Assessing the quality of rules with a new monotonic interestingness measure *Z*

Roman Słowiński

Poznań University of Technology,

Institute for Systems Research, PAS, Poland

Salvatore Greco

University of Catania, Italy

Izabela Szczęch

Poznań University of Technology, Poland

ICAISC 2008, Zakopane

Presentation plan

- **Introduction**
- **Basic quantitative characteristics of rules**
- **Bayesian confirmation measures**
- **Normalization of confirmation measures**
- ***Z*-measure**
- **Symmetry properties of confirmation measures**
- **Other desired properties of confirmation measures**
- **Confirmation perspective on support – anti-support** Pareto **border**
- **Practical application of the results**
- **Conclusions**

Introduction - motivations

The number of rules
induced from datasets is usually quite large

- **overwhelming for human comprehension,**
- **many rules are irrelevant or obvious (low practical value)**

rule evaluation – interestingness (attractiveness) measures
(e.g. support, confidence, gain, lift)

- **each measure was proposed to capture different characteristics of rules**
- **the number of proposed measures is very large**
- **easiness of interpretation of NORMALIZED measures**

Introduction - motivations

The choice of interestingness measure for a certain application is a difficult problem

- **the users expectations vary,**
- **the number of proposed measures is overwhelming**

properties of interestingness measures, which reflect users′ expectations towards the behavior of measures in particular situations

need to analyze measures with respect to their properties

Introduction – rule induction

- **Patterns in form of rules are induced from a data table**
- ***S*=⟨*U*, *A*⟩ – *data table*, where *U* and *A* are finite, non-empty sets**
  U* – *universe*; *A* – set of *attributes
- ***S*=⟨*U*, *C*, *D*⟩ – *decision table*, where C – set of *condition attributes*,**
  ***D* – set of *decision attributes*, $C \cap D = \varnothing$**

- ***Decision rule* or *association rule* induced from *S***
  is a *consequence relation*: f ® y *read as* *if* f *then* y
  where f is condition (evidence or premise)
  and y is conclusion (hypothesis or decision)
  formula built from attribute-value pairs (*q*,*v*)
- **If the division into independent and dependent attributes is fixed, then rules are regarded as decision rules, otherwise as association rules.**

Introduction – rule induction

Characterization of nationalities

U	*Height*	*Hair*	*Eyes*	*Nationality*	Support
1	*tall*	*blond*	*blue*	*Swede*	270
2	*medium*	*dark*	*hazel*	*German*	90
3	*medium*	*blond*	*blue*	*Swede*	90
4	*tall*	*blond*	*blue*	*German*	360
5	*short*	*red*	*blue*	*German*	45
6	*medium*	*dark*	*hazel*	*Swede*	45
		C		**D**	

Introduction – rule induction

- **Decision rules** induced from „characterization of nationalities":
  1) If (*Height*, *tall*), then (*Nationality*, *Swede*)
  2) If (*Height*, *medium*) and (*Hair*, *dark*), then (*Nationality*, *German*)
  3) If (*Height*, *medium*) and (*Hair*, *blond*), then (*Nationality*, *Swede*)
  4) If (*Height*, *tall*), then (*Nationality*, *German*)
  5) If (*Height*, *short*), then (*Nationality*, *German*)
  6) If (*Height*, *medium*) and (*Hair*, *dark*), then (*Nationality*, Swede)

Certainty and coverage factors

Rule number	Certainty	Coverage	Support	Strength
1	0.43	0.67	270	0.3
2	0.67	0.18	90	0.1
3	1.00	0.22	90	0.1
4	0.57	0.73	360	0.4
5	1.00	0.09	45	0.05
6	0.33	0.11	45	0.05

43% *tall* people are *Swede*

67% *Swede* are *tall*

certain rules

Introduction – rule interestingness measures

- The number of rules generated from massive datasets can be very large and only a few of them are likely to be **useful**
- In all practical applications, like **medical practice, market basket**, it is crucial to know **how good the rules are**
- To measure the relevance and utility of rules, quantitative measures called **attractiveness** or **interestingness measures**, have been proposed

  (e.g. support, confidence, lift, gain, conviction, Piatetsky-Shapiro,…)
- **There is no evidence which measure(s) is (are) the best**

Basic quantitative characteristics of rules

- **Notation:**
  - $sup(\mathbf{0})$ **is the number of all objects from *U*, having property °**

    e.g. $sup(\phi)$ **,** $sup(\psi)$

  - **Support of rule $\phi \rightarrow \psi$ in *S*:**

$$sup(\phi \rightarrow \psi) = sup(\phi \wedge \psi)$$

  - **Anti-support of rule $\phi \rightarrow \psi$ in *S*:**

$$anti\text{-}sup(\phi \rightarrow \psi) = sup(\phi \wedge \neg\psi)$$

Basic quantitative characteristics of rules

- **Confidence of rule $\phi \rightarrow \psi$ in *S* (Łukasiewicz, 1913):**

$$conf(\varphi \rightarrow \psi) = \frac{sup(\varphi \wedge \psi)}{sup(\varphi)}$$

- **Coverage of rule $\phi \rightarrow \psi$ in *S*:**

$$cov(\varphi \rightarrow \psi) = \frac{sup(\varphi \wedge \psi)}{sup(\psi)}$$

Basic quantitative characteristics of rules

- **Why a new measure is required for decision rules in addition to strength, certainty, coverage, … ?**
- **Example** **(inspired by Popper, 1959)**
  Consider the possible result of rolling a dice: 1,2,3,4,5,6.

  ψ **="the result is 6"** $\neg\psi$ **="the result is *not* 6"**

  ϕ **="the result is an even number (i.e. 2 or 4 or 6)"**

  - $\phi \to \psi$, ***conf***$(\phi\to\psi)$ **= 1/3**
  - $\phi \to\neg\psi$, ***conf***$(\phi\to\neg\psi)$ **= 2/3**
- **Probability that the result is 6 is: 1/6**
  Probability that the result is *not* 6 is: 5/6
- **Information** ϕ **increases the probability of** ψ **from 1/6 to 1/3, and decreases the probability of** $\neg\psi$ **from 5/6 to 2/3**
- **In conclusion:** ϕ **confirms** ψ **and disconfirms** $\neg\psi$**,** **<u>independently</u> of the fact that** ***conf***$(\phi\to\psi)$ **<** ***conf***$(\phi\to\neg\psi)$

Bayesian confirmation measures

- Among widely studied interestingness measures, there is a group of *Bayesian confirmation measures*
- Measures of confirmation quantify the strength of confirmation that premise ϕ gives to conclusion ψ
- An attractiveness *c* measure has the property of confirmation (i.e. is a confirmation measure) if is satisfies the following condition:

$$c(\phi \rightarrow \psi)\begin{cases} > 0 \ \ if \ \ Pr(\psi|\phi) > Pr(\psi) \\ = 0 \ \ if \ \ Pr(\psi|\phi) = Pr(\psi) \\ < 0 \ \ if \ \ Pr(\psi|\phi) < Pr(\psi) \end{cases}$$

- Its meaning is different from a simple statistics of co-occurrence of properties ϕ and ψ in universe *U*

Bayesian confirmation measures

- ***c*(ϕ, ψ)>0** means that property ψ is satisfied **more frequently** when ϕ is satisfied (then, this frequency is *conf*(ϕ, ψ)), rather than generically in *S* (where the frequency is *Pr*(ψ)),

- ***c*(ϕ, ψ)=0** means that property ψ is satisfied **with the same frequency** whether ϕ is satisfied or not

- ***c*(ϕ, ψ)<0** means that property ψ is satisfied **less frequently** when ϕ is satisfied, rather than generically

Bayesian confirmation measures

- Under „the closed world assumption" adopted in inductive reasoning, and because ***U*** is a finite set, it is legitimate to estimate probabilities in the following way: $Pr(\psi) = \frac{sup(\psi)}{|U|}$

$$c(\phi \to \psi)\begin{cases} > 0 & if \;\; Pr(\psi|\phi) > Pr(\psi) \\ = 0 & if \;\; Pr(\psi|\phi) = Pr(\psi) \\ < 0 & if \;\; Pr(\psi|\phi) < Pr(\psi) \end{cases}$$

$$c(\phi \to \psi)\begin{cases} > 0 & if \;\; \frac{sup(\phi \to \psi)}{sup(\phi)} > \frac{sup(\psi)}{|U|} \\ = 0 & if \;\; \frac{sup(\phi \to \psi)}{sup(\phi)} = \frac{sup(\psi)}{|U|} \\ < 0 & if \;\; \frac{sup(\phi \to \psi)}{sup(\phi)} < \frac{sup(\psi)}{|U|} \end{cases}$$

Rival Bayesian confirmation measures

- **The condition**

$$c(\phi \to \psi)\begin{cases} > 0 \ \ if \ \dfrac{sup\,(\phi \to \psi)}{sup\,(\phi)} > \dfrac{sup\,(\psi)}{|U|} \\ = 0 \ \ if \ \dfrac{sup\,(\phi \to \psi)}{sup\,(\phi)} = \dfrac{sup\,(\psi)}{|U|} \\ < 0 \ \ if \ \dfrac{sup\,(\phi \to \psi)}{sup\,(\phi)} < \dfrac{sup\,(\psi)}{|U|} \end{cases}$$

does not put any constraint on the value to be assigned to confirmatory arguments (as long as they are positive) or disconfirmatory arguments (as long as they are negative)

- **There are many alternative, non-equivalent measures of Bayesian confirmation with different scales**

Rival Bayesian confirmation measures

- **Notation:** ***a***=$sup(\phi \to \psi)$, ***b***=$sup(\neg\phi \to \psi)$, ***c***=$sup\,(\phi \to \neg\psi)$, ***d***=$sup(\neg\phi \to \neg\psi)$
- **Among popular confirmation measures there are:**

$$D(\phi \to \psi) = \frac{a}{a+c} - \frac{(a+b)}{a+b+c+d}$$ **(Carnap 1950/1962)**

$$S(\phi \to \psi) = \frac{a}{a+c} - \frac{b}{b+d}$$ **(Christensen 1999)**

$$M(\phi \to \psi) = \frac{a}{a+b} - (a+c)$$ **(Mortimer 1988)**

$$N(\phi \to \psi) = \frac{a}{a+b} - \frac{c}{c+d}$$ **(Nozick 1981)**

$$C(\phi \to \psi) = \frac{a-(a+c)(a+b)}{a+b+c+d}$$ **(Carnap 1950/1962)**

$$R(\phi \to \psi) = \frac{a(a+b+c+d)}{(a+c)(a+b)} - 1$$ **(Finch 1960)**

Normative requirement for confirmation measures

- **Crupi, Tentori and Gonzalez (2007)** **have considered the confirmation measures from the viewpoint of classical deductive logic introducing function *v* such that for any argument (ϕ,ψ):**
  - ***v* assigns it the same positive value (e.g., 1) iff ϕ entails ψ, i.e. ϕ a ψ,**
  - **an equivalent value of opposite sign (e.g., -1) iff ϕ entails the negation of ψ, i.e. ϕ a $\neg\psi$, and**
  - **value 0, otherwise.**

Normative requirement for confirmation measures

- **The relationship between the logical implication or refutation of ψ by ϕ, and the conditional probability of ψ** subject to ϕ **requires that any Bayesian confirmation measure *c*($\phi\rightarrow\psi$) agrees with *v*(ϕ,ψ) in the following sense:**

  (Ex_1): $\;\; if \;\; v(\phi_1,\psi_1) > v(\phi_2,\psi_2), \; then \;\; c(\phi_1 \rightarrow \psi_1) > c(\phi_2 \rightarrow \psi_2).$

  Ex_1 guarantees that any conclusively confirmatory argument (ϕ a ψ) is assigned a higher value of *c*($\phi\rightarrow\psi$) than any argument which is *not* conclusively confirmatory,
  and any conclusively disconfirmatory argument (ϕ a $\neg\psi$) is assigned a lower value of *c*($\phi\rightarrow\psi$) than any argument which is *not* conclusively disconfirmatory

- **Crupi et al. have proved that neither of the above mentioned confirmation measures satisfies principle (Ex_1)**

- **However, their further analysis has unveiled a normalization approach that makes those measures fulfill this principle**

Proof that *D* measure is inconsistent with principle (Ex_1)

- **(Ex_1):** $if\ \ v(\phi_1,\psi_1) > v(\phi_2,\psi_2),\ then\ \ D(\phi_1 \to \psi_1) > D(\phi_2 \to \psi_2)$.
- ***D* measure:** $D(\phi \to \psi) = \dfrac{a}{a+c} - \dfrac{(a+b)}{a+b+c+d} = p(\psi \mid \phi) - p(\psi)$
- **Proof by counterexample:**
  - **Suppose that *x*, *y*, *z* are pairwise logically incompatible,**
    that $p(x)+p(y)+p(z)=1$ and that $p(z)>0.5$
    For instance, a fair dice is rolled:
    $x=1$; $y=2$; $z=$score greater than 2
  - **Let it be the case that:**
    $e_1=y$; $e_2=(x\vee y)$; $h_1=(y\vee z)$; $h_2=x$
  - **Notice that e_1 a h_1 (because having *y* entails having ($y\vee z$))**
    and e_2 Ø a h_2 (because having ($x\vee y$) does not necessarily entail *x*)
    and, therefore, $v(e_1, h_1)=1 > v(e_2, h_2)=0$

Proof that *D* measure is inconsistent with principle (Ex_1)

- x=1; y=2; z=score greater than 2; $e_1=y$; $e_2=(x \vee y)$; $h_1=(y \vee z)$; $h_2=x$

Continuation of proof:

- From definition, $D(e \rightarrow h)=p(h|e)-p(h)$
- Since $p(h_1|e_1)=1$ //as $p((y \vee z)|y)=1$//
  we obtain: $D(e_1 \rightarrow h_1)=p(h_1|e_1) - p(h_1)=1 - p(h_1)=p(\neg h_1)=p(x)$
- Moreover, $D(e_2 \rightarrow h_2)=p(h_2|e_2) - p(h_2)=p(x)/[p(x)+p(y)]-p(x)$
- It follows that $D(e_1 \rightarrow h_1)<D(e_2 \rightarrow h_2)$
  because $[p(x)+p(y)]=[p(1)+p(2)]=1/6+1/6=1/3 < 0.5$

Conclusions:

- $v(e_1, h_1) > v(e_2, h_2)$ while $D(e_1 ® h_1) < D(e_2 ® h_2)$
- *D* measure is inconsistent with principle (Ex_1)

Normalization of confirmation measures

Normalization of confirmation measures

- **The normalization approach distinguishes between two completely different situations:**
  - **situation α in which confirmation occurs:** $Pr(\psi \mid \phi) \geq Pr(\psi)$ **which, under „the closed world assumption", can be estimated as:**

$$\frac{sup(\phi \rightarrow \psi)}{sup(\phi)} \geq \frac{sup(\psi)}{|U|}$$

  - **situation β in which disconfirmation occurs:** $Pr(\psi \mid \phi) < Pr(\psi)$ **which, under „the closed world assumption", can be estimated as:**

$$\frac{sup(\phi \rightarrow \psi)}{sup(\phi)} < \frac{sup(\psi)}{|U|}$$

Normalization of confirmation measures

- **The normalization is made by dividing the confirmation measures by:**
  - **the maximum possible value in case of confirmation, and**
  - **the absolute minimum possible value in case of disconfirmation.**

- **„Bayesian" approach to** determining **the maximum** a**nd minimum values:**
  - **the evidence confirms the hypothesis, if the hypothesis is more frequent with the evidence rather than with ¬evidence, and**
  - **the evidence disconfirms the hypothesis, if ¬hypothesis is more frequent with the evidence rather than with ¬evidence.**

„Bayesian" approach to normalization

- E.g. consider rule $\phi \rightarrow \psi$: ***if x is a raven then x is black***

  a=$sup(\phi \rightarrow \psi)$ – the number of objects in ***U*** which are black ravens

  b=$sup(\neg\phi \rightarrow \psi)$ – the no. of objects in ***U*** which are black non-ravens

  c=$sup\ (\phi \rightarrow \neg\psi)$ – the no. of objects in ***U*** which are non-black ravens

  d=$sup(\neg\phi \rightarrow \neg\psi)$ – the no. of objects in ***U*** which are non-black non-ravens

„Bayesian" approach to normalization

- The **maximal support of the rule** will be obtained when:
  - *a* takes over all observations from *b*
    (i.e. only ravens are black, $a := a + b$)
  - $b := 0$ (i.e. there are no black non-ravens)
  - $c := 0$ (i.e. there are no non-black ravens)
  - *d* takes over all observations from *c*
    (i.e. each non-raven is not black, $d := c + d$)

$a = sup(\phi \rightarrow \psi)$

$b = sup(\neg\phi \rightarrow \psi)$

$c = sup(\phi \rightarrow \neg\psi)$

$d = sup(\neg\phi \rightarrow \neg\psi)$

„Bayesian" approach to normalization

- The minimal support of the rule will be obtained when:
  - $a := 0$ (i.e. there are no black ravens),
  - b takes over all observations from a
    (i.e. each non-raven is black, $b := a + b$)
  - c takes over all observations from d
    (i.e. each raven is not black, $c := c + d$)
  - $d := 0$ (i.e. there are no non-black non-ravens)

$a = sup(\phi \rightarrow \psi)$

$b = sup(\neg\phi \rightarrow \psi)$

$c = sup\ (\phi \rightarrow \neg\psi)$

$d = sup(\neg\phi \rightarrow \neg\psi)$

Normalized confirmation measures

- **Notation:** ***a*=**$sup(\phi \rightarrow \psi)$ **,** ***b*=**$sup(\neg\phi \rightarrow \psi)$**,** ***c*=**$sup\ (\phi \rightarrow \neg\psi)$**,** ***d*=**$sup(\neg\phi \rightarrow \neg\psi)$
- **Among confirmation measures that were normalized and analysed by Crupi et al. there are:**

$$D(\phi \rightarrow \psi) = \frac{a}{a+c} - \frac{(a+b)}{|U|}$$ **(Carnap 1950/1962)**

$$S(\phi \rightarrow \psi) = \frac{a}{a+c} - \frac{b}{b+d}$$ **(Christensen 1999)**

$$M(\phi \rightarrow \psi) = \frac{a}{a+b} - (a+c)$$ **(Mortimer 1988)**

$$N(\phi \rightarrow \psi) = \frac{a}{a+b} - \frac{c}{c+d}$$ **(Nozick 1981)**

$$C(\phi \rightarrow \psi) = \frac{a-(a+c)(a+b)}{a+b+c+d}$$ **(Carnap 1950/1962)**

$$R(\phi \rightarrow \psi) = \frac{a(a+b+c+d)}{(a+c)(a+b)} - 1$$ **(Finch 1960)**

Normalized confirmation measures

- **The confirmation measures are normalized according to „Bayesian" approach:**

$$D(\phi \to \psi) = \frac{a}{a+c} - \frac{(a+b)}{a+b+c+d}$$

$$D_{norm}(\phi \to \psi) = \begin{cases} \dfrac{\dfrac{a}{a+c} - \dfrac{(a+b)}{a+b+c+d}}{\dfrac{c+d}{a+b+c+d}} = \dfrac{ad-bc}{(a+c)(c+d)} \\ \text{in case of confirmation} \\ \\ \dfrac{\dfrac{a}{a+c} - \dfrac{(a+b)}{a+b+c+d}}{\dfrac{a+b}{a+b+c+d}} = \dfrac{ad-bc}{(a+c)(a+b)} \\ \text{in case of disconfirmation} \end{cases}$$

Z-measure

- **It can be observed that:**

$$D_{norm} = S_{norm} = M_{norm} = N_{norm} = C_{norm} = R_{norm} = G_{norm}.$$

- **Crupi et al. have therefore proposed to call them all by one name:**

_Z_-measure

$$Z(\varphi \to \psi) = \begin{cases} \dfrac{ad - bc}{(a+c)(c+d)} \\ \text{in case of confirmation} \\ \\ \dfrac{ad - bc}{(a+c)(a+b)} \\ \text{in case of disconfirmation} \end{cases}$$

Z-measure

- It has been proved that *Z*, and all confirmation measures equivalent to it, satisfy principle (Ex_1).

- Thus, *Z* is surely a valuable tool for measuring the confirmation of decision or association rules induced from datasets.

Symmetry properties of confirmation measures

Properties of symmetry

- **Properties of symmetry (Carnap 1962, Eells & Fitelson 2000):**
  - **Evidence symmetry (ES):** $c(\varphi \rightarrow \psi) = -c(\neg\varphi \rightarrow \psi)$
  - **Inversion symmetry (IS):** $c(\varphi \rightarrow \psi) = c(\psi \rightarrow \varphi)$
  - **Hypothesis symmetry (HS):** $c(\varphi \rightarrow \psi) = -c(\varphi \rightarrow \neg\psi)$
  - **Total symmetry (TS):** $c(\varphi \rightarrow \psi) = -c(\neg\varphi \rightarrow \neg\psi)$

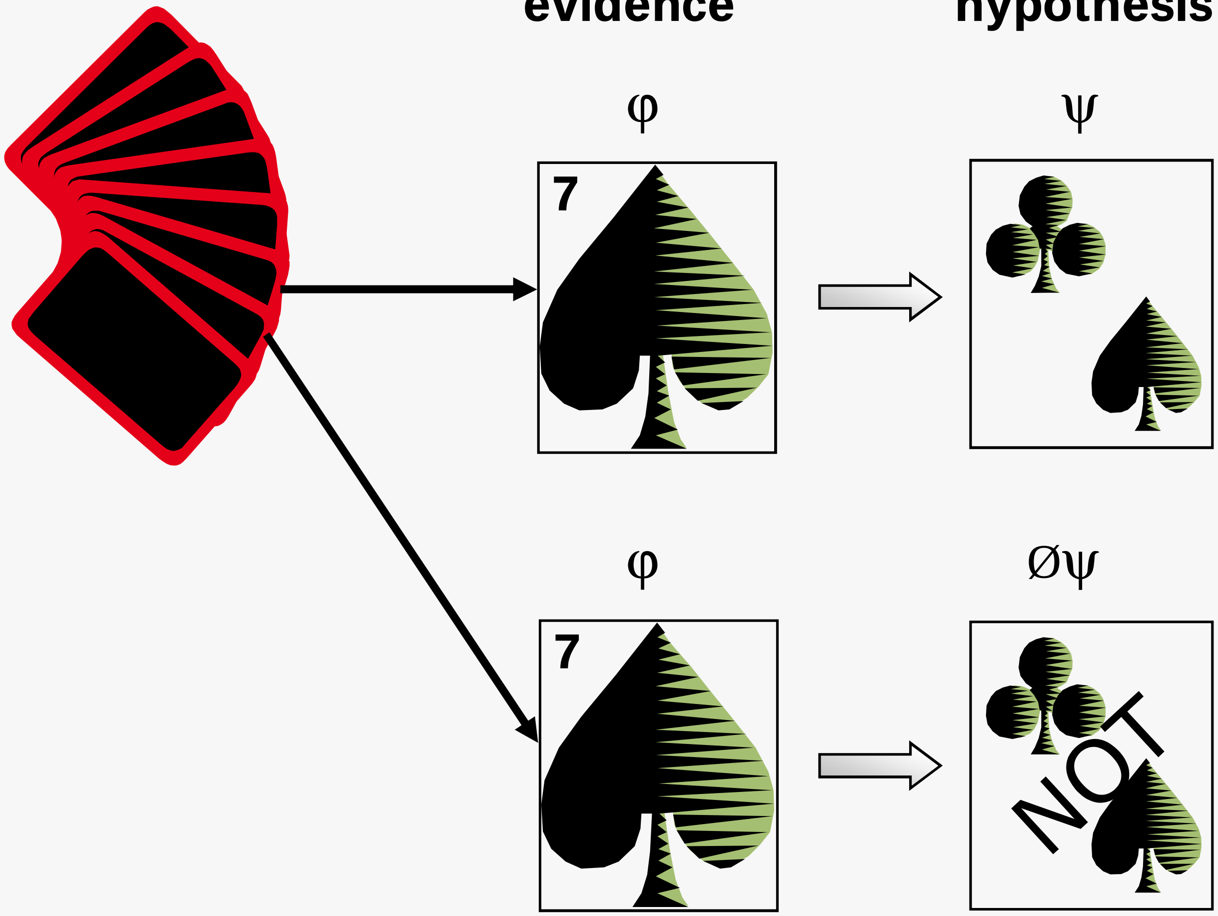
- **Given IS, ES $\equiv$ HS, and ES$\wedge$HS $\Rightarrow$ TS**
- **Only hypothesis symmetry (HS) is desirable**

  HS: the impact of φ on ψ should be of the same strength, but of the opposite sign, as the impact of φ on $\neg\psi$

Bayesian confirmation measures – *Hypothesis Symmetry* (HS)

evidence **hypothesis (*„the card is black"*)**

φ ψ

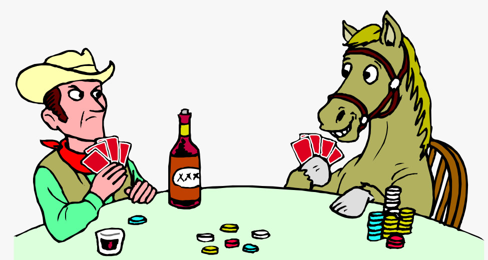


φ is conclusive for ψ

φ Øψ **(*„the card is not black"*)**

φ is negatively conclusive for Øψ

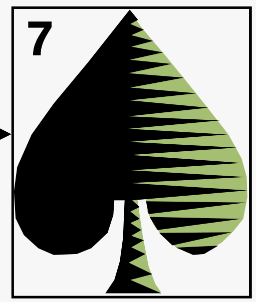
$$c(\varphi \rightarrow \psi) = -c(\varphi \rightarrow \neg\psi)$$

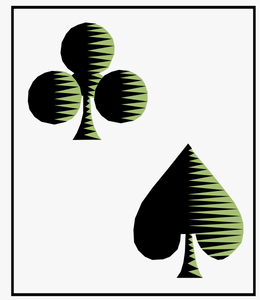


Bayesian confirmation measures – *Evidence Symmetry* (ES)

evidence **hypothesis („*the card is black*")**

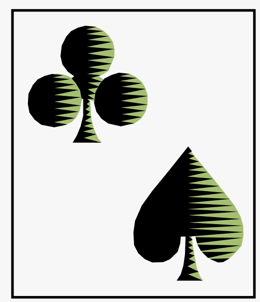
φ ψ



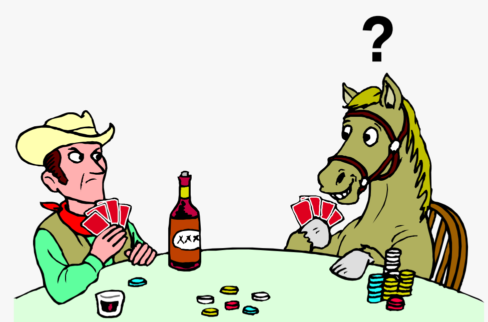


φ is conclusive for ψ

Øφ ψ



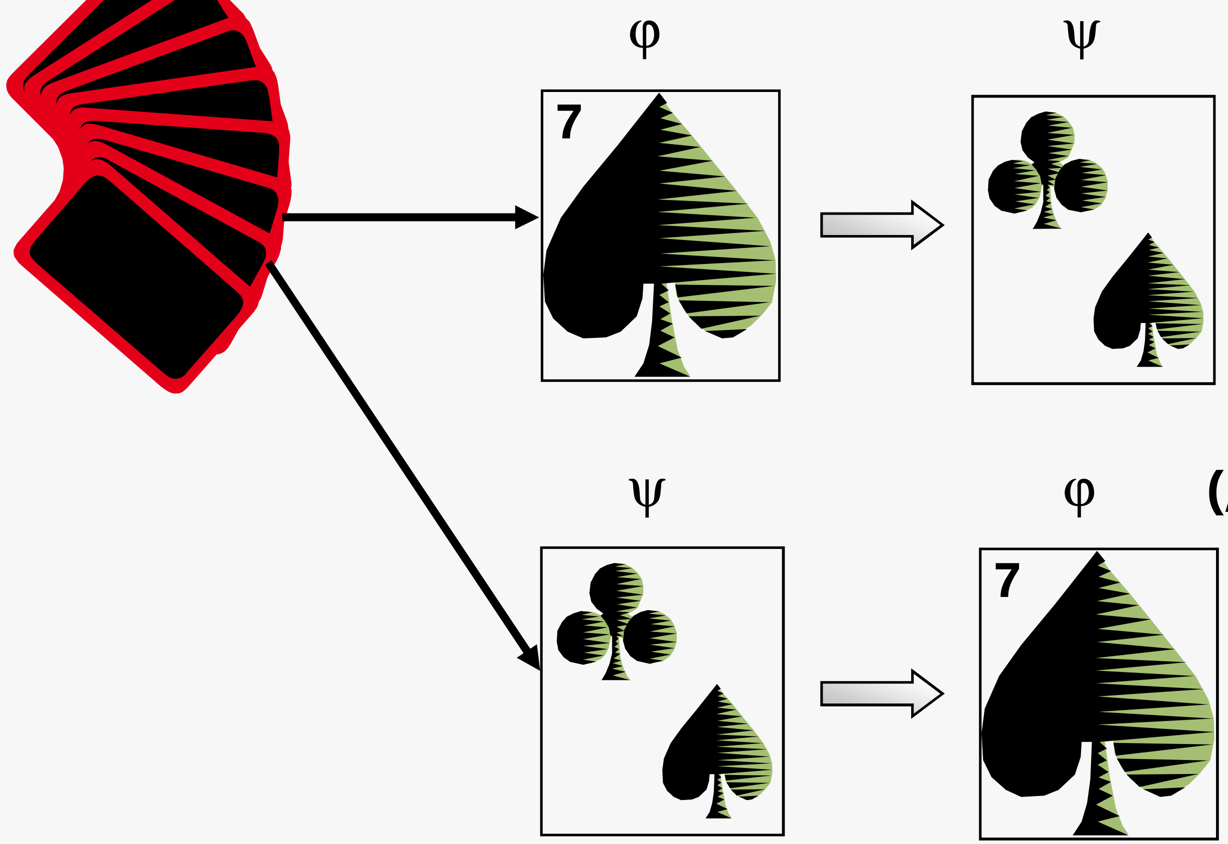
Øφ is useless for ψ



$$c(\varphi \rightarrow \psi) = -c(\neg\varphi \rightarrow \psi)$$

Bayesian confirmation measures – *Inversion Symmetry* (IS)

evidence **hypothesis (*„the card is black"*)**

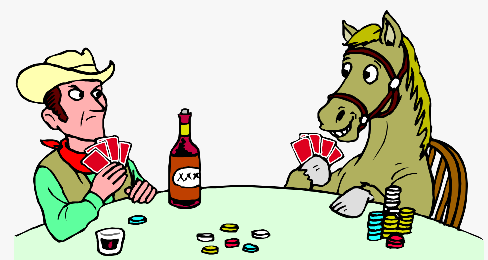


φ is conclusive for ψ

(*„the card is 7 of spades"*)

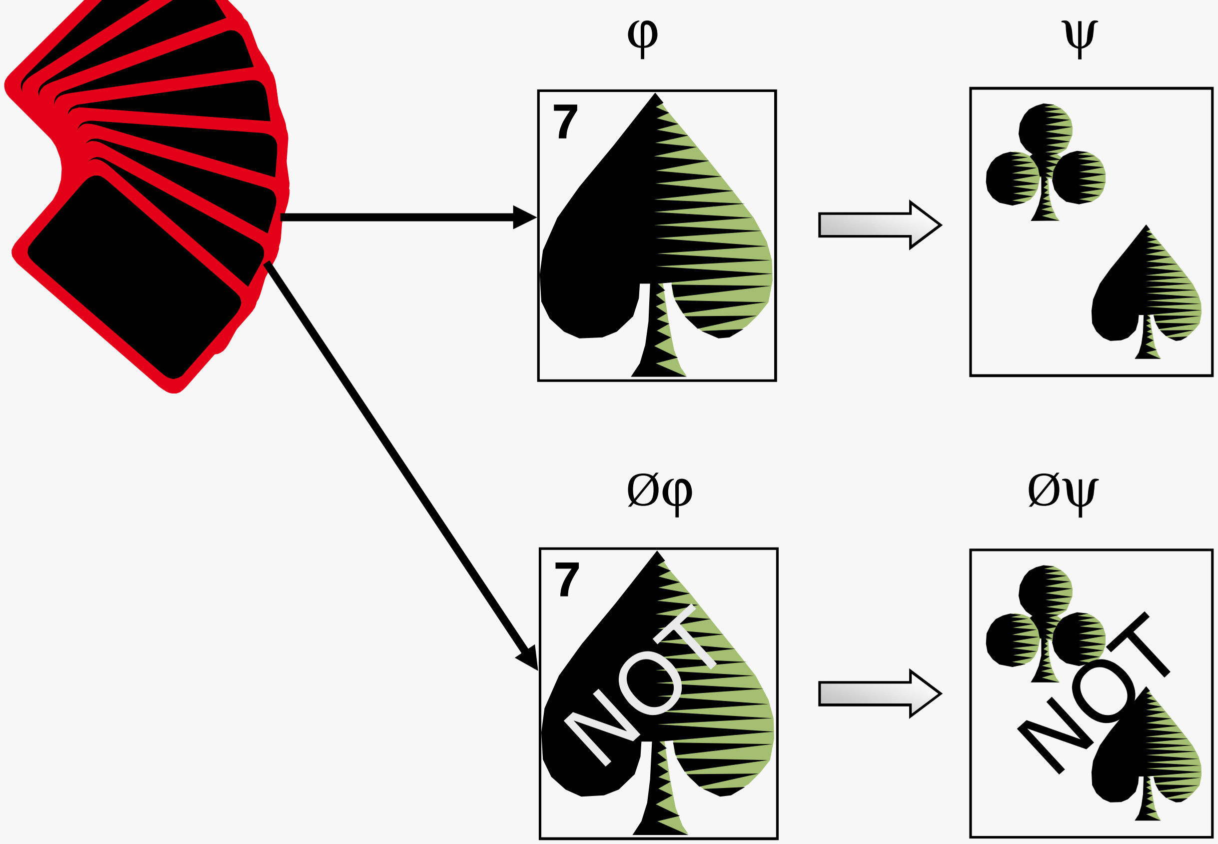
ψ is less useful for φ than *vice versa*

$$c(\varphi \rightarrow \psi) = c(\psi \rightarrow \varphi)$$



Bayesian confirmation measures – *Total Symmetry* (TS)

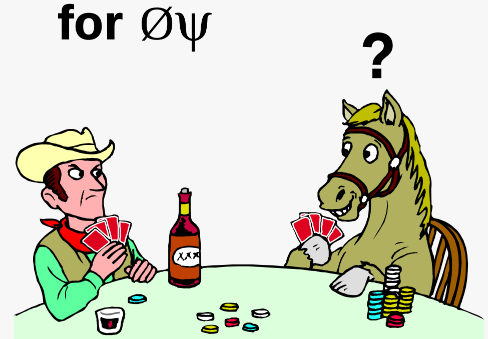
evidence **hypothesis (*„the card is black"*)**



φ is conclusive for ψ

Øφ is not conclusive for Øψ

$$c(\varphi \rightarrow \psi) = -c(\neg\varphi \rightarrow \neg\psi)$$



Property of hypothesis symmetry

- **Property of hypothesis symmetry (HS)** **(Carnap `62, Eells, Fitelson '02)**
- **An interestingness measure $c(\phi \to \psi)$ has the property HS if**

$$c(\phi \to \psi) = -c(\phi \to \neg\psi)$$

- **Interpretation: the impact of ϕ on ψ should be of the same strength, but of the opposite sign as the impact of ϕ on $\neg\psi$**
- **Example: Let us consider a rule $\phi \to \psi$:**

if x is $\boxed{}$ ***then x is*** $\boxed{}$

ϕ is conclusive for ψ and negatively conclusive for $\neg\psi$

Additional properties of symmetry

- **Properties of symmetry (Crupi et al. 2007):**
  - **It is reasonable to combine Inversion with Evidence, Hypothesis or Total Symmetry**
  - **EIS:** $c(\phi \rightarrow \psi) = -c(\psi \rightarrow \neg\phi)$
  - **HIS:** $c(\phi \rightarrow \psi) = -c(\neg\psi \rightarrow \phi)$
  - **TIS:** $c(\phi \rightarrow \psi) = c(\neg\psi \rightarrow \neg\phi)$
- **Claim of Crupi et al.: The symmetries should be considered separately for the situation of confirmation and disconfirmation**

Properties of symmetry

	symmetry	in case of confirmation	in case of disconfirmation
Eels & Fitelson	ES: $c(\phi\rightarrow\psi)=-c(\neg\phi\rightarrow\psi)$	**no**	**no**
Eels & Fitelson	HS: $(\phi\rightarrow\psi)=-c(\phi\rightarrow\neg\psi)$	**yes**	**yes**
	EIS: $c(\phi\rightarrow\psi)=-c(\psi\rightarrow\neg\phi)$	**no**	**yes**
	HIS: $c(\phi\rightarrow\psi)=-c(\neg\psi\rightarrow\phi)$	**yes**	**no**
Eels & Fitelson	IS: $c(\phi\rightarrow\psi)=c(\psi\rightarrow\phi)$	**no**	**yes**
Eels & Fitelson	TS: $c(\phi\rightarrow\psi)=c(\neg\phi\rightarrow\neg\psi)$	**no**	**no**
	TIS: $c(\phi\rightarrow\psi)=c(\neg\psi\rightarrow\neg\phi)$	**yes**	**no**

c(Jack→face) ≠ c(face→Jack)

c(7→face) = c(face→7)

Properties of symmetry

- A measure has symmetry properties as in the table if it possesses:
  - **Hypothesis symmetry (HS)** both in case of confirmation and disconfirmation, and
  - **Inverse symmetry (IS)** in case of disconfirmation but NOT in case of confirmation

- ***Z*-measure satisfies the above properties as required** and, therefore, is a good measure from the point of view of symmetry properties

Other desired properties of confirmation measures

Property M

- **Property M** of monotonicity*
- An interestingness measure $F(a, b, c, d)$ has the property M if it is a function non-decreasing with respect to a and d and non-increasing with respect to b and c

  where:

  $a=sup(\phi\rightarrow\psi)$ - the no. of objects in U for which ϕ and ψ hold together

  $b=sup(\neg\phi\rightarrow\psi)$

  $c=sup(\phi\rightarrow\neg\psi)$

  $d=sup(\neg\phi\rightarrow\neg\psi)$

* Greco, S., Pawlak, Z., Słowiński, R.: „Can Bayesian confirmation measures be useful for rough set decision rules?", *Engineering Applications of Artificial Intelligence*, 17 (2004):345–361

Interpretation of the property M

- **E.g. consider rule $\phi \rightarrow \psi$:**

  if x is a raven then x is black

  - **non-decreasing with respect to *a* –**

    the more *black ravens* we observe,

    the more credible becomes the rule

  - **non-increasing with respect to *b* –**

    the more *black non-ravens* we observe,

    the less credible becomes the rule

  - **non-increasing with respect to *c***

  - **non-decreasing with respect to *d***

Results of the conducted analysis with respect to property M

- ***Theorem*:**

  The new *Z*-measure has the property M

- ***Proof outline:***
  we have proved that the *Z*-measure
  both in case of confirmation and disconfirmation is:
  - ***non-decreasing* with respect to *a***
  - ***non-increasing* with respect to *b***
  - ***non-increasing* with respect to *c***
  - ***non-decreasing* with respect to *d***

Practical application of the results

Support – Anti-support Pareto border

- ***Theorem*** ***:**

  For a set of rules with the same conclusion,

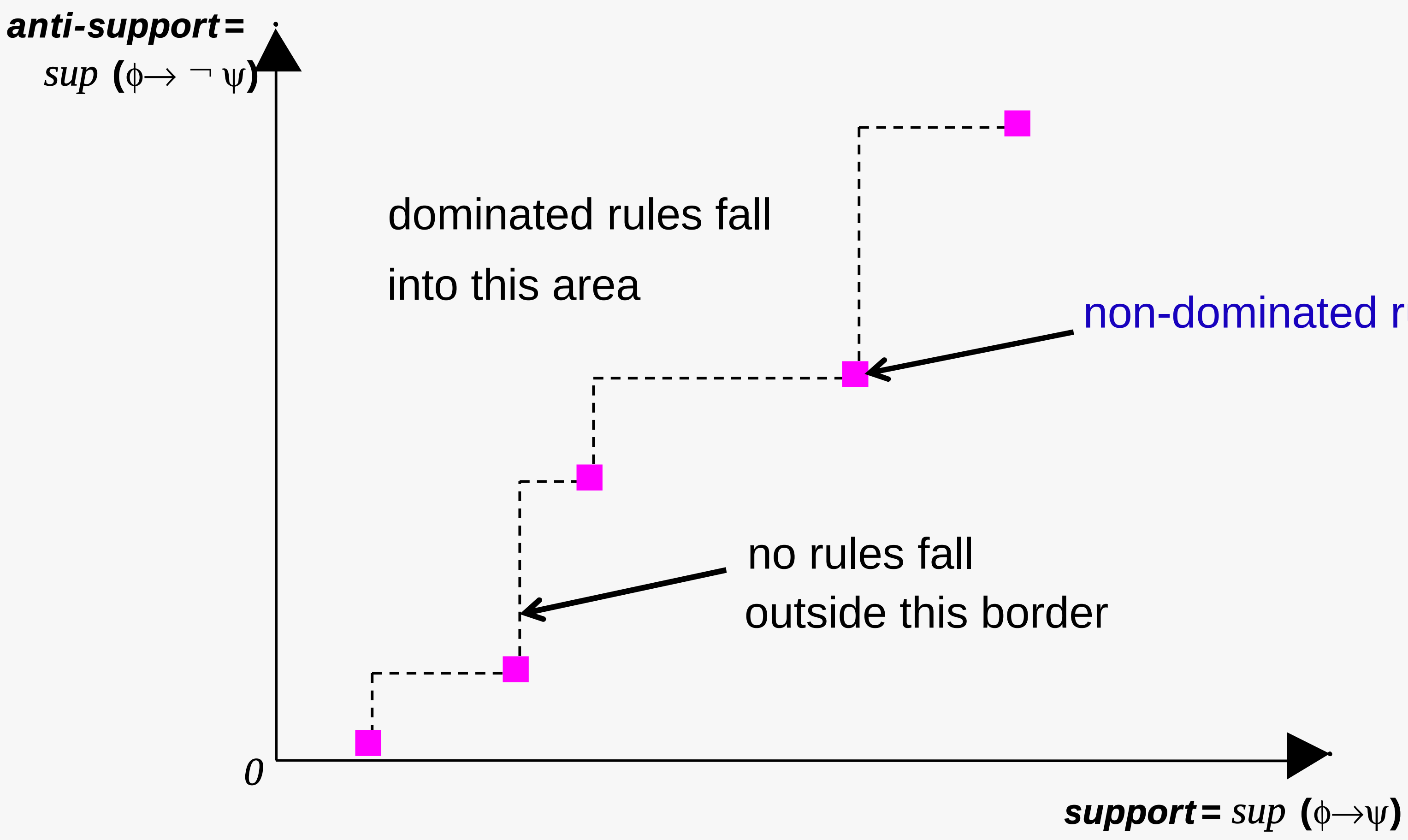
  due to (anti) monotonic dependencies between measures of support and anti-support on one hand and any interestingness measure with property M on the other hand

  the best rules according to any measure with the property M must reside on the support – anti-support Pareto optimal border

- **The support – anti-support Pareto border is a set of non-dominated rules with respect to support and anti-support**

* Brzezińska I., Greco S., Słowiński R.: "Mining Pareto-Optimal Rules with Respect to Support and Confirmation or Support and Anti-Support", *Engineering Applications of Artificial Intelligence*, 20 (2007): 587-600

Support – Anti-support Pareto border



The best rules according to any measure with the property M must reside on the support – anti-support Pareto border

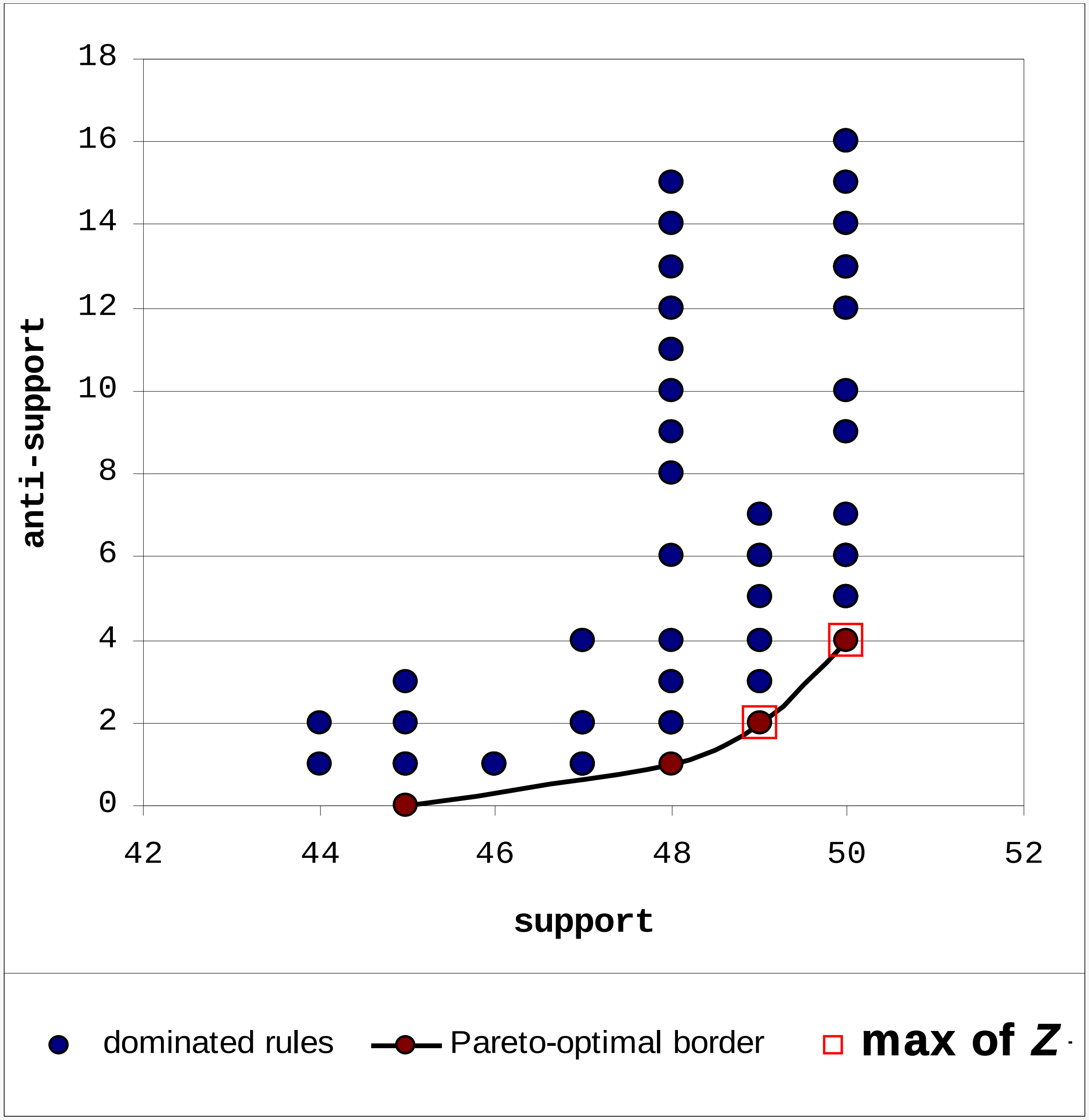
Practical application of the results

- **Since the *Z*-measure satisfies the property M**
  we can conclude that
  rules optimal with respect to *Z* will be found in the set of
  Pareto-optimal rules wrt support and anti-support
  (considering rules with the same conclusion)

Practical application of the results

- Possession of property M implies potential efficiency improvement:
  - one can concentrate on mining only the support–anti-support Pareto set instead of conducting rule evaluation separately wrt Z-measure or any other measure with property M
  - rules optimal wrt Z-measure or any other measure with property M can be mined from the support–anti-support Pareto set instead of searching the set of all rules
  - due to relationship between anti-support and any measure with property M, the rule order wrt anti-support (for fixed value of support) is the same for any other measure with M

Practical application of the results



Dataset: ***busses*** **about technical state of buses**

Set of 85 rules with the same conclusion

The rule order wrt anti-support (for fixed value of support) is the same for any measure with M

Confirmation perspective on support – anti-support border

- **Is there a curve separating rules with negative value of any confirmation measure in the support–anti-support space?**
- ***Theorem**:**

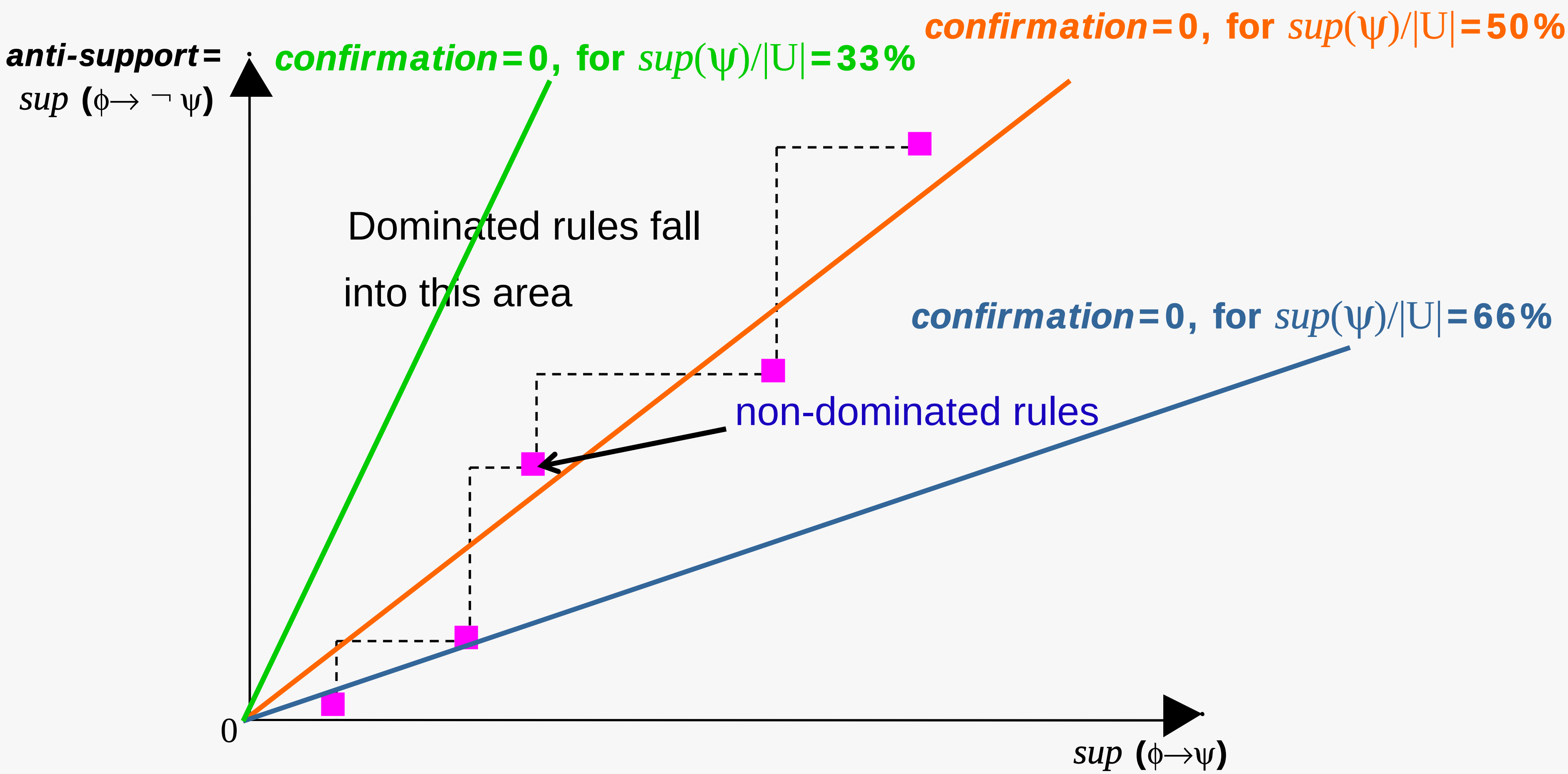
Rules lying above the linear function:

$$sup(\phi \to \psi)[|U|/sup(\psi) - 1]$$

have a negative value of any confirmation measure

For those rules, the premise only disconfirms the conclusion!

Confirmation perspective on support - anti-support border

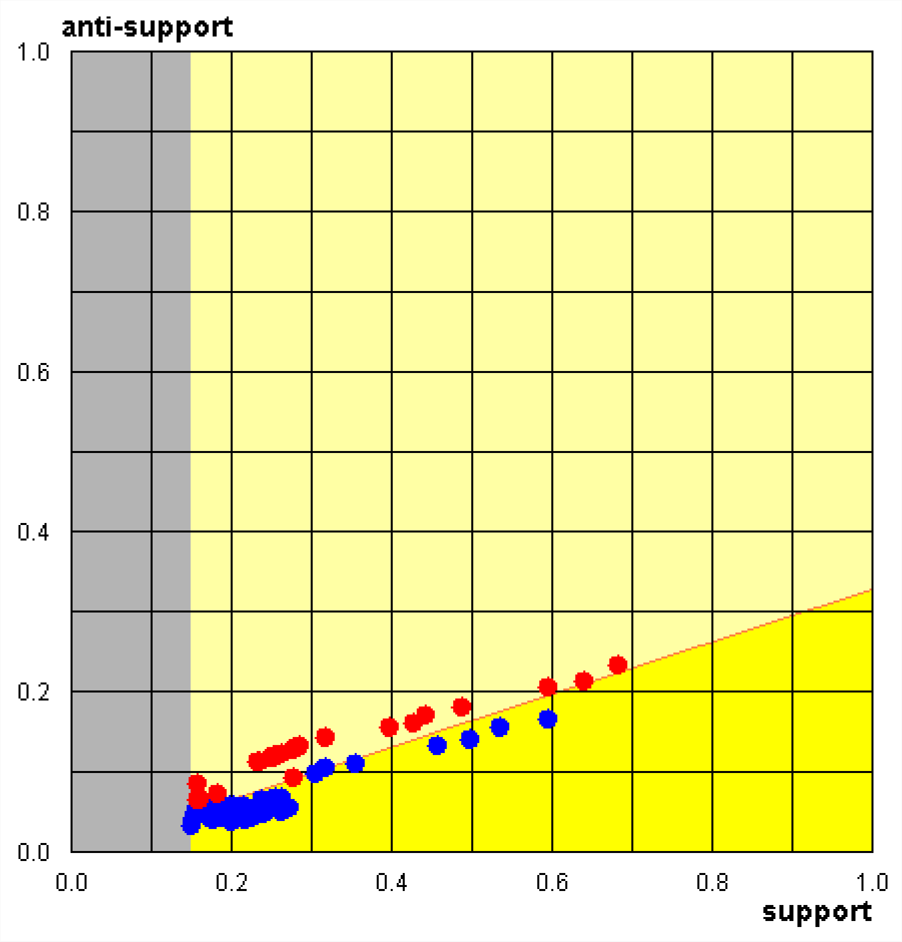


For rules lying above the curve for which *confirmation*=0 the premise only disconfirms the conclusion

Computational experiment: general info about the dataset

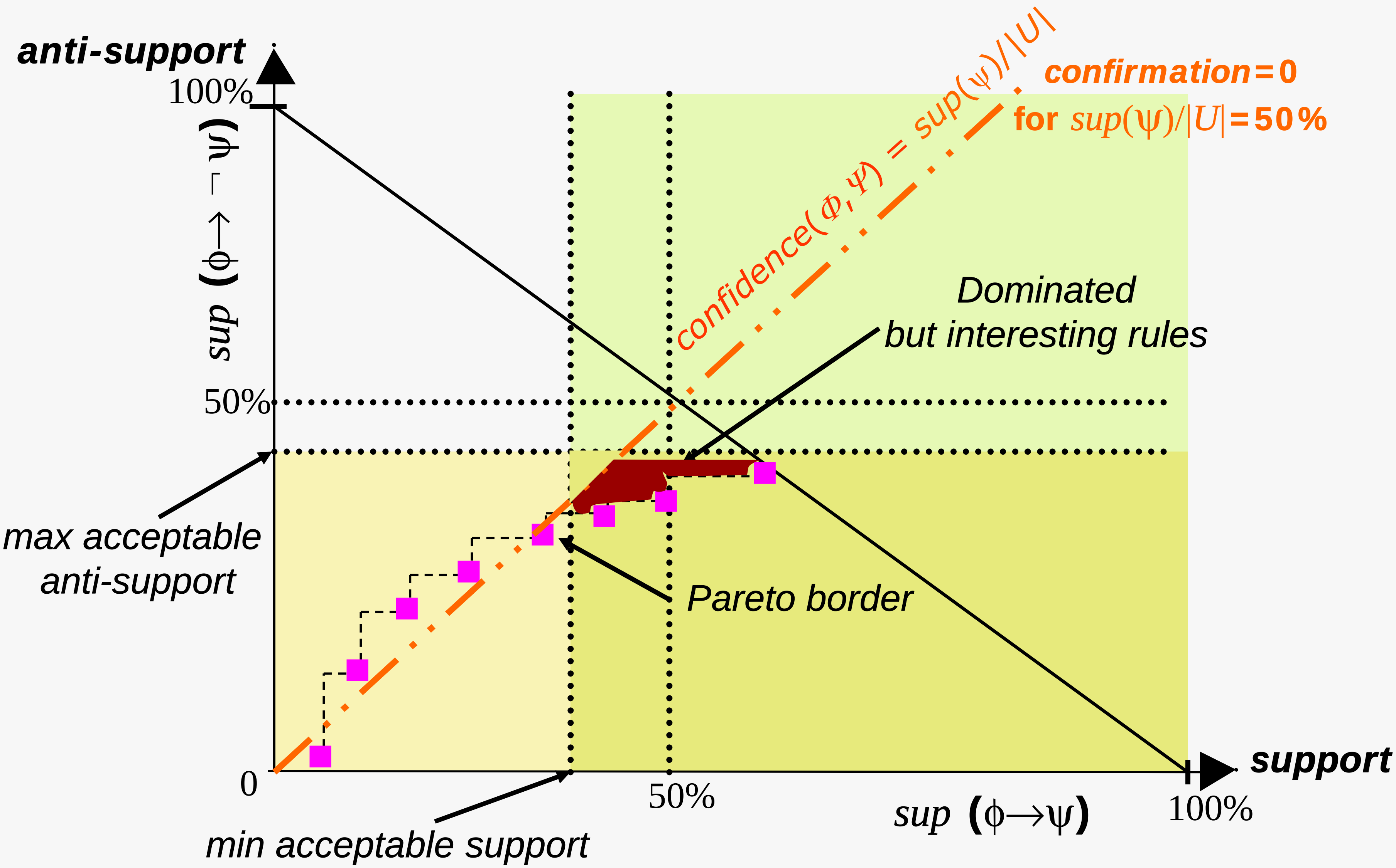
- Dataset ***adult***, created in ′96 by B. Becker & R. Kohavi from census database
- 32 561 instances
- 9 nominal attributes
  - workclass: Private, Local-gov, etc.;
  - education: Bachelors, Some-college, etc.;
  - marital-status: Married, Divorced, Never-married, et.;
  - occupation: Tech-support, Craft-repair, etc.;
  - relationship: Wife, Own-child, Husband, etc.;
  - race: White, Asian-Pac-Islander, etc.;
  - sex: Female, Male;
  - native-country: United-States, Cambodia, England, etc.;
  - salary: >50K, <=50K
- throughout the experiment, *sup*($\phi \rightarrow \psi$) is denotes relative rule support [0,1]

Support - anti-support (workclass=Private)



- ● indicates rules with negative confirmation
- even some rules from the Pareto border need to be discarded

Confirmation perspective on support – anti-support border



Conclusions

Conclusions

- Normalization of rival Bayesian confirmation measures unifies them into one Z-measure
- Z-measure has all the desired symmetry properties
- Z-measure has property M of monotonicity
- Possession of property M shows relationship between Z-measure and measures of support and anti-support

  and allows potential efficiency gains while searching for rules optimal with respect to Z-measure

Thank you!