Università
Ca' Foscari
Venezia

**Department
of Management**

Working Paper Series

Emilio Celotto

**Sensitivity and simmetry of
Confirmation Measures**

**Working Paper n. 22/2014
October 2014**

ISSN: 2239-2734

This Working Paper is published under the auspices of the Department of Management at Università Ca' Foscari Venezia. Opinions expressed herein are those of the authors and not those of the Department or the University. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional nature.

Sensitivity and simmetry of Confirmation Measures

EMILIO CELOTTO

emi.web@tin.it

Dept. of Management
Università Ca' Foscari Venezia

(October 2014)

Abstract

Bayesian Confirmation Measures can be used in ranking inductive rules obtained as an output when using Association Rules in database mining. Many measures of this kind have been defined and the choice of the right one can be considered as a further challenge to be faced. In this paper we try to classify them observing their sensitivity and simmetry properties.

Keywords: Multi-Criteria, Association rules, Confirmation measure, Data Mining.

JEL Classification Numbers: C10, C44, C80

1. Introduction

Multi-Criteria methodologies whose outputs are inductive rules have been actively studied, like Decision Rules or Association Rules, and many papers have dealt with the application of those methods to many diverse fields. Association Rules were applied, e.g., to Financial Markets [16], Internet (optimization of search engine results) [5], Network Security [21], Artificial Intelligence [15], Image Analysis [12].

Though in different ways, inductive rules need always to be extracted from a dataset which is often rather large, usually resulting in a very high number of rules. Not seldom a huge set of rules leads to an excess of information and needs to be sieved in order to be actually useful. The outcome of a sieving algorithm is a smaller set of rules which require to be ranked with respect to their importance, which is usually defined by means of the so called Interestingness Measures. A very large variety of them have been proposed by the specialized literature (see, e.g., [8]).

Inductive rules are usually expressed in the form $E \rightarrow H$, where the premise E is called also evidence and the conclusion H hypothesis. When different rules must be compared and ranked, a natural way to give them a score is based on at least two quantities, prior probability $P(H)$ and posterior probability $P(H|E)$, which allow the evaluation of the degree to which E supports or contradicts H (or, the “weight of evidence E in favour or against hypothesis H ” [9]). The comparison of the two probabilities usually takes place by means of a function $c(H,E)$ called confirmation measure; we focus here on Bayesian Confirmation Measures, which are required to satisfy the following minimal requirements:

- i. $c(H,E) > 0$ in case of confirmation, i.e. if $P(H|E) > P(H)$;
- ii. $c(H,E) = 0$ in case of indifference, i.e. if $P(H|E) = P(H)$;
- iii. $c(H,E) < 0$ in case of disconfirmation, i.e. if $P(H|E) < P(H)$.

After recalling some Confirmation Measures in Section 2, in Sections 3 and 4 we will outline some of the main problems related to the sensitivity on end-of-scale values and to symmetry properties of those measures.

2. Confirmation Measures

Let us consider a short repertoire of some well known Confirmation Measures.

Kemeny's f [13], Rescher's Z [19], Carnap's d [2], Keynes's r [14] are well known measures, their analytical expression are reported in Table 1 along with the more recently introduced Shogenji's J [20].

Table 1. Definitions of some Confirmation Measures

Name	$c(H,E)$	Author
f	$\frac{P(E H)-P(E \neg H)}{[P(E H)+P(E \neg H)]}$	Kemeny [13]
Z	$\frac{[P(H E)-P(H)]/[1-P(H)]}{[P(H E)-P(H)]/[P(H)]}$ in case of confirmation in case of disconfirmation	Rescher [19]
d	$P(H E)-P(H)$	Carnap [2]
l	$\log[P(E H)/P(E \neg H)]$	Good [9]
η	$\frac{[P(H E)-P(H)]/[P(H E)+P(H)]}{[P(H E)-P(H)]/[P(H E)+P(H)]}$	Popper [18]
J	$1-\log[P(H E)]/\log[P(H)]$	Shogenji [20]
r	$\log[P(H E)/P(H)]$	Keynes [14]

Let us now recall that two confirmation measures are said to be ordinally equivalent if the rank-order they induce is strictly preserved – despite the specific values assigned to the probabilities $P(H|E)$ and $P(H)$ – whichever of them have been chosen. In other words, the difference between them is only due to the chosen measure unit but the rank is preserved:

Definition. Two confirmation measures $c(H,E)$ and $c^*(H,E)$ are said ordinally equivalent when for any two couples (H_1,E_1) and (H_2,E_2) :

$$c(H_1,E_1) < c(H_2,E_2) \text{ if and only if } c^*(H_1,E_1) < c^*(H_2,E_2).$$

To prove ordinal equivalence we introduce a normalization transformation useful to obtain an infinite set of normalized measures (i.e., with range given by the interval $[-1,1]$) starting from a single unbounded measure (i.e., with range $(-\infty,+\infty)$).

Given $c(H,E)$ an unbounded confirmation measure, we can define the strictly increasing transformation

$$N_b(c(H,E)) = 1 - 2/[1+b^{c(H,E)}] \text{ with } b > 1.$$

It can be easily proved that:

- if $c(H,E) \rightarrow +\infty$ then $N_b[c(H,E)] \rightarrow +1$
- if $c(H,E) = 0$ then $N_b[c(H,E)] = 0$
- if $c(H,E) \rightarrow -\infty$ then $N_b[c(H,E)] \rightarrow -1$.

For each chosen value of b , $N_b[c(H,E)]$ is therefore a normalized confirmation measure ordinally equivalent to $c(H,E)$.

Observe that the transformation can equivalently be rewritten as

$$N_b(c(H,E)) = (b^{c(H,E)} - 1) / (b^{c(H,E)} + 1) = \tanh[\ln(b)/2 * c(H,E)]$$

where $\tanh$ is the hyperbolic tangent $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$.

$N_b[c(H,E)]$ provides an infinite set of normalized confirmation measures from which it is possible to select a particular one by just selecting a specific value of the parameter b .

Let us consider Kemeny's f and Good's l , for example. Without loss of generality we can consider Good's l defined using the logarithm with base 10 (as originally defined in [8]). Therefore Kemeny's f and Good's l are immediately recognised as being ordinally equivalent $f(H,E) = N_b[l(H,E)]$ with $b=10$. If, instead, one considers Good's l defined by means of the natural logarithm, the two measures are again recognised as ordinally equivalent since $f(H,E) = N_b[l(H,E)]$ with $b=e$, or $f(H,E) = \tanh[l(H,E)/2]$ ¹.

Another example of the power of the above defined transformation can be obtained by considering Keynes's r and Popper's η . Popper calls it the "explanatory power of E with respect to H "; later the same measure was used in [17] and [10] and was called $c_r^*(H,E)$ in [6]. Assuming in Keynes's r the natural logarithm it can be easily verified that $\eta(H,E) = N_b[r(H,E)]$ with $b=e$. Therefore also Keynes's r and Popper's η are immediately recognised as ordinally equivalent.

3. Sensitivity of Confirmation measures

We address now very briefly a problem that should be more deeply investigated, i.e. the sensitivity of Confirmation Measures for what concerns end-of-scale values of $P(H|E)$ and $P(H)$. As an example we will compare the behaviour of three confirmation measures: Kemeny's f , Rescher's Z and a new one, Z_g , defined below.

Let us first consider the effect of moving around the values $P(H|E)=P(H)=0$ on confirmation measures' scores, i.e., sensitivity to low range values.

It is easy to understand this effect by means of numeric examples. Consider for instance, $P(H|E)=0.01$ and $P(H)=0.0000001$; in this case $Z=0.01$ whilst $f=0.99998$. It means that, Z does not overestimate the resulting small absolute value gain $P(H|E)-P(H)=0.0099999$, while f reaches nearly 1, the highest score. On the other hand, when $P(H|E)=0.0000001$ and $P(H)=0.01$ one has $Z=-0.99999$ and $f=-0.99998$. Therefore, both Z and f overestimate the small absolute value loss $P(H|E)-P(H)=-0.0099999$.

To solve the overestimation problem we propose a new confirmation measure Z_g as follows:

¹ Rather strangely, in the paper by Good [9] the author wrote that $f(H,E)$ can be obtained from $l(H,E)$ by means of the increasing function $\sinh[l(H,E)/2]$ where $\sinh$ is the hyperbolic sin $\sinh(x) = (e^x - e^{-x})/2$. But $\sinh[l(H,E)/2]$ corresponds to the expression $[P(E|H) - P(E|\neg H)] / \{2[P(E|H)P(E|\neg H)]^{1/2}\}$ which does not correspond to $f(H,E)$.

$$Z_g = (P(H|E)-P(H))/(1-P(H)) \text{ in case of confirmation}$$

$$Z_g = (P(H|E)-P(H))/(1-P(H|E)) \text{ in case of disconfirmation}$$

As a matter of fact, Z_g mimics the definition of Z in case of confirmation.

If we consider high range values, we encounter a similar problem around $P(H|E)=P(H)=1$.

In this case, the use of Carnap's d offers a way to overcome the overestimation problem. In fact, consider the couple of values $P(H|E)=0.99999$ and $P(H)=0.99990$, the small difference $P(H|E)-P(H)=0.00009$ results in three rather different evaluations when using the three confirmation measures f , Z and d : the scores are respectively $f=0.81820$, $Z=0.90000$ and $d=0.00009$. If, instead, $P(H|E)=0.99990$ and $P(H)=0.99999$ the difference $P(H|E)-P(H)=-0.00009$ leads to the following evaluations: $f=-0.81820$, $Z=-0.00009$ and $d=0.00009$. Therefore, only Carnap's d gives a little score when considering small variations around $P(H|E)=1$, $P(H)=1$. On the other hand, when $P(H|E)=0.99999$ and $P(H)=0.01$ a big increase $P(H|E)-P(H)=0.98999$ obtains a value close to 1 from each one of the three measures. Analogously when $P(H|E)=0.01$ and $P(H)=0.99999$ a big decrease $P(H|E)-P(H)=-0.98999$ obtains nearly -1 from each of the measures. The last numeric examples concern the sensitivity around $P(H|E)=1$, $P(H)=0$ and, respectively, $P(H|E)=0$, $P(H)=1$.

A systematic method to measure sensitivity was proposed in [4]: the variation of the values of a confirmation measure can be investigated varying $P(H|E)$ given a fixed value of $P(H)$. We suggest also two different approaches, consisting in fixing $P(H|E)$ and varying $P(H)$ and in fixing $P(H|E)+P(H)$ varying $P(H|E)$.

In the examples reported below we will consider also the normalized version of Shogenji's J , as defined in [1]:

$$J'(H,E) = 2^{J(H,E)} - 1 .$$

Figure 1 reports a comparison of Kemeny's f , Shogenji's J' , Rescher's Z and Popper's η , for 3 fixed values of $P(H)$ (0.2, 0.5 and 0.8) varying $P(H|E)$.

The variation of the same measures with respect to $P(H)$ is illustrated by Figure 2 while Figure 3 represents the four measures when the sum $P(H|E) + P(H)$ is fixed and $P(H|E)$ varies.

Figure 1. Sensitivity visualization for Kemeny's f , Shogenji's J' , Rescher's Z , Popper's η , for $P(H) = 0.2$ (red line), 0.5 (green line) and 0.8 (blue line).

Figure 2. Sensitivity visualization for Kemeny's f , Shogenji's J' , Rescher's Z , Popper's η , for $P(H|E) = 0.2$ (red line), 0.5 (green line) and 0.8 (blue line).

Figure 3. Sensitivity visualization for Kemeny's f , Shogenji's J' , Rescher's Z , Popper's η , for $P(H|E) + P(H) = 0.4$ (red line), 1.0 (green line) and 1.6 (blue line).

4. Symmetries of Confirmation measures

In this Section we examine the links between symmetries of Confirmation measures and those of Interestingness measures. Confirmation measures' symmetries were actually introduced by Carnap [2] but investigated also recently (see, e.g., [8], [11]). The four main symmetries are formulated with respect to E and H as follows:

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):	$c(H,E) = c(\neg H,\neg E)$

On the other hand, [22] defines five properties of interestingness measures, and here we analyse three of them, Symmetry under Variable Permutation (SVP), Antisymmetry under Row/Column Permutation (ARP/ACP) and Inversion Invariance (INI) which show important links with the four classical confirmation measures' symmetries. We follow the very clear interpretation of Tan's properties illustrated in [7], a general survey on Interestingness Measures for Data Mining. SVP, ARP/ACP and INI properties, indeed, can be considered useful in better understanding the difference between statistical and data mining approaches, which focus on association (or independence) between variables and, respectively, look for hidden patterns in data.

Let us observe that since Interestingness measures can also be defined starting from contingency tables, operating on contingency tables as they were matrices, it becomes easy to express some important properties of interestingness measures.

A rule $E \rightarrow H$ defined over a finite number N of cases, and the corresponding degree of association between premise E and conclusion H , can be expressed referring to a 2×2 contingency table (see Table 2), containing the values of $\text{supp}(E \& H)$, $\text{supp}(\neg E \& H)$, $\text{supp}(E \& \neg H)$, $\text{supp}(\neg E \& \neg H)$ where $a+b+c+d=N$. The contingency table represents the relation between E and H as a numerical expression of the rule $E \rightarrow H$.

Table 2. A 2×2 contingency table

	H	$\neg H$
E	$a=\text{supp}(E \& H)$	$c=\text{supp}(E \& \neg H)$
$\neg E$	$b=\text{supp}(\neg E \& H)$	$d=\text{supp}(\neg E \& \neg H)$

Property 1: Symmetry under Variable Permutation (SVP)

A confirmation measure satisfies SVP when $c(H,E) = c(E,H)$, i.e. the rules $E \rightarrow H$ and $H \rightarrow E$ receive the same evaluation. Variable permutation means therefore to exchange the role of E and H and SVP corresponds to Commutativity Symmetry (CS). This property means that the degree of association between E and H does not depend on how the roles of premise and conclusion are assigned.

In terms of contingency tables the property means that, given Table 2 (E and $\neg E$ on the rows) one can immediately write Table 3 (with E and $\neg E$ as columns).

Table 3. Variable permutation: from $E \rightarrow H$ to $H \rightarrow E$

	E	$\neg E$
H	a	b
$\neg H$	c	d

Property 2: Antisymmetry under Row/Column Permutation (ARP/ACP)

Property 2 is double in the sense that it can be written in two ways, i.e., with respect to rows and to columns, respectively.

A confirmation measure satisfies ARP (Row Permutation) when $c(H,E) = -c(H,\neg E)$.

This means that $E \rightarrow H$ and $\neg E \rightarrow H$ receive exactly the opposite evaluation.

Starting from Table 2, ARP means the contingency matrix reported on Table 4.

Table 4 – Row permutation: from $E \rightarrow H$ to $\neg E \rightarrow H$

	H	$\neg H$
$\neg E$	b	d
E	a	c

In a similar way, a Confirmation measure satisfies ACP (Column Permutation) when $c(H,E) = -c(\neg H,E)$. Therefore, ACP corresponds to Hypothesis Symmetry (HS).

In this case the rules $E \rightarrow H$ and $E \rightarrow \neg H$ are the ones which obtain exactly the opposite evaluation. ACP corresponds to have the contingency Table 5 as soon as one has Table 2.

Table 5 – Column permutation: from $E \rightarrow H$ to $E \rightarrow \neg H$

	$\neg H$	H
E	c	a
$\neg E$	d	b

Property 3: Inversion Invariance (INI)

A confirmation measure satisfies INI when $c(H,E) = c(\neg H, \neg E)$. Therefore, INI corresponds to Total Symmetry (TS). In this case $\neg E \rightarrow \neg H$ obtains the same value of $E \rightarrow H$. Inversion Invariance means to have the contingency matrix defined in Table 6.

Table 6 – Inversion: from $E \rightarrow H$ to $\neg E \rightarrow \neg H$

	$\neg H$	H
$\neg E$	d	b
E	c	a

INI can be also obtained from ARP/ACP by applying, for instance, first Row Permutation and then Column Permutation.

Correspondence between the four classical symmetries and Tan's properties

Table 7 summarizes the correspondences between the four classical symmetries and Tan properties ('yes' stands for corresponding, 'no' stands for not-corresponding).

Table 7 – Classical symmetries vs. Tan properties

	CS	ES	HS	TS
SVP	Yes	no	no	no
ARP/ACP	No	yes	yes	no
INI	No	no	no	yes

Therefore, Evidence Symmetry, Commutativity Symmetry and Total Symmetry characterize statistical association measures – where the order of E and H is not important - while only Hypothesis Symmetry – corresponding to ACP - is a desirable property in selecting Bayesian confirmation measures.

5. Conclusions and further research

Bayesian Confirmation Measures, Interestingness Measures and statistical Association Measures have many similarities. In this paper we start to investigate their relationships, which has both theoretical and practical meaning, for instance it has many possible outcomes in considering rules' selection.

The investigation of a wider set of measures and eventually find further theoretical properties is needed and a general framework to study general properties of Confirmation Measures is our current main research topic.

References

- [1] Atkinson, D. (2012) Confirmation and justification. *Synthese*, 184 (1), 49-61.
- [2] Carnap R. (1962) *Logical foundations of probability*. Second edition. University of Chicago Press, Chicago, IL.
- [3] Crupi V., Festa R., Buttasi C. (2010) *Towards a grammar of Bayesian confirmation*, in: M. Suárez, M. Dorato, M. Rédei (Eds.), *Epistemology and Methodology of Science*, Springer, Dordrecht, 2010, pp. 73–93.
- [4] Crupi V., Tentori K., Gonzalez M. (2007) On Bayesian measures of evidential support: theoretical and empirical issues, *Philosophy of Science*, 74, 229– 252.
- [5] Du Ya Jun, Li Hai Ming (2010) Strategy for mining association rules for web pages based on formal concept analysis, *Applied Soft Computing*, 10 (3), 772-783.
- [6] Festa R. (1999) Bayesian confirmation. In: Galavotti M., Pagnini A. (eds) *Experience, reality, and scientific explanation*. Kluwer, Dordrecht, pp 55–87.
- [7] Geng L., Hamilton H. J. (2006) Interestingness Measures for Data Mining: A Survey, *ACM Computing Surveys*, Vol. 38, No. 3, Article 9, Publication date: September 2006.
- [8] Glass D.H. (2013) Confirmation measures of association rule interestingness, *Knowledge-Based Systems* 44, 65–77.
- [9] Good I.J., (1985) Weight of Evidence: A Brief Survey, *Bayesian Statistics 2*, 249–270, *Proceedings of the Valencia International Meetings on Bayesian Statistics*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds), Elsevier, Dordrecht.
- [10] Greco S., Słowiński R., Szczęch I. (2009) Analysis of monotonicity properties of some rule interestingness measures, *Control and Cybernetics* 38 (1), 9–25.
- [11] Greco S., Słowiński R., Szczęch I.: (2012) Properties of rule interestingness measures and alternative approaches to normalization of measures. *Information Sciences* 216, 1–16, Elsevier , Dordrecht.
- [12] Karabatak Murat, Ince M. Cevdet, Sengur Abdulkadir (2011) Wavelet domain association rules for efficient texture classification, *Applied Soft Computing*, 11 (1), 32-38.
- [13] Kemeny J., Oppenheim, P. (1952) Degrees of factual support. *Philosophy of Science* 19, 307–324.
- [14] Keynes J., (1921) *A Treatise on Probability*. Macmillan, London.
- [15] Kuo R.J., Chao C.M., Chiu Y.T. (2011) Application of particle swarm optimization to association rule mining, *Applied Soft Computing*, 11 (1), 326-336.
- [16] Paranjape-Voditel Preeti, Deshpande Umesh (2013) A stock market portfolio recommender system based on association rule mining, *Applied Soft Computing* 13 (2), 1055-1063.
- [17] Pawlak Z. (2004) Some issues on Rough Sets. *Transactions on Rough Sets I*. LNCS 3100, 1–58.

- [18] Popper K.R. (1959) *The Logic of Scientific Discovery*. Hutchinson, London.
- [19] Rescher N. (1958) A theory of evidence, *Philosophy of Science* 25, 83–94.
- [20] Shogenji T. (2012) The Degree of Epistemic Justification and the Conjunction Fallacy, *Synthese* 184 (1), 29-48.
- [21] Tajbakhsh A., Rahmati M., Mirzaei A. (2009) Intrusion detection using fuzzy association rules, *Applied Soft Computing* 9 (2), 462-469.
- [22] Tan P., Kumar V., Srivastava J. (2002) Selecting the right interestingness measure for association patterns. In *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002)*. Edmonton, Canada. 32–41.