Discretisation Effects in Naive Bayesian Networks

Roel Bertens, Linda C. van der Gaag, and Silja Renooij

Department of Information and Computing Sciences, Utrecht University
P.O. Box 80.089, 3508 TB, The Netherlands
{R.Bertens,L.C.vanderGaag,S.Renooij}@uu.nl

Abstract. Naive Bayesian networks are often used for classification problems that involve variables of a continuous nature. Upon capturing such variables, their value ranges are modelled as finite sets of discrete values. While the output probabilities and conclusions established from a Bayesian network are dependent of the actual discretisations used for its variables, the effects of choosing alternative discretisations are largely unknown as yet. In this paper, we study the effects of changing discretisations on the probability distributions computed from a naive Bayesian network. We demonstrate how recent insights from the research area of sensitivity analysis can be exploited for this purpose.

1 Introduction

Naive Bayesian networks are being used for a large range of classification problems. These networks in essence are probabilistic graphical models of restricted topology, describing a joint probability distribution over a set of stochastic variables. Efficient algorithms are available for computing any prior or posterior probability of interest over the variables of a network, and over its main output variable more specifically. Most of these algorithms assume all variables to be discrete. A classification problem under study however, may involve variables which are of a continuous nature. For capturing such variables, their value ranges should be modelled as finite sets of discrete values. Several different methods are available for automated discretisation of continuous-valued variables in general; for an overview of such methods, we refer to [1]. For Bayesian-network modelling, these general methods unfortunately tend to yield unsatisfactory results [2]. Yet, while the output probabilities established from a Bayesian network are dependent of the actual ways in which its variables are discretised [3], the effects of choosing alternative discretisations are largely unknown.

In this paper, we study the effects of changing the discretisations of continuous-valued feature variables on the posterior probability distributions computed from a naive Bayesian network. We note that discretising a continuous variable amounts to setting one or more threshold values to split its value range into intervals. Choosing an alternative discretisation thus amounts to changing one or more of these threshold values. From the conditional probability table for the variable at hand it is now readily seen that changing even a single threshold

S. Greco et al. (Eds.): IPMU 2012, Part III, CCIS 299, pp. 161–170, 2012.

value will result in changes in the values of many of the parameter probabilities involved. These parameter values do not change independently: their changes are functionally related through the change in threshold value. We will demonstrate how this functional dependency allows exploiting recent insights from the research area of sensitivity analysis of Bayesian networks in general [4], to efficiently study the effects of changing discretisations. Throughout the paper, we will illustrate our findings using real-world breast-cancer screening data.

The paper is organised as follows. In Sect. 2, we introduce our notations and briefly review sensitivity analysis of Bayesian networks in general. In Sect. 3, we establish functions that describe the effects of changing the discretisation of a feature variable on the probability distributions computed from a naive Bayesian network. The paper ends with our concluding observations in Sect. 4.

2 Preliminaries

We introduce our notational conventions and review recent insights from the field of sensitivity analysis of Bayesian networks in general.

2.1 Naive Bayesian Networks

We consider joint probability distributions $\Pr(\mathbf{V})$ over sets $\mathbf{V}$ of discrete stochastic variables. For our notations, we will use (possibly indexed) upper-case letters V to denote single variables, and bold-faced upper-case letters $\mathbf{V}$ to indicate sets of variables. The possible values of a variable V are denoted by (indexed or primed) small letters v_i; we will write v and $\bar{v}$ more specifically, for the two values of a binary variable V. Bold-faced small letters $\mathbf{v}$ are used to denote joint value combinations for the variables from a set $\mathbf{V}$.

A Bayesian network in general is a probabilistic graphical model describing a joint probability distribution $\Pr(\mathbf{V})$ over the set of variables $\mathbf{V}$. The variables from $\mathbf{V}$ are modelled as nodes in a directed acyclic graph, and the (in)dependency relation among them is captured by arcs. Associated with each variable V in the graph are parameter probabilities $p(V \mid \pi(V))$ from the distribution Pr which jointly describe the influence of the possible values of the parents $\pi(V)$ of V on the probabilities over V itself; these parameter probabilities constitute the conditional probability table of the variable V. A naive Bayesian network now is a Bayesian network of highly restricted topology, consisting of a single class variable C and one or more feature variables E_i. In its graphical structure, all feature variables are connected directly with the class variable, and are unconnected otherwise; the feature variables are thereby modelled as mutually independent given the class variable. Naive Bayesian networks are commonly used for computing posterior probability distributions $\Pr(C \mid \mathbf{e})$ over the possible values of the class variable, given a joint value combination $\mathbf{e}$ for the set $\mathbf{E}$ of feature variables.

2.2 Sensitivity Analysis

Sensitivity analysis is a general technique for studying the effects of parameter variation on the output of a mathematical model. For Bayesian networks more specifically, sensitivity analysis amounts to investigating the effects of varying the values of one or more parameter probabilities on an output probability of interest; to this end, tailored algorithms have been developed [5,6].

In a one-way sensitivity analysis of a Bayesian network, a single parameter probability p is being varied as x and the other parameter probabilities p' from the same conditional probability distribution are co-varied as $\frac{(1-x)}{(1-p)} \cdot p'$. The effects of this variation are described by a mathematical function $f(x)$ which expresses the output probability of interest in terms of the parameter under study. For a marginal probability of interest, this sensitivity function $f(x)$ is linear in the parameter being varied. For a conditional probability of interest, the effects of parameter variation are described by a fraction of two linear functions. The function $f(x)$ then essentially is a fragment of one of the branches of a rectangular hyperbola [5]. Since both the parameter under study and the probability of interest are restricted to values from $[0, 1]$, the range of points is effectively constrained to just a fragment of the hyperbola; the two-dimensional space of feasible points in general is termed the unit window.

In the sequel, we will use higher-order sensitivity analyses in which multiple parameter probabilities are being varied simultaneously. In general, in an n-way sensitivity analysis in which n parameters are being varied, a marginal probability of interest is described by a multi-linear function in these parameters. For a conditional probability of interest, the sensitivity function again is a fraction of two such functions. For example, a two-way sensitivity function that expresses a posterior probability of interest $\Pr(c \mid \mathbf{e})$ in terms of two parameter probabilities which are being varied as x and y, has the following form:

$$f_{\Pr(c|\mathbf{e})}(x, y) = \frac{f_{\Pr(c,\mathbf{e})}(x, y)}{f_{\Pr(\mathbf{e})}(x, y)} = \frac{a_1 \cdot x \cdot y + a_2 \cdot x + a_3 \cdot y + a_4}{b_1 \cdot x \cdot y + b_2 \cdot x + b_3 \cdot y + b_4}$$

where the constants $a_i, b_i, i = 1, \ldots, 4$, are built from the non-varied parameters of the network under study. The two parameter probabilities and the output probability of interest again are restricted to the $[0, 1]$-range, which defines a three-dimensional space of feasible points called the unit cube.

3 Studying the Effects of Discretisation

The basic formalism of naive Bayesian networks requires all included variables to be discrete. Upon modelling domain knowledge, variables which take their value from an intrinsically continuous value range will therefore have to be discretised. Such a discretisation amounts to splitting the variable's value range into two or more disjoint intervals and associating each such interval with a value of a (newly defined) discrete variable. In Sect. 3.1, we will study binary discretisations in view of a binary class variable; in Sect. 3.2, we extend our results to naive Bayesian networks including non-binary variables in general.

3.1 Binary Discretisation in Two-Class Naive Bayesian Networks

We consider a continuous feature variable E and address its binary discretisation, that is, we assume that the value range of E is split into two intervals by means of a threshold value t. Slightly abusing notation, we will write $E < t$ and $E \geq t$ for the two values of the (now discretised) variable E, and use e' to indicate either of these values. Upon including the discretised variable E as a feature variable in a naive Bayesian network with the binary class variable C, a conditional probability table is constructed for E with the parameter probabilities $p(E < t \mid c)$ and $p(E \geq t \mid c)$, and the probabilities $p(E < t \mid \bar{c})$ and $p(E \geq t \mid \bar{c})$. It is now readily seen that changing the discretisation of E by choosing a different threshold value t, will affect *all* parameters from this table. Since these parameter probabilities do not stem all from the same conditional distribution, we must conclude that we cannot study the effects of changing E's discretisation by conducting a one-way sensitivity analysis. It is not necessary however, to use a full four-way sensitivity analysis in all parameters involved either. We observe that by varying the parameter probability $p(E < t \mid c)$, the variation of $p(E \geq t \mid c)$ is covered by standard co-variation; similarly, variation of $p(E \geq t \mid \bar{c})$ is handled by varying $p(E < t \mid \bar{c})$. A two-way sensitivity analysis thus should suffice for studying the effects of changing the discretisation of E on the output probabilities computed from a naive Bayesian network.

In Sect. 2, we reviewed the general form of a two-way sensitivity function expressing an output probability $\Pr(c \mid \mathbf{e})$ computed from a Bayesian network in terms of two parameter probabilities being varied as x and y:

$$f_{\Pr(c|\mathbf{e})}(x, y) = \frac{f_{\Pr(c,\mathbf{e})}(x, y)}{f_{\Pr(\mathbf{e})}(x, y)} = \frac{a_1 \cdot x \cdot y + a_2 \cdot x + a_3 \cdot y + a_4}{b_1 \cdot x \cdot y + b_2 \cdot x + b_3 \cdot y + b_4}$$

For studying the effects of changing the discretisation of our feature variable E, the two parameter probabilities to be varied are $p(E < t \mid c)$ and $p(E < t \mid \bar{c})$ (or their complements). We note that these parameter probabilities stem from different conditional distributions, that is, they are conditioned on different values of the class variable. As a consequence, the two parameters have no interaction effects and the constants a_1 and b_1 are equal to zero. The independency properties of a naive Bayesian network even further constrain the general form of the function, as is shown in the following proposition.

Proposition 1. *Let C be the binary class variable of a naive Bayesian network which further includes the set $\mathbf{E}$ of feature variables. Let $\Pr(c \mid \mathbf{e})$ be the network's probability of interest, for a joint combination of observed values $\mathbf{e}$ for $\mathbf{E}$. Now, let $x = p(e' \mid c)$ and $y = p(e' \mid \bar{c})$ be the parameter probabilities for the observed value e' of the binary feature variable E. Then, the two-way sensitivity function expressing $\Pr(c \mid \mathbf{e})$ in x and y is of the form*

$$f_{\Pr(c|\mathbf{e})}(x, y) = \frac{a \cdot \Pr(c) \cdot x}{a \cdot \Pr(c) \cdot x + a' \cdot \Pr(\bar{c}) \cdot y}$$

where a and a' are constants.

Proof. Using Bayes' theorem and exploiting the independency properties of a naive Bayesian network, we find for our probability of interest $\Pr(c \mid \mathbf{e})$ that

$$\begin{aligned}\Pr(c \mid \mathbf{e}) &= \frac{\Pr(\mathbf{e} \mid c) \cdot \Pr(c)}{\Pr(\mathbf{e} \mid c) \cdot \Pr(c) + \Pr(\mathbf{e} \mid \bar{c}) \cdot \Pr(\bar{c})} \\ &= \frac{\prod_{e'_k \in \mathbf{e}} \Pr(e'_k \mid c) \cdot \Pr(c)}{\prod_{e'_k \in \mathbf{e}} \Pr(e'_k \mid c) \cdot \Pr(c) + \prod_{e'_k \in \mathbf{e}} \Pr(e'_k \mid \bar{c}) \cdot \Pr(\bar{c})}\end{aligned}$$

The result follows with $a = \prod_{e'_k \in \mathbf{e} \setminus e'} \Pr(e'_k \mid c)$ and $a' = \prod_{e'_k \in \mathbf{e} \setminus e'} \Pr(e'_k \mid \bar{c})$. $\square$

We note that the constants a and a' in the sensitivity function stated above are readily computed from the parameter probabilities of the feature variables in the naive Bayesian network; the two-way sensitivity function can in fact be established without the need of any propagation, as a result of the conditional independencies holding among the feature variables. We further note that if the probability of interest pertains to the value c of the class variable C, then the numerator of the sensitivity function does not include the parameter probability being varied as y; similarly, for a probability of interest involving $\bar{c}$, the numerator does not include x. We observe that if the value e' specified in the parameters x and y for E differs from the actually observed value, then both the numerator and the denominator of the sensitivity function include an additional constant. Alternatively, we can choose the complements of x and y as the parameters to be varied, which will again result in a function of the above form.

We illustrate the form of the two-way sensitivity function derived above by means of a simple naive Bayesian network for classifying mammographic images.

Example 1. To distinguish between benign and malignant mass lesions, a simple naive Bayesian network was constructed from breast-cancer screening data from the UCI Data Repository [7]. The available data involved several discrete variables modelling properties of the mass lesions seen in mammographic images, and a continuous variable describing the age of a patient. The naive Bayesian network was constructed with the class variable *Severity*, with the values *benign* and *malignant*; the continuous variable *Age* and the five-valued variable *Shape* were selected for its feature variables. We now suppose that we are interested in the output probability $\Pr(\mathit{Severity} = \mathit{benign} \mid \mathit{Age} < t, \mathit{Shape} = 4)$ for the class variable. In our analysis, we further focus on the effects of varying the two parameter probabilities $x = p(\mathit{Age} < t \mid \mathit{Severity} = \mathit{benign})$ and $y = p(\mathit{Age} < t \mid \mathit{Severity} = \mathit{malignant})$ associated with the feature variable *Age*.

To establish the two-way sensitivity function which describes our output probability of interest in terms of the two parameter probabilities being varied, we need to determine the prior probability of a mass lesion being benign and the conditional probabilities of a shape-4 mass for benign lesions and for malignant lesions respectively. We computed these probabilities from the data collection after removal of the five cases for which no value for the variable *Age* was available. The prior probability of a benign lesion was found to be $\Pr(\mathit{Severity} = \mathit{benign}) =$

0.54. For the variable *Shape*, we found $p(Shape = 4 \mid Severity = benign) = 0.16$ and $p(Shape = 4 \mid Severity = malignant) = 0.71$. With these probabilities, we determined the two-way sensitivity function $f_{benign}(x, y)$ for the output probability of interest to be

$$f_{benign}(x, y) = \frac{0.54 \cdot 0.16 \cdot x}{0.54 \cdot 0.16 \cdot x + 0.46 \cdot 0.71 \cdot y}$$

Figure 1(a) shows the fragment of the function $f_{benign}(x, y)$ that lies within the unit cube; the function $f_{malignant}(x, y)$ describing the effects of varying the same parameter probabilities x and y on the complementary output probability $\Pr(Severity = malignant \mid Age < t, Shape = 4)$ is shown in Fig. 1(b). From Fig. 1(a), we can read for example that a relatively small probability $\Pr(Severity = benign \mid Age < t, Shape = 4)$ of a shape-4 mass lesion being benign in younger patients will be found for small values of the parameter x. □

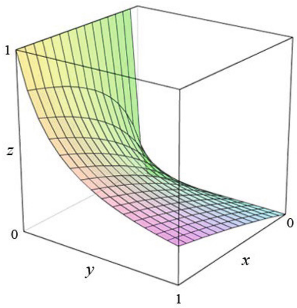
In Proposition 1, we stated the general form of a two-way sensitivity function which expresses an output probability computed from a two-class naive Bayesian network in terms of two parameter probabilities of a binary feature variable. This two-way function specifies a value for the output probability for each combination of values for the two parameters. We now recall that our aim is to use sensitivity analysis as a means for studying the effects of changing the binary discretisation of a continuous-valued feature variable. In view of such a discretisation, the two parameters under study are not unrelated, as is assumed in a two-way sensitivity analysis in general. We observe that since varying the threshold value t in a binary discretisation affects all parameter probabilities of its feature variable, the two parameters under study are dependent of t, and are in fact varied as $x(t)$ and $y(t)$. As a result of this dependency, their variation is related through a function $h(t) = (x(t), y(t))$. From the way in which discretisations are formalised, we have that this function $h(t)$ cannot be any arbitrary function. The following lemma in fact shows that the function is either monotonically non-decreasing or monotonically non-increasing in each of the dimensions of its co-domain.

Lemma 1. *Let C be a binary class variable, let E be a continuous-valued feature variable, and let t be a threshold value for binary discretisation of E. Let $x(t) = p(E < t \mid c)$ and $y(t) = p(E < t \mid \bar{c})$ be parameter probabilities of E, and let h be the function with $h(t) = (x(t), y(t))$. Then, h is monotonically non-decreasing in both dimensions of its co-domain.*

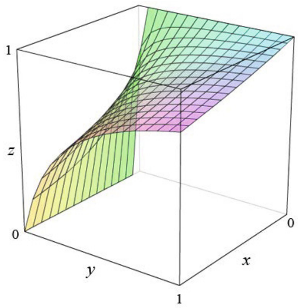
Proof. The property stated in the lemma derives from the interdependency of test characteristics in epidemiology [8], and is easily verified by observing that as the threshold t is shifted to larger values of the continuous variable E, then the probability $p(E < t \mid C)$ cannot decrease, regardless of the value of C. □

From the lemma, we have that the function $h(t)$ is monotonically non-decreasing in any output dimension pertaining to the value $E < t$ of the feature variable E; it is monotonically non-increasing in a dimension pertaining to $E \geq t$.

For studying the overall effect of changing a binary discretisation, we must now explicitly take the induced relation between the two parameter probabilities



(a) $f_{benign}(x, y)$



(b) $f_{malignant}(x, y)$

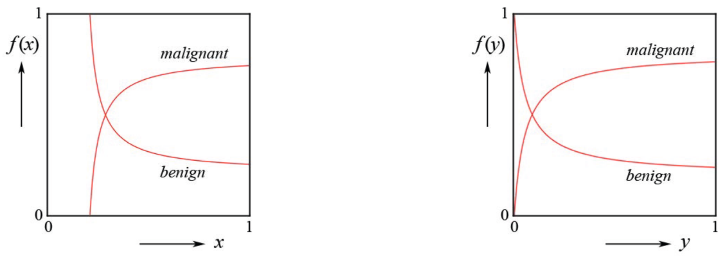
Fig. 1. Two-way sensitivity functions for the class variable *Severity* given $Age < t$ and $Shape = 4$, with the parameters $x = p(Age < t \mid Severity = benign)$ and $y = p(Age < t \mid Severity = malignant)$, assuming independent variation

into account in the sensitivity function under study. Based upon considerations of practicability, we propose to approximate this relation by $y(t) = g(x(t))$ for some function g. Note that by doing so, the dimensionality of the sensitivity function is reduced and its ease of interpretation is enhanced. Studying the effects of changing a discretisation then requires the function

$$f_{\Pr(c|\mathbf{e})}(x(t), g(x(t))) = \frac{a_1 \cdot x(t) + a_2 \cdot g(x(t)) + a_3}{b_1 \cdot x(t) + b_2 \cdot g(x(t)) + b_3}$$

where the constants involved are again built from the non-varied parameters of the network under study. We note that this function is a function in a single parameter probability, but not a one-way sensitivity function; to simplify our notations, we will again omit the explicit dependency of the parameter probabilities $x(t)$ and $y(t)$ on t and write x and y for short. We note in addition that the function g that is chosen to approximate the induced relation between the parameter probabilities x and y cannot be arbitrarily shaped, but should preserve the monotonicity properties of its underlying function h; g is further defined by knowledge of the problem at hand, as is shown in the following example.

Example 2. We consider again, from Example 1, the problem of establishing the severity of mass lesions from mammographic images. From the available data, we approximated the true relation between the parameter probabilities $x = p(Age < t \mid Severity = benign)$ and $y = p(Age < t \mid Severity = malignant)$ by a linear function: by means of linear regression of y on x, we constructed the function $y = 1.00 \cdot x - 0.21$; note that this function preserves the property of non-decreasing values of y for increasing values of x. We now recall that the surface $f_{benign}(x, y)$ from Fig. 1(a) described the probability of interest $\Pr(Severity = benign \mid Age < t, Shape = 4)$ in terms of the two parameters x and y under the assumption of independent variation. By intersecting this surface with the plane $y = 1.00 \cdot x - 0.21$, we therefore find the function that expresses the probability of interest in terms of x taking its actual, albeit approximated, variation effect



(a) $f_{benign}(x)$ and $f_{malignant}(x)$ (b) $f_{benign}(y)$ and $f_{malignant}(y)$

Fig. 2. Dimension-reduced functions for the class variable *Severity* given $Age < t$ and $Shape = 4$, taking the variational dependency of x and y into account

with y into consideration. The intersection curve thus describes the sensitivity of the output probability to changes occasioned in x as a result of varying the discretisation threshold t. The function capturing the intersection curve is

$$f_{benign}(x) = \frac{0.09 \cdot x}{0.09 \cdot x + 0.33 \cdot (1.00 \cdot x - 0.21)}$$

Figure 2(a) displays this function, along with the function for the complement of the probability of interest. We observe that the depicted functions do not specify a value for the probability of interest for the smaller values of the parameter x. This finding originates from the approximated variational dependency of x and y: for small values of x, there are no matching values $g(x)$ for y within the feasible range $[0, 1]$. Note that the finding underlines our earlier observation that the depicted functions are no one-way sensitivity functions, but dimension-reduced two-way sensitivity functions instead. Figure 2(b) again shows the two intersection functions, this time from the perspective of the parameter y; the variational dependency of x and y was now approximated by linear regression of x on y, which resulted in $x = 0.90 \cdot y + 0.25$. □

3.2 Discretisation in Naive Bayesian Networks in General

Thus far, we assumed the class variable of a naive Bayesian network to be binary and considered binary discretisations only. We will now argue that our results are readily generalised to naive Bayesian networks in general, that is, to naive Bayesian networks which include an n-ary class variable and in which the value range of a continuous variable is split into multiple disjoint intervals.

Non-binary Class Variables. We consider an n-ary class variable C with the possible values c_j, $j = 1, \ldots, n$, $n \geq 2$, and assume that we construct a binary discretisation for our continuous-valued feature variable E through a threshold value t as before. Changing the discretisation of E by choosing a different threshold value will again affect all parameter probabilities specified for E. These parameter probabilities now pertain to n different conditional probability distributions, that is, to n distributions over E conditioned on all possible class values.

To study the effects of changing E's discretisation therefore, we have to vary as x_j the parameter probabilities $p(E < t \mid c_j)$ for all $j = 1, \ldots, n$. The sensitivity function describing the effects of this variation on an output probability of interest thus is an n-way sensitivity function. Despite its higher dimensionality, this sensitivity function again is highly constrained by the independency properties of a naive Bayesian network. For an output probability $\Pr(c_k \mid \mathbf{e})$ for some value c_k, $1 \leq k \leq n$, of the class variable C, the sensitivity function in $x_1, \ldots, x_n$ more specifically has the following form:

$$f_{\Pr(c_k|\mathbf{e})}(x_1, \ldots, x_n) = \frac{a_k \cdot \Pr(c_k) \cdot x_k}{a_1 \cdot \Pr(c_1) \cdot x_1 + \ldots + a_n \cdot \Pr(c_n) \cdot x_n}$$

where a_j, $j = 1, \ldots, n$, again are constants; the proof and conditions of this property are analogous to those of Proposition 1. The sensitivity function stated above again assumes independent variation of its parameters $x_1, \ldots, x_n$, as with n-way analyses in general. As before however, these parameters are mutually related through a function $h(t) = (x_1(t), \ldots, x_n(t))$ which is either monotonically non-decreasing or monotonically non-increasing in each of the dimensions of its co-domain. To take the variational relation among the parameters into account, we propose again to approximate this relation by choosing a single focal parameter x_i and to functionally relate each other parameter x_j, $j = 1, \ldots, n, j \neq i$, to x_i by constructing a function g_j with $x_j = g_j(x_i)$ which preserves the monotonicity properties of the underlying function h. A dimension-reduced sensitivity function then results, showing the overall effects of changing a discretisation on a computed class probability.

Non-binary Discretisations. We address a continuous-valued feature variable E for which we construct an m-ary discretisation, that is, whose value range is split into $m \geq 3$ disjoint intervals; for ease of exposition, we assume the class variable again to be binary. We observe that constructing an m-ary discretisation amounts to setting threshold values t_j, $j = 1, \ldots, m-1$, with $t_j < t_{j+1}$. For such a discretisation, we consider changing just a single threshold value t_k, $1 \leq k \leq m-1$, keeping all other thresholds at their original values. We feel that changing multiple threshold values simultaneously would not just complicate the details of our analysis, but would also yield impractical results. Now, changing the threshold value t_k of the discretisation of our feature variable E will again affect its conditional probability table. Not all parameter probabilities will be influenced by the change, however: only the parameter probabilities $p(t_{k-1} \leq E < t_k \mid C)$ and $p(t_k \leq E < t_{k+1} \mid C)$ will be affected, for each possible value of the class variable. We recall that with binary discretisations we could handle the relation between the affected parameter probabilities from the same conditional distribution by standard co-variation, which allowed us to reduce the dimensionality of the sensitivity function. For m-ary discretisations, the commonly assumed co-variation scheme no longer applies, however: if the parameter probability $p(t_{k-1} \leq E_i < t_k \mid c)$ is varied as x, then the parameter $p(t_k \leq E_i < t_{k+1} \mid c)$ is varied as $1 - x - \sum_{j=1,\ldots,k-2,k+1,\ldots,m-2} p(t_j \leq E_i < t_{j+1} \mid c)$ and all other parameters $p(t_j \leq E_i < t_{j+1} \mid c)$, $j = 1, \ldots, m-1$,

$j \neq k-1, k$, from the same distribution are kept constant. It is readily seen however, that this scheme of variation will again result in a two-way sensitivity function of the form stated in Proposition 1. By taking the variational relation between the two parameter probabilities into account as before therefore, again a dimension-reduced sensitivity function results that allows studying the overall effects of the change in discretisation on a class probability of interest.

4 Conclusions and Further Research

Focusing on naive Bayesian networks, we studied the effects of changing the discretisation of a network's continuous feature variable on the posterior probabilities computed for its class variable. We showed that recent insights from sensitivity analysis of Bayesian networks in general serve to analytically describe these effects. We argued more specifically that changing the discretisation of a feature variable affects multiple parameter probabilities, and showed how the relation that is thus induced among these parameters can be explicitly taken into account for establishing a dimension-reduced sensitivity function that shows the overall effects of the change of discretisation on a class probability of interest. We currently are extending our results to Bayesian network classifiers in general and are studying changes in discretisation that induce a change of the most likely class value. We hope to be able to report our further insights in the discretisation effects in Bayesian network classifiers in the near future.

References

1. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: Russel, S.J. (ed.) Proceedings of the 12th International Conference on Machine Learning, pp. 194–202. Morgan Kaufmann, CA (1995)
2. Uusitalo, L.: Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling 203, 312–318 (2007)
3. Myllymäki, P., Silander, T., Tirri, H., Uronen, P.: B-Course: a web-based tool for Bayesian and causal data analysis. International Journal of Artificial Intelligence Tools 11, 369–387 (2002)
4. Coupé, V.M.H., van der Gaag, L.C.: Properties of sensitivity analysis of Bayesian belief networks. Annals of Mathematics and Artificial Intelligence 36, 323–356 (2002)
5. van der Gaag, L.C., Renooij, S., Coupé, V.M.H.: Sensitivity analysis of probabilistic networks. In: Lucas, P., Gámez, J., Salmerón, A. (eds.) Advances in Probabilistic Graphical Models. STUDFUZZ, vol. 214, pp. 103–124. Springer, NY (2007)
6. Kjærulff, U., van der Gaag, L.C.: Making sensitivity analysis computationally efficient. In: Boutilier, C., Goldszmidt, M. (eds.) Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, pp. 317–325. Morgan Kaufmann, CA (2000)
7. `http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass`
8. Dawson-Saunders, B., Trapp, R.G.: Basic & Clinical Biostatistics. McGraw-Hill, NY (2001)