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

Arthur Choi[a], Yexiang Xue[b,1], Adnan Darwiche[c]

[a] *Computer Science Department, University of California, Los Angeles*
[b] *Department of Computer Science, Cornell University*
[c] *Computer Science Department, University of California, Los Angeles*

Abstract

We consider in this paper the robustness of decisions based on probabilistic thresholds. To this effect, we propose the *same-decision probability* as a query that can be used as a confidence measure for threshold-based decisions. More specifically, the same-decision probability is the probability that we would have made the same threshold-based decision, had we known the state of some hidden variables pertaining to our decision.

We study a number of properties about the same-decision probability. First, we analyze its computational complexity. We then derive a bound on its value, which we can compute using a variable elimination algorithm that we propose. Finally, we consider decisions based on noisy sensors in particular, showing through examples that the same-decision probability can be used to reason about threshold-based decisions in a more refined way.

Keywords: Bayesian networks, robust decision making, computational complexity of reasoning, sensitivity analysis, exact inference, variable elimination

1. Introduction

There has been an increased interest recently in providing assurances on the results of probabilistic reasoning systems. Clear examples come from the

Email addresses: `aychoi@cs.ucla.edu` (Arthur Choi), `yexiang@cs.cornell.edu` (Yexiang Xue), `darwiche@cs.ucla.edu` (Adnan Darwiche)

[1]Part of this research was conducted while the author was a visiting student at the University of California, Los Angeles.

many results on sensitivity analysis, which is concerned with the sensitivity of probabilistic queries with respect to changes in the model parameters; see, e.g., Chan (2009), van der Gaag et al. (2007), van der Gaag and Coupé (1999), Charitos and van der Gaag (2006) and Kwisthout and van der Gaag (2008). For example, some of 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 made under 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 a probability $Pr(d \mid \mathbf{e})$ surpasses a given threshold T, where $\mathbf{e}$ represents evidence or observations. This is the prototypical scenario in which Bayesian networks are employed to support decision making in practice, for example, in domains such as diagnosis (Pauker and Kassirer, 1980) and (binary) classification (Friedman et al., 1997).[2]

The same-decision probability is based on a few simple ideas. Let $\mathbf{H}$ be a subset of the unobserved variables that pertain to the hypothesis d upon which our decision is based. For example, the variables $\mathbf{H}$ may represent the hidden state of a system, such as health modes of components in a diagnostic application. The variables $\mathbf{H}$ could also represent observations yet to be made, such as medical tests. Now, if we knew the true states of our variables $\mathbf{H}$, we would stand to make a better informed decision based on the probability $Pr(d \mid \mathbf{e}, \mathbf{h})$. As it stands, the probability $Pr(d \mid \mathbf{e})$ can already be viewed as the expectation of $Pr(d \mid \mathbf{e}, \mathbf{h})$ with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{e})$. Now, different scenarios $\mathbf{h}$ may confirm or contradict our decision based on the probability $Pr(d \mid \mathbf{e}, \mathbf{h})$, but these scenarios may be likely or unlikely, according to $Pr(\mathbf{h} \mid \mathbf{e})$. The same-decision probability is then the probability that we would have made the same threshold-based decision, had we known the true state $\mathbf{h}$ of our hidden variables $\mathbf{H}$.

We show a number of results about this proposed quantity. First, we formally define the same-decision probability, and then analyze its compu-

[2]In principle, analogous definitions of the "same-decision probability" can be proposed for decisions based on utilities, or decisions that are not necessarily binary (yes/no). For the purposes of introducing the "same-decision probability" as a new query for Bayesian networks, we restrict our attention in this paper to the simple and transparent case of threshold-based decisions.

tational complexity, showing that the same-decision probability is a $\mathrm{PP}^{\mathrm{PP}}$–complete problem. Second, we propose a bound on the same-decision probability using the one-sided Chebyshev inequality, which requires only the variance of $Pr(d \mid \mathbf{e}, \mathbf{h})$ with respect to the distribution $Pr(\mathbf{h} \mid \mathbf{e})$. 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 further consider the same-decision probability in scenarios where we are making threshold-based decisions based on the readings of noisy sensors. In particular, we propose to explicate the causal mechanisms that govern the behaviors of noisy sensors. We can then consider the probability that we would have made the same threshold-based decision, had we known the latent causal mechanisms that led to our sensor readings. 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 under noisy sensor readings. In particular, we illustrate how the same-decision probability is able to distinguish scenarios that are otherwise indistinguishable, based on the probability $Pr(d \mid \mathbf{e})$ alone.

2. An Introductory Example

In the rest of the paper, we use standard notation for variables and their instantiations. In particular, variables are denoted by upper case letters (X) and their instantiations by lower case letters (x). Moreover, sets of variables are denoted by bold upper case letters ($\mathbf{X}$) and their instantiations by bold lower case letters ($\mathbf{x}$).

Before we formally define the same-decision probability, we first describe a simple example, to highlight the basic ideas that underlie the same-decision probability as a way to quantify the robustness of threshold-based decisions (van der Gaag and Coupé, 1999; Charitos and van der Gaag, 2006). Again, such decisions are the prototypical context in which Bayesian networks are employed to support decision making in practice. These include classical applications such as diagnosis (Hamscher et al., 1992), troubleshooting (Heckerman et al., 1995a), classification (Friedman et al., 1997), and probabilistic planning (Littman et al., 1998). For example, in health diagnosis, physicians are commonly put in situations where they must commit to performing a test or administering a treatment. Based on their (possibly subjective) belief surpassing some (possibly subjective) threshold (Pauker and Kassirer, 1980),

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

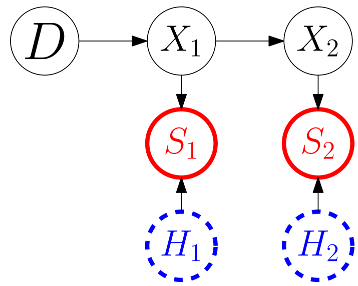


Figure 1: A simple Bayesian network, under sensor readings $\{S_1 = +, S_2 = +\}$. Here $(+)$ indicates a positive sensor reading for a sensor variable S_i, or a positive outcome for a decision variable D or auxiliary variable X_i; similarly, $(-)$ indicates a negative reading or outcome. Variables H_1 and H_2 represent the health of sensors S_1 and S_2. On the left is the posterior on the decision variable D. Network CPTs are given in Figure 2.

a physician will commit to one of these choices. As another example, in systems diagnosis, a diagnostician, in the process of troubleshooting, must decide whether or not they should perform one of many tests, or stop the process of testing and perform a repair (or some other intervention) (Lu and Przytula, 2006; Agosta et al., 2008, 2010). Again, this decision is typically made based on a diagnostician's beliefs about the health state of the system, and the extent to which they are certain or uncertain about it. In this section, we highlight an example of a threshold-based decision made under a simple but generally applicable context, where observations are given by noisy sensor readings. This is also a scenario which we shall revisit in more depth in Section 6.

Consider now the Bayesian network in Figure 1, which models a scenario involving a variable D of interest, and two noisy sensors S_1 and S_2 that bear (indirectly) on a hypothesis d. The probability $Pr(d \mid s_1, s_2)$ then represents a belief in the hypothesis d, given sensor readings s_1, s_2. We want to use this Bayesian network to support a decision on the basis that this belief exceeds a certain threshold, $Pr(d \mid s_1, s_2) \geq T$. Figure 1 shows a particular reading of the 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.

Note that in Figure 1 (and further Figure 2), we modeled the health of our sensors through variables H_1 and H_2, which dictate the behavior of our sensors. Suppose we knew the sensors' state of health, in which case, we would know how to interpret the readings of our sensors. For example, we

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

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

X_1	X_2	$Pr(X_2 \mid X_1)$
$+$	$+$	0.9
$+$	$-$	0.1
$-$	$+$	0.1
$-$	$-$	0.9

H_i	X_i	S_i	$Pr(S_i \mid H_i, X_i)$
t	$+$	$+$	1.0
t	$-$	$+$	0.0
p	$+$	$+$	1.0
p	$-$	$+$	1.0
n	$+$	$+$	0.0
n	$-$	$+$	0.0
l	$+$	$+$	0.0
l	$-$	$+$	1.0

H_i	$Pr(H_i)$
t	0.81
p	0.09
n	0.09
l	0.01

Figure 2: The CPTs for the Bayesian network given in Figure 1. Note that for the CPTs of variables S_i, only the lines for the case $S_i\!=\!+$ are given, since $Pr(S_i\!=\!-|H_i, X_i) = 1 - Pr(S_i\!=\!+|H_i, X_i)$. Moreover, we model the following health states for our sensors: the state $H_i\!=\!\mathsf{t}$ says that the sensor is truthful, the state $H_i\!=\!\mathsf{l}$ says the sensor is lying, the state $H_i\!=\!\mathsf{p}$ says the sensor is stuck with a positive reading, and the state $H_i\!=\!\mathsf{n}$ says the sensor is stuck with a negative reading. We consider noisy sensors further in Section 6.

would know whether we could trust their readings, or otherwise ignore some of them. We would then be able to, and would prefer to, make a better informed decision based on the probability $Pr(d \mid s_1, s_2, \mathbf{h})$ instead of the probability $Pr(d \mid s_1, s_2)$, where $\mathbf{h}$ represents the state of our variables H_1 and H_2, for example, $\mathbf{h} = \{H_1\!=\!\mathsf{t}, H_2\!=\!\mathsf{t}\}$.

Consider Table 1, which enumerates all of the possible health states $\mathbf{h}$ of our example, where we have nine scenarios with non-zero probability. 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 actual health states that dictate the readings of

Table 1: Scenarios $\mathbf{h}$ for sensor readings $\mathbf{e} = \{S_1 = +, S_2 = +\}$ for the network in Figure 1, where $\mathbf{H} = \{H_1, H_2\}$. Cases above the threshold $T = 0.6$ are in bold.

$\mathbf{h}$	H_1	H_2	$Pr(\mathbf{h} \mid s_1, s_2)$	$Pr(d \mid s_1, s_2, \mathbf{h})$
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

our sensors. For this example, this probability is:

$$0.781071 + 0.096429 + 0.096429 + 0.001071 = 0.975$$

indicating a relatively robust decision.

3. Same-Decision Probability

Suppose we have a Bayesian network conditioned on evidence $\mathbf{e}$, and that we are interested in making a decision depending on whether the probability of some hypothesis d surpasses some threshold T. There may be hidden, latent, or otherwise unobserved variables $\mathbf{H}$ that pertain to our hypothesis d. If we did have access to the true joint state $\mathbf{h}$, we would certainly want to make a better informed decision based on whether the probability $Pr(d \mid \mathbf{e}, \mathbf{h})$ surpasses the threshold T. In the absence of this knowledge, we can still reason about the possible scenarios $\mathbf{h}$.

Consider the fact that different scenarios $\mathbf{h}$ may confirm or contradict our decision based on the probability $Pr(d \mid \mathbf{e}, \mathbf{h})$. These scenarios may be likely or unlikely, according to $Pr(\mathbf{h} \mid \mathbf{e})$. However, what if the scenarios $\mathbf{h}$ that contradict our decision, where $Pr(d \mid \mathbf{e}, \mathbf{h}) < T$, have a low probability $Pr(\mathbf{h} \mid \mathbf{e})$? In this case, we have a degree of confidence in our original decision based on $Pr(d \mid \mathbf{e}) \geq T$, in the sense that even if we were able to discover the state of our unobserved variables $\mathbf{H}$, it is unlikely that we would have made

Table 2: Scenarios $\mathbf{h}$ for sensor readings $\mathbf{e} = \{S_1 = +, S_2 = -\}$ for the network in Figure 1, where $\mathbf{H} = \{H_1, H_2\}$. Cases above the threshold $T = 0.6$ are in bold.

$\mathbf{h}$	H_1	H_2	$Pr(\mathbf{h} \mid s_1, s_2)$	$Pr(d \mid s_1, s_2, \mathbf{h})$
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

a different decision. The same-decision probability is then the *probability that we would have made the same decision had we known the states of our unobserved variables* $\mathbf{H}$.

Definition 1. *Let $\mathcal{N}$ be a Bayesian network that is conditioned on evidence $\mathbf{e}$, where we are further given a hypothesis d, a threshold T, and a set of unobserved variables $\mathbf{H}$. Suppose we are making a decision that is confirmed by the threshold $Pr(d \mid \mathbf{e}) \geq T$. The **same-decision probability** in this scenario is*

$$\mathcal{P}(Pr(d \mid \mathbf{e}, \mathbf{H}) \geq T) = \sum_{\mathbf{h}} [Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T] Pr(\mathbf{h} \mid \mathbf{e}), \qquad (1)$$

where we have the indicator function

$$[Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T] = \begin{cases} 1 & \text{if } Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T \\ 0 & \text{otherwise.} \end{cases}$$

For the remainder of the paper, we shall denote the same-decision probability by $\mathcal{P}(Q(\mathbf{H}) \geq T)$, for reasons that we discuss below.

Consider the following observation. When we are making a decision based on whether $Pr(d \mid \mathbf{e}) \geq T$, even if the state $\mathbf{h}$ of our variables $\mathbf{H}$ is unknown, we are in fact averaging over all possible scenarios $\mathbf{h}$ when we make a decision.

That is,

$$Pr(d \mid \mathbf{e}) = \sum_{\mathbf{h}} Pr(d \mid \mathbf{e}, \mathbf{h}) Pr(\mathbf{h} \mid \mathbf{e}) \qquad (2)$$
$$= \sum_{\mathbf{h}} Q(\mathbf{h}) Pr(\mathbf{h} \mid \mathbf{e}).$$

Here, we denote $Pr(d \mid \mathbf{e}, \mathbf{h})$ using $Q(\mathbf{h})$ to emphasize our view on the probability $Pr(d \mid \mathbf{e})$ as an expectation $\mathrm{E}[\, Q(\mathbf{H}) \,]$ with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{e})$ over unobserved variables $\mathbf{H}$. We remark that the same-decision probability $\mathcal{P}(Q(\mathbf{H}) \geq T)$ is also an expectation, as in Equation 2. We view Equation 1, however, as the expected decision based on $Pr(d \mid \mathbf{e}, \mathbf{h})$, with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{e})$ over unobserved variables $\mathbf{H}$.

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 instantiations $\mathbf{h}$. 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 scenarios $\mathbf{h}$, 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.

The following theorem now highlights the complexity of computing the same decision probability.

Theorem 1. *The problem of deciding whether the same-decision probability is greater than some given probability p is* $\mathrm{PP}^{\mathrm{PP}}$*–complete.*

This complexity result indicates that computing the same-decision probability is computationally quite challenging, even more so than computing MAP in Bayesian networks, for example, which is only $\mathrm{NP}^{\mathrm{PP}}$–complete (Park and Darwiche, 2004). In particular, the complexity classes NP, PP, and the corresponding classes assuming a PP oracle, are related in the following way:

$$\mathrm{NP} \subseteq \mathrm{PP} \subseteq \mathrm{NP}^{\mathrm{PP}} \subseteq \mathrm{PP}^{\mathrm{PP}}$$

where the complexity class $\mathrm{NP}^{\mathrm{PP}}$ already contains the entire Polynomial Hierarchy (Toda, 1991). The proof of Theorem 1 is included in the Appendix, together with some further comments on the complexity class $\mathrm{PP}^{\mathrm{PP}}$ (Allender and Wagner, 1990).

Since the same-decision probability is a natural problem that is of practical interest, and given that it is $\mathrm{PP}^{\mathrm{PP}}$–complete, studying the same-decision probability could help analyze the complexity of other reasoning problems for Bayesian networks that may also be $\mathrm{PP}^{\mathrm{PP}}$–complete (as MAP has proved useful for analyzing $\mathrm{NP}^{\mathrm{PP}}$–complete problems). See Umans (2000) for natural problems in the Polynomial Hierarchy, and also Kwisthout (2009) for natural problems in probabilistic reasoning, for a variety of other complexity classes.

4. Approximating the Same-Decision Probability

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 $Pr(d \mid \mathbf{e})$ is an expectation $\mathrm{E}[\,Q(\mathbf{H})\,]$ with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{e})$, where $Q(\mathbf{h}) = Pr(d \mid \mathbf{e}, \mathbf{h})$. 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:

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

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.

For an example of our bound, consider again the example from Figure 1 and Table 1. We have mean $\mathrm{E}[\,Q(\mathbf{H})\,] = 0.880952$ and variance

$\mathrm{Var}[\,Q(\mathbf{H})\,] = 0.005823$. We can thus state that $\mathcal{P}(Q(\mathbf{H}) \geq 0.6) \geq 0.931289$. Recall that the exact same-decision probability here is .975. On the other hand, if we take the same network, but are given conflicting sensor readings $\mathbf{e} = \{S_1 = +, S_2 = -\}$, as in Table 2, then we have mean $\mathrm{E}[\,Q(\mathbf{H})\,] = 0.631147$ and variance $\mathrm{Var}[\,Q(\mathbf{H})\,] = 0.114755$. The mean is much closer to our threshold, and our variance is much higher than when our readings were consistent. We can only state that $\mathcal{P}(Q(\mathbf{H}) \geq 0.6) \geq 0.008383$. Recall that the same-decision probability is 0.601229 for this example, so the Chebyshev inequality provides a weak bound here. However, the more extreme the bound is, the more confident we can be about its tightness.

5. Computing the Variance

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

Consider now the variance:

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

We need two quantities to compute this variance. First, we need the quantity $Pr(d \mid \mathbf{e})$, 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 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 this number is too large.

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:

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

Note that the term $Pr(\mathbf{e})$ is readily available using variable elimination and can be computed together with $Pr(d \mid \mathbf{e})$. Hence, we just need the sum $\sum_{\mathbf{h}} \frac{Pr(d,\mathbf{e},\mathbf{h})^2}{Pr(\mathbf{e},\mathbf{h})}$, 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 $\mathbf{e}$ and use variable elimination to sum out variables $\mathbf{Y}$, we get a set of factors that represents the following distribution:

$$Pr(\mathbf{H}, \mathbf{e}) = \prod_a \psi_a(\mathbf{X}_a).$$

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 $d, \mathbf{e}$ to obtain a set of factors whose product represents the following distribution:

$$Pr(\mathbf{H}, d, \mathbf{e}) = \prod_a \phi_a(\mathbf{X}_a).$$

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 $\mathbf{h}, d, \mathbf{e}$, we then have

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

where $\mathbf{x}_a$ is an instantiation of variables $\mathbf{X}_a$ consistent with instantiation $\mathbf{h}, d, \mathbf{e}$. We now compute a new set of factors

$$\chi_a(\mathbf{X}_a) = \frac{\phi_a(\mathbf{X}_a)^2}{\psi_a(\mathbf{X}_a)}$$

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]

[3]Formally, our summation should be over instantiations $\mathbf{h}$ where $Pr(\mathbf{e}, \mathbf{h}) > 0$. Note that if $Pr(\mathbf{e}, \mathbf{h}) = 0$ then $Pr(d, \mathbf{e}, \mathbf{h}) = 0$. Hence, if we define $x/0 = 0$, then our summation

Algorithm 1 Variance by Variable Elimination

input:

$\mathcal{N}$: a Bayes net with distribution Pr

D, d: a decision variable and a decision state

$\mathbf{E}, \mathbf{e}$: a set of observed variables $\mathbf{E}$ and evidence $\mathbf{e}$

$\mathbf{H}$: a set of unobserved variables $\mathbf{H}$

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

main:

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

Algorithm 2 Variable Elimination [VE]

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

output: a set of factors 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 $\mathcal{S}$

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 $\textsc{ve}(\mathcal{S}_1, \mathbf{Y}, \pi)$ and $\textsc{ve}(\mathcal{S}_2, \mathbf{Y}, \pi)$, 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).

We finally stress that the algorithm we proposed in this section has applicability beyond that of bounding the same-decision probability. In particular, any conditional probability of the form $Pr(d \mid \mathbf{e})$, where D is a network variable and $\mathbf{E}$ is a set of network variables, can always be interpreted as an expectation with respect to the distribution $Pr(\mathbf{H} \mid \mathbf{e})$ for some other set of network variables $\mathbf{H}$. Our algorithm can therefore be used to compute the variance of this expectation under the same complexity.

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 $\psi_a(\mathbf{x}_a) = 0$. This is typically the convention used in the implementation and analysis of jointree algorithms (Lauritzen and Spiegelhalter, 1988; Jensen et al., 1990; Huang and Darwiche, 1996).

[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.

6. On the Semantics of Noisy Sensors

In the remainder of this paper, we consider threshold-based decisions where our observations $\mathbf{e}$ correspond to readings from noisy sensors. We considered such a scenario in our example from Section 2. We propose, in particular, to explicate the causal mechanisms that govern the behavior of sensors, and then consider the same-decision probability with respect to these causal mechanisms. In Section 7, we illustrate through examples how the same-decision probability can be used to distinguish scenarios involving noisy sensors, that we could otherwise not distinguish using the probability $Pr(d \mid \mathbf{e})$ alone. Our goal, in this section, is to show how we can augment a sensor so that its causal mechanisms are modeled explicitly.

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 $\{+, -\}$.[5] Suppose further that we are given the false positive f_p and false negative f_n rates of the sensor:

$$Pr(S{=}+ \mid X{=}-) = f_p \qquad\qquad Pr(S{=}- \mid X{=}+) = f_n.$$

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	$Pr(S \mid H, X)$
t	$+$	$+$	1
t	$-$	$+$	0
l	$+$	$+$	0
l	$-$	$+$	1

H	X	S	$Pr(S \mid H, X)$
p	$+$	$+$	1
p	$-$	$+$	1
n	$+$	$+$	0
n	$-$	$+$	0

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

We observe that these Boolean function have commonly used diagnostic interpretations, describing the behavior of a sensor. We will indeed assume these interpretations in the rest of this paper, for convenience:

- the state $H = \mathsf{t}$ indicates the sensor is "truthful,"
- the state $H = \mathsf{l}$ indicates the sensor is "lying,"
- the state $H = \mathsf{p}$ indicates the sensor is "stuck positive," and
- the state $H = \mathsf{n}$ indicates the sensor is "stuck negative."

Note that any stochastic model can be emulated by a functional model with stochastic inputs (Pearl, 2009; Druzdzel and Simon, 1993).

6.1. Assumptions about Causal Mechanisms

To reason about our augmented sensor model $X \to S \leftarrow H$, 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 $X \to S$, when variable H has been marginalized out:

$$Pr(S=+ \mid X=-) = \sum_H Pr(S=+ \mid H, X=-)Pr(H) = f_p \tag{3}$$

$$Pr(S=- \mid X=+) = \sum_H Pr(S=- \mid H, X=+)Pr(H) = f_n. \tag{4}$$

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 3 and 4, immediately commits us to the following distribution over causal mechanisms:

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

$$Pr(H = \mathsf{l}) = f_p f_n. \tag{5}$$

This assumption, with Equations 3 and 4, 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 (Heckerman et al., 1995b).[6] 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 false 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.

6.2. 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

[6] 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.

each refers to a hypothesis about which sensors are truthful, which are lying and which are irrelevant.

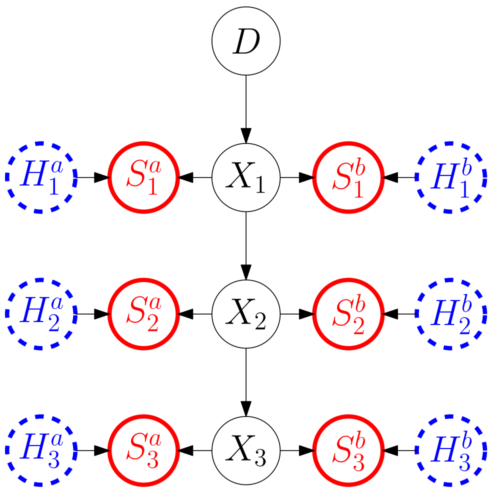
Note that our example network of Figure 1, from Section 2, corresponds to sensor models $X \to S \leftarrow H$ expanded from sensor models $X \to S$ with parameters $f_p = f_n = 0.1$. Table 1 depicts the nine causal mechanisms corresponding to two positive sensor readings in the network of Figure 1. The table also depicts the posterior distribution over these mechanisms, suggesting that the leading scenario, by a large margin, 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.

Before we close this section, we point out that the probability $Pr(d \mid \mathbf{s})$ is actually invariant to any assumption we made about the causal mechanisms governing sensor readings, i.e., about the prior distribution $Pr(\mathbf{H})$. In other words, as long as the distribution on variable H satisfies Equations 3 and 4, the probability $Pr(d \mid \mathbf{s})$ will have the same value, regardless of which particular distribution we choose for variable H. This is indeed true for the probability of any event that does not mention the auxiliary variables H. It is not true, however, for the same-decision probability, which we shall see in the following section.

7. Examples

Consider the Bayesian network in Figure 3, which depicts a chain $D \to X_1 \to X_2 \to X_3$ 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 $Pr(D{=}+ \mid \mathbf{e}) \geq T$ for some sensor readings $\mathbf{e}$ 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 false negatives, which is needed to induce a distribution over causal mechanisms. We also provide the results of these computations under the second assumption where the "lying" causal mechanism has zero probability (in brackets). As we discussed earlier, the



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

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

Figure 3: A Bayesian network with six sensors. Variables S_i^a and S_i^b represent redundant sensors for variable X_i. All sensors have the same false positive and negative rates of $f_p = f_n = .05$. Variables X_i all have the same CPTs. (only the one for variable X_1 is shown).

two results are expected to be very close since the false positive and negative rates are small. This is also 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	$S_2^a{=}+$	$S_2^a{=}+$ $S_2^b{=}+$
hypothesis probability	55.34%	60.01%
decision confidence	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	$S_2^a\!=\!+,\ S_2^b\!=\!+$	$S_1^a\!=\!+,\ S_1^b\!=\!-$ $S_2^a\!=\!+,\ S_2^b\!=\!+$
hypothesis probability	60.01%	60.01%
decision confidence	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	$S_1^a\!=\!+,\ S_1^b\!=\!-$ $S_2^a\!=\!+,\ S_2^b\!=\!+$	$S_2^a\!=\!+,\ S_2^b\!=\!+$ $S_3^a\!=\!+,\ S_3^b\!=\!-$
hypothesis probability	60.01%	60.01%
decision confidence	79.97%[80.07%]	99.48%[99.48%]

In this case, the sensor conflict increases the same-decision probability just slightly (from 99.22% to 99.48%).[7] The next example shows what happens when we get two negative readings but at different sensor locations.

	Scenario 5	Scenario 6
readings	$S_1^a\!=\!-,\ S_1^b\!=\!-$ $S_2^a\!=\!+,\ S_2^b\!=\!+$	$S_2^a\!=\!+,\ S_2^b\!=\!+$ $S_3^a\!=\!-,\ S_3^b\!=\!-$
hypothesis probability	4.31%	57.88%
decision confidence	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

[7]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.

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 consider the exact computation of the same-decision probability a direction for further research. 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 set of unobserved variables. We analyzed the complexity of computing the same-decision probability, showing that the corresponding decision problem is complete for the complexity class $\mathrm{PP}^{\mathrm{PP}}$. In response, 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. Finally, we proposed to explicate the causal mechanisms that govern the readings of sensors, which allows us to use the same-decision probability to reason about decisions under noisy sensors in a more refined way.

Acknowledgments

This work has been partially supported by ONR grant #N000141210423.

Appendix A. On the Complexity of Same-Decision Probability

In this section, let $\mathcal{N}$ denote a Bayesian network that induces a distribution Pr over a set of variables $\mathbf{X}$. Let $\mathbf{E} \subseteq \mathbf{X}$ denote a set of observed variables, and let $\mathbf{e}$ (the evidence) denote an instantiation of the variables $\mathbf{E}$. Similarly, let $\mathbf{H} \subseteq \mathbf{X} - \mathbf{E}$ denote a set of unobserved (hidden) variables, and let $\mathbf{h}$ denote an instantiation of $\mathbf{H}$. Let $D \in \mathbf{X}$ be a variable of interest, where $D \notin \mathbf{H}$ and $D \notin \mathbf{E}$. Note that the variable D, together with the sets of variables $\mathbf{H}$ and $\mathbf{E}$ may only mention a subset of the variables $\mathbf{X}$. That is, $\{D\} \cup \mathbf{H} \cup \mathbf{E} \subseteq \mathbf{X}$, where the containment may be strict.

Consider now the following decision problem for Bayesian networks over variables $\mathbf{X}$.

> **D-SDP**: Given a decision based on probability $Pr(d \mid \mathbf{e})$ surpassing a threshold T, a set of unobserved variables $\mathbf{H}$, and a probability p, is the same-decision probability:
>
> $$\mathcal{P}(Pr(d \mid \mathbf{e}, \mathbf{H}) \geq T) = \sum_{\mathbf{h}} [Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T] Pr(\mathbf{h} \mid \mathbf{e})$$
>
> greater than p?

We show here that decision problem **D-SDP** is $\mathrm{PP}^{\mathrm{PP}}$–complete. Intuitively, typical problems in PP are counting (or enumeration) problems (e.g., counting the number of satisfying assignments in a given CNF formula). Intuitively, problems in $\mathrm{PP}^{\mathrm{PP}}$ are counting problems that have counting subproblems (the PP oracle). Note that $\mathrm{PP}^{\mathrm{PP}}$ is the second level of the counting hierarchy (Allender and Wagner, 1990). Moreover, $\mathrm{PP}^{\mathrm{PP}}$ is the counting analogue of the class $\mathrm{NP}^{\mathrm{PP}}$, the latter of which includes a number of Bayesian network queries as complete problems, including MAP (Park and Darwiche, 2004), multi-parameter sensitivity analysis (Kwisthout and van der Gaag, 2008), and optimization of decision theoretic value of information (Krause and Guestrin, 2009). For a review on the complexity of reasoning in Bayesian networks, see, for example, Park and Darwiche (2004), Kwisthout (2009), and Darwiche (2009).

First, we show that **D-SDP** is in $\mathrm{PP}^{\mathrm{PP}}$, by providing a probabilistic polynomial-time algorithm, with access to a PP oracle, that answers the decision problem **D-SDP** correctly with probability greater than $\frac{1}{2}$. Our algorithm and its proof of correctness is based on those from (Darwiche, 2009), showing that the decision problem **D-MAR** is contained in PP:

> **D-MAR**: Given query variables $\mathbf{Q} \subseteq \mathbf{X}$, an instantiation $\mathbf{q}$, and a probability p, is $Pr(\mathbf{q} \mid \mathbf{e}) > p$?

We first observe that the same-decision probability can be viewed more simply as the probability $Pr(\beta \mid \mathbf{e})$ of an event β, where $\beta = \bigvee_{\mathbf{h}: Pr(d|\mathbf{e},\mathbf{h}) \geq T} \mathbf{h}$. We now specify a probabilistic polytime algorithm for deciding if $Pr(\beta \mid \mathbf{e}) > p$.

1. Define the following probabilities as a function of p:

$$a(p) = \begin{cases} 1 & \text{if } p < \frac{1}{2} \\ 1/(2p) & \text{otherwise} \end{cases}$$

$$b(p) = \begin{cases} (1-2p)/(2-2p) & \text{if } p < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

2. Sample a complete instantiation $\mathbf{x}$ from the Bayesian network, with probability $Pr(\mathbf{x})$. We can do this in linear time, using forward sampling.
3. If $\mathbf{x}$ is compatible with $\mathbf{e}$, we test whether $Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T$ using our PP–oracle,[8] where $\mathbf{h}$ is the projection of instantiation $\mathbf{x}$ onto the variables $\mathbf{H}$. We can do this since our test is an instance of **D-MAR**, which is PP–complete.
4. Declare $Pr(\beta \mid \mathbf{e}) > p$ according to the following probabilities
   - $a(p)$ if instantiation $\mathbf{x}$ is compatible with $\mathbf{e}$, and $Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T$;
   - $b(p)$ if instantiation $\mathbf{x}$ is compatible with $\mathbf{e}$, and $Pr(d \mid \mathbf{e}, \mathbf{h}) < T$;
   - $\frac{1}{2}$ if instantiation $\mathbf{x}$ is not compatible with $\mathbf{e}$.

Theorem 2. *This procedure will declare $Pr(\beta \mid \mathbf{e}) > p$ correctly with probability greater than $\frac{1}{2}$.*

[8] Equivalently, we can test whether $Pr(\neg d \mid \mathbf{e}, \mathbf{h}) > 1 - T$, using an oracle for **D-MAR**.

Proof The probability of declaring $Pr(\beta \mid \mathbf{e}) > p$ is

$$r = a(p)Pr(\beta, \mathbf{e}) + b(p)Pr(\neg\beta, \mathbf{e}) + \frac{1}{2}[1 - Pr(\mathbf{e})]$$

noting that the probability that sample $\mathbf{x}$ is compatible with $\mathbf{e}$ is $Pr(\mathbf{e})$, and then given this, the probability that $Pr(d \mid \mathbf{e}, \mathbf{h}) \geq T$ is $Pr(\beta \mid \mathbf{e})$ (by the definition of β). It remains to show that $r > \frac{1}{2}$ iff $Pr(\beta \mid \mathbf{e}) > p$.

The remainder of the proof mirrors the proof of Theorem 11.5 in (Darwiche, 2009), which we reproduce here for completeness. First, $r > \frac{1}{2}$ iff

$$a(p)Pr(\beta \mid \mathbf{e}) + b(p)Pr(\neg\beta \mid \mathbf{e}) > \frac{1}{2}.$$

We consider two cases, $p < \frac{1}{2}$ and $p \geq \frac{1}{2}$, which are the two cases in the definitions of $a(p)$ and $b(p)$.

If $p < \frac{1}{2}$, then the following inequalities are equivalent:

$$\begin{aligned}
a(p)Pr(\beta \mid \mathbf{e}) + b(p)Pr(\neg\beta \mid \mathbf{e}) &> \frac{1}{2} \\
Pr(\beta \mid \mathbf{e}) + \frac{1-2p}{2-2p}\left[1 - Pr(\beta \mid \mathbf{e})\right] &> \frac{1}{2} \\
Pr(\beta \mid \mathbf{e})\left[1 - \frac{1-2p}{2-2p}\right] &> \frac{1}{2} - \frac{1-2p}{2-2p} \\
Pr(\beta \mid \mathbf{e})\frac{1}{2-2p} &> \frac{p}{2-2p} \\
Pr(\beta \mid \mathbf{e}) &> p.
\end{aligned}$$

Otherwise, if $p \geq \frac{1}{2}$, then the following inequalities are equivalent:

$$\begin{aligned}
a(p)Pr(\beta \mid \mathbf{e}) + b(p)Pr(\neg\beta \mid \mathbf{e}) &> \frac{1}{2} \\
\frac{1}{2p}Pr(\beta \mid \mathbf{e}) &> \frac{1}{2} \\
Pr(\beta \mid \mathbf{e}) &> p.
\end{aligned}$$

Thus, $r > \frac{1}{2}$ iff $Pr(\beta \mid \mathbf{e}) > p$. □

Having just shown that **D-SDP** is in $\mathrm{PP}^{\mathrm{PP}}$, it remains to show that **D-SDP** is $\mathrm{PP}^{\mathrm{PP}}$–hard. Given a propositional sentence α over Boolean variables $X_1, \ldots, X_n$, consider the following decision problem.

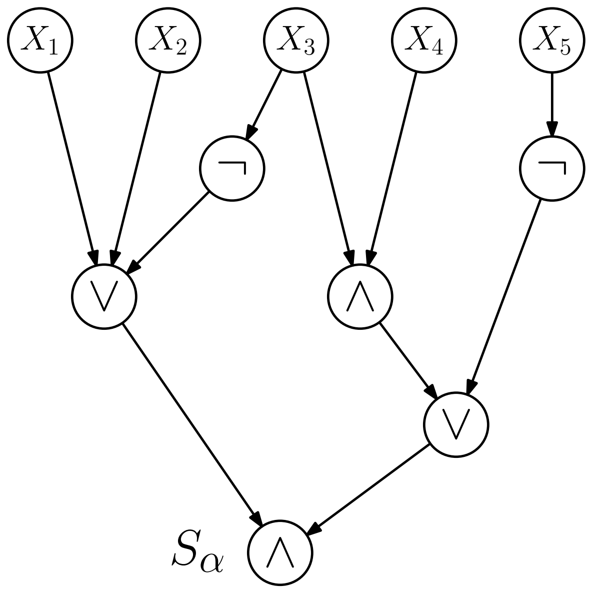


Figure A.4: A Bayesian network representing the following sentence in propositional logic: $\alpha = (X_1 \vee X_2 \vee \neg X_3) \wedge ((X_3 \wedge X_4) \vee \neg X_5)$

MAJ-MAJ-SAT: Given some number k where $1 \leq k \leq n$, are there a majority of instantiations $x_1, \ldots, x_k$, where a majority of instantiations $x_{k+1}, \ldots, x_n$ have instantiations $x_1, \ldots, x_n$ that satisfy α?

For a given instantiation $x_1, \ldots, x_k$, we can ask if a majority of instantiations $x_{k+1}, \ldots, x_n$ lead to satisfying assignments $x_1, \ldots, x_n$ (which is a **MAJ-SAT** subproblem). For the problem **MAJ-MAJ-SAT**, we ask if there are a majority of such instantiations $x_1, \ldots, x_k$. Given that **MAJ-MAJ-SAT** is complete for $\mathrm{PP}^{\mathrm{PP}}$ (Wagner, 1986), we want to reduce instances of **MAJ-MAJ-SAT** to instances of **D-SDP**.

Given propositional sentence α, we assume the typical Bayesian network N_α representing it; see, e.g., Section 11.3 of Darwiche (2009). This network has root nodes $X_1, \ldots, X_n$ and a leaf node S_α representing the value of the sentence α. Nodes X_i have uniform priors, and each logical operator appearing in sentence α is represented using the appropriate deterministic CPT. Figure A.4 illustrates an example.

Theorem 3. *There are a majority of instantiations $x_1, \ldots, x_k$, where a majority of instantiations $x_{k+1}, \ldots, x_n$ have instantiations $x_1, \ldots, x_n$ that satisfy*

α iff the same-decision probability $\mathcal{P}(Pr(S_\alpha \mid X_1, \ldots, X_k) > \frac{1}{2})$ is greater than $\frac{1}{2}$.

Proof Consider the same-decision probability for a decision based on the threshold $Pr(S_\alpha{=}\mathsf{true}) > \frac{1}{2}$, with respect to variables $X_1, \ldots, X_k$ (or equivalently, based on the threshold $Pr(S_\alpha{=}\mathsf{false}) \leq \frac{1}{2}$):

$$\begin{aligned}
&\mathcal{P}(Pr(S_\alpha{=}\mathsf{true} \mid X_1, \ldots, X_k) > \frac{1}{2}) \\
&= \sum_{x_1,\ldots,x_k} \Big[Pr(S_\alpha{=}\mathsf{true} \mid x_1, \ldots, x_k) > \frac{1}{2}\Big] \cdot Pr(x_1, \ldots, x_k) \\
&= \frac{1}{2^k} \sum_{x_1,\ldots,x_k} \Big[Pr(S_\alpha{=}\mathsf{true} \mid x_1, \ldots, x_k) > \frac{1}{2}\Big].
\end{aligned}$$

Note that $Pr(S_\alpha \mid x_1, \ldots, x_n) = 1$ if $x_1, \ldots, x_n$ satisfies α, and zero otherwise. Moreover, $Pr(x_1, \ldots, x_n) = \prod_{i=1}^n Pr(x_i) = \frac{1}{2^n}$. Thus,

$$\begin{aligned}
Pr(x_1, \ldots, x_k, S_\alpha{=}\mathsf{true}) &= \sum_{x_{k+1},\ldots,x_n} Pr(x_1, \ldots, x_n, S_\alpha{=}\mathsf{true}) \\
&= \sum_{x_{k+1},\ldots,x_n} Pr(S_\alpha{=}\mathsf{true} \mid x_1, \ldots, x_n) Pr(x_1, \ldots, x_n) \\
&= \sum_{\substack{x_{k+1},\ldots,x_n \\ x_1,\ldots,x_n \models \alpha}} Pr(x_1, \ldots, x_n) = \frac{c}{2^n}
\end{aligned}$$

where c is the number of instantiations $x_{k+1}, \ldots, x_n$ for which the instantiation $x_1, \ldots, x_n$ satisfies α. Since $Pr(x_1, \ldots, x_k) = \frac{1}{2^k}$, we have that

$$Pr(S_\alpha{=}\mathsf{true} \mid x_1, \ldots, x_k) = \frac{c}{2^{n-k}}$$

which is the fraction of such instantiations $x_{k+1}, \ldots, x_n$. Thus, there are a majority of such instantiations iff $\frac{c}{2^{n-k}} > \frac{1}{2}$.

Finally, the same-decision probability is:

$$\mathcal{P}(Pr(S_\alpha{=}\mathsf{true} \mid X_1, \ldots, X_k) > \frac{1}{2}) = \frac{b}{2^k}$$

where b is the number of instantiations $x_1, \ldots, x_k$ for which the majority of instantiations $x_{k+1}, \ldots, x_n$ have instantiations $x_1, \ldots, x_n$ that satisfy α. Thus, there are a majority of such instantiations $x_1, \ldots, x_k$ iff the same-decision probability is greater than $\frac{1}{2}$, i.e., iff $\frac{b}{2^k} > \frac{1}{2}$. $\square$

Theorem 3 establishes that **D-SDP** is $\mathrm{PP}^{\mathrm{PP}}$–hard. Theorem 2 establishes **D-SDP** is in $\mathrm{PP}^{\mathrm{PP}}$. Hence, we have Theorem 1, and **D-SDP** is $\mathrm{PP}^{\mathrm{PP}}$–complete.

References

Agosta, J. M., Gardos, T., Druzdel, M. J., 2008. Query-based diagnostics. In: The Fourth European Workshop on Probabilistic Graphical Models (PGM).

Agosta, J. M., Khan, O. Z., Poupart, P., 2010. Evaluation results for a query-based diagnostics application. In: The Fifth European Workshop on Probabilistic Graphical Models (PGM).

Allender, E., Wagner, K. W., 1990. Counting hierarchies: Polynomial time and constant depth circuits. Bulletin of the EATCS 40, 182–194.

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

Charitos, T., van der Gaag, L. C., 2006. Sensitivity analysis for threshold decision making with dynamic networks. In: Proceedings of the 22nd Conference in Uncertainty in Artificial Intelligence (UAI).

Darwiche, A., 2009. Modeling and Reasoning with Bayesian Networks. Cambridge University Press.

Dechter, R., 1996. Bucket elimination: A unifying framework for probabilistic inference. In: Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI). pp. 211–219.

Druzdzel, M. J., Simon, H. A., 1993. Causality in Bayesian belief networks. In: Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI). pp. 3–11.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Machine Learning 29 (2-3), 131–163.

Hamscher, W., Console, L., de Kleer, J., 1992. Readings in model-based diagnosis. Morgan Kaufmann.

Heckerman, D., Breese, J. S., Rommelse, K., 1995a. Decision-theoretic troubleshooting. Communications of the ACM 38 (3), 49–57.

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

Huang, C., Darwiche, A., 1996. Inference in belief networks: A procedural guide. International Journal of Approximate Reasoning 15 (3), 225–263.

Jensen, F. V., Lauritzen, S., Olesen, K., 1990. Bayesian updating in recursive graphical models by local computation. Computational Statistics Quarterly 4, 269–282.

Krause, A., Guestrin, C., 2009. Optimal value of information in graphical models. Journal of Artificial Intelligence Research 35, 557–591.

Kwisthout, J., 2009. The computational complexity of probabilistic networks. Ph.D. thesis, University of Utrecht.

Kwisthout, J., van der Gaag, L. C., 2008. The computational complexity of sensitivity analysis and parameter tuning. In: Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence (UAI). pp. 349–356.

Lauritzen, S. L., Spiegelhalter, D. J., 1988. Local computations with probabilities on graphical structures and their application to expert systems. Journal of Royal Statistics Society, Series B 50 (2), 157–224.

Littman, M. L., Goldsmith, J., Mundhenk, M., 1998. The computational complexity of probabilistic planning. Journal of Artificial Intelligence Research 9, 1–36.

Lu, T.-C., Przytula, K. W., 2006. Focusing strategies for multiple fault diagnosis. In: Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS). pp. 842–847.

Park, J., Darwiche, A., 2004. Complexity results and approximation strategies for MAP explanations. Journal of Artificial Intelligence Research 21, 101–133.

Pauker, S. G., Kassirer, J. P., 1980. The threshold approach to clinical decision making. New England Journal of Medicine 302 (20), 1109–1117.

Pearl, J., 2009. Causality: Models, Reasoning and Inference, 2nd Edition. Cambridge University Press.

Toda, S., 1991. PP is as hard as the polynomial-time hierarchy. SIAM Journal on Computing 20 (5), 865–877.

Umans, C., 2000. Approximability and completeness in the polynomial hierarchy. Ph.D. thesis, University of California, Berkeley.

van der Gaag, L., Renooij, S., Coupé, V., 2007. Sensitivity analysis of probabilistic networks. In: Lucas, P., Gámez, J., Salmerón, A. (Eds.), Advances in Probabilistic Graphical Models. Vol. 214 of Studies in Fuzziness and Soft Computing. Springer Berlin/Heidelberg, pp. 103–124.

van der Gaag, L. C., Coupé, V. M. H., 1999. Sensitive analysis for threshold decision making with Bayesian belief networks. In: 6th Congress of the Italian Association for Artificial Intelligence (AI*IA). pp. 37–48.

Wagner, K. W., 1986. The complexity of combinatorial problems with succinct input representation. Acta Informatica 23 (3), 325–356.

Zhang, N. L., Poole, D., 1996. Exploiting causal independence in Bayesian network inference. Journal of Artificial Intelligence Research 5, 301–328.