Same-Decision Probability: A Confidence Measure for Threshold-Based Decisions under Noisy Sensors

Adnan Darwiche and Arthur Choi
Computer Science Department
University of California, Los Angeles
{*darwiche, aychoi*}*@cs.ucla.edu*

Abstract

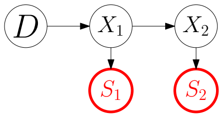
We consider in this paper the robustness of decisions based on probabilistic thresholds under noisy sensor readings. In particular, we consider the stability of these decisions under different assumptions about the causal mechanisms that govern the output of a sensor. To this effect, we propose the *same-decision probability* as a query that can be used as a confidence measure for threshold-based decisions, and study some of its properties.

1 Introduction

There has been an increased interest recently in providing assurances on the results of probabilistic reasoning systems. One clear example are the many results on sensitivity analysis, which is concerned with providing guarantees on the relationship between probabilistic queries and model parameters; see, e.g., (Chan, 2009; Kwisthout and van der Gaag, 2008). These results include specific bounds on the changes in probabilistic queries that could result from perturbing model parameters.

We consider another class of assurances in this paper, which is concerned with quantifying the robustness of threshold-based decisions against noisy observations, where we propose a specific notion, called the *same-decision probability.* Our proposed notion is cast in the context of Bayesian networks where the goal is to make a decision based on whether some probability $Pr(d \mid \mathbf{s})$ passes a given threshold T, where $\mathbf{s}$ represents the readings of noisy sensors. The same-decision probability is based on the following key observation. If one were to know the specific causal mechanisms $\mathbf{h}$ that govern the behavior of each noisy sensor (and hence, allow us to precisely interpret each sensor reading), then one should clearly make the decision based on the probability $Pr(d \mid \mathbf{s}, \mathbf{h})$ instead of the probability $Pr(d \mid \mathbf{s})$. In fact, the probability $Pr(d \mid \mathbf{s})$ can be seen as simply the expectation of $Pr(d \mid \mathbf{s}, \mathbf{h})$ with respect to the distribution $Pr(\mathbf{h} \mid \mathbf{s})$ over causal mechanisms. The same-decision probability is then the probability that we would have made the same threshold-based decision had we known the specific causal mechanism $\mathbf{h}$. More precisely, it is the expected decision based on $Pr(d \mid \mathbf{s}, \mathbf{h})$, with respect to the distribution $Pr(\mathbf{h} \mid \mathbf{s})$ over sensor causal mechanisms.

We show a number of results about this proposed quantity. First, we show that a standard Bayesian network does not contain all of the information necessary to pinpoint the distribution $Pr(\mathbf{h} \mid \mathbf{s})$ which is needed for completely defining the same decision probability. We formulate, however, two assumptions, each of which is sufficient to induce this distribution. Second, we propose a bound on the same-decision probability using the one-sided Chebyshev inequality, which requires only the variance of $Pr(d \mid \mathbf{s}, \mathbf{h})$ with respect to the distribution $Pr(\mathbf{h} \mid \mathbf{s})$. Third, we propose a variable elimination algorithm that computes this variance in time and space that are exponential only in the constrained treewidth of the given network. We conclude with a number of concrete examples that illustrate the utility of our proposed confidence measure in quantifying the robustness of threshold-based decisions.



D	$Pr(D \mid S_1{=}+, S_2{=}+)$
$+$	0.880952
$-$	0.119048

D	$Pr(D)$
$+$	0.5
$-$	0.5

D	X_1	$Pr(X_1 \mid D)$
$+$	$+$	0.9
$+$	$-$	0.1
$-$	$+$	0.1
$-$	$-$	0.9

X_i	S_i	$Pr(S_i \mid X_i)$
$+$	$+$	0.9
$+$	$-$	0.1
$-$	$+$	0.1
$-$	$-$	0.9

Figure 1: A simple Bayesian network, under sensor readings $\{S_1{=}+, S_2{=}+\}$. Variables S_1 and S_2 represent noisy sensor readings, and they have the same CPT $Pr(S_i|X_i)$. Variables X_1 and X_2 also have the same CPTs (only the one for variable X_1 is shown).

Table 1: Causal mechanisms for sensor readings $\mathbf{s} = \{S_1{=}+, S_2{=}+\}$ for the network in Fig. 1. Cases above threshold $T = 0.6$ are in bold.

$\mathbf{h}$	H_1	H_2	$Pr(\mathbf{h}|s_1, s_2)$	$Pr(d|s_1, s_2, \mathbf{h})$
1	t	t	0.781071	**0.90**
2	p	t	0.096429	**0.82**
3	l	t	0.001071	0.10
4	t	p	0.096429	**0.90**
5	p	p	0.021429	0.50
6	l	p	0.001190	0.10
7	t	l	0.001071	**0.90**
8	p	l	0.001190	0.18
9	l	l	0.000119	0.10

2 An Introductory Example

Consider the Bayesian network in Figure 1, which models a scenario involving a hypothesis D of interest and two noisy sensors S_1 and S_2 that bear on this hypothesis. The network captures the false positive and false negative rates of these sensors, where each sensor S_i is meant to measure the state of variable X_i. A typical usage of this and similar networks involves the computation of our belief in the hypothesis given some sensor readings, $Pr(d \mid s_1, s_2)$. This belief can then be the basis of a decision that depends on whether this probability exceeds a certain threshold, $Pr(d \mid s_1, s_2) \geq T$. Scenarios such as this are typical in applications such as diagnosis (Hamscher et al., 1992), troubleshooting (Heckerman et al., 1995a), and probabilistic planning (Littman et al., 1998).

Figure 1 shows a particular reading of two sensors and the resulting belief $Pr(D{=}+ \mid S_1{=}+, S_2{=}+)$. If our threshold is $T = 0.6$, then our computed belief confirms the decision under consideration. This decision, however, is based on the readings of two noisy sensors. Suppose now that our model had explicated the causal mechanisms that led to the sensor readings we observed, as depicted in Table 1 (we discuss how to obtain such a model in the next section). This table depicts a distribution over causal mechanisms $Pr(\mathbf{h} \mid s_1, s_2)$. Assuming a particular causal mechanism $\mathbf{h}$ is the active one, we also have a refined belief $Pr(d \mid s_1, s_2, \mathbf{h})$ on the hypothesis d. In fact, the original belief $Pr(d \mid s_1, s_2)$ can now be seen as the expectation of the refined beliefs with respect to the distribution over causal mechanisms:

$$Pr(d|s_1, s_2) = \sum_{\mathbf{h}} Pr(d|s_1, s_2, \mathbf{h}) Pr(\mathbf{h}|s_1, s_2).$$

We show that this is the case in general, later.

Suppose now that we knew the specific causal mechanism $\mathbf{h}$ that governs our sensor readings. We would then be able to (and would prefer to) make a decision based on the probability $Pr(d \mid s_1, s_2, \mathbf{h})$ instead of the probability $Pr(d \mid s_1, s_2)$, which again, is only an average over possible mechanisms $\mathbf{h}$. Consider for example Table 1 which enumerates all nine causal mechanisms. In only four of these cases does the probability of the hypothesis pass the given threshold (in bold), leading to the same decision. In the other five scenarios, a different decision would have been made. Clearly, the extent to which this should be of concern will depend on the likelihood of these last five scenarios. As such, we propose to quantify the confidence in our decision using the *same-decision probability:* the probability that we would have made the same decision had we known the causal mechanisms governing a sensor's readings. For this example, this probability is 0.975, indicating a relatively robust decision.

3 Noisy Sensors

In this section, we show how we can augment a sensor so that its causal mechanisms are modeled explicitly. The ultimate goal is to construct models like the one in Table 1, which are needed for defining the same-decision probability.

Consider a Bayesian network fragment $X \rightarrow S$, where S represents a sensor that bears on variable X and suppose that both S and X take values in $\{+, -\}$.[1] Suppose further that we are given the false positive f_p and false negative f_n rates of the sensor:

$$Pr(S{=}+|X{=}-) = f_p,\ Pr(S{=}-|X{=}+) = f_n.$$

Our augmented sensor model is based on a functional interpretation of the causal relationship between a sensor S and the event X that it bears on. This causal perspective in turn is based on Laplace's conception of natural phenomena (Pearl, 2009, Section 1.4). In particular, we assume that the output of a sensor S is a deterministic function that depends on the state of X, and that the stochastic nature of the sensor arises from the uncertainty in *which* functional relationship manifests itself.

We propose to expand the above sensor model into $X \rightarrow S \leftarrow H$, where variable H is viewed as a selector for one of the four possible Boolean functions mapping X to S, which we ascribe the labels $\{\mathsf{t}, \mathsf{l}, \mathsf{p}, \mathsf{n}\}$:

H	X	S	$Pr(S\|H, X)$	H	X	S	$Pr(S\|H, X)$
t	$+$	$+$	1	p	$+$	$+$	1
t	$-$	$+$	0	p	$-$	$+$	1
l	$+$	$+$	0	n	$+$	$+$	0
l	$-$	$+$	1	n	$-$	$+$	0

We observe that these Boolean function have commonly used diagnostic interpretations, describing the behavior of a sensor. The state $H{=}\mathsf{t}$ indicates the sensor is truthful, $H{=}\mathsf{l}$ indicates it is lying, $H{=}\mathsf{p}$ indicates it is stuck positive and $H{=}\mathsf{n}$ indicates it is stuck negative. Note that any stochastic model can be emulated by a functional one, with stochastic inputs (Pearl, 2009; Druzdzel and Simon, 1993).

To reason about our augmented sensor model $X \rightarrow S \leftarrow H$, we need to specify a prior distribution $Pr(H)$ over causal mechanisms. Moreover, we need to specify one that yields a model equivalent to the original model $X \rightarrow S$, when variable H has been marginalized out:

$$\begin{aligned} &Pr(S{=}+|X{=}-) \\ &= \textstyle\sum_H Pr(S{=}+|H, X{=}-)Pr(H) = f_p \quad (1) \\ &Pr(S{=}-|X{=}+) \\ &= \textstyle\sum_H Pr(S{=}-|H, X{=}+)Pr(H) = f_n \quad (2) \end{aligned}$$

There is not enough information in the given Bayesian network to identify a unique prior $Pr(H)$. However, if we make some assumptions about this prior, we may be able to pin down a unique one. We make two such proposals here.

For our first proposal, assume that the probability $Pr(H{=}\mathsf{l})$ that a sensor lies is zero, which is a common assumption made in the diagnostic community. This assumption, along with Equations 1 and 2, immediately commits us to the following distribution over causal mechanisms:

H	t	p	n	l
$Pr(H)$	$1 - f_p - f_n$	f_p	f_n	0

For our second proposal, consider the event $\alpha_p = \{H{=}\mathsf{p} \vee H{=}\mathsf{l}\}$ which denotes the materialization of a causal mechanism that produces a false positive behavior by the sensor. That is, if α_p holds, the sensor will report a positive reading when variable X is negative. Moreover, the event $\alpha_n = \{H{=}\mathsf{n} \vee H{=}\mathsf{l}\}$ denotes the materialization of a causal mechanism that produces a false negative behavior by the sensor. Now, if we further assume that the false positive and negative mechanisms of the sensor are independent, we get $Pr(\alpha_p, \alpha_n) = Pr(\alpha_p)Pr(\alpha_n)$. Since α_p, α_n is equivalent to $H{=}\mathsf{l}$, we now get

$$Pr(H{=}\mathsf{l}) = f_p f_n. \quad (3)$$

This assumption, with Equations 1 and 2, commits us to the following CPT:

H	$Pr(H)$
t	$(1 - f_p)(1 - f_n)$
p	$f_p(1 - f_n)$
n	$(1 - f_p)f_n$
l	$f_p f_n$

The assumption is similar to parameter independence used in learning Bayesian networks

[1]Our discussion focuses on sensors over binary variables, but generalizing to multi-valued variables is not difficult; see also (Druzdzel and Simon, 1993).

Table 2: Causal mechanisms for sensor readings $\mathbf{s} = \{S_1 = +, S_2 = -\}$ for the network in Fig. 1. Cases above threshold $T = 0.6$ are in bold.

$\mathbf{h}$	H_1	H_2	$Pr(\mathbf{h}\|s_1, s_2)$	$Pr(d\|s_1, s_2, \mathbf{h})$
1	t	t	0.268893	**0.90**
2	p	t	0.298770	0.18
3	l	t	0.029877	0.10
4	t	n	0.298770	**0.90**
5	p	n	0.066393	0.50
6	l	n	0.003689	0.10
7	t	l	0.029877	**0.90**
8	p	l	0.003689	**0.82**
9	l	l	0.000041	0.10

(Heckerman et al., 1995b).[2] Interestingly, under this assumption (and $f_p + f_n < 1$), as the probabilities of $H = \mathsf{p}$ and $H = \mathsf{n}$ go to zero (i.e., the sensor does not get stuck), the probability of $H = \mathsf{l}$ also goes to zero, therefore, implying that the sensor must be truthful.

Note that the two assumptions discussed above become equivalent as the false positive and negative rates of a sensor approach zero. In fact, as we shall illustrate later, the same-decision probability is almost the same when these rates are small, which is the more interesting case. We stress here, however, that the same decision-probability, as a notion, is independent of the specific assumption adopted — and so are the corresponding computational results we shall present later on computing and bounding this probability.

4 Beliefs Based On Noisy Sensors

Suppose now that we have observed the values of n sensors. For a sensor with a positive reading, the three possible states are $\{\mathsf{t}, \mathsf{l}, \mathsf{p}\}$, since the probability $Pr(H = \mathsf{n})$ that a sensor is stuck-negative is zero when we have a positive reading. Similarly, for a sensor with a negative reading, the three possible states are $\{\mathsf{t}, \mathsf{l}, \mathsf{n}\}$. Hence, we have (at most) 3^n sensor states that have non-zero probability. Each one of these 3^n states are causal mechanisms, and each refers to a hypothesis about which sensors are truthful, which are lying and which are irrelevant. Table 1 depicts the nine causal mechanisms corresponding to two positive sensor readings in the network of Figure 1. The table also depicts the posterior distribution over these mechanisms, suggesting that the overwhelming leading scenario is the one in which the two sensors are truthful ($\mathbf{h}_1$). Table 2 depicts the nine causal mechanisms assuming two conflicting sensor readings.

Given a reading $\mathbf{s}$ of sensors $\mathbf{S}$, and letting $\mathbf{h}$ range over the causal mechanisms, we now have:

$$\begin{aligned} Pr(d \mid \mathbf{s}) &= \sum_{\mathbf{h}} Pr(d \mid \mathbf{h}, \mathbf{s}) Pr(\mathbf{h} \mid \mathbf{s}) \quad (4) \\ &= \sum_{\mathbf{h}} Q(\mathbf{h}) Pr(\mathbf{h} \mid \mathbf{s}). \end{aligned}$$

We thus view the probability $Pr(d \mid \mathbf{s})$ as an expectation $\mathrm{E}[\,Q(\mathbf{H})\,]$ with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{s})$ over causal mechanisms, where $Q(\mathbf{h}) = Pr(d \mid \mathbf{h}, \mathbf{s})$.

Table 1 depicts the posterior over causal mechanisms given two positive sensor readings in the network of Figure 1. We have $Pr(D = + \mid S_1 = +, S_2 = +) = 0.880952$ in this case, which one can easily verify as also being the expectation of $Pr(D = + \mid S_1 = +, S_2 = +, \mathbf{h})$ with respect to the distribution $Pr(\mathbf{h} \mid S_1 = +, S_2 = +)$. Table 2 depicts another posterior over causal mechanisms given two conflicting sensor readings. We have $Pr(D = + \mid S_1 = +, S_2 = -) = 0.631147$ in this case.

5 Same-Decision Probability

As mentioned in the introduction, one is usually interested in making a decision depending on whether the probability of some hypothesis d is no less than some threshold T. Assuming that we know the correct causal mechanism $\mathbf{h}$ governing the sensors readings $\mathbf{s}$, we clearly want to make this decision based on whether the probability $Q(\mathbf{h}) = Pr(d \mid \mathbf{s}, \mathbf{h})$ is no less than threshold T. However, as we usually do not know the correct causal mechanism, we end up averaging over all such hypotheses, leading to the expectation $Pr(d \mid \mathbf{s})$, and then making a decision depending on whether $Pr(d \mid \mathbf{s}) \geq T$.

[2] Namely, using a Dirichlet prior on the CPT of S in the original model $X \rightarrow S$ would basically assume independent false positive and false negative rates.

Our interest now is in quantifying our confidence in such a decision given that we do not know the correct causal mechanism. Since $Q(\mathbf{H})$ is a random variable, we propose to quantify such a confidence using the following, which we call the *same-decision probability:*

$$\mathcal{P}(Q(\mathbf{H}) \geq T) = \sum_{\mathbf{h}} [Q(\mathbf{h}) \geq T] Pr(\mathbf{h} \mid \mathbf{s}), \quad (5)$$

where $[Q(\mathbf{h}) \geq T]$ is an indicator function that is 1 if $Q(\mathbf{h}) \geq T$ and 0 otherwise. This is the *probability that we would have made the same decision had we known the correct causal mechanisms governing the sensor readings.*

Consider now Equation 5 in relation to Equation 4. Both equations define expectations with respect to the same distribution $Pr(\mathbf{h} \mid \mathbf{s})$. In Equation 4, the resulting expectation is the probability $Pr(d \mid \mathbf{s})$. In Equation 5, the expectation is the same-decision probability. One key difference between the two expectations is that the one in Equation 4 is invariant to the specific distributions used for variables H, as long as these distributions satisfy Equations 1 and 2. However, the expectation in Equation 5 — that is, the same-decision probability — is indeed dependent on the specific distributions used for variables H.

Consider now Table 1, which corresponds to two positive sensor readings in Figure 1. Assuming a threshold of $T = 0.60$, a decision is confirmed given that we have $Pr(D{=}+ \mid S_1{=}+, S_2{=}+) = 0.880952 \geq T$. We make the same decision, however, in only four of the nine causal mechanisms. These probabilities add up to 0.975; hence, the same-decision probability is 0.975. Consider now Table 2, which corresponds to two conflicting sensor readings. The decision is also confirmed here since $Pr(D{=}+ \mid S_1{=}+, S_2{=}-) = 0.631147 \geq T$. Again, we make the same decision in four causal mechanisms, although they are now less likely scenarios. The same-decision probability is only 0.601229, suggesting a smaller confidence in the decision in this case.

Although computing the same-decision probability may be computationally difficult, the one-sided Chebyshev inequality can be used to bound it. According to this inequality, if V is a random variable with expectation $\mathrm{E}[\,V\,] = \mu$ and variance $\mathrm{Var}[\,V\,] = \sigma^2$, then for any $a > 0$:

$$\mathcal{P}(V \geq \mu - a) \geq 1 - \frac{\sigma^2}{\sigma^2 + a^2}$$

Recall now that the probability $Pr(d \mid \mathbf{s})$ is an expectation $\mathrm{E}[\,Q(\mathbf{H})\,]$ with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{s})$, where $Q(\mathbf{h}) = Pr(d \mid \mathbf{h}, \mathbf{s})$. Suppose that $\mathrm{E}[\,Q(\mathbf{H})\,] \geq T$ and a decision has been confirmed accordingly. The same-decision probability is simply the probability of $Q(\mathbf{H}) \geq T$, where $Q(\mathbf{H})$ is a random variable. Using the Chebyshev inequality, we get the following bound on the same-decision probability:

$$\mathcal{P}(Q(\mathbf{H}) \geq T) \geq 1 - \frac{\mathrm{Var}[\,Q(\mathbf{H})\,]}{\mathrm{Var}[\,Q(\mathbf{H})\,] + [Pr(d|\mathbf{s}) - T]^2}$$

Suppose now that $\mathrm{E}[\,Q(\mathbf{H})\,] \leq T$ and a decision has been confirmed accordingly. The same-decision probability in this case is the probability of $Q(\mathbf{H}) \leq T$. Using the Chebyshev inequality now to bound $\mathcal{P}(V \leq \mu + a)$, we get the same bound for the same-decision probability $\mathcal{P}(Q(\mathbf{H}) \leq T)$. To compute these bounds, we need the variance $\mathrm{Var}[\,Q(\mathbf{H})\,]$. We provide an algorithm for this purpose in the next section.

6 Computing the Variance

Let $\mathbf{S}$ and $\mathbf{H}$ be any two disjoint sets of variables in a Bayesian network, with neither set containing variable D. The probability $Pr(d \mid \mathbf{s})$ can be interpreted as an expectation of $Q(\mathbf{h}) = Pr(d \mid \mathbf{s}, \mathbf{h})$ with respect to a distribution $Pr(\mathbf{h} \mid \mathbf{s})$. We propose in this section a general algorithm for computing the variance of such expectations.

Consider now the variance:

$$\begin{aligned}\mathrm{Var}[\,Q(\mathbf{H})\,] &= \mathrm{E}[\,Q(\mathbf{H})^2\,] - \mathrm{E}[\,Q(\mathbf{H})\,]^2 \\ &= \Big[\sum_{\mathbf{h}} Pr(d \mid \mathbf{h}, \mathbf{s})^2 Pr(\mathbf{h} \mid \mathbf{s})\Big] - Pr(d \mid \mathbf{s})^2\end{aligned}$$

We need two quantities to compute this variance. First, we need the quantity $Pr(d \mid \mathbf{s})$, which can be computed using standard algorithms for Bayesian network inference, such as variable elimination (Zhang and Poole, 1996; Dechter, 1996; Darwiche, 2009). The other

Algorithm 1 Variance by Variable Elimination

input:
$\mathcal{N}$: a Bayes net with distribution Pr
D, d: decision variable and decision state
$\mathbf{S}, \mathbf{s}$: set of sensor variables and readings
$\mathbf{H}$: set of health variables for sensors $\mathbf{S}$

output: a factor that contains $\sum_{\mathbf{h}} \frac{Pr(d,\mathbf{h},\mathbf{s})^2}{Pr(\mathbf{h},\mathbf{s})}$

main:
$\mathcal{S}_1 \leftarrow$ factors of $\mathcal{N}$ under observations $d, \mathbf{s}$
$\mathcal{S}_2 \leftarrow$ factors of $\mathcal{N}$ under observations $\mathbf{s}$
$\mathbf{Y} \leftarrow$ all variables in $\mathcal{N}$ but variables $\mathbf{H}$
$\pi \leftarrow$ an ordering of variables $\mathbf{Y}$
$\mathcal{S}_1 \leftarrow \textsc{ve}(\mathcal{S}_1, \mathbf{Y}, \pi)$
$\mathcal{S}_2 \leftarrow \textsc{ve}(\mathcal{S}_2, \mathbf{Y}, \pi)$
$\mathcal{S} \leftarrow \{\chi_a \mid \chi_a = \frac{\phi_a^2}{\psi_a}$ for $\phi_a \in \mathcal{S}_1, \psi_a \in \mathcal{S}_2\}$
$\pi \leftarrow$ an ordering of variables $\mathbf{H}$
$\mathcal{S} \leftarrow \textsc{ve}(\mathcal{S}, \mathbf{H}, \pi)$
return $\prod_{\psi \in \mathcal{S}} \psi$

quantity involves a summation over instantiations $\mathbf{h}$. Naively, we could compute this sum by simply enumerating over all instantiations $\mathbf{h}$, using again the variable elimination algorithm to compute the relevant quantities for each instantiation $\mathbf{h}$. However, the number of instantiations $\mathbf{h}$ is exponential in the number of variables in $\mathbf{H}$ and will thus be impractical when the number of such variables is large enough.

However, with a suitably augmented variable elimination algorithm, we can compute this summation more efficiently, and thus the variance. First, consider the following alternative form for the summation:

$$\sum_{\mathbf{h}} Pr(d \mid \mathbf{h}, \mathbf{s})^2 Pr(\mathbf{h} \mid \mathbf{s}) = \frac{1}{Pr(\mathbf{s})} \sum_{\mathbf{h}} \frac{Pr(d,\mathbf{h},\mathbf{s})^2}{Pr(\mathbf{h},\mathbf{s})}.$$

Note that the term $Pr(\mathbf{s})$ is readily available using variable elimination and can be computed together with $Pr(d \mid \mathbf{s})$. Hence, we just need the sum $\sum_{\mathbf{h}} \frac{Pr(d,\mathbf{h},\mathbf{s})^2}{Pr(\mathbf{h},\mathbf{s})}$, which, as we show next, can be computed using an augmented version of variable elimination.[3]

Let $\mathbf{Y}$ denote all variables in the Bayesian network excluding variables $\mathbf{H}$. If we set evidence $\mathbf{s}$ and use variable elimination to sum out variables $\mathbf{Y}$, we get a set of factors that represents the following distribution: $Pr(\mathbf{H}, \mathbf{s}) = \prod_a \psi_a(\mathbf{X}_a)$. Here, ψ_a are the factors remaining from variable elimination after having eliminated variables $\mathbf{Y}$.

We can similarly run the variable elimination algorithm with evidence $\mathbf{s}, d$ to obtain a set of factors whose product represents the following distribution: $Pr(\mathbf{H}, d, \mathbf{s}) = \prod_a \phi_a(\mathbf{X}_a)$. Using the same variable ordering when eliminating variables $\mathbf{Y}$, we can ensure a one-to-one correspondence between factors in both factorizations: each pair of factors ψ_a and ϕ_a will be over the same set of variables $\mathbf{X}_a$ for a given index a. For each instantiation $\mathbf{h}, d, \mathbf{s}$, we have

$$\frac{Pr(\mathbf{h}, d, \mathbf{s})^2}{Pr(\mathbf{h}, \mathbf{s})} = \prod_a \frac{\phi_a(\mathbf{x}_a)^2}{\psi_a(\mathbf{x}_a)},$$

where $\mathbf{x}_a$ is an instantiation of variables $\mathbf{X}_a$ consistent with instantiation $\mathbf{h}, d, \mathbf{s}$. We now compute a new set of factors $\chi_a(\mathbf{X}_a) = \frac{\phi_a(\mathbf{X}_a)}{\psi_a(\mathbf{X}_a)}$ and run the variable elimination algorithm a third time to eliminate variables $\mathbf{H}$ from the factors $\chi_a(\mathbf{X}_a)$. The result will be a trivial factor that contains the quantity of interest.[4]

Algorithm 2 Variable Elimination [$\textsc{ve}$]

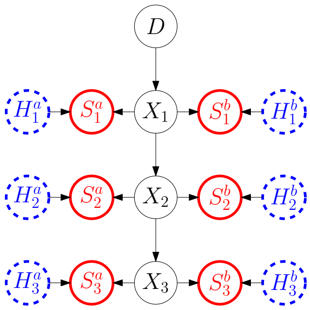
input:
$\mathcal{S}$: set of factors
$\mathbf{Y}$: variables to eliminate in factor set $\mathcal{S}$
π: ordering of variable $\mathbf{Y}$

output: factor set where variables $\mathbf{Y}$ are eliminated

main:
for $i = 1$ to length of order π **do**
  $\mathcal{S}_i \leftarrow$ factors in $\mathcal{S}$ containing variable $\pi(i)$
  $\psi_i \leftarrow \sum_{\pi(i)} \prod_{\psi \in \mathcal{S}_i} \psi$
  $\mathcal{S} \leftarrow \mathcal{S} - \mathcal{S}_i \cup \{\psi_i\}$
return $\mathcal{S}$

[3] Formally, our summation should be over instantiations $\mathbf{h}$ where $Pr(\mathbf{h}, \mathbf{s}) > 0$. Note that if $Pr(\mathbf{h}, \mathbf{s}) = 0$ then $Pr(d, \mathbf{h}, \mathbf{s}) = 0$. Hence, if we define $x/0 = 0$, then our summation is simply over all instantiations $\mathbf{h}$. In Algorithm 1, we thus define factor division such that $\phi_a(\mathbf{x}_a)^2/\psi_a(\mathbf{x}_a) = 0$ when $\psi_a(\mathbf{x}_a) = 0$.

[4] According to the formulation of variable elimination in (Darwiche, 2009), a trivial factor is a factor over the empty set of variables and contains one entry. It results from eliminating all variables from a set of factors.



D	$Pr(D)$
$+$	0.25
$-$	0.75

D	X_1	$Pr(X_1 \mid D)$
$+$	$+$	0.9
$+$	$-$	0.1
$-$	$+$	0.1
$-$	$-$	0.9

Figure 2: All sensors have $f_p = f_n = .05$. Variables X_i all have the same CPTs.

Algorithm 1 provides pseudo-code that implements this procedure. Note that on Line 7, there is a one-to-one correspondence between the factors of $\mathcal{S}_1$ and $\mathcal{S}_2$ as we have a one-to-one correspondence between the factors passed to $\text{VE}(\mathcal{S}_1, \mathbf{Y}, \pi)$ and $\text{VE}(\mathcal{S}_2, \mathbf{Y}, \pi)$, and since each call eliminates the same set of variables using the same variable order. Algorithm 1 must eliminate variables $\mathbf{H}$ last, so the complexity of the algorithm is exponential in the *constrained treewidth* (Darwiche, 2009). This is analogous to the complexity of variable elimination for computing MAP, where variables $\mathbf{H}$ are MAP variables (Park and Darwiche, 2004).

7 Examples

Consider the Bayesian network in Figure 2, which depicts a chain D, X_1, X_2, X_3 with two sensors S_i^a and S_i^b attached to each node X_i. Our goal here is to make a decision depending on whether $Pr(D{=}+ \mid \mathbf{s}) \geq T$ for some sensor reading $\mathbf{s}$ and threshold $T = 0.5$. We will next consider a number of sensor readings, each leading to the same decision but a different same-decision probability. Our purpose is to provide concrete examples of this probability, and to show that it can discriminate among sensor readings that not only lead to the same decision, but also under *very similar probabilities* for the hypothesis of interest. The examples will also shed more light on the tightness of the one-sided Chebyshev bound proposed earlier.

Our computations in this section assume the independence between the mechanisms governing false positives and negatives, which is needed to induce a distribution over causal mechanisms (as in Section 3). We also provide the results under the second assumption proposed where the lying causal mechanism has zero probability (in brackets). As we discussed earlier, we expect the two results to be very close since the false positive and negative rates are small, which is confirmed empirically here.

We start by observing that $Pr(D{=}+) = 25\%$. Suppose now that we have a positive reading for sensor S_2^a. We now have the *hypothesis probability* $Pr(D{=}+ \mid S_2^a{=}+) = 55.34\%$ and the decision is confirmed given our threshold. The same-decision probability is 86.19%. From now on, we will say that our *decision confidence* is 86.19% in this case.

The following table depicts what happens when we obtain another positive sensor reading.

	Scenario 1	Scenario 2
sensor readings	$S_2^a{=}+$	$S_2^a{=}+$ $S_2^b{=}+$
hyp. prob.	55.34%	60.01%
dec. conf.	86.19%[85.96%]	99.22%[99.19%]

Note how the decision confidence has increased significantly even though the change in the hypothesis probability is relatively modest. The following table depicts a scenario when we have two more sensor readings that are conflicting.

	Scenario 2	Scenario 3
readings	$S_2^a{=}+,\ S_2^b{=}+$	$S_1^a{=}+,\ S_1^b{=}-$ $S_2^a{=}+,\ S_2^b{=}+$
hyp. prob.	60.01%	60.01%
dec. conf.	99.22%[99.19%]	79.97%[80.07%]

Note how the new readings keep the hypothesis probability the same, but reduce the decision confidence significantly. This is mostly due to raising the probability of some causal mechanism under which we would make a different decision. The following table depicts a conflict between a different pair of sensors.

	Scenario 3	Scenario 4
readings	$S_1^a{=}+,\ S_1^b{=}-$ $S_2^a{=}+,\ S_2^b{=}+$	$S_2^a{=}+,\ S_2^b{=}+$ $S_3^a{=}+,\ S_3^b{=}-$
hyp. prob.	60.01%	60.01%
dec. conf.	79.97%[80.07%]	99.48%[99.48%]

In this case, the sensor conflict increases the same-decision probability slightly (from 99.22%

in Scenario 2 to 99.48%).[5] The next example shows what happens when we get two negative readings but at different sensor locations.

	Scenario 5	Scenario 6
readings	$S_1^a = -$, $S_1^b = -$ $S_2^a = +$, $S_2^b = +$	$S_2^a = +$, $S_2^b = +$ $S_3^a = -$, $S_3^b = -$
hyp. prob.	4.31%	57.88%
dec. conf.	98.73%[98.70%]	95.25%[95.23%]

When the negative sensors are close to the hypothesis, they reduce the hypothesis probability significantly below the threshold, leading to a high confidence decision. When the readings are further away from the hypothesis (and dominated by the two positive readings), they reduce the hypothesis probability, yet keep it above the threshold. The decision confidence is also reduced, but remains relatively high. Finally, consider the table below which compares the decision confidence, the bound on the confidence, and the variance used to compute the bound.

Scenario	confidence		bound	variance
1	86.19%	$\geq$	15.53%	$1.54 \cdot 10^{-2}$
2	99.22%	$\geq$	90.50%	$1.05 \cdot 10^{-3}$
3	79.97%	$\geq$	11.05%	$8.06 \cdot 10^{-2}$
4	99.48%	$\geq$	88.30%	$1.32 \cdot 10^{-3}$
5	98.73%	$\geq$	98.02%	$4.22 \cdot 10^{-3}$
6	95.25%	$\geq$	34.73%	$1.16 \cdot 10^{-2}$

Note that our decision confidence is high when our bound on the same-decision probability is high. Moreover, the one-sided Chebyshev inequality may provide only weak bounds, which may call for exact computation of the same-decision probability. We computed this quantity through exhaustive enumeration here, yet an algorithm that is exponential only in the constrained treewidth could open new possibilities for reasoning about threshold-based decisions.

8 Conclusion

We considered in this paper the robustness of decisions based on probabilistic thresholds under noisy sensor readings. In particular, we suggested a confidence measure for threshold-based decisions which corresponds to the probability that one would have made the same decision if one had knowledge about a sensor's causal mechanisms. We used the one-sided Chebyshev inequality to bound this probability, which requires computing the variance of a conditional probability with respect to the marginal distribution over a subset of network variables. We also proposed a variable elimination algorithm for computing this variance, whose complexity is exponential only in the constrained treewidth of the given network. We finally provided a number of concrete examples showing the utility of our proposed confidence measure in quantifying the robustness of threshold-based decisions.

[5] Knowing that sensor S_3^b is lying, or that S_3^a is telling the truth, is enough to confirm our decision given our threshold. The conflicting sensor readings thus introduce new scenarios under which the decision is confirmed, although these scenarios are very unlikely.

References

Hei Chan. 2009. *Sensitivity Analysis of Probabilistic Graphical Models: Theoretical Results and Their Applications on Bayesian Network Modeling and Inference.* VDM Verlag.

Adnan Darwiche. 2009. *Modeling and Reasoning with Bayesian Networks.* Cambridge University Press.

Rina Dechter. 1996. Bucket elimination: A unifying framework for probabilistic inference. In *UAI*, pages 211–219.

Marek J. Druzdzel and Herbert A. Simon. 1993. Causality in Bayesian belief networks. In *UAI*, pages 3–11.

Walter Hamscher, Luca Console, and Johan de Kleer. 1992. *Readings in model-based diagnosis.* MK.

David Heckerman, John S. Breese, and Koos Rommelse. 1995a. Decision-theoretic troubleshooting. *CACM*, 38(3):49–57.

David Heckerman, Dan Geiger, and David Maxwell Chickering. 1995b. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243.

Johan Kwisthout and Linda C. van der Gaag. 2008. The computational complexity of sensitivity analysis and parameter tuning. In *UAI*, pages 349–356.

Michael L. Littman, Judy Goldsmith, and Martin Mundhenk. 1998. The computational complexity of probabilistic planning. *JAIR*, 9:1–36.

James Park and Adnan Darwiche. 2004. Complexity results and approximation strategies for MAP explanations. *JAIR*, 21:101–133.

Judea Pearl. 2009. *Causality: Models, Reasoning and Inference.* Cambridge University Press, 2nd edition.

Nevin Lianwen Zhang and David Poole. 1996. Exploiting causal independence in Bayesian network inference. *JAIR*, 5:301–328.