Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 2, May 2026, pp. 426~441
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i2.pp426-441

An extended relational database model and algebra with interval probability valued attributes and tuples

Hoa Nguyen, Thi Nhi Tran
Faculty of Information Technology, Saigon University, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Aug 28, 2025
Revised Feb 24, 2026
Accepted Mar 4, 2026

Keywords:

EPRDB algebra
EPRDB data model
Interval probability
Probabilistic interpretation
Probabilistic relation
Probabilistic value
Tuple membership

ABSTRACT

This paper introduces an extended relational database model and algebra, named EPRDB, where both the attribute and tuple of a relation may take values associated with interval probabilities for modelling and computing uncertain and imprecise information. To build EPRDB, three key methods are employed: i) probabilistic values and intervals are used for representing uncertain and imprecise valued attributes and tuple membership degrees; ii) the probabilistic interpretations of binary relations on sets and operators on probability intervals are proposed for computing and querying the uncertain degree of relations on value domains of attributes; and iii) the combination strategies of probabilistic intervals and values are defined for manipulating probabilistic relational tuples. Then, the EPRDB data model including fundamental concepts and components such as the schema, probabilistic relation, functional dependency, and key is extended with interval probability valued attributes and tuples such that it is coherent and consistent with the classical relational data model. The EPRDB algebra including the set of basic probabilistic relational algebraic operations is developed corresponding to the EPRDB data model. A set of the properties of the algebraic operations is also formulated and proven. The new proposed EPRDB model and algebra can represent and deal effectively with uncertain and imprecise information in practical applications.

Corresponding Author:

Hoa Nguyen
Faculty of Information Technology, Saigon University
Ho Chi Minh City, Vietnam
Email: nguyenhoa@sgu.edu.vn

1. INTRODUCTION

Although the classical relational database model (CRDB) is useful to model, design, and implement large-scale systems [1], [2], it is limited to represent and deal with uncertain and imprecise information pervaded in the real world. This is due to the fact that the relations in CRDB are defined as classical sets where the information of elements of the sets (i.e., the relations) is certain and precise. In other words, both the attribute value and relational tuple membership (i.e., records) in classical relations of CRDB must be defined certainly and precisely. As such, CRDB model cannot represent a piece of information like "Mary may be diseased and it is about 50% likely that she has either hepatitis or cirrhosis". Here, Mary is uncertain to be a patient (diseased) and also uncertain which of the two diseases it is "hepatitis" or "cirrhosis". This has motivated and attracted much research effort for the past several years to extend and generalize the CRDB model to deal with uncertain and imprecise information.

Currently, there have been many non-classical database models including probabilistic relational database models (PRDB) studied and built to overcome the limitations of CRDB model. For instance, Eiter *et al.* [3] presented a PRDB model to compute the uncertain degree of attribute values of tuples in a probabilistic relation, and Suciu [4] proposed another PRDB model that can compute the uncertain membership degree of each tuple in a probabilistic relation. Some works extended PRDB models with fuzzy sets to enhance the ability of the representation and computation of uncertain and imprecise data, such as [5], [6]. However, no database model would be so universal that it could include all measures and tackle all aspects of uncertainty and imprecision of information in the real world.

Probabilistic database models have also been used in many real applications, such as [7]-[11]. Particularly, probabilistic databases were applied for detecting faulty sensors [7]. Queries over the relational cross model were processed by using uncertain databases [8] and probabilistic queries were employed to express and handle uncertain multidimensional data [9]. A probabilistic database was implemented using deep-learning infrastructure for the integration of the probabilistic logical reasoning [10], and an image retrieval application was developed using probabilistic databases [11].

Probabilistic relational database models are developed and built as extensions of CRDB based on the probability theory. There are two main classes of PRDB models proposed and built corresponding to two uncertain data levels to extend the CRDB model to a PRDB model: (1) at the relation level, CRDB relations (i.e., sets of tuples) are extended to PRDB relations such that each tuple is associated with a probability to represent its uncertain membership degree in the relations; or (2) at the attribute level, CRDB relational attributes are extended to PRDB relational attributes such that each attribute is associated with a probability to define the uncertainty degree of the values that it may take.

For the first PRDB model class, such as [12]-[16], where each tuple of a relation was associated with a probability in the interval [0, 1] to represent the uncertainty membership degree of that tuple for the relation. More generally, the model [17] defined formally a probabilistic database as a probability space on a σ-algebra measurable under a probability measure where each tuple of a relation in database instances was associated with a probability in the interval [0, 1]. However, in many natural situations, we cannot know precisely the probability that we can only estimate it as an approximate number in a subinterval of [0, 1]. The models [18]-[20] were extended with probability intervals associated with each tuple to overcome the shortcoming of the models in [12]-[17]. Yan and Ma [21] used possibility distributions inferred from fuzzy sets to flexibly represent the uncertainty membership degree of tuples of a probabilistic relation. Nevertheless, in general the PRDB models in the first class did not express and deal with the uncertainty of attribute values of relations.

For the second PRDB model class, such as [3], [9], [22], [23], where the value of a relational attribute was associated with a probability in the interval [0, 1] to represent the uncertain degree for that attribute taking the value. However, in many real cases, we cannot define precisely the probability for attribute values. We can only estimate it to be an approximate number in a subinterval of [0, 1]. The models [24], [25] overcame the restriction by using a pair of lower and upper-bound probability distribution functions to represent the possibility for an attribute taking a value in a set with a computed probability interval from the distribution function pair. However, when the probabilistic relations in databases [24], [25] have many attributes, the number of generated probability distribution functions is too large to lead to low performance in manipulating data. The researchers [26], [27] overcame the shortcoming of the models [24], [25] by using probability intervals on a set to represent attribute values. Nevertheless, in general the PRDB models in the second class did not represent and deal with the uncertainty of tuple memberships of relations.

As presented above, the attribute value and tuple membership of a relation in a practical database may be uncertain and should be represented and computed in real applications. Up to now, many PRDB models have been built for uncertain information. The models in the first class cannot represent and compute uncertain degrees of the relational attribute values whereas the models in the second class cannot represent and compute uncertain degrees of the relational tuple memberships. To our knowledge, there are few works to develop the extended PRDB models that can represent and deal with both the uncertain attribute value and tuple membership of a relation in a database. A few such extended models were proposed in [28]-[30]. However, limitations and shortcomings still exist in the models. Particularly, Lee [28] used probability distributions as $\{(v_1, p_1),..., (v_m, p_m)\}$ and probabilistic subintervals of [0, 1] to express the uncertainty of the attribute value and tuple membership, respectively of relations. Nguyen [29] employed uniform distributions as $\{(v_1, 1/m), ... , (v_m, 1/m)\}$ and probabilistic subintervals of [0, 1] to represent the uncertainty of the attribute value and tuple membership of probabilistic relations. The shortcoming of the models [28], [29] is that, in some real situations, we cannot know definitely the probability pi or a uniform probability distribution for the value vi. Cao [30] whereas used fuzzy sets and probabilistic subintervals of [0, 1] to describe the imprecision of the attribute value and the uncertainty of the tuple membership of relations. The limitation of the model [30] is that the tuple attribute of a relation certainly takes only a fuzzy set value. In other words, the model [30] did not represent the uncertainty of the attribute values of a relation. Thus, one can say that the ability to represent and deal with diversified uncertain information of the models [28]-[30] has been limited in the real world applications. This has motivated the development of more general database models to overcome limitations and shortcomings of the existing PRDB models.

In this paper, we propose a new extended PRDB model integrating interval probabilities, denoted EPRDB, that combines both uncertain data levels of the relational attribute value and tuple membership, extends CRDB model, and overcomes the shortcomings of the models [28]-[30] for representing and handling diversified uncertain information in practice.

To build EPRDB, we first adapt probabilistic values (i.e., a distribution of probabilistic intervals on a value set) in [31] with data types for representing uncertain attribute values of relations, propose a probabilistic interpretation of binary relations on sets, and use operators on probability intervals and combination strategies of probabilistic values in [26] for computing and combining the interval probability of the uncertain attribute values and tuple memberships in data queries and manipulations. Then, the probabilistic data model and relational algebraic operations are extended coherently and consistently with those of CRDB for EPRDB. The EPRDB model is also an extension of the models in [26], [28], [29] with uncertain attribute values and tuple memberships.

The probabilistic value used to represent the relational attribute in our EPRDB model is a distribution of probabilistic intervals on a value set having the form $\{(v_1, [l_1, u_1]), \ldots, (v_m, [l_m, u_m])\}$. Thus, the probabilistic values are the extensions of probability distributions and uniform distributions $\{(v_1, p_1), \ldots, (v_m, p_m)\}$ and $\{(v_1, 1/m), \ldots, (v_m, 1/m)\}$, respectively in the models [28], [29]. Moreover, each fuzzy set value v that represents a relational attribute value in the model [30] is also considered a special probability distribution $\{(v, 1.0)\}$. Hence, the built EPRDB model and algebra based on the probabilistic values and interval probability valued tuples can represent and deal more generally and flexibly with uncertain information to overcome the shortcomings of the existing PRDB models including the models [28]-[30]. The EPRDB model can compute, manipulate and query effectively uncertain and imprecise information in the real database applications.

The mathematical basis to build EPRDB is presented in Section 2. The proposed methodology for extending the data model and developing probabilistic relational algebraic operations on EPRDB is introduced in Section 3. Section 4 presents the achieved results and discussions of the EPRDB model. Finally, Section 5 concludes the paper and outlines further research directions.

2. BASIC PROBABILITY DEFINITIONS AND NOTIONS

Basic probability definitions and notions are presented in this section as a mathematical base to build EPRDB for representing and manipulating uncertain information.

2.1. Probabilistic values

The notion of probabilistic value is the fundamental concept for representing uncertain relational attribute values in EPRDB. The probabilistic value in [31] is adapted to data types in EPRDB and defined as follows.

Definition 1. Let τ be a data type and D be the domain of τ, a *probabilistic value* on the domain D of τ is a finite set of pairs $\{(v_1, [l_1, u_1]), \ldots, (v_m, [l_m, u_m])\}$, where $v_i \in D$ and $0 \le l_i \le u_i \le 1$, for every $i = 1, 2, \ldots, m$.

Informally, a probabilistic value $pv = \{(v_1, [l_1, u_1]), \ldots, (v_m, [l_m, u_m])\}$ assigns each element $v_i \in V = \{v_1, \ldots, v_m\}$ a probability interval $[l_i, u_i]$ to represent the uncertainty degree of v_i in V. Thus, a probabilistic value represents both the uncertainty of its value and the imprecision of the probability of that value. A probabilistic value $pv = \{(v_1, [l_1, u_1]), \ldots, (v_m, [l_m, u_m])\}$ corresponds with a probability distribution p over $V = \{v_1, \ldots, v_m\}$ such that $p(v_i) \in [l_i, u_i]$ and $\Sigma_{vi \in V}\, p(v_i) \le 1$.

We note that a probabilistic value can be denoted by $pv = \{(v, I) | v \in D, I = [l, u] \subseteq [0, 1]\}$.

Example 1. Suppose a patient's disease is diagnosed as cirrhosis with a probability between 0.3 and 0.5 or cholecystitis with a probability between 0.5 and 0.7. Then, this information may be represented by the probabilistic value {(cirrhosis, [0.3, 0.5]), (cholecystitis, [0.5, 0.7])}.

2.2. Probabilistic interpretation of binary relations on sets

The probabilistic interpretation of binary relations on sets is a proposed probabilistic measure for computing the probability of binary relations on sets of uncertain attribute values in EPRDB. The probabilistic interpretation of binary relations on sets is defined below.

Definition 2. Let U and V be value domains, A and B be subsets of U and V respectively, and θ be a binary relation in $\{=, \neq, \le, \ge, <, >\}$. The probabilistic interpretation of the relation $A\ \theta\ B$, denoted $Pr(A\ \theta\ B)$, is a value in [0, 1] that is defined by:

$$Pr(A\ \theta\ B) = \Sigma_{u \in A,\, v \in B}\, P(u\ \theta\ v),$$

where θ is assumed to be valid on $(U \times V)$ and $P(u\ \theta\ v)$ is the conditional probability of $u\ \theta\ v$ given $u \in A$ and $v \in B$.

We note that Definition 2 can be applied to elements of U and V because each element in a set also considered a subset of that set.

2.3. Probabilistic combination strategies

In many real situations, the probability of an event may not be defined or computed exactly [32]. Therefore, a probability interval can be used instead of a precise single probability value. Let two events e_1 and e_2 have probabilities in the intervals $[l_1, u_1]$ and $[l_2, u_2]$, respectively. Then, the probability intervals of the conjunction event $e_1 \wedge e_2$, disjunction event $e_1 \vee e_2$, and difference event $e_1 \wedge \neg e_2$ can be computed by alternative strategies. To combine probability intervals in EPRDB, we use the conjunction, disjunction, and difference strategies defined in [25] and [31] as in Table 1, where ⊗, ⊕, and ⊖ denote the conjunction, disjunction, and difference operators, respectively.

In the following sections, the notation $[l_1, u_1] \leq [l_2, u_2]$ is used to denote $l_1 \leq l_2$ and $u_1 \leq u_2$, whereas the notation $[l_1, u_1] \subseteq [l_2, u_2]$ is used to denote $l_2 \leq l_1$ and $u_1 \leq u_2$. In addition, a single probability value p can be treated as the probability interval $[p, p]$, the operations $p.[l, u]$ and $[l_1, u_1].[l_2, u_2]$ computed as $[p.l, p.u]$ and $[l_1.l_2, u_1.u_2]$, respectively.

Table 1. Definitions of probabilistic combination strategies

Strategy	Operators
Independence	$([l_1, u_1] \otimes_{in} [l_2, u_2]) = [l_1 . l_2, u_1 . u_2]$
	$([l_1, u_1] \oplus_{in} [l_2, u_2]) = [l_1 + l_2 - (l_1 . l_2), u_1 + u_2 - (u_1 . u_2)]$
	$([l_1, u_1] \ominus_{in} [l_2, u_2]) = [l_1 . (1 - u_2), u_1 . (1 - l_2)]$
Positive correlation (when e_1 implies e_2, or e_2 implies e_1)	$([l_1, u_1] \otimes_{pc} [l_2, u_2]) = [min(l_1, l_2), min(u_1, u_2)]$
	$([l_1, u_1] \oplus_{pc} [l_2, u_2]) = [max(l_1, l_2), max(u_1, u_2)]$
	$([l_1, u_1] \ominus_{pc} [l_2, u_2]) = [max(0, l_1 - u_2), max(0, u_1 - l_2)]$
Mutual exclusion (when e_1 and e_2 are mutually exclusive)	$([l_1, u_1] \otimes_{me} [l_2, u_2]) = [0, 0]$
	$([l_1, u_1] \oplus_{me} [l_2, u_2]) = [min(1, l_1 + l_2), min(1, u_1 + u_2)]$
	$([l_1, u_1] \ominus_{me} [l_2, u_2]) = [l_1, min(u_1, 1 - l_2)]$

2.4. Conjunction, disjunction, and difference of probabilistic values

The conjunction, disjunction, and difference of probabilistic values are used to combine the probability of the uncertain attribute values and tuple memberships in probabilistic relational algebraic operations on EPRDB. The conjunction, disjunction, and difference of probabilistic values are introduced in [26] and defined as follows.

Definition 3. Let pv_1 and pv_2 be two probabilistic values and ⊗ be a probabilistic conjunction strategy. The *conjunction* of pv_1 and pv_2 under ⊗, denoted by $pv_1 \otimes pv_2$, is the probabilistic value pv defined by $pv = \{(v, I_1 \otimes I_2) \mid (v, I_1) \in pv_1, (v, I_2) \in pv_2\}$.

Example 2. Let pv_1 = {(*hepatitis*, [0.4, 0.6]), (*cholecystitis*, [0.4, 0.6])} and pv_2 = {(*hepatitis*, [1.0, 1.0])} be probabilistic values, then $pv_1 \otimes_{in} pv_2$ under the independence probabilistic conjunction strategy is the probabilistic value pv = {(*hepatitis*, [0.4, 0.6])}.

Definition 4. Let pv_1 and pv_2 be two probabilistic values and ⊕ be a probabilistic disjunction strategy. The *disjunction* of pv_1 and pv_2 under ⊕, denoted by $pv_1 \oplus pv_2$, is the probabilistic value pv defined by $pv = \{(v, I_1) \mid (v, I_1) \in pv_1 \text{ and } \neg\exists I_2, (v, I_2) \in pv_2\} \cup \{(v, I_2) \mid (v, I_2) \in pv_2 \text{ and } \neg\exists I_1, (v, I_1) \in pv_1\} \cup \{(v, I_1 \oplus I_2) \mid (v, I_1) \in pv_1 \text{ and } (v, I_2) \in pv_2\}$.

Example 3. Let pv_1 = {(*hepatitis*, [0.2, 0.5]), (*cholecystitis*, [0.3, 0.6])} and pv_2 = {(*hepatitis*, [0.3, 0.5]), (*pancreatitis*, [0.2, 0.6])} be probabilistic values, then $pv_1 \oplus_{in} pv_2$ under the independence probabilistic disjunction strategy is the probabilistic value pv = {(*cholecystitis*, [0.3, 0.6]), (*pancreatitis*, [0.2, 0.6]), (*hepatitis*, [0.44, 0.75])}.

Definition 5. Let pv_1 and pv_2 be two probabilistic values and ⊖ be a probabilistic difference strategy. The *difference* of pv_1 and pv_2 under ⊖, denoted by $pt_1 \ominus pt_2$, is the probabilistic value pv defined by $pv = \{(v, I_1) \mid (v, I_1) \in pv_1 \text{ and } \neg\exists I_2, (v, I_2) \in pv_2\} \cup \{(v, I_1 \ominus I_2) \mid (v, I_1) \in pv_1 \text{ and } (v, I_2) \in pv_2\}$.

3. PROPOSED METHOD

The proposed EPRDB model including the data model and probabilistic relational algebra is defined and built by extending the CRDB model [1] using the probability theory presented above.

3.1. EPRDB data model

The EPRDB data model is a structure including fundamental concepts and data elements such as the schema, relation, database, functional dependency, and the schema key extended from the CRDB data model with probability data notions for representing uncertain information in the real world.

3.1.1. EPRDB schemas and relations

An EPRDB schema consists of a set of relational attributes that defines the sets of data tuples associated with probabilistic intervals. The EPRDB schema is an extension of the CRDB schema with the uncertain relational attribute value and tuple membership. The EPRDB schema is defined as follows.

Definition 6. An *EPRDB schema* is a pair $R = (U, \wp)$, where

1. $U = \{A_1, A_2, ..., A_k\}$ is a set of pairwise different attributes.
2. $\wp$ is a function that maps each $(pv_1, pv_2, ..., pv_k) \in \Im(D_1)\times\Im(D_2)\times\ldots\times\Im(D_k)$ to a subinterval of the interval $[0, 1]$, $\Im(D_i)$ is the set of all probabilistic values on the domain D_i of the attribute A_i, for $i = 1, \ldots, k$.

For simplicity, we can use the notation $R(U, \wp)$ and R to denote the schema $R = (U, \wp)$. The domain of an attribute A is denoted by $dom(A)$.

The EPRDB relation is extended from the CRDB relation and that of the data models [26], [28], [29] with uncertain attribute values and tuple memberships as the following definition.

Definition 7. Let $U = \{A_1, A_2, \ldots, A_k\}$ be a set of k pairwise different attributes. An *EPRDB relation* r over the schema $R(U, \wp)$ is a finite set $\{t_1, t_2,\ldots, t_n \mid \wp(t_i) = [\alpha_i, \beta_i] \subseteq [0, 1]\}$, where each $t_i = (pv_{i1}, pv_{i2}, \ldots, pv_{ik}) \in \Im(D_1)\times\Im(D_2)\times\ldots\times\Im(D_k)$, $\wp(t_i) = [\alpha_i, \beta_i]$ represents the uncertain membership degree of t_i in r, $\Im(D_j)$ is the set of all probabilistic values on the domain D_j of the attribute A_j for $j =1, 2,\ldots, k$, $i = 1, 2, \ldots, n$.

As in the CRDB model, each element $t_i = (pv_{i1}, pv_{i2}, \ldots, pv_{ik})$ in the relation r over $R(U, \wp)$ is also called a tuple on U. The probabilistic value $pv_{ij} = \{(v_{ij}, [l_{ij}, u_{ij}])| \; v_{ij} \in D_j, [l_{ij}, u_{ij}]) \subseteq [0, 1]\}$ represents the uncertain and imprecise value of the attribute A_j of the tuple t_i and $\wp(t_i) = [\alpha_i, \beta_i]$ represents the uncertain membership degree of the tuple t_i in r. We write $t_i.A_j$ or $t_i[A_j]$ to denote pv_{ij} and $[t_i]$ to replace $(V_{i1}, V_{i2}, \ldots, V_{ik})$, where $V_{ij} = \{v_{ij} \mid (v_{ij}, [l_{ij}, u_{ij}]) \in pv_{ij}\}$. For each set of attributes $L \subseteq \{A_1, A_2, ..., A_k\}$, the symbol $t_i[L]$ denotes the rest of the tuple t_i after eliminating the values of attributes not belonging to L. In addition, if we only care about a unique relation over a schema then we can unify the relation's name and its schema's name.

We note that in many real situations, the probability of an event may not be defined or computed exactly [32]. Then, a probability interval can be used instead of a precise single probability value. The probability interval represents an imprecise and proximate probability for the exact probability of the event. This is the reason for representing the uncertain degree of the relational attribute value and tuple membership in the EPRDB model by an interval probability instead of a precise single probability as in PRDB models [12]-[16], [22], [23], [28], [29]. In other words, the EPRDB model is an extension to overcome the limitations of the PRDB models that represent uncertain information by precise single probabilities.

Example 4. In the database about patients at the clinic of a hospital, a simple EPRDB relation, named PATIENT, over the EPRDB schema PATIENT({P_NAME, P_AGE, P_DISEASE, D_COST}, $\wp$) can be given as Table 2.

Table 2. Relation PATIENT

P_ID	P_NAME	P_AGE	P_DISEASE	D_COST	$\wp$
P202	George	{(72, [1, 1])}	{(lung cancer, [1, 1])}	{($35, [0.5, 0.5]), ($40, [0.5, 0.5])}	[1, 1]
P226	Mary	{(24, [0.5, 0.5]), (25, [0.5, 0.5])}	{(cirrhosis, [0.3, 0.5]), (hepatitis, [0.5, 0.7])}	{($10, [0.4, 0.6]), ($11, [0.4, 0.6])}	[0.9, 1]
P315	Blair	{(56, [1, 1])}	{(duodenitis, [0.5, 0.5]), (gastritis, [0.5, 0.5])}	{($6, [0.3, 0.6]), ($7, [0.4, 0.7])}	[0.8, 1]
P318	Selena	{(21, [1, 1])}	{(hepatitis, [0.6, 0.7]), cholecystitis, [0.3, 0.4]}	{($10, [0.5, 0.5]), ($11, [0.5, 0.5])}	[0.8, 0.9]
P424	Kate	{(18, [1, 1])}	{(bronchitis, [0.4, 0.5]), (angina, [0.5, 0.6])}	{($8, [0.3, 0.5]), ($9, [0.5, 0.7])}	[0.7, 0.8]
P523	Paul	{(56, [1, 1])}	{(duodenitis, [0.4, 0.5]), (gastritis, [0.5, 0.6])}	{($6, [0.3, 0.5]), ($7, [0.5, 0.7])}	[0.4, 0.5]

In the relation, the attributes P_ID, P_NAME, P_AGE, P_DISEASE, and D_COST describe the information about the identifier, name, age, disease, and daily treatment cost of each patient, respectively. In reality, while diagnosing, the doctors can be unsure of the disease of patients. Also, the daily treatment cost for patients is not sure even though the patients learn about their diseases. For instance, the information of the patient Blair, represented by the third tuple (i.e., t_3), says that Blair's age is 56, the patient's disease may be duodenitis or gastritis with a probability of 0.5, and Blair has to pay a daily treatment cost of $6 with a probability between 0.3 and 0.6 or $7 with a probability between 0.4 and 0.7. Meanwhile, the probability interval [0.8, 1] that is the value of $\wp(t_3)$ expresses the degree of uncertainty for the patient Blair (i.e., the third tuple) belonging to the relation PATIENT.

We note that, for simplicity, each probabilistic value $\{(v, [1, 1])\}$ will be represented as a single value v (such as probabilistic values for the attribute P_ID). Because if an attribute takes such a probabilistic value, then it only takes a value v with the probability of 1.0 (Definition 1). In other words, the attribute certainly takes the value v. The EPRDB relational database is a set of probabilistic relations over EPRDB schemas defined as follows.

Definition 8. An EPRDB relational database over a set of attributes is a set of EPRDB relations corresponding to the set of their EPRDB schemas.

3.1.2. EPRDB functional dependencies

Functional dependencies play an essential role in CRDB. The probabilistic functional dependent concept in EPRDB is an extension of that in CRDB [1], [2] with probabilistic interval valued attributes. We first define the probability measure to determine the equal degree of two probabilistic values of the same attribute for two different tuples in an EPRDB relation.

Definition 9. Let $R(U, \wp)$ be an EPRDB schema, r be a relation over R and t_1 and t_2 be two tuples in r, A be an attribute of U, and $\otimes$ be a probabilistic conjunction strategy. The *probability interval* for the values of the attribute A of two tuples t_1 and t_2 to be equal under $\otimes$, denoted by $p(t_1.A =_\otimes t_2.A)$, is $\bigoplus_{i=1}^{m}\bigoplus_{j=1}^{n}(([l_{1i}, u_{1i}] \otimes [l_{2j}, u_{2j}]).Pr(v_{1i} = v_{2j}))$, where $t_1.A = \{(v_{11}, [l_{11}, u_{11}]), \ldots, (v_{1m}, [l_{1m}, u_{1m}])\}$, $t_2.A = \{(v_{21}, [l_{21}, u_{21}]), \ldots, (v_{2n}, [l_{2n}, u_{2n}])\}$ and $\oplus$ is the mutual exclusion probabilistic disjunction operator.

The probabilistic functional dependency in EPRDB is an extension of the functional dependency in CRDB with uncertain and imprecise valued attribute sets as below.

Definition 10. Let $R(U, \wp)$ be an EPRDB schema, r be any relation over R, $\otimes$ be a probabilistic conjunction strategy, X and Y be two non-empty subsets of U. An *EPRDB functional dependency* of Y on X under $\otimes$, denoted by $X \rightarrow_\otimes Y$, holds if and only if

$\forall t_1, t_2 \in r$: $\otimes_{A\in X}\, p(t_1.A =_\otimes t_2.A) \leq \otimes_{A\in Y}\, p(t_1.A =_\otimes t_2.A)$.

It is easy to see that for every EPRDB schema $R(U, \wp)$, then $U \rightarrow_\otimes Y$ with $Y \subseteq U$ under all probabilistic conjunction strategies.

Example 5. In every relation r over the schema PATIENT with the set of attributes U = {P_ID, P_NAME, P_AGE, P_DISEASE, D_COST} in Example 4, the values of the attribute P_ID that describe the identifiers of patients are single and pairwise different. Thus, for two tuples $t_1, t_2 \in r$ and an attribute $A \in U$, then $p(t_1.A =_\otimes t_2.A) \geq 0$ and $p(t_1.\text{P\_ID} =_\otimes t_2.\text{P\_ID}) = 0$. So, $p(t_1.\text{P\_ID} =_\otimes t_2.\text{P\_ID}) \leq \otimes_{A\in Y}\, p(t_1.A =_\otimes t_2.A)$ with $Y \subseteq U$, by Definition 10, there is the EPRDB functional dependency P_ID $\rightarrow_\otimes Y$ in the schema PATIENT under all probabilistic conjunction strategies.

As in CRDB [1], [2], the keys of a schema in EPRDB are the basis for recognizing a tuple of a probabilistic relation. In the model and management systems of the classical relational database [1], [2], key attributes cannot take the null value. Similarly, in EPRDB, we assume that the value of each key attribute is always definite and unique. The concept of the key of EPRDB schemas is defined using the probabilistic functional dependency as follows.

Definition 11. Let $R(U, \wp)$ be an EPRDB schema, r be any relation over R, and $\otimes$ be a probabilistic conjunction strategy. A set of attributes $K \subseteq U$ is a *key* of R under $\otimes$ if the value of the attributes of K is definite and there is a probabilistic functional dependency $K \rightarrow_\otimes U$ such that there does not exist any proper subset of K holding this property.

Example 6. In the relation PATIENT above, if we assume that each patient has a unique identifier corresponding to the value of the attribute P_ID, then P_ID is a key of the schema PATIENT under all probabilistic conjunction strategies.

3.2. EPRDB algebra

As the CRDB algebra [1], [2], the EPRDB algebra is a set of basic probabilistic relational algebraic operations such as the selection, projection, Cartesian product, join, intersection, union, and difference. The EPRDB algebra or the probabilistic relational algebra is an extension of the CRDB algebra with uncertain attribute values and tuple memberships of relations to compute, manipulate, and query uncertain and imprecise information on the EPRDB data model.

3.2.1. Selection

The selection operation of EPRDB is an extension of that of CRDB with uncertain valued attributes and tuple memberships. Before defining the selection operation, we introduce the formal syntax and semantics of selection expressions and conditions.

Definition 12. Let R be an EPRDB schema and X be a set of relational tuple variables. Then selection expressions are inductively defined and have one of the following forms:

a. $x.A\ \theta\ v$, where $x \in X$, A is an attribute in R, θ is a binary relation from $\{=, \neq, \leq, \geq, <, >\}$, and v is a value in the domain of A.
b. $x.A_1 =_{\otimes} x.A_2$, where $x \in X$, A_1, and A_2 are two attributes in R, and $\otimes$ is a probabilistic conjunction strategy.
c. $\alpha \otimes \beta$, where α and β are selection expressions on the same relational tuple variable, and $\otimes$ is a probabilistic conjunction strategy.
d. $\alpha \oplus \beta$, where α and β are selection expressions on the same relational tuple variable, and $\oplus$ is a probabilistic disjunction strategy.

Definition 13. Let R be an EPRDB schema. Then selection conditions are inductively defined as follows:

– If α is a selection expression and $[l, u]$ is a subinterval of $[0, 1]$, then $(\alpha)[l, u]$ is a selection condition.
– If φ and ω are selection conditions on the same tuple variable, then $\neg\varphi$, $(\varphi \wedge \omega)$, $(\varphi \vee \omega)$ are selection conditions.

Example 7. Given the schema PATIENT in Example 4, the selection of "all patients who are over 50 years old with a probability of at least 0.9 or have lung cancer and pay the daily treatment cost not less than 35 USD with a probability between 0.4 and 0.6" can be done using the selection condition (x.P_AGE > 50)[0.9, 1] $\vee$ (x.P_DISEASE = lung cancer $\otimes$ x.D_COST $\geq$ 35)[0.4, 0.6].

The probabilistic interpretation (i.e., a probabilistic measure) of selection expressions in EPRDB is proposed and defined for computing and combining the probabilities of uncertain attribute values and tuple memberships of relations as below.

Definition 14. Let $R(U, \wp)$ be an EPRDB schema, r be a relation over R, x be a tuple variable, and t be a tuple in r. The *probabilistic interpretation of selection expressions* with respect to R, r and t, denoted by $Prob_{R,r,t}$, is the partial mapping from the set of all selection expressions to the set of all closed subintervals of [0, 1] that is inductively defined as follows:

a. $Prob_{R,r,t}(x.A\ \theta\ v) = (\bigoplus_{i=1}^{k}[l_i, u_i].Pr(v_i\ \theta\ v)).\wp(t)$, where $t.A = \{(v_1, [l_1, u_1]), \ldots, (v_k, [l_k, u_k])\}$ and $\oplus$ is the mutual exclusion probabilistic disjunction operator.
b. $Prob_{R,r,t}(x.A_1 =_{\otimes} x.A_2) = (\bigoplus_{i=1}^{m}\bigoplus_{j=1}^{n}([l_{1i}, u_{1i}] \otimes [l_{2j}, u_{2j}]).Pr(v_{1i} = v_{2j})).\wp(t)$, where $t.A_1 = \{(v_{11}, [l_{11}, u_{11}]), \ldots, (v_{1m}, [l_{1m}, u_{1m}])\}$, $t.A_2 = \{(v_{21}, [l_{21}, u_{21}]), \ldots, (v_{2n}, [l_{2n}, u_{2n}])\}$ and $\oplus$ is the mutual exclusion probabilistic disjunction operator.
c. $Prob_{R,r,t}(\alpha \otimes \beta) = Prob_{R,r,t}(\alpha) \otimes Prob_{R,r,t}(\beta)$.
d. $Prob_{R,r,t}(\alpha \oplus \beta) = Prob_{R,r,t}(\alpha) \oplus Prob_{R,r,t}(\beta)$.

We note that the mutual exclusion probabilistic disjunction operator $\oplus_{me}$ is used in items 1 and 2 of Definition 14 because the intervals $[l_1, u_1], \ldots, [l_k, u_k]$ represent a probability distribution function over $\{v_1, \ldots, v_k\}$, likewise for $[l_{11}, u_{11}], \ldots, [l_{1m}, u_{1m}]$ and $[l_{21}, u_{21}], \ldots, [l_{2n}, u_{2n}]$. Intuitively, $Prob_{R,r,t}(x.A\ \theta\ v)$ is the probability interval for the attribute A of the tuple t with the uncertainty membership degree in r being $\wp(t)$ having a value v_i such that $v_i\ \theta\ v$, while $Prob_{R,r,t}(x.A_1 =_{\otimes} x.A_2)$ is the probability interval for the attributes A_1 and A_2 of the tuple t with the uncertainty membership degree in r being $\wp(t)$ having values v_{1i} and v_{2j}, respectively, such that $v_{1i} = v_{2j}$.

Example 8. Let R denote the schema PATIENT and r denote the relation PATIENT in Example 4. Consider the second tuple in r, denoted by t_2. We have

$Prob_{R,r,t2}(x.\text{P\_DISEASE} = \text{hepatitis})$
$= ([0.3, 0.5].Pr(\text{cirrhosis} = \text{hepatitis}) \oplus_{me} [0.5, 0.7].Pr(\text{hepatitis} = \text{hepatitis})).\wp(t_2)$
$= ([0.3, 0.5]\times 0.0 \oplus_{me} [0.5, 0.7]\times 1.0).[0.9, 1] = ([0, 0] \oplus_{me} [0.5, 0.7]).[0.9, 1]$
$= [0.5, 0.7].[0.9, 1] = [0.45, 0.7]$.

The satisfaction of the selection conditions for the tuples of relations in EPRDB is an extension of that in CRDB with probability intervals as below.

Definition 15. Let R be an EPRDB schema, r be a relation over R, and $t \in r$. The *satisfaction of selection conditions* under $Prob_{R,r,t}$ is defined as follows:

a. $Prob_{R,r,t} \vDash (\alpha)[l, u]$ if and only if (iff) $Prob_{R,r,t}(\alpha) \subseteq [l, u]$.
b. $Prob_{R,r,t} \vDash \neg\varphi$ iff $Prob_{R,r,t} \vDash \varphi$ does not hold.
c. $Prob_{R,r,t} \vDash \varphi \wedge \omega$ iff $Prob_{R,r,t} \vDash \varphi$ and $Prob_{R,r,t} \vDash \omega$.
d. $Prob_{R,r,t} \vDash \varphi \vee \omega$ iff $Prob_{R,r,t} \vDash \varphi$ or $Prob_{R,r,t} \vDash \omega$.

Now, the selection operation on relations for querying uncertain information in EPRDB is defined as follows.

Definition 16. Let R be an EPRDB schema, r be a relation over R, and φ be a selection condition over a tuple variable x. The selection on r with respect to φ, denoted by $\sigma_\varphi(r)$, is the relation $r^* = \{t \in r \mid prob_{R,r,t} \vDash \varphi\}$ over R, including all satisfied tuples of the selection condition φ.

Example 9. Let r denote the relation PATIENT in Example 4, and R denote its schema. The query "Find all patients who are over 20 years old with a probability of at least 0.8 and have hepatitis, and pay the daily treatment cost not less than 10 USD with a probability between 0.3 and 0.7" can be done by the selection operation σ_φ(PATIENT), where $\varphi = (x.\text{P\_AGE} > 20)[0.8, 1] \wedge (x.\text{P\_DISEASE} = \text{hepatitis} \otimes_{in} x.\text{D\_COST} \geq 10)[0.3, 0.7]$.

There are two patients denoted by the second and fourth tuples (t_2 and t_4) of the relation PATIENT in Example 4 satisfied φ, because:

For t_2, we have $Prob_{R,r,t2}(x.\text{P\_AGE} > 20) = ([0.5, 0.5]\times Pr(24 > 20) \oplus_{me} [0.5, 0.5]\times Pr(25 > 20))$. $\wp(t_2) = ([0.5, 0.5]\times 1.0 \oplus_{me} [0.5, 0.5]\times 1.0).[0.9, 1] = [1.0, 1.0].[0.9, 1] = [0.9, 1] \subseteq [0.8, 1]$.

$Prob_{R,r,t2}(x.\text{D\_COST} \geq 10) = ([0.4, 0.6]\times Pr(10 \geq 10) \oplus_{me} [0.4, 0.6]\times Pr(11 \geq 10))$. $\wp(t_2) = ([0.4, 0.6]\times 1.0 \oplus_{me} [0.4, 0.6]\times 1.0).[0.9, 1] = ([0.4, 0.6] \oplus_{me} [0.4, 0.6]).[0.9, 1] = [0.8, 1].[0.9, 1] = [0.72, 1]$.

From the result of the computation in Example 8, we have $Prob_{R,r,t2}(x.\text{P\_DISEASE} = \text{hepatitis} \otimes_{in} x.\text{D\_COST} \geq 10) = [0.45, 0.7] \otimes_{in} [0.72, 1] = [0.324, 0.7] \subseteq [0.3, 0.7]$. Thus, t_2 satisfies φ.

For t_4, we have $Prob_{R,r,t_4}(x.\text{P\_AGE} > 20) = ([1.0, 1.0]\times Pr(21 > 20))$. $\wp(t_4) = ([1, 1]\times 1.0).[0.8, 0.9] = [0.8, 0.9] \subseteq [0.8, 1]$.

$Prob_{R,r,t_4}(x.\text{P\_DISEASE} = \text{hepatitis}) = ([0.6, 0.7].Pr(\text{hepatitis} = \text{hepatitis}) \oplus_{me} [0.3, 0.4].Pr(\text{cholecystitis} = \text{hepatitis}))$. $\wp(t_4) = ([0.6, 0.7]\times 1.0 \oplus_{me} [0.3, 0.4]\times 0.0).[0.8, 0.9] = ([0.6, 0.7] \oplus_{me} [0, 0]).[0.8, 0.9] = [0.6, 0.7].[0.8, 0.9] = [0.48, 0.63]$.

$Prob_{R,r,t4}(x.\text{D\_COST} \geq 10) = ([0.5, 0.5]\times Pr(10 \geq 10) \oplus_{me} [0.5, 0.5]\times Pr(11 \geq 10))$. $\wp(t_4) = ([0.5, 0.5]\times 1.0 \oplus_{me} [0.5, 0.5]\times 1.0).[0.8, 0.9] = ([0.5, 0.5] \oplus_{me} [0.5, 0.5]).[0.8, 0.9] = [1, 1].[0.8, 0.9] = [0.8, 0.9]$.

Hence, we have $Prob_{R,r,t_4}(x.\text{P\_DISEASE} = \text{hepatitis} \otimes_{in} x.\text{D\_COST} \geq 10) = [0.48, 0.63] \otimes_{in} [0.8, 0.9] = [0.384, 0.567] \subseteq [0.3, 0.7]$. Thus, t_4 satisfies φ.

For the other tuples, one has $Prob_{R,r,ti}(x.\text{P\_DISEASE} = \text{hepatitis} \otimes_{in} x.\text{D\_COST} \geq 10) = [0, 0] \not\subset [0.3, 0.7]$, $\forall i \neq 2, 4$. Thus, the result of the query is shown in Table 3.

Table 3. Relation σ_φ(PATIENT)

P_ID	P_NAME	P_AGE	P_DISEASE	D_COST	$\wp$
P226	Mary	{(24, [0.5, 0.5]), (25, [0.5, 0.5])}	{(cirrhosis, [0.3, 0.5]), (hepatitis, [0.5, 0.7])}	{($10, [0.4, 0.6]), ($11, [0.4, 0.6])}	[0.9, 1]
P318	Selena	{(21, [1, 1])}	{(hepatitis, [0.6, 0.7]), cholecystitis, [0.3, 0.4]}	{($10, [0.5, 0.5]), ($11, [0.5, 0.5])}	[0.8, 0.9]

3.2.2. Projection

The projection of an EPRDB relation on a set of attributes is an extension of that of a CRDB relation with uncertain attribute values and tuple memberships such that the projected tuples having the same value should be coalesced into a tuple in the result relation by a probabilistic disjunction strategy. The projection operation of an EPRDB relation is defined as follows.

Definition 17. Let $R(U, \wp)$ be an EPRDB schema, r be a relation over R, L be a subset of attributes of U, $\oplus$ be a probabilistic disjunction strategy. The *projection* of r on L under $\oplus$, denoted by $\Pi_{L\oplus}(r)$, is the relation r^* over the schema R^* determined by:

1. $R^* = (L, \wp^*)$, where $\wp^*$ is the mapping from $\Im(D_1)\times\Im(D_2)\times\ldots\times\Im(D_m)$ to the set of all intervals on [0, 1], $m = |L|$, D_i is the domain of $A_i \in L$, $\Im(D_i)$ is the set of all probabilistic values on D_i, for $i = 1, \ldots, m$.
2. $r^* = \{t^* \mid t^*.A = u.A \oplus\ldots\oplus w.A,\ \wp^*(t^*) = \wp(u) \oplus\ldots\oplus \wp(w),\ \forall A \in L, \exists\, u, \ldots, w \in r$ such that $[u[L]] = \ldots = [w[L]]\}$.

Example 10. Consider the relation PATIENT over the schema PATIENT({P_ID, P_NAME, P_AGE, P_DISEASE, D_COST}, $\wp$) as in Table 2, then the projection of it on the set of the attributes L = {P_AGE, P_DISEASE, D_COST} under $\oplus_{in}$ is the relation $\Pi_{L\oplus in}$(PATIENT) over the schema R^*({P_AGE, P_DISEASE, D_COST}, $\wp^*$) computed as in Table 4.

Table 4. Relation $\Pi_{\{P\_AGE,\ P\_DISEASE,\ D\_COST\}\oplus_{in}}$(PATIENT)

P_AGE	P_DISEASE	D_COST	$\wp^*$
{(72, [1, 1])}	{(lung cancer, [1, 1])}	{($35, [0.5, 0.5]), ($40, [0.5, 0.5])}	[1, 1]
{(24, [0.5, 0.5]), (25, [0.5, 0.5])}	{(cirrhosis, [0.3, 0.5]), (hepatitis, [0.5, 0.7])}	{($10, [0.4, 0.6]), ($11, [0.4, 0.6])}	[0.9, 1]
{(21, [1, 1])}	{(hepatitis, [0.6, 0.7]), cholecystitis, [0.3, 0.4]}	{($10, [0.5, 0.5]), ($11, [0.5, 0.5])}	[0.8, 0.9]
{(18, [1, 1])}	{(bronchitis, [0.4, 0.5]), (angina, [0.5, 0.6])}	{($8, [0.3, 0.5]), ($9, [0.5, 0.7])}	[0.7, 0.8]
{(56, [1, 1])}	{(duodenitis, [0.7, 0.75]), (gastritis, [0.75, 0.8])}	{($6, [0.51, 0.8]), ($7, [0.7, 0.91])}	[0.88, 1]

Note that in the relation PATIENT, we have $[t_3[L]] = [t_6[L]]$, thus two tuples t_3 and t_6 are projected on L and coalesced into the tuple t_5 under the independence probabilistic disjunction strategy $\oplus_{in}$ in Table 4.

3.2.3. Cartesian product

For the Cartesian product of two EPRDB relations, as in CRDB, we assume the set of attributes of their schemas are disjoint, and every k-tuple t = (pv_1, pv_2, …, pv_k) of probabilistic values is an unordered list. The Cartesian product of two EPRDB relations is extended from that of two CRDB relations with uncertain attribute values and tuple memberships as follows.

Definition 18. Let U_1, U_2 be two sets of attributes that do not have any common element, $R_1(U_1, \wp_1)$, $R_2(U_2, \wp_2)$ be two EPRDB schemas, r_1, r_2 be two relations over R_1 and R_2, respectively and $\otimes$ be a probabilistic conjunction strategy. The *Cartesian product* of r_1 and r_2, denoted by $r_1 \times_{\otimes} r_2$, is the relation r over R, determined by:

1. $R = (\boldsymbol{U}, \wp)$, where $\boldsymbol{U} = \boldsymbol{U_1} \cup \boldsymbol{U_2}$, $\wp$ is the mapping from $\Im(D_1)\times\Im(D_2)\times\ldots\times\Im(D_n)$ to the set of all intervals on [0, 1], $n = |\boldsymbol{U}|$, D_i is the domain of $A_i \in \boldsymbol{U}$, and $\Im(D_i)$ is the set of all probabilistic values on D_i, for $i = 1, \ldots, n$.
2. $r = \{t \mid t.A = t_1.A$ if $A \in \boldsymbol{U_1}$, $t.A = t_2.A$ if $A \in \boldsymbol{U_2}$, $t_1 \in r_1$, $t_2 \in r_2$, $\wp(t) = \wp_1(t_1) \otimes \wp_2(t_2)\}$.

3.2.4. Join

The join of two EPRDB relations is an extension of the natural join of two CRDB relations with probabilistic values and uncertain tuple memberships as the following definition.

Definition 19. Let U_1 and U_2 be two sets of attributes such that if they have the same name attributes, respectively, in those two sets, then such attributes have the same value domain. Let $R_1(U_1, \wp_1)$ and $R_2(U_2, \wp_2)$ be two EPRDB schemas, r_1 and r_2 be two relations over R_1 and R_2, respectively, and $\otimes$ be a probabilistic conjunction strategy. The *join* of r_1 and r_2 under $\otimes$, denoted by $r_1 \bowtie_{\otimes} r_2$, is the relation r over the schema R, determined by:

1. $R = (U, \wp)$, where $U = U_1 \cup U_2$, $\wp$ is the mapping from $\Im(D_1)\times\Im(D_2)\times\ldots\times\Im(D_n)$ to the set of all intervals on [0, 1], $n = |U|$, D_i is the domain of $A_i \in U$, and $\Im(D_i)$ is the set of all probabilistic values on D_i, for $i = 1, \ldots, n$.
2. $r = \{t \mid t.A = t_1.A$ if $A \in U_1 - U_2$, $t.A = t_2.A$ if $A \in U_2 - U_1$, $t.A = t_1.A \otimes t_2.A$ if $A \in U_1 \cap U_2$, $\wp(t) = \wp_1(t_1) \otimes \wp_2(t_2)$, $t_1 \in r_1$, $t_2 \in r_2\}$.

Example 11. Given two probabilistic relations $PATIENT_1$ and $PATIENT_2$ over two EPRDB schemas $PATIENT_1$({P_ID, P_DISEASE}, $\wp_1$) and $PATIENT_2$({P_NAME, P_DISEASE}, $\wp_2$), respectively as in Tables 5 and 6, then the result of the join of them under the probabilistic conjunction strategy $\otimes_{in}$ is the relation $PATIENT_1 \bowtie_{\otimes in}$ $PATIENT_2$ over the schema $PATIENT_1 \bowtie_{\otimes in}$ $PATIENT_2$(P_ID, P_NAME, P_DISEASE}, $\wp$) computed as in Table 7.

Table 5. Relation $PATIENT_1$

P_ID	P_DISEASE	$\wp_1$
P521	{(bronchitis, [0.2, 0.4]), (bronchiectasis, [0.6, 0.8])}	[0.9, 1]
P628	{(cholecystitis, [0.5, 0.7]), (gallstone, [0.3, 0.5])}	[0.8, 0.9]

Table 6. Relation $PATIENT_2$

P_NAME	P_DISEASE	$\wp_2$
Peter	{(bronchiectasis, [1, 1])}	[1, 1]
Alice	{(cholecystitis, [0.4, 0.5]), (cirrhosis, [0.5, 0.6])}	[0.9, 1]

Table 7. Relation $\text{PATIENT}_1 \bowtie_{\otimes_{in}} \text{PATIENT}_2$

P_ID	P_NAME	P_DISEASE	℘
P521	Peter	{(bronchiectasis, [0.6, 0.8)}	[0.9, 1]
P628	Alice	{(cholecystitis, [0.2, 0.35])}	[0.72, 0.9]

3.2.5. Intersection, union and difference

The intersection, union, and difference of two EPRDB relations over the same schema are EPRDB relations over that schema, where two tuples that have the same key respectively of those two relations, should be coalesced into a tuple in the result relation by a probabilistic combination strategy. Here, two tuples have the same key value like two identical tuples in CRDB. Thus, the operations are the extensions of the intersection, union, and difference of two CRDB relations with probabilistic attribute values and uncertain tuple memberships. The intersection, union, and difference of two EPRDB relations are defined as follows.

Definition 20. Let $R(U, \wp)$ be an EPRDB schema, r_1 and r_2 be two relations over R, K be a key of R, and $\otimes$ be a probabilistic conjunction strategy. The *intersection* of r_1 and r_2 under $\otimes$, denoted by $r_1 \cap_{\otimes} r_2$, is the EPRDB relation r over R defined by $r = \{t \mid t.A = t_1.A \otimes t_2.A, A \in U, \wp(t) = \wp(t_1) \otimes \wp(t_2), t_1 \in r_1, t_2 \in r_2,$ such that $t_1[K] = t_2[K]\}$.

We note that the value of each key attribute is definite under Definition 11. Thus, the notation $t_1[K] = t_2[K]$ can be used in Definition 20. Moreover, we can uniquely determine a tuple of a relation under every key of the relation. So, the result relation is unique under all the keys.

Example 12. Given two EPRDB relations DIAGNOSE_1 and DIAGNOSE_2 over the same schema DIAGNOSE({P_ID, D_ID, P_DISEASE, D_COST}, ℘) as in Tables 8 and 9, where {P_ID, D_ID} is the key of this schema. Then, the intersection of DIAGNOSE_1 and DIAGNOSE_2 under $\otimes_{in}$ is the relation $\text{DIAGNOSE}_1 \cap_{\otimes_{in}} \text{DIAGNOSE}_2$ computed as in Table 10.

Table 8. Relation DIAGNOSE_1

P_ID	D_ ID	P_DISEASE	D_COST	℘
P226	D014	{(lung cancer, [0.3, 0.6]), (tuberculosis, [0.4, 0.7])}	{($30, [0.3, 0.4]), ($35, [0.6, 0.7])}	[0.8, 0.9]
P255	D020	{(hepatitis, [0.3, 0.8]), (pancreatitis, [0.2, 0.7])}	{($8, [0.6, 1])}	[0.9, 1]

Table 9. Relation DIAGNOSE_2

P_ID	D_ ID	P_DISEASE	D_COST	℘
P228	D016	{(lung cancer, [1, 1])}	{($30, [1, 1])}	[1, 1]
P255	D020	{(hepatitis, [0.4, 0.8]), (cholecystitis, [0.2, 0.6])}	{($7, [0.2, 0.4]), ($8, [0.4, 0.8])}	[0.8, 1]
P262	D022	{(dyspepsia, [1, 1])}	{($5, [1, 1])}	[1, 1]

Table 10. Relation $\text{DIAGNOSE}_1 \cap_{\otimes_{in}} \text{DIAGNOSE}_2$

P_ID	D_ ID	P_DISEASE	D_COST	℘
P255	D020	{(hepatitis, [0.12, 0.64])}	{($8, [0.24, 0.8])}	[0.72, 1]

We note that the tuple t_2 in Table 8 and the tuple t_2 in Table 9 have the same key value coalesced into the tuple t_1 under $\otimes_{in}$ in Table 10.

Definition 21. Let $R(U, \wp)$ be an EPRDB schema, r_1 and r_2 be two relations over R, K be a key of R, $\oplus$ be a probabilistic disjunction strategy. The *union* of r_1 and r_2 under $\oplus$, denoted by $r_1 \cup_{\oplus} r_2$, is the EPRDB relation r over R defined by $r = \{t \mid t = t_1 \in r_1$ such that $\forall t_2 \in r_2, t_1[K] \neq t_2[K], \wp(t) = \wp(t_1)\} \cup \{t \mid t = t_2 \in r_2$ such that $\forall t_1 \in r_1, t_2[K] \neq t_1[K], \wp(t) = \wp(t_2)\} \cup \{t \mid t.A = t_1.A \oplus t_2.A, A \in U, \wp(t) = \wp(t_1) \oplus \wp(t_2), t_1 \in r_1, t_2 \in r_2,$ such that $t_1[K] = t_2[K]\}$.

Definition 22. Let $R(U, \wp)$ be an EPRDB schema, r_1 and r_2 be two relations over R, K be a key of R, and $\ominus$ be a probabilistic difference strategy. The *difference* of r_1 and r_2 under $\ominus$, denoted by $r_1 -_{\ominus} r_2$, is the EPRDB relation r over R defined by $r = \{t \mid t = t_1 \in r_1$ such that $\forall t_2 \in r_2, t_1[K] \neq t_2[K], \wp(t) = \wp(t_1)\} \cup \{t \mid t.A = t_1.A \ominus t_2.A, \wp(t) = \wp(t_1) \ominus \wp(t_2), t_1 \in r_1, t_2 \in r_2, A \in U$ such that $t_1[K] = t_2[K]\}$.

We can see that, as in Definition 20, the result relation computed by the definitions 21 and 22 does not depend on choosing the key of its schema.

3.2.6. Property of algebraic operations

The basic properties of the EPRDB algebraic operations are the extensions of those of the CRDB algebraic operations with uncertain attribute values and tuple memberships. These properties say that the EPRDB model is sound and coherent.

Proposition 1. Let R be an EPRDB schema, r be a relation over R, and φ and ω be two selection conditions. Then,

$$\sigma_{\varphi}(\sigma_{\omega}(r)) = \sigma_{\omega}(\sigma_{\varphi}(r)) \tag{1}$$

Proof: Let $s = \sigma_{\omega}(r)$. By Definition 15 and 16, we have

$\sigma_{\varphi}(\sigma_{\omega}(r)) = \{t \in s \mid Prob_{R,s,t} \vDash \varphi\}$

$= \{t \in r \mid (Prob_{R,r,t} \vDash \omega) \wedge (Prob_{R,s,t} \vDash \varphi)\}$

$= \{t \in r \mid (Prob_{R,r,t} \vDash \omega) \wedge (Prob_{R,r,t} \vDash \varphi)\}$

$= \{t \in r \mid Prob_{R,r,t} \vDash \varphi \wedge \omega\} = \sigma_{\varphi \wedge \omega}(r)$.

Thus, the equation $\sigma_{\varphi}(\sigma_{\omega}(r)) = \sigma_{\varphi \wedge \omega}(r)$ is proven. The equation $\sigma_{\omega}(\sigma_{\varphi}(r)) = \sigma_{\omega \wedge \varphi}(r)$ is similarly proven, since $\omega \wedge \varphi \Leftrightarrow \varphi \wedge \omega$. So, Proposition 1 is proven.

Proposition 2. Let R be an EPRDB schema, r be a relation over R, and $\oplus$ be a probabilistic disjunction strategy, A and B be two subsets of attributes of R, $A \subseteq B$. Then,

$$\Pi_{A\oplus}(\Pi_{B\oplus}(r)) = \Pi_{A\oplus}(r) \tag{2}$$

Proof: Because $A \subseteq B$, so $A \cap B = A$. From Definition 17, it is easy to see that the sides of (2) are the relations over the same schema with the set of attributes $A \cap B = A$. By the property of the projection of the classical relations, Definition 7, and Definition 17, it follows that two classical sets of tuples which are collected respectively from two relations $\Pi_{A\oplus}(\Pi_{B\oplus}(r))$ and $\Pi_{A\cap B\oplus}(r)$ are the same. Moreover, by Definition 17, the probabilistic disjunction in two sides of (2) is executed on the same set of probabilistic intervals of the membership degrees of tuples of r. From that $\Pi_{A\oplus}(\Pi_{B\oplus}(r)) = \Pi_{A\cap B\oplus}(r) = \Pi_{A\oplus}(r)$ under the probabilistic disjunction strategy $\oplus$. Thus, the equation (2) is proven.

Proposition 3. Let R_1, R_2, and R_3 be the EPRDB schemas such that if they have the same name attributes, then such attributes have the same value domain, r_1, r_2, and r_3 be relations over R_1, R_2, and R_3, respectively, $\otimes$ be a probabilistic conjunction strategy. Then,

$$r_1 \bowtie_{\otimes} r_2 = r_2 \bowtie_{\otimes} r_1 \tag{3}$$

$$(r_1 \bowtie_{\otimes} r_2) \bowtie_{\otimes} r_3 = r_1 \bowtie_{\otimes} (r_2 \bowtie_{\otimes} r_3) \tag{4}$$

In (3) and (4) say that the join of EPRDB relations is commutative and associative.

Proof: It is easy to see that $r_1 \bowtie_{\otimes} r_2$ and $r_2 \bowtie_{\otimes} r_1$ are two relations over the same schema. By the property of the join of the classical relations, Definition 7, and Definition 19, it follows that two classical sets of tuples which are collected respectively from two relations $r_1 \bowtie_{\otimes} r_2$ and $r_2 \bowtie_{\otimes} r_1$ are the same. By Definition 3, the conjunction of probabilistic values is commutative (due to the commutativity of probabilistic conjunction strategies). Also, the conjunction of probabilistic intervals of the membership degrees of tuples is commutative. From that leading the join of tuples of the probabilistic relations and r_1 and r_2 has commutativity. So, by Definition 19, it follows that $r_1 \bowtie_{\otimes} r_2 = r_2 \bowtie_{\otimes} r_1$ under the probabilistic conjunction strategy $\otimes$. Thus, in (3) is proven.

By Definition 19, the results of two sides in (4) are the relations over the same schema. By the property of the join of the classical relations, Definition 7, and Definition 19, it follows that two classical sets of tuples which are collected respectively from two relations $(r_1 \bowtie_{\otimes} r_2) \bowtie_{\otimes} r_3$ and $r_1 \bowtie_{\otimes} (r_2 \bowtie_{\otimes} r_3)$ are the same. By Definition 3, the conjunction of probabilistic values is associative. Also, the conjunction of probabilistic intervals of the membership degrees of tuples is associative. By Definition 19 and from the associativity of the classical relational natural join, it follows that the join of EPRDB relations is associative.

Thus, it results in $(r_1 \bowtie_\otimes r_2) \bowtie_\otimes r_3 = r_1 \bowtie_\otimes (r_2 \bowtie_\otimes r_3)$ under the probabilistic conjunction strategy ⊗. in (4) is proven.

Because the Cartesian product (Definition 18) is a particular case of the join, it yields the straight result of Proposition 3.

Corollary 1. Let R_1, R_2, and R_3 be EPRDB schemas such that they do not have the same name attributes, r_1, r_2, and r_3 be relations over R_1, R_2, and R_3, respectively and ⊗ be a probabilistic conjunction strategy. Then,

$$r_1 \times_\otimes r_2 = r_2 \times_\otimes r_1 \quad (5)$$

$$(r_1 \times_\otimes r_2) \times_\otimes r_3 = r_1 \times_\otimes (r_2 \times_\otimes r_3) \quad (6)$$

Proposition 4. Let R be an EPRDB schema, r_1, r_2, and r_3 be relations over R. Let ⊗/⊕ be a probabilistic conjunction/disjunction strategy. Then,

$$r_1 \cap_\otimes r_2 = r_2 \cap_\otimes r_1 \quad (7)$$

$$(r_1 \cap_\otimes r_2) \cap_\otimes r_3 = r_1 \cap_\otimes (r_2 \cap_\otimes r_3) \quad (8)$$

$$r_1 \cup_\oplus r_2 = r_2 \cup_\oplus r_1 \quad (9)$$

$$(r_1 \cup_\oplus r_2) \cup_\oplus r_3 = r_1 \cup_\oplus (r_2 \cup_\oplus r_3) \quad (10)$$

In (7), (8), (9), and (10) say that the intersection and union of relations in EPRDB are commutative and associative.

Proof: From the commutativity and associativity of the probabilistic conjunction strategies, it follows that the conjunction of probabilistic values has the commutativity and associativity (Definition 3). Also, the conjunction of probabilistic intervals of the membership degrees of tuples has commutativity and associativity. So, the intersection of EPRDB relations r_1, r_2, and r_3 under the probabilistic conjunction strategy ⊗ and every chosen key also has commutativity and associativity. From that, by Definition 20, we have in (7) and (8).

From the commutativity and associativity of the probabilistic disjunction strategies, it follows that the disjunction of probabilistic values has the commutativity and associativity (Definition 4). Also, the disjunction of probabilistic intervals of the membership degrees of tuples has commutativity and associativity. So, the union of EPRDB relations r_1, r_2, and r_3 under the probabilistic disjunction strategy ⊕ and every chosen key also has commutativity and associativity. From that, by Definition 21, we have in (9) and (10).

For ending this section, we note that the computing complexity of EPRDB algebraic operations is a polynomial under the size of probabilistic relations. Moreover, EPRDB algebraic operations are as effective as CRDB algebraic operations. Indeed, regarding the selection operation, since the computation time that a tuple holds or does not hold a selection condition is bounded above by some constant under the maximum cardinality of probabilistic values representing relational attribute values in the selection condition (Definition 14 and 15). Then, the cost for the selection of each tuple in an EPRDB relation (Definition 16) also is some constant (i.e. O(1)). Consequently, the computing time complexity of the selection operation on an EPRDB relation having n tuples is $O(n)$. Similarly, the computing time complexity of Cartesian product and join operations on two EPRDB relations having n and m tuples is $O(nm)$. Thus, we can say that the performance of EPRDB model in computing and manipulating uncertain information is good and can apply it in practice.

4. RESULTS AND DISCUSSION

As presented in previous sections, we can see that the EPRDB model is an extension of the CRDB model and both the first and second classes of PRDB models with interval probability valued attributes and tuples. Moreover, the EPRDB model also has the capability of manipulating uncertain data more effectively than some PRDB models in the second class, such as [24], [25]. A more detailed discussion of the obtained results is as below.

4.1. Extension of EPRDB in representing and handling data

As introduced above, there are two main classes of the PRDB models corresponding to two uncertain data levels in the relations that can be extended from the CRDB model. The first class, denoted by C-1PRDB, defines a probabilistic relation as a set of tuples whose membership degree is represented by a single probability value or a probability subinterval of [0, 1] and attribute associated with a unique value, such as [12]-[20]. In other words, the C-1PRDB models are the extensions of the CRDB model with uncertain membership degree of relational tuples. The C-1PRDB algebraic operations are defined by directly extending the CRDB algebraic operations based on computing and combining single probability values or intervals of tuples in the C-1PRDB relations. The limitation of the C-1PRDB models is that they cannot represent and deal with the uncertainty of attribute values of relations.

The second class, denoted by C-2PRDB, defines a probabilistic relation as a set of tuples whose membership is certain and attribute associated with a single probability value, such as [3], [9], [22], [23] or a pair of lower and upper-bound probability distribution functions α, β on a value set V represented by $\langle V, \alpha, \beta \rangle$, such as [24], [25] or a distribution of probability intervals represented by $\{(v_1, [l_1, u_1]),..., (v_m, [l_m, u_m])\}$, such as [26], [27]. In other words, the C-2PRDB models are the extensions of the CRDB model with uncertain relational attribute values. The C-2PRDB algebraic operations are defined by extending the CRDB algebraic operations and employing operators on single probabilities or probability distributions for computing and combining probabilities of attribute values in the C-2PRDB relations. The limitation of the C-2PRDB models is that they cannot represent and deal with the uncertainty of the tuple membership of relations.

In practice, both the relational attribute value and tuple membership can be uncertain. However, the C-1PRDB models as well as the C-2PRDB ones cannot represent and deal with both the uncertain relational attribute value and tuple membership. To overcome the limitation, the models [28], [29] respectively extended the CRDB model such that relational tuples are associated with probabilistic intervals and attribute values with probability distributions represented by $(v_1, p_1),..., (v_m, p_m)\}$ or uniform distributions represented by $\{(v_1, 1/m), ... , (v_m, 1/m)\}$. The algebraic operations on the models are defined by extending the CRDB algebraic operations and employing operators on probability distributions or uniform distributions for manipulating and combining probabilities of tuples and attribute values of relations. However, the shortcoming of the models [28], [29] is that, in some real situations, we cannot know definitely the probability p_i or a uniform probability distribution for the value v_i.

To overcome the restriction of the C-1PRDB and C-2PRDB models, and the models [28], [29], the proposed EPRDB model extends the CRDB model with two uncertain data levels of the attribute value and tuple membership of probabilistic relations. In the EPRDB model, the uncertain relational tuple membership degree is represented by an interval probability of the interval [0, 1] and the relational attribute is represented by an interval probability distribution $\{(v_1, [l_1, u_1]), …, (v_m, [l_m, u_m])\}$. The EPRDB algebraic operations are defined by extending the CRDB algebraic operations using the probabilistic interpretations of binary relations on sets and the combination strategies of probabilistic intervals of attribute values and tuples in the EPRDB relations.

It is easy to see that a probability distribution $\{(v_1, p_1),..., (v_m, p_m)\}$ or a uniform distribution $\{(v_1, 1/m), ... , (v_m, 1/m)\}$ in the models [28], [29] is a particular case of an interval probability distribution $\{(v_1, [l_1, u_1]), …, (v_m, [l_m, u_m])\}$ in our EPRDB model. Thus, the EPRDB model is an extension of PRDB models [28], [29] with probabilistic values or interval probability distributions (Definition 1 and 7). Moreover, by associating interval probabilities with tuples, EPRDB allows representing and computing both uncertainties of relational attribute values and tuples. In other words, the proposed EPRDB model is a general extension of the CRDB, C-1PRDB, and C-2PRDB models with interval probabilities to represent and deal with both uncertainties of attribute values and tuples of probabilistic relations in practice. Figure 1 illustrates the extension of EPRDB in comparison with the CRDB, C-1PRDB, and C-2PRDB models.

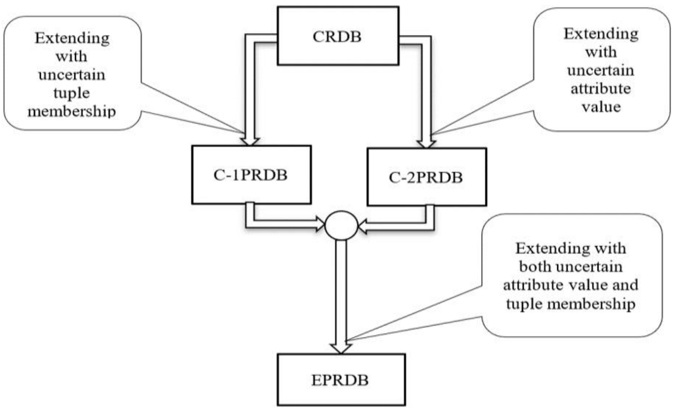

Figure 1. Extension of EPRDB

4.2. Efficiency of EPRDB in manipulating and computing data

Because the attribute value of EPRDB relations is a probabilistic value (definition 1), the computation and manipulation on the EPRDB data model are more effective than those on the C-2PRDB data models in [24], [25] where the attribute value is the probability distribution function pairs of a set of values. Moreover, the computation complexity of EPRDB algebraic operations is a polynomial under the size of probabilistic relations, and it is as effective as that of CRDB algebraic operations. Indeed, regarding the selection operation, since the computation time that a tuple holds or does not hold a selection condition is bounded above by some constant (Definition 14 and 15), then the cost for the selection of each tuple in an EPRDB relation (Definition 16) is also some constant or O(1). Thus, the time complexity of the selection operation on an EPRDB relation with *n* tuples is O(*n*). For the projection operation, from Definition 17, it is easy to see that the time for the probabilistic combination of the duplicate value tuples under a probabilistic disjunction strategy is a constant. Hence, the computation complexity of the projection operation on an EPRDB relation having *n* tuples is O(*n*). With the Cartesian product operation, by Definition 18, we can see that the time for the probabilistic combination of the membership degrees of two tuples under a probabilistic conjunction strategy is a constant. So, the computation complexity of the Cartesian product operation of two EPRDB relations having *n* and *m* tuples is O(*nm*). Similarly, the computation complexity of the join, intersection, union, and difference operations of two EPRDB relations having *n* and *m* tuples is O(*nm*). Thus, the performance of the EPRDB model in computing and manipulating uncertain data is good and can be applied in practice.

5. CONCLUSION

We have presented a new PRDB model, named EPRDB that extends the CRDB model with interval probability valued relational attributes and tuple memberships for representing and dealing with uncertain and imprecise information. As discussed, and compared to the existing PRDB models, the proposed EPRDB model is a general extension of the C-1PRDB and C-2PRDB models with two uncertain data levels of the attribute value and tuple membership of probabilistic relations. The uniqueness of our proposed EPRDB model is that it can represent and handle both uncertain relational tuples and imprecise attribute values associated with interval probabilities. In EPRDB, a relation is a set of uncertain tuples whose attributes may take one of imprecise values represented by a probabilistic value or an interval probability distribution on a value set. The fundamental concepts of the relational schema, probabilistic functional dependency, key as well as the set of basic probabilistic relational algebraic operations in EPRDB have been extended consistently with those in CRDB using the probabilistic interpretation of binary relations on sets, combination strategies of probabilistic intervals, and the conjunction, disjunction, and difference of probabilistic values. Basic properties of the EPRDB algebraic operations are proposed and proven completely to demonstrate that the EPRDB model is sound and coherent. The built EPRDB model can effectively manipulate and deal with uncertain and imprecise data.

Towards applying EPRDB, we will build a management system for EPRDB with the familiar querying and manipulating language like SQL that is able to represent and handle uncertain and imprecise information in the real world.

FUNDING INFORMATION

This research was funded by Saigon University (SGU). The authors acknowledge SGU for supporting this work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hoa Nguyen	✓	✓		✓	✓	✓		✓	✓	✓			✓	✓
Thi Nhi Tran		✓				✓	✓	✓	✓		✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] E. F. Codd, "A relational model of data for large shared data banks," *Communications of the ACM*, vol. 13, no. 6, pp. 377–387, Jun. 1970, doi: 10.1145/362384.362685.

[2] E. Sciore, *Database design and implementation*. in data-centric systems and applications. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-33836-7.

[3] T. Eiter, T. Lukasiewicz, and M. Walter, "A data model and algebra for probabilistic complex values," *Annals of Mathematics and Artificial Intelligence*, vol. 33, no. 2-4, pp. 205-252, Dec. 2001, doi: 10.1023/A:1013121110704.

[4] D. Suciu, "Probabilistic databases for all," in *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, New York, NY, USA: ACM, Jun. 2020, pp. 19–31. doi: 10.1145/3375395.3389129.

[5] Z. Ma and L. Yan, "Data modeling and querying with fuzzy sets: a systematic survey," *Fuzzy Sets and Systems*, vol. 445, pp. 147–183, Sep. 2022, doi: 10.1016/j.fss.2022.01.006.

[6] H. Nguyen, "An extended type-2 fuzzy relational database model for aggregate and grouping operations," *Malaysian Journal of Science and Advanced Technology*, pp. 150-157, Aug. 2023, doi: 10.56532/mjsat.v3i3.169.

[7] A. Ali, S. Talpur, and S. Narejo, "Detecting faulty sensors by analyzing the uncertain data using probabilistic database," in *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, IEEE, Jan. 2020, pp. 1–7. doi: 10.1109/iCoMET48670.2020.9074069.

[8] V. V. Kheradkar and S. K. Shirgave, "Query processing over relationalcross model in uncertain and probabilistic databases," in *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, IEEE, Feb. 2023, pp. 763-769. doi: 10.1109/ICAIS56108.2023.10073827.

[9] J. Bernad, C. Bobed, and E. Mena, "Uncertain probabilistic range queries on multidimensional data," *Information Sciences*, vol. 537, pp. 334–367, Oct. 2020, doi: 10.1016/j.ins.2020.05.068.

[10] W. W. Cohen, F. Yang, and K. R. Mazaitis, "TensorLog: a probabilistic database implemented using deep-learning infrastructure," *Journal of Artificial Intelligence Research*, vol. 67, pp. 285-325, Feb. 2020, doi: 10.1613/JAIR.1.11944.

[11] F. Yunus, P. Karmakar, P. Senellart, T. Abdessalem, and S. Bressan, "Using a Probabilistic database in an image retrieval application," in *EDBT 2025-28th International Conference on Extending Database Technology*, 2025, pp. 1106-1109. doi: 10.48786/edbt.2025.100.

[12] Y. Li, J. Chen, and L. Feng, "Dealing with uncertainty: a survey of theories and practices," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 11, pp. 2463-2482, Nov. 2012, doi: 10.1109/TKDE.2012.179.

[13] T. Ge, A. Dekhtyar, and J. Goldsmith, "Uncertain data: Representations, query processing, and applications," in *Studies in Fuzziness and Soft Computing*, vol. 304, 2013, pp. 67-108. doi: 10.1007/978-3-642-37509-5_4.

[14] T. Friedman and G. Van den Broeck, "Symbolic querying of vector spaces: probabilistic databases meets relational embeddings," in *Conference on Uncertainty in Artificial Intelligence*, 2020, pp. 1268-1277.

[15] İ. İ. Ceylan, A. Darwiche, and G. Van den Broeck, "Open-world probabilistic databases: Semantics, algorithms, complexity," *Artificial Intelligence*, vol. 295, p. 103474, Jun. 2021, doi: 10.1016/j.artint.2021.103474.

[16] H. Debbi, "Explaining query answers in probabilistic databases," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 4, pp. 140-152, Dec. 2023, doi: 10.9781/ijimai.2023.07.005.

[17] M. Grohe and P. Lindner, "Infinite probabilistic databases," *Logical Methods in Computer Science* , vol. 18, no. 1, pp. 34:1-34:43, Feb. 2022, doi: 10.46298/LMCS-18(1:34)2022.

[18] W. Zhao, A. Dekhtyar, and J. Goldsmith, "Databases for interval probabilities," *International Journal of Intelligent Systems*, vol. 19, no. 9, pp. 789-815, Sep. 2004, doi: 10.1002/int.20025.

[19] R. Ross, V. S. Subrahmanian, and J. Grant, "Aggregate operators in probabilistic databases," *Journal of the ACM*, vol. 52, no. 1, pp. 54-101, Jan. 2005, doi: 10.1145/1044731.1044734.

[20] C. Zhang, Z. Mei, B. Wu, Z. Zhao, J. Yu, and Q. Wang, "Query with assumptions for probabilistic relational databases," *Tehnicki Vjesnik*, vol. 27, no. 3, pp. 923-932, Jun. 2020, doi: 10.17559/TV-20191123110408.

[21] L. Yan and Z. M. Ma, "A probabilistic nested relational database model with fuzzy probability measures," in *Proceedings of the 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, New Orleans, Louisiana, USA, 2019, pp. 253-257.

[22] K. Papaioannou, M. Theobald, and M. Boehlen, "Supporting set operations in temporal-probabilistic databases," in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, IEEE, Apr. 2018, pp. 1180-1191. doi: 10.1109/ICDE.2018.00109.

[23] A. Gilad, A. Imber, and B. Kimelfeld, "The consistency of probabilistic databases with independent cells," in *Proceedings of the 26th International Conference on Database Theory (ICDT 2023)*, Athens, Greece, 2023, pp. 22:1-22:19.

[24] H. Nguyen, T. N. Nguyen, and T. T. N. Tran, "A probabilistic relational database model with uncertain multivalued attributes," *ICIC Express Letters*, vol. 16, no. 3, pp. 241-248, 2022, doi: 10.24507/icicel.16.03.241.

[25] H. Nguyen, "Extending probabilistic relational database model with uncertain multivalued attributes," *International Journal of Innovative Computing, Information and Control*, vol. 18, no. 5, pp. 1477-1492, 2022, doi: 10.24507/ijicic.18.05.1477.

[26] H. Nguyen and D. N. Le, "A relational database model with interval probability valued attributes for uncertain and imprecise information," *ECTI Transactions on Computer and Information Technology*, vol. 18, no. 3, pp. 307-318, 2024, doi: 10.37936/ecti-cit.2024183.255697.

[27] H. Nguyen, "Extended relational database model for interval probability set-valued attributes," *International Journal of Information Technology and Computer Science Applications*, vol. 3, no. 1, pp. 11-24, Feb. 2025, doi: 10.58776/ijitcsa.v3i1.174.

[28] S. K. Lee, "An extended relational database model for uncertain and imprecise information," in *Proceedings of the 18th International Conference on Very Large Data Bases*, 1992, pp. 211-220.

[29] H. Nguyễn, "Extending relational database model for uncertain information," *Journal of Computer Science and Cybernetics*, vol. 35, no. 4, pp. 355-372, Oct. 2019, doi: 10.15625/1813-9663/35/4/13907.
[30] T. H. Cao, "A relational database model and algebra integrating fuzzy attributes and probabilistic tuples," *Fuzzy Sets and Systems*, vol. 445, pp. 123-146, Sep. 2022, doi: 10.1016/j.fss.2021.10.017.
[31] V. Biazzo, R. Giugno, T. Lukasiewicz, and V. S. Subrahmanian, "Temporal probabilistic object bases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 921-939, Jul. 2003, doi: 10.1109/TKDE.2003.1209009.
[32] W. Briggs, *Uncertainty*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-39756-6.

BIOGRAPHIES OF AUTHORS

Hoa Nguyen received his Ph.D. degree in Computer Science at Vietnam National University, Ho Chi Minh City, Vietnam, in 2008. He is currently an associate professor at Information Technology Faculty, Saigon University, Vietnam. His research interest includes imprecise and uncertain knowledge representation, soft computing, fuzzy and probabilistic databases. He can be contacted at email: nguyenhoa@sgu.edu.vn.

Thi Nhi Tran received her Master's degree in Computer Science at Saigon University, Vietnam, in 2025. She is currently a data analyst at Vietmy SSU Company, Vietnam. Her research interest includes computational science, data mining and probabilistic databases. She can be contacted at email: ttnhi@hv.sgu.edu.vn.