Sensitivity of Multi-dimensional Bayesian Classifiers

Janneke H. Bolt

Silja Renooij

Technical Report UU-CS-2014-024
October 2014

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

Sensitivity of Multi-dimensional Bayesian Classifiers

Janneke H. Bolt and Silja Renooij

Department of Information and Computing Sciences, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
{j.h.bolt,s.renooij}@uu.nl

1 INTRODUCTION

One-dimensional Bayesian network classifiers (OBCs) are popular tools for classification [**?**]. An OBC is a Bayesian network [**?**] consisting of just a *single* class variable and several feature variables. Multi-dimensional Bayesian network classifiers (MBCs) were introduced to generalise OBCs to multiple class variables [**?**, **?**].

Classification performance of OBCs is known to be rather good. Experimental results that support this observation were substantiated by a study of the sensitivity properties of naive OBCs [**?**]. In this paper we investigate the sensitivity of MBCs. We present sensitivity functions for the outcome probabilities of interest of an MBC and use these functions to study the sensitivity value. This value captures the sensitivity of an output probability to small changes in a parameter. We compare MBCs to OBCs in this respect and conclude that an MBC will on average be even more robust to parameter changes than an OBC.

2 PRELIMINARIES

We denote variables by $\mathbf{A} = \{A_1, \ldots, A_n\}$, a value assignment to A_i by a_i, and a joint value assignment to $\mathbf{A}$ by $\mathbf{a}$. The signs $\sim$ and $\nsim$ indicate compatibility and non-compatibility of assignments, respectively. For example, $abc \sim ab$ and $abc \nsim a\bar{b}$. The subscript '$_0$' indicates original values of parameters and outcome probabilities prior to a parameter shift.

An MBC [**?**, **?**] is a Bayesian network in which the variables $\mathbf{A}$ are divided into a set of class variables $\mathbf{C}$ and a set of feature variables $\mathbf{F}$. With $\pi_{\mathbf{F_i}}$ and $\pi_{\mathbf{C_i}}$ we denote, respectively, feature and class parents of a node A_i; instantiations are indicated by $\pi_{\mathbf{f_i}}$ and $\pi_{\mathbf{c_i}}$. In an MBC, feature variables are not allowed to have class children, that is, $\pi_{\mathbf{F_i}} = \emptyset$ for $A_i \in \mathbf{C}$. An MBC is used to classify a joint value assignment $\mathbf{f}$, that is, to assess $\mathrm{argmax}_{\mathbf{c}} \Pr(\mathbf{c} \mid \mathbf{f})$. An OBC is a special case of an MBC with a single class variable. We will consider MBCs with mutually independent class variables (root class variables) and without restriction on the dependencies between the feature variables.

In a sensitivity analysis, parameters x of a network are varied and some probability of interest as a function of the varied parameters is computed. From this *sensitivity function* $f(x)$, sensitivity properties such as the *sensitivity value* can be computed. The sensitivity value [**?**] is the absolute value of the first derivative of the function at the original assessment x_0 of parameter x, that is, $\left|\frac{df}{dx}(x_0)\right|$. This value

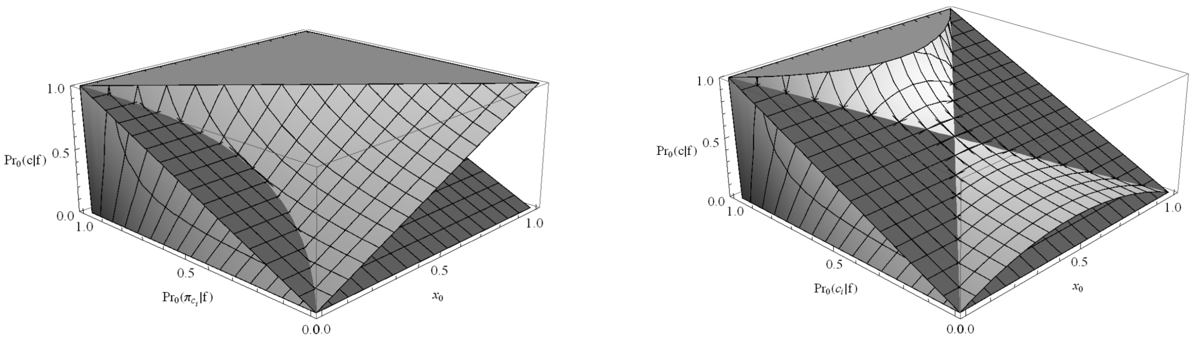

Figure 1: MBC unit sensitivity values. Left: The surface $\Pr_0(\mathbf{c} \mid \mathbf{f}) = x_0/(1 - \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f}))$ (light grey) and the plane $\Pr_0(\mathbf{c} \mid \mathbf{f}) = \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$. Right: The surface $\Pr_0(\mathbf{c} \mid \mathbf{f}) = (x_0 \cdot (1 - x_0))/(1 - \Pr(c_i \mid \mathbf{f}))$ (light grey) and the plane $\Pr_0(\mathbf{c} \mid \mathbf{f}) = \Pr_0(c_i \mid \mathbf{f})$.

captures the impact of local parameter changes on the outcome probability. A sensitivity value of 1 basically marks the transition from a low to a high sensitivity.

3 SENSITIVITY FUNCTIONS FOR MBCS

We propose sensitivity functions $f_{\Pr(\mathbf{c}|\mathbf{f})}(x)$ for the outcome probabilities $\Pr(\mathbf{c} \mid \mathbf{f})$ of an MBC. Proofs are based on the standard assumption of proportional co-variation of parameters pertaining to the same conditional distribution of x. The proofs will be presented in a forthcoming technical report.

Proposition 1. Let $x = \Pr(f_i \mid \pi_{\mathbf{f_i}} \pi_{\mathbf{c_i}})$ be a parameter of a feature variable for which $\mathbf{f} \sim \pi_{\mathbf{f_i}}$. Then $f_{\Pr(\mathbf{c}|\mathbf{f})}(x)$ has one of the following forms:

	$\mathbf{c} \sim \pi_{\mathbf{c_i}}$	$\mathbf{c} \nsim \pi_{\mathbf{c_i}}$
$\mathbf{f} \sim f_i$	$\frac{x \cdot \Pr_0(\mathbf{c}\|\mathbf{f})}{(x - x_0) \cdot \Pr_0(\pi_{\mathbf{c_i}}\|\mathbf{f}) + x_0}$	$\frac{x_0 \cdot \Pr_0(\mathbf{c}\|\mathbf{f})}{(x - x_0) \cdot \Pr_0(\pi_{\mathbf{c_i}}\|\mathbf{f}) + x_0}$
$\mathbf{f} \nsim f_i$	$\frac{(1-x) \cdot \Pr_0(\mathbf{c}\|\mathbf{f})}{(x_0 - x) \cdot \Pr_0(\pi_{\mathbf{c_i}}\|\mathbf{f}) + 1 - x_0}$	$\frac{(1-x_0) \cdot \Pr_0(\mathbf{c}\|\mathbf{f})}{(x_0 - x) \cdot \Pr_0(\pi_{\mathbf{c_i}}\|\mathbf{f}) + 1 - x_0}$

Note that if $\mathbf{c} \sim \pi_{\mathbf{c_i}}$ then $\Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f}) \geq \Pr_0(\mathbf{c} \mid \mathbf{f})$ and if $\mathbf{c} \nsim \pi_{\mathbf{c_i}}$, then $\Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f}) \leq 1 - \Pr_0(\mathbf{c} \mid \mathbf{f})$. Also note that if $\mathbf{f} \nsim \pi_{\mathbf{f_i}}$, then $\Pr(\mathbf{c} \mid \mathbf{f})$ is constant with respect to x.

Proposition 2. Let $x = \Pr(c_i)$ be a parameter of a root class variable. Then $f_{\Pr(\mathbf{c}|\mathbf{f})}(x)$ has one of the following forms:

$$\frac{x \cdot (1 - x_0) \cdot \Pr_0(\mathbf{c} \mid \mathbf{f})}{x \cdot (1 - x_0) \cdot \Pr_0(c_i \mid \mathbf{f}) + (1 - x) \cdot x_0 \cdot (1 - \Pr_0(c_i \mid \mathbf{f}))}, \text{ if } \mathbf{c} \sim c_i$$

$$\frac{(1 - x) \cdot x_0 \cdot \Pr_0(\mathbf{c} \mid \mathbf{f})}{x \cdot (1 - x_0) \cdot \Pr_0(c_i \mid \mathbf{f}) + (1 - x) \cdot x_0 \cdot (1 - \Pr_0(c_i \mid \mathbf{f}))}, \text{ if } \mathbf{c} \nsim c_i$$

Note that if $\mathbf{c} \sim c_i$ we have that $\Pr_0(c_i \mid \mathbf{f}) \geq \Pr_0(\mathbf{c} \mid \mathbf{f})$ and if $\mathbf{c} \nsim c_i$ we have that $\Pr_0(c_i \mid \mathbf{f}) \leq 1 - \Pr_0(\mathbf{c} \mid \mathbf{f})$.

Table 1: MBC Sensitivity values.

x	$\mathbf{c} \sim \pi_{\mathbf{c_i}}$ (or $\mathbf{c} \sim c_i$)	$\mathbf{c} \nsim \pi_{\mathbf{c_i}}$ (or $\mathbf{c} \nsim c_i$)
$\Pr(f_i \mid \pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})$, with $\mathbf{f} \sim f_i$	$\frac{\Pr_0(\mathbf{c}\mid\mathbf{f})\cdot(1-\Pr_0(\pi_{\mathbf{c_i}}\mid\mathbf{f}))}{x_0}$	$\frac{\Pr_0(\mathbf{c}\mid\mathbf{f})\cdot\Pr_0(\pi_{\mathbf{c_i}}\mid\mathbf{f})}{x_0}$
$\Pr(f_i \mid \pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})$, with $\mathbf{f} \nsim f_i$	$\frac{\Pr_0(\mathbf{c}\mid\mathbf{f})\cdot(1-\Pr_0(\pi_{\mathbf{c_i}}\mid\mathbf{f}))}{1-x_0}$	$\frac{\Pr_0(\mathbf{c}\mid\mathbf{f})\cdot\Pr_0(\pi_{\mathbf{c_i}}\mid\mathbf{f})}{1-x_0}$
$\Pr(c_i)$	$\frac{\Pr_0(\mathbf{c}\mid\mathbf{f})\cdot(1-\Pr_0(c_i\mid\mathbf{f}))}{x_0\cdot(1-x_0)}$	$\frac{\Pr_0(\mathbf{c}\mid\mathbf{f})\cdot\Pr_0(c_i\mid\mathbf{f})}{x_0\cdot(1-x_0)}$

4 SENSITIVITY OF MBCS

As mentioned, a sensitivity value > 1 basically indicates a high sensitivity of some outcome to a local parameter change. From the above sensitivity functions we derive general expressions for the sensitivity value $\left|\frac{df}{dx}(x_0)\right|$; these expressions are shown in Table **??**. We then establish for which percentage of combinations of their terms x_0, $\Pr_0(\mathbf{c} \mid \mathbf{f})$, and $\Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$ or $\Pr_0(c_i \mid \mathbf{f})$ a sensitivity value > 1 will be found, given that the terms are chosen independently; and we compare MBCs in general to the special case of OBCs, that is, MBCs with just one class variable, in this respect. Since the terms in the expressions are in fact related, we then reflect on the effect of the dependencies between the terms.

Feature parameters

Consider the sensitivity value from Table **??** for $x = \Pr(f_i \mid \pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})$ with $\mathbf{c} \sim \pi_{\mathbf{c_i}}$ and $\mathbf{f} \sim f_i$. We note that for this case a sensitivity value of 1 entails that $\Pr_0(\mathbf{c} \mid \mathbf{f}) = \frac{x_0}{(1-\Pr_0(\pi_{\mathbf{c_i}}\mid\mathbf{f}))}$. The curved surface in the left graph of Figure **??** captures this relationship. Combinations of values of these terms above this surface result in a sensitivity value > 1. The graph in addition shows the plane $\Pr_0(\mathbf{c} \mid \mathbf{f}) = \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$, giving the boundary of allowed combinations: since $\Pr_0(\mathbf{c} \mid \mathbf{f}) \leq \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$, only value combinations below this plane are feasible. We find that approximately 8% of the feasible combinations has a sensitivity value > 1. We observe that high local sensitivity can only be found given low values of x_0, and is mostly found given non-extreme values of $\Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$ close to $\Pr_0(\mathbf{c} \mid \mathbf{f})$. In the special case of an OBC, $\Pr(\mathbf{c} \mid \mathbf{f}) = \Pr(\pi_{\mathbf{c_i}} \mid \mathbf{f})$. The boundary plane thus represents an OBC. In an OBC now approximately 17% of arbitrarily chosen combinations of terms will result in a sensitivity value > 1. For an OBC this percentage thus is considerably higher than for MBCs in general. The other feature parameter functions yield analogous results.

The 8% and 17% mentioned above hold for independently chosen combinations of terms. The terms, however, are in fact related. Given a feature parameter $x = \Pr(f_i \mid \pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})$ with $\mathbf{c} \sim \pi_{\mathbf{c_i}}$ and $\mathbf{f} \sim f_i$, for example, x_0 is positively related to both $\Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$ and $\Pr_0(\mathbf{c} \mid \mathbf{f})$. Now, roughly spoken, since in an OBC $\mathbf{C_i}$, includes just a single parent whereas in an MBC $\mathbf{C_i}$ may include more parents, in an OBC x_0 will on average have more influence on $\Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$ than in an MBC. At the same time, in an OBC we have that $\Pr_0(\mathbf{c} \mid \mathbf{f}) = \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$ whereas in an MBC, whatever the effect of the dependency between x_0 and $\Pr_0(\mathbf{c} \mid \mathbf{f})$ we have that $\Pr_0(\mathbf{c} \mid \mathbf{f}) \leq \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f})$. As a result of the dependencies between the terms, therefore, the percentages mentioned above will differ, however, the proportion of combinations of terms that results in a sensitivity value > 1, will remain smaller for MBCs in general than for the special case of an OBC.

All in all we conclude that MBCs will on average be less sensitive to feature parameter changes than OBCs.

Root class parameters

The sensitivity of MBCs, compared to OBCs, to shifts in a root class parameter is analysed analogously. The results are presented graphically in the right graph of Figure **??**. Independently chosen, the percentage of feasible combinations of terms with a sensitivity value > 1 is approximately 20% for MBCs and 50% for OBCs. Again, for an OBC this percentage thus is considerably higher than for an MBC in general. We observe that a high local sensitivity is mostly found at either high or low values of x_0, combined with non-extreme values of $\mathrm{Pr}_0(c_i \mid \mathbf{f})$ which are close to $\mathrm{Pr}_0(\mathbf{c} \mid \mathbf{f})$. The same results are found for root class parameters with $\mathbf{c} \nsim c_i$.

Again, as for the feature parameters the terms of the sensitivity functions are in fact related. Analogous arguments as for the feature parameters substantiate that also their dependency will not undo the lower sensitivity of MBCs in general compared the special case of an OBC, as found for independent terms.

All in all we conclude that MBCswill on average be less sensitive to root class parameter changes than OBCs.

Acknowledgements. This research was supported by the Netherlands Organisation for Scientific Research.

APPENDIX

In the following proofs each Pr is a function of x whereas each original value Pr_0 is constant w.r.t. x. With $\pi_{\mathbf{c_i}}$ we indicate an instantiation of the class parents $\pi_{\mathbf{C_i}}$ of F_i and with $\pi_{\mathbf{f_i}}$ we indicate an instantiation of the feature parents $\pi_{\mathbf{F_i}}$ of F_i. In proving Proposition 1 we will use the following property

$$\sum_{\mathbf{C}\sim\pi_{\mathbf{c_i}}} \mathrm{Pr}(\mathbf{f} \mid \mathbf{C}) \cdot \mathrm{Pr}(\mathbf{C}) = \sum_{\mathbf{C}\sim\pi_{\mathbf{c_i}}} \mathrm{Pr}(\mathbf{f}\,\mathbf{C}) = \mathrm{Pr}(\mathbf{f}\pi_{\mathbf{c_i}}) = \mathrm{Pr}(\mathbf{f} \mid \pi_{\mathbf{c_i}}) \cdot \mathrm{Pr}(\pi_{\mathbf{c_i}})$$

where $\mathbf{C} \sim \pi_{\mathbf{c_i}}$ selects the probabilities $\mathrm{Pr}(\mathbf{f}\,\mathbf{C})$ with instantiation $\pi_{\mathbf{c_i}}$ for the subset $\pi_{\mathbf{C_i}}$ of $\mathbf{C}$. Moreover we will use the following MDC property

$$\mathrm{Pr}(\mathbf{f} \mid \mathbf{c}) \quad = \quad \mathrm{Pr}(f_1, \ldots, f_m \mid \mathbf{c}) = \prod_i \mathrm{Pr}(f_i \mid \mathbf{c}) = \prod_i \mathrm{Pr}(f_i \mid \pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})$$

Proof of Proposition 1. Consider a parameter $x = \mathrm{Pr}(f_i \mid \pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})$ with $\mathbf{f} \sim \pi_{\mathbf{f_i}}$. For the case where $\mathbf{c} \sim \pi_{\mathbf{c_i}}$ and $\mathbf{f} \sim f_i$ we find

$$\begin{aligned}\mathrm{Pr}(\mathbf{c} \mid \mathbf{f}) &= \frac{\mathrm{Pr}(\mathbf{f} \mid \mathbf{c}) \cdot \mathrm{Pr}(\mathbf{c})}{\mathrm{Pr}(\mathbf{f})} = \frac{\mathrm{Pr}(\mathbf{f} \mid \mathbf{c}) \cdot \mathrm{Pr}(\mathbf{c})}{\sum_{\mathbf{C}} \mathrm{Pr}(\mathbf{f} \mid \mathbf{C}) \cdot \mathbf{C}} \\ &= \frac{\mathrm{Pr}(\mathbf{f} \mid \mathbf{c}) \cdot \mathrm{Pr}(\mathbf{c})}{\sum_{\mathbf{C}\sim\pi_{\mathbf{c_i}}} \mathrm{Pr}(\mathbf{f} \mid \mathbf{C}) \cdot \mathrm{Pr}(\mathbf{C}) + \sum_{\mathbf{C}\nsim\pi_{\mathbf{c_i}}} \mathrm{Pr}(\mathbf{f} \mid \mathbf{C}) \cdot \mathrm{Pr}(\mathbf{C})}\end{aligned}$$

Replacing the parameter in the product by x, this equals:

$$
\begin{aligned}
& \frac{\frac{x \cdot \Pr_0(\mathbf{f}|\mathbf{c}) \cdot \Pr_0(\mathbf{c})}{x_0}}{\frac{x \cdot \sum_{\mathbf{C} \sim \pi_{\mathbf{c_i}}} \Pr_0(\mathbf{f}|\mathbf{C}) \cdot \Pr_0(\mathbf{C})}{x_0} + \sum_{\mathbf{C} \nsim \pi_{\mathbf{c_i}}} \Pr_0(\mathbf{f} \mid \mathbf{C}) \cdot \Pr_0(\mathbf{C})} \\
= \; & \frac{\frac{x \cdot \Pr_0(\mathbf{fc})}{x_0}}{\frac{x \cdot \sum_{\mathbf{C} \sim \pi_{\mathbf{c_i}}} \Pr_0(\mathbf{fC})}{x_0} + \Pr_0(\mathbf{f}) - \frac{x_0 \cdot \sum_{\mathbf{C} \sim \pi_{\mathbf{c_i}}} \Pr_0(\mathbf{fC})}{x_0}} \\
= \; & \frac{\frac{x \cdot \Pr_0(\mathbf{fc})}{x_0}}{\frac{(x - x_0) \cdot \Pr_0(\mathbf{f}\pi_{\mathbf{c_i}})}{x_0} + \Pr_0(\mathbf{f})}
\end{aligned}
$$

Dividing all terms by $\Pr_0(\mathbf{f})$, this equals:

$$
\frac{\frac{x \cdot \Pr_0(\mathbf{c}|\mathbf{f})}{x_0}}{\frac{(x - x_0) \cdot \Pr_0(\pi_{\mathbf{c_i}}|\mathbf{f})}{x_0} + 1} = \frac{x \cdot \Pr_0(\mathbf{c} \mid \mathbf{f})}{(x - x_0) \cdot \Pr_0(\pi_{\mathbf{c_i}} \mid \mathbf{f}) + x_0}
$$

For the case where $\mathbf{c} \nsim \pi_{\mathbf{c_i}}$ and $\mathbf{f} \sim f_i$ the proof is analogous, but with a constant numerator with respect to the sensitivity parameter.

The sensitivity functions $f_{\Pr(\mathbf{c}|\mathbf{f})}(x)$ with $\mathbf{f} \nsim f_i$ can be derived from their counterparts for $\mathbf{f} \sim f_i$. In that case, a parameter change only influences the probability $\Pr(\mathbf{c} \mid \mathbf{f})$ indirectly as a result of the co-variation of the other parameters from $\Pr(F_i \mid \pi_{\mathbf{f_i}} \pi_{\mathbf{c_i}})$. x and x_0 then have to be replaced by $(1 - x) \cdot \frac{\Pr(f_i|\pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})}{1 - \Pr(f_i|\pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})}$ and $(1 - x_0)\frac{\Pr(f_i|\pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})}{1 - \Pr(f_i|\pi_{\mathbf{f_i}}\pi_{\mathbf{c_i}})}$, respectively, which gives the results stated in Proposition 1.

Proof of Proposition 2. For a parameter $x = \Pr(c_i)$ with $\mathbf{c} \sim c_i$, we find

$$
\Pr(c_i \mid \mathbf{f}) = \frac{\Pr(\mathbf{f} \mid c_i) \cdot \Pr(c_i)}{\sum_{C_i} \Pr(\mathbf{f} \mid C_i) \cdot \Pr(C_i)}
$$

Replacing the parameter in the product by x, this equals:

$$
\frac{x \cdot \Pr_0(\mathbf{f} \mid c_i)}{x \cdot \Pr_0(\mathbf{f} \mid c_i) + \frac{1 - x}{1 - x_0} \cdot \sum_{C_i \setminus c_i} \Pr_0(\mathbf{f} \mid C_i) \cdot \Pr_0(C_i)}
$$

Using that $\sum_{C_i \setminus c_i} \Pr_0(\mathbf{f} \mid C_i) \cdot \Pr_0(C_i) = \sum_{C_i} \Pr_0(\mathbf{f}C_i)) - \Pr_0(\mathbf{f} \mid c_i) \cdot x_0 = \Pr_0(\mathbf{f}) - \Pr_0(\mathbf{f} \mid c_i) \cdot x_0$, and dividing both numerator and denominator by $\Pr_0(\mathbf{f} \mid c_i)$, this equals:

$$
\frac{x}{x + \frac{1 - x}{1 - x_0} \cdot \left(\frac{\Pr_0(\mathbf{f})}{\Pr(\mathbf{f}|c_i)} - x_0\right)}
$$

Multiplying the innermost fraction in the denominator by $\frac{x_0}{\Pr_0(c_i)} (= 1)$, this equals:

$$
\begin{aligned}
& \frac{x}{x + \frac{1 - x}{1 - x_0} \cdot \left(\frac{x_0}{\Pr_0(c_i|\mathbf{f})} - x_0\right)} \\
= \; & \frac{x \cdot \Pr_0(c_i \mid \mathbf{f}) \cdot (1 - x_0)}{x \cdot (1 - x_0) \cdot \Pr_0(c_i \mid \mathbf{f}) + (1 - x) \cdot x_0 \cdot (1 - \Pr_0(c_i \mid \mathbf{f}))}
\end{aligned}
$$

Slightly abusing notation, we have that $\Pr(\mathbf{c} \mid \mathbf{f}) = \Pr(\mathbf{c}\backslash c_i \mid c_i\mathbf{f}) \cdot \Pr(c_i \mid \mathbf{f})$. Since $\Pr(\mathbf{c}\backslash c_i \mid c_i\mathbf{f})$ is constant w.r.t. x, we also have that $\Pr(\mathbf{c}\backslash c_i \mid c_i\mathbf{f}) = \frac{\Pr_0(\mathbf{c}|\mathbf{f})}{\Pr_0(c_i|\mathbf{f})}$ and we thus can write $\Pr(\mathbf{c} \mid \mathbf{f}) = \frac{\Pr_0(\mathbf{c}|\mathbf{f})}{\Pr_0(c_i|\mathbf{f})} \cdot \Pr(c_i \mid \mathbf{f})$. The result now follows by multiplying the expression derived above by $\frac{\Pr_0(\mathbf{c}|\mathbf{f})}{\Pr_0(c_i|\mathbf{f})}$.

For a parameter $\Pr(c_i^*)$ with $\mathbf{c} \nsim c_i^*$ the proof follows exactly the same steps, but with a different numerator: $x \cdot \Pr_0(\mathbf{f} \mid c_i)$ is replaced by $\frac{1-x}{1-x_0} \cdot \Pr_0(\mathbf{f} \mid c_i^*) \cdot \Pr_0(c_i^*)$.

References

[1] L.C. van der Gaag and P.R. de Waal. 2006. Multi-dimensional Bayesian Network Classifiers. In: M Studeny and J Vomlel (Eds.), *Proceedings of the Third European Workshop in Probabilistic Graphical Models*, 107-114.

[2] M. Friedman, D. Geiger and M. Goldschmidt. 1997. Bayesian network classifiers. In: Machine Learning, 29, 131-163.

[3] K.B. Laskey. 1995. Sensitivity analysis for probability assessments in Bayesian networks. In: *IEEE Transactions on Systems, Man and Cybernetics*, 25: 901-909.

[4] J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Palo Alto.

[5] S. Renooij and L.C. van der Gaag. 2008. Evidence and scenario sensitivities in naive Bayesian classifiers. In: *International Journal of Approximate Reasoning*, 49(2): 398-416.

[6] P.R. de Waal and L.C. van der Gaag. 2007. Inference and learning in multi-dimensional Bayesian network classifiers. In: K. Mellouli (ed). *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 501-511.