

The background of the slide is a complex, abstract composition. It features a dark, reddish-brown base with a network of thin, light-colored lines forming a web-like structure. Scattered throughout are small, colorful dots in shades of green, blue, and orange. A prominent white banner with a subtle geometric pattern runs horizontally across the center, containing the title text. To the left of the banner, there is a smaller, rectangular inset image showing a different pattern of dots and lines. The overall aesthetic is technical and modern.

# Null Invariance Measures

# Interestingness Measures & Null-Invariance

- ❑ **Null invariance**: Value does not change with the # of null-transactions
- ❑ A few interestingness measures: Some are null invariant

Measure	Definition	Range	Null-Invariant
$\chi^2(A, B)$	$\sum_{i,j=0,1} \frac{(e(a_i b_j) - o(a_i b_j))^2}{e(a_i b_j)}$	$[0, \infty]$	No
$Lift(A, B)$	$\frac{s(A \cup B)}{s(A) \times s(B)}$	$[0, \infty]$	No
$AllConf(A, B)$	$\frac{s(A \cup B)}{\max\{s(A), s(B)\}}$	$[0, 1]$	Yes
$Jaccard(A, B)$	$\frac{s(A \cup B)}{s(A) + s(B) - s(A \cup B)}$	$[0, 1]$	Yes
$Cosine(A, B)$	$\frac{s(A \cup B)}{\sqrt{s(A) \times s(B)}}$	$[0, 1]$	Yes
$Kulczynski(A, B)$	$\frac{1}{2} \left( \frac{s(A \cup B)}{s(A)} + \frac{s(A \cup B)}{s(B)} \right)$	$[0, 1]$	Yes
$MaxConf(A, B)$	$\max\left\{ \frac{s(A)}{s(A \cup B)}, \frac{s(B)}{s(A \cup B)} \right\}$	$[0, 1]$	Yes

*$\chi^2$  and lift are not null-invariant*

*Jaccard, cosine, AllConf, MaxConf, and Kulczynski are null-invariant measures*

# Null Invariance: An Important Property

□ Why is null invariance crucial for the analysis of massive transaction data?

□ Many transactions may contain neither milk nor coffee!

milk vs. coffee contingency table

	<i>milk</i>	$\neg milk$	$\Sigma_{row}$
<i>coffee</i>	<i>mc</i>	$\neg mc$	<i>c</i>
$\neg coffee$	<i>m</i> $\neg c$	$\neg m$ $\neg c$	$\neg c$
$\Sigma_{col}$	<i>m</i>	$\neg m$	$\Sigma$

□ Lift and  $\chi^2$  are not null-invariant: not good to evaluate data that contain too many or too few null transactions!

□ Many measures are not null-invariant!

Null-transactions  
w.r.t. *m* and *c*

Data set	<i>mc</i>	$\neg mc$	<i>m</i> $\neg c$	$\neg m$ $\neg c$	$\chi^2$	<i>Lift</i>
<i>D</i> <sub>1</sub>	10,000	1,000	1,000	100,000	90557	9.26
<i>D</i> <sub>2</sub>	10,000	1,000	1,000	100	0	1
<i>D</i> <sub>3</sub>	100	1,000	1,000	100,000	670	8.44
<i>D</i> <sub>4</sub>	1,000	1,000	1,000	100,000	24740	25.75
<i>D</i> <sub>5</sub>	1,000	100	10,000	100,000	8173	9.18
<i>D</i> <sub>6</sub>	1,000	10	100,000	100,000	965	1.97