Confirmation Measures of Association Rule
Interestingness

David H. Glass
School of Computing and Mathematics,
University of Ulster,
Newtownabbey, Co. Antrim,
BT37 0QB, UK
dh.glass@ulster.ac.uk

Author’s final version of the paper.
doi:10.1016/j.knosys.2013.01.021

Abstract

This paper considers advantages of measures of confirmation or evidential support in the context of interestingness of association rules. In particular, it is argued that the way in which they characterize positive/negative association has advantages over other measures such as null-invariant measures. Several properties are reviewed and proposed as requirements for an adequate confirmation measure in a data mining context. While none of the well-known confirmation measures satisfy all of these requirements, two new measures are proposed which do and one of these is shown to have a further advantage. Some results suggest that these measures are relatively stable when the number of null transactions varies.

Keywords: confirmation, evidential support, interestingness, association rule, probability
1 Introduction

A central goal in data mining is the discovery of interesting associations in databases. In many cases these associations are expressed in rules of the form \( A \Rightarrow B \). In a database of transactions an example would be that customers who buy product \( A \) tend to buy product \( B \). An important research question concerns what constitutes an interesting rule and to this end numerous measures of the ‘interestingness’ or strength of association rules have been proposed (for surveys see [1, 2, 3]). The study of such measures is an active area of research and finds widespread application (see for example [4, 5, 6, 7, 8]).

In their review, Geng and Hamilton [2] distinguish between objective, subjective and semantic measures of interestingness. Objective measures are based on data and are typically motivated by statistical considerations, subjective measures take into account not only the data but also the knowledge and interests of users [9], while semantic measures additionally taken into account utility and actionability [10]. The focus in this paper is on objective measures, but even in this case many measures have been proposed. One feature of research in this area is that many properties of such measures have been discussed and compared in the literature either to aid the selection of an appropriate measure for a given context or to determine their theoretical or computational merits (see [3, 11, 12, 13, 14, 15]).

In a very different context, philosophers of science have devoted a lot of attention to measures of confirmation or evidential support (see for example [16, 17]). Confirmation measures are based on probability theory and are such that the degree to which evidence \( E \) confirms or supports hypothesis \( H \) is positive if there is a positive probabilistic dependence between \( E \) and \( H \), zero if \( E \) and \( H \) are independent, and negative if they are negatively dependent. As in the case of interestingness measures, various proposals have been made concerning the desirable properties of confirmations measures. In fact, the two fields are related since a number of interestingness measures are confirmation measures and the role of probabilistic (in)dependence has often been discussed in the context of association rules (see particularly [11]).

Confirmation measures been proposed in the context of data mining [18] and a similar approach was adopted in the context of fuzzy association rules [19]. The current paper builds on this work by exploring further the merits of confirmation measures in a data mining context. A number of properties which have been proposed for confirmation measures in the philosophy lit-
erature are reviewed and their relevance for association rules considered. In particular, three key properties are identified for confirmation measures of interestingness. It turns out that none of the confirmation measures in either the philosophical or data mining literature satisfies all these properties, but two new measures are proposed which do satisfy them.

The structure of the rest of the paper is as follows. Section 2 reviews confirmation measures and some of the main properties considered in the philosophical literature. Section 3 explores confirmation measures in the context of association rules and defines a number of confirmation measures. Section 4 compares confirmation measures with null-invariant measures [20] and section 5 proposes a number of requirements for interestingness measures and evaluates existing confirmation measures in light of them. Section 6 proposes two new measures that satisfy all the proposed requirements, explores their stability as the number of null-transactions varies and provides a comparison with recent work by Greco et al. [21]. Finally, conclusions are presented in section 7.

2 Confirmation measures

The assumption underlying work on probabilistic measures of confirmation (or evidential support) is that the confirmation of an hypothesis provided by a piece of evidence is a matter of degree. Confirmation is also understood to relate to the impact of the evidence on the probability of the hypothesis rather than simply being the posterior probability of the hypothesis given the evidence. Thus, the goal is to find a measure that quantifies this impact in a satisfactory way. A confirmation measure can be defined as follows:

**Definition 1.** A confirmation measure of the degree to which a piece of evidence $E$ confirms an hypothesis $H$, denoted $c(H, E)$, is a measure which satisfies the following properties:

(i) $c(H, E) > 0$ if $Pr(H|E) > Pr(H)$

(ii) $c(H, E) = 0$ if $Pr(H|E) = Pr(H)$

(iii) $c(H, E) < 0$ if $Pr(H|E) < Pr(H)$

where $Pr$ is a probability function.
A large number of confirmation measures have been discussed in the philosophical literature. These include:

\[ d(H, E) = Pr(H|E) - Pr(H), \]

\[ r(H, E) = \log \left[ \frac{Pr(H|E)}{Pr(H)} \right], \]

\[ l(H, E) = \log \left[ \frac{Pr(E|H)}{Pr(E|\neg H)} \right], \]

\[ k(H, E) = \frac{Pr(E|H) - Pr(E|\neg H)}{Pr(E|H) + Pr(E|\neg H)}, \]

\[ s(H, E) = Pr(H|E) - Pr(H|\neg E), \]

\[ b(H, E) = Pr(E, H) - Pr(E)Pr(H). \]

(1)

While, these measures have received a lot of attention in the philosophical literature (see [17] and references therein for further discussion of these measures), some are also widely used in other contexts, particularly the likelihood ratio measure, \( l \). As we shall see later, a number of other confirmation measures are also used as measures of interestingness.

With so many confirmation measures on offer and the fact that it is straightforward to generate new measures, one may wonder whether there is any reason to prefer one measure over another. Eells and Fitelson [22] propose the following symmetry considerations to help discriminate between the various measures:

- Evidence Symmetry (ES): \( c(H, E) = -c(H, \neg E) \)
- Commutativity Symmetry (CS): \( c(H, E) = c(E, H) \)
- Hypothesis Symmetry (HS): \( c(H, E) = -c(\neg H, E) \)
- Total Symmetry (TS): \( c(H, E) = c(\neg H, \neg E) \).

They present a number of arguments to show that a desirable property for a confirmation measure is that it satisfy what we shall call the symmetry requirement which can be stated as follows:
Symmetry Requirement: $c(H, E)$ should satisfy HS, but not ES, CS nor TS.

They use the following example. Suppose a card is randomly drawn from a standard deck. If $E$ represents the card being the seven of spades and $H$ represents it being black, it seems clear that $E$ confirms $H$ much more than $H$ confirms $E$ and so CS is not desirable. Similarly, $\neg E$ tells us very little about $H$ so ES is not desirable either. By contrast HS does seem desirable (or at least reasonable) since it states that if $E$ confirms $H$ to a certain degree it should disconfirm $\neg H$ to the same degree. It follows from these considerations that TS is undesirable. Eells and Fitelson go on to show that $r$, $s$ and $b$ fail to satisfy these requirements whereas $d$, $l$ and $k$ do satisfy them. This gives some reason to prefer and $d$, $l$ and $k$ over the other measures. Crupi et al. [23] proposed an extension of this these symmetry requirements which will be considered in section 5.

A further desideratum discussed by Fitelson [24] is that of logicality\(^1\) and concerns the relationship between confirmation and logic. The idea is that a confirmation measure should be a quantitative generalization of logical entailment and so logicality can be expressed as

Logicality Requirement: $c(H, E)$ should be maximal (minimal) when $E \models H$ ($E \models \neg H$) provided $c(H, E)$ is defined.

In fact, this desideratum on its own would rule out all the above measures except $l$ and $k$, which are closely related to each other since they are ordinally equivalent in the sense that they provide the same orderings of evidence-hypothesis pairs.

Several other proposals, which are related to logicality, can also be considered. Crupi et al. [23] consider a function $v$ such that $v(H, E)$ has the same positive value (e.g. +1) if $E$ entails $H$, $v(H, E)$ has the same negative value (e.g. -1) if $E$ refutes $\neg H$ and $v(H, E) = 0$ otherwise. For a confirmation measure $C$, they then define principle (Ex\(_1\)):

(Ex\(_1\)) If $v(H_1, E_1) > v(H_2, E_2)$ then $c(H_1, E_1) > c(H_2, E_2)$.

(Ex\(_1\)) guarantees that $c(H, E)$ will be greater in cases where $E$ entails $H$ than in other cases and $c(H, E)$ will be less in cases where $E$ entails $\neg H$ than in

\(^1\)This idea as well as hypothesis symmetry can also be found in the paper by Kemeny and Oppenheim [25].
other cases. While this is similar in certain respects to logicality, they point out that the two concepts are independent in the sense that a measure can satisfy (Ex1) and not logicality or vice versa.

Greco et al. [21] have proposed a modification of principle (Ex1), which they call (weak Ex1). It is defined as follows:

\[(\text{weak Ex}_1) \quad \text{If } v(H_1, E_1) > v(H_2, E_2) \text{ and } v(H_1, \neg E_1) < v(H_2, \neg E_2) \text{ then } c(H_1, E_1) > c(H_2, E_2).\]

(weak Ex1) differs from (Ex1) in that it guarantees that \(c(H, E)\) will be greater in cases where \(E\) entails \(H\) and \(\neg E\) entails \(\neg H\) than in cases where these entailments do not hold. They also propose a modification to logicality called (weak L) which will be considered in section 5.1.

Fitelson [24] also discusses another property, which we shall call the posterior requirement, that can be stated as follows:

**Posterior Requirement:** If \(Pr(H|E_1) \geq Pr(H|E_2)\), then \(c(H, E_1) \geq c(H, E_2)\).

It is worth noting that it follows from this requirement that if \(Pr(H|E_1) = Pr(H|E_2)\), then \(c(H, E_1) = c(H, E_2)\). It is easy to show that \(d, r, l, \text{ and } k\) all satisfy this requirement, while \(s\) and \(b\) do not.

Each of the above requirements seems plausible, with the posterior requirement in particular being widely accepted among Bayesian confirmation theorists as Fitelson points out. One possibility, however, is that there are different notions of confirmation and so while the posterior requirement might be suitable for one conception, it might not be for another. And, if this is correct, the same might be true of the other requirements as well. Even if this is not the case, the goal in this paper is not to address the philosophical issues surrounding confirmation, but to investigate the suitability of confirmation measures to quantify the strength of association rules. This means that the appropriate requirements may differ from those in the philosophical case.

3 Confirmation measures and association rule strength

Following Agrawal et al. [26], let \(I = \{i_1, i_2, \ldots, i_n\}\) be a set of binary attributes called items and \(D = \{t_1, t_2, \ldots, t_m\}\) a database of transactions,
where each transaction $t$ is represented as a binary vector with $t[k] = 1$ if $t$ contains the item $i_k$ and $t[k] = 0$ otherwise. Associate with each subset of items in $I$ a standard formal language $L$ where formulas consist of items combined by means of logical connectives. We say that a transaction $t$ satisfies a formula $A$ if $A$ evaluates to true when items in $A$ are assigned true if $t[k] = 1$ and false if $t[k] = 0$. Let $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$ and let $A$ be a formula in the language associated with $X$ and $B$ a formula in the language associated with $Y$. An association rule is an implication of the form $A \Rightarrow B$. Let $||A||$ denote the number of transactions satisfying $A$.

The general idea of an association rule $A \Rightarrow B$ is to indicate that records possessing attribute $A$ also tend to possess attribute $B$. The goal is to find association rules which are considered sufficiently interesting as defined by one or more measures. The most common measures are support defined by

$$supp(A \Rightarrow B) = \frac{||A \land B||}{|L|} \quad (2)$$

and confidence defined by

$$conf(A \Rightarrow B) = \frac{||A \land B||}{||A||} \quad (3)$$

The support is the proportion of transactions satisfying $A$ and $B$ while confidence is the proportion of transactions satisfying $A$ which also satisfy $B$. Despite the fact that the set of records in question will not in general be a random sample and so relative frequencies cannot be equated with probabilities, it is common practice to take the support as an estimate of the probability of $A$ and $B$ occurring together, $Pr(A, B)$, and confidence as an estimate of the conditional probability of $B$ given $A$, $Pr(B|A)$.

In addition to these measures, numerous other measures of ‘interestingness’ have been discussed in the literature [1, 2]. Among these are measures corresponding to the confirmation measures defined in section 2. Of particular relevance to the current work is that of Greco et al. [18], who focus on confirmation measures and point out their merits, and Glass [19] who adopts a similar approach in the case of fuzzy association rules.

Just as the support of attributes $A$ and $B$ corresponds to their joint probability and confidence of $A \Rightarrow B$ to the probability of $B$ given $A$, so the relative frequency of $A$, i.e. $||A||/|L|$, corresponds to the probability of $A$. To simplify the notation and to make comparison with confirmation measures more explicit we represent the relative frequency by a function $Fr$ so that
\[ Fr(A) = \frac{|A|}{|D|}, \]  
\[ Fr(A, B) = \frac{|A \land B|}{|D|} \left( = \text{supp}(A, B) \right), \]  
and
\[ Fr(B|A) = \frac{Fr(A, B)}{Fr(A)} = \frac{|A \land B|}{|A|} \left( = \text{conf}(A \Rightarrow B) \right). \]

This enables us to define confirmation measures for association rules and so, for example, we can define a measure corresponding to the measure \( d \),
\[ d(A \Rightarrow B) = Fr(B|A) - Fr(B). \]

More generally we can define a confirmation measure for association rules as follows:

**Definition 2.** A confirmation measure of the strength of an association rule \( A \Rightarrow B \) in a database \( D \), denoted \( c(A \Rightarrow B) \), is a measure which satisfies

(i) \( c(A \Rightarrow B) > 0 \) if \( Fr(B|A) > Fr(B) \)

(ii) \( c(A \Rightarrow B) = 0 \) if \( Fr(B|A) = Fr(B) \)

(iii) \( c(A \Rightarrow B) < 0 \) if \( Fr(B|A) < Fr(B) \)

where \( Fr(B|A) \) is the confidence of \( A \Rightarrow B \) and \( Fr(B) \) is the relative frequency of \( B \).

As [18] point out this can be interpreted as follows:

- \( c(A \Rightarrow B) > 0 \) means that \( B \) is satisfied more frequently when \( A \) is satisfied than it is generically.

- \( c(A \Rightarrow B) = 0 \) means that \( B \) is satisfied with the same frequency when \( A \) is satisfied as it is generically.

- \( c(A \Rightarrow B) < 0 \) means that \( B \) is satisfied less frequently when \( A \) is satisfied than it is generically.
Since relative frequency is a probability measure, all of the confirmation measures defined in section 2 have a counterpart confirmation measure for the strength of association rules which is obtained by replacing $Pr$ with $Fr$. For simplicity of notation, however, we will use $Pr$ and assume that probabilities should be defined in terms of relative frequencies.

The key advantage of using confirmation measures to quantify the strength of an association rule is due to their sensitivity to probabilistic dependence. In particular, the important factor in determining the strength of a rule $A \Rightarrow B$ is not the extent to which $B$ is satisfied when $A$ is satisfied, but the extent to which $B$ is more likely to be satisfied when $A$ is satisfied than when $A$ is not satisfied. Indeed, from the interpretation of $c(A \Rightarrow B) > 0$ given above and the fact that relative frequencies are probabilities, it follows that:

$$c(A \Rightarrow B) > 0 \text{ means that } B \text{ is satisfied more frequently when } A \text{ is satisfied than when } A \text{ is not satisfied.}$$

The following example illustrates the point.

**Example 1.** Suppose we have the following database of transactions involving milk ($M$), bread ($B$) and eggs ($E$) and wish to compare the association rules $M \Rightarrow B$ and $M \Rightarrow E$.

<table>
<thead>
<tr>
<th>Milk</th>
<th>Bread</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

According to the confidence measure the rule $M \Rightarrow B$ is stronger than $M \Rightarrow E$ since $conf(M \Rightarrow B) = 3/4 > 1/2 = conf(M \Rightarrow E)$. By contrast, according to the difference measure, $d$, given in equation (7) $M \Rightarrow E$ is stronger since $conf(M \Rightarrow B) = 3/4 - 7/8 = -1/8 < 1/8 = 1/2 - 3/8 = conf(M \Rightarrow E)$. 

9
Even though most people who bought milk did in fact buy bread (3 out of 4), the confidence measure fails to take into account the fact that the proportion of people buying bread who did not buy milk is even higher (4 out of 4). Similarly, even though the proportion of people buying eggs who bought milk is not very high (2 out of 4), the confidence measure fails to take into account the fact that the proportion of people buying eggs who did not buy milk is lower (1 out of 4). In fact, the relationship between milk and bread is one of negative dependence in the database so that someone who buys milk is less likely to buy bread than someone who does not buy milk. By contrast, milk and eggs are positively related. Thus, this difference between a confirmation measure and confidence does not arise because of a judicious choice of $d$ as the confirmation measure. The negative/positive dependencies means that $M \Rightarrow E$ is stronger than $M \Rightarrow B$ according to all confirmation measures.

Furthermore, for any measure which is not a confirmation measure it will be possible to find a similar example where an association rule involving negatively dependent (or independent) items is found to be stronger than another rule involving items which are positively dependent. We consider this to provide a strong motivation for confirmation measures.

Granting that confirmation measures have certain merits, we are still left with the question as to which confirmation measure to choose. This is not a trivial question since although all confirmation measures will agree on whether a particular association rule has a positive, negative or zero value, in general they will not agree on how to order different association rules. Thus, supposing that items $A$ and $B$ are positively dependent and so are $A$ and $C$, one measure might assign a greater value to the rule $A \Rightarrow B$ than to $A \Rightarrow C$ while another measure might reverse this ordering. This issue will be considered below.

In order to assist the analysis in later sections, we follow Greco et al. [18] in defining, for a rule $A \Rightarrow B$, $a = ||A \land B||$, i.e. the number of transactions satisfying both $A$ and $B$, and similarly $b = ||\neg A \land B||$, $c = ||A \land \neg B||$ and $d = ||\neg A \land \neg B||$. Table 1 presents this information in the form of a contingency table, which in some case we will represent as $[a \; c; \; b \; d]$.

The confirmation measures defined earlier are expressed both in terms of probabilities and in terms of $a, b, c$ and $d$ in table 2. In addition, a number of other confirmation measures which are used as measures of interestingness (see for example [2]). For most of these measures it is straightforward to establish that they are in fact confirmation measures.
Table 1: Contingency table for the rule $A \Rightarrow B$.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$\neg B$</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$a$</td>
<td>$c$</td>
<td>$a + c$</td>
</tr>
<tr>
<td>$\neg A$</td>
<td>$b$</td>
<td>$d$</td>
<td>$b + d$</td>
</tr>
<tr>
<td>Sum</td>
<td>$a + b$</td>
<td>$c + d$</td>
<td>$a + b + c + d$</td>
</tr>
</tbody>
</table>

4 Against null-invariance

A number of papers have emphasized null-invariant measures of interestingness [13, 20, 27, 28]. In the case of a rule of the form $A \Rightarrow B$, a null transaction is a transaction which contains neither $A$ nor $B$. A measure of interestingness is said to be null-invariant if it does not depend on the number of null transactions. Suppose we are considering the relationship between coffee and milk (i.e. the rule coffee $\Rightarrow$ milk). It is tempting to think that a measure of interestingness should be null-invariant since it might not be immediately obvious what transactions involving neither milk nor coffee have to do with such a rule. There is a serious problem with such an approach, however, as the following discussion (based on an example by Wu et al. [20]) is intended to show.

As in the previous section, let $a$ be the number of transactions involving both coffee and milk, $b$ milk but no coffee, $c$ coffee but no milk and $d$ neither coffee nor milk. Table 3 presents values for $a$, $b$, $c$ and $d$ for four example data sets. It also presents the value of rule strength according to the measure $k$ from section 3, which is taken as an example of a confirmation measure, and the cosine measure, which for a rule $A \Rightarrow B$ is defined as

$$cosine(A \Rightarrow B) = \frac{Pr(A, B)}{\sqrt{Pr(A)Pr(B)}} = \frac{a}{\sqrt{(a+c)(a+b)}}.$$ 

The cosine measure is one of five null-invariant measures considered by Wu et al. [20] and is used here simply as an example of such a measure; any of the other null-invariant measures could be used to make the same point. Also, no claim is being made that the $k$ measure is the best confirmation measure to use, it is just used to illustrate the nature of association, if any, between milk and coffee.

Note that the cosine measure gives the same result for data sets $D_1$, $D_2$.
and $D_3$, which is what we expect since it is a null-invariant measure and the values for $a$, $b$ and $c$ are unchanged for these data sets. The values for the $k$ measure highlight a problem for cosine. In data set $D_1$, the positive value of the $k$ measure indicates a positive association between milk and coffee, while in data set $D_2$ the value of zero for the $k$ measure indicates that there is no association (i.e. probabilistic independence) and in $D_3$ the negative value of the $k$ measure indicates a negative association. The failure of the cosine measure to distinguish these cases represents a serious problem for cosine as a measure of association.

A further problem is illustrated by data set $D_4$. Wu et al. [20] judge this to be a case of negative association because the ratio of $a:(a+c)$ is 1:11, i.e. the number of transactions including both milk and coffee is only a small proportion of the total number of transactions including coffee. However, defining probabilities in terms of frequencies, $Pr(\text{coffee}) = (a + c)/(a + b + c + d)$, which for $D_4$ gives $Pr(\text{coffee}) = 1100/102$, $100 \simeq 0.0108$ and $Pr(\text{milk})$ also has the same value. If transactions of coffee and milk are independent, then we would expect $Pr(\text{coffee}, \text{milk}) = Pr(\text{coffee}) \times Pr(\text{milk}) \simeq 0.000116$. This means that the expected number of transactions including both coffee and milk would be about 12 and hence the value of 100 in $D_4$ indicates a very positive association. This is reflected in a positive value of the $k$ measure, but the cosine measure gives a very low value of just 0.09. This is particularly worrying when compared with $D_3$ because in $D_3$ cosine gives a very high value in the case of a negative association and in $D_4$ it gives a very low value in the case of a positive association.

A final point concerns the idea of a neutral association, which is neither negative nor positive. According to Wu et al., $D_5$ represents a such a neutral association. The reason for this seems to be that the $Pr(\text{milk}|\text{coffee}) = 0.5$, yet as also noted by Wu et al. a central motivation for measures of interestingness has been that such a conditional probability is limited because if milk is a very popular item on its own then the value of 0.5 could represent a negative association. Compare data sets $D_5$ and $D_6$. Although $a = b = c$ in both cases and so $Pr(\text{milk}|\text{coffee}) = 0.5$ in both cases, in $D_5$, $Pr(\text{milk}) = (a + b)/(a + b + c + d) \simeq 0.019$ indicating a very strong positive association between coffee and milk, whereas in $D_6$ the majority of transactions include milk, with $Pr(\text{milk}) \simeq 0.645$, indicating a negative association. This is clearly reflected in the values for the $k$ measure, but not in those for cosine. Since all null-invariant measures will agree in cases $D_5$ and $D_6$ it seems clear that they fail to distinguish between cases of negative and positive association.
in an appropriate way.

The following result clearly brings out the importance of null transactions in the context of confirmation measures. Recall that $a = |A \land B|$, $b = |\neg A \land B|$, $c = |A \land \neg B|$ and $d = |\neg A \land \neg B|$. Hence $d$ is the number of null transactions.

**Proposition 1.** For a confirmation measure of the strength of an association rule $A \Rightarrow B$, denoted $c(A \Rightarrow B)$

(i) $c(A \Rightarrow B) > 0$ iff $ad - bc > 0$,

(ii) $c(A \Rightarrow B) = 0$ iff $ad - bc = 0$,

(iii) $c(A \Rightarrow B) < 0$ iff $ad - bc < 0$.

**Proof.** Consider (i). From definition 2 we know that $c(A \Rightarrow B) > 0$ iff

$$\frac{Pr(B|A)}{Pr(B)} > \frac{a}{a+c} > \frac{a+b}{a+b+c+d}$$

$$\Leftrightarrow \frac{ad - bc}{a+b+c+d} > 0.$$ 

(ii) and (iii) can be proved in the same way. \hfill \square

This point is closely related to the property of monotonicity which will be discussed in section 5 since an increase in the number of null-transactions can effect a transition from non-confirmation to confirmation [18]. It is worth noting that the relevance of null-transactions to measures of interestingness is related to Hempel’s raven paradox [29], according to which observation of a non-black, non-raven provides support for the hypothesis that all ravens are black since logically this is equivalent to ‘all non-black things are non-ravens’. This paradox has been discussed at considerable length in the literature and is generally accepted by Bayesians who seek to resolve the paradox by showing that the hypothesis is much more strongly supported by black ravens than by non-black, non-ravens [30]. For a discussion of this topic in the context of monotonicity see [18].

Null-invariant measures no doubt have their uses. For example, in some contexts high posterior probability, rather than positive probabilistic dependence, may be what is of interest to the user and so confidence (perhaps in conjunction with support) will be the appropriate measure(s). However, they have serious problems as measures of association. In particular, they cannot
distinguish a neutral point between negative and positive association in an appropriate way. By contrast, confirmation measures have a very clearly defined neutral point in terms of probabilistic independence, but as the above result shows, null-transactions are essential to confirmation measures.

There is nevertheless an important issue concerning null-transactions. Wu et al. [20] point out that the number of null-transactions is usually huge and unstable. Thus, it will be important to consider the impact of changes in the number of null-transactions on proposed measures. We will return to this issue in section 6.2.

5 Selecting a confirmation measure

In this section we consider a range of possible requirements to be satisfied by confirmation measures in the context of association rule strength. These include the requirements noted in section 2 as well as the monotonicity requirement proposed by Greco et al. [18].

5.1 Logicality and related requirements

In the context of association rules, the logicality requirement amounts to saying that, provided it is defined, $c(A \Rightarrow B)$ should be maximal when there are no counterexamples to the rule, i.e. when $Pr(A, \neg B) = 0$, and minimal when there are no counterexamples to the rule $A \Rightarrow \neg B$, i.e. when $Pr(A, B) = 0$. But should logicality be a requirement in the context of association rule strength? According to [3], a measure take on its maximal value, when there are no counterexamples, but the following example suggests otherwise.\footnote{More accurately, they claim that it should be constant and maximal or possibly infinite. For measures such as $l$ we make the assumption that $\log(c/0) = \infty$ for positive $c$ and $\log(0) = -\infty$.}

**Example 2.** Suppose we have the following database of transactions involving milk (M), bread (B) and eggs (E) and wish to compare the association rules $M \Rightarrow E$ and $B \Rightarrow E$. 
Since there are no counterexamples to either rule logicality requires that \( c(M \Rightarrow E) \) and \( c(B \Rightarrow E) \) are both maximal.

One can certainly understand the motivation for logicality: if confirmation is a quantitative generalization of logical entailment, then in any case where evidence entails an hypothesis confirmation should be maximal. Yet in terms of the strength of association rules there seems to be a clear difference between the two rules in example 2 since \( M \Rightarrow E \) accounts for all cases of \( E \), whereas \( B \Rightarrow E \) does not. Thus, it seems reasonable that in this example \( c(M \Rightarrow E) \) should be maximal, but not \( c(B \Rightarrow E) \). This motivates the following alternative to logicality in the context of rule interestingness:

**Maximality/Minimality Requirement:** \( c(A \Rightarrow B) \) should be maximal

iff \( Pr(A, \neg B) = Pr(\neg A, B) = 0 \) and minimal iff \( Pr(A, B) = Pr(\neg A, \neg B) = 0 \), provided \( c(A \Rightarrow B) \) is defined.\(^3\)

Strictly speaking, the above example only motivates the weaker claim that \( c(A \Rightarrow B) \) should be maximal iff \( Pr(A, \neg B) = Pr(\neg A, B) = 0 \). However, this condition also means that \( c(A \Rightarrow \neg B) \) should be maximal iff \( Pr(A, B) = Pr(\neg A, \neg B) = 0 \), and given hypothesis symmetry it follows that \( c(A \Rightarrow B) \) should be minimal iff \( Pr(A, B) = Pr(\neg A, \neg B) = 0 \).

The maximality/minimality requirement is closely related to a proposal by Greco et al. [21] called (weak L). Essentially, (weak L) provides the same sufficient conditions for maximality and minimality, but does not require that they are also necessary. Nevertheless, their reasons for proposing (weak L) are similar to those given here, which seem to justify the stronger requirement that a measure should not be maximal unless both \( Pr(A, \neg B) = 0 \) and \( Pr(\neg A, B) = 0 \).

In the rest of the paper, the maximality/minimality requirement will be adopted instead of logicality or (weak L). Furthermore, (Ex1) as discussed in

---

\(^3\)In terms of probabilistic confirmation measures, this would amount to replacing entailment with mutual entailment and making this a necessary as well as sufficient condition so that \( c(H, E) \) should be maximal (minimal) iff \( E \vdash H \) and \( H \vdash E \) (\( E \vdash \neg H \) and \( \neg H \vdash E \)).
section 2 will be rejected as a requirement for confirmation measures in the context of rule interestingness since in this case there is no convincing reason to think that confirmation should always be greater when A entails B, i.e. \( Pr(B|A) = 1 \) than in cases where \( Pr(B|A) < 1 \). The discussion here and in [21] suggests that \( Pr(B|\neg A) \) should also be taken into account. Drawing on one of their examples, suppose that for the rule \( A \Rightarrow B \), \( Pr(B|A) = 1 \) and \( Pr(B|\neg A) = 0.99 \), while for the rule \( A' \Rightarrow B' \), \( Pr(B'|A') = 0.99 \) and \( Pr(B'|\neg A') = 0 \). It is far from clear that a higher value should be assigned to \( A \Rightarrow B \) than to \( A' \Rightarrow B' \). (weak Ex₁) does seem acceptable, but given that the measures proposed in this paper which satisfy maximality/minimality also satisfy (weak Ex₁) it will not be used as a further requirement.

In the notation at the end of section 2, the requirement that \( Pr(A, \neg B) = Pr(\neg A, B) = 0 \) can be expressed as \( b = c = 0 \), while \( Pr(A, B) = Pr(\neg A, \neg B) = 0 \) can be expressed as \( a = d = 0 \). From this it can be shown that \( s \), \( \phi \) and \( cs \) are the only confirmation measures that satisfy this requirement. In particular, the \( l \) and \( k \) measures favoured by Greco et al. [18] fail to satisfy it. These results, however, present a problem since \( s \), \( \phi \) and \( cs \) do not satisfy the symmetry requirement. We will return to this problem later.

5.2 Symmetry requirement

This was discussed in the context of probabilistic confirmation measures in section 2 and in the context of an association rule \( A \Rightarrow B \) can be expressed as

\[
c(A \Rightarrow B) \text{ should satisfy HS, but not ES, CS nor TS,}
\]

where HS is the claim that \( c(A \Rightarrow B) = -c(\neg A \Rightarrow B) \), ES the claim that \( c(A \Rightarrow B) = -c(\neg A \Rightarrow B) \), CS the claim that \( c(A \Rightarrow B) = c(B \Rightarrow A) \) and TS the claim that \( c(A \Rightarrow B) = c(\neg A \Rightarrow \neg B) \). Consider the symmetry considerations in light of the contingency table for the rule \( \text{Milk} \Rightarrow \text{Bread} \) in table 4.

First, consider commutative symmetry. In this example, everyone who buys milk also buys bread whereas overall only 50% buy bread. By contrast, only 10% of people who buy bread also buy milk. Hence, intuitively milk provides much stronger confirmation of bread than vice versa and so we would expect the rule \( \text{Milk} \Rightarrow \text{Bread} \) to have a much greater value than the rule \( \text{Bread} \Rightarrow \text{Milk} \) in this case. As such, this scenario seems to provide a counterexample to commutative symmetry.
Now consider evidence symmetry. Does not purchasing milk count against bread as much as purchasing milk counts for it? We have already seen that everyone who buys milk also buys bread, but of those who do not buy milk almost as many (45) buy bread as do not (50). Hence, not purchasing milk only counts very weakly against bread. Thus, this scenario also seems to provide a counterexample to evidence symmetry. Similarly, it seems to provide a counterexample to total symmetry since not purchasing milk only counts very weakly for the non-purchase of bread.

Finally, consider hypothesis symmetry. Does the purchase of milk count against the non-purchase of bread as much as it counts for the purchase of bread? Just as purchasing milk provides strong confirmation of the claim that bread was also purchased, it provides strong disconfirmation of the claim that bread was not purchased. Of course, while a single scenario can function as a counterexample to the other symmetries it cannot establish hypothesis symmetry. But even stating hypothesis symmetry makes it seem almost obvious since it is difficult conceptually to distinguish between confirmation of bread and disconfirmation of not-bread.

Hence, we agree with Greco et al. [18] that the arguments presented by Eells and Fitelson [22] in favour of the symmetry requirement carry over to the current context. We note that property 1 discussed by Tan et al. [28] corresponds to commutative symmetry and so should be rejected. Property 3 of Tan et al. [28] is the requirement that measures should be antisymmetric under permutations of the rows or columns of the contingency table. Permutations of the rows results in swapping \(A\) and \(\neg A\) in the case of a rule \(A \Rightarrow B\) and so the claim that it should be antisymmetric is that \(c(A \Rightarrow B) = -c(\neg A \Rightarrow B)\), i.e. evidence symmetry. Hence, this part of property 3 should be rejected. Antisymmetry under column permutation, however, corresponds to hypothesis symmetry and so this part of property 3 should be accepted.

Property 2 of Tan et al. [28] is the requirement that measures should be invariant under row and column scaling of the contingency table. However, if we represent a contingency table for the rule \(A \Rightarrow B\) by the matrix \([a \ c; \ b \ d]\), we can scale the first row by \(b/c\) to get \([ab/c \ b; \ b \ d]\) and then the first column by \(c/b\) to get \([a \ b; \ c \ d]\), which corresponds to the contingency table for the rule \(B \Rightarrow A\). Hence, row and column scaling invariance would correspond to commutative symmetry and so should be rejected. This does not preclude the possibility that one of row or column scaling invariance should be accepted. We will return to this point later.
An extension to the symmetry requirement of Eells and Fitelson [22] has been proposed by Crupi et al. [23] and adopted by Greco et al. [21] in the context of rule interestingness. This extension is based on the approach of Crupi et al. to view confirmation as an extension of classical logic. Briefly, they argue that confirmation measures should mirror the symmetry properties of the function $v$ considered earlier in the discussion of (Ex$_1$). Consider again commutative symmetry. Crupi et al. agree with Eells and Fitelson that it does not hold in general, but they argue that it should hold in the case of disconfirmation. Their argument for this is based on the fact that $E$ refutes $H$, i.e. $E$ entails $\neg H$, if and only if $H$ refutes $E$. By extension this is applied to the general case. Given their conception of confirmation as a generalization of logical entailment, it seems very plausible that symmetry should apply since, given the logical equivalence between $E$ refuting $H$ and $H$ refuting $E$, the symmetry applies in the limiting case.

However, it is not so clear that it should apply in the context of rule interestingness. Given the maximality / minimality requirement, the fact that $E$ refutes $H$ in a given case is not sufficient for minimality so the limiting case no longer applies. If $E$ refutes $H$, then $Pr(H|E) = 0$ and $Pr(E|H) = 0$, but does it follow that the disconfirmation of $E$ by $H$ should be the same as that of $H$ by $E$? As noted earlier, in the case of rule interestingness it is reasonable to think that the confirmation (or disconfirmation) of $H$ by $E$ should depend not only on $Pr(H|E)$ but also on $Pr(H|\neg E)$. But in that case, the disconfirmation of $H$ by $E$ and $E$ by $H$ when $E$ refutes $H$ would require not only that $Pr(H|E) = Pr(E|H) = 0$ but also that $Pr(H|\neg E) = Pr(E|\neg H)$ and in general this is not the case. For these reasons the extended symmetry principles of Crupi et al. will not be adopted here, but only the original symmetry requirement of [22].

Eells and Fitelson show that of the various confirmation measures presented in 2, $d$, $l$, and $k$ satisfy these symmetry requirements while $r$, $s$ and $b$ do not. Results for the other measures are presented in table 5.

5.3 Posterior requirement

As in the case of logicality, the posterior requirement seems very plausible if the goal is a probabilistic notion of confirmation which generalizes logical entailment. However, it is not so clear that it is also appropriate when the goal is a confirmation measure to quantify the strength of an association rule. Example 2 illustrates the problem. In section 2 we noted that it follows from
the posterior requirement that if the posterior probability of an hypothesis given two different pieces of evidence is the same, then the confirmation of the hypothesis by the two pieces of evidence is also the same. In fact, the logicality and posterior requirements are closely linked. Logicality is a special case of this consequence of the posterior requirement in the case where the posterior probability is one with the additional stipulation that confirmation should be maximal in this case. In example 2, \( Pr(E|M) = Pr(E|B) = 1 \), yet as we noted above the rule \( c(M \Rightarrow E) \) is stronger than the rule \( c(B \Rightarrow E) \). So by rejecting logicality, we have already rejected the posterior requirement.

Does the same kind of problem arise even if the relative frequencies are not equal to one? Consider the following extended version of the database from example 2.

<table>
<thead>
<tr>
<th>Milk</th>
<th>Bread</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In this case, \( Pr(E|M) = Pr(E|B) = 1/2 \), yet it could be argued that \( M \Rightarrow E \) is stronger than the rule \( B \Rightarrow E \) since the former accounts for both cases where eggs were purchased. In association rule mining it is common to use support in addition to confidence when considering rules and on these grounds \( c(M \Rightarrow E) \) would be preferred in this case because, although both rules have the same confidence, \( c(M \Rightarrow E) \) has greater support since \( Pr(M, E) = 1/3 > 1/6 = Pr(B, E) \). However, it must be remembered that in general just focussing on confidence and support ignores probabilistic dependence and so is problematic.

Given the way in which the conditions for maximality differ from those of logicality in the case, one possibility would be to attempt to modify the posterior requirement by taking support into account. If we focus on two rules with the same consequent and equal posteriors, e.g. rules \( A_1 \Rightarrow B \) and \( A_2 \Rightarrow B \) with \( Pr(B|A_1) = Pr(B|A_2) \), the difference on support depends on the difference in the relative frequencies of the antecedents since \( Pr(A_1, B) = Pr(B|A_1) \times Pr(A_1) \) and \( Pr(A_2, B) = Pr(B|A_2) \times Pr(A_2) \). The role of support could motivate the following requirement for a confirmation measure:
Equal Posterior Requirement: If $Pr(B|A_1) = Pr(B|A_2)$ and $Pr(B|A_1) > Pr(B)$, then $c(A_1 \Rightarrow B) > c(A_2 \Rightarrow B)$ iff $Pr(A_1) > Pr(A_2)$.

Since this requirement is in opposition to the posterior requirement set out in section 2 it means that since $d$, $r$, $l$ and $k$ all satisfied it, they all fail to satisfy the current requirement. Since $s$ and $b$ can be expressed as

\[ s(A \Rightarrow B) = \frac{Pr(B|A) - Pr(B)}{1 - Pr(A)} \quad (8) \]

and

\[ b(A \Rightarrow B) = [Pr(B|A) - Pr(B)] \times Pr(A) \quad (9) \]

respectively, it is clear that $s$ and $b$ both satisfy the equal posterior requirement.

However, just as the original posterior requirement was questionable, so is this equal posterior requirement. As can be seen from the example above, the greater support for rule $M \Rightarrow E$ is only achieved at the expense of a greater number of counterexamples to the rule. By contrast, in the maximality condition of the last sub-section, greater support occurred without the introduction of counterexamples. Hence, while the equal posterior requirement might be desirable in certain contexts, we do not propose it as a necessary requirement for an adequate measure.

5.4 Monotonicity requirement

Greco et al. [18] propose a further requirement which, in the terminology and notation used in this paper, can be stated as follows.

Monotonicity Requirement: A confirmation measure of the strength of an association rule $A \Rightarrow B$, denoted $c(A \Rightarrow B)$, is a function of $||A \land B||$, $||\neg A \land B||$, $||A \land \neg B||$ and $||\neg A \land \neg B||$, which is non-decreasing with respect to $||A \land B||$ and $||\neg A \land B||$ and non-increasing with respect to $||\neg A \land B||$.

\footnote{Note that the requirement only applies whenever $Pr(B|A_1) > Pr(B)$, i.e. $A_1$ and $B$ (and hence also $A_2$ and $B$) have a positive dependence. If they have a negative dependence, it seems that a higher value of $Pr(A_1)$ than $Pr(A_2)$ would decrease the strength of the rule since any benefit arising from a greater proportion of instances of the rule would be outweighed by a corresponding greater proportion of counterexamples to the rule.}
One line of reasoning to support monotonicity can be based on proposition 1. Recall from section 3 that $a = ||A \land B||$, $b = ||\neg A \land B||$, $c = ||A \land \neg B||$ and $d = ||\neg A \land \neg B||$. As shown by Greco et al. [18], a change from negative dependence to positive dependence when $b$, $c$ and $d$ are kept fixed can only occur if there is an increase in $a$ and similarly for $d$, while if $a$, $c$ and $d$ are kept fixed such a change can only occur if there is a decrease in $b$ and similarly for $c$. Clearly, this follows trivially from proposition 1 and provides some motivation for accepting monotonicity (see [18] for further discussion in support of monotonicity).

Note that monotonicity is concerned with the number of instances, counterexamples, etc. of the rule rather than probabilities or relative frequencies. Furthermore, although there is a result in terms of probabilities that is similar to proposition 1, there is no straightforward counterpart to monotonicity since, for example, it is impossible for the probability $Pr(A, B)$ to change while $Pr(\neg A, B)$, $Pr(A, \neg B)$ and $Pr(\neg A, \neg B)$ remain fixed.\(^5\)

It is worth comparing monotonicity with the following requirement for an interestingness measure $F$ of the rule $A \Rightarrow B$ proposed by Geng and Hamilton [2]:

**GH:** $F$ is an increasing measure of support if the margins in the contingency table are fixed.

and two further requirements which Geng and Hamilton attribute to Piatetsky-Shapiro [11]:

**PS1:** $F$ monotonically increases with $Pr(A, B)$ when $Pr(A)$ and $Pr(B)$ remain the same.

**PS2:** $F$ monotonically decreases with $Pr(A)$ (or $Pr(B)$) when $Pr(A, B)$ and $Pr(B)$ (or $Pr(A)$) remain the same.

In the discussion which follows we make the assumption that the probabilities in PS1 and PS2 are to be defined in terms of frequencies from the contingency table for $A \Rightarrow B$ as discussed in section 3. Interestingly, the original requirements proposed by Piatetsky-Shapiro [11] were not expressed in terms of probabilities and, in fact, the original version of PS1 is equivalent to GH. However, here we consider the relationship between GH and the probabilistic version of PS1 since it is more general in nature.

\(^5\)Suppose that we take $a = Pr(A, B)$ and similarly for $b$, $c$ and $d$, then proposition 1 is still correct.
Proposition 2. If a measure \( F \) satisfies PS1, it also satisfies GH.

Proof. This follows from the fact that support is defined as \( Pr(A, B) \) and if margins in the contingency table are fixed then \( Pr(A) \) and \( Pr(B) \) are fixed.

To establish a result in the opposite direction, we need the following definition:

Definition 3. A measure \( F \) is said to be invariant under overall scaling if both rows (or both columns) of the corresponding contingency table are scaled by a constant factor.

Clearly, any measure defined in terms of probability and where the probabilities in turn are defined in terms of the frequencies from the contingency table will satisfy this property. This enables us to establish the following result:

Lemma 1. If a measure \( F \) satisfies GH and is invariant under overall scaling, it also satisfies PS1.

Proof. Suppose that we wish to compare \( F \) for two rules \( A \Rightarrow B \) and \( A' \Rightarrow B' \), that \( F \) satisfies GH and is invariant under overall scaling and that \( Pr(A', B') \geq Pr(A, B) \), \( Pr(A') = Pr(A) \) and \( Pr(B') = Pr(B) \). We need to show that \( F(A' \Rightarrow B') \geq F(A \Rightarrow B) \).

Represent the contingency table for \( A \Rightarrow B \) by \([X, \alpha - X; \beta - X, N - a - b + X]\) and the contingency table for \( A' \Rightarrow B' \) by \([X', \alpha' - X'; \beta' - X', N' - \alpha' - \beta' + X']\).

Since \( Pr(A) = Pr(A') \) and \( Pr(B) = Pr(B') \) then \( \alpha/N = \alpha'/N' \) and \( \beta/N = \beta'/N' \).

We can set \( N' = kN \) for some constant \( k \). Then \( \alpha' = k\alpha \) and \( \beta' = k\beta \).

Since \( Pr(A', B') \geq Pr(A, B) \) it follows that \( X'/N \geq X/N \) and so \( X' \geq kX \).

Let \( X' = kX + \Delta \).

This means that the contingency table for \( A' \Rightarrow B' \) can be expressed as \([kX + \Delta, k\alpha - kX - \Delta; k\beta - kX - \Delta, kN - k\alpha - k\beta + kX + \Delta]\).

Define a further rule \( A'' \Rightarrow B'' \) with the contingency table \([kX, k\alpha - kX; k\beta - kX, kN - k\alpha - k\beta + kX]\). Since \( F \) is invariant under overall scaling, \( F(A'' \Rightarrow B'') = F(A \Rightarrow B) \).

It is clear that the margins of the contingency table for \( A' \Rightarrow B' \) are the same as those for \( A'' \Rightarrow B'' \) and that \( Pr(A', B') \geq Pr(A'', B'') \). Hence, since \( F \) satisfies GH, we have \( F(A' \Rightarrow B') \geq F(A'' \Rightarrow B'') = F(A \Rightarrow B) \).
Hence $F$ satisfies PS1.

**Theorem 1.** If a measure $F$ satisfies monotonicity, it also satisfies GH.

*Proof.* Consider a rule $A \Rightarrow B$ with contingency table $[X \; \alpha - X; \; \beta - X \; N - \alpha - \beta + X]$ and another rule $A' \Rightarrow B'$ with contingency table $[X + \Delta \; \alpha - X - \Delta; \; \beta - X - \Delta \; N - \alpha - \beta + X + \Delta]$ with $\Delta \geq 0$. Since the margins of the contingency table are the same in both cases and $Pr(A'B') \geq Pr(A, B)$ we need to show that $F(A' \Rightarrow B') \geq F(A \Rightarrow B)$ given that $F$ satisfies monotonicity.

Define a further rule $A'' \Rightarrow B''$ with the contingency table $[X \; \alpha - X - \Delta; \; \beta - X - \Delta \; N - \alpha - \beta + X]$. Since $F$ satisfies monotonicity it is a non-increasing function of $b = ||\neg A \land B||$ and $c = ||A \land \neg B||$ and clearly $\alpha - X \geq \alpha - x - \Delta$ and $\beta - x \geq \beta - X - \Delta$, then $F(A'' \Rightarrow B'') \geq F(A \Rightarrow B)$. Also $F$ is a non-decreasing function of $a = ||A \land B||$, $d = ||\neg A \land \neg B||$, hence $F(A', B') \geq F(A'', B'')$. Hence, $F(A' \Rightarrow B') \geq F(A \Rightarrow B)$ and so $F$ satisfies GH.

**Corollary 1.** If a measure $F$ satisfies monotonicity and it is invariant under overall scaling, it also satisfies PS1.

*Proof.* This follows immediately from lemma 1 and theorem 1.

Notice, however, that the converse of theorem 1 is false and so a measure can satisfy GH (and PS1) without satisfying monotonicity.

**Example 3.** Consider the ratio measure $r = \log[Pr(A, B)/(Pr(A)Pr(B))]$. It is clear from its definition that it satisfies PS1 and hence GH. However, consider the rule $A \Rightarrow B$ with contingency table $[1 \; 0; \; 0 \; 1]$ and another rule $A' \Rightarrow B'$ with contingency table $[2 \; 0; \; 0 \; 1]$. Clearly, $||A' \land B'|| > ||A \land B||$ and yet $r(A \Rightarrow B) = \log 2 > \log 3/2 = r(A' \Rightarrow B')$ and so the ratio measure does not satisfy monotonicity.

This means that monotonicity is logically stronger than GH (and PS1) and so it will be used as one of the requirements for a confirmation measure. For similar reasons PS2 is not included as a separate requirement. For further discussion of the relationship between monotonicity and other measures, particularly confidence, support and anti-support (which for a rule $A \Rightarrow B$ is the support of $A \Rightarrow \neg B$) and for a way to mine all the rules maximizing any confirmation measure that satisfies monotonicity see [31].
Monotonicity is related to our maximality / minimality requirement since the latter entails a special case of monotonicity. Suppose a confirmation measure \( c(A \Rightarrow B) \) satisfies the maximality / minimality requirement. In cases where \( a = d = 0 \), then \( c(A \Rightarrow B) \) must be a non-increasing function of \( b \) and \( c \) and in cases where \( b = c = 0 \), then \( c(A \Rightarrow B) \) must be a non-decreasing function of \( a \) and \( d \). Thus, monotonicity is consistent with the maximality / minimality requirement. Given this relationship and the above discussion, we think that monotonicity is a suitable requirement for a confirmation measure in the context of association rules.

Monotonicity is also related to property 2 proposed by Tan et al. [28], which is the property of row and column scaling invariance. This was discussed in section 5.2 where the combination of both row and column scaling invariance was rejected, but can one or other of them be accepted? Consider a rule \( A \Rightarrow B \) with the contingency table \([a \ c; \ 0 \ d]\) and a scaling of the first column by a constant \( k > 1 \) giving \([ka \ c; \ 0 \ d]\). Consider a specific case where \( c = a \) i.e. the number of instances of the rule and the number of counterexamples is the same. Column scaling invariance in this case would not necessarily be in conflict with monotonicity since the latter does not require that a measure be a strictly increasing function of \( a \) in all cases, but this does seem like a case where we would expect the strength of the rule to increase since scaling the first column increases the number of instances while leaving the number of counterexamples the same and so, since we are also assuming \( b = 0 \), it seems like an unreasonable requirement that the strength of the rule should remain unchanged. This suggests that column scaling invariance should be rejected.

A similar argument can also be deployed in the case of row scaling invariance. If the contingency table is instead \([a \ 0; \ b \ d]\) a scaling of the first row by \( k > 1 \) gives \([ka \ 0; \ b \ d]\). This means that the number of instances of the rule \( A \Rightarrow B \) increases while \( b \), the number of transactions involving \( B \) but not \( A \), remains the same. As in the previous case, row scaling invariance in this case would not necessarily be in conflict with monotonicity since the latter does not require that a measure be a strictly increasing function of \( a \) in all cases, but this does seem like a case where it is not unreasonable to think that the strength of the rule should increase. Notice that in this case \( c = 0 \), which means there are no counterexamples to the rule. Had the posterior requirement been accepted, the measure should yield its maximum value for any value of \( a > 0 \), but since maximality has been proposed instead, there is no good reason to think that the measure should yield the same value for all.
positive values of $a$. This suggests that row scaling invariance should also be rejected.

As shown by Greco et al. [18], $l$, $k$ and $s$ all satisfy monotonicity, while $d$, $r$ and $b$ fail to satisfy it. Results for all the measures are shown in table 5.

**5.5 Summary of Results**

On the basis of the foregoing discussion we propose the following desiderata for a measure to quantify the strength of association rules. It should satisfy the following requirements:

(i) confirmation measure (see definition 2),

(ii) symmetry requirement (see section 5.2),

(iii) maximality / minimality requirement (see section 5.1),

(iv) monotonicity (see section 5.4).

Table 5 summarizes the results for all the measures defined in section 3. As noted earlier, all the measures considered here are confirmation measures, but apart from that there is a lot of disagreement. In terms of the probabilistic conception of confirmation found in the philosophical literature, the symmetry requirement is important and this is also important in the context of association rules. Of all measures only $d$, $l$, $k$ and $cf$ satisfy symmetry. Apart from symmetry, it is not at all obvious that the same desiderata are appropriate in the association rule context as in the philosophical context. In terms of association rules, monotonicity has been proposed by Greco et al. [18] in addition to symmetry and so this counts in favour of $l$, $k$ and $cf$. It is also interesting to note that these same three measures satisfy the logicality and posterior requirements (see section 2) which are important in the philosophical literature. Thus, although different considerations have come into play in previous work there is agreement in terms of the best measures.

The current work undermines this consensus due to the maximality / minimality requirement. It has been argued that this requirement is appropriate in the context of association rules, but only $s$, $\phi$ and $cs$ satisfy this requirement. However, none of these measures satisfies the symmetry requirement and $cs$ also does not satisfy monotonicity. So unfortunately none of the measures defined in section 3 satisfies all of the desiderata.
6 Two new confirmation measures

Does the fact that none of these measures satisfies all the desiderata mean that a compromise is necessary? Or is it possible to find an alternative measure that satisfies all the desiderata? Fortunately, it turns out that this is indeed possible. Here we consider two new measures obtained by combining existing measures. First, by combining the measures $k$ and $s$ we propose a new confirmation measure:

$$ks(A \Rightarrow B) = \frac{1}{2} \left( k(A \Rightarrow B) + s(A \Rightarrow B) \right). \quad (10)$$

Second, by combining the measures $k$ and $\phi$ we propose another new measure:

$$k\phi(A \Rightarrow B) = \frac{1}{2} \left( k(A \Rightarrow B) + \phi(A \Rightarrow B) \right). \quad (11)$$

Both of these measures satisfy all of the desiderata as shown in the following theorem.

**Theorem 2.** The measures $ks(A \Rightarrow B)$ and $k\phi(A \Rightarrow B)$ are (i) confirmation measures that satisfy (ii) the symmetry requirement, (iii) the maximality / minimality requirement and (iv) monotonicity.

**Proof.** (i) $ks$ and $k\phi$ are confirmation measures. This follows straightforwardly from the fact that $ks$ and $k\phi$ are linear combinations of two confirmation measures.

(ii) $ks$ and $k\phi$ satisfy the symmetry requirement.

(a) $ks$ and $k\phi$ satisfy hypothesis symmetry. First, $ks$.

$$ks(A \Rightarrow B) = \frac{1}{2} \left[ k(A \Rightarrow B) + s(A \Rightarrow B) \right],$$

$$= \frac{1}{2} \left[ -k(A \Rightarrow \neg B) - s(A \Rightarrow \neg B) \right],$$

(since $k$ and $s$ satisfy hypothesis symmetry [22])

$$= -ks(A \Rightarrow \neg B).$$

The result for $k\phi$ can be established in the same way, but we need
to show that $\phi$ satisfies hypothesis symmetry.

\[
\phi(A \Rightarrow B) = \frac{Pr(A, B) - Pr(A)Pr(B)}{\sqrt{Pr(A)Pr(B)Pr(\neg A)Pr(\neg B)}},
\]

\[
= \frac{Pr(A) - Pr(A, \neg B) - Pr(A)Pr(B)}{\sqrt{Pr(A)Pr(\neg B)Pr(\neg A)Pr(B)}},
\]

\[
= \frac{Pr(\neg B) - Pr(A, \neg B)}{\sqrt{Pr(A)Pr(\neg B)Pr(\neg A)Pr(B)}},
\]

\[
= -\phi(A \Rightarrow \neg B).
\]

(b) $ks$ and $k\phi$ do not satisfy evidence symmetry.

Consider the cards example in section 2 where a card is randomly drawn from a standard deck and $A$ is the evidence that the card is the seven of spades and $B$ the hypothesis that the card is black. In this case we have $ks(A \Rightarrow B) = \frac{1}{2}[1 + \frac{26}{51}] = \frac{77}{102} \neq \frac{27}{102} = -\frac{1}{2}[\frac{1}{51} - \frac{26}{51}] = -ks(\neg A \Rightarrow B)$. Similarly, $k\phi(A \Rightarrow B) = \frac{1}{2}[1 + \frac{1}{\sqrt{51}}] = \frac{51 + \sqrt{51}}{102} \neq \frac{1 + \sqrt{51}}{102} = -\frac{1}{2}[\frac{1}{51} - \frac{1}{\sqrt{51}}] = -k\phi(\neg A \Rightarrow B)$.

(c) $ks$ and $k\phi$ do not satisfy commutativity symmetry.

In the same cards example, $ks(A \Rightarrow B) = \frac{1}{2}[1 + \frac{26}{51}] = \frac{77}{102} \neq \frac{77}{102} = \frac{1}{2}[\frac{13}{38} + \frac{1}{26}] = ks(B \Rightarrow A)$. Similarly, $k\phi(A \Rightarrow B) = \frac{1}{2}[1 + \frac{1}{\sqrt{51}}] = \frac{51 + \sqrt{51}}{102} \neq \frac{19}{76} + \frac{\sqrt{51}}{102} = \frac{1}{2}[\frac{13}{38} + \frac{1}{\sqrt{51}}] = k\phi(B \Rightarrow A)$.

(d) $ks$ and $k\phi$ do not satisfy total symmetry.

In the same cards example, $ks(A \Rightarrow B) = \frac{1}{2}[1 + \frac{26}{51}] = \frac{77}{102} \neq \frac{77}{102} = \frac{1}{2}[\frac{1}{51} + \frac{26}{51}] = ks(\neg A \Rightarrow \neg B)$. Similarly, $k\phi(A \Rightarrow B) = \frac{1}{2}[1 + \frac{1}{\sqrt{51}}] = \frac{51 + \sqrt{51}}{102} \neq \frac{1 + \sqrt{51}}{102} = \frac{1}{2}[\frac{1}{51} + \frac{1}{\sqrt{51}}] = k\phi(\neg A \Rightarrow \neg B)$.

(iii) $ks$ and $k\phi$ satisfy the maximality / minimality requirement.

Necessity. Suppose $ks(A \Rightarrow B) = 1$ (i.e. is maximal). By definition of $ks$, $s(A \Rightarrow B) = 1$ and since $s$ satisfies the maximality / minimality requirement it follows that $Fr(A, \neg B) = Fr(\neg A, B) = 0$. Similarly, if $ks(A \Rightarrow B) = -1$ (i.e. is minimal), then $s(A \Rightarrow B) = -1$ and it follows that $Fr(A, B) = Fr(\neg A, \neg B) = 0$.

Sufficiency. Suppose $Fr(A, \neg B) = Fr(\neg A, B) = 0$. Since $s$ satisfies the maximality / minimality requirement it follows that $s(A \Rightarrow B) = 1$.
and since \( k(A \Rightarrow B) = 1 \) if \( Fr(A, \neg B) = 0 \) then \( ks(A \Rightarrow B) = 1 \).

Similarly, if \( Fr(A, B) = Fr(\neg A, \neg B) = 0 \), then \( s(A \Rightarrow B) = -1 \) and since \( k(A \Rightarrow B) = -1 \) if \( Fr(A, B) = 0 \) then \( ks(A \Rightarrow B) = -1 \).

The result for \( k\phi \) can be established in the same way.

(iv) Monotonicity

This follows straightforwardly from the fact that \( ks \) and \( k\phi \) are linear combinations of two confirmation measures which satisfy monotonicity.

This completes the proof. \( \square \)

6.1 Monotonicity revisited

Both of the new measures satisfy all the requirements, but in this section we consider the issue of monotonicity again to illustrate some advantages of the new measures and also to discriminate between them. Consider the transactions in the table below.

<table>
<thead>
<tr>
<th>Milk</th>
<th>Bread</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

We consider the strength of the rules \( M \Rightarrow E \) and \( B \Rightarrow E \) according to each of the confirmation measures that satisfy monotonicity, first considering just the first five transactions in the table and then all six. The results are presented in table 6.

Since the sixth transaction is an instance of both of the rules, monotonicity tells us that the rules should be no weaker when all six transactions are included compared to the first five. Consider first the rule \( M \Rightarrow E \). In this case, \( b \), the number of transactions involving \( E \) but not \( M \), is zero. In moving from five to six transactions \( a \), the number of instances of the rule increases by one while the number of counterexamples remains fixed at one. Although monotonicity does not require that the strength of the rule should be strictly greater with the addition of the extra transaction, it seems very reasonable that it should do so, yet this only occurs for \( s, cf, \phi, ks \) and \( k\phi \).
The reason for the failure of the other measures to get what is intuitively the correct ordering here differs. From table 2 it is easy to see that $rr$, $or$, $yq$ and $yy$ take on their maximum values when $b = 0$. This, however, is an odd result. We already know that these measures do not satisfy the maximality / minimality requirement, but if a measure is to violate this requirement, the most plausible condition for maximality would be when $c = 0$, i.e. there are no counterexamples, in keeping with logicality. $l$ and $k$ are not maximal when $b = 0$, but from their definitions in table 2 it is easy to see that when $b = 0$ they are independent of $a$.

Consider now rule $B \Rightarrow E$. In this case, $c$, the number of counterexamples to the rule, is zero. In moving from five to six transactions $a$, the number of instances of the rule increases by one while the number of transactions involving $E$ but not $B$ remains fixed at one. Should the strength of the rule be strictly greater with the addition of the extra transaction even though this is not required by monotonicity? This is not as clear as in the previous case. Given that reasons can be given in support of logicality and that logicality results in a measure having its maximum value when $c = 0$, this provides a reason for thinking that the strength of $B \Rightarrow E$ should be maximal whether five or six transactions are taken into consideration.

The measures $l$, $k$, $cf$, $or$, $yq$ and $yy$ all take on their maximum values when $c = 0$ and hence the extra transaction has no effect even though it is an instance of the rule. However, as we have seen, even though a rationale can be given for logicality, we have argued that it is inappropriate in the context of association rules. Instead, the maximality / minimality requirement states that a measure should be maximal if and only if $b = c = 0$. Given this requirement and monotonicity, it seems reasonable that the strength of the rule should be strictly greater with the addition of the extra transaction. Measures $rr$, $s$ and $ks$ are not maximal when $c = 0$, but they give an equal value for $B \Rightarrow E$ in the two cases. From their definitions in table 2 it is easy to see that when $c = 0$ these measures are independent of $a$.

The only measures that give what is intuitively the correct result in this case are $\phi$ and $k\phi$. Given that $k\phi$ satisfies all four requirements, this suggests that it is to be preferred to the other measures considered. In particular, it is to be preferred to $ks$ even though it also satisfies all the requirements. These considerations which have enabled us to discriminate between $k\phi$ and $ks$ suggest that the four requirements should be considered as necessary, but not sufficient requirements. It is also worth noting that although a simple example has been used to illustrate these differences, they are in fact based
on the underlying properties of the different measures.

## 6.2 Comparison with recent work

In a recent article, Greco et al. [21] provide a very interesting discussion of confirmation measures of rule interestingness. As noted earlier, they adopt the symmetry requirements proposed by Crupi et al. [23] and propose their own modified versions of (Ex₁) and logicality called (weak Ex₁) and (weak L) respectively. In light of these requirements, they advocate the following confirmation measures, which for a rule \( A \Rightarrow B \) can be written as:

\[
Z(\text{A} \Rightarrow \text{B}) = \frac{Pr(\text{B} | \text{A}) - Pr(\text{B})}{1 - Pr(\text{B})} = \frac{ad - bc}{(a + c)(c + d)} \quad \text{if} \quad Pr(\text{B} | \text{A}) > Pr(\text{B}) \\
= \frac{Pr(\text{B} | \text{A}) - Pr(\text{B})}{Pr(\text{B})} = \frac{ad - bc}{(a + c)(a + b)} \quad \text{if} \quad Pr(\text{B} | \text{A}) \leq Pr(\text{B})
\]

(12)

\[
\mathcal{A}(\text{A} \Rightarrow \text{B}) = \frac{Pr(\text{B}) - Pr(\text{B} | \neg \text{A})}{Pr(\text{B})} = \frac{ad - bc}{(a + b)(b + d)} \quad \text{if} \quad Pr(\text{B} | \text{A}) > Pr(\text{B}) \\
= \frac{Pr(\text{B}) - Pr(\text{B} | \neg \text{A})}{1 - Pr(\text{B})} = \frac{ad - bc}{(b + d)(c + d)} \quad \text{if} \quad Pr(\text{B} | \text{A}) \leq Pr(\text{B})
\]

(13)

The \( Z \) measure has also been favoured by Crupi et al. and is the certainty factor (see table 2), which was originally proposed by Shortliffe and Buchanan [32] and has been widely used in expert systems, while the \( \mathcal{A} \) measure is a further measure proposed by Greco et al. As noted in table 5, \( Z \) satisfies the symmetry and monotonicity requirements, but does not satisfy the maximality/minimality requirement since according to this requirement a measure should be maximal if and only if \( b = c = 0 \) and minimal if and only if \( a = d = 0 \) whereas \( Z \) is maximal if \( c = 0 \) and minimal if \( a = 0 \). Similarly, \( \mathcal{A} \) also fails this requirement since it is maximal if \( b = 0 \) and minimal if \( d = 0 \).

It is worth comparing two pairs of cases considered by Greco et al. In case 1 the number of transactions for a rule \( A \Rightarrow B \) are \( a = 100, b = 99, c = 0 \) and \( d = 1 \) while in case 2 \( a = 0, b = 1, c = 100 \) and \( d = 99 \). The measures yield the following results: \( Z(\text{A} \Rightarrow \text{B}) = 1 \) in case 1 and \( Z(\text{A} \Rightarrow \text{B}) = -1 \)
in case 2, while \( \mathcal{A}(A \Rightarrow B) = 1/199 \) in case 1 and \( \mathcal{A}(A \Rightarrow B) = -1/199 \) in case 2. Greco et al. point out that \( Z \) gives a bad representation in these cases. The reason for this is that the maximal and minimal values yielded by \( Z \) seem unwarranted in this case. The maximal value is unwarranted since although \( Pr(B|A) = 1 \) in case 1, \( Pr(B|\neg A) = 99/100 \) and they argue that both of these factors are relevant. The minimal value is unwarranted since although \( Pr(B|A) = 0 \) in case 2, \( Pr(B|\neg A) = 1/100 \). Comparing the measures proposed in this paper we find that \( \kappa s(A \Rightarrow B) = 101/200 \) in case 1 and \( \kappa s(A \Rightarrow B) = -101/200 \) in case 2, while \( \kappa \phi(A \Rightarrow B) = 1/2(1 + \frac{1}{\sqrt{199}}) \) in case 1 and \( \kappa \phi(A \Rightarrow B) = -1/2(1 + \frac{1}{\sqrt{199}}) \) in case 2.

Now consider the second pair of cases. In case 3 the number of transactions for a rule \( A \Rightarrow B \) are \( a = 100, b = 0, c = 99 \) and \( d = 1 \) while in case 4 \( a = 99, b = 1, c = 100 \) and \( d = 0 \). The measures yield the following results: \( Z(A \Rightarrow B) = 1/199 \) in case 3 and \( Z(A \Rightarrow B) = -1/199 \) in case 4, while \( \mathcal{A}(A \Rightarrow B) = 1 \) in case 3 and \( \mathcal{A}(A \Rightarrow B) = -1 \) in case 4. Greco et al. point out that \( \mathcal{A} \) gives a bad representation in these cases. The reason for this is that the maximal and minimal values yielded by \( \mathcal{A} \) seem unwarranted in this case. The maximal value is unwarranted since although \( Pr(B|\neg A) = 0 \) in case 3, \( Pr(B|A) = 100/199 \). The minimal value is unwarranted since although \( Pr(B|\neg A) = 0 \) in case 4, \( Pr(B|A) = 99/199 \). Comparing the measures proposed in this paper we find that \( \kappa s(A \Rightarrow B) = 101/398 \) in case 3 and \( \kappa s(A \Rightarrow B) = -101/398 \) in case 4, while \( \kappa \phi(A \Rightarrow B) = \frac{1+\sqrt{199}}{398} \) in case 3 and \( \kappa \phi(A \Rightarrow B) = -\frac{1+\sqrt{199}}{398} \) in case 4.

In these cases the \( \kappa s \) and \( \kappa \phi \) measures give encouraging results and do not suffer from the problems of \( Z \) in the first pair of cases or \( \mathcal{A} \) in the second pair of cases. Greco et al. also propose several ways of combining \( Z \) and \( \mathcal{A} \) which also give reasonable results for the cases considered above, although if the arguments proposed in this paper are correct these measures fail to satisfy desirable symmetry properties or satisfy properties such as \((\text{Ex}_1)\) which has been rejected here. However, there is clearly more scope for combining the measures \( Z \) and \( \mathcal{A} \) in different ways and further scope for exploring the relationship between the work of Greco et al. and the current work.

### 6.3 Null transactions revisited

As noted in section 3, there is a legitimate concern about the role of null transactions. For a rule \( A \Rightarrow B \), null transactions are those in which neither...
A nor $B$ occurs and in this paper the number of null transactions is represented by $d$. In many scenarios $d$ is very large and so, for example, the rules under consideration have very small support. The concern is that in such cases measures which are null invariant may be very sensitive to fluctuations in $d$. To investigate this concern, two scenarios are considered which have low support. In case 1, $a = b = c = 1000$, while $d = 100,000$ and so the support is approximately 1%; in case 2, $a = 100$, $b = c = 1000$ and again $d = 100,000$ and so the support is approximately 0.1%. In each case, results are presented in table 7 for the new measure $k\phi$, which in this case is equivalent to $ks$ since $b = c$, and for the $b$ measure for comparison. The value of $d$ is varied in order to determine how much the values of the two measures vary.

In case 1 the results for $k\phi$ are very stable when $d$ is changed from its initial value of 100,000. When $d$ is decreased or increased by 20% (to 80,000 and 120,000 respectively) the result for $k\phi$ changes by less than 1%. Even when $d$ is halved to 50,000 or doubled to 50,000 the result for $k\phi$ changes by less than 3.5%. By contrast, the results for the $b$ measure change much more dramatically. In fact, they change almost in inverse proportion to $d$ with, for example, an increase of 87% when $d$ is halved and a decrease of 48% when $d$ is doubled.

A similar pattern is found in case 2 although the results for $k\phi$ vary slightly more this time. When $d$ is decreased or increased by 20% the result for $k\phi$ changes by less than 1%. The maximum change of 19% occurs when $d$ is halved to 50,000. As in case 1, the percentage changes in $b$ are much greater for $k\phi$ even though they are slightly lower than in case 1.

These results illustrate that while confirmation measures can be highly sensitive to the number of null transactions, this depends on the measure in question and the rules under consideration. It might be thought that the results for $k\phi$ are less stable in case 2 than in case 1 because of the lower support, but this is not the case. It is easy to show that if the initial value of $d$ in case 2 were 1,000,000 instead of 100,000 the results would be much more stable as $d$ is varied. In fact, the results suggest that $k\phi$ is more stable for higher values of confirmation and less stable for lower values. This would certainly make sense since, given proposition 1, if $ad \approx bc$ even modest changes in $d$ could change a positive association to a negative one. And stability for high values of confirmation is what is needed since we are interested in finding rules with high values.
7 Conclusion

In data mining there has been a great deal of investigation of measures to quantify the interestingness of association rules. The focus in this paper has been on a particular class of objective measures known as confirmation measures. Since such measures are based on well defined notions of probabilistic dependence, they are ideally suited to the task. In particular, it has been argued that they have advantages over null-invariant measures since they characterize positive/neutral/negative dependence in an appropriate way.

Three further properties have been discussed: the symmetry requirement, the maximality/minimality requirement and monotonicity. These properties have been motivated through simple examples, which are intended to illustrate fundamental features that are desirable in many data mining contexts. They have also been compared in detail with other properties proposed in the literature. Although none of the existing confirmations satisfies all these requirements, two new measures have been proposed which do. Some advantages of these measures have been considered in the context of monotonicity by comparing them with other measures which also satisfy this property. This discussion highlights a particular advantage of one of the new measures. It has also been shown that these new measures are relatively stable when the number of null transactions varies.

A preliminary comparison has also been made with recent proposals by Greco et al. [21], but there is certainly scope for a more detailed comparison both in terms of suitable properties for measures of interestingness and the particular measures proposed. Furthermore, in both their work and Crupi et al. [23] normalization approaches have been applied to generate measures satisfying the properties they propose from rules which do not. It would be interesting to see whether similar approaches might be used to generate other measures satisfying the properties proposed in this paper.

The discussion of differences between measures which satisfy monotonicity suggests that there are further reasonable requirements for confirmation measures of interestingness. A goal of future research will be to explore such extensions of monotonicity as well as other properties in the literature to see whether the list of requirements can be extended further. Other combinations of existing measures and/or new measures will also be investigated to determine whether they have advantages over the measures proposed here.
Acknowledgments

The author would like to thank anonymous reviewers for very helpful suggestions.

References


[31] I. Brzezinska, S. Greco, R. Slowinski, Mining pareto-optimal rules with respect to support and confirmation or support and anti-support, Engineering Applications of Artificial Intelligence 20 (2007) 587–600.

Table 2: Definitions of confirmation measures. †These measures have a value of one when A and B are probabilistically independent and so the log of these measures has been taken to ensure that they conform to the definition of a confirmation measure.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Probability</th>
<th>Number of Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference, d</td>
<td>( Pr(B</td>
<td>A) - Pr(B) )</td>
</tr>
<tr>
<td>Log-Ratio, r</td>
<td>( \log \frac{Pr(B</td>
<td>A)}{Pr(B)} )</td>
</tr>
<tr>
<td>Log-Likelihood, l</td>
<td>( \log \frac{Pr(A</td>
<td>B)}{Pr(A</td>
</tr>
<tr>
<td>Kemeny-Oppenheim, k</td>
<td>( Pr(A</td>
<td>B) - Pr(A</td>
</tr>
<tr>
<td>Normalized Difference, ( s )</td>
<td>( Pr(B</td>
<td>A) - Pr(B</td>
</tr>
<tr>
<td>Carnap, b</td>
<td>( Pr(A,B) - Pr(A)Pr(B) )</td>
<td>( \frac{ad-bc}{(a+b)(a+c)} )</td>
</tr>
<tr>
<td>Certainty Factor, ( cf )</td>
<td>( \frac{Pr(B</td>
<td>A) - Pr(B)}{Pr(B) - Pr(A</td>
</tr>
<tr>
<td>Relative Risk†, ( rr )</td>
<td>( log \frac{Pr(B</td>
<td>A)}{Pr(B</td>
</tr>
<tr>
<td>Odds Ratio†, ( or )</td>
<td>( log \frac{Pr(A,B)Pr(\neg A</td>
<td>B)}{Pr(\neg A,B)Pr(A</td>
</tr>
<tr>
<td>Conviction†, ( conv )</td>
<td>( log \frac{Pr(A)Pr(\neg B)}{Pr(\neg A)Pr(\neg B)} )</td>
<td>( \log \frac{a(a+b)(c+d)}{(a+c)(a+b+c+d)} )</td>
</tr>
<tr>
<td>Yule’s Q, ( yq )</td>
<td>( \frac{Pr(A,B)Pr(\neg A,\neg B) - Pr(\neg A,B)Pr(A,\neg B)}{Pr(A,B)Pr(\neg A,\neg B) + Pr(\neg A,B)Pr(A,\neg B)} )</td>
<td>( \log \frac{ad-bc}{ad+bc} )</td>
</tr>
<tr>
<td>Yule’s Y, ( yy )</td>
<td>( \sqrt{Pr(A,B)Pr(\neg A,\neg B) - Pr(\neg A,B)Pr(A,\neg B)} ) ( \sqrt{Pr(A,B)Pr(\neg A,\neg B) + Pr(\neg A,B)Pr(A,\neg B)} )</td>
<td>( \log \frac{ad-bc}{ad+bc} )</td>
</tr>
<tr>
<td>( \phi )-Coefficient, ( \phi )</td>
<td>( \frac{Pr(A,B)Pr(\neg A,\neg B) - Pr(\neg A,B)Pr(A,\neg B)}{\sqrt{Pr(A,B)Pr(\neg A,\neg B) + Pr(\neg A,B)Pr(A,\neg B)}} )</td>
<td>( \log \frac{(a+c)(a+b)(b+d)(c+d)}{a+b+c+d} )</td>
</tr>
<tr>
<td>One-Way Support, ( ows )</td>
<td>( Pr(B</td>
<td>A)\log \frac{Pr(A,B)}{Pr(B</td>
</tr>
<tr>
<td>Two-Way Support, ( tws )</td>
<td>( Pr(A,B)\log \frac{Pr(A,B)}{Pr(\neg A,\neg B)} )</td>
<td>( \log \frac{a(a+b+c+d)}{(a+c)(a+b)} )</td>
</tr>
<tr>
<td>Klosgen, ( kl )</td>
<td>( \sqrt{Pr(A,B)\max(Pr(B</td>
<td>A) - Pr(B), Pr(A</td>
</tr>
<tr>
<td>Collective Strength, ( cs )</td>
<td>( \log \left[ \frac{Pr(A,B)+Pr(\neg A\neg B)}{Pr(A)Pr(B)+Pr(\neg A)Pr(\neg B)} \right] \times \frac{1-Pr(A,B)-Pr(\neg A\neg B)}{1-Pr(A)Pr(B)} )</td>
<td>( \log \frac{a(a+b+c+d)}{(a+c)(a+b)(b+d)(c+d)} )</td>
</tr>
</tbody>
</table>

37
Table 3: A comparison of a confirmation and null-invariant measure.

<table>
<thead>
<tr>
<th>Data set</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>k measure</th>
<th>Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>10,000</td>
<td>1,000</td>
<td>1,000</td>
<td>100,000</td>
<td>0.978</td>
<td>0.91</td>
</tr>
<tr>
<td>$D_2$</td>
<td>10,000</td>
<td>1,000</td>
<td>1,000</td>
<td>100</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>$D_3$</td>
<td>10,000</td>
<td>1,000</td>
<td>1,000</td>
<td>10</td>
<td>-0.043</td>
<td>0.91</td>
</tr>
<tr>
<td>$D_4$</td>
<td>100</td>
<td>1,000</td>
<td>1,000</td>
<td>100,000</td>
<td>0.804</td>
<td>0.09</td>
</tr>
<tr>
<td>$D_5$</td>
<td>1,000</td>
<td>1,000</td>
<td>1,000</td>
<td>100,000</td>
<td>0.961</td>
<td>0.5</td>
</tr>
<tr>
<td>$D_6$</td>
<td>1,000</td>
<td>1,000</td>
<td>1,000</td>
<td>100</td>
<td>-0.29</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 4: Contingency table for the rule $Milk \Rightarrow Bread$.

<table>
<thead>
<tr>
<th></th>
<th>Bread</th>
<th>¬Bread</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>¬Milk</td>
<td>45</td>
<td>50</td>
<td>95</td>
</tr>
<tr>
<td>Sum</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 5: Summary of requirements met by various measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Confirmation Measure</th>
<th>Symmetry</th>
<th>Maximality / Minimality</th>
<th>Monotonicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference, $d$</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Log-Ratio, $r$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Log-Likelihood, $l$</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Kemeny-Oppenheim, $k$</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Normalized Difference, $s$</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Carnap, $b$</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Certainty Factor, $cf$</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Relative Risk, $rr$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Odds Ratio, $or$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Conviction, $conv$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Yule’s Q, $yq$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Yule’s Y, $yy$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>$\phi$-Coefficient, $\phi$</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>One-Way Support, $ows$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Two-Way Support, $tws$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Klosgen, $kl$</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Collective Strength, $cs$</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>
Table 6: A comparison of all the measures that satisfy monotonicity.

<table>
<thead>
<tr>
<th>Confirmation Measure</th>
<th>( M \Rightarrow E ) (5 transactions)</th>
<th>( M \Rightarrow E ) (6 transactions)</th>
<th>( B \Rightarrow E ) (5 transactions)</th>
<th>( B \Rightarrow E ) (6 transactions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l )</td>
<td>( \log(3) )</td>
<td>( \log(3) )</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>( k )</td>
<td>( 1/2 )</td>
<td>( 1/2 )</td>
<td>( 1 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( s )</td>
<td>( 2/3 )</td>
<td>( 3/4 )</td>
<td>( 3/4 )</td>
<td>( 3/4 )</td>
</tr>
<tr>
<td>( cf )</td>
<td>( 4/9 )</td>
<td>( 1/2 )</td>
<td>( 1 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( rr )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \log(4) )</td>
<td>( \log(4) )</td>
</tr>
<tr>
<td>( or )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>( yq )</td>
<td>( 1 )</td>
<td>( 1 )</td>
<td>( 1 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( yy )</td>
<td>( 1 )</td>
<td>( 1 )</td>
<td>( 1 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( \phi )</td>
<td>( 2/3 )</td>
<td>( \sqrt{2}/2 \approx 0.707 )</td>
<td>( \sqrt{6}/4 \approx 0.612 )</td>
<td>( \sqrt{2}/2 \approx 0.707 )</td>
</tr>
<tr>
<td>( ks )</td>
<td>( 7/12 \approx 0.583 )</td>
<td>( 5/8 = 0.625 )</td>
<td>( 7/8 )</td>
<td>( 7/8 )</td>
</tr>
<tr>
<td>( k\phi )</td>
<td>( 7/12 \approx 0.583 )</td>
<td>( \frac{1+\sqrt{7}}{4} \approx 0.604 )</td>
<td>( \frac{4+\sqrt{5}}{8} \approx 0.806 )</td>
<td>( \frac{2+\sqrt{2}}{4} \approx 0.854 )</td>
</tr>
</tbody>
</table>

Table 7: A comparison of how two measures depend on the number of null transactions.

<table>
<thead>
<tr>
<th>Confirmation Measure</th>
<th>( d = 100000 )</th>
<th>( d = 50000 )</th>
<th>( d = 80000 )</th>
<th>( d = 120000 )</th>
<th>( d = 200000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k\phi )</td>
<td>0.726</td>
<td>0.702</td>
<td>0.720</td>
<td>0.730</td>
<td>0.738</td>
</tr>
<tr>
<td>( b )</td>
<td>0.0093</td>
<td>0.0174</td>
<td>0.0115</td>
<td>0.0079</td>
<td>0.0048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confirmation Measure</th>
<th>( d = 100000 )</th>
<th>( d = 50000 )</th>
<th>( d = 80000 )</th>
<th>( d = 120000 )</th>
<th>( d = 200000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k\phi )</td>
<td>0.442</td>
<td>0.358</td>
<td>0.420</td>
<td>0.458</td>
<td>0.491</td>
</tr>
<tr>
<td>( b )</td>
<td>0.00086</td>
<td>0.00147</td>
<td>0.00104</td>
<td>0.00074</td>
<td>0.00047</td>
</tr>
</tbody>
</table>