

Approximation of Frequentness Probability of Itemsets in Uncertain Data

Toon Calders
Eindhoven University of Technology
The Netherlands

Calin Garboni
University of Antwerp
Belgium

Bart Goethals
University of Antwerp
Belgium

Abstract—Mining frequent itemsets from transactional datasets is a well known problem with good algorithmic solutions. Most of these algorithms assume that the input data is free from errors. Real data, however, is often affected by noise. Such noise can be represented by uncertain datasets in which each item has an existence probability. Recently, *Bernecker et al. (2009)* proposed the *frequentness probability*; i.e., the probability that a given itemset is frequent, to select itemsets in an uncertain database. A dynamic programming approach to evaluate this measure was given as well. We argue, however, that for the setting of *Bernecker et al. (2009)*, that assumes independence between the items, already well-known statistical tools exist. We show how the frequentness probability can be approximated extremely accurately using a form of the central limit theorem. We experimentally evaluated our approximation and compared it to the dynamic programming approach. The evaluation shows that our approximation method is extremely accurate even for very small databases while at the same time it has much lower memory overhead and computation time.

I. INTRODUCTION

In frequent itemset mining, the considered transaction dataset is typically represented as a binary matrix M where each line represents a transaction and every column corresponds to an item. An element M_{ij} represents the presence or the absence of the item j in transaction i by the value 1 or 0 respectively as in Table I (left). This is the basic traditional model, where we are certain that an item is present or absent in a transaction. For this type of data many algorithms have been proposed for mining frequent itemsets; i.e., sets of columns of M that have all ones in at least a given number of transactions (see e.g. [6] for an overview on frequent itemset mining).

In several applications, however, an item is not present or absent in a transaction, but rather the probability of it being in the transaction is given. This is the case for data collected from experimental measurements susceptible to noise. For example, in satellite picture data the presence of an object or feature can be expressed more faithfully by a probability score when it is obtained by subjective human interpretation or an image segmentation tool.

Such data is called uncertain data and Table I (right) presents a popular type of uncertain database. This example dataset consists of 4 transactions and 3 items. For every transaction, a score between 0 and 1 is given to reflect the probability that the item is present in the transaction.

TID	a	b	c
t_1	1	1	0
t_2	0	0	1
t_3	1	1	0
t_4	1	1	0

TID	a	b	c
t_1	0.9	0.8	0.2
t_2	0.1	0.1	0.9
t_3	0.6	0.8	0.3
t_4	0.9	0.9	0.2

Table I
CERTAIN DATASET (LEFT) AND UNCERTAIN DATASET (RIGHT)

For example, the existence probability of 0.9 associated to item a in the first transaction represents that there is a 90% probability that a is present in transaction t_1 and 10% probability that it is absent. Table I (left) actually represents an instantiation of the uncertain dataset depicted in Table I (right). Such instantiations are called *possible worlds*. There are $2^{|T|*|I|}$ possible worlds, where $|T|$ is the total number of transactions and $|I|$ the total number of items in the dataset. Under the assumption that presence and absence of the different items is statistical independent, the probability of a possible world is obtained by simply multiplying the individual item probabilities. We will call this model the *independent uncertain database model*. For example, the probability of the world in Table I (left) is $0.9 \times 0.8 \times 0.8 \times 0.9 \times \dots \times 0.8 = 0.0914$. A far less probable world is obtained if we take the complement of Table I (left); i.e., we switch the ones to zeroes and the zeroes to ones. The probability of this world is 1.92×10^{-10} . The probabilities of all possible worlds sum up to 1.

Mining frequent patterns from this kind of datasets is more difficult than mining from traditional transaction datasets. After all, computing the support of an itemset now has to take the existence probabilities of the items into consideration. To provide information about the frequency of an itemset, two approaches exist. One is based on the *expected support* as introduced by Chui et al. [5]. For every itemset, its expected support is computed and those for which it exceeds a minimum threshold are reported as frequent. The second one is called *frequentness probability* and it was introduced in [2]. For a particular itemset, it takes into consideration the probability distribution of the support and it gives its probability of being frequent at a given minimum support threshold. The existing methods have the drawback of being computationally costly and being exposed to rounding errors when dealing with low probability values.

For example, the dynamic programming approach proposed in [2] to compute the frequentness probability requires $\mathcal{O}(\text{minsup}|D|)$ computation steps, where minsup is the minimal support threshold and $|D|$ the size of the dataset.

This paper is similar in spirit as [3]. In [3] we showed that for the independent uncertain database model a simple sampling technique gives very good results when computing the expected frequency of an itemset. In this paper we show that also for estimating the itemset frequency distribution there are good statistical tools available. This time, however, we do not rely on sampling and Hoeffding's inequality to bound the error on the mean, but instead we use a Normal approximation. Concretely, we show that the frequentness probability of itemsets can be approximated by a normal distribution function using a weak version of the Central Limit Theorem. Even for small datasets this approximation turns out to be surprisingly accurate. In contrast to the exact computation, the approximation scales linearly in the number of transactions and is independent of any minimal support or minimal frequentness thresholds.

We would like to stress that the main core of this paper is about how the frequentness probability computation for a *single* itemset can be optimized using standard statistical techniques. As such, the conclusions of this paper extend to every data mining method that is based on frequentness properties of itemsets. The main message of this paper therefore is: *In the probabilistic database model that assumes independence between items in a transaction database, standard statistical tools such as the Central Limit Theorem can and should be used to characterize the frequency distribution of the itemsets.*

II. FREQUENTNESS PROBABILITY

In this section we formally define the notion of the uncertain database, frequency distribution, expected frequency, and frequentness probability. The definitions given here are equivalent to those in [5], [2].

Definition 1 (Uncertain database): Let T be a set of transaction identifiers, and I a set of items.

An *uncertain database* D over $T \times I$ is a total function from $T \times I$ to the interval $[0, 1]$.

A *possible world* W of D is a subset of $T \times I$. The *probability of the possible world* W given the uncertain database D , denoted $P_D(W)$, is defined as

$$P_D(W) := \left(\prod_{(t,i) \in W} D(t,i) \right) \left(\prod_{(t,i) \in (T \times I) \setminus W} (1 - D(t,i)) \right)$$

Let \mathcal{W}_D denote the set of all possible worlds of D .

The support of an itemset $X \subseteq I$ in a possible world W is defined as the number of transactions in W that contain X ; i.e.: $\text{sup}(X, W) := |\{t \in T \mid \forall x \in X : (t, x) \in W\}|$. It is easy to see that P_D describes a probability distribution over the possible worlds of the uncertain dataset. Computing

the frequency of an itemset in uncertain datasets is based on this distribution. Among all possible worlds, we don't now which one is the true world; rather, P_D assigns the probability of being the true world. Given this, the expected support of an itemset $X \subseteq I$ is:

$$\text{expSup}(X) := \sum_{W \in \mathcal{W}_D} P_D(W) \text{sup}(X, W) ,$$

which is equivalent to the definition in Chui et al. [5]:

$$\sum_{t \in T} \prod_{x \in X} P_D(t, x) .$$

That is, because all item occurrences are independent, we can rewrite the expression with the possible worlds into one that only involves for every transaction in T , the probability that it contains the itemset X .

The distribution of the support of an itemset $X \subseteq I$ can now be computed as follows:

$$P_D(\text{sup}(X) = i) = \sum_{\substack{W \in \mathcal{W}_D \\ \text{sup}(X, W) = i}} P_D(W) ,$$

which is equivalent to [2]:

$$\sum_{\substack{S \subseteq T \\ |S|=i}} \left(\prod_{t \in S} P_D(X \subseteq t) \right) \left(\prod_{t \in T \setminus S} (1 - P_D(X \subseteq t)) \right) ,$$

where $P_D(X \subseteq t)$ denotes $\prod_{x \in X} P_D(t, x)$. The probability of X having a support of i can thus be computed by summing over all subsets $S \subseteq T$ of size i , the probability that the transactions in S contain X , and all others do not.

A. The Frequentness Probability Problem

For given minimal support minsup , the *frequentness probability of itemset* X [2] is now defined as:

$$P_D(\text{sup}(X) \geq \text{minsup}) ,$$

which equals

$$\sum_{\substack{S \subseteq T \\ |S| \geq \text{minsup}}} \left(\prod_{t \in S} P_D(X \subseteq t) \right) \left(\prod_{t \in T \setminus S} (1 - P_D(X \subseteq t)) \right) ,$$

In [2], the following *Probabilistic Frequent Itemset Mining (PFIM)* problem was proposed: given an uncertain database D , a minimal support threshold minsup and a minimal frequentness probability τ , find all itemsets X with a frequentness probability of at least τ ; i.e., $P_D(\text{sup}(X) \geq \text{minsup}) \geq \tau$.

For solving the PFIM problem, in [2] a dynamic programming approach is proposed based on the following observation: let $T = \{t_1, \dots, t_n\}$, and let D_j denote the restriction of the uncertain database D to its first j

transactions t_1, \dots, t_j . Following the notations of [2], we denote $P_{D_j}(\text{sup}(X) \geq i)$ by $P_{\geq i,j}$. Then,

$$P_{\geq i,j} = P_D(X \subseteq t_j) \cdot P_{\geq i-1,j-1} + (1 - P_D(X \subseteq t_j)) \cdot P_{\geq i,j-1}.$$

By incremental computation of the values $P_{\geq i,j}$, the frequentness probability $P_D(\text{sup}(X) \geq \text{minsup}) = P_{\geq \text{minsup}, |T|}$ can be computed in $\mathcal{O}(\text{minsup}|T|)$. Different optimizations are possible, such as 0-1 optimization and early pruning for itemsets not meeting the frequentness probability. For details we refer to the paper [2]. The main take-away message we want to stress here is that the algorithm of [2] is exact, computes the frequentness probability for one itemset X , and has complexity $\mathcal{O}(\text{minsup}|T|)$. For computing all itemsets that satisfy the frequentness probability threshold, the monotonicity of frequentness is used: the frequentness probability of an itemset is at most as high as the frequentness probability of its subsets. This follows trivially from the fact that in every possible world where an itemset is frequent, also all its subsets are frequent.

III. APRIORIAPPROX: A METHOD FOR APPROXIMATING THE FREQUENTNESS PROBABILITY

In this section we propose an alternative method for approximating the frequentness probability, by approximating the frequency distribution of an itemset with a normal distribution using a weak version of the Central Limit Theorem. Every transaction can be considered as a single coin toss; either the transaction contains the itemset X , or it doesn't. The final support of the itemset X is the sum of the outcomes of these coin tosses. We can now use a weak form of the central limit theorem, known as the *Lyapunov's Central Limit Theorem*: Let X_1, X_2, \dots be an infinite sequence of stochastic variables, and let, for all numbers N , s_N^2 denote $\sum_{k=1}^N \sigma_k^2$ (σ_k^2 denotes the variance of variable X_k). The theorem states that if for some $\delta > 0$ the following two conditions hold:

- 1) $E[|X_k|^{2+\delta}]$ is finite for all k , and
- 2) $\lim_{N \rightarrow \infty} \frac{1}{s_N^{2+\delta}} \sum_{i=1}^N E[|X_i - \mu_i|^{2+\delta}] = 0$

then the Central Limit Theorem still holds; i.e.,

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n}$$

converges in distribution to a standard normal random variable as n goes to infinity.

In our case, X_k is a stochastic variable denoting if transaction t_k contains itemset X . X_k follows a Bernoulli distribution; $P_D(X_k = 0) = 1 - \prod_{x \in X} D(t_k, x)$, and $P_D(X_k = 1) = \prod_{x \in X} D(t_k, x)$. Therefore, $\sigma_k^2 = P_D(X_k = 1)(1 - P_D(X_k = 1))$, and $s_N^2 = \sum_{k=1}^N P_D(X_k = 1)(1 - P_D(X_k = 1))$. The support can be expressed by the probabilistic variable $S = \sum_{i=1}^{|T|} X_i$, and the expected support equals $E[S] = \text{expSup}(X)$.

Algorithm 1 Apriori Framework

Require: $D, \sigma, \text{minfreqprob}$

Ensure: $F(D, \sigma, \text{freqprob})$

```

1:  $C_1 := \{i | i \in I\}$ 
2:  $k := 1$ 
3: while  $C_k$  is not empty do
4:   for all transactions  $t$  in  $D$  do
5:     for Candidates in  $C_k$  do
6:       if  $X \subset t$  then
7:          $\text{update}(X)$ 
8:       end if
9:     end for
10:  end for
11:   $F_k := \{X \in C_k | X.\text{freqprob} \geq \text{minfreqprob}\}$ 
12:   $C_{k+1} := \text{generateCandidates}(F_k)$ 
13:   $k++$ 
14: end while
```

We will show that the two conditions hold for $\delta = 1$.

Clearly, for all k , $E[|X_k|^3] \leq 1$, as X_k is either 0 or 1, and it also holds that:

$$\lim_{N \rightarrow \infty} \frac{1}{s_N^{2+\delta}} \sum_{i=1}^N E[|X_i - \mu_i|^{2+\delta}] = 0$$

Therefore, $\frac{\sum_{i=1}^N (X_i - \mu_i)}{s_N}$ converges to a normal distribution for increasing N . For $N = |T|$, we get:

$$\frac{\sum_{i=1}^N (X_i - \mu_i)}{s_N} = \frac{S - \text{expSup}(X)}{\sqrt{s_N^2}}$$

Hence, for sufficiently large databases T , $\frac{S - \text{expSup}(X)}{\sqrt{s_N^2}}$ converges in probability to the standard normal distribution, and thus:

$$P(S \geq \text{minsup}) = \Phi\left(\frac{\text{minsup} - 0.5 - \text{expSup}(X)}{\sqrt{s_N^2}}\right)$$

A. Computing All Itemsets

Computing the frequentness probability can be done in a systematic manner thanks to its anti-monotonic property. For enumerating the candidates we use an Apriori-like breadth-first strategy depicted in the Algorithm 1. This approach consists in iteratively generating, counting and pruning the candidates. The differences from the traditional Apriori are in the counting step (line 7) and in pruning criterium (line 11). For the counting step, instead of incrementing the support, we have to update two variables: one for the expected support and the second one for

$$s_n^2 = \sum_{k=1}^n P_D(X_k = 1)(1 - P_D(X_k = 1))$$

	# trans.	Avg. length	# items	density
KOSARAK	990980	8.1	119	0.02%
CONNECT4	67557	43.0	129	33,3%
T40I10D100K	100000	39.6	942	4.2%
T25I15D320K	320000	26.28	994	0.02%

Figure 1. Datasets Summary

The approximation of frequentness probability will be computed after the dataset is scanned using the cumulative distribution function Φ :

$$freqprob(X) = \Phi\left(\frac{minsup - 0.5 - expSup(X)}{s_n}\right)$$

The pruning step relies on the frequentness probability counted as above. The pruning criterium is based on the condition that the frequentness probability is lower than the minimal frequentness threshold.

IV. EXPERIMENTS

The experiments were conducted on a GNU/Linux machine with a 2.1GHz CPU Dual Core and 3.5 Gb of main memory. We use the datasets from [1]. Connect4 and Kosarak are available on the FIMI repository¹. T40I10D100K and T25I15D320K were generated using the IBM synthetic data generator². A brief description of these datasets is available in 1. The original datasets were transformed by Aggarwal et al. [1] into uncertain datasets by assigning to every item in every transaction existential probabilities according to the normal distribution $N(\mu, \sigma^2)$, where μ and σ were randomly and independently generated with values ranging between $[0.87, 0.99]$ and $[1/21, 1/12]$ respectively.

For producing the results, we implemented the AprioriApprox method described in Section III. In order to evaluate our method in terms of execution time and accuracy of the results, we have implemented the PFIM method based on the Probabilistic Frequent Itemset Mining approach proposed in [2] and discussed in Section II-A. The same Apriori framework has been used with the difference that, for every itemset, a vector containing the non-zero existence probabilities of the itemset was stored. This is the 0-1 optimization since, for large sparse datasets, it is important not to carry vectors of the size of the dataset consisting mainly of null values. Based on this vector, the exact frequentness probability is computed according to the dynamic programming scheme. We consider this implementation as base-line method for validating the accuracy of our approximation.

In Figure 2 we compare the execution times of the two approaches. The X-axis represents the varying minsup, in

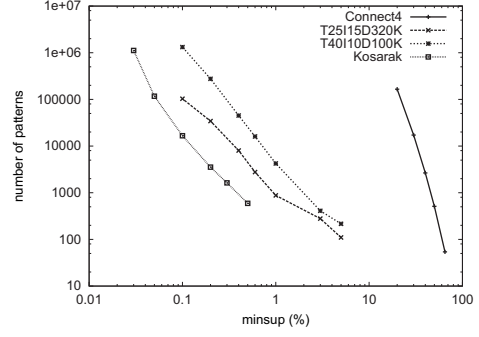


Figure 3. Number of patterns for varying support

percentage of the dataset size, and the Y-axis the execution time. The minimum frequentness probability is set to 0.8. We also used, for the Kosarak and datasets the early pruning optimization which speeds up the base-line method as it can be seen in Figure 2(b) and 2(c), but is still few orders of magnitude slower than AprioriApprox. As expected, in terms of execution time, AprioriApprox always outperforms PFIM. The reason is that our method does not overload the memory by constructing the vectors as PFIM. Moreover, PFIM requires $\mathcal{O}(minsup|D|)$ computation steps, where $|D|$ is the size of the dataset and $minsup$ is the minimum support. So, for higher minsup, more computations are needed, which is not the case of AprioriApprox. For the cases where PFIM overloaded the memory, we used a technique to avoid this, by splitting the set of candidates into batches such that by one traversal of the dataset only the candidates in the active batch were counted and then pruned, without overloading the memory. However, executions exceeding 2 hours were stopped.

Figure 3 shows the number of patterns for different values of minsup. Both X and Y axes are on log scale. We report the number of itemsets obtained with the approximation and, for the cases where the PFIM gave results, we compare the error introduced and the number of False Positives and False negatives which are depicted in Table II.

The error in approximation is depicted in Figures 5 and 6. The average error and the maximum error for every collection of patterns are represented. These errors were calculated only for the runs where PFIM succeeded since this gives the exact value of the frequentness probability. Almost exactly the same collections of patterns were generated, the average error in support being less than 0.0005 and never exceeding 0.0035. Very few false positives and false negatives were generated as can be seen in Table II. This was the case only for itemsets for which the frequentness probability was very close to the minimum probability threshold.

The next experiment investigates the impact of the number of tuples in the dataset on the accuracy of the approximation. We took the Kosarak dataset and we randomly sampled

¹<http://fimi.cs.helsinki.fi/>

²http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/datasets/syndata.html

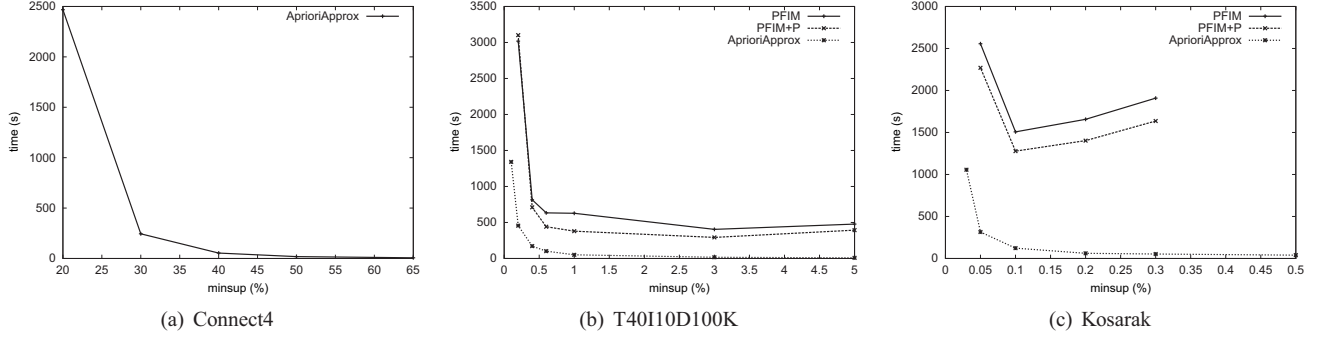


Figure 2. Execution Time

minsup	T40I10D100K		minsup	Kosarak	
0.01%	FP	-	0.03%	FP	-
	FN	-		FN	-
	TP	-		TP	-
0.02%	FP	47	0.05%	FP	10
	FN	4		FN	4
	TP	276438		TP	116973
0.04%	FP	1	0.1%	FP	0
	FN	1		FN	0
	TP	45155		TP	16672
0.06%	FP	0	0.2%	FP	0
	FN	0		FN	0
	TP	16072		TP	3535
1%	FP	0	0.3%	FP	0
	FN	0		FN	0
	TP	4238		TP	1626
3%	FP	0	0.5%	FP	-
	FN	0		FN	-
	TP	410		TP	-
5%	FP	0	0.8%	FP	-
	FN	0		FN	-
	TP	216		TP	-

Table II
CONFUSION MATRICES

it in order to obtain datasets with different number of transactions. From each of this newly generated datasets we extract the itemsets that satisfy the frequentness probability threshold set to 0.8 as in the previous experiments, given a minimum support of 0.2%. The results are presented in Table IV. As it can be seen, the approximation remains accurate, even for a very low number of transactions.

V. RELATED WORK

Recent work on frequent itemset mining in uncertain data focuses on methods for efficiently computing the expected support. This methods inherit the breadth-first and depth-first approaches from traditional frequent itemset mining and adapt the data structures to the probabilistic model. *U-Apriori* [5] represents a baseline algorithm for mining frequent itemsets from uncertain datasets. Because of its generate and test strategy, level by level, the method does not scale well. *UCP-Apriori* [4] is based on the decremental

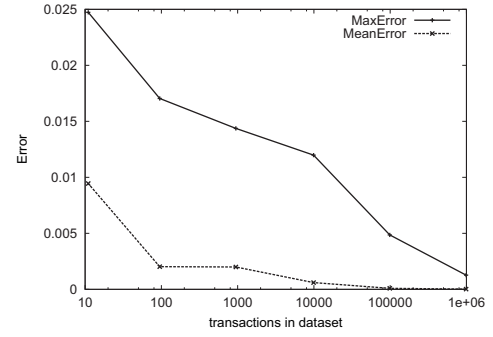


Figure 4. Max and mean approximation error for varying dataset size

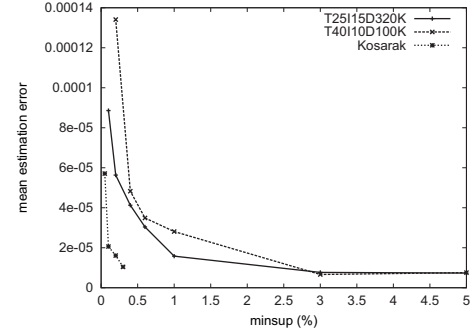


Figure 5. Mean approximation error for varying support

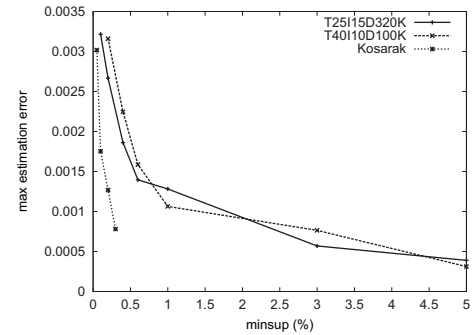


Figure 6. Max approximation error for varying support

	#sample tuples	False Positives	False Negatives	True Positives
Kosarak	990980	0	0	3535
Kosarak0.1	98869	0	0	3459
Kosarak0.01	9896	0	0	3223
Kosarak0.001	948	4	0	420
Kosarak0.0001	95	5	0	82
Kosarak0.00001	11	0	0	4

Table III
THE IMPACT OF THE DATASET SIZE ON THE APPROXIMATION FOR KOSARAK

pruning technique which consists in maintaining an upper bound of the support and decrementing it while scanning the dataset. The itemset is pruned as soon as its most optimistic value falls below the threshold. This approach represents the state of the art for mining frequent patterns from uncertain data with a generate-and-prune strategy. *UF-growth* [8] extends the FP-Growth algorithm [7]. Aggarwal et al. [1] extend several existing classical frequent itemset mining algorithms for deterministic data sets, and compare their relative performance in terms of efficiency and memory usage. The *UH-mine* algorithm, proposed in their paper, provides the best trade-offs. The algorithm is based on the pattern growth paradigm. The main difference with *UF-growth* is the data structure used which is an hyperlinked array. The limitations of these existing methods are the ones inherited from the original methods. The size of the data for the level-wise generate-and-test techniques affects their scalability and the pattern-growth techniques require a lot of memory for accommodating the dataset in the data structures, such as the FP-tree, especially when the transactions do not share many items. In the case of uncertain data, not only the items have to be shared for a better compression but also the existence probabilities, which is often not the case. Another approach for the independent uncertain database model is proposed in [3]. It is shown here that a simple sampling technique gives very good results when computing the expected frequency of an itemset.

VI. CONCLUSION

The focus of this work is approximating the frequentness probability of itemsets in uncertain datasets based on a special form of the Central Limit Theorem. Using this technique, the frequentness probability can be computed with minimal computational efforts and the results are highly accurate, actually invalidating the use of complicated algorithms for computing the exact solution. Similarly as in the paper [3], again we show that in the independent model one can easily apply simple statistical tools to reach excellent results. Basically, one could argue that for those independent cases statisticians solved the problems already a long time ago. A logical next step is now to see if these results can be extended to more complicated probability models. Clearly the statistical analysis of such cases will become much more challenging.

ACKNOWLEDGEMENTS

We are grateful to Charu C. Aggarwal, Yan Li, Jianyong Wang and Jing Wang for making available the datasets.

REFERENCES

- [1] C. C. Aggarwal, Y. Li, J. Wang, and J. Wang, "Frequent pattern mining with uncertain data," in *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2009, pp. 29–38.
- [2] T. Bernecker, H.-P. Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic frequent itemset mining in uncertain databases," in *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2009, pp. 119–128.
- [3] T. Calders, C. Garboni, and B. Goethals, "Efficient pattern mining from uncertain data with sampling," in *Proceedings of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-2010)*, 2010.
- [4] C. K. Chui and B. Kao, "A decremental approach for mining frequent itemsets from uncertain data," in *PAKDD*, ser. Lecture Notes in Computer Science, T. Washio, E. Suzuki, K. M. Ting, and A. Inokuchi, Eds., vol. 5012. Springer, 2008, pp. 64–75.
- [5] C. K. Chui, B. Kao, and E. Hung, "Mining frequent itemsets from uncertain data," in *PAKDD*, ser. Lecture Notes in Computer Science, Z.-H. Zhou, H. Li, and Q. Yang, Eds., vol. 4426. Springer, 2007, pp. 47–58.
- [6] B. Goethals, "Frequent set mining," in *In The Data Mining and Knowledge Discovery Handbook*. Springer, 2005, ch. 17, pp. 377–397.
- [7] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach," *Data Min. Knowl. Discov.*, vol. 8, no. 1, pp. 53–87, 2004.
- [8] C. K.-S. Leung, M. A. F. Mateo, and D. A. Brajczuk, "A tree-based approach for frequent pattern mining from uncertain data," in *PAKDD*, ser. Lecture Notes in Computer Science, T. Washio, E. Suzuki, K. M. Ting, and A. Inokuchi, Eds., vol. 5012. Springer, 2008, pp. 653–661.
- [9] T. Washio, E. Suzuki, K. M. Ting, and A. Inokuchi, Eds., *Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference, PAKDD 2008, Osaka, Japan, May 20-23, 2008 Proceedings*, ser. Lecture Notes in Computer Science, vol. 5012. Springer, 2008.