On the Revision of Probabilistic Beliefs using Uncertain Evidence

Hei Chan and Adnan Darwiche

Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095
{hei,darwiche}@cs.ucla.edu

Abstract

We revisit the problem of revising probabilistic beliefs using uncertain evidence, and report results on several major issues relating to this problem: How should one specify uncertain evidence? How should one revise a probability distribution? How should one interpret informal evidential statements? Should, and do, iterated belief revisions commute? And what guarantees can be offered on the amount of belief change induced by a particular revision? Our discussion is focused on two main methods for probabilistic revision: Jeffrey's rule of probability kinematics and Pearl's method of virtual evidence, where we analyze and unify these methods from the perspective of the questions posed above.

Key words: belief revision, uncertain evidence, belief networks, probabilistic reasoning

1 Introduction

We consider in this paper the problem of revising probabilistic beliefs using uncertain evidence, where beliefs are represented by a probability distribution. There are two main methods for revising probabilistic beliefs in this case. The first method is known as *Jeffrey's rule* and is based on the principle of *probability kinematics,* which can be viewed as a principle for minimizing belief change [8]. The second method is called the *virtual evidence method* and was proposed by Pearl in the context of belief networks, even though it can

---

* A shorter version of this paper appeared in the *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 99–105.

be easily generalized to arbitrary probability distributions, and is based on recasting uncertain evidence as hard evidence on some virtual event [11]. We will analyze both of these methods with respect to the following questions:

(1) How should one specify uncertain evidence?
(2) How should one revise a probability distribution?
(3) How should one interpret informal evidential statements?
(4) Should, and do, iterated belief revisions commute?
(5) What guarantees can be offered on the amount of belief change induced by a particular revision?

To answer the first question, we note that uncertain evidence must be specified as a formal constraint on posterior beliefs. This could be an absolute constraint on posterior beliefs, or a relative constraint on how posterior beliefs should relate to prior beliefs. Yet, the constraint induced by a given evidence will usually not define posterior beliefs uniquely, hence, we need to adopt a principle that commits us to a unique set of posterior beliefs that satisfy the given constraint. This principle will then define the belief revision method. But even if we choose a method for specifying evidence formally, and a method for revising beliefs, there is still the problem of interpreting informal evidential statements, which are usually specified in natural language. These statements may not map directly to our formal specification language for evidence, and it is this process of interpretation that appears to underlie most of the controversies on revision methods.

Our main findings can be summarized as follows. First, Jeffrey's rule and Pearl's method both revise beliefs using the principle of probability kinematics. Whereas Jeffrey's rule explicitly commits to this principle, Pearl's method is based on a different revision principle, yet it implicitly implies the principle of probability kinematics, leading to the same revision method as that of Jeffrey's. The difference between Jeffrey's rule and Pearl's method is in the way uncertain evidence is specified. Jeffrey requires uncertain evidence to be specified in terms of the *effect* it has on beliefs once accepted, which is a function of both evidence strength and beliefs held before the evidence is obtained. Pearl, on the other hand, requires uncertain evidence to be specified in terms of its *strength* only. Despite this difference, we will show that one can easily translate between the two methods of specifying evidence and provide the equations for carrying out this translation.

The multiplicity of methods for specifying evidence also raises an important question: how should informal statements about evidence be captured formally using available methods? For example, what should the following statement translate to: "Seeing these clouds, I believe there is an 80% chance that it will rain?" We will discuss the differences in interpreting informal evidential statements, where we emphasize its subtlety and show how it appears to be

the culprit in reaching different conclusions by different revision methods.

As to the question of iterated belief revisions, it is well known that Jeffrey's rule does not commute, and hence, the order in which different pieces of uncertain evidence are accepted matters [4]. This has long been perceived as a problem, until clarified recently by the work of Wagner who observed that the method of specifying evidence used by Jeffrey's rule is dependent on what is believed before the evidence is obtained, and hence, should not be commutative to start with [13]. Wagner proposed a method for specifying evidence, based on the notion of a *Bayes factor*, and argued that this method specifies only the strength of evidence, and is independent of the beliefs held before attaining the evidence. Wagner argued that when evidence is specified in this particular way, iterated revisions should commute. He even showed that combining this method for specifying evidence with the principle of probability kinematics leads to a revision rule that commutes. We will actually show that Pearl's method of virtual evidence specifies evidence according to Bayes factors, exactly as proposed by Wagner, and hence, corresponds exactly to the proposal he calls for. Therefore, the results we will discuss in this paper unify the two main methods of probabilistic belief revision proposed by Jeffrey and Pearl, and show that differences between them amount only to a difference in the protocol for specifying uncertain evidence.

Our last set of results relate to the problem of providing guarantees on the amount of belief change induced by a revision. We have recently proposed a distance measure for bounding belief change, and showed how one can use it to provide such guarantees [1]. We will demonstrate how this distance measure can be computed when one distribution is obtained from another using the principle of probability kinematics. The guarantees provided by this distance measure can be realized when applying either Jeffrey's rule or Pearl's method, since they both perform revisions based on the principle of probability kinematics.

Proofs of all theorems in this paper can be found in the Appendix.

2 Probability Kinematics and Jeffrey's Rule

Consider the problem of revising a probability distribution Pr given uncertain evidence relating to a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$. One method of specifying uncertain evidence is through the *effect* that it would have on beliefs once accepted. Specifically, according to this method, we have to specify evidence by providing the following set of proba-

bilities:

$$Pr'(\gamma_i) = q_i, \text{ for } i = 1, \ldots, n,$$

where Pr' denotes the new probability distribution that results from accepting the given evidence. To revise the distribution Pr, we must therefore choose a unique posterior distribution Pr' that satisfies the above constraint. The principle of *probability kinematics*, which we define next, assumes that the conditional belief in any event α given any γ_i remains unchanged.

Definition 1 *[8] Suppose that two probability distributions Pr and Pr' disagree on the probabilities they assign to a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$. The distribution Pr' is said to be obtained from Pr by probability kinematics on $\gamma_1, \ldots, \gamma_n$, iff for every event α in the probability space:*

$$Pr(\alpha \mid \gamma_i) = Pr'(\alpha \mid \gamma_i), \text{ for } i = 1, \ldots, n. \tag{1}$$

This concept was proposed by Jeffrey [8] to capture the notion that even though Pr and Pr' disagree on the probabilities of events $\gamma_1, \ldots, \gamma_n$, they agree on their relevance to every event α.

We now introduce the belief revision method of *Jeffrey's rule* [8], which can be viewed as being comprised of two components: a suggestion to specify uncertain evidence as a constraint on the posterior probabilities of events $\gamma_1, \ldots, \gamma_n$; and a proposal to choose the posterior distribution using the principle of probability kinematics.

Definition 2 (Jeffrey's Rule) *Given an original distribution Pr and some uncertain evidence bearing on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, and assuming that such evidence is specified by the set of posterior probabilities:*

$$Pr'(\gamma_i) = q_i, \text{ for } i = 1, \ldots, n, \tag{2}$$

the new posterior distribution Pr' proposed by Jeffrey's rule is as follows:

$$Pr'(\alpha) \stackrel{def}{=} \sum_{i=1}^{n} q_i \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}. \tag{3}$$

Theorem 1 *The posterior distribution Pr' given in Equation 3 is the one and only distribution that satisfies the constraint in Equation 2 and that is obtained from Pr by probability kinematics on $\gamma_1, \ldots, \gamma_n$.*

We stress here that we are drawing a distinction between the principle of probability kinematics and the revision method of Jeffrey's rule, which are often considered synonymous. As we have mentioned, Jeffrey's rule arises from a combination of two proposals:

(1) the principle of probability kinematics;
(2) the specification of uncertain evidence using a posterior distribution.

It is possible for one to combine the principle of probability kinematics with other methods for specifying evidence, as we will discuss later.

We now show an example of using Jeffrey's rule.

Example 1 (Due to Jeffrey) *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 want to know whether, on the next day, the cloth will be sold (s), or not sold ($\overline{s}$). Our original state of belief is given by the distribution Pr:*

$$Pr(s, c_g) = .12, \ Pr(s, c_b) = .12, \ Pr(s, c_v) = .32,$$
$$Pr(\overline{s}, c_g) = .18, \ Pr(\overline{s}, c_b) = .18, \ Pr(\overline{s}, c_v) = .08.$$

Therefore, our original state of belief on the color of the cloth is given by $(Pr(c_g), Pr(c_b), Pr(c_v)) = (.3, .3, .4)$. Assume that we now inspect the cloth by candlelight, and conclude that our new state of belief on the color of the cloth should be $(Pr'(c_g), Pr'(c_b), Pr'(c_v)) = (.7, .25, .05)$. If we revise our beliefs by applying Jeffrey's rule (Equation 3), we get the new distribution Pr':

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

3 Virtual Evidence and Pearl's Method

The problem of revising a probability distribution using uncertain evidence can be approached from a different perspective than that of the principle of probability kinematics. For example, when we have uncertain evidence about some mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, we can recast this evidence as *hard evidence* on some *virtual event* η, where the relevance of $\gamma_1, \ldots, \gamma_n$ to η is uncertain. According to this approach, the uncertainty regarding evidence on $\gamma_1, \ldots, \gamma_n$ is now interpreted as uncertainty in the relevance of $\gamma_1, \ldots, \gamma_n$ to the virtual event η, and this uncertainty is specified by the likelihood of γ_i given this virtual evidence η, $Pr(\eta \mid \gamma_i)$, for $i = 1, \ldots, n$.

This belief revision method, called the *virtual evidence method*, is defined explicitly as follows.

Definition 3 (Virtual Evidence Method) *Given an original distribution Pr and some uncertain evidence η bearing on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, and assuming that such evidence is specified by $\lambda_1, \ldots, \lambda_n$ such that:*

$$Pr(\eta \mid \gamma_1) : \ldots : Pr(\eta \mid \gamma_n) = \lambda_1 : \ldots : \lambda_n, \tag{4}$$

the revised distribution proposed by the virtual evidence method is $Pr(\cdot \mid \eta)$. Moreover, this method assumes that for every event α in the probability space, we have:

$$Pr(\eta \mid \gamma_i, \alpha) = Pr(\eta \mid \gamma_i), \ for \ i = 1, \ldots, n. \tag{5}$$

That is, the virtual event η depends only on the events $\gamma_1, \ldots, \gamma_n$ and is independent of every event α given γ_i, for $i = 1, \ldots, n$

Note that the likelihoods $Pr(\eta \mid \gamma_1), \ldots, Pr(\eta \mid \gamma_n)$ are not essential for the virtual evidence method, but the likelihood ratios $Pr(\eta \mid \gamma_1) : \ldots : Pr(\eta \mid \gamma_n)$ are. Note also that the assumption given by Equation 5 is needed to uniquely define the posterior distribution $Pr(\cdot \mid \eta)$ as shown by the following theorem.

Theorem 2 *Given the constraint in Equation 4, and the assumption of Equation 5, we have:*

$$Pr(\alpha \mid \eta) = \frac{\sum_{i=1}^{n} \lambda_i Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}. \tag{6}$$

Hence, under the assumption of Equation 5, the virtual evidence method is able to reduce the incorporation of uncertain evidence into that of incorporating certain evidence using Bayes' conditioning.

The virtual evidence method is a generalization of Pearl's method of virtual evidence, which Pearl proposed in the context of Bayesian belief networks [11]. The closed form of this method as given by Equation 6 for arbitrary probability distributions is original as far as we know.

We now show an example of using the virtual evidence method.

Example 2 (Due to Pearl) *Assume that we are concerned with whether the alarm of Mr. Holmes' house is triggered (values a and $\bar{a}$), and whether there*

is burglary at his house (values b and $\overline{b}$). Our original state of belief is given by the distribution Pr:

$$Pr(a, b) = .000095, \; Pr(a, \overline{b}) = .009999,$$
$$Pr(\overline{a}, b) = .000005, \; Pr(\overline{a}, \overline{b}) = .989901.$$

This means that on any given day, there is a burglary at Mr. Holmes' house with probability $Pr(b) = 1 \times 10^{-4}$. One day, Mr. Holmes receives a call from his neighbor, Mrs. Gibbons, saying she may have heard the alarm of his house being triggered. Since Mrs. Gibbons suffers from a hearing problem, Mr. Holmes concludes that there is an 80% chance that Mrs. Gibbons did hear the alarm triggered. This can be interpreted as follows: The probability that Mrs. Gibbons will make the call given that the alarm has triggered is 4 times the probability that Mrs. Gibbons will make the call given that the alarm did not trigger. This uncertain evidence can be recast as hard evidence on the virtual event η (the event of Mrs. Gibbons calling), with likelihood ratios $\lambda_a : \lambda_{\overline{a}} = Pr(\eta \mid a) : Pr(\eta \mid \overline{a}) = 4 : 1$. We can apply Equation 6 and obtain the new distribution $Pr(\cdot \mid \eta)$:

$$Pr(a, b \mid \eta) \approx .000369, \; Pr(a, \overline{b} \mid \eta) \approx .038820,$$
$$Pr(\overline{a}, b \mid \eta) \approx .000005, \; Pr(\overline{a}, \overline{b} \mid \eta) \approx .960806.$$

Therefore, the new probability that there is a burglary at Mr. Holmes' house after revising our beliefs given this piece of virtual evidence is $Pr(b \mid \eta) \approx 3.74 \times 10^{-4}$.

4   Comparing the Revision Methods

From the illustrations of the two belief revision methods, Jeffrey's rule and Pearl's method of virtual evidence, we can see that a belief revision method can be broken into two parts: a formal constraint that is used to specify the uncertain evidence, and a principle of belief revision that commits to a unique distribution among many which satisfy the evidential constraint. For example, Jeffrey's rule specifies evidence using posterior probabilities, while Pearl's method specifies evidence using likelihood ratios; Jeffrey's rule obeys the principle of probability kinematics explicitly, while in Pearl's method, beliefs are revised by conditioning on a virtual event η. In this section, we will compare the two revision methods with respect to these two parts, and consequently show how we can translate between the two methods. Moreover,

we will also show how we can apply these two methods in the context of belief networks.

4.1 Pearl's Method and Probability Kinematics

Our first result shows that the principle of belief revision underlying Pearl's method does indeed satisfy the principle of probability kinematics. Therefore, it actually uses the same belief revision principle as Jeffrey's rule, and what they only differ in is how uncertain evidence is specified.

Theorem 3 *The probability distribution $Pr(\cdot \mid \eta)$ given in Definition 3 is obtained from the original distribution Pr by probability kinematics on the set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$.*

Theorem 3 clarifies one misconception that Jeffrey's rule and Pearl's method are two different belief revision methods. In fact, what they differ in is not how beliefs are revised as they both obey the principle of probability kinematics, but in the constraint that is used to specify evidence. We will next show how we can easily translate between the two different evidence specification methods.

4.2 From Pearl's Method to Jeffrey's Rule

First, we show how we can translate from the evidential constraint used by Pearl's method into one used by Jeffrey's rule.

Theorem 4 *Suppose that we have an original distribution Pr and some uncertain evidence η bearing on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, and suppose further that such evidence is specified by likelihood ratios $\lambda_1, \ldots, \lambda_n$ such that:*

$$Pr(\eta \mid \gamma_1) : \ldots : Pr(\eta \mid \gamma_n) = \lambda_1 : \ldots : \lambda_n.$$

The new posterior distribution $Pr(\cdot \mid \eta)$ proposed by the virtual evidence method (Definition 3) can be obtained using Jeffrey's rule (Definition 2) given that the uncertain evidence is specified by the following set of posterior probabilities:

$$Pr'(\gamma_i) = q_i = Pr(\gamma_i \mid \eta), \ \text{for } i = 1, \ldots, n. \tag{7}$$

We now illustrate the translation from Pearl's method to Jeffrey's rule in Theorem 4 by revisiting Example 2.

Example 3 *In Example 2, the new distribution $Pr(\cdot \mid \eta)$ is obtained from the original distribution Pr by applying revision using Pearl's method. By Equation 7, the equivalent distribution $Pr' = Pr(\cdot \mid \eta)$ can be obtained by applying Jeffrey's rule, given uncertain evidence specified by the following set of posterior probabilities:*

$$
\begin{aligned}
Pr'(a) &= Pr(a \mid \eta) \\
&= \frac{\lambda_a Pr(a)}{\lambda_a Pr(a) + \lambda_{\overline{a}} Pr(\overline{a})} \quad \text{(using Equation 6)} \\
&= \frac{4 \times .010094}{4 \times .010094 + 1 \times .989906} \\
&\approx .039189; \\
Pr'(\overline{a}) &= Pr(\overline{a} \mid \eta) \\
&= \frac{\lambda_{\overline{a}} Pr(\overline{a})}{\lambda_a Pr(a) + \lambda_{\overline{a}} Pr(\overline{a})} \quad \text{(using Equation 6)} \\
&= \frac{1 \times .989906}{4 \times .010094 + 1 \times .989906} \\
&\approx .960811.
\end{aligned}
$$

4.3   From Jeffrey's Rule to Pearl's Method

We now show how we can translate from the evidential constraint used by Jeffrey's rule into one used by Pearl's method.

Theorem 5 *Suppose that we have an original distribution Pr and some uncertain evidence bearing on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, and suppose further that such evidence is specified by a set of posterior probabilities:*

$$
Pr'(\gamma_i) = q_i, \ \text{for } i = 1, \ldots, n.
$$

The new posterior distribution Pr' proposed by Jeffrey's rule (Definition 2) can be obtained using the virtual evidence method (Definition 3) given that the uncertain evidence is specified by the following likelihood ratios:

$$
\lambda_1 : \ldots : \lambda_n = \frac{q_1}{Pr(\gamma_1)} : \ldots : \frac{q_n}{Pr(\gamma_n)}. \tag{8}
$$

We now illustrate the translation from Jeffrey's rule to Pearl's method in Theorem 5 by revisiting Example 1.

Example 4 *In Example 1, the new distribution Pr' is obtained from the original distribution Pr by applying revision using Jeffrey's rule. By Equation 8, the equivalent distribution $Pr(\cdot \mid \eta) = Pr'$ can be obtained by applying Pearl's method, given virtual evidence η specified by the following likelihood ratios:*

$$
\begin{aligned}
\lambda_g : \lambda_b : \lambda_v &= \frac{Pr'(c_g)}{Pr(c_g)} : \frac{Pr'(c_b)}{Pr(c_b)} : \frac{Pr'(c_v)}{Pr(c_v)} \\
&= \frac{.7}{.3} : \frac{.25}{.3} : \frac{.05}{.4} \\
&= 7 : 2.5 : .375.
\end{aligned}
$$

4.4 Revision in Belief Networks

In this subsection, we describe the procedure of applying belief revision in belief networks [11]. A *belief network* (or Bayesian 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 [11,10]. The CPT of variable X with parents $\mathbf{U}$ defines a set of conditional probabilities of the form $Pr(x \mid \mathbf{u})$, where x is a value of variable X, and $\mathbf{u}$ is an instantiation of parents $\mathbf{U}$. Given a network structure and the set of CPTs, a unique probability distribution is defined, and we can compute any probabilistic queries by performing inference on the network.

4.4.1 Pearl's Method in Belief Networks

The method of revision by virtual evidence was first introduced by Pearl in the context of belief networks [11]. Suppose that we have some virtual evidence η bearing on variable Y in a belief network, which has values $y_1, \ldots, y_n$. This virtual evidence is represented in the belief network by adding an auxiliary variable Z, and a directed edge $Y \rightarrow Z$, where one value of Z, say z, corresponds to the virtual event η. This ensures the key assumption described by Equation 5, that the virtual event z is independent of every event α given y_i, i.e., $Pr(z \mid y_i, \alpha) = Pr(z \mid y_i)$, for $i = 1, \ldots, n$, which follows from the independence semantics of belief networks [11]. The uncertainty of evidence is quantified by the likelihood ratios $\lambda_1, \ldots, \lambda_n$, and the CPT of variable Z is assigned such that $Pr(z \mid y_1) : \ldots : Pr(z \mid y_n) = \lambda_1 : \ldots : \lambda_n$.[1] Finally, we accommodate the presence of the virtual event z by asserting the observation $Z = z$ in the belief network. We now show a simple example.

Example 5 *We can represent our probability distribution in Example 2 us-*

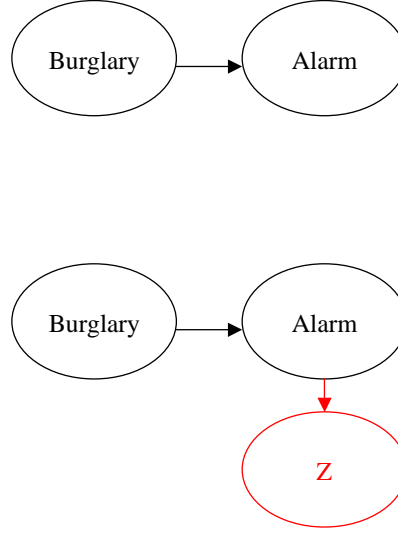[1] There are more than one CPT for Z that satisfies this condition.

Fig. 1. Applying the virtual evidence method in the belief network of Example 5, by adding an auxiliary variable Z as a child of variable *Alarm*.

ing a belief network with two variables: Alarm, which represents whether the alarm of Mr. Holmes' house is triggered (values a and $\bar{a}$); and Burglary, which represents whether there is a burglary at his house (values b and $\bar{b}$). To represent the influence between the two variables, there is a directed edge from Burglary to Alarm. The CPTs of Alarm and Burglary are given by: $Pr(a \mid b) = .95$, meaning the alarm is triggered if there is a burglary with probability .95; $Pr(a \mid \bar{b}) = .01$, meaning the alarm is triggered if there is no burglary with probability .01; and $Pr(b) = 1 \times 10^{-4}$, meaning on any given day, there is a burglary at Mr. Holmes' house with probability 1×10^{-4}. The probability distribution generated by this belief network is the same as our state of belief Pr shown in Example 2.

Now suppose we know that Mr. Holmes receives a call from his neighbor, Mrs. Gibbons, saying she may have heard the alarm of his house being triggered, and concludes that there is an 80% chance that Mrs. Gibbons did hear the alarm triggered. This uncertain evidence can be recast as hard evidence on the virtual event η, with likelihood ratios $\lambda_a : \lambda_{\bar{a}} = 4 : 1$. To incorporate this virtual evidence into the belief network, we add the auxiliary variable Z, and the directed edge Alarm $\rightarrow Z$ (See Figure 1), where the value z of Z corresponds to the virtual event η, and the CPT of Z is assigned such that $Pr(z \mid a) : Pr(z \mid \bar{a}) = 4 : 1$. For example, we can assign $Pr(z \mid a) = .4$ and $Pr(z \mid \bar{a}) = .1$. After asserting the observation $Z = z$ in the belief network, we can easily compute any probabilistic queries by performing inference. For example, the probability that there is a burglary at Mr. Holmes' house is now $Pr(b \mid z) \approx 3.74 \times 10^{-4}$.

4.4.2 Jeffrey's Rule in Belief Networks

There is no known proposal for applying Jeffrey's rule in the context of belief networks. However, because of our earlier results on the translation between Jeffrey's rule and Pearl's method, we immediately get a proposal for this purpose, as we can first translate the evidential constraint used by Jeffrey's rule into one used by the virtual evidence method using Theorem 5, and then perform belief revision by the procedure shown above. We now show a simple example.

Example 6 *We can represent our probability distribution in Example 1 using a belief network with two variables: Color, which represents the color of the cloth (with values c_g, c_b and c_v); and Sold, which represents whether the cloth is sold on the next day (with values s and $\overline{s}$). To represent the influence between the two variables, there is a directed edge from Color to Sold. The CPTs of the variables are given by: $(Pr(c_g), Pr(c_b), Pr(c_v)) = (.3, .3, .4)$; and $Pr(s \mid c_g) = .4$, $Pr(s \mid c_b) = .4$, and $Pr(s \mid c_v) = .8$. The probability distribution generated by this belief network is the same as our state of belief Pr shown in Example 1.*

Now suppose that we inspect the cloth by candlelight, and conclude that our new state of belief on the color of the cloth should be $(Pr'(c_g), Pr'(c_b), Pr'(c_v)) = (.7, .25, .05)$. To incorporate this uncertain evidence into the belief network, we first have to interpret the inspection of the cloth by candlelight as virtual evidence. In Example 4, we show how we can translate this evidential constraint into one used by the virtual evidence method. The uncertain evidence can now be recast as virtual event η, with likelihood ratios $\lambda_g : \lambda_b : \lambda_v = 7 : 2.5 : .375$. We can now follow the procedure of incorporating virtual evidence: we first add an auxiliary variable Z as a child of variable Color, then assign the CPT of Z with values consistent with the likelihood ratios, and finally assert the observation of the virtual event.

5 Interpreting Evidential Statements

We now turn our attention to the investigation of the evidence specification protocols adopted by Jeffrey's rule and Pearl's method in relation to the problem of formally interpreting evidential statements.[2] Consider the following statement as an example:

   "Looking at this evidence, I am willing to bet $2 : 1$ that David is not the

---

[2] This section is a summary of Pearl's discussions on this issue [12], in the context of the approach we take in this paper by dividing the belief revision process into a evidence specification method and a revision principle.

killer."

This statement can be formally interpreted using either protocol. For example, if α denotes the event that David is not the killer, this statement can be interpreted in two ways:

(1) After accepting the evidence, the probability that David is not the killer becomes twice the probability that David is the killer: $Pr'(\alpha) = 2/3$ and $Pr'(\overline{\alpha}) = 1/3$;

(2) The probability that I will see this evidence η given that David is not the killer is twice the probability that I will see it given that David is the killer: $Pr(\eta \mid \alpha) : Pr(\eta \mid \overline{\alpha}) = 2 : 1$.

The first interpretation translates directly into a formal piece of evidence, Jeffrey's style, and can be characterized as an "All things considered" interpretation since it is a statement about the agent's final beliefs, which are a function of both his prior beliefs and the evidence [12]. On the other hand, the second interpretation translates directly into a formal piece of evidence, Pearl's style, and can be characterized as a "Nothing else considered" interpretation since it is a statement about the evidence only [12].

The two interpretations can lead to contradictory conclusions about the evidence. For example, if we use the "Nothing else considered" approach to interpret our statement, we will conclude that the evidence is against David being the killer. However, if we use the "All things considered" interpretation, it is not clear whether the evidence is for or against David being the killer, unless we know the original probability that David is the killer. If, for example, David is one of four suspects who are equally likely to be the killer, our original state of belief is $Pr(\alpha) = 3/4$. Therefore, this evidence has actually increased the probability that David is the killer! Because of this, Pearl argued for the "Nothing else considered" interpretation, as it provides a summary of the evidence alone, and discussed how people tend to use betting odds to quantify their beliefs even when they are based on the evidence only [12].

Example 2 provides another opportunity to illustrate the subtlety involved in interpreting evidential statements. The evidential statement in this case is "Mr. Holmes concludes that there is an 80% chance that Mrs. Gibbons did hear the alarm triggered." Interpreting this statement using the "All things considered" approach gives us the conclusion that $Pr'(a) : Pr'(\overline{a}) = 4 : 1$, where a denotes the event that the alarm has been triggered. This interpretation assumes that the $4 : 1$ ratio applies to the posterior beliefs in a and $\overline{a}$, after Mr. Holmes has accommodated the evidence provided by Mrs. Gibson. However, in Example 2, this statement was given a "Nothing else considered" interpretation, as by Pearl [11, Page 44–47], where the $4 : 1$ ratio is taken as a quantification of the evidence strength, i.e., the statement is interpreted

13

as $Pr(\eta \mid a) : Pr(\eta \mid \overline{a}) = 4 : 1$, where η is the evidence. In fact, the two interpretations will lead to two different probability distributions, and hence, give us different results for probabilistic queries. For example, if we use the "All things considered" approach in interpreting this evidential statement, the probability of having a burglary is $Pr'(b) = 7.53 \times 10^{-3}$, which is much larger than the probability we get using the "Nothing else considered" approach in Example 2, which is $Pr(b \mid \eta) = 3.74 \times 10^{-4}$.

From the discussions above, the formal interpretation of evidential statements appears to be a non–trivial task, which can be sensitive to context and communication protocols. Regardless of how this is accomplished though, we need to stress that the process of mapping an informal evidential statement into a revised probability distribution involves three distinct elements:

(1) One must adopt a formal method for specifying evidence;
(2) One must interpret the informal evidential statement as a formal piece of evidence, according to the evidence specification method;
(3) One must apply a revision, by mapping the original probability distribution and formal piece of evidence into a new distribution, according to a belief revision principle.

As we have shown previously, Jeffrey's rule and Pearl's method both employ the same belief revision principle, i.e., the principle of probability kinematics. Moreover, although they adopt different formal methods of specifying evidence, one can translate between the two methods.

6 Virtual Evidence and Bayes Factors

In this section, we aim to clarify the virtual evidence method by relating it to some classical concepts in probability theory. Before doing so, we first define the notion of odds.

Definition 4 *Given events α and β, the odds of α given β is defined as:*[3]

$$O(\alpha \mid \beta) \stackrel{def}{=} \frac{Pr(\alpha \mid \beta)}{Pr(\overline{\alpha} \mid \beta)}. \tag{9}$$

In quantifying the strength of some evidence η on a hypotheses γ, we often compute the ratio of the odds of γ before and after accepting the evidence, $O(\gamma \mid \eta)/O(\gamma)$. This ratio is called the *odds factor in favor of γ by η* [6], and

---

[3] Here, the odds are defined only if $Pr(\beta) \neq 0$.

its logarithm is called the *weight of evidence η in favor of γ* [6]. We now show a classical result of how this odds factor can be computed.

Theorem 6 *[5,6] The odds factor $O(\gamma \mid \eta)/O(\gamma)$ is equal to the likelihood ratio $Pr(\eta \mid \gamma)/Pr(\eta \mid \overline{\gamma})$, i.e.:*

$$\frac{O(\gamma \mid \eta)}{O(\gamma)} = \frac{Pr(\eta \mid \gamma)}{Pr(\eta \mid \overline{\gamma})}. \tag{10}$$

Therefore, from Equation 10, if we are given virtual evidence η on events γ and $\overline{\gamma}$, the likelihood ratio $Pr(\eta \mid \gamma)/Pr(\eta \mid \overline{\gamma})$ serves as a measure of the strength of the virtual evidence η on the hypotheses γ.

If we are given the general case where we have uncertain evidence on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$, where $n > 2$, we need to expand our notion of odds factor in the following sense.

Definition 5 *The odds of α_i against α_j given β is defined as the probability ratio:*

$$O(\alpha_i : \alpha_j \mid \beta) \stackrel{def}{=} \frac{Pr(\alpha_i \mid \beta)}{Pr(\alpha_j \mid \beta)}. \tag{11}$$

We can now in turn define the *odds factor in favor of γ_i against γ_j by η* with the ratio $O(\gamma_i : \gamma_j \mid \eta)/O(\gamma_i : \gamma_j)$. This ratio of new–to–old odds is called the *Bayes factor* [5,6,9]. The formal definition of the Bayes factor is given as follows.

Definition 6 *Given two distributions Pr and Pr', the Bayes factor for events γ_i and γ_j is defined as:*

$$F_{Pr',Pr}(\gamma_i : \gamma_j) \stackrel{def}{=} \frac{O'(\gamma_i : \gamma_j)}{O(\gamma_i : \gamma_j)} = \frac{Pr'(\gamma_i)/Pr'(\gamma_j)}{Pr(\gamma_i)/Pr(\gamma_j)}. \tag{12}$$

In fact, one can revise beliefs using the notion of Bayes factors. We first specify uncertain evidence on a set of mutually exclusive and exhaustive events $\gamma_1, \ldots, \gamma_n$ by providing the Bayes factor for every pair of events γ_i and γ_j, then commit to the principle of probability kinematics for belief revision. One interesting property of this method of specification is that Bayes factors do not constrain the prior distribution Pr, i.e., any uncertain evidence specified by Bayes factors is compatible with every distribution Pr. [4] Hence, they are suitable for a "Nothing else considered" interpretation of evidential statements.

---

[4] This is not true if we use ratios of probabilities instead of ratios of odds. For example, if $Pr'(\alpha) = 2Pr(\alpha)$, we must have $Pr(\alpha) \le .5$ because $Pr'(\alpha) \le 1$ [13].

In fact, we can show that this revision method using Bayes factors corresponds to the virtual evidence method. This has a number of implications. First, it provides an alternative and more classical semantics for the virtual evidence method. Second, it again confirms that the virtual evidence method obeys the principle of probability kinematics. Finally, it shows that revisions by the virtual evidence method are commutative, as we will illustrate later. The following theorem shows how we can easily find the Bayes factors when we specify uncertain evidence using virtual evidence.

Theorem 7 *Let $Pr(\cdot \mid \eta)$ and Pr be the two distributions given in Definition 3. We then have:*

$$F_{Pr(\cdot|\eta),Pr}(\gamma_i : \gamma_j) = \frac{\lambda_i}{\lambda_j}, \ for \ i,j = 1, \ldots, n. \tag{13}$$

Therefore, we can obtain the same distribution as $Pr(\cdot \mid \eta)$ if we specify the uncertain evidence with the Bayes factor λ_i/λ_j for every pair of events γ_i and γ_j, as shown in Equation 13, and then revise our beliefs according to the principle of probability kinematics. The advantage of using the virtual evidence method for specifying uncertain evidence is that we only need to specify the n likelihood ratios $\lambda_1, \ldots, \lambda_n$ in order to define the n^2 Bayes factors that are necessary for belief revision.

6.1 Reasoning about Evidence

As we have said before, the virtual evidence method can be interpreted as a "Nothing else considered" revision method, and does not depend on one's prior beliefs. In fact, this specification of evidence can be reasoned about and interpreted even when we do not have any prior beliefs. We will illustrate this by the following example due to Halpern and Pucella [7].

Assume that Alice has two coins, a fair one and a double–headed one. If she non–probabilistically chooses one of them and tosses it repeatedly, what is the probability of landing heads in a single toss? Without knowing which coin she chooses (and how she chooses it), the only conclusion that can be drawn is that the probability is either 1/2 (if the fair coin is chosen), or 1 (if the double–headed coin is chosen).

Now assume that we know the results of the first 100 tosses, and all of them landed heads. What is the probability that the next coin toss lands heads? We can again conclude that it is still either 1/2 or 1 depending on which coin is used, as either coin cannot be ruled out from our observation. This is hardly useful because no matter how many of these consecutive tosses that

landed heads we witness, the probability that the next toss will land heads remains unchanged, when in fact the probability of the coin being double–headed should increase.

However, this piece of evidential information can be easily expressed and used if we interpret it as virtual evidence. If we denote the event γ as the coin being double–headed, the event $\overline{\gamma}$ as the coin being fair, and the virtual event η as the coin landing heads in a single toss, we can quantify the evidence strength by computing the likelihood ratio:

$$\frac{Pr(\eta \mid \gamma)}{Pr(\eta \mid \overline{\gamma})} = \frac{1}{1/2} = 2.$$

Therefore, this piece of evidence is in favor of the coin being double–headed, no matter what our prior beliefs are. If we witness 100 tosses landing heads, the likelihood ratio of this observation is 2^{100}, which means it is very strongly in favor of the coin being double–headed. Obviously, it is still not possible to determine the posterior probability of the coin being double–headed without knowing its prior probability. For example, we will still believe the coin tossed is unlikely to be double–headed if its prior probability is 10^{-100}.

The advantage of specifying evidential information using virtual evidence is that the evidence can be shared among different agents with different prior beliefs (even those without prior beliefs), and it will be interpreted the same way by the different agents because the specification depends only on the evidence but not the prior beliefs. The likelihood ratios specified in the virtual evidence method capture completely whether the uncertain evidence is in favor or against a hypotheses, and also its strength. Recently, Halpern and Pucella proposed a logic for reasoning about evidence [7], which essentially views evidence as a confirmation function from the prior beliefs before making the observation, to the posterior beliefs after making the observation. The measure of evidence they use is the likelihood ratio, because it is the only function that does not assume that we have any prior beliefs on the hypotheses.

6.2 Commutativity of Iterated Revisions

We now discuss the problem of the commutativity of iterated revisions, i.e., whether the order in which we accept uncertain evidence matters. [5]

---

[5] There is a key distinction between iterated revisions using certain evidence versus uncertain evidence. In the former case, pieces of evidence may be logically inconsistent, which adds another dimension of complexity to the problem [3], leading to different properties and treatments.

It is well known that iterated revisions by Jeffrey's rule are not commutative [4]. As a simple example, assume that we are given a piece of uncertain evidence which suggests that the probability of event α is .7, followed by another piece of uncertain evidence which suggests that the probability of α is .8. After accepting both pieces of evidence in this particular order using Jeffrey's rule, we believe that the probability of α is .8. However, if the reversed order of revision is employed, we believe that the probability of α is .7. In general, even if we are given pieces of uncertain evidence on different events, iterated revisions by Jeffrey's rule are not commutative.

This was viewed as a problematic aspect of Jeffrey's rule for a long time, until clarified recently by Wagner [13]. First, Wagner observed and stressed that the evidence specification method adopted by Jeffrey is suitable for the "All things considered" interpretation of evidential statements. Moreover, he argued convincingly that when evidential statements carry this interpretation, they must not be commutative to start with. So the lack of commutativity is not a problem of the revision method, but a property of the evidence specification method.

On the other hand, revisions by the virtual evidence method is commutative, and this is supported by Wagner, which suggested specifying evidence based on Bayes factors leads to commutativity [13]. Interestingly enough, he showed that when evidence is specified by Bayes factors and the revision method obeys the principle of probability kinematics, belief revision becomes commutative. [6] These two properties are satisfied by the virtual evidence method, as shown in earlier sections.

7 Bounding Belief Change Induced by Probability Kinematics

One important question relating to belief revision is that of measuring the extent to which a revision disturbs existing beliefs. We have recently proposed a distance measure defined between two probability distributions which can be used to bound the amount of belief change induced by a revision [1]. We review this measure next and then use it to provide guarantees on any revision

---

[6] Wagner shows not only that the representation of uncertain evidence using Bayes factors is sufficient for commutativity, but in a large number of cases, necessary.

which is based on the principle of probability kinematics. [7]

Definition 7 *[1] Let Pr and Pr' be two probability distributions over the same set of worlds ω. We define a measure $D(Pr, Pr')$ as follows:*

$$D(Pr, Pr') \stackrel{def}{=} \ln \max_{\omega} \frac{Pr'(\omega)}{Pr(\omega)} - \ln \min_{\omega} \frac{Pr'(\omega)}{Pr(\omega)}, \tag{14}$$

where $0/0$ is defined as 1. This measure can also be expressed using Bayes factors:

$$D(Pr, Pr') = \ln \max_{\omega_i, \omega_j} F_{Pr', Pr}(\omega_i : \omega_j). \tag{15}$$

This measure satisfies the three properties of distance: positiveness, symmetry, and the triangle inequality. It is useful to compute this distance measure between two probability distributions as it allows us to bound the difference in the beliefs captured by them.

Theorem 8 *[1,2] Let Pr and Pr' be two probability distributions over the same set of worlds. Let α and β be two events. We have the following bound:*

$$e^{-D(Pr, Pr')} \leq \frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \leq e^{D(Pr, Pr')}, \tag{16}$$

where $O(\alpha \mid \beta)$ is the odds of α given β under Pr, and $O'(\alpha \mid \beta)$ is the odds of α given β under Pr'. The bound is tight in the sense that for every pair of distributions Pr and Pr', there are events α and β such that:

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

According to Theorem 8, if we are given a belief revision method, and are able to compute the distance measure between the original and revised distributions, we can get a tight bound on the new probability of any conditional event after the belief revision, given its original probability, using Equation 16. Compared with other popular measures such as KL-divergence and Euclidean

---

[7] The results in this section are reformulations and generalizations of previous results [1,2], and are inspired by a new understanding of Jeffrey's rule and Pearl's method as two belief revision methods based on the principle of probability kinematics, and the understanding of Pearl's method in terms of Bayes factors. Proofs of theorems in this section are included in the Appendix for the purpose of completeness.

distance, this distance measure is the only one that can bound belief change in the precise way of providing a tight bound on the new probability of any conditional event [1,2].

We now compute the distance measure for belief revision methods based on the principle of probability kinematics.

Theorem 9 *[1,2] If Pr' is obtained from Pr by probability kinematics on $\gamma_1, \ldots, \gamma_n$, the distance measure between Pr and Pr' is given by:*

$$D(Pr, Pr') = \ln \max_{i=1}^{n} \frac{Pr'(\gamma_i)}{Pr(\gamma_i)} - \ln \min_{i=1}^{n} \frac{Pr'(\gamma_i)}{Pr(\gamma_i)}. \tag{17}$$

Using Equation 17, we can easily compute the distance measure for revisions based on Jeffrey's rule and Pearl's method.

Corollary 1 *If Pr' is obtained from Pr by applying Jeffrey's rule, given uncertain evidence specified by the set of posterior probabilities $Pr'(\gamma_i) = q_i$, for $i = 1, \ldots, n$, the distance measure between Pr and Pr' is given by:*

$$D(Pr, Pr') = \ln \max_{i=1}^{n} \frac{q_i}{Pr(\gamma_i)} - \ln \min_{i=1}^{n} \frac{q_i}{Pr(\gamma_i)}. \tag{18}$$

Corollary 2 *If $Pr(\cdot \mid \eta)$ is obtained from Pr by applying Pearl's method, given virtual evidence η specified by likelihood ratios $\lambda_1, \ldots, \lambda_n$, the distance measure between Pr and $Pr(\cdot \mid \eta)$ is given by:*

$$D(Pr, Pr(\cdot \mid \eta)) = \ln \max_{i=1}^{n} \lambda_i - \ln \min_{i=1}^{n} \lambda_i. \tag{19}$$

The significance of Corollaries 1 and 2 is that we can compute the distance measure easily in both cases. For Jeffrey's rule, we can compute the distance measure by knowing only the prior and posterior probabilities of events $\gamma_1, \ldots, \gamma_n$. For Pearl's method, we can compute the distance measure by knowing only the likelihood ratios $\lambda_1, \ldots, \lambda_n$. For both revision methods, the distance measure can be computed in constant time from the uncertain evidence, and we can guarantee a bound on the belief change due to the fact that they both obey the principle of probability kinematics, without explicitly knowing the original and posterior distributions.

We close this section by showing that the principle of probability kinematics is optimal in a very precise sense: it commits to a probability distribution that minimizes the distance measure.

Theorem 10 *[1,2] The distribution Pr' obtained from Pr by probability kinematics on $\gamma_1, \ldots, \gamma_n$ is optimal in the following sense. Among all possible dis-*

tributions that agree with Pr' on the probabilities of events $\gamma_1, \ldots, \gamma_n$, Pr' is the closest to Pr according to the distance measure given by Definition 7.

8 Conclusion

In this paper, we analyzed two main methods for revising probability distributions given uncertain evidence: Jeffrey's rule and Pearl's method of virtual evidence. We were able to analyze the process of belief revision according to three different aspects: the formal specification of evidence, the belief revision principle, and the interpretation of informal evidential statements. We showed that both Jeffrey's rule and Pearl's method obey the belief revision principle of probability kinematics, with the difference in the manner in which they specify uncertain evidence: a set of posterior probabilities are used in Jeffrey's rule, while likelihood ratios are used in Pearl's method. We also showed how we can easily translate between the two specification of uncertain evidence, and with this result, we can implement both methods in the context of belief networks by adding virtual evidence nodes. For the much debated problem of interpreting informal evidential statements, we emphasized that the two methods commit to two different interpretations of evidence, and thus can lead to different conclusions about the evidential statement.

Moreover, we showed that the virtual evidence method can be reformulated in terms of Bayes factors, which implies a number of results, including the ability to reason and share evidential information among agents with different prior beliefs, and the commutativity of iterated revisions. Finally, we showed that revisions based on the principle of probability kinematics are optimal in a very precise way, and pointed to a distance measure for bounding belief change due to these revisions. Our bounds included Jeffrey's rule and Pearl's method as special cases.

References

[1] Hei Chan and Adnan Darwiche. A distance measure for bounding probabilistic belief change. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 539–545. AAAI Press, Menlo Park, 2002.

[2] Hei Chan and Adnan Darwiche. A distance measure for bounding probabilistic belief change. To be published in *International Journal of Approximate Reasoning*, 2004.

[3] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial Intelligence* 87:1–29, 1997.

[4] Persi Diaconis and Sandy L. Zabell. Updating subjective probability. *Journal of the American Statistical Association* 77:822–830, 1982.

[5] Irving J. Good. *Probability and the Weighing of Evidence*. Charles Griffin, London, 1950.

[6] Irving J. Good. *Good Thinking: the Foundations of Probability and Its Applications*. University of Minnesota Press, Minneapolis, 1983.

[7] Joseph Y. Halpern and Riccardo Pucella A logic for reasoning about evidence. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, pages 297–304. Morgan Kaufmann, San Francisco, 2003.

[8] Richard C. Jeffrey. *The Logic of Decision*. McGraw–Hill, New York, 1965. (2nd edition) University of Chicago Press, Chicago, 1983. (Paperback correction) 1990.

[9] Richard C. Jeffrey. *Probability and the Art of Judgement*. Cambridge University Press, Cambridge, 1992.

[10] Finn V. Jensen. *Bayesian Networks and Decision Graphs*. Springer–Verglag, New York, 2001.

[11] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, 1988.

[12] Judea Pearl. On two pseudo–paradoxes in Bayesian analysis. *Annals of Mathematics and Artificial Intelligence* 32:171–177, 2001.

[13] Carl Wagner. Probability kinematics and commutativity. *Philosophy of Science* 69:266–278, 2002.

A   Proofs

Proof of Theorem 1   Given the distribution Pr' in Equation 3, since events $\gamma_1, \ldots, \gamma_n$ are mutually exclusive, we have:

$$Pr'(\gamma_i) = q_i \frac{Pr(\gamma_i)}{Pr(\gamma_i)} = q_i, \text{ for } i = 1, \ldots, n.$$

satisfying Equation 2, and:

$$Pr'(\alpha, \gamma_i) = q_i \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}, \text{ for } i = 1, \ldots, n.$$

Therefore, Pr' is obtained from Pr by probability kinematics on $\gamma_1, \ldots, \gamma_n$, i.e., it satisfies Equation 1, since:

$$
\begin{aligned}
Pr'(\alpha \mid \gamma_i) &= \frac{Pr'(\alpha, \gamma_i)}{Pr'(\gamma_i)} \\
&= \frac{q_i \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}}{q_i} \\
&= \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)} \\
&= Pr(\alpha \mid \gamma_i), \text{ for } i = 1, \ldots, n.
\end{aligned}
$$

On the other hand, if there is a distribution Pr' that satisfies both Equations 1 and 2, the probability of event α under Pr' must be:

$$
\begin{aligned}
Pr'(\alpha) &= \sum_{i=1}^{n} Pr'(\alpha \mid \gamma_i) Pr'(\gamma_i) \\
&= \sum_{i=1}^{n} Pr(\alpha \mid \gamma_i) q_i \\
&= \sum_{i=1}^{n} q_i \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}.
\end{aligned}
$$

Proof of Theorem 2   We want to prove that the distribution given by Equation 6 is the unique distribution $Pr(\cdot \mid \eta)$ identified by the virtual evidence method in Definition 3. First of all, if there is a distribution $Pr(\cdot \mid \eta)$ that satisfies both Equations 4 and 5, the probability of event α under distribution $Pr(\cdot \mid \eta)$ must be:

23

$$Pr(\alpha \mid \eta) = \frac{Pr(\alpha, \eta)}{Pr(\eta)}$$

$$= \frac{\sum_{i=1}^{n} Pr(\alpha, \eta, \gamma_i)}{Pr(\eta)}$$

$$= \frac{\sum_{i=1}^{n} Pr(\eta \mid \gamma_i, \alpha) Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} Pr(\eta \mid \gamma_j) Pr(\gamma_j)}$$

$$= \frac{\sum_{i=1}^{n} Pr(\eta \mid \gamma_i) Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} Pr(\eta \mid \gamma_j) Pr(\gamma_j)}$$

$$= \frac{\sum_{i=1}^{n} (\lambda_i/k) Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} (\lambda_j/k) Pr(\gamma_j)}$$

$$= \frac{\sum_{i=1}^{n} \lambda_i Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}.$$

On the other hand, given a distribution which satisfies Equation 6, since events $\gamma_1, \ldots, \gamma_n$ are mutually exclusive, we have:

$$Pr(\eta \mid \gamma_1) : \ldots : Pr(\eta \mid \gamma_n) = \frac{Pr(\eta, \gamma_1)}{Pr(\gamma_1)} : \ldots : \frac{Pr(\eta, \gamma_n)}{Pr(\gamma_n)}$$

$$= \frac{Pr(\gamma_1 \mid \eta) Pr(\eta)}{Pr(\gamma_1)} : \ldots : \frac{Pr(\gamma_n \mid \eta) Pr(\eta)}{Pr(\gamma_n)}$$

$$= \frac{\frac{\lambda_1 Pr(\gamma_1)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)} Pr(\eta)}{Pr(\gamma_1)} : \ldots : \frac{\frac{\lambda_n Pr(\gamma_n)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)} Pr(\eta)}{Pr(\gamma_n)}$$

$$= \lambda_1 : \ldots : \lambda_n,$$

satisfying Equation 4. Therefore, $\lambda_i = k Pr(\eta \mid \gamma_i)$, where k is a constant, and Equation 5 is also satisfied since:

$$Pr(\eta \mid \gamma_i, \alpha) = \frac{Pr(\eta, \gamma_i, \alpha)}{Pr(\gamma_i, \alpha)}$$

$$= \frac{Pr(\gamma_i, \alpha \mid \eta) Pr(\eta)}{Pr(\gamma_i, \alpha)}$$

$$= \frac{\frac{\lambda_i Pr(\gamma_i, \alpha)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)} Pr(\eta)}{Pr(\gamma_i, \alpha)}$$

$$= \frac{\lambda_i Pr(\eta)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}$$

$$= \frac{k Pr(\eta \mid \gamma_i) Pr(\eta)}{\sum_{j=1}^{n} k Pr(\eta \mid \gamma_j) Pr(\gamma_j)}$$

$$= \frac{Pr(\eta \mid \gamma_i) Pr(\eta)}{\sum_{j=1}^{n} Pr(\eta \mid \gamma_j) Pr(\gamma_j)}$$

$$= \frac{Pr(\eta \mid \gamma_i)Pr(\eta)}{Pr(\eta)}$$
$$= Pr(\eta \mid \gamma_i), \ \text{for } i = 1, \dots, n.$$

Proof of Theorem 3 Given the distribution $Pr(\cdot \mid \eta)$ in Equation 6, since events $\gamma_1, \dots, \gamma_n$ are mutually exclusive, we have:

$$Pr(\gamma_i \mid \eta) = \frac{\lambda_i Pr(\gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}, \ \text{for } i = 1, \dots, n.$$

and:

$$Pr(\alpha, \gamma_i \mid \eta) = \frac{\lambda_i Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}, \ \text{for } i = 1, \dots, n.$$

Therefore, $Pr(\cdot \mid \eta)$ is obtained from Pr by probability kinematics on $\gamma_1, \dots, \gamma_n$, i.e., it satisfies Equation 1, since:

$$Pr(\alpha \mid \gamma_i, \eta) = \frac{Pr(\alpha, \gamma_i \mid \eta)}{Pr(\gamma_i \mid \eta)}$$
$$= \frac{\frac{\lambda_i Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}}{\frac{\lambda_i Pr(\gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}}$$
$$= \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}$$
$$= Pr(\alpha \mid \gamma_i), \ \text{for } i = 1, \dots, n.$$

Proof of Theorem 4 From Equations 6 and 7, we have:

$$Pr'(\gamma_i) = q_i = \frac{\lambda_i Pr(\gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)},$$

for $i = 1, \dots, n$. We can substitute the set of posterior probabilities into Jeffrey's rule (Equation 3), and get:

$$Pr'(\alpha) = \sum_{i=1}^{n} q_i \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}$$
$$= \sum_{i=1}^{n} \frac{\lambda_i Pr(\gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)} \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)}$$

25

$$= \frac{\sum_{i=1}^{n} \lambda_i Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}.$$

This is exactly the distribution obtained by the virtual evidence method (Equation 6), with likelihood ratios $\lambda_1, \ldots, \lambda_n$.

Proof of Theorem 5 From Equation 8, we have $\lambda_i = kq_i/Pr(\gamma_i)$, where k is a constant, for $i = 1, \ldots, n$, and we can substitute the likelihood ratios $\lambda_1, \ldots, \lambda_n$ into Pearl's method (Equation 6), and get:

$$
\begin{aligned}
Pr(\alpha \mid \eta) &= \frac{\sum_{i=1}^{n} \lambda_i Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} \lambda_j Pr(\gamma_j)}. \\
&= \frac{\sum_{i=1}^{n} (kq_i/Pr(\gamma_i)) Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} (kq_j/Pr(\gamma_j)) Pr(\gamma_j)}. \\
&= \frac{\sum_{i=1}^{n} (q_i/Pr(\gamma_i)) Pr(\alpha, \gamma_i)}{\sum_{j=1}^{n} q_j} \\
&= \sum_{i=1}^{n} q_i \frac{Pr(\alpha, \gamma_i)}{Pr(\gamma_i)},
\end{aligned}
$$

since $\sum_{j=1}^{n} q_j = \sum_{j=1}^{n} Pr'(\gamma_j) = 1$. This is exactly the distribution obtained by Jeffrey's rule (Equation 3), with the set of posterior probabilities $Pr'(\gamma_i) = q_i$, for $i = 1, \ldots, n$.

Proof of Theorem 6 We have:

$$
\begin{aligned}
\frac{O(\gamma \mid \eta)}{O(\gamma)} &= \frac{Pr(\gamma \mid \eta)/Pr(\bar{\gamma} \mid \eta)}{Pr(\gamma)/Pr(\bar{\gamma})} \\
&= \frac{Pr(\eta, \gamma)/Pr(\eta, \bar{\gamma})}{Pr(\gamma)/Pr(\bar{\gamma})} \\
&= \frac{Pr(\eta, \gamma)/Pr(\gamma)}{Pr(\eta, \bar{\gamma})/Pr(\bar{\gamma})} \\
&= \frac{Pr(\eta \mid \gamma)}{Pr(\eta \mid \bar{\gamma})}.
\end{aligned}
$$

Proof of Theorem 7 We have:

$$F_{Pr',Pr}(\gamma_i : \gamma_j) = \frac{Pr'(\gamma_i)/Pr'(\gamma_j)}{Pr(\gamma_i)/Pr(\gamma_j)}$$

$$= \frac{\left(\frac{\lambda_i Pr(\gamma_i)}{\sum_{k=1}^{n} \lambda_k Pr(\gamma_k)}\right) / \left(\frac{\lambda_j Pr(\gamma_j)}{\sum_{k=1}^{n} \lambda_k Pr(\gamma_k)}\right)}{Pr(\gamma_i)/Pr(\gamma_j)}$$

$$= \frac{\lambda_i}{\lambda_j}, \text{ for } i, j = 1, \ldots, n.$$

Proof of Theorem 8   If Pr and Pr' do not have the same support,[8] we have $D(Pr, Pr') = \infty$, and thus:

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

If they have the same support, let $r_\omega = Pr'(\omega)/Pr(\omega)$ for every world ω. The odds ratio $O'(\alpha \mid \beta)/O(\alpha \mid \beta)$ can be expressed as:

$$\frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} = \frac{Pr'(\alpha \mid \beta)/Pr'(\overline{\alpha} \mid \beta)}{Pr(\alpha \mid \beta)/Pr(\overline{\alpha} \mid \beta)}$$

$$= \frac{Pr'(\alpha, \beta)/Pr'(\overline{\alpha}, \beta)}{Pr(\alpha, \beta)/Pr(\overline{\alpha}, \beta)}$$

$$= \frac{(\sum_{\omega \models \alpha, \beta} Pr'(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} Pr'(\omega))}{(\sum_{\omega \models \alpha, \beta} Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} Pr(\omega))}$$

$$= \frac{(\sum_{\omega \models \alpha, \beta} r_\omega Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} r_\omega Pr(\omega))}{(\sum_{\omega \models \alpha, \beta} Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} Pr(\omega))}.$$

We now introduce $\max_\omega r_\omega$ and $\min_\omega r_\omega$ to get the upper bound on the odds ratio:

$$\frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \leq \frac{(\sum_{\omega \models \alpha, \beta} (\max_\omega r_\omega) Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} (\min_\omega r_\omega) Pr(\omega))}{(\sum_{\omega \models \alpha, \beta} Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} Pr'(\omega))}$$

$$= \frac{\max_\omega r_\omega}{\min_\omega r_\omega}$$

$$= \frac{\max_\omega Pr'(\omega)/Pr(\omega)}{\min_\omega Pr'(\omega)/Pr(\omega)}.$$

Similarly, we can also get the lower bound on the odds ratio:

$$\frac{O'(\alpha \mid \beta)}{O(\alpha \mid \beta)} \geq \frac{(\sum_{\omega \models \alpha, \beta} (\min_\omega r_\omega) Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} (\max_\omega r_\omega) Pr(\omega))}{(\sum_{\omega \models \alpha, \beta} Pr(\omega))/(\sum_{\omega \models \overline{\alpha}, \beta} Pr'(\omega))}$$

---

[8] Two probability distributions Pr and Pr' have the same support, if for every world ω, $Pr(\omega) = 0$ iff $Pr'(\omega) = 0$.

$$= \frac{\min_\omega r_\omega}{\max_\omega r_\omega}$$

$$= \frac{\min_\omega Pr'(\omega)/Pr(\omega)}{\max_\omega Pr'(\omega)/Pr(\omega)}.$$

We note that both results can be expressed using our distance measure:

$$e^{D(Pr,Pr')} = \frac{\max_\omega Pr'(\omega)/Pr(\omega)}{\min_\omega Pr'(\omega)/Pr(\omega)};$$

$$e^{-D(Pr,Pr')} = \frac{\min_\omega Pr'(\omega)/Pr(\omega)}{\max_\omega Pr'(\omega)/Pr(\omega)}.$$

Therefore, we have the following inequality: [9]

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

The bound is tight in the sense that for every pair of distributions Pr and Pr', there are events $\alpha = \omega_i$ and $\beta = \omega_i \vee \omega_j$, where $\omega_i = \arg\max_\omega r_\omega$ and $\omega_j = \arg\min_\omega r_\omega$, such that:

$$O'(\alpha \mid \beta) = \frac{\sum_{\omega \models \alpha,\beta} r_\omega Pr(\omega)}{\sum_{\omega \models \overline{\alpha},\beta} r_\omega Pr(\omega)}$$

$$= \frac{r_{\omega_i} Pr(\omega_i)}{r_{\omega_j} Pr(\omega_j)}.$$

Since $O(\alpha \mid \beta) = Pr(\omega_i)/Pr(\omega_j)$ and $e^{D(Pr,Pr')} = (\max_\omega r_\omega)/(\min_\omega r_\omega) = r_{\omega_i}/r_{\omega_j}$, we have:

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

Similarly, we can get:

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

Proof of Theorem 9 If the two sets of probabilities $Pr(\gamma_1), \ldots, Pr(\gamma_n)$ and $Pr'(\gamma_1), \ldots, Pr'(\gamma_n)$ do not have the same support, there must exist some

---

[9] If both $O'(\alpha \mid \beta)$ and $O(\alpha \mid \beta)$ takes on either 0 or ∞, the theorem still holds if we define both 0/0 and ∞/∞ as 1.

world ω where $Pr(\omega) = 0$ and $Pr'(\omega) \neq 0$ or $Pr(\omega) \neq 0$ and $Pr'(\omega) = 0$, and hence, the distributions Pr and Pr' also do not have the same support, giving us $D(Pr, Pr') = \infty = \ln \max_{i=1}^{n} (Pr'(\gamma_i)/Pr(\gamma_i)) - \ln \min_{i=1}^{n} (Pr'(\gamma_i)/Pr(\gamma_i))$. Otherwise, given a world ω where $\omega \models \gamma_i$, from Equation 1, we have:

$$\frac{Pr'(\omega)}{Pr(\omega)} = \frac{Pr'(\omega \mid \gamma_i) Pr'(\gamma_i)}{Pr(\omega \mid \gamma_i) Pr(\gamma_i)}$$
$$= \frac{Pr'(\gamma_i)}{Pr(\gamma_i)}.$$

Therefore, the distance measure can be computed by:

$$D(Pr, Pr') = \ln \max_{\omega} \frac{Pr'(\omega)}{Pr(\omega)} - \ln \min_{\omega} \frac{Pr'(\omega)}{Pr(\omega)}$$
$$= \ln \max_{i=1}^{n} \frac{Pr'(\gamma_i)}{Pr(\gamma_i)} - \ln \min_{i=1}^{n} \frac{Pr'(\gamma_i)}{Pr(\gamma_i)}.$$

Proof of Theorem 10 Let Pr'' be any distribution that satisfies the constraint, $Pr''(\gamma_i) = Pr'(\gamma_i)$, for $i = 1, \ldots, n$. We would like to prove that $D(Pr, Pr'') \geq D(Pr, Pr')$, where Pr' is obtained from Pr by probability kinematics on $\gamma_1, \ldots, \gamma_n$. If Pr and Pr'' do not have the same support, we have $D(Pr, Pr'') = \infty \geq D(Pr, Pr')$. If they have the same support, let $j = \arg \max_{i=1}^{n} (Pr'(\gamma_i)/Pr(\gamma_i))$ and $k = \arg \min_{i=1}^{n} (Pr'(\gamma_i)/Pr(\gamma_i))$. If $r_{\max} = \max_{\omega} (Pr''(\omega)/Pr(\omega))$, we can write the following inequality:

$$r_{\max} Pr(\gamma_j) = \sum_{\omega \models \gamma_j} r_{\max} Pr(\omega)$$
$$\geq \sum_{\omega \models \gamma_j} \frac{Pr''(\omega)}{Pr(\omega)} Pr(\omega)$$
$$= \sum_{\omega \models \gamma_j} Pr''(\omega)$$
$$= Pr''(\gamma_j)$$
$$= Pr'(\gamma_j).$$

This gives us $r_{\max} \geq Pr'(\gamma_j)/Pr(\gamma_j)$. Similarly, if $r_{\min} = \min_{\omega} (Pr''(\omega)/Pr(\omega))$, we can write the following inequality:

$$r_{\min} Pr(\gamma_k) = \sum_{\omega \models \gamma_k} r_{\min} Pr(\omega)$$

$$\leq \sum_{\omega \models \gamma_k} \frac{Pr''(\omega)}{Pr(\omega)} Pr(\omega)$$

$$= \sum_{\omega \models \gamma_k} Pr''(\omega)$$

$$= Pr''(\gamma_k)$$

$$= Pr'(\gamma_k).$$

This gives us $r_{\min} \leq Pr'(\gamma_k)/Pr(\gamma_k)$. Therefore, the distance measure between Pr and Pr'' is:

$$
\begin{aligned}
D(Pr, Pr'') &= \ln \max_{\omega} \frac{Pr''(\omega)}{Pr(\omega)} - \ln \min_{\omega} \frac{Pr''(\omega)}{Pr(\omega)} \\
&= \ln r_{\max} - \ln r_{\min} \\
&\geq \ln \frac{Pr'(\gamma_j)}{Pr(\gamma_j)} - \ln \frac{Pr'(\gamma_k)}{Pr(\gamma_k)} \\
&= \ln \max_{i=1}^{n} \frac{Pr'(\gamma_i)}{Pr(\gamma_i)} - \ln \min_{i=1}^{n} \frac{Pr'(\gamma_i)}{Pr(\gamma_i)} \\
&= D(Pr, Pr').
\end{aligned}
$$

Therefore, the distribution Pr' gives us the smallest distance among all possible distributions that agree with Pr' on the probabilities of events $\gamma_1, \ldots, \gamma_n$.