**2 Problem Formulation**

The uncertain data model applied in this paper is based on the possible worlds semantic with existential uncertain items.

We first introduce some basic concepts and then define the problem of mining probabilistic high utility itemsets in uncertain databases.

Definition 1. An uncertain item is an item i ∈ I whose presence in a transaction T ∈ D is defined by an existential probability P ( i ∈ T) ∈ (0, 1). A certain item is an item where P (i ∈ T) ∈ {0,1}. I is the set of all possible items, each item ip has a unit value v (i). An itemset X ⊆ I. P (X ∈ T) = ∏i∈x P (i ∈ T).

Definition 2. An uncertain transaction T is a transaction that contains uncertain items. A transaction database D containing uncertain transactions is called an uncertain transaction database.

Definition 3. Given an uncertain transaction database D, the expected support E (ip) of an item ip is defined as E (ip) = ∑T∈D P (ip ⊆ T), the expected support E (X) of an item X is defined as E (X) = ∑T∈D P (X ⊆ T).

In the certain database, the utility of an itemset X in transaction T is denoted as U (X, T) = ∑i∈X v(i), the utility of an itemset X in database CD is denoted as U (X, CD) = ∑X∈T∩T∈D U (X, T). If U (X, CD) ≥ *minutil*, itemset X is called a high utility itemset. The parameter minutil is the user specified minimum confidence in the utility of an itemset. They also apply to the uncertain database.

Definition 4. The expected utility of item X in transaction T is denoted as EU (X, T) = U (X) \* P (X, T), the expected utility of itemset X in uncertain transaction database D is denoted as

EU (X, D) = ∑X∈T∩T∈D EU (X, T). For example, the expected utility of {A} in transaction 1 is

EU ({A}, transaction1) = 3 \* 0.8 = 2.4, the expected utility of {A} in database is EU ({A}, D) = 2.4 + 2.7 + 1.8 = 6.9.

Definition 5. The probabilistic utility of itemset in uncertain transaction database D is denoted as up (X) = ∑1≤j≤|W|∩U(X,wj)≥minutil pj, where minutil is the utility threshold.

Definition 6. A probabilistic utility itemset (PUI) in uncertain transaction database is an itemset with a utility probability of at least *put,* up (I) ≥ put.

The parameter minutil is the user specified minimum confidence in the probabilistic utility of an itemset.

We are now able to specify the Mining Probabilistic High Utility Itemsets (MPHU) problem as follows; given an uncertain transaction database T, a minimum utility threshold *minutil* and a utility probability threshold *put*.

For example, given utility threshold minutil = 5 and probabilistic utility threshold put = 0.9, v (A) = 3, up ({A}) = ∑1≤j≤|W|∩U(X,wj)≥minutil pj = 0.8\*0.6\*0.9 + (1-0.8)\*0.6\*0.9 + 0.8\*(1-0.6)\*0.9 + 0.8\*0.6\*(1-0.9) = 0.876. Due to up({A}) < put, itemset {A} is not a probabilistic utility itemset.