minsup = 2

Database

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

$\widehat{C}_1$

TID	Set-of-Itemsets
100	{ {1}, {3}, {4} }
200	{ {2}, {3}, {5} }
300	{ {1}, {2}, {3}, {5} }
400	{ {2}, {5} }

L_1

Itemset	Support
{1}	2
{2}	3
{3}	3
{5}	3

C_2

Itemset	Support
{1 2}	1
{1 3}	2
{1 5}	1
{2 3}	2
{2 5}	3
{3 5}	2

$\widehat{C}_2$

TID	Set-of-Itemsets
100	{ {1 3} }
200	{ {2 3}, {2 5}, {3 5} }
300	{ {1 2}, {1 3}, {1 5}, {2 3}, {2 5}, {3 5} }
400	{ {2 5} }

L_2

Itemset	Support
{1 3}	2
{2 3}	2
{2 5}	3
{3 5}	2

C_3

Itemset	Support
{2 3 5}	2

$\widehat{C}_3$

TID	Set-of-Itemsets
200	{ {2 3 5} }
300	{ {2 3 5} }

L_3

Itemset	Support
{2 3 5}	2

Figure 12.3
Example.

12.3 Generating Rules

The association rules that we consider are somewhat more general than in (Agrawal *et al.* 1993) in that we allow a consequent to have more than one item. In this chapter we give an efficient generalization of the algorithm in (Agrawal *et al.* 1993).

For every large itemset l, we output all rules $a \Rightarrow (l - a)$, where a is a subset of l, such that the ratio support(l)/support(a) is at least *minconf*. The support of any subset $\tilde{a}$ of a must be as great as the support of a. Therefore, the confidence of the rule $\tilde{a} \Rightarrow (l - \tilde{a})$ cannot be more than the confidence of $a \Rightarrow (l - a)$. Hence, if a did not yield a rule involving all the items in l with a as the antecedent, neither will $\tilde{a}$. It follows that for a rule $(l - a) \Rightarrow a$ to hold, all rules of the form $(l - \tilde{a}) \Rightarrow \tilde{a}$ must also hold, where $\tilde{a}$ is a non-empty subset of a. For example, if the rule $AB \Rightarrow CD$ holds, then the rules $ABC \Rightarrow D$ and $ABD \Rightarrow C$ must also hold.

This characteristic is similar to the property that if an itemset is large