Data Mining and Analysis: Fundamental Concepts and Algorithms

datamininingbook.info

Mohammed J. Zaki[1] Wagner Meira Jr.[2]

[1]Department of Computer Science
Rensselaer Polytechnic Institute, Troy, NY, USA

[2]Department of Computer Science
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Chap. 12: Pattern and Rule Assessment

Rule Assessment Measures: Support and Confidence

Support: The *support* of the rule is defined as the number of transactions that contain both X and Y, that is,

$$sup(X \longrightarrow Y) = sup(XY) = |\mathbf{t}(XY)|$$

The *relative support* is the fraction of transactions that contain both X and Y, that is, the empirical joint probability of the items comprising the rule

$$rsup(X \longrightarrow Y) = P(XY) = rsup(XY) = \frac{sup(XY)}{|\mathbf{D}|}$$

Confidence: The *confidence* of a rule is the conditional probability that a transaction contains the consequent Y given that it contains the antecedent X:

$$conf(X \longrightarrow Y) = P(Y|X) = \frac{P(XY)}{P(X)} = \frac{rsup(XY)}{rsup(X)} = \frac{sup(XY)}{sup(X)}$$

Example Dataset: Support and Confidence

Tid	Items
1	*ABDE*
2	*BCE*
3	*ABDE*
4	*ABCE*
5	*ABCDE*
6	*BCD*

Frequent itemsets: $minsup = 3$

sup	*rsup*	Itemsets
3	0.5	*ABD*, *ABDE*, *AD*, *ADE* *BCE*, *BDE*, *CE*, *DE*
4	0.67	*A*, *C*, *D*, *AB*, *ABE*, *AE*, *BC*, *BD*
5	0.83	*E*, *BE*
6	1.0	*B*

Rule confidence

Rule	*conf*
$A \longrightarrow E$	1.00
$E \longrightarrow A$	0.80
$B \longrightarrow E$	0.83
$E \longrightarrow B$	1.00
$E \longrightarrow BC$	0.60
$BC \longrightarrow E$	0.75

Rule Assessment Measures: Lift, Leverage and Jaccard

Lift: Lift is defined as the ratio of the observed joint probability of X and Y to the expected joint probability if they were statistically independent, that is,

$$lift(X \longrightarrow Y) = \frac{P(XY)}{P(X) \cdot P(Y)} = \frac{rsup(XY)}{rsup(X) \cdot rsup(Y)} = \frac{conf(X \longrightarrow Y)}{rsup(Y)}$$

Leverage: Leverage measures the difference between the observed and expected joint probability of XY assuming that X and Y are independent

$$leverage(X \longrightarrow Y) = P(XY) - P(X) \cdot P(Y) = rsup(XY) - rsup(X) \cdot rsup(Y)$$

Jaccard: The Jaccard coefficient measures the similarity between two sets. When applied as a rule assessment measure it computes the similarity between the tidsets of X and Y:

$$\begin{aligned} jaccard(X \longrightarrow Y) &= \frac{|\mathbf{t}(X) \cap \mathbf{t}(Y)|}{|\mathbf{t}(X) \cup \mathbf{t}(Y)|} \\ &= \frac{P(XY)}{P(X) + P(Y) - P(XY)} \end{aligned}$$

Lift, Leverage, Jaccard, Support and Confidence

Rule			*lift*
AE	$\longrightarrow$	BC	0.75
CE	$\longrightarrow$	AB	1.00
BE	$\longrightarrow$	AC	1.20

Rule			*rsup*	*lift*	*leverage*
ACD	$\longrightarrow$	E	0.17	1.20	0.03
AC	$\longrightarrow$	E	0.33	1.20	0.06
AB	$\longrightarrow$	D	0.50	1.12	0.06
A	$\longrightarrow$	E	0.67	1.20	0.11

Rule			*rsup*	*conf*	*lift*
E	$\longrightarrow$	AC	0.33	0.40	1.20
E	$\longrightarrow$	AB	0.67	0.80	1.20
B	$\longrightarrow$	E	0.83	0.83	1.00

Rule			*rsup*	*lift*	*jaccard*
A	$\longrightarrow$	C	0.33	0.75	0.33

Contingency Table for X and Y

	Y	$\neg Y$	
X	$sup(XY)$	$sup(X\neg Y)$	$sup(X)$
$\neg X$	$sup(\neg XY)$	$sup(\neg X \neg Y)$	$sup(\neg X)$
	$sup(Y)$	$sup(\neg Y)$	$\lvert\mathbf{D}\rvert$

Rule Assessment Measures: Conviction

Define $\neg X$ to be the event that X is not contained in a transaction, that is, $X \not\subseteq t \in \mathcal{T}$, and likewise for $\neg Y$. There are, in general, four possible events depending on the occurrence or non-occurrence of the itemsets X and Y as depicted in the contingency table.

Conviction measures the expected error of the rule, that is, how often X occurs in a transaction where Y does not. It is thus a measure of the strength of a rule with respect to the complement of the consequent, defined as

$$conv(X \longrightarrow Y) = \frac{P(X) \cdot P(\neg Y)}{P(X \neg Y)} = \frac{1}{lift(X \longrightarrow \neg Y)}$$

If the joint probability of $X \neg Y$ is less than that expected under independence of X and $\neg Y$, then conviction is high, and vice versa.

Rule Conviction

Rule	*rsup*	*conf*	*lift*	*conv*
$A \longrightarrow DE$	0.50	0.75	1.50	2.00
$DE \longrightarrow A$	0.50	1.00	1.50	∞
$E \longrightarrow C$	0.50	0.60	0.90	0.83
$C \longrightarrow E$	0.50	0.75	0.90	0.68

Rule Assessment Measures: Odds Ratio

The odds ratio utilizes all four entries from the contingency table. Let us divide the dataset into two groups of transactions – those that contain X and those that do not contain X. Define the odds of Y in these two groups as follows:

$$odds(Y|X) = \frac{P(XY)/P(X)}{P(X\neg Y)/P(X)} = \frac{P(XY)}{P(X\neg Y)}$$
$$odds(Y|\neg X) = \frac{P(\neg XY)/P(\neg X)}{P(\neg X\neg Y)/P(\neg X)} = \frac{P(\neg XY)}{P(\neg X\neg Y)}$$

The odds ratio is then defined as the ratio of these two odds:

$$oddsratio(X \longrightarrow Y) = \frac{odds(Y|X)}{odds(Y|\neg X)} = \frac{P(XY) \cdot P(\neg X\neg Y)}{P(X\neg Y) \cdot P(\neg XY)}$$
$$= \frac{sup(XY) \cdot sup(\neg X\neg Y)}{sup(X\neg Y) \cdot sup(\neg XY)}$$

If X and Y are independent, then odds ratio has value 1.

Odds Ratio

Let us compare the odds ratio for two rules, $C \longrightarrow A$ and $D \longrightarrow A$. The contingency tables for A and C, and for A and D, are given below:

	C	$\neg C$
A	2	2
$\neg A$	2	0

	D	$\neg D$
A	3	1
$\neg A$	1	1

The odds ratio values for the two rules are given as

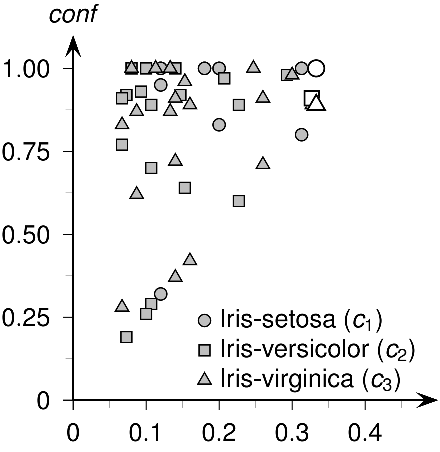
$$oddsratio(C \longrightarrow A) = \frac{sup(AC) \cdot sup(\neg A \neg C)}{sup(A \neg C) \cdot sup(\neg AC)} = \frac{2 \times 0}{2 \times 2} = 0$$

$$oddsratio(D \longrightarrow A) = \frac{sup(AD) \cdot sup(\neg A \neg D)}{sup(A \neg D) \cdot sup(\neg AD)} = \frac{3 \times 1}{1 \times 1} = 3$$

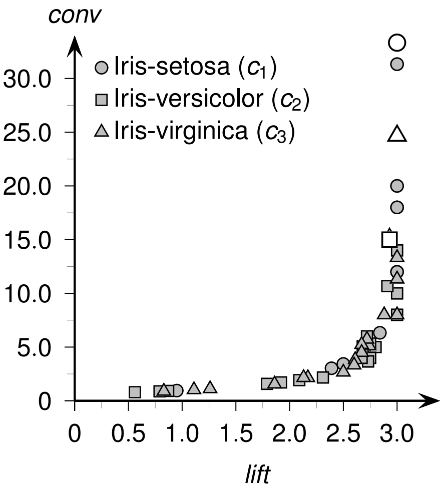
Iris Data: Discretization

Attribute	Range or value	Label
Sepal length	4.30–5.55	sl_1
	5.55–6.15	sl_2
	6.15–7.90	sl_3
Sepal width	2.00–2.95	sw_1
	2.95–3.35	sw_2
	3.35–4.40	sw_3
Petal length	1.00–2.45	pl_1
	2.45–4.75	pl_2
	4.75–6.90	pl_3
Petal width	0.10–0.80	pw_1
	0.80–1.75	pw_2
	1.75–2.50	pw_3
Class	Iris-setosa	c_1
	Iris-versicolor	c_2
	Iris-virginica	c_3

Iris: Support vs. Confidence, and Conviction vs. Lift



(a) Support vs. confidence



(b) Lift vs. conviction

Iris Data: Best Class-specific Rules

Best Rules by Support and Confidence

Rule	*rsup*	*conf*	*lift*	*conv*
$\{pl_1, pw_1\} \longrightarrow c_1$	0.333	1.00	3.00	33.33
$pw_2 \longrightarrow c_2$	0.327	0.91	2.72	6.00
$pl_3 \longrightarrow c_3$	0.327	0.89	2.67	5.24

Best Rules by Lift and Conviction

Rule	*rsup*	*conf*	*lift*	*conv*
$\{pl_1, pw_1\} \longrightarrow c_1$	0.33	1.00	3.00	33.33
$\{pl_2, pw_2\} \longrightarrow c_2$	0.29	0.98	2.93	15.00
$\{sl_3, pl_3, pw_3\} \longrightarrow c_3$	0.25	1.00	3.00	24.67

Pattern Assessment Measures: Support and Lift

Support: The most basic measures are support and relative support, giving the number and fraction of transactions in $\mathbf{D}$ that contain the itemset X:

$$sup(X) = |\mathbf{t}(X)| \qquad\qquad rsup(X) = \frac{sup(X)}{|\mathbf{D}|}$$

Lift: The *lift* of a k-itemset $X = \{x_1, x_2, \ldots, x_k\}$ is defined as

$$lift(X, \mathbf{D}) = \frac{P(X)}{\prod_{i=1}^{k} P(x_i)} = \frac{rsup(X)}{\prod_{i=1}^{k} rsup(x_i)}$$

Generalized Lift: Assume that $\{X_1, X_2, \ldots, X_q\}$ is a q-partition of X, i.e., a partitioning of X into q nonempty and disjoint itemsets X_i. Define the generalized lift of X over partitions of size q as follows:

$$lift_q(X) = \min_{X_1,\ldots,X_q} \left\{ \frac{P(X)}{\prod_{i=1}^{q} P(X_i)} \right\}$$

This is, the least value of lift over all q-partitions X.

Let Θ be some rule assessment measure. We generate all possible rules from X of the form $X_1 \longrightarrow X_2$ and $X_2 \longrightarrow X_1$, where the set $\{X_1, X_2\}$ is a 2-partition, or a bipartition, of X.

We then compute the measure Θ for each such rule, and use summary statistics such as the mean, maximum, and minimum to characterize X.

For example, if Θ is rule lift, then we can define the average, maximum, and minimum lift values for X as follows:

$$AvgLift(X) = \underset{X_1, X_2}{\text{avg}} \left\{ lift(X_1 \longrightarrow X_2) \right\}$$

$$MaxLift(X) = \max_{X_1, X_2} \left\{ lift(X_1 \longrightarrow X_2) \right\}$$

$$MinLift(X) = \min_{X_1, X_2} \left\{ lift(X_1 \longrightarrow X_2) \right\}$$

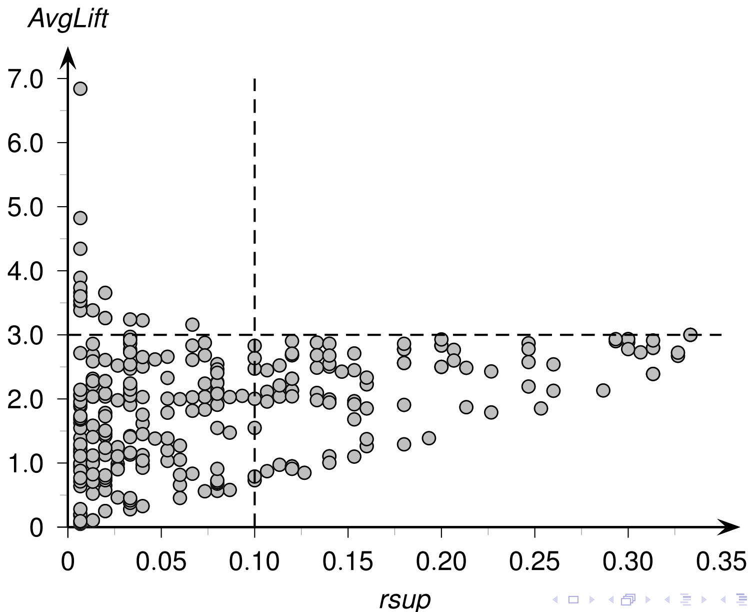
Iris Data: Support Values for $\{pl_2, pw_2, c_2\}$ and its Subsets

Itemset	*sup*	*rsup*
$\{pl_2, pw_2, c_2\}$	44	0.293
$\{pl_2, pw_2\}$	45	0.300
$\{pl_2, c_2\}$	44	0.293
$\{pw_2, c_2\}$	49	0.327
$\{pl_2\}$	45	0.300
$\{pw_2\}$	54	0.360
$\{c_2\}$	50	0.333

Rules Generated from $\{pl_2, pw_2, c_2\}$

Bipartition	Rule	*lift*	*leverage*	*conf*
$\{\{pl_2\}, \{pw_2, c_2\}\}$	$pl_2 \longrightarrow \{pw_2, c_2\}$	2.993	0.195	0.978
	$\{pw_2, c_2\} \longrightarrow pl_2$	2.993	0.195	0.898
$\{\{pw_2\}, \{pl_2, c_2\}\}$	$pw_2 \longrightarrow \{pl_2, c_2\}$	2.778	0.188	0.815
	$\{pl_2, c_2\} \longrightarrow pw_2$	2.778	0.188	1.000
$\{\{c_2\}, \{pl_2, pw_2\}\}$	$c_2 \longrightarrow \{pl_2, pw_2\}$	2.933	0.193	0.880
	$\{pl_2, pw_2\} \longrightarrow c_2$	2.933	0.193	0.978

Iris: Relative Support and Average Lift of Patterns



Comparing Itemsets: Maximal Itemsets

An frequent itemset X is *maximal* if all of its supersets are not frequent, that is, X is maximal iff

$$sup(X) \geq minsup, \text{ and for all } Y \supset X, sup(Y) < minsup$$

Given a collection of frequent itemsets, we may choose to retain only the maximal ones, especially among those that already satisfy some other constraints on pattern assessment measures like lift or leverage.

Iris: Maximal Patterns for Average Lift

Pattern	Avg. lift
$\{sl_1, sw_2, pl_1, pw_1, c_1\}$	2.90
$\{sl_1, sw_3, pl_1, pw_1, c_1\}$	2.86
$\{sl_2, sw_1, pl_2, pw_2, c_2\}$	2.83
$\{sl_3, sw_2, pl_3, pw_3, c_3\}$	2.88
$\{sw_1, pl_3, pw_3, c_3\}$	2.52

Closed Itemsets and Minimal Generators

An itemset X is *closed* if all of its supersets have strictly less support, that is,

$$sup(X) > sup(Y), \text{ for all } Y \supset X$$

An itemset X is a *minimal generator* if all its subsets have strictly higher support, that is,

$$sup(X) < sup(Y), \text{ for all } Y \subset X$$

If an itemset X is not a minimal generator, then it implies that it has some redundant items, that is, we can find some subset $Y \subset X$, which can be replaced with an even smaller subset $W \subset Y$ without changing the support of X, that is, there exists a $W \subset Y$, such that

$$sup(X) = sup(Y \cup (X \setminus Y)) = sup(W \cup (X \setminus Y))$$

One can show that all subsets of a minimal generator must themselves be minimal generators.

Closed Itemsets and Minimal Generators

sup	Closed Itemset	Minimal Generators
3	*ABDE*	*AD*, *DE*
3	*BCE*	*CE*
4	*ABE*	*A*
4	*BC*	*C*
4	*BD*	*D*
5	*BE*	*E*
6	*B*	*B*

Comparing Itemsets: Productive Itemsets

An itemset X is *productive* if its relative support is higher than the expected relative support over all of its bipartitions, assuming they are independent. More formally, let $|X| \geq 2$, and let $\{X_1, X_2\}$ be a bipartition of X. We say that X is productive provided

$$rsup(X) > rsup(X_1) \times rsup(X_2), \text{ for all bipartitions } \{X_1, X_2\} \text{ of } X$$

This immediately implies that X is productive if its minimum lift is greater than one, as

$$MinLift(X) = \min_{X_1, X_2} \left\{ \frac{rsup(X)}{rsup(X_1) \cdot rsup(X_2)} \right\} > 1$$

In terms of leverage, X is productive if its minimum leverage is above zero because

$$MinLeverage(X) = \min_{X_1, X_2} \left\{ rsup(X) - rsup(X_1) \times rsup(X_2) \right\} > 0$$

Comparing Rules

Given two rules $R : X \longrightarrow Y$ and $R' : W \longrightarrow Y$ that have the same consequent, we say that R is *more specific* than R', or equivalently, that R' is *more general* than R provided $W \subset X$.

Nonredundant Rules: We say that a rule $R : X \longrightarrow Y$ is *redundant* provided there exists a more general rule $R' : W \longrightarrow Y$ that has the same support, that is, $W \subset X$ and $sup(R) = sup(R')$.

Improvement and Productive Rules: Define the *improvement* of a rule $X \longrightarrow Y$ as follows:

$$imp(X \longrightarrow Y) = conf(X \longrightarrow Y) - \max_{W \subset X}\left\{conf(W \longrightarrow Y)\right\}$$

A rule $R : X \longrightarrow Y$ is *productive* if its improvement is greater than zero, which implies that for all more general rules $R' : W \longrightarrow Y$ we have $conf(R) > conf(R')$.

Fisher Exact Test for Productive Rules

Let $R : X \longrightarrow Y$ be an association rule. Consider its generalization $R' : W \longrightarrow Y$, where $W = X \setminus Z$ is the new antecedent formed by removing from X the subset $Z \subseteq X$.

Given an input dataset **D**, conditional on the fact that W occurs, we can create a 2×2 contingency table between Z and the consequent Y

W	Y	$\neg Y$	
Z	a	b	$a+b$
$\neg Z$	c	d	$c+d$
	$a+c$	$b+d$	$n = sup(W)$

where

$$a = sup(WZY) = sup(XY) \qquad b = sup(WZ\neg Y) = sup(X\neg Y)$$
$$c = sup(W\neg ZY) \qquad d = sup(W\neg Z\neg Y)$$

Fisher Exact Test for Productive Rules

Given a contingency table conditional on W, we are interested in the odds ratio obtained by comparing the presence and absence of Z, that is,

$$oddsratio = \frac{a/(a+b)}{b/(a+b)} \Bigg/ \frac{c/(c+d)}{d/(c+d)} = \frac{ad}{bc}$$

Under the null hypothesis H_0 that Z and Y are independent given W the odds ratio is 1. If we further assume that the row and column marginals are fixed, then a uniquely determines the other three values b, c, and d, and the probability mass function of observing the value a in the contingency table is given by the hypergeometric distribution.

$$P\Big(a \big| \; (a+c), (a+b), n\Big) = \frac{(a+b)!\;(c+d)!\;(a+c)!\;(b+d)!}{n!\;a!\;b!\;c!\;d!}$$

Fisher Exact Test: P-value

Our aim is to contrast the null hypothesis H_0 that $oddsratio = 1$ with the alternative hypothesis H_a that $oddsratio > 1$.
The *p-value* for a is given as

$$p\text{-}value(a) = \sum_{i=0}^{\min(b,c)} P(a+i \mid (a+c),(a+b),n)$$

$$= \sum_{i=0}^{\min(b,c)} \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{n!\,(a+i)!\,(b-i)!\,(c-i)!\,(d+i)!}$$

which follows from the fact that when we increase the count of a by i, then because the row and column marginals are fixed, b and c must decrease by i, and d must increase by i, as shown in the table below:

W	Y	$\neg Y$	
Z	$a+i$	$b-i$	$a+b$
$\neg Z$	$c-i$	$d+i$	$c+d$
	$a+c$	$b+d$	$n = sup(W)$

Fisher Exact Test: Example

Consider the rule $R: pw_2 \longrightarrow c_2$ obtained from the discretized Iris dataset. To test if it is productive, because there is only a single item in the antecedent, we compare it only with the default rule $\emptyset \longrightarrow c_2$. We have

$$a = sup(pw_2, c_2) = 49 \qquad b = sup(pw_2, \neg c_2) = 5$$
$$c = sup(\neg pw_2, c_2) = 1 \qquad d = sup(\neg pw_2, \neg c_2) = 95$$

with the contingency table given as

	c_2	$\neg c_2$	
pw_2	49	5	54
$\neg pw_2$	1	95	96
	50	100	150

Thus the *p-value* is given as
$p\text{-}value = \sum_{i=0}^{\min(b,c)} P(a+i \mid (a+c),(a+b),n) = 1.51 \times 10^{-32}$

Since the *p-value* is extremely small, we can safely reject the null hypothesis that the odds ratio is 1. Instead, there is a strong relationship between $X = pw_2$ and $Y = c_2$, and we conclude that $R: pw_2 \longrightarrow c_2$ is a productive rule.

Permutation Test for Significance: Swap Randomization

A *permutation* or *randomization* test determines the distribution of a given test statistic Θ by randomly modifying the observed data several times to obtain a random sample of datasets, which can in turn be used for significance testing.

The *swap randomization* approach maintains as invariant the column and row margins for a given dataset, that is, the permuted datasets preserve the support of each item (the column margin) as well as the number of items in each transaction (the row margin).

Given a dataset **D**, we randomly create *k* datasets that have the same row and column margins. We then mine frequent patterns in **D** and check whether the pattern statistics are different from those obtained using the randomized datasets. If the differences are not significant, we may conclude that the patterns arise solely from the row and column margins, and not from any interesting properties of the data.

Swap Randomization

Given a binary matrix $\mathbf{D} \subseteq \mathcal{T} \times \mathcal{I}$, the swap randomization method exchanges two nonzero cells of the matrix via a *swap* that leaves the row and column margins unchanged.

Consider any two transactions $t_a, t_b \in \mathcal{T}$ and any two items $i_a, i_b \in \mathcal{I}$ such that $(t_a, i_a), (t_b, i_b) \in \mathbf{D}$ and $(t_a, i_b), (t_b, i_a) \notin \mathbf{D}$, which corresponds to the 2×2 submatrix in $\mathbf{D}$, given as

$$\mathbf{D}(t_a, i_a; t_b, i_b) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

After a swap operation we obtain the new submatrix

$$\mathbf{D}(t_a, i_b; t_b, i_a) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

where we exchange the elements in $\mathbf{D}$ so that $(t_a, i_b), (t_b, i_a) \in \mathbf{D}$, and $(t_a, i_a), (t_b, i_b) \notin \mathbf{D}$. We denote this operation as $Swap(t_a, i_a; t_b, i_b)$.

Algorithm SwapRandomization

SwapRandomization(t, $\mathbf{D} \subseteq \mathcal{T} \times \mathcal{I}$):

while $t > 0$ **do**
  Select pairs $(t_a, i_a), (t_b, i_b) \in \mathbf{D}$ randomly
  if $(t_a, i_b) \notin \mathbf{D}$ *and* $(t_b, i_a) \notin \mathbf{D}$ **then**
    $\mathbf{D} \leftarrow \mathbf{D} \setminus \{(t_a, i_a), (t_b, i_b)\} \cup \{(t_a, i_b), (t_b, i_a)\}$
  $t = t - 1$
return D

Swap Randomization Example

Tid	Items A	B	C	D	E	Sum
1	1	1	0	1	1	4
2	0	1	1	0	1	3
3	1	1	0	1	1	4
4	1	1	1	0	1	4
5	1	1	1	1	1	5
6	0	1	1	1	0	3
Sum	4	6	4	4	5	

(a) Input binary data **D**

Tid	Items A	B	C	D	E	Sum
1	1	1	1	0	1	4
2	0	1	1	0	1	3
3	1	1	0	1	1	4
4	1	1	0	1	1	4
5	1	1	1	1	1	5
6	0	1	1	1	0	3
Sum	4	6	4	4	5	

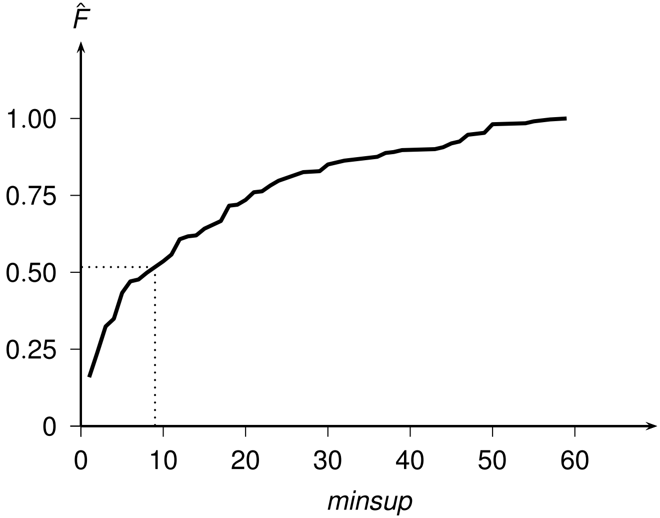
(b) $Swap(1, D; 4, C)$

Tid	Items A	B	C	D	E	Sum
1	1	1	1	0	1	4
2	1	1	0	0	1	3
3	1	1	0	1	1	4
4	0	1	1	1	1	4
5	1	1	1	1	1	5
6	0	1	1	1	0	3
Sum	4	6	4	4	5	

(c) $Swap(2, C; 4, A)$

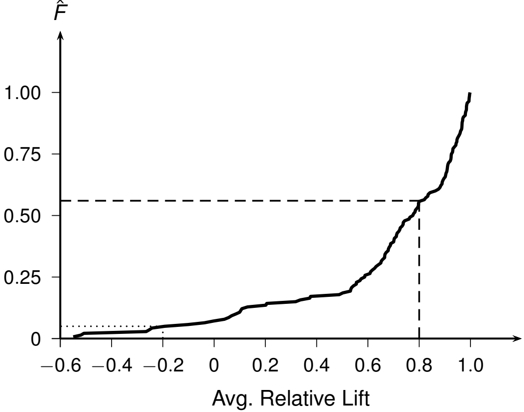
CDF for Number of Frequent Itemsets: Iris

$k = 100$ swap randomization steps



CDF for Average Relative Lift: Iris

$k = 100$ swap randomization steps



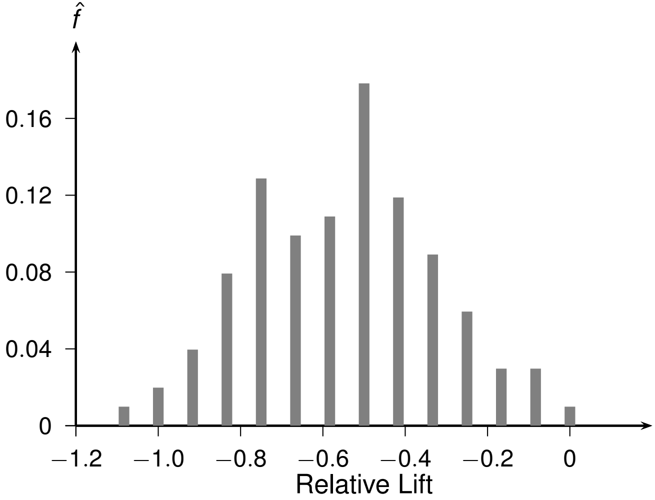
The relative lift statistic is

$$rlift(X, \mathbf{D}, \mathbf{D}_i) = \frac{sup(X, \mathbf{D}) - sup(X, \mathbf{D}_i)}{sup(X, \mathbf{D})} = 1 - \frac{sup(X, \mathbf{D}_i)}{sup(X, \mathbf{D})}$$

$\mathbf{D}_i$ is ith swap randomized dataset obtained after k steps.

PMF for Relative Lift: $\{sl_1, pw_2\}$

$k = 100$ swap randomization steps



Bootstrap Sampling for Confidence Interval

We can generate k bootstrap samples from $\mathbf{D}$ using sampling *with replacement*. Given pattern X or rule $R : X \longrightarrow Y$, we can obtain the value of the test statistic in each of the bootstrap samples; let θ_i denote the value in sample $\mathbf{D}_i$.

From these values we can generate the empirical cumulative distribution function for the statistic

$$\hat{F}(x) = \hat{P}\left(\Theta \leq x\right) = \frac{1}{k}\sum_{i=1}^{k} I(\theta_i \leq x)$$

where I is an indicator variable that takes on the value 1 when its argument is true, and 0 otherwise.

Given a desired confidence level α (e.g., $\alpha = 0.95$) we can compute the interval for the test statistic by discarding values from the tail ends of $\hat{F}$ on both sides that encompass $(1-\alpha)/2$ of the probability mass.

Bootstrap Confidence Interval Algorithm

BOOTSTRAP-CONFIDENCEINTERVAL(X, α, k, $\mathbf{D}$):

for $i \in [1, k]$ **do**

  $\mathbf{D}_i \leftarrow$ sample of size n with replacement from $\mathbf{D}$

  $\theta_i \leftarrow$ compute test statistic for X on $\mathbf{D}_i$

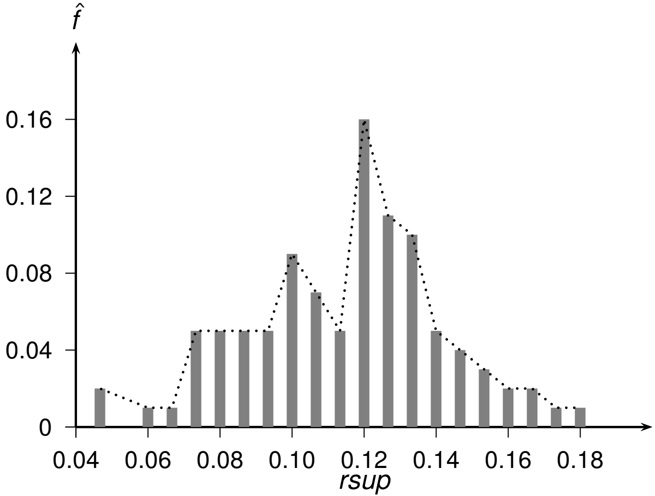
$\hat{F}(x) = P(\Theta \leq x) = \frac{1}{k}\sum_{i=1}^{k} I(\theta_i \leq x)$

$v_{(1-\alpha)/2} = \hat{F}^{-1}\big((1-\alpha)/2\big)$

$v_{(1+\alpha)/2} = \hat{F}^{-1}\big((1+\alpha)/2\big)$

return $\big[v_{(1-\alpha)/2}, v_{(1+\alpha)/2}\big]$

Empirical PMF for Relative Support: Iris



Empirical CDF for Relative Support: Iris