Appendix A. Counting strategy

To determine the optimal sequence to count itemset, we evaluate the information gain after counting of each itemset. We definition some symbols in table 4 first.

Table 4. Symbol definitions

N	Number of items (default value = 1k)
X	An itemset
$K(\mathbf{X})$	$\lvert\mathbf{X}\rvert$, the number of items in **X**
$P_f(\mathbf{X})$	The probability that **X** is a large itemset
$P_i(\mathbf{X})$	The probability that **X** is not a large itemset
T	The average number of items in a transaction (default value = 5)
$E(\mathbf{X})$	The expectation number of itemsets which can be determined frequency or not after counting X

For any itemset **X**, there are $2^{K(\mathbf{X})}$-2 subsets of **X** and $2^{(\mathbf{N}-K(\mathbf{X}))}$ supersets which contain **X**. So, the expectation number of itemsets which can be determined frequency or not after counting **X** $E(\mathbf{X})$ is:

$P_f(2^K-1) + P_i(2^{(\mathbf{N}-K)}+1)$

$= P_f(2^k-1) + P_i(2^{(1000-K)}+1)$, where **N** = 1000

$K(\mathbf{X}) = 1$:

$E(\mathbf{X}) = P_f(2-1) + P_i(2^{(1000-1)}+1)$

$= P_f + 2^{999}P_i + P_i \fallingdotseq 2^{999}P_i$

$K(\mathbf{X}) = 2$:

$E(\mathbf{X}) = P_f(4-1) + P_i(2^{(1000-2)}+1)$

$= 3P_f + 2^{998}P_i + P_i \fallingdotseq 2^{998}P_i$

The probability of a transaction contains a 1-itemset is:

$C_4^{999} / C_5^{1000} = 0.005$

The probability of a transaction contains a 2-itemset is:

$C_{3}^{998} / C_{5}^{1000} \fallingdotseq 0.00002$

Since $P_f(\mathbf{X}) \to 0$ when $K(\mathbf{X}) > 1$, we begin a level-wise searching from 1-itemset.

Appendix B. Proof sketches

Proof of Lemma 2. (G_j, r) is a *multi-dimension association rule w.r.t full match* in **MD**, that means r is an association rule hold in every element segmentation belong to $T[G_j]$. Suppose $T[G_j]$ is composed of n element segmentations : $T[E_1]$, $T[E_2]$, …. $T[E_n]$, r can be represented as **X**→**Y**, where **X** and **Y** are two disjoint itemset. The counts of **X**∪**Y** in $T[E_1]$, $T[E_2]$, …, $T[E_n]$ is C(r, $T[E_1]$), C(r, $T[E_2]$), …, $C(T[E_n])$ respectively, and the support of r in $T[E_1]$, $T[E_2]$, …, $T[E_n]$ is $S(r, T[E_1])$, $S(r, T[E_2])$, …, $S(r, T[E_n])$ respectively. Since r is holds in $T[E_1]$, $T[E_2]$, …, $T[E_n]$, for all i (i = 1 to n) $minsup(|T[E_i]|) \le C(r, T[E_i])$.

So, the support of **X**∪**Y** in $T[G_j]$ $= \frac{\sum_i C(r, T[E_i])}{\sum_i |T[E_i]|} \ge \frac{\sum_i minsup(|T[E_i]|)}{\sum_i |T[E_i]|} = minsup$.

So, **X**∪**Y** is a large itemset in $T[G_j]$, the proof of the confidence is similar.

Proof of Lemma 3. Since (G_j, r) is a *multi-dimension association rule w.r.t full match* in **MD**, and $T[G_j]$ = **MD**, r must be an association rule in every *element segmentation* of **MD**. Thus, (G_j, r) is a *multi-dimension association rule w.r.t full match* in **MD** for any *generalized patterns* of **MD**. According to Lemma 2, r must be an association rule in $T[G_j]$ for any G_j of **MD**.