



Interestingness Measures: Lift and χ^2

Interestingness Measure: Lift

- Measure of dependent/correlated events: **lift**

$$lift(B, C) = \frac{c(B \rightarrow C)}{s(C)} = \frac{s(B \cup C)}{s(B) \times s(C)}$$

Lift is more telling than s & c

	B	¬ B	Σ _{row}
C	400	350	750
¬ C	200	50	250
Σ _{col.}	600	400	1000

- Lift(B, C) may tell how B and C are correlated

- Lift(B, C) = 1: B and C are independent
- > 1: positively correlated
- < 1: negatively correlated

- For our example, $lift(B, C) = \frac{400/1000}{600/1000 \times 750/1000} = 0.89$
 $lift(B, \neg C) = \frac{200/1000}{600/1000 \times 250/1000} = 1.33$

- Thus, B and C are negatively correlated since $lift(B, C) < 1$;
 B and ¬C are positively correlated since $lift(B, \neg C) > 1$

Interestingness Measure: χ^2

- Another measure to test correlated events: χ^2

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

- General rules

- $\chi^2 = 0$: independent
- $\chi^2 > 0$: correlated, either positive or negative, so it needs additional test

Now, $\chi^2 = \frac{(400 - 450)^2}{450} + \frac{(350 - 300)^2}{300} + \frac{(200 - 150)^2}{150} + \frac{(50 - 100)^2}{100} = 55.56$

- χ^2 shows B and C are negatively correlated since the expected value is 450 but the observed is only 400
- χ^2 is also more telling than the support-confidence framework

distribution, $\frac{C}{\chi^2} = \frac{250}{250} = 1 \therefore \frac{B,C}{B \cdot C} = 3$
- 2745a Ua

	B	$\neg B$	Σ_{row}
C	400 (450)	350 (300)	750
$\neg C$	200 (150)	50 (100)	250
Σ_{col}	600	400	1000


Expected value

Observed value

Lift and χ^2 : Are They Always Good Measures?

- ❑ Null transactions: Transactions that contain neither B nor C
- ❑ Let's examine the dataset D
 - ❑ BC (100) is much rarer than B¬C (1000) and ¬BC (1000), but there are many ¬B¬C (100000)
 - ❑ Unlikely B & C will happen together!
- ❑ But, $\text{Lift}(B, C) = 8.44 \gg 1$ (Lift shows B and C are strongly positively correlated!)
- ❑ $\chi^2 = 670$: Observed(BC) \gg expected value (11.85)
- ❑ *Too many null transactions may "spoil the soup"!*

	B	¬ B	Σ_{row}
C	100	1000	1100
¬ C	1000	100000	101000
$\Sigma_{\text{col.}}$	1100	101000	102100

 **null transactions**

Contingency table with expected values added

	B	¬ B	Σ_{row}
C	100 (11.85)	1000	1100
¬ C	1000 (988.15)	100000	101000
$\Sigma_{\text{col.}}$	1100	101000	102100