Chi-squared computation for association rules: preliminary results

Technical Report BC-CS-2003-01
July 2003

Sergio A. Alvarez
Computer Science Dept.
Boston College
Chestnut Hill, MA 02467 USA
alvarez@cs.bc.edu

ABSTRACT

Chi squared analysis is useful in determining the statistical significance level of association rules. We show that the chi squared statistic of a rule may be computed directly from the values of confidence, support, and lift (interest) of the rule in question. Our results facilitate pruning of rule sets obtained using standard association rule mining techniques, allow identification of statistically significant rules that may have been overlooked by the mining algorithm, and provide an analytical description of the relationship between confidence and support in terms of chi–squared and lift.

1. INTRODUCTION

An association rule is a rule of the form

$$A \Rightarrow B \tag{1}$$

where A and B are *itemsets*, that is, sets of items that appear in a database of *transactions*. This terminology is that of "market basket" analysis; each transaction itemset represents the set of items that are purchased together in a single retail transaction. An association rule such as that in Eq. 1 is meant to represent the statement that transactions that contain the itemset A are likely to also contain the itemset B, at least within a particular database of transactions. Association rules have been widely used within data mining since the development of the famous Apriori association rule mining algorithm [1], [2]. Various evaluation measures have been proposed to assess the degree to which an association rule applies to or is of interest in a given context. See [11] for a review and further references. Confidence and support are the most commonly used measures in part because of their central role in the Apriori algorithm. In [4], it is shown that the best rules according to many other measures can be found among those that lie along an upper "confidence–support border".

Despite the abundance of alternative evaluation measures, chi-squared analysis, a classical technique in statistics for determining the closeness of two probability distributions, continues to be one of the most widely used tools for statistical significance testing in many scientific circles, e.g. bioinformatics [7, 9]. It was suggested in [5] that chi-squared analysis be used to assess the statistical significance level of the dependence between antecedent and consequent in association rules. [5] also describes a mining algorithm that uses chi-square significance levels to prune the search for itemsets during mining. It has been acknowledged that the chi-square statistic does not by itself measure the strength of the dependence of the antecedent and consequent of a rule, and that an additional measure (such as confidence) is needed for this purpose; [5] uses the *interest* (also known as *lift*) associated with individual cells in the contingency table of the antecedent and consequent. Using chi-squared analysis to prune the set of mined rules was proposed in [8], where lift is again used as a measure of dependence for individual cells of the contingency tables.

Contributions of this paper

In the present paper we observe that the value of the chi-squared statistic for the itemset pair (A, B) viewed as a pair of binary–valued random variables may in fact be calculated directly in terms of the standard measures of confidence, support, and lift of the single rule $A \Rightarrow B$. We express the values of the four cells of the observed and expected contingency tables for (A, B) in terms of the values of these three measures for a single cell, and find a closed–form expression for the chi-square statistic

in terms of these values. Our results allow the statistical significance of associations to be estimated after mining has been completed using a standard mining algorithm, as long as the values of support, confidence, and lift of the mined rules are available. Also, generation of rule variants with improved lift values becomes possible. Finally, we are able to give an analytical description of the relationship between confidence and support with chi–squared and lift as parameters. We show in particular that a confidence–support tradeoff occurs for confidence values greater than a threshold value which we express in closed form. Our results extend to any set of three independent measures that are sufficient to describe the probabilities of all boolean combinations of two events A and B. We present related results involving classical information retrieval measures in [3].

Relation to previous work

We briefly discuss the differences between the treatment of chi-squared analysis in the present paper as compared with the papers [5], [8], [4].

First, [5] addresses the use of chi-squared significance levels in the mining process itself, while we focus on computing chi-squared levels independently of mining, e.g. for use in post–pruning. In computing the chi-square statistic in [5], the authors separately consider each of the items that appear in a rule, which leads to high–dimensional contingency tables (one dimension per item). Here, we instead aggregate the items of the antecedent and the items of the consequent separately, and consider the resulting two–dimensional contingency tables. This allows us to draw conclusions based on aggregate values that would be available from a standard association rule mining algorithm. Also, the use of lower dimensional tables makes it less likely that the expected cell values will be small enough to violate the assumptions required for validity of chi–squared significance testing.

[8] addresses the use of chi–squared analysis for pruning the rule set and finding a set of representative rules after mining, and appears to rely on two–dimensional contingency tables (the latter point is not completely clear in [8]). However, like [5], [8] also presents the lift as a measure that must be computed independently of the chi-squared statistic, and separately for each cell of the contingency table, in order to measure the strength of the dependence among all possible boolean combinations of the variables present in a rule. Neither [5] nor [8] offer insights or results regarding a possible connection between the chi-squared statistic and other metrics used for mining. In contrast, we focus on this very issue. We show that the chi–squared statistic may be expressed directly as a function of confidence, support, and lift of the single original rule $A \Rightarrow B$, and indeed of any three metrics that uniquely determine the probabilities of the four boolean combinations of two events.

[4] studies monotonicity of various evaluation measures relative to the partial order $\leq_{sc}$ defined on the set of rules by the support and confidence measures jointly:

$$R \leq_{sc} S \equiv supp(R) \leq supp(S) \text{ and } conf(R) \leq conf(S)$$

They point out that the chi–squared statistic is not monotonic relative to this partial order, exhibiting instead a "switching behavior" at a critical confidence value. The analysis in [4] uses an expression for chi–squared as a convex function of support and confidence (see Appendix A in [4]). However, this expression also depends on the confidence value of the rules $A \Rightarrow \overline{B}$ and $\emptyset \Rightarrow B$, not just that of the original rule $A \Rightarrow B$. [4] does not provide analytical relationships among measures for a given rule, focusing instead on convexity properties. This yields existence results (such as the existence of a switching point for the confidence) but falls short of providing the precise quantitative information that becomes available through the analytical results of the present paper.

2. CHI-SQUARED ANALYSIS

Consider two binary–valued random variables A and B. Chi–square analysis is a standard statistical technique that allows one to gauge the degree of dependence between the variables A and B; see, e.g. [6], pages 154–161. In this section we outline this approach.

2.1 Contingency tables

Computing the chi-square statistic for the pair of variables (A, B) requires constructing two contingency tables. The observed contingency table for (A, B) has four cells, corresponding to the four possible boolean combinations of A, B. The value in each cell is the number of observations (samples) that match the boolean combination for that cell. These values may be expressed in terms of the total number of samples n and of the observed relative frequencies (probabilities) corresponding to the four boolean combinations as shown in Table 1.

Table 1: Observed contingency table for (A, B)

	B	$\overline{B}$
A	$n\ P(A \cap B)$	$n\ P(A \cap \overline{B})$
$\overline{A}$	$n\ P(\overline{A} \cap B)$	$n\ P(\overline{A} \cap \overline{B})$

Chi-square analysis dictates that the observed contingency table should be compared with that which would be obtained asymptotically as $n \to \infty$ if the variables A and B were statistically independent. The latter table is shown in Table 2.

Table 2: Expected contingency table for (A, B)

	B	$\overline{B}$
A	$n\ P(A)\ P(B)$	$n\ P(A)\ (1 - P(B))$
$\overline{A}$	$n\ (1 - P(A))\ P(B)$	$n\ (1 - P(A))\ (1 - P(B))$

2.2 The chi-squared statistic

The chi square statistic is defined in terms of the entries of the observed contingency table (Table 1) and the expected contingency table (Table 2) as follows.

$$\chi^2 = \sum_{0 \le i,j \le 1} \frac{(observed_{i,j} - expected_{i,j})^2}{expected_{i,j}}, \qquad (2)$$

Thus, χ^2 represents a summed normalized square deviation of the observed values from the corresponding expected values. Statistical significance levels corresponding to specific χ^2 values may be found in tables of the χ^2 distribution. Here, we are dealing with binary–valued attributes and so the number of degrees of freedom is 1. A summary table of the χ^2 distribution with 1 degree of freedom containing minimum χ^2 values for selected significance levels appears in Table 3.

Table 3: Selected χ^2 significance levels

Tail probability	.10	.05	.025	.01
Min. χ^2	2.706	3.841	5.024	6.635

For example, Table 3 shows that a chi-squared value of 6 has a significance level better than .025. This means that the residual probability that a chi-squared value of 6 would be observed if the underlying variables are actually independent is less than .025. We note that chi–squared significance levels are actually a large sample approximation. Validity of the chi–squared test requires that the cells of the expected contingency table (Table 2) contain values greater than 5; otherwise, Fisher's exact test may be needed in order to obtain accurate significance values.

3. CHI-SQUARED AS A FUNCTION OF CONFIDENCE, SUPPORT, AND LIFT

The statistical significance of an association rule may be gauged through chi-square analysis. In the approach to this problem presented in [5], the presence or absence of each item that appears in a rule is viewed as a random variable. This requires one dimension for each item, leading to high–dimensional contingency tables. Here, we adopt an alternate approach. For a rule $A \Rightarrow B$, we aggregate the items of the antecedent A and, separately, the items of the consequent B. In other words, we view the boolean product over each of these itemsets as a single binary–valued random variable. This allows us to deal with two–dimensional contingency tables regardless of the number of items that appear in a rule. One advantage of using lower–dimensional tables is that it becomes easier to achieve the minimum cell counts required for validity of chi–squared analysis.

In this section we show how to express the values that appear in the contingency tables associated with the pair (A, B) consisting of (the boolean products over) the antecedent and consequent of the rule $A \Rightarrow B$ in terms of the standard association rule measures of confidence (conf), support (supp), and lift (also known as interest) of the rule $A \Rightarrow B$. The basic observation is that the probabilities of the various events in the boolean algebra generated by A and B may be expressed in terms of confidence, support, and lift (see Lemma 3.1). We focus on these particular measures because they are widely used. However, our results extend to any set of three measures the collective values of which uniquely determine the probabilities of all boolean combinations of the two variables A and B.

Lemma 3.1. *The values of support, confidence, and lift of the rule $A \Rightarrow B$ satisfy the following identities (whenever the denominators are nonzero):*

$$\begin{aligned} P(A \cap B) &= supp \\ P(A) &= \frac{supp}{conf} \\ P(B) &= \frac{conf}{lift} \end{aligned} \qquad (3)$$

Proof. The proof is straightforward, using the following definitions of the three measures: [1]

$$\begin{aligned} conf &= \frac{P(A \cap B)}{P(A)} \\ supp &= P(A \cap B) \\ lift &= \frac{P(A \cap B)}{P(A)P(B)} \end{aligned} \qquad (4)$$

Theorem 3.2. *The contingency tables for the pair of binary random variables (A, B) corresponding to the association rule $A \Rightarrow B$ are given in terms of the confidence, support, and lift of this rule as shown in Tables 4 and 5. We assume here that $conf$ is nonzero.*

Table 4: Observed contingency table for $A \Rightarrow B$

	B	$\overline{B}$
A	$n\ supp$	$n\ \frac{supp}{conf}(1 - conf)$
$\overline{A}$	$n\ (\frac{conf}{lift} - supp)$	$n\ (1 - \frac{supp}{conf}\ (1 - conf) - \frac{conf}{lift})$

Table 5: Expected contingency table for $A \Rightarrow B$

	B	$\overline{B}$
A	$n\ \frac{supp}{lift}$	$n\ \frac{supp}{conf}(1 - \frac{conf}{lift})$
$\overline{A}$	$n\ (1 - \frac{supp}{conf})\ \frac{conf}{lift}$	$n\ (1 - \frac{supp}{conf})(1 - \frac{conf}{lift})$

Proof. We observe that Lemma 3.1 allows us to express the probabilities that appear in the observed contingency table (Table 1) and the expected contingency

[1]Note that the definition of lift used here coincides with that used in [4]; an extra n factor appears in [4] because they use the term support for the number of transactions that support a given rule rather than the percentage of such transactions in the database.

table (Table 2) in terms of the confidence, support, and lift of the rule $A \Rightarrow B$. This is because the probabilities in Eq. 3 uniquely determine the probabilities of all boolean combinations of A and B. In particular, we have the following identities:

$$\begin{aligned} P(A \cap B) &= supp \\ P(A \cap \overline{B}) &= P(A) - P(A \cap B) \\ &= \frac{supp}{conf} - supp \\ P(\overline{A} \cap B) &= P(B) - P(A \cap B) \\ &= \frac{conf}{lift} - supp \\ P(\overline{A} \cap \overline{B}) &= 1 - P(A \cap B) - P(A \cap \overline{B}) - P(\overline{A} \cap B) \\ &= 1 - \frac{supp}{conf} - \frac{conf}{lift} + supp \end{aligned} \tag{5}$$

The identities of Eq. 5 provide expressions for all of the cells of the observed contingency table (Table 1) in terms of the confidence, support, and lift of the rule $A \Rightarrow B$. Substituting these expressions into Table 1 and simplifying them, we obtain Table 4, which is equivalent to Table 1. The entries of the expected contingency table (Table 2) may similarly be re-expressed in terms of the confidence, support, and lift of the underlying rule $A \Rightarrow B$ as shown in Table 5. All expressions shown are valid when the denominators are nonzero; since $lift \geq conf$, it is sufficient that $conf$ be nonzero. This completes the proof of Theorem 3.2.

Note.

Although Theorem 3.2 as stated above expresses the contingency table entries in terms of confidence, support, and lift, the proof of Theorem 3.2 shows that an analogous result will be obtained for any set of measures which determine the probabilities of all boolean combinations of A and B. In fact, it is clear that the analog of Eq. 3 suffices in order to obtain all cells of the contingency tables in terms of a given set of measures.

4. EXACT CHI-SQUARED FORMULA

We will now obtain a closed–form expression for the chi-squared statistic of Eq. 2 in terms of the confidence, support, and lift of the single rule $A \Rightarrow B$. Comments on limiting cases of the expression are provided afterwards.

Theorem 4.1. *The chi-square statistic of Eq. 2 satisfies the following equality (whenever the expression on the right-hand side is well defined).* [2]

$$\chi^2 = n\ (lift - 1)^2 \frac{supp\ conf}{(conf - supp)\ (lift - conf)} \tag{6}$$

Proof. A simple calculation using the expressions in Tables 4 and 5 shows that the numerator term in Eq. 2 is the same for all four cell positions. This is due to the fact that the marginals, that is, the row and column sums, must be the same in both tables. In fact, for all values of i, j we have:

$$(observed_{i,j} - expected_{i,j})^2 = n^2 supp^2 \left(\frac{lift - 1}{lift}\right)^2 \tag{7}$$

We may now compute the chi-square statistic as in Eq. 2, by using respectively Eq. 7 for the numerators in Eq. 2 and the appropriate cell expressions in Table 5 for the denominators in Eq. 2. The result is a product:

$$\begin{aligned} \chi^2 = n\ supp^2 \left(\frac{lift - 1}{lift}\right)^2 \\ \left(\begin{array}{l} \frac{lift}{supp} + \frac{conf}{supp}\ \frac{lift}{lift - conf} \\ + \frac{lift}{conf}\ \frac{conf}{conf - supp} \\ + \frac{conf\ lift}{(conf - supp)\ (lift - conf)} \end{array} \right) \end{aligned} \tag{8}$$

After some manipulation, the large parenthesized sum factor in Eq. 8 reduces to the following expression:

$$\frac{lift^2\ conf}{supp\ (lift - conf)\ (conf - supp)} \tag{9}$$

Replacing the parenthesized sum in Eq. 8 by the equivalent expression of Eq. 9 leads to the desired result in Eq. 6, and the proof of Theorem 4.1 is complete.

Corollary 4.2. *The chi-square statistic is bounded above as shown in Eq. 10.*

$$\chi^2 \leq n\ (lift - 1) \frac{supp\ conf}{conf - supp} \tag{10}$$

Proof. Since $conf \leq 1$, the bound in Eq. 10 follows from Eq. 6 by replacing the factor $lift - conf$ in the denominator by $lift - 1$.

We may also extract from Theorem 4.1 an expression for the χ^2 value of a rule in terms of lift and of the odds ratios of the antecedent and consequent. This expression, shown in Eq. 11, is equivalent to the relationship between χ^2 and the ϕ coefficient that we discuss and provide an independent proof of in Appendix A.

Corollary 4.3.

$$\chi^2(A \Rightarrow B) = n\ (lift - 1)^2 \frac{P(A)P(B)}{(1 - P(A))(1 - P(B))} \tag{11}$$

Proof. If we express all terms in the fraction on the right-hand side of Eq. 6 in terms of $supp$, we find for $supp \neq 0$:

$$\chi^2 = n\ (lift - 1)^2 \frac{supp\ \frac{supp}{P(A)}}{\left(\frac{supp}{P(A)} - supp\right)\left(\frac{supp}{P(A)P(B)} - \frac{supp}{P(A)}\right)}, \tag{12}$$

[2] The denominator in Eq. 6 is non-negative by Eq. 4, but is zero for rules $A \Rightarrow B$ such that A and B are mutually exclusive or one of A, B appears in all transactions.

Simplification of Eq. 12 yields Eq. 11. Although we derived Eq. 11 assuming that $supp \neq 0$, the resulting expression is valid even when $supp = 0$.

Indeed, letting $p = P(A)$ and $q = P(B)$ denote the probabilities of mutually exclusive events A and B, a direct calculation shows that the correct χ^2 value is:

$$\chi^2 = n\frac{pq}{(1-p)(1-q)}, \tag{13}$$

which agrees with Eq. 11 since $(lift-1)^2 = 1$ when $supp = 0$.

Singular cases in Theorem 4.1

Theorem 4.1 describes a relation among the four variables χ^2/n, $conf$, $supp$, and $lift$. At most points of the four–dimensional space defined by these variables, Eq. 6 constitutes a single–dimensional constraint that leaves three remaining degrees of freedom for the four variables. However, according to the implicit function theorem from multivariable calculus (e.g. [10]), this ideal situation breaks down at the *singular points* at which the gradient vector of the right-hand side of Eq. 6 viewed as a mapping from the three variables $conf$, $supp$, $lift$ to the real numbers is zero. We find the singular points of Eq. 6 in Lemma 4.4 below. We will need this result in section 6, where we explore the relationship between confidence and support.

Lemma 4.4. *The singular points of Eq. 6 are those for which* $lift = 0, 1$.

Proof. The proof of Lemma 4.4 reduces to calculating the three partial derivatives of the right-hand side of Eq. 6 and then setting them simultaneously equal to zero. This calculation is straightforward but somewhat lengthy, so we omit the details. We will note only that the quadratic factor $(lift-1)^2$ in Eq. 6 leads to a singularity when $lift = 1$. The other singular points are associated with the simultaneous satisfaction of the conditions $conf = supp$ and $lift = conf$. This occurs when $lift = 0$ (A and B mutually exclusive) and also when $P(A) = P(B) = 1$. Note that $lift = 1$ in the latter case.

5. INCREASING RULE LIFT

Our results in Theorems 3.2 and 4.1 can be used in at least the following two ways in order to improve the output of a standard association rule mining algorithm:

1. For ruleset pruning
2. To identify additional rules with higher lift

Since the first of these applications is perhaps obvious, we will comment only on the second.

Notice first that the χ^2 significance level of all "boolean variants" $A \Rightarrow \overline{B}$, $\overline{A} \Rightarrow B$, $\overline{A} \Rightarrow \overline{B}$ of a given rule $A \Rightarrow B$ is exactly the same as that of the original rule. This is because by the definition in Eq. 2, the χ^2 statistic is an aggregate value based on all cells of the observed and expected contingency tables for the pair of random variables associated with the antecedent and consequent of the rule. [3] Therefore, all of the boolean variants can be considered to have similar statistical significance; they may be ranked according to correlation strength as measured by lift. Nonetheless, it is possible for one of the three boolean variants of a given rule to have greater lift than the original rule, but smaller support. If the rule mining algorithm at hand fits within the support/confidence framework, and more generally unless lift is used as a criterion in guiding the search for rules, the mining process may miss some of these rules entirely. In this section we will describe one situation in which this phenomenon can occur. Details concerning a second situation appear in Appendix B of this paper.

Theorem 5.1. *Suppose that the rule* $A \Rightarrow B$ *satisfies*

$$\begin{aligned} lift(A \Rightarrow B) &< 1 \\ conf(A \Rightarrow B) &> 0.5 \end{aligned} \tag{14}$$

Then the rule $A \Rightarrow \overline{B}$ *with negated consequent satisfies:*

$$\begin{aligned} lift(A \Rightarrow \overline{B}) &= \frac{1-conf}{1-\frac{conf}{lift}} > 1 \\ supp(A \Rightarrow \overline{B}) &= \frac{supp}{conf}(1-conf) < supp(A \Rightarrow B) \end{aligned} \tag{15}$$

Proof of Theorem 5.1. The lift of the rule $A \Rightarrow \overline{B}$ equals

$$lift(A \Rightarrow \overline{B}) = \frac{P(A \cap \overline{B})}{P(A)P(\overline{B})} = \frac{1-conf}{1-\frac{conf}{lift}}, \tag{16}$$

where we have used Theorem 3.2 to rewrite the probabilities in terms of the confidence, support, and lift of the original rule $A \Rightarrow B$. Eq. 16 shows that the rule $A \Rightarrow \overline{B}$ will have lift greater than 1 if the lift of the latter is less than 1. This proves the first half of Eq. 15.

By Theorem 3.2, the support of the new rule $A \Rightarrow \overline{B}$ is given in terms of the support of the original rule by:

$$supp(A \Rightarrow \overline{B}) = \frac{supp}{conf}(1-conf) \tag{17}$$

The above quantity is less than $supp$ if $conf$ is greater than 0.5. This completes the proof.

5.1 Comments

We note the following points in connection with Theorem 5.1.

1. If a rule $A \Rightarrow B$ found by a mining algorithm such as Apriori satisfies the assumptions of Theorem 5.1, then its boolean variation $A \Rightarrow \overline{B}$ may

[3] It is also possible to prove by a direct but laborious calculation that Eq. 6 yields exactly the same χ^2 values for all boolean variants of the rule $A \Rightarrow B$.

well be overlooked during mining because by Eq. 15 the support of the latter rule may fall below the minimum support threshold used for mining. By Eq. 15, the missed rule has greater lift than the corresponding rule found by the mining algorithm in this case.

2. Eq. 15 allows one to recover the lift and support values of the rule $A \Rightarrow \overline{B}$ if this rule is missed by the mining algorithm, using only information about the original rule $A \Rightarrow B$. We note that the confidence value of the rule $A \Rightarrow \overline{B}$ may be similarly expressed using the appropriate identities in Eq. 5. Namely, the confidences of the two rules are complementary:

$$conf(A \Rightarrow B) + conf(A \Rightarrow \overline{B}) = 1 \qquad (18)$$

Example 5.1. Assume that the original rule $A \Rightarrow B$ satisfies:

$$conf = 0.6,\ supp = 0.2,\ lift = 0.8 \qquad (19)$$

We note that for the specific values given in Eq. 19, and assuming that the number of training instances is $n = 100$, Theorem 4.1 shows that the chi-squared statistic for the rules $A \Rightarrow B$ and $A \Rightarrow \overline{B}$ in this case is $\chi^2 = 6$, which corresponds to a significance level (residual probability) better than .025 according to Table 3.

Substituting the values given in Eq. 19 into Eq. 15 shows that the lift of the rule $A \Rightarrow \overline{B}$ here is 1.6, which exceeds that of the original rule by a significant margin. However, by Eq. 15 the support of the new rule is only 0.133 and may fall below the minimum support threshold used for mining; if so, it will be missed by a mining algorithm such as Apriori.

Example 5.2. In order to obtain an experimental illustration of the fact that a standard mining algorithm can fail to find some rules with significant lift, we ran the Apriori algorithm as implemented in the Weka system [12], using the Weather dataset provided with Weka. The minimum support level for mining was set at 0.2, which requires a support count of 3 or greater for this dataset. We note that the small size of this dataset (14 instances) leads to small χ^2 values for many of the rules. However, the phenomenon illustrated by this example extends to larger datasets.

1. Apriori found the rule:

$$outlook = rainy \Rightarrow play = yes$$
$$(supp = 3/14,\ conf = .6,\ lift = .93)$$

However, because of the minimum support constraint, Apriori missed the following higher lift variation of the above rule in which the consequent has been negated:

$$outlook = rainy => play = no$$
$$(supp = 2/14,\ conf = .4,\ lift = 1.13)$$

2. Apriori found the rule

$$play = yes => outlook = rainy$$
$$(supp = 3/14,\ conf = .33,\ lift = .93)$$

Complementing the consequent yields the following disjunctive rule which has higher lift:

$$play = yes => outlook = sunny \text{ or } overcast$$
$$(supp = 6/14,\ conf = .67,\ lift = 1.04)$$

The latter rule was of course not found by Apriori because of its disjunctive form. The closest rule that Apriori could find is:

$$play = yes => outlook = overcast$$
$$(supp = 4/14,\ conf = .44,\ lift = 1.56)$$

Notice that this rule found by Apriori has lower confidence than the rule with the full disjunction in the consequent, which Apriori did not find.

5.2 Discussion

Theorem 5.1 allows one to compensate for mining omissions such as those illustrated in the above examples. Indeed, one need only carry out the lift computations for the boolean variant $A \Rightarrow \overline{B}$ of each mined rule $A \Rightarrow B$ with lift less than 1, as in Eq. 15. This can be done after mining is complete. Any missed variants with significant lift values may be recovered in this way. We note, however, that such rules will have low confidence unless the minimum confidence threshold is set low during mining. This is because the confidence values of the two rules involved are complementary as shown in Eq. 18. Thus, this technique may be useful mainly in situations such as scientific discovery (e.g. [7, 9]) in which a significant lift value can be more important than high rule confidence.

6. ANALYTICAL DESCRIPTION OF THE RELATION BETWEEN CONFIDENCE AND SUPPORT

The very elegant paper [4] shows that the minimal rules mined according to several evaluation measures lie along an "upper confidence/support border". This border exhibits a tradeoff between support and confidence. However, the precise dependence between support and confidence along the border will depend on the statistics of the database and may be difficult to quantify in general; [4] contains no analytical results in this regard.

The results of the present paper allow us to say more about the relation between support and confidence for sets of rules that satisfy constraints on the values of other evaluation measures. For example, the result in Theorem 4.1 may be viewed as a relation between support and confidence, with χ^2/n and lift as parameters. If these parameters are held fixed, the resulting relation describes a curve in the support/confidence plane, as long as one stays away from the singular points of Eq. 6 as identified in Lemma 4.4.

By solving for the support in Eq. 6, we can find the dependence of support on confidence explicitly.

Lemma 6.1. *If $lift \neq 0, 1$, then supp may be expressed in terms of $conf$ as follows:* [4]

$$supp = \frac{conf * (lift - conf)}{\left(\frac{n(lift-1)^2}{\chi^2} - 1\right) conf + lift} \quad (20)$$

Eq. 20 expresses the support analytically as an explicit function of confidence with χ^2/n and lift as parameters. An example plot of Eq. 20 produced in MATLAB using the values $n = 100$, $\chi^2 = 6$, $lift = 2.5$ is shown in Fig. 1. Fig. 1 shows a tradeoff (inverse relationship)

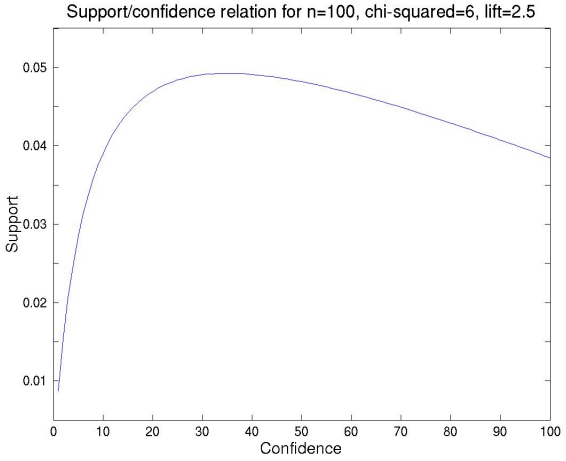

Figure 1: Support/confidence relationship

between confidence and support for confidence values above 40% or so. For smaller values of the confidence, a direct relationship between support and confidence is observed; in the latter case, higher confidence at the same χ^2 significance level would be obtained by negating the consequent of the rule.

Our results enable us to show that the change from a direct to an inverse relationship between confidence and support observed in Fig. 1 occurs in general, and we are able to determine precisely where this change takes place. We summarize this result in Theorem 6.2, which we prove in Appendix C. We exclude the case $\chi^2/n = 1$ in the statement of Theorem 6.2 because in the presence of the constraint $lift \neq 0$ it would only allow rules of the form $A \Rightarrow A$ and therefore would force $conf = 1$.

Theorem 6.2. *Assume that $\chi^2/n \neq 1$ and $lift \neq 0, 1$ are fixed. Then support is a downward concave function of confidence on the interval $0 < conf < 1$, with a unique maximum at the following value $conf^*$.*

$$conf^* = \frac{lift}{1 + \sqrt{\frac{n(lift-1)^2}{\chi^2}}} \quad (21)$$

Example 6.1. For instance, using the values $n = 100$, $\chi^2 = 6$, $lift = 2.5$ associated with Fig. 1, Theorem 6.2 yields to three digits:

$$conf^* = .351,$$

which certainly agrees with Fig. 1.

The existence as demonstrated above of a confidence value that separates regions of monotonicity and anti–monotonicity of the support as a function of the confidence is related to the fact (see [4]) that the chi–squared statistic exhibits a similar "switching" behavior relative to the partial order $\leq_{sc}$ described in [4] (see the discussion of related work in the Introduction of the present paper). As shown above, our results allow us to quantify such phenomena precisely in terms of the values of standard evaluation measures for a single rule $A \Rightarrow B$.

7. CONCLUSIONS

Chi-squared analysis is known to be useful in assessing the statistical significance of association rules. Previous work in this direction has either viewed the chi-squared statistic as being independent of other measures used to evaluate rules, or has relied on measures for several related rules in addition to the rule of interest. We have shown that all values contained in the two–dimensional contingency tables needed for computation of the chi-squared statistic may in fact be directly computed from the confidence, support, and lift values of the association rule in question, and we have presented a closed–form expression for the chi-squared statistic in terms of these three widely used measures. This allows easy computation of chi-squared values after mining, facilitates pruning of the mined rule set based on statistical significance, provides a way of recovering rules with significant lift values that may have been missed by a standard mining algorithm, and enables an analytical description of the confidence–support relationship with chi–squared and lift levels as parameters. In particular, we prove that a tradeoff between confidence and support occurs for confidence values above a certain threshold if chi–squared and lift are held constant, and we provide the value of this threshold in closed form. Our results extend easily to other sets of measures that collectively determine the probabilities $P(A)$, $P(B)$, and $P(A \cap B)$ for a given association rule $A \Rightarrow B$.

[4] If $lift = 1$, then $\chi^2/n = 0$ in Eq. 6, and *supp* cannot be determined from the remaining parameter $conf$. Indeed, if A and B are independent, with $P(A) = p$, $P(B) = q$, then $supp(A \Rightarrow B) = pq$, $conf(A \Rightarrow B) = q$.

8. REFERENCES

[1] R. Agrawal, T. Imielinski and A.N. Swami. Mining Association Rules between Sets of Items in Large Databases, in Peter Buneman and Sushil Jajodia (eds.), *Proc. 1993 ACM SIGMOD International Conference on Management of Data*, 207–216, 1993.

[2] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules, in J.B. Bocca, M. Jarke, and C. Zaniolo (eds.), *Proc. 20th International Conference on Very Large Databases (VLDB94)*, 487–499, 1994.

[3] S.A. Alvarez. "An Analytical Relation among Precision, Recall, and Classification Accuracy in Information Retrieval", Technical Report BC–CS–2002-01, Computer Science Department, Boston College, June 2002

[4] R.J. Bayardo, Jr. and R. Agrawal. Mining the most interesting rules. In *Proc. Fifth Intl. SIGKDD Conf. Knowledge Discovery and Data Mining (KDD1999)*, 145–154, 1999.

[5] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In J. M. Peckman (ed.), *Proc. ACM SIGMOD Conference on Management of Data (SIGMOD'97)*, pages 265 – 276, ACM, May 1997.

[6] M.G. Bulmer. *Principles of Statistics*, Dover, 1979

[7] A. Collins, C. Lonjou, and N.E. Morton. Genetic epidemiology of single nucleotide polymorphisms. *Proc. National Academy of Sciences*, 96(26), 15173–15177, Dec. 1999

[8] B. Liu, W. Hsu, Y. Ma. Pruning and summarizing the discovered associations. In *Proc. Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD1999)*, 125–134, Aug. 1999.

[9] G. Pesole, S. Liuni, M. D'Souza. PatSearch: a pattern matcher software that finds functional elements in nucleotide and protein sequences and assesses their statistical significance. *Bioinformatics*, 16(5), 439–450, 2000.

[10] M. Spivak. *Calculus on Manifolds*, Benjamin, 1965

[11] P.-N. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. In *Proc. Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2002)*, 32–41, July 2002.

[12] Ian H. Witten, Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kauffman, 1999

APPENDIX

A. CHI–SQUARED AND PHI

In this appendix we describe the close relationship that exists between the chi–squared statistic and the phi coefficient, a normalized directional measure of association that has previously been used in data mining (see e.g. [11]). The relationship between the two measures is well known to statisticians but appears not to be widely recognized in the data mining community.

The phi coefficient for two binary random variables A and B is defined as follows.

$$\phi = \frac{P(A \cap B) - P(A)P(B)}{\sqrt{P(A)P(B)(1 - P(A))(1 - P(B))}} \tag{22}$$

The phi coefficient of Eq. 22 is exactly the same as the Pearson correlation coefficient for a large sample distributed according to the given joint probability distribution (note that the denominator of Eq. 22 is precisely the product of the standard deviations of the variables A and B). It follows in particular that the phi coefficient takes values between -1 and $+1$.

The numerator of the phi coefficient in Eq. 22 may be expressed in terms of the determinant of the observed contingency table $O(A, B)$ for the variables A and B (Table 1).

Lemma A.1.

$$P(A \cap B) - P(A)P(B) = \frac{1}{n^2} \det O(A, B) \tag{23}$$

Proof. The result follows easily by using the identities:

$$P(A \cap \overline{B}) = P(A) - P(A \cap B)$$
$$P(\overline{A} \cap B) = P(B) - P(A \cap B)$$
$$P(\overline{A} \cap \overline{B}) = 1 - (P(A) + P(B) - P(A \cap B)),$$

to express each of the cells of Table 1 in terms of $P(A)$, $P(B)$, and $P(A \cap B)$, before computing the numerator of Eq. 22.

We will use Eq. 23 in the proof of the following result, which shows that χ^2 and ϕ are very closely related.

Theorem A.2.

$$\frac{\chi^2}{n} = \phi^2 \tag{24}$$

Proof. We will expand the left-hand side of Eq. 24 and show that it reduces to the phi coefficient on the right-hand side.

First write the observed contingency table (Table 1) in the notation shown in Table 6.

In this notation, the expected contingency table (Table 2) is as shown in Table 7.

Table 6: Observed contingency table for (A, B)

	B	$\overline{B}$
A	a	b
$\overline{A}$	c	d

Table 7: Expected contingency table for (A, B)

	B	$\overline{B}$
A	$\frac{1}{n}(a+b)(a+c)$	$\frac{1}{n}(a+b)(b+d)$
$\overline{A}$	$\frac{1}{n}(c+d)(a+c)$	$\frac{1}{n}(c+d)(b+d)$

Using the definition of the chi–squared statistic (Eq. 2), we now obtain:

$$\begin{aligned}\frac{\chi^2}{n} &= \frac{1}{n}\sum_{0\le i,j\le 1}\frac{(observed_{i,j}-expected_{i,j})^2}{expected_{i,j}} \\ &= \frac{((a+b)(a+c)-na)^2}{n^2(a+b)(a+c)} + \frac{((a+b)(b+d)-nb)^2}{n^2(a+b)(b+d)} \\ &+ \frac{((c+d)(a+c)-nc)^2}{n^2(c+d)(a+c)} + \frac{((c+d)(b+d)-nd)^2}{n^2(c+d)(b+d)}\end{aligned} \tag{25}$$

Expand the squares in Eq. 25 and collect like terms:

$$\begin{aligned}\frac{\chi^2}{n} &= \frac{(a+b)(a+c)}{n^2} + \frac{(a+b)(b+d)}{n^2} \\ &+ \frac{(c+d)(a+c)}{n^2} + \frac{(c+d)(b+d)}{n^2} \\ &- \frac{2}{n}(a+b+c+d) \\ &+ \frac{a^2}{(a+b)(a+c)} + \frac{b^2}{(a+b)(b+d)} \\ &+ \frac{c^2}{(c+d)(a+c)} + \frac{d^2}{(c+d)(b+d)} \\ &= \frac{1}{n^2}(a+b+c+d)^2 \\ &- \frac{2}{n}(a+b+c+d) \\ &+ \frac{a^2(b+d)(c+d) + b^2(a+c)(c+d)}{(a+b)(a+c)(b+d)(c+d)} \\ &+ \frac{c^2(a+b)(b+d) + d^2(a+b)(a+c)}{(a+b)(a+c)(b+d)(c+d)}\end{aligned} \tag{26}$$

Since the sum $a+b+c+d$ of all the entries in each contingency table is the total number of instances n, Eq. 26 simplifies as follows, where all terms have been expressed relative to a common denominator:

$$\begin{aligned}\frac{\chi^2}{n} &= -\frac{(a+b)(a+c)(b+d)(c+d)}{(a+b)(a+c)(b+d)(c+d)} \\ &+ \frac{a^2(b+d)(c+d) + b^2(a+c)(c+d)}{(a+b)(a+c)(b+d)(c+d)} \\ &+ \frac{c^2(a+b)(b+d) + d^2(a+b)(a+c)}{(a+b)(a+c)(b+d)(c+d)}\end{aligned} \tag{27}$$

We expand the numerator of the initial negative term in Eq. 27 and use the result to cancel parts of the positive terms that follow it. For example, notice that the numerator of the negative term includes the quantity $a^2(b+d)(c+d)$ as a summand, and that the latter exactly cancels the first half of the numerator of the first positive term in Eq. 27. At the end of this long cancelation and simplification process we arrive at the following:

$$\frac{\chi^2}{n} = \frac{(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \tag{28}$$

The numerator in Eq. 28 is the square of the determinant of the observed contingency table, while the denominator is n^4 times the product of the marginals. Thus, in light of the determinant identity for the ϕ coefficient in Eq. 23, we see that Eq. 28 is equivalent to Eq. 24. This completes the proof of the Theorem.

B. MORE ON FINDING RULES WITH INCREASED LIFT

We continue our discussion from section 5 concerning rules with high lift that may be overlooked by a mining algorithm operating within the support/confidence framework and that may be recovered after mining using the results of the present paper. Theorem 5.1 identified one situation in which this may occur. Another situation in which a significant rule may be missed by the mining algorithm due to low support is described by the following result.

Theorem B.1. *Suppose that the rule* $A \Rightarrow B$ *satisfies*

$$lift(A \Rightarrow B) > 1 \tag{29}$$

In order for the boolean variant $\overline{A} \Rightarrow \overline{B}$ *(with negated antecedent and consequent) to have increased lift and reduced support as follows:*

$$\begin{aligned} lift(\overline{A} \Rightarrow \overline{B}) &> lift(A \Rightarrow B) \\ supp(\overline{A} \Rightarrow \overline{B}) &< supp(A \Rightarrow B)\end{aligned} \tag{30}$$

it is necessary and sufficient that the following quadratic inequality hold:

$$conf^2 - lift\ conf + lift\ supp > 0 \tag{31}$$

Furthermore, it is sufficient that either one of the following two conditions hold:

$$\begin{aligned} conf(A \Rightarrow B) &> lift(A \Rightarrow B) - supp(A \Rightarrow B) \\ lift(A \Rightarrow B) &< 4\ supp(A \Rightarrow B)\end{aligned} \tag{32}$$

Proof. Using Theorem 3.2, we express the lift and support of the new rule $\overline{A} \Rightarrow \overline{B}$ in terms of the values $conf$, $supp$, and $lift$ of the three basic measures for the

original rule $A \Rightarrow B$:

$$\begin{aligned} & supp(\overline{A} \Rightarrow \overline{B}) \\ & = 1 - \frac{supp}{conf}\ (1 - conf) - \frac{conf}{lift} \\ & lift(\overline{A} \Rightarrow \overline{B}) \\ & = \frac{1 - \frac{supp}{conf}\ (1 - conf) - \frac{conf}{lift}}{(1 - \frac{supp}{conf})(1 - \frac{conf}{lift})} \end{aligned} \tag{33}$$

Since we wish to compare the lift and support of the new rule with the corresponding values for the original rule $A \Rightarrow B$, we subtract the appropriate value from each of the quantities in Eq. 33 above. After some simplification, we obtain the following results:

$$\begin{aligned} & supp(\overline{A} \Rightarrow \overline{B})\ -\ supp \\ & = -\ \frac{conf^2\ -\ lift\ conf\ +\ supp\ lift}{conf\ lift} \\ & lift(\overline{A} \Rightarrow \overline{B})\ -\ lift \\ & = \frac{(conf^2\ -\ lift\ conf\ +\ supp\ lift)(lift - 1)}{(conf - supp)\ (lift - conf)} \end{aligned} \tag{34}$$

Both denominators in Eq. 34 are positive, as is the term $lift - 1$ because of Eq. 29, so the signs of the two differences are determined by that of the quadratic term in the numerator, which they share (note, however, that the supp expression in Eq. 34 has an extra leading minus sign). This term is precisely the expression that which appears in Eq. 31. This proves the first claim of the Theorem.

In order to obtain sufficient conditions for Eq. 31 to hold, we observe first that the quadratic numerator expression that appears in Eq. 34 is positive when $conf$ is at positive or negative infinity. The roots of the expression are given by:

$$\begin{aligned} conf &= \frac{1}{2}\left(lift\ \pm\ \sqrt{lift^2\ -\ 4\ lift\ supp}\right) \\ &= \frac{lift}{2}\left(1\ \pm\ \sqrt{1\ -\ 4\ \frac{supp}{lift}}\right) \end{aligned} \tag{35}$$

Assume first that the expression inside the square root in Eq. 35 is negative. This occurs when the second condition in Eq. 32 holds. Then the quadratic expression in Eq. 31 is positive for all values of $conf$, which shows that the second condition in Eq. 32 is sufficient for the new rule to have increased lift and reduced support as claimed.

Now assume that the expression inside the square root in Eq. 35 is positive or zero. Because of the downward convexity of the graph of the square root function, we have the following inequality valid for all x such that the argument inside the square root below is non-negative:

$$\sqrt{1 + 2x} \leq 1 + x \tag{36}$$

Using Eq. 36 in Eq. 35, we see that the two roots in Eq. 35 lie inside the interval with the following endpoints:

$$\begin{aligned} conf &= \frac{lift}{2}\left(1\ \pm\ \left(1\ -\ 2\ \frac{supp}{lift}\right)\right) \\ &= \{supp,\ lift - supp\} \end{aligned} \tag{37}$$

Because of the positivity of the quadratic expression at positive and negative infinity, this implies that it is sufficient to have either $conf < supp$ or $conf > lift - supp$. The first of these two conditions is impossible. The second, however, may occur, and corresponds exactly to the first condition that appears in Eq. 32. This condition is therefore sufficient for the new rule to have increased lift and reduced support. This completes the proof of the Theorem.

C. THE SUPPORT/CONFIDENCE RELATIONSHIP

We provide a proof of the concavity of the graph of support as a function of confidence and of the expression for the unique maximum point of this graph as stated in Theorem 6.2.

Proof of Theorem 6.2. Partial differentiation of Eq. 20 with the parameters χ^2/n and $lift$ held constant yields:

$$\frac{\partial supp}{\partial conf}\ = \\ -\ \frac{\left(\frac{n(lift-1)^2}{\chi^2} - 1\right) conf^2\ +\ 2\ lift\ conf\ -\ lift^2}{\left(\left(\frac{n(lift-1)^2}{\chi^2} - 1\right) conf\ +\ lift\right)^2} \tag{38}$$

Since $lift \neq 0, 1$, the expression inside the outer square in the denominator is strictly greater than $lift - conf$ and by Eq. 4 is therefore strictly positive on the interval $0 < conf < 1$.

In order to determine the concavity of the graph of support as a function of confidence, we differentiate Eq. 38. Direct calculation yields Eq. 39.

$$\frac{1}{2}\frac{\partial^2 supp}{\partial conf^2}\ = -\ \frac{\frac{n\ lift^2 (lift-1)^2}{\chi^2}}{\left(\left(\frac{n(lift-1)^2}{\chi^2} - 1\right) conf\ +\ lift\right)^3} \tag{39}$$

The quantity on the right-hand side of Eq. 39 is negative for $lift \neq 1$, which proves that the graph of support as a function of confidence is concave downward as claimed.

The concavity of the graph on the interval $0 < conf < 1$ implies that it has at most one maximum point in that interval. We will show that there is indeed a maximum point and we will determine its location. A maximum point will occur at a value of $conf$ for which the derivative in Eq. 38 is zero. Since the denominator of Eq. 38 is strictly positive on the interval $0 < conf < 1$, this will occur precisely at a root of the numerator.

The numerator of Eq. 38 is quadratic in the confidence,

with two roots given by:

$$conf = \frac{lift}{\frac{n(lift-1)^2}{\chi^2} - 1}\left(-1 \pm \sqrt{\frac{n(lift-1)^2}{\chi^2}}\right) \quad (40)$$

By using the factorization identity $a^2-b^2 = (a+b)(a-b)$ in Eq. 40, we obtain the following simplified expression for the roots:

$$conf = \frac{lift}{1 \pm \sqrt{\frac{n(lift-1)^2}{\chi^2}}} \quad (41)$$

The fact that the derivative in Eq. 38 has two roots instead of one may at first seem to contradict the concavity of the graph of support as a function of confidence. That this is not the case is due to the fact that exactly one of the two roots will lie inside the interval $0 < conf < 1$. This fact, stated as Lemma C.1 below, completes the proof of Theorem 6.2.

Lemma C.1. *In Eq. 41, the value*

$$conf^* = \frac{lift}{1 + \sqrt{\frac{n(lift-1)^2}{\chi^2}}} \quad (42)$$

lies in the open interval $0 < conf < 1$*, but the value*

$$conf^{**} = \frac{lift}{1 - \sqrt{\frac{n(lift-1)^2}{\chi^2}}} \quad (43)$$

does not; in fact, $conf^{**}$ *lies outside the closed interval* $0 \leq conf \leq 1$*.*

Proof of Lemma C.1. Using the relationship between χ^2/n and the ϕ coefficient described in Appendix A, we rewrite the roots in Eq. 42 and Eq. 43 as follows:

$$\begin{aligned} conf^+ &= \frac{lift}{1 + \frac{(lift-1)}{|\phi|}} = \frac{1}{\frac{1}{lift} + \frac{1}{|\phi|} - \frac{1}{lift\ |\phi|}} \\ conf^- &= \frac{lift}{1 - \frac{(lift-1)}{|\phi|}} = \frac{1}{\frac{1}{lift} - \frac{1}{|\phi|} + \frac{1}{lift\ |\phi|}} \end{aligned} \quad (44)$$

Note that since by convention the square root in Eq. 42 and Eq. 43 is positive, the root $conf^*$ will be either $conf^+$ or $conf^-$ depending on the sign of $lift - 1$:

$$conf^* = \begin{cases} conf^+ & \text{if } lift > 1 \\ conf^- & \text{if } lift < 1 \end{cases} \quad (45)$$

Recall that ϕ may be interpreted as a Pearson correlation and therefore takes values between -1 and $+1$. Its absolute value $|\phi|$ takes values between 0 and 1. Since we assume that $lift \neq 1$, we exclude the value $\phi = 0$ which occurs only when the antecedent and consequent are independent. Because of Theorem A.2, the statement of Theorem 6.2 also excludes the values $\phi = \pm 1$. Therefore, we have $0 < |\phi| < 1$ in the present context.

We now consider two cases, depending on whether the lift of the rule is greater than 1 or less than 1.

1. If $lift > 1$, then $1/lift < 1$, and taking into account also that $1/|\phi| > 1$ here, the reciprocals of the roots satisfy:

$$\begin{aligned} \frac{1}{conf^+} &= \frac{1}{lift} + \frac{1}{|\phi|}\left(1 - \frac{1}{lift}\right) \\ &> \frac{1}{lift} + \left(1 - \frac{1}{lift}\right) \\ &= 1 \\ \frac{1}{conf^-} &= \frac{1}{lift}\left(1 + \frac{1}{|\phi|}\right) - \frac{1}{|\phi|} \\ &< \left(1 + \frac{1}{|\phi|}\right) - \frac{1}{|\phi|} \\ &= 1 \end{aligned} \quad (46)$$

   It follows from Eq. 46 that $conf^+$ lies in the interval $0 < conf < 1$ in this case and that $conf^-$ lies outside the closed interval $0 \leq conf \leq 1$. By Eq. 45, this proves the statement of Lemma C.1 in this case.

2. If $lift < 1$, then $1/lift > 1$, and taking into account also that $1/|\phi| > 1$, the reciprocals of the roots satisfy:

$$\begin{aligned} \frac{1}{conf^+} &= \frac{1}{lift} + \frac{1}{|\phi|}\left(1 - \frac{1}{lift}\right) \\ &< \frac{1}{lift} + \left(1 - \frac{1}{lift}\right) \\ &= 1 \\ \frac{1}{conf^-} &= \frac{1}{lift}\left(1 + \frac{1}{|\phi|}\right) - \frac{1}{|\phi|} \\ &> \left(1 + \frac{1}{|\phi|}\right) - \frac{1}{|\phi|} \\ &= 1 \end{aligned} \quad (47)$$

   By Eq. 47 that the roles of the two roots have been reversed: $conf^-$ lies inside the open interval $0 < conf < 1$ in this case and $conf^+$ lies outside the closed interval $0 \leq conf \leq 1$. However, note that the root that lies inside the open interval is $conf^*$ in this case as well, by Eq. 45. This completes the proof of Lemma C.1.