}} \left(\frac{1}{1+e^{\beta_{jc1}}}\right)^{\delta_{jc2}} = \prod_{j=1}^{n} \prod_{c=1}^{C_j} \frac{\left(e^{\beta_{jc1}}\right)^{\delta_{jc1}}}{(1+e^{\beta_{jc1}})^{\sum_s \delta_{jcs}}}. \tag{4.2}$$

The proof of the proposition can be found in the appendix.

Proposition 2. *Let $\mathcal{A}$ be any complete additive Bayesian network model in a domain $\boldsymbol{X}$. The Jacobian for the transformation from $\boldsymbol{\theta}_{\mathcal{B}}$ to $\boldsymbol{\beta}_{\mathcal{A}}$: $J^{\boldsymbol{\theta}_{\mathcal{B}} \overset{h}{\to} \boldsymbol{\beta}_{\mathcal{A}}}$, is:*

$$J^{\boldsymbol{\theta}_{\mathcal{B}} \overset{h}{\to} \boldsymbol{\beta}_{\mathcal{A}}} = \prod_{j=1}^{n} \prod_{c=1}^{C_j} \theta_{jc1}(1-\theta_{jc1}) = \prod_{j=1}^{n} \prod_{c=1}^{C_j} \left[\frac{h^{-1}(\beta_{jc1})}{d\beta_{jc1}}\right] = \prod_{j=1}^{n} \prod_{c=1}^{C_j} \frac{e^{\beta_{jc1}}}{(1+e^{\beta_{jc1}})^2}. \tag{4.3}$$

The proof of the proposition can be also found in the appendix.

With this proposition, we introduce the 'suitable' conjugate prior and we show how it can be linked to the generalized prior for the additive parameters (4.2) presented previously.

Proposition 3. *The generalized prior found in* (4.2) *can be linked to the Diaconis and Ylvisaker (1979) conjugate priors for GLMs, introduced by Chen and Ibrahim (2003):*

$$\pi(\boldsymbol{\beta}_{\mathcal{A}}) \propto \prod_{j=1}^{n} \exp\left\{ \sum_{i=1}^{m} a_j \left(b_{ij} \boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j - \log(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}) \right) \right\}. \tag{4.4}$$

Proof. Starting with the expression (4.2), we show how it can be reduced to the form (4.4):

$$\begin{aligned} \pi(\boldsymbol{\beta}_{\mathcal{A}}) &\propto \prod_{j=1}^{n} \prod_{c=1}^{C_j} \frac{\left(e^{\beta_{jc1}}\right)^{\delta_{jc1}}}{\left(1 + e^{\beta_{jc1}}\right)^{\sum_s \delta_{jcs}}} \\ &= \prod_{j=1}^{n} \exp\left\{ \sum_s \delta_{jcs} \left(\frac{\delta_{jc1}}{\sum_s \delta_{jcs}} \sum_{c=1}^{C_j} \beta_{jc1} - \prod_{c=1}^{C_j} \log(1 + e^{\beta_{jc1}}) \right) \right\}. \end{aligned}$$

The sum of all the possible parent combinations is nothing else than the sum of all the observations in the design matrix; hence, we get:

$$\begin{aligned} \pi(\boldsymbol{\beta}_{\mathcal{A}}) &\propto \prod_{i=1}^{m} \prod_{j=1}^{n} \exp\left\{ \sum_s \delta_{jcs} \left(\frac{\delta_{jc1}}{\sum_s \delta_{jcs}} \boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j - \log(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}) \right) \right\} \\ &= \prod_{j=1}^{n} \exp\left\{ \sum_{i=1}^{m} \sum_s \delta_{jcs} \left(\frac{\delta_{jc1}}{\sum_s \delta_{jcs}} \boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j - \log(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}) \right) \right\} \qquad (4.5) \\ &= \prod_{j=1}^{n} \exp\left\{ \sum_{i=1}^{m} a_j \left(b_{ij} \boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j - \log(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}) \right) \right\}. \qquad (4.6) \end{aligned}$$

Here, we let a_j and b_{ij} be $a_j = \bigcup_{c=1}^{C_j} \left(\sum_s \delta_{jcs} \right)$ and $b_{ij} = \bigcup_{c=1}^{C_j} \left(\frac{\delta_{jc1}}{\sum_s \delta_{jcs}} \right), \forall i.$ □

This proposition shows how the prior (4.5) can be adapted for the ABN models.

GLM models are a particular class of regression models, in which the outcome variable belongs to a distribution from the exponential family. The prior (4.4) is conjugate for GLM. We then indicate its distribution using $\mathcal{E}$, for the link with the exponential family:

$$\boldsymbol{\beta}_{\mathcal{A}} \sim \mathcal{E}(a_{\mathcal{A}}; \boldsymbol{b}_{\mathcal{A}}). \tag{4.7}$$

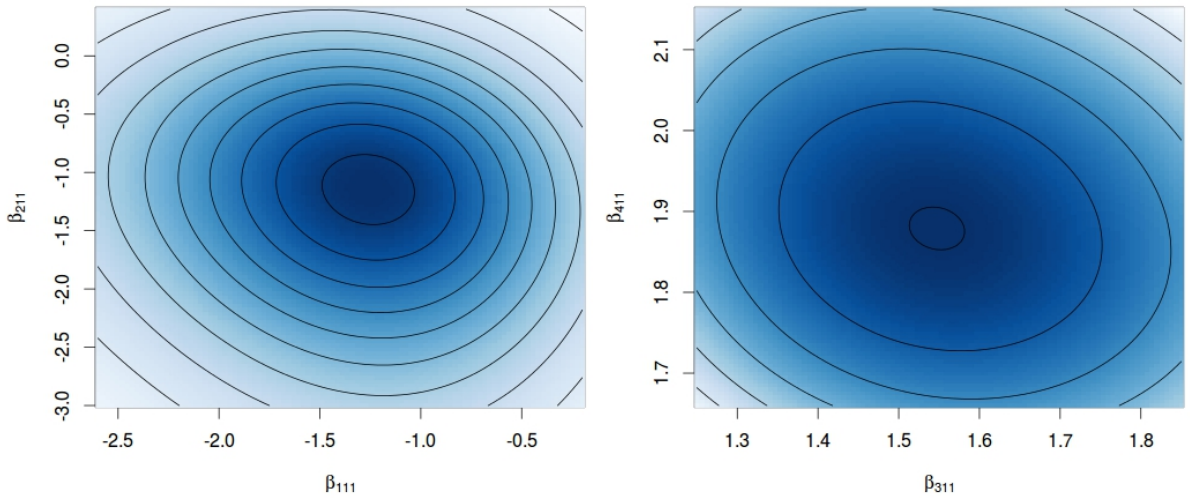


Figure 2: Bivariate representation, based on contour lines, of prior (4.4). The data consist of 434 observations, with 95 positive counts for β_{111}, 56 positive counts for β_{211}, 124 positive counts for β_{311} and 202 positive counts for β_{411}. On the left, the bivariate prior for the variables $(\beta_{111}, \beta_{211})$ is shown for the hyperparameters choice: $(a_{\beta_{111}\beta_{211}} = 1/10, b_{\beta_1\beta_2} = 1/5)$. On the right, the bivariate prior for the variables $(\beta_{311}, \beta_{411})$ is shown, for the hyper-parameters choice: $(a_{\beta_{311}\beta_{411}} = 5/2, b_{\beta_{311}\beta_{411}} = 9/10)$.

$$a_{\mathcal{A}} = \bigcup_{j=1}^{n} a_j; \quad a_j = \bigcup_{c=1}^{C_j} a_{jc}. \tag{4.8}$$

$$\boldsymbol{b}_{\mathcal{A}} = \bigcup_{j=1}^{n} \boldsymbol{b}_j; \quad \boldsymbol{b}_j = (b_{1j}, \ldots, b_{mj})^T; \quad \forall i \; b_{ij} = \bigcup_{c=1}^{C_j} b_{jc}. \tag{4.9}$$

The two prior hyperparameters a_j and b_{ij} have the following meaning:

- $a_j > 0 \; \forall j$ is seen as a prior sample size, that controls the *dispersion*;
- $0 < b_{ij} < 1 \; \forall i, j$ can be seen as the marginal mean of X_j, that affects the prior mode, and plays a role in the *symmetry*.

A bivariate representation of the prior (4.4) can be found in Figure 2, based on a subset of the **ex5.dag.data** data from the R-package *abn* (Lewis et al., 2015).

In order to show the prior's suitability for additive Bayesian network models, we continue to list its important properties.

4.1 Conjugacy and main simplifying assumptions

We proceed with the characterization of our prior, showing that it is conjugate with respect to the multivariate Bernoulli likelihood.

Proposition 4. *The prior* (4.4) *is conjugate with respect to the multivariate Bernoulli likelihood. The posterior distribution takes the following form:*

$$\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{D}) \propto \prod_{j=1}^{n} \exp\left\{ \sum_{i=1}^{m} \{a_j + 1\} \left(\left\{ \frac{a_j b_j + x_{ij}}{a_j + 1} \right\} \boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j - \log(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}) \right) \right\}. \quad (4.10)$$

Specifically, we see that (4.10) *belong to an* $\mathcal{E}$ *distribution:*

$$(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{D}) \sim \mathcal{E}\left(a_{\mathcal{A}|\mathcal{D}}; \boldsymbol{b}_{\mathcal{A}|\mathcal{D}}\right) = \mathcal{E}\left(a_{\mathcal{A}} + 1; \frac{a_{\mathcal{A}} b_{\mathcal{A}} + x_{ij}}{a_{\mathcal{A}} + 1}\right). \quad (4.11)$$

Proof. The result follows from a straightforward multiplication of the multivariate Bernoulli likelihood for node j: $L_j = \prod_{i=1}^{m} \left(\frac{e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{x_{ij}} \left(\frac{1}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{1 - x_{ij}}$, and the prior in (4.4), followed by recognition of the resulting posterior in (4.10). □

This result is a generalization, in a multivariate framework, of the one found in Chen and Ibrahim (2003), Theorem 2.2.

In order to characterize the prior specific to additive Bayesian networks $\mathcal{A}$ properly, we showed that the prior (4.4) satisfies the six main assumptions for standard Bayesian network models $\mathcal{B}$, as presented in Heckerman et al. (1995); Geiger and Heckerman (2002):

A4. *Multinomial sample*

This point indicates that the parameters follow a multinomial distribution. It is fulfilled because the Bernoulli distribution is a special case of the multinomial one. $\boldsymbol{\beta}_{\mathcal{A}}$ is a one-to-one transformation of the *multinomial parameters* $\boldsymbol{\theta}_{\mathcal{B}}$.

Parameter independence

The prior can be factorized across the nodes present in the domain, as well as across all the possible parent combinations. From the previous expressions (4.2) and (4.4), it is possible to see this factorization easily. Hence, we get both:

A5. *Global independence*: $\pi(\boldsymbol{\beta}_{\mathcal{A}}|\mathcal{S}) = \prod_{j=1}^{n} \pi(\boldsymbol{\beta}_j|\mathcal{S})$.

A6 *Local independence*: $j = 1, \ldots, n$, $\pi(\boldsymbol{\beta}_j|\mathcal{S}) = \prod_{c=1}^{C_j} \pi(\boldsymbol{\beta}_{jc}|\mathcal{S})$.

An important consequence of parameter independence is that, for each configuration of the discrete parents, we can update the parameters in the local distributions independently. This also means that if we have *local conjugacy*, i.e., the distributions of β_{jcs} belong to a conjugate family, because of parameter independence, we then have *global conjugacy*, i.e., $P(\boldsymbol{\beta}_{\mathcal{A}})$, the joint distribution of $\boldsymbol{\beta}_{\mathcal{A}}$ belongs

to a conjugate family. With this we can show that *parameter independence* is a *conjugate property*, meaning that, as a result of the *parameter independence* and the *conjugacy property*, we also have a *posterior independence* property.

A7. *Dirichlet*

This assumption is linked to the distribution of the prior, which should belong to a Dirichlet distribution. With the *Prop.* 1, we showed that our prior is a generalization of the Dirichlet distribution. Hence, this property is satisfied.

A8. *Prior modularity*

Given that the two network structures $\mathcal{S}_1$ and $\mathcal{S}_2$ are related to the two models $\mathcal{A}_1$ and $\mathcal{A}_2$ such that both are feasible ($P(\mathcal{S}_1) > 0$ and $P(\mathcal{S}_2) > 0$), if a variable X_j has the same parents in $\mathcal{A}_1$ and $\mathcal{A}_2$, then it also has the same prior, specifically:

$$\pi(\boldsymbol{\beta}_{jc}|\mathcal{S}_1) = \pi(\boldsymbol{\beta}_{jc}|\mathcal{S}_2), \quad c = 1, \ldots, C_j.$$

This last assumption is additional to the 6 previously stated, because it is relevant only if one is interested in *score equivalent* networks.

A9 *Likelihood equivalence*

Given two additive network models $\mathcal{A}_1$ and $\mathcal{A}_2$ such that both are possible ($P(\mathcal{A}_2) > 0$ and $P(\mathcal{A}_2) > 0$), if $\mathcal{A}_1$ and $\mathcal{A}_1$ are equivalent, then they have the same likelihood. This assumption is shown in the supplementary material.

5 Marginal likelihood computation

In the following section, the advantages brought about by the new prior and its property (conjugacy) are described. The issue related to the computation of the *marginal likelihood* and its related easy form of expression will be explained. Starting from the simplest scenario of an orphan node case, all the details are explained in order to arrive at a generalized form for an arbitrary number of parents.

As explained in the previous sections, the main difficulty in moving towards an additive parametrization is the computation of the network score, the *marginal likelihood*. So far, we are only aware of work Lewis (2012) using the Laplace approximation to calculate the *marginal likelihood*.

As a result of the *conjugacy property* of the prior (4.4), obtaining the resulting product inside the integrand of the network score is straightforward. However, the normalizing constant, obtained by the computation of the integral (3.2) is not trivial to obtain. Later, we will exploit different scenarios to achieve the easiest results.

Choosing the prior (4.4), the *marginal likelihood* at each node j becomes:

$$P(\mathcal{D}_j|\mathcal{S}) = \int_{-\infty}^{+\infty} \prod_{i=1}^{m} \left(\frac{e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{x_{ij}} \left(\frac{1}{1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}} \right)^{1 - x_{ij}} \frac{\left(e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j} \right)^{a_j b_{ij}}}{\left(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j} \right)^{a_j}} d\boldsymbol{\beta}_j. \tag{5.1}$$

From the above expression, we have the data set $\mathcal{D}_j$, the parameter vector $\boldsymbol{\beta}_j$, the design matrix Z_{ij} and the related prior hyperparameters a_j and b_{ij} described previously.

5.1 Orphan node case analytic form solution

We start with the simplest case of no parents for the related node j: the 'orphan node' case. For ease of notation, we denote $j = 1$ as the orphan node with which we are dealing. Specifically, we have $x_{1+} = \frac{\sum_{i:X_{1,i}=1} X_{1,i}}{m} \leq 1$ and $\boldsymbol{z}_{i1} = 1$, $b_{i1} = b_1 \ \forall i$. Due to the absence of parents, we have $\boldsymbol{\beta}_1 = \beta_{111}$; therefore, we indicate $\boldsymbol{\beta}_1$ with β_1 to simplify the notation because it is in a univariate case. For this one-dimensional case, the posterior density for the orphan node 1 is:

$$\begin{aligned}
\pi(\boldsymbol{\beta}_1|\mathcal{D}) &\propto \prod_{i=1}^{m} \left(\frac{e^{\boldsymbol{z}_{i1}^T \boldsymbol{\beta}_1}}{1 + e^{\boldsymbol{z}_{i1}^T \boldsymbol{\beta}_1}} \right)^{x_{i1}} \left(\frac{1}{1 + e^{\boldsymbol{z}_{i1}^T \boldsymbol{\beta}_1}} \right)^{1-x_{i1}} \frac{\left(e^{\boldsymbol{z}_{i1}^T \boldsymbol{\beta}_1} \right)^{a_1 b_{i1}}}{\left(1 + e^{\boldsymbol{z}_{i1}^T \boldsymbol{\beta}_1} \right)^{a_1}} \\
&= \prod_{i=1}^{m} \frac{\left(e^{\beta_1} \right)^{a_1 b_1 + x_{i1}}}{\left(1 + e^{\beta_1} \right)^{a_1 + 1}} = \frac{e^{\beta_1 m (a_1 b_1 + x_{1+})}}{\left(1 + e^{\beta_1} \right)^{m(a_1+1)}}. \qquad (5.2) \\
&\Rightarrow \ \boldsymbol{\beta}_1|\mathcal{D} \sim \mathcal{E}_1 \left(a_1 + 1; \frac{a_1 b_1 + x_{i1}}{a_1 + 1} \right).
\end{aligned}$$

Let the posterior hyperparameters be $\alpha_{1,1} := m(a_1 b_1 + x_{1+})$ and $\gamma_{1,1} := m(a_1 + 1)$; hence, the resulting *ABN marginal likelihood* for the orphan node 1, linked to (5.2), is:

$$P(\mathcal{D}_1|\mathcal{S}) \propto \int_{-\infty}^{+\infty} \frac{e^{\beta_1 m (a_1 b_1 + x_{1+})}}{\left(1 + e^{\beta_1} \right)^{m(a_1+1)}} d\beta_1 = \int_{-\infty}^{+\infty} \frac{e^{\beta_1 \cdot \alpha_{1,1}}}{\left(1 + e^{\beta_1} \right)^{\gamma_{1,1}}} d\beta_1. \qquad (5.3)$$

From Abramowitz and Stegun (1992); Gradshteyn and Ryzhik (1965), (5.3) reduces to:

$$P(\mathcal{D}_1|\mathcal{S}) \propto \frac{{}_2F_1(\gamma_{1,1}, \gamma_{1,1} - \alpha_{1,1}, 1 + \gamma_{1,1} - \alpha_{1,1}, -1)}{\gamma_{1,1} - \alpha_{1,1}} + \frac{{}_2F_1(\gamma_{1,1}, \alpha_{1,1}, 1 + \alpha_{1,1}, -1)}{\alpha_{1,1}}. \qquad (5.4)$$

where ${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$, is the *hypergeometric function*.

A requirement for the integral convergence is $\alpha_{1,1} < \gamma_{1,1}$, which holds because $m a_1 b_1 + m x_{1+} < m a_1 + m$, since $0 < b_1 < 1$ and $x_{1+} \leq 1$.

5.2 Orphan node case Laplace approximation

In the literature by Bernardo and Smith (2000); Sabanés Bové and Held (2011); Lewis (2012); Held and Sabanés Bové (2014), the integral (5.3) has been computed using the Laplace approximation method. Therefore, the results of the integral (5.4) are also calculated and compared using this technique.

We define $h(\boldsymbol{\beta}_1) = h(\beta_1) = -\frac{1}{m}\{\alpha_{1,1} \cdot \beta_1 - \gamma_{1,1} \cdot \log(1 + e^{\beta_1})\} = -\frac{1}{m}\log\{L(\beta_1) \cdot \pi(\beta_1)\} = -\frac{1}{m}\{l(\beta_1) + \log\{\pi(\beta_1)\}\}$, where $l(\beta_1) = \log L(\beta_1)$ is the log-likelihood function. We then have:

$$\tilde{P}(\mathcal{D}_1|\mathcal{S}) \propto \int_{-\infty}^{+\infty} \exp(-m \cdot h(\beta_1))d\beta_1 = \exp(-m \cdot h(\beta_1^*))\sqrt{\frac{2\pi}{mk}}, \tag{5.5}$$

where β_1^* is the maximum of $h(\beta)$ and $k = \dfrac{d^2h(\beta_1^*)}{d\beta^2}$ denotes the curvature of $h(\beta)$. Using the previous definition of the function $h(\beta_1)$, we obtain the following results based on the Laplace approximation method:

$$\frac{dh(\beta_1)}{d\beta_1} = -\frac{1}{m}\Big(\alpha_{1,1} - \gamma_{1,1}\frac{e^{\beta_1}}{1 + e^{\beta_1}}\Big); \quad \frac{dh(\beta_1^*)}{d\beta_1} = 0 \Rightarrow \beta_1^* = \log\Big(\frac{\alpha_{1,1}}{\gamma_{1,1} - \alpha_{1,1}}\Big). \tag{5.6}$$

$$k^{-1} = \Big(\frac{d^2h(\beta_1^*)}{d\beta_1^2}\Big)^{-1} = \frac{m}{\gamma_{1,1}}\frac{(1 + e^{\beta_1^*})^2}{e^{\beta_1^*}} = \frac{1}{1 + a_1}\frac{(1 + e^{\beta_1^*})^2}{e^{\beta_1^*}} = (1 + a_1)^{-1}\theta_{111}^{-1}\theta_{112}^{-1}. \tag{5.7}$$

Therefore (5.5), based on the results (5.6), (5.7), becomes:

$$\tilde{P}(\mathcal{D}_1|\mathcal{S}) = \exp\Big(-m \cdot h(\beta_1^*)\Big)\sqrt{\frac{2\pi}{mk}} \propto \frac{\alpha_{1,1}^{\alpha_{1,1}-1/2}}{\gamma_{1,1}^{\gamma_{1,1}-1/2}}(\gamma_{1,1} - \alpha_{1,1})^{\gamma_{1,1}-\alpha_{1,1}-1/2}. \tag{5.8}$$

The choice of a *conjugate* prior leads to an analytic form solution for the posterior mode β_1^* as (5.6), without the requirement for a numerical solution approximation via the Newton method, as has been done to date.

Generally speaking, the posterior density is determined as $\pi(\boldsymbol{\beta}|\mathcal{D}) \propto e^{l(\boldsymbol{\beta})}\pi(\boldsymbol{\beta})$. Letting $\bar{l}_m(\boldsymbol{\beta}) = \frac{l(\boldsymbol{\beta})}{m} = \dfrac{1}{m}\sum_1^m \log f(X_i|\boldsymbol{\beta})$, the law of large numbers yields that for $m \to \infty$, $\bar{l}_m(\boldsymbol{\beta}) = \mathbf{E}_{\boldsymbol{\beta}}\{\log f(X|\boldsymbol{\beta})\} = -H(\boldsymbol{\beta})$, $H(\boldsymbol{\beta})$ is the entropy of the density $f(\cdot|\boldsymbol{\beta})$. Thus, the variation in the posterior density $\pi(\boldsymbol{\beta}|\mathcal{D})$ will be dominated by the contribution from the likelihood function for sufficiently large m. Expanding $l(\boldsymbol{\beta})$ around the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ yields:

$$\pi(\boldsymbol{\beta}|\mathcal{D}) \propto e^{m \cdot \bar{l}_m(\hat{\boldsymbol{\beta}})}\pi(\hat{\boldsymbol{\beta}})e^{-(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T I_m(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})/2} \propto e^{-(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T I_m(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})/2}, \tag{5.9}$$

where $I_m(\hat{\boldsymbol{\beta}}) = m \cdot I(\hat{\boldsymbol{\beta}})$ is the observed Fisher information matrix, so for large m, the posterior distribution of $\boldsymbol{\beta}|\mathcal{D}$ is approximately:

$$\boldsymbol{\beta}|\mathcal{D} \sim \mathcal{N}_n(\hat{\boldsymbol{\beta}}, I_m(\hat{\boldsymbol{\beta}})^{-1}) = \mathcal{N}_n(\hat{\boldsymbol{\beta}}, I(\hat{\boldsymbol{\beta}})^{-1}/m). \tag{5.10}$$

A more accurate approximation is obtained by expanding around the posterior mode $\boldsymbol{\beta}^*$, as we did when looking for the maximum of the function $h(\boldsymbol{\beta})$, yielding approximately:

$$\boldsymbol{\beta}|\mathcal{D} \sim \mathcal{N}_n(\boldsymbol{\beta}^*, I_m(\boldsymbol{\beta}^*)^{-1}) = \mathcal{N}_n(\boldsymbol{\beta}^*, I(\boldsymbol{\beta}^*)^{-1}/m), \tag{5.11}$$

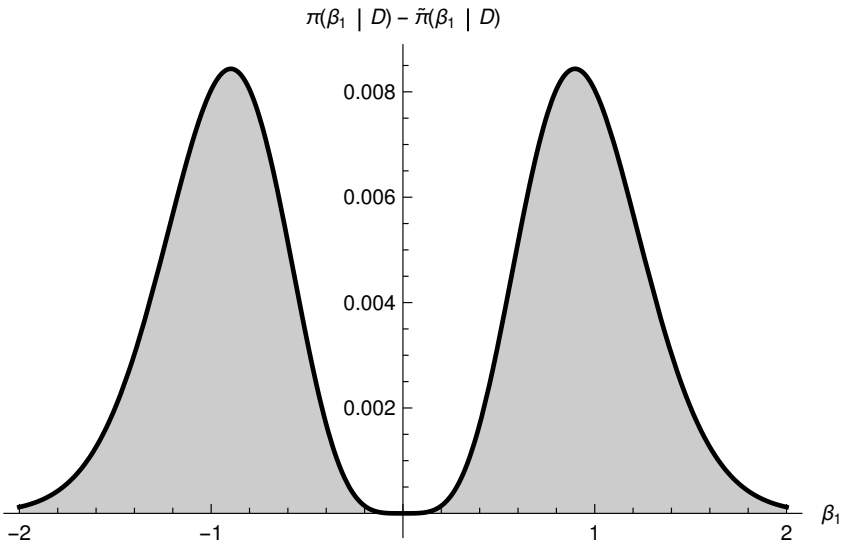


Figure 3: Difference between $\pi(\boldsymbol{\beta}_1|\mathcal{D})$ and $\tilde{\pi}(\boldsymbol{\beta}_1|\mathcal{D})$, for $\alpha_{1,1} = 10$ and $\gamma_{1,1} = 20$, corresponding to data values $m = 10$ and $x_{1+} = 5$, with hyperparameters $a_1 = 1$ and $b_1 = \frac{1}{2}$.

for large m.

Hence, from (5.11), we can see that, generally speaking, the posterior $\pi(\boldsymbol{\beta}|\mathcal{D})$ can be approximated, for sufficiently large m, to a Gaussian distribution with a mean of $\boldsymbol{\beta}^*$ and a covariance matrix of $I_m(\boldsymbol{\beta}^*)^{-1}$. This result is a generalization for the posterior of a theorem present in Chen and Ibrahim (2003). Thus, we have that (5.2), for sufficiently large m, using the notation (5.7), becomes:

$$\beta_1|\mathcal{D} \sim \mathcal{N}(\beta_1^*, (m \cdot k)^{-1}) = \mathcal{N}\Big(\log\Big[\Big(\frac{\gamma_{1,1}}{\alpha_{1,1}} - 1\Big)^{-1}\Big], \alpha_{1,1}^{-1} \cdot \Big(1 - \frac{\alpha_{1,1}}{\gamma_{1,1}}\Big)^{-1}\Big). \tag{5.12}$$

The approximation error of the Laplace method is in the order of $\mathcal{O}(m^{-1})$. However, we can check this result using further graphical controls.

In order to evaluate how good the approximation is graphically, we plot the difference between the 'original' posterior (5.2) and the 'approximated' one (5.12), as in Figure 3. From Figure 3, we can see that the difference is null at the posterior mode β_1^* point around that the second-order Taylor expansion has been developed, while the difference increases within roughly two standard deviations.

Moreover, we plot the committed error, indicated via $\mathsf{E}_1(\alpha_{1,1}, \gamma_{1,1})$, as the difference between the 'exact' (5.4) and the 'approximated' (5.8) value of the marginal likelihood.

The resulting plot, on a $\log_{10}$ scale of $\mathsf{E}_1(\alpha_{1,1}, \gamma_{1,1})$ for $\alpha_{1,1} \in (0, 50)$ and $\gamma_{1,1} \in (1, 50)$ can be found in Figure 4. From Figure 4, we can see that the error $\mathsf{E}_1(\alpha_{1,1}, \gamma_{1,1})$ reaches the smallest value when $\alpha_{1,1} = 4/5 \cdot \gamma_{1,1}$; while $\mathsf{E}_1(\alpha_{1,1}, \gamma_{1,1})$ becomes larger when $\alpha_{1,1} = 0$ or $\alpha_{1,1} = \gamma_{1,1}$, corresponding to the limits of the feasibility condition for the convergence of the integral (5.3).

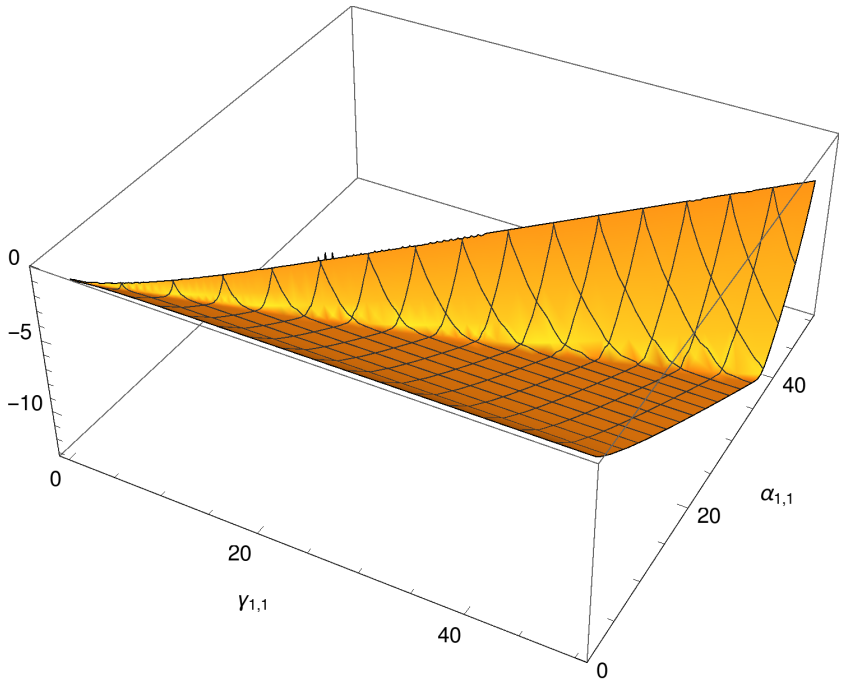


Figure 4: $\log_{10}(\mathsf{E}_1(\alpha_{1,1}, \gamma_{1,1}))$ for $\alpha_{1,1} \in (0, 50)$ and $\gamma_{1,1} \in (1, 50)$.

5.3 Generalized node case, analytic and approximated results

The results obtained for an orphan node case can be generalized for all the other possible combinations of parent cases for an ABN model.

Let us consider the resulting *ABN marginal likelihood* for a general node j from (4.11), (4.10) and (5.4); thus, we get:

$$P(\mathcal{D}_j|\mathcal{S}) = \int_{-\infty}^{+\infty} \prod_{i=1}^{m} \frac{\left(e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}\right)^{a_j b_j + x_{ij}}}{\left(1 + e^{\boldsymbol{z}_{ij}^T \boldsymbol{\beta}_j}\right)^{a_j+1}} d\boldsymbol{\beta}_j = \int_{-\infty}^{+\infty} \prod_{c=1}^{C_j} \frac{e^{\beta_{jc1} \cdot \alpha_{j,c}}}{(1 + e^{\beta_{jc1}})^{\gamma_{j,c}}} d\beta_{jc1} \tag{5.13}$$

$$= \prod_{c=1}^{C_j} \left[\frac{{}_2F_1(\gamma_{j,c}, \gamma_{j,c} - \alpha_{j,c}, 1 + \gamma_{j,c} - \alpha_{j,c}, -1)}{\gamma_{j,c} - \alpha_{j,c}} + \frac{{}_2F_1(\gamma_{j,c}, \alpha_{j,c}, 1 + \alpha_{j,c}, -1)}{\alpha_{j,c}} \right]. \tag{5.14}$$

The main requirement for the integral convergence is $\alpha_{j,c} < \gamma_{j,c} \quad \forall j, c$, which always holds due to the data structure. In fact, $\alpha_{j,c} = Pa_{jc+} \cdot a_j \cdot b_j + X_{jc+}$ and $\gamma_{j,c} = Pa_{jc+} \cdot (a_j + 1)$, where Pa_{jc+} represents the sum of all the parent configurations c for node j and X_{jc+} indicates the sum of all $X_j = 1$ that corresponds to the parent configurations c, because $X_{jc+} \leq Pa_{jc+}$ and $0 < b_j < 1 \Rightarrow \alpha_{j,c} < \gamma_{j,c} \quad \forall j, c$. Moreover, the results from the Laplace approximation can also be generalized for a general node j:

$$\boldsymbol{\beta}_j^* = \beta_{jc1}^* = \log\left(\frac{\alpha_{j,c}}{\gamma_{j,c} - \alpha_{j,c}}\right). \tag{5.15}$$

$$k_j^{-1} = m^{C_j} \prod_{c=1}^{C_j} \frac{1}{\gamma_{j,c}} \frac{(1+e^{\beta^*_{jc1}})^2}{e^{\beta^*_{jc1}}}. \tag{5.16}$$

If we consider node $j = 3$ in Figure 1 which has two parents ($P_j = 2$), and we express the additive transformed parameters β_{jcs} in terms of the parameters $\phi_{j,p-1}$, linked to the marginal effect of each covariate X_p, we get:

$$\beta_{jcs} = \sum_{p=1}^{c} \Big[\phi_{j,p-1} + \sum_{p=2,\ p/X_{p-1} \in \mathbf{Pa}_j}^{\min(c,P_j+1)} \phi_{j,p-1} X_{p-1} + \sum_{p=P_j+2}^{c} \phi_{j,p-1} \Big]. \tag{5.17}$$

$$P(\mathcal{D}_3|\mathcal{S}) = \int_{-\infty}^{\infty} \frac{e^{\beta_{311}\cdot\alpha_{3,1}}\ e^{\beta_{321}\cdot\alpha_{3,2}}\ e^{\beta_{331}\cdot\alpha_{3,3}}\ e^{\beta_{341}\cdot\alpha_{3,4}}}{(1+e^{\beta_{311}})^{\gamma_{3,1}}\ (1+e^{\beta_{321}})^{\gamma_{3,2}}\ (1+e^{\beta_{331}})^{\gamma_{3,3}}\ (1+e^{\beta_{341}})^{\gamma_{3,4}}} d\boldsymbol{\beta}_3 =$$

$$\int_{-\infty}^{\infty} \frac{e^{\phi_{3,0}\cdot\alpha_{3,1}}\ e^{(\phi_{3,0}+\phi_{3,1})\cdot\alpha_{3,2}}\ e^{(\phi_{3,0}+\phi_{3,2})\cdot\alpha_{3,3}}\ e^{(\phi_{3,0}+\phi_{3,1}+\phi_{3,2}+\phi_{3,3})\cdot\alpha_{3,4}}}{(1+e^{\phi_{3,0}})^{\gamma_{3,1}}(1+e^{\phi_{3,0}+\phi_{3,1}})^{\gamma_{3,2}}(1+e^{\phi_{3,0}+\phi_{3,2}})^{\gamma_{3,3}}(1+e^{\phi_{3,0}+\phi_{3,1}+\phi_{3,2}+\phi_{3,3}})^{\gamma_{3,4}}} d\boldsymbol{\beta}_3 =$$

$$\int_{-\infty}^{\infty} \frac{e^{\phi_{3,0}\cdot\lambda_{3,0}}\ e^{\phi_{3,1}\cdot\lambda_{3,1}}\ e^{\phi_{3,2}\cdot\lambda_{3,2}}\ e^{\phi_{3,3}\cdot\lambda_{3,3}}}{(1+e^{\phi_{3,0}})^{\omega_{3,0}}(1+e^{\phi_{3,0}+\phi_{3,1}})^{\omega_{3,1}}(1+e^{\phi_{3,0}+\phi_{3,2}})^{\omega_{3,2}}(1+e^{\phi_{3,0}+\phi_{3,1}+\phi_{3,2}+\phi_{3,3}})^{\omega_{3,3}}} d\boldsymbol{\beta}_3.$$

It is then straightforward to see the relationship of $\lambda_{j,c-1}$ and $\omega_{j,c-1}$ coefficients for the 'marginal' parameters $\phi_{j,c-1}$ with the one between $\alpha_{j,c}$ and $\gamma_{j,c}$ coefficients for the additive transformed parameters β_{jcs}. Therefore, we get:

$$\begin{cases} \lambda_{3,0} = \alpha_{3,1}+\alpha_{3,2}+\alpha_{3,3}+\alpha_{3,4}, \\ \lambda_{3,1} = \alpha_{3,2}+\alpha_{3,4}, \\ \lambda_{3,2} = \alpha_{3,3}+\alpha_{3,4}, \\ \lambda_{3,3} = \alpha_{3,4}. \end{cases} \Rightarrow \begin{cases} \alpha_{3,1} = \lambda_{3,0}-\lambda_{3,1}-\lambda_{3,2}+\lambda_{3,3}, \\ \alpha_{3,2} = \lambda_{3,1}-\lambda_{3,3}, \\ \alpha_{3,3} = \lambda_{3,2}-\lambda_{3,3}, \\ \alpha_{3,4} = \lambda_{3,3}. \end{cases} \tag{5.18}$$

As a result of the expression for the $\phi_{j,c-1}$ in terms of β_{jcs} and the related equation for the coefficients, we can derive the results (5.15) and (5.16) in terms of $\phi_{j,c-1}$ in a straightforward way by applying the equations (5.18).

5.4 Comparison with previous parameter priors for ABN models

A possible question that can arise is why a new choice of parameter prior is needed and is useful for ABN models. Some suggestions have already been proposed and implemented in R (R Core Development Team, 2015), in the R-package *abn* (Lewis et al., 2015) for the additive Bayesian network models. The main issue is that the prior selected in the R-package *abn* is an uninformative Gaussian prior with a zero mean and a large variance for each of the regression parameters across all parts of the model. Unfortunately, this specification of the parameter prior leads to some problems when cases of complete data separation occur. Moreover, the choice of a prior that is too simple and uninformative may not give appropriate results, as stated by Lindley's paradox Lindley (1957).

In order to demonstrate the aforementioned issue, we compared the posterior densities resulting from the prior chosen in the R-package *abn* to the conjugate prior introduced previously. We used a subset of the **ex5.dag.data** data set present in the R-package *abn*, from that we selected the binary variables $\{\mathbf{b1}, \mathbf{b2}, \mathbf{b3}, \mathbf{b4}, \mathbf{b5}, \mathbf{b6}\}$. Without conducting a model search, we focused on the parameter-learning step, starting with a given additive Bayesian network structure, $\mathcal{A}_E$, as illustrated in Figure 5, followed by checking and drawing the related posterior density.

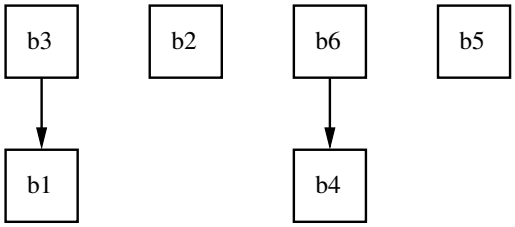


Figure 5: Example of ABN model $\mathcal{A}_E$ with data set **ex5.dag.data** from R-package *abn*

$b1 \backslash b3$	0	**1**
0	308	124
1	2	**0**

$b4 \backslash b6$	0	**1**
0	348	56
1	30	**0**

Table 1: Summary statistics for the data couple combination $(b1, b3)$ left and $(b4, b6)$ right.

The explanatory model is linked to the data structure in which two cases of complete data separation occur, as can be seen in Table 1. We then fitted the model $\mathcal{A}_E$ to the data and plotted the resulting posterior density. In Figure 6, it is straightforward to see that the choice of an uninformative Gaussian prior with a zero mean and a large variance, as used in the R-package *abn*, leads to strange features (right skewness and unexpected tail behaviour around the 0), while this does not happen for posterior density resulting from the prior conjugate choice.

Comparing the two densities, we can see how the posterior densities resulting from the prior conjugate choice (4.4) reach their peaks at the MLE estimates, while this does not occur for the posterior densities resulting from the R-package *abn*. Moreover, due to the complete case separation situation, the tails and the skewness of the posterior densities from the R-package *abn* are a signal of the main problem arising from this choice of prior. The *abn* posterior densities do not even integrate to one, a necessary requirement for a function in order to be a density.

A plausible justification for the improvement of the posterior density when using our suggested conjugate prior is connected to the prior's dependency on its parent variables. The introduced parameter prior is highly dependent on the design matrix which, in terms of Bayesian network models, is represented by the influence of the parents variables $\mathbf{Pa}_j$. From the summary statistics in Table 1, it is easy to observe

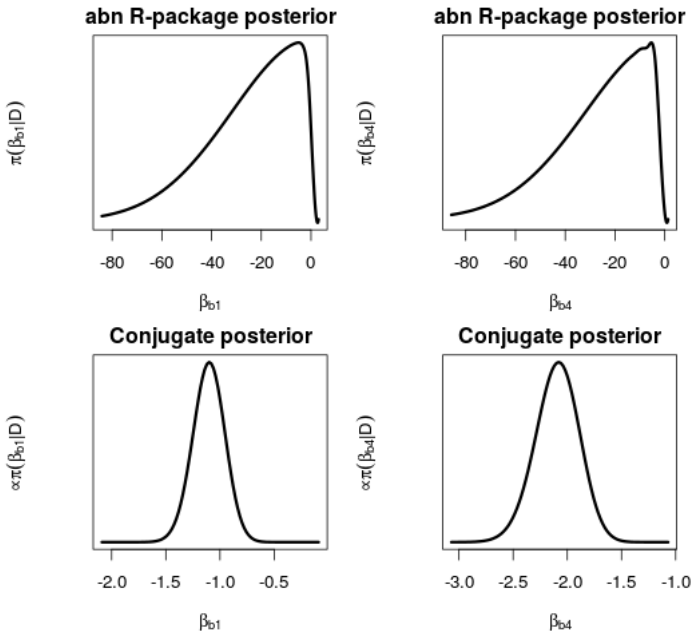


Figure 6: Comparison between the posteriors of ABN models using the R-package *abn* and the conjugate prior (4.4). On the left, the marginal posterior density for node $b1$ is shown with hyperparameters $(a_{b1}, b_{b1}) = (5/4, 1/5)$, while on the right, the marginal posterior density for node $b4$ is shown with hyperparameters $(a_{b4}, b_{b4}) = (7/2, 1/7)$.

that $Pa_{b12+} = \sum_{i=1}^{m} b3_i = 124$ and, similarly, $Pa_{b62+} = \sum_{i=1}^{m} b6_i = 56$. As indicated in (5.13), the role of the parents Pa_{jc+} for the coefficients $\alpha_{j,c}$ and $\gamma_{j,c}$ specification plays an important part. Hence, it helps to improve and resolve the situation of complete data separation.

Moreover, the introduced conjugate prior (4.4) is linked to the classical Zellner's g-prior (Zellner, 1986), as shown in (cf. Sabanés Bové and Held, 2011, Section 2.1, formula (5)), proposed for the regression coefficients as a 'reference informative prior'. Hence, as a result of his 'informative' property and connection with the g-prior, the dependence on the design matrix and parent variables is more evident, which leads to a better posterior density quality.

Also if we use *abn* implementation without an uninformative prior, we will encounter issues related to the *marginal likelihood* computation, because numerical and approximated methods are used in the R-package in order to compute the network score.

6 Discussion and conclusions

In this work, a conjugate prior for additive Bayesian networks models has been proposed. The consequences of this choice have been analysed, starting from the resulting posterior distribution in order to arrive at the computation of the marginal likelihood. Crucial assumptions of the prior have been examined, and their fulfilment has been

demonstrated. An easy expression represented by the ${}_2F_1$ *hypergeometric function* has been found for the *ABN marginal likelihood*. These achievements were possible because of the appropriate choice of the parameter prior, which has led to easier computations. This result has been compared to the usual Laplace approximation approach in order to check the differences between the two results. Recently published research, Nadarajah (2015), has shown that the ${}_2F_1$ *hypergeometric function* can be further simplified via a Truncated Beta Function, implying that an even simpler result is still possible. Moreover, the assumption of *likelihood equivalence* and the possibility of having a *score equivalent network* for *equivalent classes* have been explored and developed.

The main contribution of our work is linked to opening a door to the additive Bayesian network literature that compares these modern models in terms of Bayesian model selection (i.e. the specification of parameter priors and the computation of the resulting posterior model probabilities via the marginal likelihoods) with the classical Bayesian network models that have been developed for a longer period. This is also the first attempt to create a link between the application and the theory. So far, the additive Bayesian networks have mainly been used in an applied statistics framework, such as the veterinary epidemiologist field, without going into too much detail concerning the underlying theory. Mainly due to the meticulous notation, this is the first time that a fully described parametrization has been elucidated and a simplification of the *ABN marginal likelihood* has been found.

In this framework, further developments can be done to improve additive Bayesian networks for both discrete (i.e. multinomial, Poisson) and continuous (i.e. Gaussian) distributions. Another improvement is in terms of application, as a possible implementation of the newly introduced parameter prior in appropriate software dealing with additive Bayesian networks, such as the R package *abn*, can be a good starting point to promote and use the newly introduced methodology. It would represent nothing less than the continuation of the connection between theory and practice, albeit with a newly developed method.

Appendix A: Proofs

Proof of Proposition 1. Let $\mathcal{S}$ be any complete additive Bayesian network structure in a domain $\boldsymbol{X}$. We use the change of variable formula:

$$\pi(\boldsymbol{\beta}_{\mathcal{A}}) = \pi_{\boldsymbol{\theta}_{\mathcal{B}}}(h^{-1}(\boldsymbol{\beta}_{\mathcal{A}})) \cdot J^{\boldsymbol{\theta}_{\mathcal{B}} \overset{h}{\to} \boldsymbol{\beta}_{\mathcal{A}}}. \tag{A.1}$$

The Jacobian of interest, $J^{\boldsymbol{\theta}_{\mathcal{B}} \overset{h}{\to} \boldsymbol{\beta}_{\mathcal{A}}}$, as given by Proposition 2.

$$\begin{aligned} \pi_{\boldsymbol{\theta}_{\mathcal{B}}}(h^{-1}(\boldsymbol{\beta}_{\mathcal{A}})) &= \prod_{j=1}^{n} \prod_{c=1}^{C_j} \left(\frac{e^{\beta_{jc1}}}{1+e^{\beta_{jc1}}} \right)^{\delta_{jc1}-1} \left(\frac{1}{1+e^{\beta_{jc1}}} \right)^{\delta_{jc2}-1} \\ &= \prod_{j=1}^{n} \prod_{c=1}^{C_j} \prod_{s=1}^{S_j} \left(\frac{e^{\beta_{jc1}}}{1+e^{\beta_{jc1}}} \right)^{\delta_{jcs}-1} \left(\frac{1}{1+e^{\beta_{jc1}}} \right)^{\delta_{jcs}-1}. \end{aligned} \tag{A.2}$$

Multiplying the equation (A.2) by the Jacobian given by Proposition 2, and adjusting and collecting the respective hyperparameters δ_{jcs} of the Dirichlet density appropriately, we obtain equation (4.2). □

Proof of Proposition 2. We proceed by induction using J_n to denote the Jacobian $J^{\boldsymbol{\theta}_{\mathcal{B}} \overset{h}{\to} \boldsymbol{\beta}_{\mathcal{A}}}$ for the n-variable case. For the following proof, a special notation $\beta_{j|\mathbf{Pa}_j}$ is used, equivalent to β_{jc}, in order to place more emphasis on the related parents configurations. For $n = 1$, the orphan node case, $\mathbf{Pa}_1 = \varnothing$, the theorem holds trivially, because:

$$J^{\boldsymbol{\theta}_1 \overset{h}{\to} \boldsymbol{\beta}_1} = \frac{e^{\beta_{111}}}{(1+e^{\beta_{111}})^2} = \frac{e^{\beta_{1|\mathbf{Pa}_1}}}{(1+e^{\beta_{1|\mathbf{Pa}_1}})^2}.$$

For the induction step, let us assume that the complete Bayesian network structure has variable ordering $X_1, \ldots, X_{n+1}$.

First, a change of variables from $\boldsymbol{\beta}_{1,\ldots,n+1}$ to $\boldsymbol{\beta}_{1,\ldots,n} \cup \boldsymbol{\beta}_{n+1|1,\ldots,n}$ is performed. By definition, $\beta_{1,\ldots,n+1} = \beta_{1,\ldots,n} \cdot \beta_{n+1|1,\ldots,n}$. Thus, the Jacobian matrix consists of $(n+1)$ block matrices. There is one block for each of the j variables in the domain. The block matrix dimension depends on the number of possible parents configurations for each node j; each matrix then has dimension C_j.

Each block consists of a lower triangular matrix. The $(n+1)^{\text{th}}$ block has the following form of size C_{n+1}:

$$\begin{pmatrix} \frac{e^{\beta_{n+111}}}{\left(1+e^{\beta_{n+111}}\right)^2} & & & \\ \frac{e^{\beta_{n+111}}}{\left(1+e^{\beta_{n+111}}\right)^2} & \frac{e^{\beta_{n+121}}}{\left(1+e^{\beta_{n+121}}\right)^2} & & \\ \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \vdots \\ \frac{e^{\beta_{n+111}}}{\left(1+e^{\beta_{n+111}}\right)^2} & \frac{e^{\beta_{n+121}}}{\left(1+e^{\beta_{n+121}}\right)^2} & \cdots & \frac{e^{\beta_{n+1(C_{n+1}-1)1}}}{\left(1+e^{\beta_{n+1(C_{n+1}-1)1}}\right)^2} \quad \frac{e^{\beta_{n+1C_{n+1}1}}}{\left(1+e^{\beta_{n+1C_{n+1}1}}\right)^2} \end{pmatrix}.$$

The determinant of this block matrix is the Jacobian (J^B_{n+1}) given below:

$$\begin{aligned} J^B_{n+1} &= \frac{e^{\beta_{n+111}}}{\left(1+e^{\beta_{n+111}}\right)^2} \frac{e^{\beta_{n+121}}}{\left(1+e^{\beta_{n+121}}\right)^2} \cdots \frac{e^{\beta_{n+1(C_{n+1}-1)1}}}{\left(1+e^{\beta_{n+1(C_{n+1}-1)1}}\right)^2} \frac{e^{\beta_{n+1C_{n+1}1}}}{\left(1+e^{\beta_{n+1C_{n+1}1}}\right)^2} \\ &= \prod_{c=1}^{C_{n+1}} \frac{e^{\beta_{n+1c1}}}{\left(1+e^{\beta_{n+1c1}}\right)^2} = \prod_{c=1}^{C_{n+1}} \frac{e^{\beta_{n+1|\mathbf{Pa}_{n+1}}}}{\left(1+e^{\beta_{n+1|\mathbf{Pa}_{n+1}}}\right)^2}. \end{aligned} \tag{A.3}$$

Next, we change variables from $\boldsymbol{\beta}_{1,\ldots,n}$ to $\boldsymbol{\beta}_1 \cup \boldsymbol{\beta}_{2|1} \cup \cdots \cup \boldsymbol{\beta}_{n|1,\ldots,n-1}$, that is equivalent to $\boldsymbol{\beta}_{1|\mathbf{Pa}_1} \cup \boldsymbol{\beta}_{2|\mathbf{Pa}_2} \cup \cdots \cup \boldsymbol{\beta}_{n|\mathbf{Pa}_n}$. For simplicity, we have $\mathbf{Pa}_1 = \varnothing, \mathbf{Pa}_2 = 1, \ldots, \mathbf{Pa}_n = \{1, \ldots, n-1\}$. Using the Jacobian J_n obtained from the induction hypothesis, the

combined Jacobian for the transformation from $\boldsymbol{\beta}_{1,\ldots,n+1}$ to $\boldsymbol{\beta}_1 \cup \cdots \cup \boldsymbol{\beta}_{n+1|1,\ldots,n}$ is $J_{n+1} = J^B_{n+1} J_n$. Consequently, we have:

$$
\begin{aligned}
J_{n+1} &= \prod_{c=1}^{C_{n+1}} \frac{e^{\beta_{n+1c1}}}{(1+e^{\beta_{n+1c1}})^2} \cdot \prod_{j=1}^{n} \prod_{c=1}^{C_j} \frac{e^{\beta_{jc1}}}{(1+e^{\beta_{jc1}})^2} \\
&= \prod_{c=1}^{C_{n+1}} \frac{e^{\beta_{n+1|\mathbf{Pa}_{n+1}}}}{(1+e^{\beta_{n+1|\mathbf{Pa}_{n+1}}})^2} \cdot \prod_{j=1}^{n} \prod_{c=1}^{C_j} \frac{e^{\beta_{j|\mathbf{Pa}_j}}}{(1+e^{\beta_{j|\mathbf{Pa_j}}})^2} = \prod_{j=1}^{n+1} \prod_{c=1}^{C_j} \frac{e^{\beta_{j|\mathbf{Pa}_j}}}{(1+e^{\beta_{j|\mathbf{Pa}_j}})^2} \\
&= \prod_{j=1}^{n+1} \prod_{c=1}^{C_j} \frac{e^{\beta_{jc1}}}{(1+e^{\beta_{jc1}})^2} = \prod_{j=1}^{n} \prod_{c=1}^{C_j} \left[\frac{h^{-1}(\beta_{jc1})}{d\beta_{jc1}} \right] = \prod_{j=1}^{n} \prod_{c=1}^{C_j} \theta_{jc1}(1-\theta_{jc1}).
\end{aligned}
$$

□

Supplementary Material

Likelihood and score equivalence (http://www.some-url-address.org/dowload/0000.zip). The supplementary material contains all the proofs for and descriptions of the likelihood equivalence of a multivariate Bernoulli likelihood using an additive parametrization. Moreover, it contains the score equivalence proof, together with a suitable choice for the hyperparameters, for two equivalent additive Bayesian networks.

References

Abramowitz, M. and Stegun, I. A. (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, Inc., New York. 17

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian theory*. Wiley Series In Probability. Chichester: John Wiley & Sons, Inc., 389 edition. 2, 8, 17

Boettcher, S. G. (2004). "Learning Bayesian networks with mixed variables." Ph.D. thesis, Aalborg University - Department of Mathematical Sciences. 3, 6, 10

Buntine, W. (1991). "Theory refinement on Bayesian networks." In DAmbrosio, B., Smets, P., and Bonissone, P. (eds.), *Uncertainty in artificial intelligence: proceedings of the seventh conference*, 52–60. Morgan Kaufmann Publishers Inc. 6

Chen, M. and Ibrahim, J. (2003). "Conjugate priors for generalized linear models." *Statistica Sinica*, 13(389, 391): 461–476. 12, 13, 15, 19

Chickering, D. M. (1996). "Learning Bayesian networks is NP-complete." In Fisher, D. and Lenz, H.-J. (eds.), *Learning from Data*, volume 112 of *Lecture Notes in Statistics*, 121–130. Springer New York. 7

Chickering, D. M., Heckerman, D., and Meek, C. (2004). "Large-sample learning of Bayesian networks is NP-hard." *Journal of Machine Learning Research*, 5: 1287–1330. 7

Chow, C. and Liu, C. (1968). "Approximating discrete probability distributions with dependence trees." *Ieee Transactions On Information Theory*, 14(3): 462–467. 7

Cooper, G. F. and Herskovits, E. (1992). "A Bayesian method for the induction of probabilistic networks from data." *Machine Learning*, 9(4): 309–347. 3, 9, 10

Cowell, R. G. (2001). "Conditions under which conditional independence and scoring methods lead to identical selection of Bayesian network models." In Breese, I. J. and Koller, D. (eds.), *Proceedings of the Seventeenth International Conference on Uncertainty in Artificial Intelligence*, 91–97. Morgan Kaufmann. 7

Dai, B., Ding, S., and Wahba, G. (2013). "Multivariate Bernoulli distribution." *Bernoulli*, 19(4): 1465–1483. 6

Diaconis, P. and Ylvisaker, D. (1979). "Conjugate priors for exponential families." *Annals of Statistics*, 7(2): 269–281. 2, 11, 12, 13

Djebbari, A. and Quackenbush, J. (2008). "Seeded Bayesian Networks: Constructing genetic networks from microarray data." *Bmc Systems Biology*, 2: 57. 2

Dojer, N., Gambin, A., Mizera, A., Wilczynski, B., and Tiuryn, J. (2006). "Applying dynamic Bayesian networks to perturbed gene expression data." *Bmc Bioinformatics*, 7: 249. 1

Firestone, S. M., Lewis, F. I., Schemann, K., Ward, M. P., Toribio, J. A., and Dhand, N. K. (2013). "Understanding the associations between on-farm biosecurity practice and equine influenza infection during the 2007 outbreak in Australia." *Preventive Veterinary Medicine*, 110(1): 28–36. 2

Firestone, S. M., Lewis, F. I., Schemann, K., Ward, M. P., Toribio, J. A., Taylor, M. R., and Dhand, N. K. (2014). "Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a subsequent outbreak." *Preventive Veterinary Medicine*, 116(3): 243–251. 2

Friedman, N., Geiger, D., Goldszmidt, M., Provan, G., Langley, P., and Smyth, P. (1997). "Bayesian Network Classifiers." In *Machine Learning*, 131–163. 3

Friedman, N. and Goldszmidt, M. (1998). "Learning Bayesian networks with local structure." *Learning in Graphical Models*, 421–459. 9

Friedman, N. and Koller, D. (2003). "Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks." *Machine Learning*, 50(1-2): 95–125. 3, 6, 9

Geiger, D. and Heckerman, D. (1994). "Learning Gaussian networks." In *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, UAI'94, 235–243. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 10

— (2002). "Parameter priors for directed acyclic graphical models and the characterization of several probability distributions." *Annals of Statistics*, 30(5): 1412–1440. 15

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam. 17

Heckerman, D. (1998). "A tutorial on learning with Bayesian networks." In Jordan, M. (ed.), *Learning in graphical models*, volume 89, 301–354. 6

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). "Learning Bayesian networks: the combination of knowledge and statistical data." *Machine Learning*, 20(3): 197–243. 3, 6, 9, 10, 15

Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference*. Springer, Heidelberg. Likelihood and Bayes. 11, 17

Hodges, A. P., Dai, D. J., Xiang, Z. S., Woolf, P., Xi, C. W., and He, Y. Q. (2010). "Bayesian Network Expansion Identifies New ROS and Biofilm Regulators." *Plos One*, 5(3): e9513. 2

Jansen, R., Yu, H. Y., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S. B., Emili, A., Snyder, M., Greenblatt, J. F., and Gerstein, M. (2003). "A Bayesian networks approach for predicting protein-protein interactions from genomic data." *Science*, 302(5644): 449–453. 2

Jensen, F. V. (2001). *Bayesian network and decision graphs*. Springer-Verlag, New York. 3, 7, 11

Lam, W. and Bacchus, F. (1994). "Learning Bayesian belief networks. An approach based on the MDL principle." *Computational Intelligence*, 10: 269–293. 9

Lewis, F., Pittavino, M., and Furrer, R. (2015). *abn: modelling multivariate data with additive Bayesian networks*. R package version 0.86. 14, 21

Lewis, F. and Ward, M. (2013). "Improving epidemiologic data analyses through multivariate regression modelling." *Emerging Themes in Epidemiology*, 10(1): 4. 2

Lewis, F. I. (2012). "Bayesian networks as a tool for epidemiological systems analysis." In Sivasundaram, S. (ed.), *9th International Conference On Mathematical Problems In Engineering, Aerospace and Sciences (icnpaa 2012), Amer Inst Physics*, volume 1493 of *AIP Conference Proceedings*, 610–617. 2, 12, 16, 17

Lewis, F. I., Brulisauer, F., and Gunn, G. J. (2011). "Structure discovery in Bayesian networks: An analytical tool for analysing complex animal health data." *Preventive Veterinary Medicine*, 100(2): 109–115. 2

Lewis, F. I. and McCormick, B. J. J. (2012). "Revealing the complexity of health determinants in resource-poor settings." *American Journal of Epidemiology*, 176(11): 1051–1059. 2

Lindley, D. (1957). "A statistical paradox." *Biometrika*, 44(1-2): 187–192. 11, 21

Ludwig, A., Berthiaume, P., Boerlin, P., Gow, S., Léger, D., and Lewis, F. I. (2013). "Identifying associations in Escherichia coli antimicrobial resistance patterns using additive Bayesian networks." *Preventive Veterinary Medicine*, 110(1): 64–75. 2

Massam, H., Liu, J., and Dobra, A. (2009). "A conjugate prior for discrete hierarchical log-linear models." *Annals of Statistics*, 37(6A): 3431–3467. 11

McCormick, B., Sanchez-Vazquez, M., and Lewis, F. (2013). "Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs." *Preventive Veterinary Medicine*, 110(1): 54–63. 2

Nadarajah, S. (2015). "On the computation of Gauss hypergeometric functions." *The American Statistician*, 69(2): 146–148. 24

Needham, C. J., Bradford, J. R., Bulpitt, A. J., and Westhead, D. R. (2007). "A primer on learning in Bayesian networks for computational biology." *Plos Computational Biology*, 3(8): e129. 1

Poon, A. F. Y., Lewis, F. I., Frost, S. D. W., and Pond, S. L. K. (2008). "Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models." *Bioinformatics*, 24(17): 1949–1950. 1

Poon, A. F. Y., Lewis, F. I., Pond, S. L. K., and Frost, S. D. W. (2007a). "Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope." *Plos Computational Biology*, 3(1): e11. 1

— (2007b). "An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope." *Plos Computational Biology*, 3(11): e231. 1

R Core Development Team (2015). "R: A language and environment for statistical computing." *R Foundation for Statistical Computing, Vienna, Austria.*. 21

Rijmen, F. (2008). "Bayesian networks with a logistic regression model for the conditional probabilities." *International Journal of Approximate Reasoning*, 48(2): 659–666. 5

Rissanen, J. (1987). "Stochastic complexity." *Journal of the Royal Statistical Society, Series B*, 49(3): 223–229. With discussions. 9

Sabanés Bové, D. and Held, L. (2011). "Hyper-g priors for generalized linear models." *Bayesian Analysis*, 6(3): 387–410. 2, 17, 23

Sanchez-Vazquez, M., Nielen, M., Edwards, S., Gunn, G., and Lewis, F. (2012). "Identifying associations between pig pathologies using a multi-dimensional machine learning methodology." *BMC Veterinary Research*, 8(1): 151. 2

Schemann, K., Lewis, F. I., Firestone, S. M., Ward, M. P., Toribio, J.-A. L. M. L., Taylor, M. R., and Dhand, N. K. (2013). "Untangling the complex inter-relationships between horse managers' perceptions of effectiveness of biosecurity practices using Bayesian graphical modelling." *Preventive Veterinary Medicine*, 110(1, SI): 37–44. 2

Scutari, M. (2013). "On the prior and posterior distributions used in graphical modelling." *Bayesian Analysis*, 8(3): 505 – 532. 8

Spiegelhalter, D. J. and Lauritzen, S. L. (1990). "Sequential updating of conditional probabilities on directed graphical structures." *Networks*, 20(5): 579–605. 10

Tierney, L. and Kadane, J. B. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81(393): 82–86. 2

Ward, M. P. and Lewis, F. I. (2013). "Bayesian Graphical modelling: Applications in veterinary epidemiology." *Preventive Veterinary Medicine*, 110(1): 1–3. 2

Wilson, A. J., Ribeiro, R., and Boinas, F. (2013). "Use of a Bayesian network model to identify factors associated with the presence of the tick Ornithodoros erraticus on pig farms in southern Portugal." *Preventive Veterinary Medicine*, 110(1, SI): 45–53. 2

Zellner, A. (1986). "On assessing prior distributions and Bayesian regression analysis with g-prior distributions." In *Bayesian inference and decision techniques*, volume 6 of *Stud. Bayesian Econometrics Statist.*, 233–243. North-Holland, Amsterdam. 2, 23

Acknowledgments

The additive Bayesian network models are part of the PhD project by Marta Pittavino. She is a PhD candidate at the Epidemiology and Biostatistics PhD Program of Life Science Zurich Graduate School, and gratefully acknowledges its support.

We gratefully acknowledge funding by the Swiss National Science Foundation (SNF138562 and SNF144973) and by the Fondazione Franco e Marilisa Caligara per l'Alta Formazione Interdisciplinare.

Bayesian Analysis (0000) **00**, Number 0, pp. 1

Supplementary Material

Conjugate Priors for Additive Bayesian Networks

Marta Pittavino* and Reinhard Furrer*,†

This supplementary material contains further insights into additive Bayesian networks with all the proofs and descriptions for the likelihood equivalence of a multivariate Bernoulli likelihood, using an additive parametrization. Moreover, it contains the score equivalence proof for two equivalent binary additive Bayesian networks.

In Section 1, further descriptions of additive Bayesian networks, with a fully explanatory example, are provided. In Example 1, a simple example of the proof of the generalization of Dirichlet density for additive Bayesian networks is shown. In Section 2, the process of learning an additive Bayesian network structure is presented in more detail, with an explanation of the greedy search algorithm and their relationship with a *greedy equivalent search algorithm*. In Section 3, the notion of the *likelihood equivalence* property is introduced and demonstrated. In Section 4, a proof of the *score equivalence*, in the specific case of additive Bayesian networks for binary data, is presented.

1 Further insights into additive Bayesian networks

In particular, when considering the specific subset in Figure 1 in the original manuscript consisting of node X_3 with parent nodes X_1 and X_2, suppose that X_3 can take two possible values, $(S_3 = 2)$, and has two parents (X_1 and X_2), each of which has two possible categories. The expression for the linear predictor for each observation i is then represented by:

$$\boldsymbol{\eta}_{i3} = \begin{pmatrix} \eta_{i31} \end{pmatrix} = \boldsymbol{Z}_{i3}\boldsymbol{\beta}_3 = \boldsymbol{z}_{i3}^T\boldsymbol{\beta}_3 = \begin{bmatrix} \boldsymbol{z}_{i31} & \cdots & \boldsymbol{z}_{i34} \end{bmatrix} \begin{bmatrix} \beta_{311} \\ \beta_{321} \\ \beta_{331} \\ \beta_{341} \end{bmatrix}.$$

There are as many logistic regression parameters as there are free probability parameters $(2 \times 2 \times 1 = 4)$. For any of the 4 possible configurations of $\mathbf{Pa}_j$, one different logistic regression parameter is selected each time by pre-multiplying $\boldsymbol{\beta}_j$ with $\boldsymbol{Z}_{ij}$, one for the response category s, $s = S_j - 1 = 2 - 1 = 1$, to which it corresponds the positive variable state $X_j = x_j^1 = 1$. Each $\boldsymbol{P}(X_j = s|\mathbf{Pa}_j) = \boldsymbol{\theta}_{js}$, in particular $\boldsymbol{P}(X_j = 1|\mathbf{Pa}_j) = \boldsymbol{\theta}_{j1}$,

*Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland marta.pittavino@math.uzh.ch reinhard.furrer@math.uzh.ch

†Institute of Computational Science, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland reinhard.furrer@math.uzh.ch

 DOI: 0000

from a $\mathcal{B}$ is modelled with a $\mathcal{M}$. The corresponding conditional probabilities are obtained as before by applying the inverse of the link function to the linear predictor:

$$\begin{aligned}&\boldsymbol{\eta}_{j1} = \text{logit}(\boldsymbol{\theta}_{j1}) = h(\boldsymbol{\theta}_{j1}) : \ \textit{multinomial link function},\\ &\Rightarrow \boldsymbol{\theta}_{j1} = \frac{e^{\boldsymbol{\eta}_{j1}}}{1+e^{\boldsymbol{\eta}_{j1}}} = \frac{e^{\boldsymbol{Z}_{ij}\boldsymbol{\beta}_j}}{1+e^{\boldsymbol{Z}_{ij}\boldsymbol{\beta}_j}} = \frac{e^{\boldsymbol{z}_{ij}^T\boldsymbol{\beta}_j}}{1+e^{\boldsymbol{z}_{ij}^T\boldsymbol{\beta}_j}} = \frac{e^{\beta_{jc1}}}{1+e^{\beta_{jc1}}}.\end{aligned}$$

Moreover, in order to better understand this parametrization, a specific example is described. The structure shown in Figure 1 in the original manuscript is considered; the related additive Bayesian network model is called $\mathcal{A}_I$ and leads to the additive parameters:

$\boldsymbol{\beta}_{\mathcal{A}_I} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4, \boldsymbol{\beta}_5, \},$
For $j = \{1,2\}$, $\boldsymbol{\beta}_j = \{\boldsymbol{\beta}_{j1}\} = \{\beta_{j11} \cup \beta_{j12}\},$
$\boldsymbol{\beta}_3 = \{\boldsymbol{\beta}_{31}, \boldsymbol{\beta}_{32}, \boldsymbol{\beta}_{33}, \boldsymbol{\beta}_{34}\},$
$\boldsymbol{\beta}_{31} = \{\beta_{311} \cup \beta_{312}\}$, $\boldsymbol{\beta}_{32} = \{\beta_{321} \cup \beta_{322}\}, \boldsymbol{\beta}_{33} = \{\beta_{331} \cup \beta_{332}\}$, $\boldsymbol{\beta}_{34} = \{\beta_{341} \cup \beta_{342}\},$
For $j = \{4,5\}$, $\boldsymbol{\beta}_j = \{\boldsymbol{\beta}_{j1}, \boldsymbol{\beta}_{j2}\} = \{\beta_{j11} \cup \beta_{j12}, \beta_{j21} \cup \beta_{j22}\}.$

Specifically, the following reparametrization occurs, proceeding with a description of the node that follows an increase in parent's orders:

$$\text{For} j = \{1,2\}, \beta_{j11} = h(\theta_{111}) = \log\left(\frac{\theta_{111}}{1-\theta_{111}}\right),\ \theta_{111} = p(X_1{=}1 \| \mathbf{Pa}_1{=}1{=}\varnothing, \boldsymbol{\theta}_{11}),$$

$$j = \{4,5\}, \beta_{j11} = h(\theta_{j11}) = \log\left(\frac{\theta_{j11}}{1-\theta_{j11}}\right), \theta_{j11} = p(X_j{=}1 \| \mathbf{Pa}_j{=}1{=}\{X_3{=}0\}, \boldsymbol{\theta}_{j1}),$$

$$j = \{4,5\}, \beta_{j21} = h(\theta_{j21}) = \log\left(\frac{\theta_{j21}}{1-\theta_{j21}}\right), \theta_{j21} = p(X_j{=}1 \| \mathbf{Pa}_j{=}2{=}\{X_3{=}1\}, \boldsymbol{\theta}_{j2}).$$

For node $j = 3$ further explanation is provided because two parents are involved, implying more parents configuration that require carefulness:

$$\beta_{311} = h(\theta_{311}) = \log\left(\frac{\theta_{311}}{1-\theta_{311}}\right), \theta_{311} = p(X_3 = 1 \| \mathbf{Pa}_3 = 1 = \{X_1 = 0, X_2 = 0\}, \boldsymbol{\theta}_{31}),$$

$$\beta_{321} = h(\theta_{321}) = \log\left(\frac{\theta_{321}}{1-\theta_{321}}\right), \theta_{321} = p(X_3 = 1 \| \mathbf{Pa}_3 = 2 = \{X_1 = 1, X_2 = 0\}, \boldsymbol{\theta}_{32}),$$

$$\beta_{331} = h(\theta_{331}) = \log\left(\frac{\theta_{331}}{1-\theta_{331}}\right), \theta_{331} = p(X_3 = 1 \| \mathbf{Pa}_3 = 3 = \{X_1 = 0, X_2 = 1\}, \boldsymbol{\theta}_{33}),$$

$$\beta_{341} = h(\theta_{341}) = \log\left(\frac{\theta_{341}}{1-\theta_{341}}\right), \theta_{341} = p(X_3 = 1 \| \mathbf{Pa}_3 = 4 = \{X_1 = 1, X_2 = 1\}, \boldsymbol{\theta}_{34}).$$

Example 1. *In order to better understand the essence of the proposition regarding the generalization of the Dirichlet density presented in the paper, a simple example with* 2 *random variables is provided.*

For $n = 2$, considering the complete hypothesized additive Bayesian network structure model $\mathcal{S} = \mathcal{A}_1$, where $\mathcal{A}_1 = \{X_1 \to X_3\}$, a subset example of Figure 1 in the original manuscript.

From the change of variable formula and considering $\mathcal{S} = \mathcal{A}_1$, we get:

$$\pi(\boldsymbol{\beta}_{\mathcal{A}_1}) = \prod_{j=1}^{n} \pi_{\boldsymbol{\theta}_j}(h^{-1}(\boldsymbol{\beta}_j)) \cdot J^{\boldsymbol{\theta}_j \overset{h}{\to} \boldsymbol{\beta}_j} = \prod_{j=1}^{n} \pi(\boldsymbol{\beta}_j|\mathcal{A}_1) = \prod_{j=1}^{2} \pi(\boldsymbol{\beta}_j|\mathcal{A}_1).$$

$$\begin{aligned}
\pi(\boldsymbol{\beta}_1) &= \pi_{\boldsymbol{\theta}_1}(h^{-1}(\boldsymbol{\beta}_1)) \cdot J^{\boldsymbol{\theta}_1 \overset{h}{\to} \boldsymbol{\beta}_1} = \pi(\boldsymbol{\beta}_1|\mathcal{A}_1) = \pi(\boldsymbol{\beta}_{11}|\mathcal{A}_1) \\
&= \frac{\exp(\beta_{111})^{\delta_{111}-1}}{(1+\exp(\beta_{111}))^{\delta_{111}+\delta_{112}-2}} \frac{\exp(\beta_{111})}{(1+\exp(\beta_{111}))^2} = \frac{\exp(\beta_{111})^{\delta_{111}}}{(1+\exp(\beta_{111}))^{\delta 111+\delta_{112}}} \\
&= \exp\left[(\delta_{111}+\delta_{112})\left\{\left(\frac{\delta_{111}}{\delta_{111}+\delta_{112}}\right)\beta_{111} - \log(1+e^{\beta_{111}})\right\}\right].
\end{aligned}$$

$$\begin{aligned}
\pi(\boldsymbol{\beta}_3) &= \pi_{\boldsymbol{\theta}_3}(h^{-1}(\boldsymbol{\beta}_3)) \cdot J^{\boldsymbol{\theta}_3 \overset{h}{\to} \boldsymbol{\beta}_3} = \pi(\boldsymbol{\beta}_3|\mathcal{A}_1) = \pi(\boldsymbol{\beta}_{31}|\mathcal{A}_1) \cdot \pi(\boldsymbol{\beta}_{32}|\mathcal{A}_1) \\
&= \frac{\exp(\beta_{311})^{\delta_{311}-1}}{(1+\exp(\beta_{311}))^{\delta_{311}+\delta_{312}-2}} \frac{e^{\beta_{311}}}{(1+e^{\beta_{311}})^2} \\
&\quad \cdot \frac{\exp(\beta_{321})^{\delta_{321}-1}}{(1+\exp(\beta_{321}))^{\delta_{321}+\delta_{322}-2}} \frac{e^{\beta_{321}}}{(1+e^{\beta_{321}})^2} \\
&= \frac{\exp(\beta_{311})^{\delta_{311}}}{(1+\exp(\beta_{311}))^{\delta_{311}+\delta_{312}}} \cdot \frac{\exp(\beta_{321})^{\delta_{321}}}{(1+\exp(\beta_{321}))^{\delta_{321}+\delta_{322}}} \\
&= \exp\left[(\delta_{311}+\delta_{312})\left\{\left(\frac{\delta_{311}}{\delta_{311}+\delta_{312}}\right)(\beta_{311}) - \log(1+e^{\beta_{311}})\right\}\right] \\
&\quad \cdot \exp\left[(\delta_{321}+\delta_{322})\left\{\left(\frac{\delta_{321}}{\delta_{321}+\delta_{322}}\right)(\beta_{321}) - \log(1+e^{\beta_{321}})\right\}\right].
\end{aligned}$$

2 Learning a Bayesian network

Unfortunately, by following this simplified approach, three fundamental problems for learning Bayesian networks can arise:

1. The space of all Bayesian network structures is extremely large. It has been shown that the number of different structures, $f(n)$, grows more than exponentially in the number n of nodes, as represented by:

$$f(n) = \sum_{j=1}^{n} (-1)^{j+1} \frac{n!}{(n-j)!n!} 2^{j(n-j)} f(n-1). \tag{2.1}$$

2. When searching through the network structures, it is likely that the result will be several equally good candidate structures. Since a Bayesian network can represent any distribution across the set of variables over a complete graph, several candidates may appear; this implies that a Bayesian network over a complete graph cannot be the correct answer.

3. There is also the problem of overfitting: a complete graph can represent the sample distribution exactly, but $\mathcal{D}$ could have been sampled from a sparse network. Alternatively, the selected model may be so close to the sample distribution that it also covers the smallest deviances in the distribution of the original model $\mathcal{A}_1$.

2.1 Search Procedures

Given a score function, the task is to find the highest-scoring Bayesian network structure in the set of all possible network structures. In other words, the task of structural learning is reduced to a searching problem.

Researchers have developed heuristic search strategies that move around in the search space by iteratively performing small changes to the current structure. Specifically, these search methods usually work directly in the space of the Bayesian network structures; hence, each point in this *search space* corresponds to a particular DAG structure, and *search operators* need to be defined. Search operators are used to move from one structure to another, and to determine the neighbourhood of a DAG, namely those DAGs that can be reached in one step from the current DAG. The operators consist of

- *arc addition*: insert a single arc between two nonadjacent nodes;
- *arc deletion*: remove a single arc between two nodes;
- *arc reversal*: reverse the direction of a single arc.

The notation $op(\mathcal{S}, E)$ represents the result of performing the edge operation E on the structure $\mathcal{S}$; in other words $op(\mathcal{S}, E)$ is a DAG that differs from S in terms of one edge only. One important property of these operators is that they only result in local changes to the current structure, i.e. if an arc between X_i and X_j is inserted or deleted, then only the family of X_j is changed, while if an arc between X_i and X_j is reversed, the families of both X_i and X_j are changed. This property is tightly connected to the decomposability of a score function.

If we insert an edge from X_i into X_j, only the local score for X_j will change; thus, when evaluating whether such a move is beneficial, we need only to compute the score difference (or gain) $\Delta(X_i \rightarrow X_j) = score(X_j, \mathbf{Pa}_j \cup \{X_i\}, \mathcal{D}) - score(X_j, \mathbf{Pa}_j, \mathcal{D})$. Of all the possible searching procedures, a simple heuristic approach is the *greedy search*. In the next subsection, the principal steps of a greedy search algorithm will be explained.

2.2 Greedy search

The *greedy search* algorithm chooses some initial structures (usually an empty structure, a randomly chosen structure or a prior structure specified by the user), and then calculates the gain for each legal arc operation; by legal, it is meant that the resulting graph must be acyclic. Specifically, a *greedy search algorithm* consists of the following steps:

1. Let S be an initial structure.

2. Repeat

   (a) Calculate $\Delta(E)$ for each legal operation E:

   - Let $\Delta^* = \max_E \Delta(E)$ and $E^* = \text{argmax}_E \Delta(E)$.

   (b) If $\Delta^* > 0$, then

   - Set $S = op(S, E^*)$.

3. Until $\Delta^* \leq 0$.

Note that, in the previous algorithm, if the parents of two nodes do not change from one iteration to another, the gain $\Delta(X_i \rightarrow X_j)$ of any edge operation involving these two nodes will remain unchanged. This gain can therefore be cached for subsequent iterations, so that the calculations can be reused. These properties are important consequences of the decomposition property of the score function.

However, the limitation of heuristic search algorithms is that it does not guarantee the finding of an optimal global structure, but only of a local optimal structure. Various solutions have been proposed, and an example is the greedy search algorithm with multiple restarts; in other words, after a local maximum is found, the search is reinitialized with a random structure. After this first attempt, the reinitialization is repeated for a fixed number of iterations, and the best structure found, via the entire process is selected. Conversely, a valid alternative solution is to use an exact algorithm (Koivisto and Sood, 2004; Lewis et al., 2015), in which an optimal global structure is found via a reduction in the number of variables that need to be considered.

2.3 Equivalence class search and score equivalence

It can sometimes be advantageous to define the search space using a more abstract representation than that of DAGs. An example is a procedure called the *greedy equivalence search*. The search is based on the observation that data alone cannot be used to discriminate among network structures that represent the same assertions of conditional independence. We start to define this particular form of structures.

Definition 1. *Two DAG network structures, $\mathcal{S}_1$ and $\mathcal{S}_2$ are* ***equivalent*** *if they represent the same independence constraints.*

To better understand the previous definition, an example is provided. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ indicate two DAG network structures, as represented in Figure 1. They are equivalent DAGs because the same independence relations are represented: $P(X_1, X_2) = P(X_1)P(X_2|X_1) = P(X_2)P(X_1|X_2)$. This equivalence is referred to in Heckerman et al. (1995) as *Likelihood Equivalence*, which implies that, if $\mathcal{S}_1$ and $\mathcal{S}_2$ are independent equivalent networks that are related to two BN models ($\mathcal{B}_1$ and $\mathcal{B}_2$), they have the same joint likelihood $P(\boldsymbol{X}|\boldsymbol{\theta}_{\mathcal{B}_1}, \mathcal{S}_1) = P(\boldsymbol{X}|\boldsymbol{\theta}_{\mathcal{B}_2}, \mathcal{S}_2)$.

The equivalence relation is reflexive, symmetric, and transitive; hence, the relationship defines a collection of *equivalence classes*.

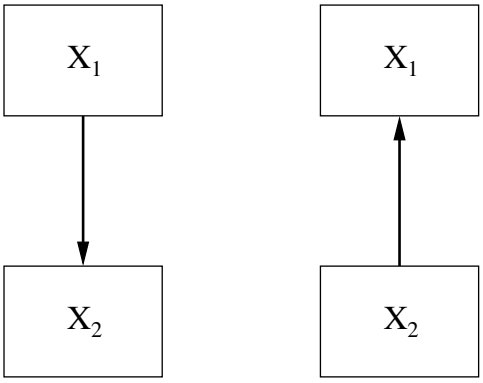


Figure 1: $\mathcal{S}_1$ and $\mathcal{S}_2$, two equivalent DAG structures.

Definition 2. *A score function that assigns the same score to equivalent structures is said to be* ***score equivalent****.*

This definition means that it is not possible to distinguish between different DAGs in an *equivalence class* from observations alone. Hence, fitting each DAG to the same dataset should give the same likelihood of observing the data in each model: *equivalent score function.*

BIC represents an example of a score equivalent function. Heckerman et al. (1995) considered the specification of prior information, such as that of equivalent network structures (Chickering, 1995), is given the same score. This was the first example of an *equivalent Bayesian score function.* In this work, a parameter prior that leads to *Score Equivalence* for ABN models is introduced.

This property means that, if we have found a particular structure using a score equivalent function, we could just as well select any other structure that is equivalent to the one identified.

In order to move around in the space of equivalence classes, where each point in the search space corresponds to an equivalence class, it is possible to identify some search operators, that are a bit more complex than the ones used in DAG spaces, due to the nature of the search space. These operators define the *neighbourhood* for an equivalence class, which is the set of structures reachable by a single change to the current structure or to one of its equivalents. An *upper neighbourhood* is one consisting of equivalence classes with fewer dependence statements, and a *lower neighbourhood* is one with more dependence statements. All the neighbourhoods are based on the definition of equivalence classes in terms of independence statements.

The two neighbourhoods are defined as the equivalence classes that can be obtained by either adding or deleting a single arc from a DAG in the current equivalence class.

Based on this specification of the search space, the **Greedy Equivalence Search Algorithm** consists of two steps:

1. Start with the equivalence class without dependencies among the variables, and

perform a greedy search upwards until a local maximum is reached.

2. Starting from the equivalence class just identified, perform a greedy search downwards until a local maximum is reached.

If the database is sufficiently large, the resulting equivalence class is guaranteed to include the Bayesian network from which the data were generated.

In the context of equivalent structures, greedy search procedures have been proposed by Chickering (2002); Chickering and Meek (2002), and are guaranteed to identify the correct structure when the amount of data becomes large. On the other hand, in the work of Heckerman et al. (1995); Geiger and Heckerman (1994); Boettcher (2004) it is shown for the discrete, the Gaussian and the conditional Gaussian case, respectively, that when a specific choice of the parameter priors is made, the marginal likelihood is the same for equivalent network structures, leading to an *Equivalent Network Score* scenario: $P(\mathcal{D}|\mathcal{S}_1) = P(\mathcal{D}|\mathcal{S}_2)$.

Finally, it should be emphasized that, even though another specification of the search space has been made, the general complexity problem present in DAG spaces has unfortunately not been solved: the number of equivalence classes also grows super exponentially in line with the number of variables.

3 Likelihood equivalence

We proceed with the characterization of our prior, showing that the *Likelihood Equivalence*, another important property, is satisfied. As the name suggests, this property is linked to the likelihood, and it prevents data from helping to identify network structures that represent the same assertions of conditional independence. In order to show that the likelihood equivalence still holds when an additive parametrization is used, we follow the description of the multivariate Bernoulli likelihood presented in Dai et al. (2013).

Specifically, we focus on the two additive Bayesian network structures $\mathcal{A}_1$ and $\mathcal{A}_2$ represented in Figure 1 above. We characterize all the possible contingency table combinations associated with the bivariate Bernoulli likelihood of this specific case, in order to show the *Likelihood Equivalence*

If we consider the bivariate Bernoulli random vector (X_1, X_2), which takes values from $(0,0), (0,1), (1,0)$ and $(1,1)$ in the Cartesian product space $\{0,1\}^2 = \{0,1\} \times \{0,1\}$ and denote $p_{ij} = P(X_1 = i, X_2 = j), i, j = 0, 1,$, its probability density function can be written as

$$\begin{aligned} P(X = x) =& p(x_1, x_2) \\ =& p_{11}^{x_1 x_2} p_{10}^{x_1(1-x_2)} p_{01}^{(1-x_1)x_2} p_{00}^{(1-x_1)(1-x_2)} \\ =& \exp\left\{ \log(p_{00}) + x_1 \log\left(\frac{p_{10}}{p_{00}}\right) + x_2 \log\left(\frac{p_{01}}{p_{00}}\right) + x_1 x_2 \log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right) \right\}. \end{aligned} \tag{3.1}$$

There are 4 possible configurations in total, depending on the two binary variables and their two possible states:

p_{11}	p_{10}
p_{01}	p_{00}

where the side condition $p_{00}+p_{10}+p_{01}+p_{11}=1$ holds to ensure it is a valid probability density function.

To simplify the notation, the natural parameters fs from general parameters are defined in Dai et al. (2013) as follows:

$$f^1 = \log\left(\frac{p_{10}}{p_{00}}\right),$$
$$f^2 = \log\left(\frac{p_{01}}{p_{00}}\right),$$
$$f^{12} = \log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right).$$

Moreover, we introduce another notation to further simply the computation:

$$e_1 = \exp(f^1) = \frac{p_{10}}{p_{00}},$$
$$e_2 = \exp(f^2) = \frac{p_{01}}{p_{00}},$$
$$e_{12} = \exp(f^{12}) = \frac{p_{11}p_{00}}{p_{10}p_{01}}.$$

It is then easy to verify the inverse of the above formula:

$$p_{11} = \frac{\exp(f^1+f^2+f^{12})}{1+\exp(f^1)+\exp(f^2)+\exp(f^1+f^2+f^{12})} = \frac{e_1\cdot e_2\cdot e_{12}}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}}, \quad (3.2)$$
$$p_{01} = \frac{\exp(f^2)}{1+\exp(f^1)+\exp(f^2)+\exp(f^1+f^2+f^{12})} = \frac{e_2}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}}, \quad (3.3)$$
$$p_{10} = \frac{\exp(f^1)}{1+\exp(f^1)+\exp(f^2)+\exp(f^1+f^2+f^{12})} = \frac{e_1}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}}, \quad (3.4)$$
$$p_{00} = \frac{1}{1+\exp(f^1)+\exp(f^2)+\exp(f^1+f^2+f^{12})} = \frac{1}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}}. \quad (3.5)$$

Using the above description of the additive parameters, we can show the *Likelihood Equivalence* property for all 4 aforementioned probability cases, as seen in Table 2.

From Table 2, we get the *Likelihood Equivalence* property:

$$p_{11}^{\mathcal{A}_1} = \theta_1\theta_{2|1} = \theta_2\theta_{1|2} = p_{11}^{\mathcal{A}_2},$$
$$p_{01}^{\mathcal{A}_1} = \theta_{\bar{1}}\theta_{2|\bar{1}} = \theta_2\theta_{\bar{1}|2} = p_{01}^{\mathcal{A}_2},$$

$$p_{10}^{\mathcal{A}_1} = \theta_1\theta_{\bar{2}|1} = \theta_{\bar{2}}\theta_{1|\bar{2}} = p_{10}^{\mathcal{A}_2},$$
$$p_{00}^{\mathcal{A}_1} = \theta_{\bar{1}}\theta_{\bar{2}|\bar{1}} = \theta_{\bar{2}}\theta_{\bar{1}|\bar{2}} = p_{00}^{\mathcal{A}_2}.$$

4 Score equivalence

In this section, we show that, with the *ABN marginal likelihood* resulting from the choice of the newly introduced conjugate prior, it is possible to have the *score equivalence property* satisfied, under a suitable choice of the prior hyperparameters. Under appropriate conditions, it is feasible to get the same *Network Score* for two additive Bayesian network structures belonging to the same *Equivalence Class*.

Proposition 1. *Given two equivalent network structures $\mathcal{A}_1$ and $\mathcal{A}_2$, as shown in Figure 1, that satisfy the* Likelihood Equivalence *property, the resulting network scores $P(\mathcal{D}|\mathcal{A}_1)$ and $P(\mathcal{D}|\mathcal{A}_2)$ are equivalent if, and only if, the sum of the positive case for the orphan node variables is the same in each model (i.e. $X_{11+}^{\mathcal{A}_1} = X_{21+}^{\mathcal{A}_2}$), and if the prior hyperparameters are chosen in the following way:*

$$(a_{\mathcal{A}}, b_{\mathcal{A}}) = \begin{cases} a_1^{\mathcal{A}_1} = a_2^{\mathcal{A}_2}, \\ a_2^{\mathcal{A}_1} = a_1^{\mathcal{A}_2}, \\ b_1^{\mathcal{A}_1} = b_2^{\mathcal{A}_2}, \\ b_2^{\mathcal{A}_1} = b_1^{\mathcal{A}_2}. \end{cases} \tag{4.1}$$

Proof. In order to show this theorem, we have to verify both of the implications.

If the two network score are equivalent, $P(\mathcal{D}|\mathcal{A}_1) = P(\mathcal{D}|\mathcal{A}_2)$, we have:

$$\begin{aligned}
&P(\mathcal{D}|\mathcal{A}_1) \\
&= \prod_{j=1}^{2}\prod_{c=1}^{C_j} \frac{{}_2F_1(\gamma_{j,c}^{\mathcal{A}_1}, \gamma_{j,c}^{\mathcal{A}_1} - \alpha_{j,c}^{\mathcal{A}_1}, 1 + \gamma_{j,c}^{\mathcal{A}_1} - \alpha_{j,c}^{\mathcal{A}_1}, -1)}{\gamma_{j,c}^{\mathcal{A}_1} - \alpha_{j,c}^{\mathcal{A}_1}} + \frac{{}_2F_1(\gamma_{j,c}^{\mathcal{A}_1}, \alpha_{j,c}^{\mathcal{A}_1}, 1 + \alpha_{j,c}^{\mathcal{A}_1}, -1)}{\alpha_{j,c}^{\mathcal{A}_1}} \\
&= \frac{{}_2F_1(\gamma_{1,1}^{\mathcal{A}_1}, \gamma_{1,1}^{\mathcal{A}_1} - \alpha_{1,1}^{\mathcal{A}_1}, 1 + \gamma_{1,1}^{\mathcal{A}_1} - \alpha_{1,1}^{\mathcal{A}_1}, -1)}{\gamma_{1,1}^{\mathcal{A}_1} - \alpha_{1,1}^{\mathcal{A}_1}} + \frac{{}_2F_1(\gamma_{1,1}^{\mathcal{A}_1}, \alpha_{1,1}^{\mathcal{A}_1}, 1 + \alpha_{1,1}^{\mathcal{A}_1}, -1)}{\alpha_{1,1}^{\mathcal{A}_1}} \\
&\quad \cdot \frac{{}_2F_1(\gamma_{2,1}^{\mathcal{A}_1}, \gamma_{2,1}^{\mathcal{A}_1} - \alpha_{2,1}^{\mathcal{A}_1}, 1 + \gamma_{2,1}^{\mathcal{A}_1} - \alpha_{2,1}^{\mathcal{A}_1}, -1)}{\gamma_{2,1}^{\mathcal{A}_1} - \alpha_{2,1}^{\mathcal{A}_1}} + \frac{{}_2F_1(\gamma_{2,1}^{\mathcal{A}_1}, \alpha_{2,1}^{\mathcal{A}_1}, 1 + \alpha_{2,1}^{\mathcal{A}_1}, -1)}{\alpha_{2,1}^{\mathcal{A}_1}} \\
&\quad \cdot \frac{{}_2F_1(\gamma_{2,2}^{\mathcal{A}_1}, \gamma_{2,2}^{\mathcal{A}_1} - \alpha_{2,2}^{\mathcal{A}_1}, 1 + \gamma_{2,2}^{\mathcal{A}_1} - \alpha_{2,2}^{\mathcal{A}_1}, -1)}{\gamma_{2,2}^{\mathcal{A}_1} - \alpha_{2,2}^{\mathcal{A}_1}} + \frac{{}_2F_1(\gamma_{2,2}^{\mathcal{A}_1}, \alpha_{2,2}^{\mathcal{A}_1}, 1 + \alpha_{2,2}^{\mathcal{A}_1}, -1)}{\alpha_{2,2}^{\mathcal{A}_1}} \\
&= \frac{{}_2F_1(\gamma_{1,1}^{\mathcal{A}_2}, \gamma_{1,1}^{\mathcal{A}_2} - \alpha_{1,1}^{\mathcal{A}_2}, 1 + \gamma_{1,1}^{\mathcal{A}_2} - \alpha_{1,1}^{\mathcal{A}_2}, -1)}{\gamma_{1,1}^{\mathcal{A}_2} - \alpha_{1,1}^{\mathcal{A}_2}} + \frac{{}_2F_1(\gamma_{1,1}^{\mathcal{A}_2}, \alpha_{1,1}^{\mathcal{A}_2}, 1 + \alpha_{1,1}^{\mathcal{A}_2}, -1)}{\alpha_{1,1}^{\mathcal{A}_2}} \\
&\quad \cdot \frac{{}_2F_1(\gamma_{1,2}^{\mathcal{A}_2}, \gamma_{1,2}^{\mathcal{A}_2} - \alpha_{1,2}^{\mathcal{A}_2}, 1 + \gamma_{1,2}^{\mathcal{A}_2} - \alpha_{1,2}^{\mathcal{A}_1}, -1)}{\gamma_{1,2}^{\mathcal{A}_2} - \alpha_{1,2}^{\mathcal{A}_2}} + \frac{{}_2F_1(\gamma_{1,2}^{\mathcal{A}_2}, \alpha_{1,2}^{\mathcal{A}_2}, 1 + \alpha_{1,2}^{\mathcal{A}_2}, -1)}{\alpha_{1,2}^{\mathcal{A}_2}} \\
&\quad \cdot \frac{{}_2F_1(\gamma_{2,1}^{\mathcal{A}_2}, \gamma_{2,1}^{\mathcal{A}_2} - \alpha_{2,1}^{\mathcal{A}_2}, 1 + \gamma_{2,1}^{\mathcal{A}_2} - \alpha_{2,1}^{\mathcal{A}_2}, -1)}{\gamma_{2,2}^{\mathcal{A}_2} - \alpha_{2,2}^{\mathcal{A}_2}} + \frac{{}_2F_1(\gamma_{2,1}^{\mathcal{A}_2}, \alpha_{2,1}^{\mathcal{A}_2}, 1 + \alpha_{2,1}^{\mathcal{A}_2}, -1)}{\alpha_{2,1}^{\mathcal{A}_2}}
\end{aligned}$$

$$= \prod_{j=2}^{1} \prod_{c=1}^{C_j} \frac{{}_2F_1(\gamma_{j,c}^{\mathcal{A}_2}, \gamma_{j,c}^{\mathcal{A}_2} - \alpha_{j,c}^{\mathcal{A}_2}, 1 + \gamma_{j,c}^{\mathcal{A}_2} - \alpha_{j,c}^{\mathcal{A}_2}, -1)}{\gamma_{j,c}^{\mathcal{A}_2} - \alpha_{j,c}^{\mathcal{A}_2}} + \frac{{}_2F_1(\gamma_{j,c}^{\mathcal{A}_2}, \alpha_{j,c}^{\mathcal{A}_2}, 1 + \alpha_{j,c}^{\mathcal{A}_2}, -1)}{\alpha_{j,c}^{\mathcal{A}_2}}$$
$$= P(\mathcal{D}|\mathcal{A}_2).$$

This implies that:

$$\begin{cases} \alpha_{1,1}^{\mathcal{A}_1} = \alpha_{2,1}^{\mathcal{A}_2}, \\ \gamma_{1,1}^{\mathcal{A}_1} = \gamma_{2,1}^{\mathcal{A}_2}, \\ \alpha_{2,1}^{\mathcal{A}_1} = \alpha_{1,1}^{\mathcal{A}_2}, \\ \gamma_{2,1}^{\mathcal{A}_1} = \gamma_{1,1}^{\mathcal{A}_2}, \\ \alpha_{2,2}^{\mathcal{A}_1} = \alpha_{1,2}^{\mathcal{A}_2}, \\ \gamma_{2,2}^{\mathcal{A}_1} = \gamma_{1,2}^{\mathcal{A}_2}. \end{cases} \Rightarrow \begin{cases} Pa_{11+}^{\mathcal{A}_1} \cdot a_1^{\mathcal{A}_1} \cdot b_1^{\mathcal{A}_1} + X_{11+}^{\mathcal{A}_1} = Pa_{21+}^{\mathcal{A}_2} \cdot a_2^{\mathcal{A}_2} \cdot b_2^{\mathcal{A}_2} + X_{21+}^{\mathcal{A}_2}, \\ Pa_{11+}^{\mathcal{A}_1} \cdot (a_1^{\mathcal{A}_1} + 1) = Pa_{21+}^{\mathcal{A}_2} \cdot (a_2^{\mathcal{A}_2} + 1), \\ Pa_{21+}^{\mathcal{A}_1} \cdot a_2^{\mathcal{A}_1} \cdot b_2^{\mathcal{A}_1} + X_{21+}^{\mathcal{A}_1} = Pa_{11+}^{\mathcal{A}_2} \cdot a_1^{\mathcal{A}_2} \cdot b_1^{\mathcal{A}_2} + X_{11+}^{\mathcal{A}_2}, \\ Pa_{21+}^{\mathcal{A}_1} \cdot (a_2^{\mathcal{A}_1} + 1) = Pa_{11+}^{\mathcal{A}_2} \cdot (a_1^{\mathcal{A}_2} + 1), \\ Pa_{22+}^{\mathcal{A}_1} \cdot a_2^{\mathcal{A}_1} \cdot b_2^{\mathcal{A}_1} + X_{22+}^{\mathcal{A}_1} = Pa_{12+}^{\mathcal{A}_2} \cdot a_1^{\mathcal{A}_2} \cdot b_1^{\mathcal{A}_2} + X_{12+}^{\mathcal{A}_2}, \\ Pa_{22+}^{\mathcal{A}_1} \cdot (a_2^{\mathcal{A}_1} + 1) = Pa_{12+}^{\mathcal{A}_2} \cdot (a_1^{\mathcal{A}_2} + 1). \end{cases}$$
$$\Leftrightarrow \begin{cases} b_1^{\mathcal{A}_1} = b_2^{\mathcal{A}_2} \wedge X_{11+}^{\mathcal{A}_1} = X_{21+}^{\mathcal{A}_2}, \\ a_1^{\mathcal{A}_1} = a_2^{\mathcal{A}_2}, \\ b_2^{\mathcal{A}_1} = b_1^{\mathcal{A}_2}, \\ a_2^{\mathcal{A}_1} = a_1^{\mathcal{A}_2}, \\ b_2^{\mathcal{A}_1} = b_1^{\mathcal{A}_2}, \\ a_2^{\mathcal{A}_1} = a_1^{\mathcal{A}_2}. \end{cases}$$

It is important to notice that the condition $Pa_{11+}^{\mathcal{A}_1} = Pa_{21+}^{\mathcal{A}_2}$ is always satisfied for the modularity property, while $Pa_{21+}^{\mathcal{A}_1} = X_{11-}^{\mathcal{A}_1} = m - X_{11+}^{\mathcal{A}_1} = m - X_{21+}^{\mathcal{A}_2} = X_{21-}^{\mathcal{A}_2} = Pa_{11+}^{\mathcal{A}_2}$ holds as a result of the condition $X_{11+}^{\mathcal{A}_1} = X_{21+}^{\mathcal{A}_2}$ and is also complementary to $Pa_{22+}^{\mathcal{A}_1} = Pa_{12+}^{\mathcal{A}_2}$. Instead, $X_{21+}^{\mathcal{A}_1} = X_{11+}^{\mathcal{A}_2}$ and $X_{22+}^{\mathcal{A}_1} = X_{12+}^{\mathcal{A}_2}$ is valid due to the data structure.

In the same line as the previous reasoning, it is easy to see that, if the following conditions are fulfilled

$$\begin{cases} b_1^{\mathcal{A}_1} = b_2^{\mathcal{A}_2} \wedge X_{11+}^{\mathcal{A}_1} = X_{21+}^{\mathcal{A}_2}, \\ a_1^{\mathcal{A}_1} = a_2^{\mathcal{A}_2}, \\ b_2^{\mathcal{A}_1} = b_1^{\mathcal{A}_2}, \\ a_2^{\mathcal{A}_1} = a_1^{\mathcal{A}_2}, \\ b_2^{\mathcal{A}_1} = b_1^{\mathcal{A}_2}, \\ a_2^{\mathcal{A}_1} = a_1^{\mathcal{A}_2}. \end{cases}$$

then $P(\mathcal{D}|\mathcal{A}_1) = P(\mathcal{D}|\mathcal{A}_2)$, and the Score Network equivalence is satisfied. □

To better understand the previous proof, we provide a short example.

Example 2. *Considering two equivalent network structures* $\mathcal{A}_1 = \{X_1 \to X_2\}$ *and* $\mathcal{A}_2 = \{X_2 \to X_1\}$*, in Figure 1, and supposing we observed three* $(= m)$ *cases:* $C_1 = \{1, 1\}, C_2 = \{1, 0\}, C_3 = \{0, 1\}$*. Let* $j = 1(2)$ *refer to variable* $X_1(X_2)$*, and let* $c = 1(2)$ *denote the false (true) state of the parent variable;* $+$ *refers to the sum of the variable*

state $X_j = 1$, while $-$ refers to the sum of the variable state $X_j = 0$. Thus, for the network structure $\mathcal{A}_1$, we have the sufficient statistics:

$$\begin{cases} X_{11+}^{\mathcal{A}_1} = Pa_{22+} = 2, \\ X_{11-}^{\mathcal{A}_1} = Pa_{21+} = 1, \\ X_{21+}^{\mathcal{A}_1} = 1, \\ X_{21-}^{\mathcal{A}_1} = 0, \\ X_{22+}^{\mathcal{A}_1} = 1, \\ X_{22-}^{\mathcal{A}_1} = 1. \end{cases}$$

Choosing the prior hyperparameters for $\mathcal{A}_1$, following condition (4.1) as $a_1^{\mathcal{A}_1} = 2$, $b_1^{\mathcal{A}_1} = 0.6$, $a_2^{\mathcal{A}_1} = 1$ and $b_2^{\mathcal{A}_1} = 0.5$, and using the formula (5.13) from the original manuscript, we get the Network Score $P(\mathcal{D}|\mathcal{A}_1) = 0.00119154$.

For the network structure $\mathcal{A}_2$, we then have the sufficient statistics:

$$\begin{cases} X_{22+}^{\mathcal{A}_2} = Pa_{12+} = 2, \\ X_{22-}^{\mathcal{A}_2} = Pa_{11+} = 1, \\ X_{11+}^{\mathcal{A}_2} = 1, \\ X_{11-}^{\mathcal{A}_2} = 0, \\ X_{12+}^{\mathcal{A}_2} = 1, \ X_{12-}^{\mathcal{A}_2} = 1. \end{cases}$$

Choosing the prior hyperparameters for $\mathcal{A}_2$, following condition (4.1) as $a_2^{\mathcal{A}_2} = 2$, $b_2^{\mathcal{A}_2} = 0.6$, $a_1^{\mathcal{A}_2} = 1$ and $b_1^{\mathcal{A}_2} = 0.5$, and using the formula (5.13) from the original manuscript, we get the Network Score $P(\mathcal{D}|\mathcal{A}_2) = 0.00119154$.

We introduced an ABN metric with this property, which is *Score Equivalent*, and it could be called an 'Additive Bayesian networks equivalent scoring metric' *ABNe metric*.

References

Boettcher, S. G. (2004). "Learning Bayesian networks with mixed variables." Ph.D. thesis, Aalborg University - Department of Mathematical Sciences. 7

Chickering, D. M. (1995). "A transformational characterization of Bayesian networks." In Besnard, I. P. and Hanks, S. (eds.), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 87–98. Morgan Kaufmann. 6

— (2002). "Optimal structure identification with greedy search." *Journal of Machine Learning Research*, 3: 507–554. 7

Chickering, D. M. and Meek, C. (2002). "Finding optimal Bayesian networks." In Darwiche, I. A. and Friedman, N. (eds.), *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, 94–102. Morgan Kaufmann. 7

Dai, B., Ding, S., and Wahba, G. (2013). "Multivariate Bernoulli distribution." *Bernoulli*, 19(4): 1465–1483. 7, 8

Geiger, D. and Heckerman, D. (1994). "Learning Gaussian networks." In *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, UAI'94, 235–243. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 7

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). "Learning Bayesian networks: the combination of knowledge and statistical data." *Machine Learning*, 20(3): 197–243. 5, 6, 7

Koivisto, M. and Sood, K. (2004). "Exact Bayesian structure discovery in Bayesian networks." *Journal of Machine Learning Research*, 5: 549–573. 5

Lewis, F., Pittavino, M., and Furrer, R. (2015). *abn: modelling multivariate data with additive Bayesian networks*. R package version 0.86. 5

().

First model	Second model
$\boldsymbol{\theta}_{\mathcal{B}_1} = \begin{cases} \theta_1 = p_{10} + p_{11}, \quad \theta_{\bar{1}} = p_{00} + p_{01}, \\ \theta_{2\vert\bar{1}} = \frac{p_{01}}{p_{00}+p_{01}}, \quad \theta_{\bar{2}\vert\bar{1}} = \frac{p_{00}}{p_{00}+p_{01}}, \\ \theta_{2\vert 1} = \frac{p_{11}}{p_{10}+p_{11}}, \quad \theta_{\bar{2}\vert 1} = \frac{p_{10}}{p_{10}+p_{11}}. \end{cases}$	$\boldsymbol{\theta}_{\mathcal{B}_2} = \begin{cases} \theta_2 = p_{01} + p_{11}, \quad \theta_{\bar{2}} = p_{00} + p_{10}, \\ \theta_{1\vert\bar{2}} = \frac{p_{10}}{p_{00}+p_{10}}, \quad \theta_{\bar{1}\vert\bar{2}} = \frac{p_{00}}{p_{00}+p_{10}}, \\ \theta_{1\vert 2} = \frac{p_{11}}{p_{01}+p_{11}}, \quad \theta_{\bar{1}\vert 2} = \frac{p_{01}}{p_{01}+p_{11}}. \end{cases}$
$\boldsymbol{\theta}_{\mathcal{B}_1} \xrightarrow{h} \boldsymbol{\beta}_{\mathcal{A}_1} = \begin{cases} \beta^{\mathcal{A}_1}_{1\bar{1}1} = \log\left(\frac{\theta_1}{1-\theta_1}\right) = \log\left(\frac{p_{10}+p_{11}}{p_{00}+p_{01}}\right), \\ \beta^{\mathcal{A}_1}_{2\bar{1}1} = \log\left(\frac{\theta_{2\vert\bar{1}}}{1-\theta_{2\vert\bar{1}}}\right) = \log\left(\frac{p_{01}}{p_{00}}\right), \\ \beta^{\mathcal{A}_1}_{221} = \log\left(\frac{\theta_{2\vert 1}}{1-\theta_{2\vert 1}}\right) = \log\left(\frac{p_{11}}{p_{10}}\right). \end{cases}$	$\boldsymbol{\theta}_{\mathcal{B}_2} \xrightarrow{h} \boldsymbol{\beta}_{\mathcal{A}_2} = \begin{cases} \beta^{\mathcal{A}_2}_{2\bar{1}1} = \log\left(\frac{\theta_2}{1-\theta_2}\right) = \log\left(\frac{p_{01}+p_{11}}{p_{00}+p_{10}}\right), \\ \beta^{\mathcal{A}_2}_{1\bar{1}1} = \log\left(\frac{\theta_{1\vert\bar{2}}}{1-\theta_{1\vert\bar{2}}}\right) = \log\left(\frac{p_{10}}{p_{00}}\right), \\ \beta^{\mathcal{A}_2}_{121} = \log\left(\frac{\theta_{2\vert 1}}{1-\theta_{2\vert 1}}\right) = \log\left(\frac{p_{11}}{p_{01}}\right). \end{cases}$
$\boldsymbol{\beta}_{\mathcal{A}_1} = \begin{cases} \beta^{\mathcal{A}_1}_{1\bar{1}1} = \log\left(\frac{p_{10}+p_{11}}{p_{00}+p_{01}}\right) = \log\left(\frac{e^1+e^1\cdot e^2\cdot e^{12}}{1+e^2}\right), \\ \beta^{\mathcal{A}_1}_{2\bar{1}1} = \log\left(\frac{p_{01}}{p_{00}}\right) = f^2, \\ \beta^{\mathcal{A}_1}_{221} = \log\left(\frac{p_{11}}{p_{10}}\right) = f^2 + f^{12}. \end{cases}$	$\boldsymbol{\beta}_{\mathcal{A}_2} = \begin{cases} \beta^{\mathcal{A}_2}_{2\bar{1}1} = \log\left(\frac{p_{01}+p_{11}}{p_{00}+p_{10}}\right) = \log\left(\frac{e^2+e^1\cdot e^2\cdot e^{12}}{1+e^2}\right), \\ \beta^{\mathcal{A}_2}_{1\bar{1}1} = \log\left(\frac{p_{10}}{p_{00}}\right) = f^1, \\ \beta^{\mathcal{A}_2}_{121} = \log\left(\frac{p_{11}}{p_{01}}\right) = f^1 + f^{12}. \end{cases}$

Table 1:: Transformation from Bayesian to additive Bayesian parameters.

First model	Second model
$p_{11}^{\mathcal{A}_1} = \theta_1\theta_{2\mid 1} = \frac{\exp(\beta_{111}^{\mathcal{A}_1})}{1+\exp(\beta_{111}^{\mathcal{A}_1})} \cdot \frac{\exp(\beta_{221}^{\mathcal{A}_1})}{1+\exp(\beta_{221}^{\mathcal{A}_1})}$ $= \frac{\frac{e_1+e_1\cdot e_2\cdot e_{12}}{1+e_2}}{1+\frac{e_1+e_1\cdot e_2\cdot e_{12}}{1+e_2}} \cdot \frac{e_2\cdot e_{12}}{1+e_2\cdot e_{12}}$ $= \frac{(e_1e_2e_{12})(1+e_2e_{12})}{(1+e_1+e_2+e_1e_2e_{12})(1+e_2e_{12})}$ $= \frac{e_1\cdot e_2\cdot e_{12}}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}} = (3.2),$	$p_{11}^{\mathcal{A}_2} = \theta_2\theta_{1\mid 2} = \frac{\exp(\beta_{211}^{\mathcal{A}_2})}{1+\exp(\beta_{211}^{\mathcal{A}_2})} \cdot \frac{\exp(\beta_{121}^{\mathcal{A}_2})}{1+\exp(\beta_{121}^{\mathcal{A}_2})}$ $= \frac{\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_1}}{1+\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_1}} \cdot \frac{e_1\cdot e_{12}}{1+e_1\cdot e_{12}}$ $= \frac{(e_1e_2e_{12})(1+e_1e_{12})}{(1+e_1+e_2+e_1e_2e_{12})(1+e_1e_{12})}$ $= \frac{e_1\cdot e_2\cdot e_{12}}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}} = (3.2),$
$p_{01}^{\mathcal{A}_1} = \theta_{\bar{1}}\theta_{2\mid \bar{1}} = \frac{1}{1+\exp(\beta_{111}^{\mathcal{A}_1})} \cdot \frac{\exp(\beta_{211}^{\mathcal{A}_1})}{1+\exp(\beta_{211}^{\mathcal{A}_1})}$ $= \frac{e_2}{1+e_2+\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_2}+\frac{e_1e_2+e_1e_2^2e_{12}}{1+e_2}}$ $= \frac{e_2}{\frac{(1+e_2)(1+e_1+e_2+e_1e_2e_{12})(1+e_2e_{12})}{(1+e_2)}}$ $= \frac{e_2}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}} = (3.3),$	$p_{01}^{\mathcal{A}_2} = \theta_2\theta_{\bar{1}\mid 2} = \frac{\exp(\beta_{211}^{\mathcal{A}_2})}{1+\exp(\beta_{211}^{\mathcal{A}_2})} \cdot \frac{1}{1+\exp(\beta_{121}^{\mathcal{A}_2})}$ $= \frac{\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_1}}{1+\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_1}+e_1e_{12}+\frac{e_1e_2e_{12}+e_1^2e_2e_{12}^2}{1+e_1}}$ $= \frac{e_2(1+e_1e_{12})}{(1+e_1+e_2+e_1e_2e_{12})(1+e_1e_{12})}$ $= \frac{e_2}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}} = (3.3).$
$p_{10}^{\mathcal{A}_1} = \theta_1\theta_{\bar{2}\mid 1} = \frac{\exp(\beta_{111}^{\mathcal{A}_1})}{1+\exp(\beta_{111}^{\mathcal{A}_1})} \cdot \frac{1}{1+\exp(\beta_{221}^{\mathcal{A}_1})}$ $= \frac{\frac{e_1+e_1\cdot e_2\cdot e_{12}}{1+e_2}}{1+\frac{e_1+e_1\cdot e_2\cdot e_{12}}{1+e_2}+e_2e_{12}+\frac{e_1e_2e_{12}+e_1e_2^2e_{12}}{1+e_2}}$ $= \frac{e_1(1+e_2e_{12})}{(1+e_1+e_2+e_1e_2e_{12})(1+e_2e_{12})}$ $= \frac{e_1}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}} = (3.4),$	$p_{10}^{\mathcal{A}_2} = \theta_{\bar{2}}\theta_{1\mid \bar{2}} = \frac{1}{1+\exp(\beta_{211}^{\mathcal{A}_2})} \cdot \frac{\exp(\beta_{111}^{\mathcal{A}_2})}{1+\exp(\beta_{111}^{\mathcal{A}_2})}$ $= \frac{e_1}{1+e_1+\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_1}+\frac{e_1e_2+e_1^2e_2e_{12}}{1+e_1}}$ $= \frac{e_1}{\frac{(1+e_1)(1+e_1+e_2+e_1e_2e_{12})(1+e_2e_{12})}{(1+e_1)}}$ $= \frac{e_1}{1+e_1+e_2+e_1\cdot e_2\cdot e_{12}} = (3.4),$
$p_{00}^{\mathcal{A}_1} = \theta_{\bar{1}}\theta_{\bar{2}\mid \bar{1}} = \frac{1}{1+\exp(\beta_{111}^{\mathcal{A}_1})} \cdot \frac{1}{1+\exp(\beta_{211}^{\mathcal{A}_1})}$ $= \frac{1}{1+e_2+\frac{e_1+e_1\cdot e_2\cdot e_{12}}{1+e_2}+\frac{e_1e_2+e_1e_2^2e_{12}}{1+e_2}}$ $= \frac{1}{\frac{(1+e_2)(1+e_1+e_2+e_1e_2e_{12})}{(1+e_2)}}$ $= \frac{1}{1+e_1+e_2+e_1e_2e_{12}} = (3.5),$	$p_{00}^{\mathcal{A}_2} = \theta_{\bar{2}}\theta_{\bar{1}\mid \bar{2}} = \frac{1}{1+\exp(\beta_{211}^{\mathcal{A}_2})} \cdot \frac{1}{1+\exp(\beta_{111}^{\mathcal{A}_2})}$ $= \frac{1}{1+\frac{e_2+e_1\cdot e_2\cdot e_{12}}{1+e_1}+e_1+\frac{e_1e_2+e_1^2e_2e_{12}^2}{1+e_1}}$ $= \frac{e_2(1+e_1e_{12})}{(1+e_1+e_2+e_1e_2e_{12})(1+e_1)}$ $= \frac{1}{1+e_1+e_2+e_1e_2e_{12}} = (3.5).$

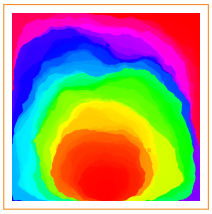
Table 2: Likelihood equivalence list.

APPENDIX A

Curriculum Vitae

Marta Pittavino

Summary of all the academic activities done during the Ph.D. experience at the University of Zurich, UZH, Switzerland



Marta Pittavino

Curriculum Vitae

Place of origin: Cuneo, CN, Italy; born on August 10th, 1987.

Education

March 2016 **Doctoral thesis in Epidemiology and Biostatistics**, *PhD Program of Life Science Zurich Graduate School*, University of Zurich, Switzerland, PhD Thesis: "Additive Bayesian Networks for Multivariate Data: Parameter Learning, Model Fitting and Applications in Veterinary Epidemiology".
Advisor: Prof. Reinhard Furrer. Jury: Prof. Paul Torgerson, UZH, Prof. Leonhard Held, UZH, Dr. Fraser Lewis, Office of Health Economics.

2009 – 2011 **Master of Science in Mathematics**, $110/110$ *cum Laude, University of Torino*, Italy, Master Thesis: "*Epidemiological models for goat breeding*".

2006 – 2009 **Bachelor of Science in Mathematics**, *University of Torino*, Italy, Undergraduate Thesis: "*Biforcazione nella speciazione: una prospettiva matematica*".

2001 – 2006 **Maturità Scientifica**, *Liceo Scientifico 'Giuseppe Peano'*, Diploma Thesis:"*Il sistema periodico*", Cuneo, Italy.

Research Interests

Biostatistics, biomathematics, applied statistics and applied mathematics for epidemiological and medical studies, veterinary and human epidemiology.

Professional Experience

Jan. 2016 – Present **Postdoctoral fellow at the International Agency for Research on Cancer (IARC)**, *World Health Organization (WHO)*, Lyon, France.

Aug. 2013 – Dec. 2015 **Research assistant at the Institute of Mathematics in Zurich**, *University of Zurich*, UZH, Switzerland.

Jan. 2012 – July 2013 **Research assistant at the Vetsuisse Faculty in Zurich**, *Section of Epidemiology*, University of Zurich, UZH, Switzerland.

June 2011 – Dec. 2011 **Internship at Det Norske Veritas**, *DNV*, Quality control and certification of products, processes and systems, Torino, Italy.

Oct. 2009 – Feb. 2010 **Student guide for the exhibition "Riflessioni e Riflessioni"**, *Regional Museum of Natural Science*, Torino, Italy, http://www.dm.unito.it/riflessioni/.

Oct. 2008 – Feb. 2009 **Student guide for the exhibition "Riflessioni e Riflessioni"**, *Regional Museum of Natural Science*, Torino, Italy, http://www.dm.unito.it/riflessioni/.

July 2007 **Supervising secondary school students at summer language school in UK**, *Cirencester*, Royal Agricultural University.

July 2006 **Supervising secondary school students at summer language school in UK**, *Broadstairs*, Kent school of English.

Teaching Experience

April 2013 – Dec. 2015 **TA, instructor and grader, University of Zurich, UZH**, *Taught lectures and graded for undergraduate and graduate level biostatistics courses*.
"Selected Topics in Statistics" (STA 380); "Practical Bioinformatics" (BIO 334); "Likelihood Inference" (STA 402); "Statistische Modellierung" (STA 121); "Applied Statististical Modelling" (STA 220); "Epidemiologie&Statistik" (EPI 134).

2009 – 2011 **Instructor for Math Olympiads, Mathesis**, *http://www.mathesistorino.it/*, Lectured, prepared problem sets, organized contest simulations for high school students.

2009 – 2010 **TA, undergraduate course instructor, University of Torino**, *Gave exercise lectures and graded for undergraduate level math and lab courses*.

Refereed Journal Publications

April 2016 Van der Lely, S., Stefanovic M., Schmidhalter M., Pittavino, M., Furrer, R., Kessler, T., and Mehnert, U., "Protocol for a prospective, randomized study on neurophysiological assessment of lower urinary tract function in a healthy cohort", (submitted).

Feb. 2016 Pittavino, M. and Dreyfus A., Heuer, C., Benschop, J., Wilson P., Collins-Emerson J., Torgerson P., Furrer R., "Comparison between Generalized Linear Modelling and Additive Bayesian Network. Identification of Factors associated with the Incidence of Antibodies against Leptospira interrogans sv Pomona in Meat Workers in New Zealand.", (submitted).

Feb. 2016 Hartnack, S., Springer, S., Pittavino, M. and Grimm, H., "Attitudes of Austrian veterinarians towards euthanasia in small animal practice: impacts of age and gender on views on euthanasia ", *BMC Veterinary Research* (2016), Volume 12, Issue 26, DOI: 10.1186/s12917-016-0649-0.

Jan. 2016 Pittavino, M. and Furrer R., "Suitable Prior for Additive Bayesian Networks", (submitted).

Jan. 2014 Pittavino, M., Ferreri, L., Giacobini, M., Bertolotti, L., Rosati, S. and Venturino, E., "A CAEV epidemiological model for goat breeding", *Applied Mathematics and Computation*, Volume 227, 156-163.

Aug. 2011 Ajraldi, V., Pittavino, M. and Venturino, E. "Modeling herd behavior in population systems", *Nonlinear analysis-real world applications*, Volume 12, Issue 4, 2319-2338, DOI: 10.1016/j.nonrwa.2011.02.002.

Research Reports and Proceedings

Jan. 2016 Pittavino, M., Lewis, F.I., and Furrer, R.: "abn: an R package for modelling multivariate data using additive Bayesian networks ", Manual for the "abn" R package. Available from the Comprehensive R Archive Network (CRAN), http://cran.r-project.org/package=abn/Vignettes.

Dec. 2012 Panero, M. and Pittavino, M. "GeoGebra e le curve di Bézier", *Proceedings of the Fifth Biannual National Conference DI.FI.MA. (Didattica della Fisica e della Matematica)*, Edited by Robutti, O. and Mosca, M., Torino, 547-556.

May 2012 Pittavino, M., Ferreri, L., Giacobini, M., Bertolotti, L., Rosati, S., and Venturino, E., "Influence of Small Ruminant Lentivirus infection on the replacement rate in goat breeding ", *Hystrix, the Italian Journal of Mammology*, 169.

Invited Talks

March 2016 *Modelling multivariate data with additive Bayesian networks*, Seminar Series in Probability and Statistics, Department of Mathematics "G.Peano", Torino, Italy.

June 2015 *Do gloves help to prevent infection?*, *"Der Gesundheit auf der Spur"*, Open Door Day of Epidemiology, Biostatistics and Prevention Institute, Zurich, Switzerland.

Dec. 2014 *Risk Factors for Leptospira Incidence in Meat Workers in New Zealand. Comparison between Generalized Linear Modelling and Additive Bayesian Network*, Welcome Home Workshop 2014, Torino, Italy.

April 2014 *Rigorous assessment of the replacement rate in disease-free and CAEV-affected goat breedings, via mathematical models*, VPHI Seminar, Bern, Switzerland.

Dec. 2013 *How could a score function vary for a Bayesian network model? Application to medical studies*, Welcome Home Workshop 2013, Torino, Italy.

Dec. 2013 *Additive Bayesian networks: a tool for Systems Epidemiology*, VPH Annual Conference 2013, Bern, Switzerland.

May 2013 *What is . . . a Bayesian Network?*, ZGSM Graduate Colloquium, Zurich, Switzerland.

Dec. 2012 *Developing Bayesian networks as a tool for Zoonotic Systems Epidemiology*, Welcome Home Workshop 2012, Torino, Italy.

Oct. 2011 *GeoGebra e le curve di Bézier*, Vth Biannual National Conference DI.FI.MA. 2011, Torino, Italy.

Conference Presentations

Aug. 2015 *Suitable Prior Distributions for Additive Bayesian Networks Models*, Joint Statistical Meeting 2015 (JSM 2015), Seattle, USA.

June 2015 *Comparison between Generalized Linear Models and Additive Bayesian Networks. Analysing Risk Factors for Leptospira Incidence in Meat Workers in New Zealand*, Joint Meeting of the International Biometric Society (IBS), Austro-Swiss and Italian Regions (IROeS 2015), Milan, Italy.

Sept. 2014 *Additive Bayesian networks as a tool for analysing zoonotic diseases in finishing pigs*, DACH Epidemiologietagung 'Tiergesundheit und Okonomie' (DACH 2014), Zurich, Switzerland.

July 2014 *Suitable Prior Distributions for Additive Bayesian Networks Models*, 27th International Biometric Conference (IBC 2014), Florence, Italy.

Poster Presentation

Oct. 2015; July 2014 — *abn: an R package for modelling data using additive Bayesian networks*, World Statistics Day 2015 and Bayesian Biostatistics 2014, Zurich, Switzerland.

May 2012 — *Influence of Small Ruminant Lentivirus infection on the replacement rate in goat breeding*, VIII Italian Congress of Teriology, Piacenza, Italy.

Awards and Grants

June 2013 — **Award "Silver Medal" and prize "Luciana Picco Botta"**, sponsored by *University of Torino*, university wide award for the best mathematical thesis of the current academic year: http://matematicalm.campusnet.unito.it/do/home.pl/View?doc=medaglia.html.

Dec. 2012 — **Scholarship "Attivitá di ricerca post laurea all'estero"**, funded by the *Caligara Foundation*, to develop research activity post graduation related to interdisciplinary scientific subjects.

May 2011 — **Scholarship "Tesi magistrale di cultura interdisciplinare"**, promoted by *Caligara Foundation*, for the development of original master thesis linked to interdisciplinary scientific topics: http://www.fondazionecaligara.it/.

June 2010 — **Grant "Soggiorno studio all'estero"**, granted by the *University College Renato Einaudi*, to allow the deserving students, living in the College, to spend two or three weeks to study English in a foreign country, http://www.collegioeinaudi.it/.

Dec. 2009 — **Grant "Homo Sapiens-Sapiens"**, fundend by *Inps*, nation wide grant to reward Italian students with the best grades in their bachelor studies.

Computer Skills

Computing Skills — LaTeX, Sweave, R, JAGS, Mathematica, MATLAB, Maple, C++, Linux, Ubuntu, Windows, Microsoft Office, Xfig, SAS.

R package — **abn**: an R package for data modelling with additive Bayesian networks. **Maintainer and Contributer**. Available from the Comprehensive R Archive Network (CRAN), http://cran.r-project.org/package=abn.

Languages

Italian	**Mothertongue**	
English	**Advanced**	
French	**Intermediate**	*B2 Level*
German	**Intermediate**	*B1 Level*

Professional Memberships

Member of the Swiss Statistical Society (SSS), the International Biometric Society (IBS) and the American Statistical Association (ASA).