$$D(Pr, Pr') \stackrel{def}{=} \ln \max_w \frac{Pr'(w)}{Pr(w)} - \ln \min_w \frac{Pr'(w)}{Pr(w)},$$

where $0/0$ is defined as 1.

We will say that two probability distributions Pr and Pr' *have the same support*, if for every world w, $Pr(w) = 0$ iff $Pr'(w) = 0$. Note that if two distributions Pr and Pr' do not have the same support, $D(Pr, Pr') = \infty$.

Our first result on the defined measure is that it satisfies the three properties of distance, hence, it is a *distance measure.*

Theorem 1 *Let Pr, Pr' and Pr'' be three probability distributions over the same set of worlds. The distance measure given in Definition 1 satisfies these three properties:*

Positiveness: $D(Pr, Pr') \geq 0$, and $D(Pr, Pr') = 0$ iff $Pr = Pr'$;

Symmetry: $D(Pr, Pr') = D(Pr', Pr)$;

Triangle Inequality: $D(Pr, Pr') + D(Pr', Pr'') \geq D(Pr, Pr'')$.

The interest in the defined distance measure stems from two reasons. First, it can be easily computed in a number of practical situations which we discuss in later sections. Second, it allows us to bound the difference in beliefs captured by two probability distributions.

Theorem 2 *Let Pr and Pr' be two probability distributions over the same set of worlds. Let α and β be two events. We then have:*

$$e^{-D(Pr,Pr')} \leq \frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \leq e^{D(Pr,Pr')},$$

where $O(\alpha \mid \beta) = Pr(\alpha \mid \beta)/Pr(\overline{\alpha} \mid \beta)$ is the odds of event α given β with respect to Pr, and $O'(\alpha \mid \beta) = Pr'(\alpha \mid \beta)/Pr'(\overline{\alpha} \mid \beta)$ is the odds of event α given β with respect to Pr'.[1] *The bound is tight in the sense that for every pair of distributions Pr and Pr', there are events α and β such that:*

$$\frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} = e^{D(Pr,Pr')}, \quad \frac{O'(\overline{\alpha} \mid \beta)}{O(\overline{\alpha} \mid \beta)} = e^{-D(Pr,Pr')}.$$

We can express the bound of Theorem 2 in two other useful forms. First, we can use logarithms:

$$|\ln O'(\alpha \mid \beta) - \ln O(\alpha \mid \beta)| \leq D(Pr, Pr'). \quad (1)$$

Second, we can use probabilities instead of odds:

$$\frac{p\, e^{-d}}{p\,(e^{-d} - 1) + 1} \leq Pr'(\alpha \mid \beta) \leq \frac{p\, e^{d}}{p\,(e^{d} - 1) + 1}, \quad (2)$$

where $p = Pr(\alpha \mid \beta)$ is the initial belief in α given β, and $d = D(Pr, Pr')$ is the distance. The bounds of $Pr'(\alpha \mid \beta)$ are plotted against p for several different values of d in Figure 1.

[1] Of course, we must have $Pr(\beta) \neq 0$ and $Pr'(\beta) \neq 0$ for the odds to be defined.

In the applications we shall discuss next, Pr is a distribution which represents some initial state of belief, and Pr' is a distribution which represents a new state of belief. The new state of belief results form applying some kind of (usually local) change to the initial state. Examples include the change in some conditional belief or the incorporation of new evidence. Our goal is then to assess the global impact of such local belief changes. According to Theorem 2, if we are able to compute the distance measure $D(Pr, Pr')$, then we can bound global belief change in a very precise sense. For example, we can use Equation 2 to compute the bound on any query $Pr'(\alpha \mid \beta)$. We will later show two applications from sensitivity analysis and belief revision where the distance measure can be computed efficiently.

But first, we need to settle a major question: Can we bound belief change in the sense given above using one of the classical probabilistic measures? We show next that this is not possible using at least two of the most commonly used measures.

Kullback-Leibler (KL) divergence One of the most common measures for comparing probability distributions is the KL-divergence (Kullback & Leibler 1951).

Definition 2 *Let Pr and Pr' be two probability distributions over the same set of worlds w. The KL-divergence between Pr and Pr' is defined as:*

$$KL(Pr, Pr') \stackrel{def}{=} -\sum_w Pr(w) \ln \frac{Pr'(w)}{Pr(w)}.$$[2]

The first thing to note about KL-divergence is that it is incomparable with our distance measure.

Example 1 *Consider the following distributions, Pr, Pr' and Pr'', over worlds w_1, w_2 and w_3:*

$$\begin{array}{lll} Pr(w_1) = .50, & Pr(w_2) = .25, & Pr(w_3) = .25; \\ Pr'(w_1) = .50, & Pr'(w_2) = .30, & Pr'(w_3) = .20; \\ Pr''(w_1) = .43, & Pr''(w_2) = .32, & Pr''(w_3) = .25. \end{array}$$

Computing the KL-divergence gives us: $KL(Pr, Pr') = .0102$ and $KL(Pr, Pr'') = .0137$. Computing our distance measure gives us: $D(Pr, Pr') = .405$ and $D(Pr, Pr'') = .398$. Therefore, according to KL-divergence, Pr' is closer to Pr than Pr'', while according to our distance measure, Pr'' is closer to Pr than Pr'.

We are now faced with two questions:

1. Can we use KL-divergence to bound belief change as we did using our distance measure? The answer is no as we show next.
2. Given that our goal is to minimize belief change, should we try to minimize our distance measure or some other measure, such as KL-divergence? We answer this question only partially in the following sections, by showing that two proposals that come from the literatures on sensitivity analysis and belief revision do correspond to the minimization of our distance measure.

[2] Note that KL-divergence is asymmetric, and is thus technically not a distance measure.

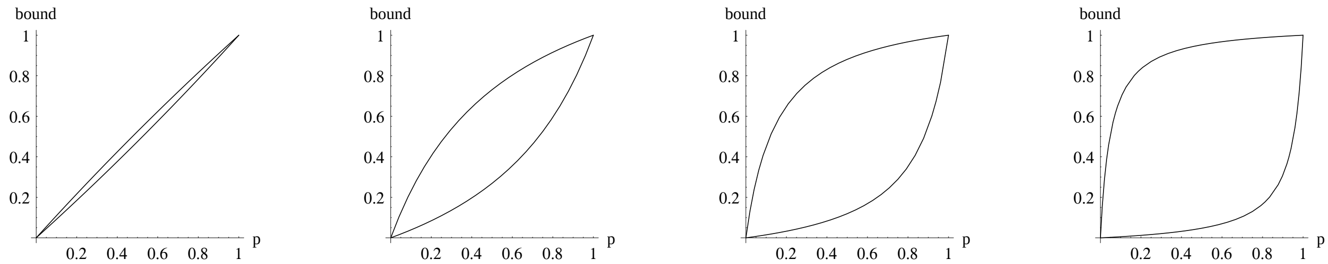

Figure 1: The bounds of $Pr'(\alpha \mid \beta)$, as given by Equation 2, plotted against the initial belief $p = Pr(\alpha \mid \beta)$ for several different values of distance $d = D(Pr, Pr')$: (from left to right) $d = .1$, $d = 1$, $d = 2$ and $d = 3$.

The following example addresses the first question.

Example 2 *Consider the following distributions, Pr and Pr', over worlds w_1, w_2 and w_3:*

$$Pr(w_1) = p, \quad Pr(w_2) = q - p, \quad Pr(w_3) = 1 - q;$$
$$Pr'(w_1) = kp, \quad Pr'(w_2) = q - kp, \quad Pr'(w_3) = 1 - q;$$

where p, q and k are constants, with $0 \le p \le q \le 1$ and $0 \le k \le q/p$. The KL-divergence between Pr and Pr' is:

$$KL(Pr, Pr') = -p \ln k - (q - p) \ln \frac{q - kp}{q - p}.$$

Assume we have events $\alpha = w_1$ and $\beta = w_1, w_2$. The odds ratio of α given β between Pr and Pr' is:

$$\frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} = \frac{k(q - p)}{q - kp}.$$

We can see that as p approaches 0, the KL-divergence also approaches 0, while the odds ratio $O'(\alpha \mid \beta)/O(\alpha \mid \beta)$ approaches k.

This example shows that we can make the KL-divergence arbitrarily close to 0, while keeping some odds ratio arbitrarily close to some constant k. In this example, we condition on event β, which has a probability of q that can be arbitrarily large. However, the probability of α, which is p according to Pr and kp according to Pr', is very small. Hence, although we have $Pr'(\alpha)/Pr(\alpha) = k$, this ratio is ignored by KL-divergence because the term $-p \ln k$ is very small as p approaches 0. More generally, the "contribution" of a world w to KL-divergence is equal to $-Pr(w) \ln(Pr'(w)/Pr(w))$. Therefore for a fixed ratio $Pr'(w)/Pr(w)$, this "contribution" becomes closer to 0 as $Pr(w)$ decreases, and becomes infinitesimal when $Pr(w)$ approaches 0.

Euclidean distance Another popular measure to compare two probability distributions is the Euclidean distance.

Definition 3 *Let Pr and Pr' be two probability distributions over the same set of worlds w. The Euclidean distance between Pr and Pr' is defined as:*

$$ED(Pr, Pr') \stackrel{def}{=} \sum_w \sqrt{(Pr'(w) - Pr(w))^2}.$$

That is, when computing the Euclidean distance, we add up the squared differences between pairs of probability values. Therefore, this measure has the same problem as KL-divergence: even if there is a large relative difference for the probability of a world with respect to Pr and Pr', it will be ignored if this probability is very small. Consequently, we cannot provide any guarantee on the ratio $O'(\alpha \mid \beta)/O(\alpha \mid \beta)$, no matter how small the Euclidean distance is (unless it is zero).

To summarize, neither KL-divergence nor Euclidean distance can be used to provide guarantees on the ratio $O'(\alpha \mid \beta)/O(\alpha \mid \beta)$, as we did in Theorem 2 using our distance measure.

Applications to sensitivity analysis

We now consider a major application of our distance measure to sensitivity analysis in belief networks (Laskey 1995; Castillo, Gutiérrez, & Hadi 1997; Kjaerulff & van der Gaag 2000; Darwiche 2000; Chan & Darwiche 2001). A belief network is a graphical probabilistic model, composed of two parts: a directed acyclic graph where nodes represent variables, and a set of conditional probability tables (CPTs), one for each variable (Pearl 1988; Jensen 2001). The CPT for variable X with parents $\mathbf{U}$ defines a set of conditional beliefs of the form $\theta_{x|\mathbf{u}} = Pr(x \mid \mathbf{u})$, where x is a value of variable X, $\mathbf{u}$ is an instantiation of parents $\mathbf{U}$, and $\theta_{x|\mathbf{u}}$ is the probability value of x given $\mathbf{u}$, and is called a network parameter.

One of the key questions with respect to belief networks is this: what can we say about the global effect of changing some parameter $\theta_{x|\mathbf{u}}$ to a new value $\theta'_{x|\mathbf{u}}$? That is, what is the effect of such a local parameter change on the value of some arbitrary query $Pr(\alpha \mid \beta)$?

Chan and Darwiche (2001) have provided a partial answer to this question, for the case where: variable X is binary (it has only two values, x and $\overline{x}$); α is the value y of some variable Y; β is the instantiation $\mathbf{e}$ of some variables $\mathbf{E}$, and neither $\theta_{x|\mathbf{u}}$ nor $\theta'_{x|\mathbf{u}}$ is extreme (equal to 0 or 1). Specifically under these conditions, they have shown that:

$$|\ln O'(y \mid \mathbf{e}) - \ln O(y \mid \mathbf{e})| \le \left| \ln \frac{\theta'_{x|\mathbf{u}}}{\theta'_{\overline{x}|\mathbf{u}}} - \ln \frac{\theta_{x|\mathbf{u}}}{\theta_{\overline{x}|\mathbf{u}}} \right|.$$

Using the above bound, Chan and Darwiche have provided a formalization of a number of intuitions relating to the sensitivity of probabilistic queries to changes in network parameters. We will now show how our distance measure can be used to derive a generalization of the above bound, which applies without any of the previously mentioned restrictions.

Suppose that our initial belief network is N and it induces a probability distribution Pr. By changing the CPT for variable X, we produce a new belief network N' that induces a probability distribution Pr'. If we are able to compute the distance between Pr and Pr', $D(Pr, Pr')$, we can then use Theorem 2 to provide a guarantee on the global effect of the local CPT change. As it turns out, the distance can be computed locally as given by the following theorem.

Theorem 3 *Let N and N' be belief networks which induce distributions Pr and Pr' respectively, and let X be a variable with parents $\mathbf{U}$ in network N. Suppose that N' is obtained from N by changing the conditional probability distribution of variable X given parent instantiation $\mathbf{u}$ from $\Theta_{X|\mathbf{u}}$ to $\Theta'_{X|\mathbf{u}}$, i.e. we change parameter $\theta_{x|\mathbf{u}}$ to $\theta'_{x|\mathbf{u}}$ for every value x. If $Pr(\mathbf{u}) > 0$, then:*

$$D(Pr, Pr') = D(\Theta_{X|\mathbf{u}}, \Theta'_{X|\mathbf{u}}).$$

The above theorem shows that the distance between the global probability distributions induced by networks N and N' is exactly the distance between the local distributions of X given $\mathbf{u}$, assuming that all other local distributions in N and N' are the same.

Theorem 3 is of great practical importance as it allows us to invoke Theorem 2 to provide a generalized sensitivity analysis formula for belief networks.

Corollary 1 *Let N and N' be belief networks which induce distributions Pr and Pr' respectively, and let X be a variable with parents $\mathbf{U}$ in network N. Suppose that N' is obtained from N by changing the conditional probability distribution of variable X given parent instantiation $\mathbf{u}$ from $\Theta_{X|\mathbf{u}}$ to $\Theta'_{X|\mathbf{u}}$, i.e. we change parameter $\theta_{x|\mathbf{u}}$ to $\theta'_{x|\mathbf{u}}$ for every value x. If $Pr(\mathbf{u}) > 0$, then:*

$$e^{-D(\Theta_{X|\mathbf{u}}, \Theta'_{X|\mathbf{u}})} \leq \frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \leq e^{D(\Theta_{X|\mathbf{u}}, \Theta'_{X|\mathbf{u}})}.$$

The bound of Chan and Darwiche is a special case of Corollary 1, when X has only two values x and $\overline{x}$. In this case, the distance $D(\Theta_{X|\mathbf{u}}, \Theta'_{X|\mathbf{u}})$ is equal to:

$$\begin{aligned} D(\Theta_{X|\mathbf{u}}, \Theta'_{X|\mathbf{u}}) &= \left| \ln \frac{\theta'_{x|\mathbf{u}}}{\theta_{x|\mathbf{u}}} - \ln \frac{\theta'_{\overline{x}|\mathbf{u}}}{\theta_{\overline{x}|\mathbf{u}}} \right| \\ &= \left| \ln \frac{\theta'_{x|\mathbf{u}}}{\theta'_{\overline{x}|\mathbf{u}}} - \ln \frac{\theta_{x|\mathbf{u}}}{\theta_{\overline{x}|\mathbf{u}}} \right|. \end{aligned}$$

We have therefore generalized their results on sensitivity analysis to arbitrary events and belief networks. We have also relaxed the condition that neither $\theta_{x|\mathbf{u}}$ nor $\theta'_{x|\mathbf{u}}$ can be extreme.

We now close this section with a final application of our distance measure. Suppose X is a variable with parents $\mathbf{U}$, values x_1, x_2 and x_3, and parameters $\theta_{x_1|\mathbf{u}} = .6$, $\theta_{x_2|\mathbf{u}} = .3$ and $\theta_{x_3|\mathbf{u}} = .1$. Suppose further that we want to change the parameter $\theta_{x_1|\mathbf{u}} = .6$ to $\theta'_{x_1|\mathbf{u}} = .8$. As a result, we will need to change the other parameters $\theta_{x_2|\mathbf{u}}$ and $\theta_{x_3|\mathbf{u}}$ so that the sum of all three parameters remains to be 1. Because X is multivalued, there are infinitely many ways to change the other two parameters and the question is: which one of them should we choose? One popular scheme, which we will call the *proportional scheme*, distributes the mass $1 - \theta'_{x_1|\mathbf{u}} = 1 - .8 = .2$ among the other two parameters proportionally to their initial values. That is, the new parameters will be $\theta'_{x_2|\mathbf{u}} = .2(.3/.4) = .15$ and $\theta'_{x_3|\mathbf{u}} = .2(.1/.4) = .05$. This scheme has been used in all approaches to sensitivity analysis we are familiar with (Laskey 1995; Kjaerulff & van der Gaag 2000; Darwiche 2000), yet without justification. As it turns out, we can use our distance measure to prove the optimality of this scheme in a very precise sense.

Theorem 4 *When changing a parameter $\theta_{x|\mathbf{u}}$ to $\theta'_{x|\mathbf{u}}$ for a multivalued variable X, the proportional scheme, i.e. the one that sets $\theta'_{x_i|\mathbf{u}} = (1 - \theta'_{x|\mathbf{u}})(\theta_{x_i|\mathbf{u}}/(1 - \theta_{x|\mathbf{u}}))$ for all $x_i \neq x$, leads to the smallest distance between the original and new distributions of X, which is given by:*

$$\begin{aligned} D(\Theta_{X|\mathbf{u}}, \Theta'_{X|\mathbf{u}}) &= \left| \ln \frac{\theta'_{x|\mathbf{u}}}{\theta_{x|\mathbf{u}}} - \ln \frac{\theta'_{\overline{x}|\mathbf{u}}}{\theta_{\overline{x}|\mathbf{u}}} \right| \\ &= \left| \ln \frac{\theta'_{x|\mathbf{u}}}{\theta'_{\overline{x}|\mathbf{u}}} - \ln \frac{\theta_{x|\mathbf{u}}}{\theta_{\overline{x}|\mathbf{u}}} \right|, \end{aligned}$$

where we define $\theta'_{\overline{x}|\mathbf{u}} = 1 - \theta'_{x|\mathbf{u}}$ and $\theta_{\overline{x}|\mathbf{u}} = 1 - \theta_{x|\mathbf{u}}$.

Theorem 4 thus justifies the use of the proportional scheme on the grounds that it leads to the tightest bound on the amount of associated belief change.

Applications to belief revision

The problem of probabilistic belief revision can be defined as follows. We are given a probability distribution Pr, which captures a state of belief and assigns a probability p to some event γ. We then obtain evidence suggesting a probability of $q \neq p$ for γ. Our goal is to change the distribution Pr to a new distribution Pr' such that $Pr'(\gamma) = q$. The are two problems here. First, usually there are many choices for Pr'. Which one should we adopt? Second, if we decide to choose the new state of belief Pr' according to some specific method, can we provide any guarantee on the amount of belief change that will be undergone as a result of moving from Pr to Pr'?

As for the first question, we will consider two methods for updating a probability distribution in the face of new evidence: Jeffrey's rule (Jeffrey 1965) and Pearl's method of virtual evidence (Pearl 1988). As for the second question, we will show next that we can indeed provide interesting guarantees on the amount of belief change induced by both methods. We present the guarantees first and then some of their applications.

Jeffrey's rule We start with Jeffrey's rule for accommodating uncertain evidence.

Definition 4 *Let Pr be a probability distribution over worlds w, and let $\gamma_1, \ldots, \gamma_n$ be a set of mutually exclusive and exhaustive events that are assigned probabilities $p_1, \ldots, p_n$, respectively, by Pr. Suppose we want to change Pr to a new distribution Pr' such that the probabilities of $\gamma_1, \ldots, \gamma_n$ become $q_1, \ldots, q_n$, respectively. Jeffrey's rule defines the new distribution Pr' as follows:*

$$Pr'(w) \stackrel{def}{=} Pr(w)\frac{q_i}{p_i}, \quad \text{if } w \models \gamma_i.$$

The main result we have about Jeffrey's rule is that the distance between probability distributions Pr and Pr' can be computed directly from the old and new probabilities of $\gamma_1, \ldots, \gamma_n$. This immediately allows us to invoke Theorem 2 as we show next.

Theorem 5 *Let Pr and Pr' be two distributions, where Pr' is obtained by applying Jeffrey's rule to Pr as given in Definition 4. We then have:*

$$D(Pr, Pr') = \ln \max_i \frac{q_i}{p_i} - \ln \min_i \frac{q_i}{p_i}.$$

We immediately get the following bound.

Corollary 2 *If O and O' are the odds functions before and after applying Jeffrey's rule as given in Definition 4, then:*

$$e^{-d} \leq \frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \leq e^{d},$$

where $d = \ln \max_i (q_i/p_i) - \ln \min_i (q_i/p_i)$.

To consider an example application of Corollary 2, we use a simple example from Jeffrey (1965).

Example 3 *Assume that we are given a piece of cloth, where its color can be one of: green (c_g), blue (c_b), or violet (c_v). We also want to know whether in the next day, the cloth will be sold (s), or remain unsold ($\overline{s}$). Our original state of belief is given by the probability distribution of the worlds Pr:*

$$\begin{array}{lll} Pr(s, c_g) = .12, & Pr(s, c_b) = .12, & Pr(s, c_v) = .32, \\ Pr(\overline{s}, c_g) = .18, & Pr(\overline{s}, c_b) = .18, & Pr(\overline{s}, c_v) = .08. \end{array}$$

Therefore, our original state of belief on the color of the cloth (c_g, c_b, c_v) is given by the distribution $(.3, .3, .4)$. Assume that we now inspect the cloth by candlelight, and we want to revise our state of belief on the color of the cloth to the new distribution $(.7, .25, .05)$ using Jeffrey's rule. The distance between the original and new distributions of the worlds can be computed by simply examining the original and new distributions on the color variable as given by Theorem 5. Specifically, the distance between the two distributions is $\ln(.7/.3) - \ln(.05/.4) = 2.93$. We can now use this distance to provide a bound on the change in any of our beliefs. Consider for example our belief that the cloth is green given that it is sold tomorrow, $Pr(c_g|s)$, which is initially $.214$. Suppose we want to find the bound on the change in this belief induced by the new evidence. Given Corollary 2 and Equation 2, we have:

$$.0144 \leq Pr'(c_g|s) \leq .836,$$

which suggests that a dramatic change in belief is possible in this case. If we actually apply Jeffrey's rule, we get the new distribution Pr':

$$\begin{array}{lll} Pr'(s, c_g) = .28, & Pr'(s, c_b) = .10, & Pr'(s, c_v) = .04, \\ Pr'(\overline{s}, c_g) = .42, & Pr'(\overline{s}, c_b) = .15, & Pr'(\overline{s}, c_v) = .01, \end{array}$$

according to which $Pr'(c_g|s) = .667$, which does suggest a dramatic change. On the other hand, if the new evidence on the color of the cloth is given by the distribution $(.25, .25, .50)$ instead, the distance between the old and new distributions will be $.406$, and our bound will be: $.153 \leq Pr'(c_g|s) \leq .290$, which is obviously much tighter as this evidence is much weaker.

We close this section by showing that Jeffrey's rule commits to a probability distribution which minimizes our distance measure. Hence, Jeffrey's rule leads to the strongest bound on the amount of belief change.

Theorem 6 *The new distribution Pr' obtained by applying Jeffrey's rule to an initial distribution Pr is optimal in the following sense. Among all possible distributions that assign probabilities $q_1, \ldots, q_n$ to events $\gamma_1, \ldots, \gamma_n$, Pr' is the closest to Pr, according to the distance measure defined in Definition 1.*

Pearl's method We now consider Pearl's method of virtual evidence. According to this method, we also have a new evidence η that bears on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, but the evidence is not specified as a set of new probabilities for these events. Instead, for each γ_i, $i \neq 1$, we are given a number λ_i which is interpreted as the ratio $Pr(\eta \mid \gamma_i)/Pr(\eta \mid \gamma_1)$. That is, λ_i represents the likelihood ratio that we would obtain evidence η given γ_i, compared with given γ_1. Note that under this interpretation, we must have $\lambda_1 = 1$.

Definition 5 *Let Pr be a probability distribution over worlds w, and let $\gamma_1, \ldots, \gamma_n$ be a set of mutually exclusive and exhaustive events that are assigned probabilities $p_1, \ldots, p_n$, respectively, by Pr. Suppose we want to change Pr to a new distribution Pr' to incorporate virtual evidence η, specified by $\lambda_1, \ldots, \lambda_n$, with $\lambda_1 = 1$ and $\lambda_i = Pr(\eta \mid \gamma_i)/Pr(\eta \mid \gamma_1)$ if $i \neq 1$. Pearl's method of virtual evidence defines the new distribution Pr' as follows:*

$$Pr'(w) \stackrel{def}{=} Pr(w)\frac{\lambda_i}{\sum_j p_j \lambda_j}, \quad \text{if } w \models \gamma_i.$$

Again, we can easily compute the distance between distributions Pr and Pr' using only local information.

Theorem 7 *Let Pr and Pr' be two distributions, where Pr' is obtained from Pr by accommodating virtual evidence as given by Definition 5. We then have:*

$$D(Pr, Pr') = \ln \max_i \lambda_i - \ln \min_i \lambda_i.$$

This immediately gives us the following bound.

Corollary 3 *If O and O' are the odds functions before and after applying Pearl's method as given in Definition 5, then:*

$$e^{-d} \leq \frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \leq e^{d},$$

where $d = \ln \max_i \lambda_i - \ln \min_i \lambda_i$.

For the special case where our evidence η bears only on $\neg\gamma$ versus γ, with $\lambda = Pr(\eta \mid \gamma)/Pr(\eta \mid \neg\gamma)$, the above bound reduces to $|\ln O'(\alpha \mid \beta) - \ln O(\alpha \mid \beta)| \leq |\ln \lambda|$. Therefore, the bound is tighter when λ is closer to 1. Clearly, when $\lambda = 1$, the evidence is trivial and the two distributions are the same.

Consider the following example from Pearl (1988).

Example 4 *On any given day, there is a burglary on any given house with probability $Pr(b) = 10^{-4}$, while the alarm of Mr. Holmes' house will go off if there is a burglary with probability $Pr(a \mid b) = .95$, and go off if there is no burglary with probability $Pr(a \mid \bar{b}) = .01$. One day, Mr. Holmes' receives a call from his neighbor, Mrs. Gibbons, saying she may have heard the alarm of his house going off. Mr. Holmes concludes that there is an 80% chance that Mrs. Gibbons did hear the alarm going off. According to Pearl's method, this evidence can be interpreted as: $\lambda = Pr(\eta \mid a)/Pr(\eta \mid \bar{a}) = 4$. Therefore, the distance between the original distribution Pr, and the new distribution Pr' which results from incorporating the virtual evidence, is* $|\ln \lambda| = |\ln 4| = 1.386$. *We can use this distance to bound the change in any of our beliefs. In particular, we may want to bound the new probability that there was a burglary at Mr. Holmes' house. Equation 2 gives us:*

$$2.50 \times 10^{-5} \leq Pr'(b) \leq 4.00 \times 10^{-4}.$$

If we actually apply Pearl's method, we get $Pr'(b) = 3.85 \times 10^{-4}$.

Our distance measure is then useful for approximate reasoning given *soft evidence*, as we can use the bound to approximate the probability of any event after the accommodation of such evidence. The approximation itself takes constant time to compute since we only need to compute the distance measure and apply Equation 2. We stress, however, that the bound becomes trivial in the case of *hard evidence* since the initial and new distributions no longer have the same support in this case, making the distance between them infinitely large.

We close this section by a final application of our distance measure, relating to the notion of *evidence strength.*

Example 5 *Going back to Example 3, we ask: What kind of evidence will assure us that our belief in the cloth being green given that it is sold tomorrow, which is now at $.214$, would not exceed $.3$? Equation 2 can be used in this case to obtain a sufficient condition on the strength of evidence which will ensure this. Specifically, Equation 2 gives us:*

$$\frac{.214\, e^{-d}}{.214\,(e^{-d}-1)+1} \leq Pr'(c_g|s) \leq \frac{.214\, e^{d}}{.214\,(e^{d}-1)+1}.$$

To ensure that $Pr'(c_g|s) \leq .3$, we must find a distance d that equates the above upper bound to $.3$. A value of $d = .454$ has this property. Hence, any piece of evidence which has a distance of no more than $.454$ from the current distribution on color, $(.3, .3, .4)$, would guarantee that $Pr'(c_g|s)$ does not exceed $.3$. Following are some pieces of evidence which satisfy this condition: $(.25, .25, .5)$, $(.25, .3, .45)$ and $(.35, .3, .35)$.

Conclusion

We proposed a distance measure between two probability distributions, which allows one to bound the amount of belief change that occurs when moving from one distribution to the other. We also contrasted the proposed measure with some well known measures, including KL-divergence, showing how they fail to be the basis for bounding belief change as is done using the proposed measure. We then presented two practical applications of the proposed distance measure: sensitivity analysis in belief networks and probabilistic belief revision. We showed how the distance measure can be easily computed in these applications, and then used it to bound global belief changes that result from either the perturbation of local conditional beliefs or the accommodation of soft evidence. Finally, we showed that two well known techniques in sensitivity analysis and belief revision correspond to the minimization of our proposed distance measure and, hence, can be shown to be optimal from that viewpoint.

Acknowledgments

This work has been partially supported by NSF grant IIS-9988543, MURI grant N00014-00-1-0617, and by DiMI grant 00-10065.

References

Castillo, E.; Gutiérrez, J. M.; and Hadi, A. S. 1997. Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part A (Systems and Humans)* 27:412–423.

Chan, H., and Darwiche, A. 2001. When do numbers really matter? In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, 65–74. San Francisco, California: Morgan Kaufmann Publishers.

Chan, H., and Darwiche, A. 2002. A distance measure for bounding probabilistic belief change. Technical Report D–124, Computer Science Department, UCLA.

Darwiche, A. 2000. A differential approach to inference in Bayesian networks. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, 123–132. San Francisco, California: Morgan Kaufmann Publishers.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Massachusetts: MIT press.

Jeffrey, R. 1965. *The Logic of Decisions*. New York: McGraw-Hill.

Jensen, F. V. 2001. *Bayesian Networks and Decision Graphs*. New York: Springer-Verglag.

Kjaerulff, U., and van der Gaag, L. C. 2000. Making sensitivity analysis computationally efficient. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, 317–325. San Francisco, California: Morgan Kaufmann Publishers.

Kullback, S., and Leibler, R. A. 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22:79–86.

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

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann Publishers.