On Probability Algebra: Classic Theory of Probability Revisited

YINGXU WANG [1,2]

[1] President, International Institute of Cognitive Informatics and Cognitive Computing (ICIC)
Director, Laboratory for Computational Intelligence, Denotational Mathematics, and Software Science
Dept. of Electrical and Computer Engineering,
Schulich School of Engineering and Hotchkiss Brain Institute
University of Calgary
2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

[2] Visiting Professor, Information Systems Lab
Stanford University, Stanford, CA. 94305, USA
http://www.ucalgary.ca/icic/ Email: yingxu@ucalgary.ca

Abstract: - It is recognized that the classic probability theory is cyclically defined among a small set of coupled operations. The sample space of probability is not merely 1-D invariant structures rather than n-D variant hyperstructure where the types of probability events encompassing those of joint or disjoint as well as dependent, independent, or mutually-exclusive ones. These fundamental properties of probability lead to a three-dimensional dynamic model of probability structures constrained by types of sample spaces, relations and dependencies of events. A reductive framework of general probability theory is rigorously derived from the independently defined model of conditional probability. This basic study reveals that the Bayes' law needs to be extended in order to fit more general contexts of variant sample spaces and complex event properties. The revisited probability theory enables an extended mathematical structure known as probability algebra for rigorous manipulating uncertainty events and causations in formal inference, qualification, quantification, and semantic analysis in contemporary fields such as cognitive informatics, computational intelligence, cognitive robots, complex systems, soft computing, semantic computing, and brain informatics.

Key-Words: - Denotational mathematics, probability theory, probability algebra, statistics, formal inference, cognitive informatics, cognitive computing, computational intelligence, semantic computing, soft computing, brain informatics, cognitive systems

1. Introduction

Probability theory is a branch of mathematics that deals with uncertainty and probabilistic norms of random events and potential causations as well as their algebraic manipulations. The development of classic theories of probability can be traced back to the work of Blaise Pascal (1623-1662) and Pierre de Fermat (1601-1665) [Todhunter, 1865; Venn, 1888; Hacking, 1975]. Many others such as Jacob Bernoulli, Reverend T. Bayes, and Joseph Lagrange had significantly contributed to classic probability theory. Theories of probability in its modern form was unified by Pierre Simon and Marquis de Laplace in the 19th century [Kolmogorov, 1933; Whitworth, 1959; Hacking, 1975; Mosteller, 1987; Bender, 1996]. Set theories [Cantor, 1874; Zadeh, 1965, 1968, 1996, 2002; Artin, 1991; Ross, 1995; Pedrycz & Gomide, 1998; Novak et al., 1999; Potter, 2004; Gowers, 2008; BISC, 2013; Wang, 2007] provide an expressive power for modeling the discourse and axioms of probability theories. A theory of fuzzy probability and its algebraic framework has been presented in [Wang, 2015e].

The philosophy of probability theory is analogy-based where large-enough experiments are required for establishing prior probabilistic estimations and norms in a certain sample space. The main methodology of classic probability theory is a black box predication for a set of uncertain phenomena of a complex system without probing into its internal mechanisms. Although the range of prior probability for any predicated event is [0, 1], the range of posterior probability is immediately reduced to {0, 1} after the given event has realized in a given probability space.

It is recognized that the classic probability theory is cyclically defined among a set of highly coupled operations on probabilities of conjunctive, disjunctive, and conditional events. This paper presents a basic study on the revisited theory of probability, which extends classic probability theory to a comprehensive set of probability operations. Some fundamental challenges and potential pitfalls of classic probability theory are formally analyzed in Section 2. The mathematical model of general probability is introduced in Section 3 based on rigorous models of the universe of discourse and sample spaces of probability. The framework of the general probability theory is embodied by probability algebra as introduced Section 4. An extended set of algebraic operators on the revisited mathematical model of probability is rigorously defined in Section 5, which extends the traditional probability operations of addition, multiplication, and condition to subtraction and division. The conventional mutually coupled probability operations are independently separated in a deductive structure on the basis of the refined model of conditional probability. Formal properties of probability and rules of algebraic operations on general probabilities are summarized in Section 6. Practical examples are provided throughout the paper for elaborating each of the fundamental definitions and operations in the general theory of probability. The revisited theory of probability can be used to solve a number of challenging problems in classic probability theory such as complex sequential, concurrent, and causal probabilities as well as real-time probabilities under highly restrictive timing constraints.

2. Pitfalls of the Classic Theory of Probability

Potential pitfalls of classic probability theory stem from the highly coupled dependency between the key probability operators and the overlooking of the variant sample spaces in probability modeling and manipulations. In order to deal with the cyclically defined framework of classic probability theory, classic literature and textbooks describe the highly coupled probability operations in various approaches merely dependent on where the loop is subjectively cut.

2.1 Highly Coupled Dependency among Probability Operators

Definition 1. The essence of probability P is a quantification function ρ that maps an event e in a sample space S into a unit interval $\mathbb{I}$ = [0, 1], which is determined by a relative ratio between the size of the event (number of expected occurrences) and the size of the sample space, i.e.:

$$P \triangleq \{(e, P(e)) \mid e \in S, P(e) = \rho : e \to \mathbb{I} = [0,1]\} \tag{1}$$

The classic theory of probability [Kolmogorov, 1933; di Finetti, 1970; Johnson & Bhattacharyya, 1996; Lipschutz & Lipson, 1997] was somehow defined on a cyclic tautology as illustrated in Fig. 1. In the framework of classic probability theory, conjunctive probability on the left-hand side is defined based on disjunctive probability on the right. Further, the disjunctive probability is dependent on conditional probability that, inversely, is defined by the disjunctive probability in an interlocked loop.

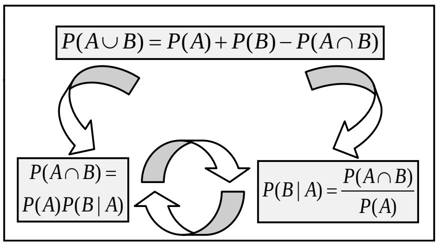

Fig. 1. The highly coupled dependency in classic probability theory

Lemma 1. The *paradox of classic theory of probability* is that none of the three basic operators for conjunctive, disjunctive, and conditional probabilities is independently definable in the framework, so that an interlocked relation among the probabilistic operators is formed, i.e.:

$$P(A \cup B) \;\to\; P(A \cap B) \;\leftrightarrow\; P(B \mid A) \tag{2}$$

The highly coupled dependency between key probability operators results in numerous problems in probabilistic reasoning, theorem proving, and applications in classic probability theories.

2.2 The Impact of Variant Sample Spaces of Probability

It is observed that, in general, the sample space of probability is dynamically variable rather than merely constant as traditionally perceived.

Example 1. Assume a bag has a black ball and a white ball denoted by two events B and W, respectively, as illustrated in Fig. 2 where the ball drawn in the previous round will not return to the bag. The probabilities for getting a black or white ball in the first trail are, $P(B)$ = $P(W)$ = 0.5, respectively. However, given the first draw was a white ball, the second trial will result in $P(W \mid W)$ =

0 and $P(B \mid W) = 1$. Or in other case, $P(B \mid B) = 0$ and $P(W \mid B) = 1$ given the first draw was a black ball. In both cases for the second trail, the sample space has been changed from $|S| = 2$ to $|S'| = 1$.

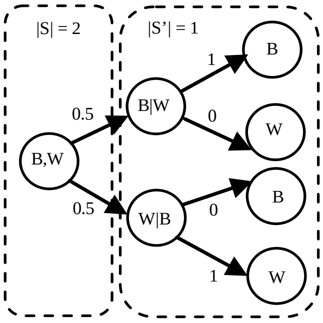

Fig. 2. Dynamic conditional probability in variant sample spaces

This explains why most illustrations and case studies on classic probability theory request that, in a typical layout, a ball drawn from the bag must be returned after each trail. However, this practice has over simplified the general context of probability theory.

Lemma 2. Let E_1 and E_2 are two sets of arbitrary events, the *sample space* of probabilistic events as the context of probability is variant in general in a series of probabilistic trails, i.e.:

$$S(E_1, E_2) \neq S(E_2 \mid E_1) \neq S(E_1 \mid E_2) \tag{3}$$

where $S(E_1, E_2)$, $S(E_2 \mid E_1)$, and $S(E_1 \mid E_2)$ represent the sample spaces for events E_1 and E_2, E_2 after E_1, and E_1 after E_2, respectively.

Proof. Assume the probabilistic sample space S includes two sets of arbitrary events E_1 and E_2. Lemma 2 is proven by considering the sizes of the sample space under four different scenarios as follows:

$$\begin{aligned}
&\forall E_1, E_2, S, E_1 \subset S \text{ and } E_2 \subset S,\\
&\quad i)\ \text{Invariant } S \text{ and independent } E:\\
&\qquad |S(E_1, E_2)| = |S^0| = |E_1| + |E_1|,\ E_1 \cap E_2 = \varnothing\\
&\quad ii)\ \text{Invariant } S \text{ and dependent } E:\\
&\qquad |S(E_1, E_2)| = |S| = |E_1| + |E_1| - |E_1 \cap E_2|,\ E_1 \cap E_2 \neq \varnothing\\
&\quad iii)\ \text{Variant } S \text{ and independent } E:\\
&\qquad |S(E_2 \mid E_1)| = |S'| = (|E_1| - 1) + |E_2|,\ E_1 \cap E_2 = \varnothing\\
&\quad iv)\ \text{Variant } S \text{ and dependent } E:\\
&\qquad |S(E_1 \mid E_2)| = |S''| = (|E_1| - 1) + (|E_2| - 1) - |E_1 \cap E_2|,\\
&\qquad\qquad E_1 \cap E_2 \neq \varnothing\\
&\Rightarrow\ |S| \neq |S^0| \neq |S'| \neq |S''|
\end{aligned} \tag{4}$$

■

According to Lemma 2, the basic assumption of classic probability theory on invariant sample spaces is a simplified special case (i) or (ii) in the general dynamic sample space, which is illustrated in Fig. 2. The finding in Lemma 2 indicates that *Bayes' law of conditional probability* is not generally true in the generally variant probability space, which will be proven in Corollary 4.

3. Mathematical Models of the General Probability Theory

The mathematical model of the extended probability theory is defined in the universe of discourse of probability and the dynamic sample space based on complex event relations. Set theory is adopted as a unified foundation for the mathematical model of probabilities and their algebraic operations.

3.1 The Universe of Discourse of General Probability

In addition to the typical axioms, as summaries in Table 3 in Section 4, the formal model of the universe of discourse for the extended theory of probability is a foundation that specifies the general context and layout of probability theory, which will be introduced after some conceptual preparations given in Definitions 2 through 4.

Definition 2. The *set of states*, Ξ, with individual bivalent probabilistic status ξ_i, $\xi_i \in \Xi$, $1 \leq i < |\Xi|$, of entities and/or causations is expressed as follows:

$$\Xi \triangleq \{\mathop{R}_{i=1}^{|\Xi|} \xi_i \mid f(\xi_i) \in [0,1] \wedge \xi_i \in \Xi\} \tag{5}$$

where $\mathop{R}_{i=1}^{n} \xi_i$ is called the *big-R notation* that denotes a set of recurrent structures or repeated behaviors [Wang, 2008b].

Definition 3. The *set of events*, $\mathfrak{E}$, is a subset of changed states in Ξ as identified by a discrete differentiation [Wang, 2007, 2014c], i.e.:

$$\begin{aligned}
\mathfrak{E} &\triangleq \frac{d\Xi}{dt} = \frac{\Xi_t \oplus \Xi_{t'}}{t - t'}\Big|_{t'=t-1} = \Xi_t \oplus \Xi_{t'}\\
&= \{\mathop{R}_{i=1}^{n} e_i,\ iff\ \frac{d\xi_i}{dt} = \xi_i^t \oplus \xi_i^{t'},\ \xi_i \in \Xi \wedge e_i \in \mathfrak{E}\}
\end{aligned} \tag{6}$$

Definition 4. The *set of probability distribution*, $\mathfrak{P}$, is a function ρ that maps each event $e_i \in \mathfrak{E}$ into the unit interval $\mathbb{I}$, i.e.:

$$\begin{aligned}\mathfrak{P}(\mathfrak{E}) &\triangleq \mathop{R}_{i=1}^{n} \mathfrak{P}(e_i) \\ &= \mathop{R}_{i=1}^{n} (\rho_i : e_i \to \mathbb{I}),\ \mathbb{I} = (0,1]\end{aligned} \quad (7)$$

The universe of discourse of probability can be modeled based on the three essences as introduced in Eqs. 5 to 7.

Definition 5. The *universe of discourse of general probability theory*, $\mathfrak{U}$, is a triple:

$$\mathfrak{U} \triangleq (\Xi, \mathfrak{E}, \mathfrak{P}) \quad (8)$$

where Ξ denotes a finite set of states, $\mathfrak{E}$ a finite set of events, and $\mathfrak{P}$ a finite set of probability distribution.

3.2 The Hyperstructure of Sample Spaces of the General Probability Theory

On the basis of the universe of discourse of general probability, a set of fundamental properties of events and sample spaces in probability theory can be formally analyzed.

Definition 6. The *relation*, *R*, between two sets of events E_1 and E_2 in $\mathfrak{U}$ is classified into the categories of *joint* and *disjoint* as follows:

$$\begin{cases} E_1 \cap E_2 \neq \varnothing & \text{// Joint} \\ E_1 \cap E_2 = \varnothing & \text{// Disjoint} \end{cases} \quad (9)$$

Definition 7. The *dependency*, *D*, between two sets of events E_1 and E_2 in $\mathfrak{U}$ is classified into the categories of *independent*, *dependent*, and *mutually-exclusive* (ME) as follows:

$$\begin{cases} E_1 \not\to E_2 & \text{// Independent} \\ E_1 \to (E_2 = E_2') & \text{// Dependent} \\ E_1 \to (E_2 = \varnothing) & \text{// Mutually-exclusive (ME)} \end{cases} \quad (10)$$

where $\to$ denotes a trigger relation or causation, ME is a special type of event dependency in which the sets of events never appear simultaneously or concurrently.

It is noteworthy that the event dependency is different from the event relation according to Definitions 7 and 6. The latter denotes that two sets of events may or may not share certain common events. However, the former represents that a set of events E_2 may or may not be influenced by another set of events E_1 in consecutive interactions via dynamic changes of the variant sample space.

Definition 8. A simple *first order sample space S* of a probabilistic layout is a set of all potential events expected in trails as a subset of the power set of the general events $S \subset \mathfrak{p}\mathfrak{E}$ in $\mathfrak{U}$, i.e.:

$$\begin{aligned} S &\triangleq \{\mathop{R}_{i=1}^{n} E_i \mid E_i \subseteq \mathfrak{p}S \subset \mathfrak{p}\mathfrak{E}\} \\ &= \{\mathop{R}_{i=1}^{n} \mathop{R}_{j=1}^{n_i} e_{ij} \mid e_{ij} \in E_i \subseteq \mathfrak{p}S \subset \mathfrak{p}\mathfrak{E}\} \end{aligned} \quad (11)$$

The sample space of probability forms the context of a given problem in probabilistic analysis and modeling.

Example 2. Let an unfair coin with 0.68 : 0.32 probabilistic weights for the events head (*H*) and tail (*T*), respectively. The simple first order sample space S_1 can be modeled according to Definition 8 as follows:

$$\begin{aligned} S_1 &= \{\mathop{R}_{i=1}^{2} e_i \mid \sum_{i=1}^{2} P_{S_1}(e_i) \equiv 1\} \\ &= \{e_1(H) = 0.68,\ e_2(T) = 0.32\} \end{aligned}$$

which is invariant, disjoint, and mutually-exclusive.

Example 3. Given a complex first oder sample space S_2 with five white balls (*W*) and five black balls (*B*) in a bag possessing 0.45 : 0.55 event probability due to the roughness between balls in different colors. S_2 can be modeled according to Definition 8 as follows:

$$\begin{aligned} S_2 &= \{\mathop{R}_{i=1}^{2} \mathop{R}_{j=1}^{5} e_{ij} \mid \sum_{i=1}^{2} \sum_{j=1}^{5} p_{S_2}(e_{ij}) \equiv 1\} \\ &= \{\mathop{R}_{j=1}^{5} e_{1j}(B) = 0.55/5),\ \mathop{R}_{j=1}^{5} e_{2j}(W) = 0.45/5)\} \\ &= \{\mathop{R}_{k=1}^{5} e_k(B) = 0.11),\ \mathop{R}_{k=6}^{10} e_k(W) = 0.09)\} \end{aligned}$$

which is disjoint, invariant or variant subject to independent or dependent events.

Corollary 1. The *size of a sample space S*, $|S|$, is determined by the number of all distinguishable or nonredounded events in *S* in $\mathfrak{U}$, i.e.:

$$|S| \triangleq |\bigcup_{i=1}^{n} E_i| = |\bigcup_{i=1}^{n} \{\mathop{R}_{j=1}^{n_i} e_{ij} \mid \mathop{R}_{i=1}^{n} \mathop{R}_{j=1}^{n_i} e_{ij} \in E_i \subseteq \mathfrak{p}S\}| \quad (12)$$

Proof. Corollary 1 is proven on the basis of Definition 8 in an invariant sample space as follows:

$$\begin{aligned} &\forall \mathop{R}_{i=1}^{n} E_i \subseteq \mathfrak{p}S \wedge \mathop{R}_{i=1}^{n} \mathop{R}_{j=1}^{n_i} e_{ij} \mid e_{ij} \in E_i \subseteq \mathfrak{p}S, \\ &|S| \triangleq \sum_{i=1}^{n} \sum_{j=1}^{n_i} |e_{ij}|,\ e_{ij} \neq e_{i'j'} \wedge e_{ij} \in E_i \wedge |e_{ij}| \equiv 1 \\ &\quad = |\bigcup_{i=1}^{n} E_i| \end{aligned} \quad (13)$$

■

Sample spaces of complex probabilistic problems are often to be higher-order hyperstructures as the context of sequential, concurrent, and conditional probabilities.

Definition 9. An *Order-m sample space* S^m is a set of potentially *m*-ary combinational events for a probabilistic structure determined by multiple Cartesian products $\overset{m}{\underset{i=1}{\times}} S_i$, $S_i \subseteq Þ\mathcal{E}$ in $\mathfrak{U}$, i.e.:

$$\begin{aligned} S^m &\triangleq \overset{m}{\underset{i=1}{\times}} S_i \\ &= \{\underset{i_1=1}{\overset{n_1}{R}}\underset{i_2=1}{\overset{n_2}{R}}...\underset{i_m=1}{\overset{n_m}{R}}(e_{i_1 i_2 ... i_m} \mid e_{i_1 i_2 ... i_k ... i_m} \in S^m \wedge e_{i_k} \in E_{i_k} \subseteq S_{i_k})\} \end{aligned}$$

when $m = 2$ or $m = 3$,

$$\begin{aligned} S^2 &= S_i \times S_j \\ &= \{\underset{i=1}{\overset{n_i}{R}}\underset{j=1}{\overset{n_j}{R}}(e_{ij} \mid e_{ij} \in S^2 \wedge e_i \in E_i \subseteq S_i \wedge e_j \in E_j \subseteq S_j)\} \\ S^3 &\triangleq S_i \times S_j \times S_k \\ &= \{\underset{i=1}{\overset{n_i}{R}}\underset{j=1}{\overset{n_j}{R}}\underset{j=k}{\overset{n_k}{R}} e_{ijk} \mid e_{ijk} \in S^3 \wedge e_i \in E_i \subseteq S_i \wedge e_j \in E_j \subseteq S_j \\ &\qquad \wedge e_k \in E_k \subseteq S_k\} \end{aligned} \tag{14}$$

It is noteworthy that there are two types of sample spaces in the revisited probability theory called the *invariant* and *variant* sample spaces, respectively. According to Lemma 2 the variant sample space is more general in probability theory where the prior and posterior sample spaces are different in each step of a series of probabilistic experiments. An example of the variant sample space is such as *n* balls in a bag where the sample space continuously reducing along a series of trails if the balls drawn will not return to the bag. Another example of the generally variant sample space is a set of bacteria where their size is exponentially increasing in a series of probabilistic experiments.

Example 4. On the basis of $S_1 = \{H = 0.68,\ T = 0.32\}$ as given in Example 2, an *invariant* 2nd-order sample space for two consecutive or concurrent tosses of the uneven coin, S_1^2, can be derived according to Definition 9 as a set of composite events $HH, HT, TH, \text{and}\, TT$ as follows:

$$\begin{aligned} S_1^2 &= \{\underset{i=1}{\overset{2}{R}}\underset{j=1}{\overset{2}{R}}(e_{ij} = e_i e_j \mid e_{ij} \in S_1^2 \wedge \sum_{i=1}^{2}\sum_{j=1}^{2} e_{ij} \equiv 1)\} \\ &= \{HH = 0.68 \bullet 0.68 = 0.46,\ HT = 0.68 \bullet 0.32 = 0.22, \\ &\quad TH = 0.32 \bullet 0.68 = 0.22,\ TT = 0.32 \bullet 0.32 = 0.10\} \end{aligned}$$

Example 5. Consider the case $S_2 = \{\underset{i=1}{\overset{5}{R}} W = 0.09, \underset{i=6}{\overset{10}{R}} B = 0.11\}$ as given in Example 3 there are five white balls (*W*) and five black balls (*B*) in a bag with 0.45 : 0.55 biased probabilistic weights in different colors. A *variant* 2nd-order sample space for two consecutive draws of the uneven balls, $S_2^{2'}$, with the set of combined events $BB, BW, WB, \text{and}\, WW$ can be formally modeled according to Definition 9 as follows:

$$\begin{aligned} S_2^{2'} &= \{\underset{i=1}{\overset{2}{R}}\underset{j=1}{\overset{2}{R}}(e_{ij} = e_i(e_j \mid e_i) \mid e_{ij} \in S_2^{2'} \wedge \sum_{i=1}^{2}\sum_{j=1}^{2} |e_{ij}| \equiv 1)\} \\ &= \{BW = B\frac{|E_{W^2}|}{|S_2^{'}|} = B\frac{|E_{W^2}|}{|S_2| - b_i} = 0.55\frac{0.45}{1-0.11} = 0.28, \\ &\quad WB = W\frac{|E_{B^2}|}{|S_2^{'}|} = W\frac{|E_{B^2}|}{|S_2| - w_i} = 0.45\frac{0.55}{1-0.09} = 0.27, \\ &\quad BB = B\frac{|E_{B^2}^{'}|}{|S_2^{'}|} = B\frac{|E_{B^2}| - b_i}{|S_2| - b_i} = 0.55\frac{0.55-0.11}{1-0.11} = 0.27, \\ &\quad WW = W\frac{|E_{W^2}^{'}|}{|S_2^{'}|} = W\frac{|E_{W^2}| - w_i}{|S_2| - w_i} = 0.45\frac{0.45-0.09}{1-0.09} = 0.18\} \end{aligned}$$

where the variant sample space $S_2^{2'}$ encompasses four complex events with individually modified probabilities by reflecting influences between a sequential or conditional event on another event in the combinations.

Contrasting Examples 4 and 5, it is obvious that an invariant sample space is merely a special case of the variant ones as formally described in Definition 9.

Theorem 1. The *variability of probabilistic sample spaces*, S, $S \subseteq \Xi$ in $\mathfrak{U}$ is general in a serial probabilistic experiments $E_1 \to E_2$ due to the removal or disappearance of an event e_i, $e_i \in E_1$ after the previous trial, i.e.:

$$\begin{aligned} &\forall S(E_1, E_2),\ E_1 \to E_2 \\ &S' \neq S, \text{ where } S = E_1 \cup E_2, \text{and} \\ &\qquad S' = E_1^{'} \cup E_2^{'} = S \setminus e_1,\ \ e_1 \in E_1 \end{aligned} \tag{15}$$

Proof. Theorem 1 can be proven on the basis of Definitions 8 and 9 as follows:

$$\forall E_1, E_2, E_1 \to E_2, \text{ and } E_1 \cap E_2 \neq \varnothing,$$
$$\begin{cases} a) \text{ Before the trail:} \\ \quad S(E_1, E_2) = E_1 \cup E_2 \\ b) \text{ After the trail } (E_1 \to E_2, e_1 \in E_1 \text{ was removed}): \\ \quad \begin{cases} E_1^{'} = E_1 \setminus e_1, \; e_1 \in E_1 \\ E_2^{'} = E_2 \end{cases} \\ \quad \Rightarrow S'(E_1^{'}, E_2^{'}) = E_1^{'} \cup E_2^{'} \\ \qquad = E_1 \cup E_2 \setminus e_1, \; e_1 \in E_1 \\ \qquad = S \setminus e_1 \\ \text{Therefore, } S'(E_1^{'}, E_2^{'}) \neq S(E_1, E_2) \end{cases} \tag{16}$$

■

It is noteworthy that the term of the variant sample space $\widetilde{S}' \neq \widetilde{S}$ as given in Theorem 1 refers to the dynamic distributions of probabilities for individual events in a series of experiments due to the impact of previous ones. Many problems and exceptions of classic probability theory stem from the overlooking of the fact of variant sample spaces as stated in Theorem 1.

3.3 Mathematical Model of the General Probability Theory

The extended probability is a mathematical structure built on the basis of set theory and the universe of discourse of probability as described in proceeding subsections.

Definition 10. The *probability* of an event $e \in E$ in a sample space *S* in $\mathfrak{U}$, denoted by $P(e \mid e \in E \subseteq S)$, is a ratio between the size of the set of the expected events $|E|$ and that of the sample space $|S|$:

$$P(e \mid e \in E \subseteq S \subset \mathfrak{E}) \triangleq \frac{|E|}{|S|} \tag{17}$$

Example 6. On the basis of the sample space $S_1 = \{H = 0.68, \; T = 0.32\}$ as given in Example 2, the probabilities of individual events $P(H)$ and $P(T)$ of the unfair coin can be determined, respectively, according to Definition 10 as follows:

$$P(H \mid H \in E_H \subset S_1) = \frac{|E_H|}{|S_1|} = \frac{0.68}{0.68 + 0.32} = 0.68$$
$$P(T \mid T \in E_T \subset S_1) = \frac{|E_T|}{|S_1|} = \frac{0.32}{1} = 0.32$$

Example 7. Reuse the sample space $S_2 = \{\underset{i=1}{\overset{5}{R}} W = 0.09, \underset{i=6}{\overset{10}{R}} B = 0.11\}$ as modeled in Example 3, the probabilities of individual events $P(B)$ and $P(W)$ of the uneven balls drawing from the bag can be determined, respectively, according to Definition 10 as follows:

$$P(B \mid B \in E_B \subseteq S_2) = \frac{|E_B|}{|S_2|} = \frac{0.11 \bullet 5}{1} = 0.55$$
$$P(W \mid W \in E_W \subseteq S_2) = \frac{|E_W|}{|S_2|} = \frac{0.09 \bullet 5}{1} = 0.45$$

Corollary 2. The *probability of the entire sample space S*, *P*(*S*), is always constrained by the unit size, i.e.:

$$P(S) = \sum_{i=1}^{|S|} P_S(e_i \mid e_i \in S) = \frac{|S|}{|S|} \equiv 1 \tag{18}$$

According to the formal models of events and sample spaces, the nature of probability is constrained by different contexts determined by the three factors in the Cartesian product, $S \times R \times D$, as described in Table 1 where *S* denotes the sample space (variant/invariant), R relation of sets of events (joint/disjoint), and D dependency of events (dependent/independent/mutually-exclusive (ME)). Therefore, the contexts of general probability are classified into four categories according to the control factors in the Cartesian product, i.e.: i) *invariant* sample space and *disjoint/*ME-*dependent* events, ii) *invariant* sample space and *joint/independent* events, iii) *variant* sample space and *disjoint/independent* events, and iv) *invariant* sample space and *joint/dependent* events.

Table 1. Contexts of Relations and Dependencies of Events in the General Probability Theory

No	Category	Definition ($S \times R \times D$)	Sample space (*S*)	Events: Relation (*R*)	Events: Dependency (*D*)
i	Disjoint/mutually-exclusive (ME) events in invariant sample space	$S \times \overline{R} \times D$	$S = S'$	$X \cap Y = \varnothing$	$X \to (Y = \varnothing)$, ME
ii	Joint/independent events in invariant sample space	$S \times R \times \overline{D}$		$X \cap Y \neq \varnothing$	$X \not\to Y$
iii	Disjoint/independent events in variant sample space	$S' \times \overline{R} \times \overline{D}$	$S \neq S'$	$X \cap Y = \varnothing$	$X \not\to Y$
iv	Joint/dependent events in variant sample space	$S' \times R \times D$		$X \cap Y \neq \varnothing$	$X \to (Y = Y')$

It will be demonstrate and proven in Section 5 that any complex probability can be expressed by an algebraic operation on the primitive single variable probabilities in the theory of general probability.

4. The Framework of the General Probability Theory and Probability Algebra

On the basis of the mathematical models of general probability and its discourse as defined in preceding sections, the framework of revisited probability theory can be established by a set of algebraic operators on formal probabilities as summarized in Table 2. It is noteworthy that traditional probability theory only covers a special case of the general probability in the invariant sample space with mainly joint and independent events.

Each operator of probability algebra in Table 2 will be formally described in the general form as an algebraic expression based on Definition 10. Then, special cases of any probability operation will be analyzed according to its properties in the three-dimensional structure constrained by $(S \times R \times D)$ as classified in Table 1. Therefore, the four combinations in $(S \times R \times D)$ form the general contexts of the revisited probability theory. This approach reveals a number of fundamental properties of both the general and classic probabilities and their manipulations, which will be formally described in Section 5. A set of basic properties of general probability in the universe of discourse $\mathfrak{U}$ are summarized in Table 3. Properties 1 and 2 in Table 3 describe the characteristics of probabilities in both singularities of the entire and empty sample spaces, respectively, where $\mathfrak{E} \sqsubset \mathfrak{U}$ denotes that $\mathfrak{E}$ is a component (dimension) of the hyperstructure $\mathfrak{U}$. It is noteworthy as specified in Property 4 and proven in Corollary 2 that the probability of the sample space $P(S) \equiv 1,\ e \in E \subseteq S \subset \mathfrak{E}$, in any given problem layout.

The mathematical model of probability, the framework of the revisited probability theory, and the formal operators of probability algebra enable rigorous analyses of the nature, properties, and rules of probabilities as well as their algebraic operations. The basic properties of probabilities provide a set of axioms for the general probability theory. On the basis of the structural properties of general probability, a comprehensive set of operations and rules of probability algebra will be derived in Section 5. This leads to the explanation that classic probability theory is a special case and compatible subsystem of the revisited probability theory in terms of both mathematical models and probability operators.

Table 2. The Framework of Contexts and Operators of Probability Algebra

No.	Operator	Definition				
		General ($a \in A \subset S \wedge b \in B \subset S'$)	**Invariant sample space (Classic)** ($S' = S$)		**Variant sample space (Revisited)** ($S' \neq S$)	
			Disjoint / ME dependent Events $A \cap B = \varnothing \wedge A \to (B = \varnothing)$	**Joint / independent events** $A \cap B \neq \varnothing \wedge A \not\to B$	**Disjoint / independent events** $A \cap B = \varnothing \wedge A \not\to B$	**Joint / dependent events** $A \cap B \neq \varnothing \wedge A \to B$
1	Condition	$P(b \mid a) \triangleq P(A \to B)$	0	$P(b)$	$P'(b) = \frac{P(b)}{1-P(a_i)}$, $P(a_i) = \frac{P(a)}{\lvert A \rvert}$	$P''(b) = \frac{P(b)-P(b_i)}{1-P(b_i)}$, $P(b_i) = \frac{P(b)}{\lvert B \rvert}$
2	Multiplication	$P(a \times b) = P(A \wedge B) = P(a)P(b \mid a)$	0	$P(a)P(b)$	$P(a)P'(b)$	$P(a)P''(b)$
3	Division	$P(b / a) = P(\frac{\lvert B \rvert}{\lvert A \rvert}) = \frac{P(b)}{P(a)},\ 0 < P(a) \geq P(b)$	0	$\frac{P(b)}{P(a)}$	$\frac{P'(b)}{P(a)}$	$\frac{P''(b)}{P(a)}$
4	Addition	$P(a+b) = P(A \vee B) = P(a)+P(b)-P(a)P(b \mid a)$	$P(a)+P(b)$	$P(a)+P(b)-P(a)P(b)$	$P(a)+P(b)-P(a)P'(b)$	$P(a)+P(b)-P(a)P''(b)$
5	Subtraction	$P(a-b) = P(A \setminus B) = P(a)-P(a)P(b \mid a)$	$P(a)$	$P(a)-P(a)P(b) = P(a)P(\bar{b})$	$P(a)-P(a)P'(b) = P(a)P'(\bar{b})$	$P(a)-P(a)P''(b) = P(a)P''(\bar{b})$
6	Complement	$P(\bar{a}) = 1 - P(a)$				

Table 3. Axiomatic Properties of Probability Algebra

No.	Property	Condition
1	$P(\mathfrak{E}) \equiv 1$	$\mathfrak{E} \sqsubset \mathfrak{U}$
2	$P(\varnothing) = 0$	$\varnothing \subset \text{þ}\mathfrak{E} \sqsubset \mathfrak{U}$
3	$0 \le P(E) \le 1$	$E \subseteq S \subset \mathfrak{E}$
4	$\sum_{i=1}^{\|S\|} P(e_i) = 1$	$e_i \in E \subseteq S \subset \mathfrak{E} \wedge P(S) \equiv 1$
5	$P(A) \le P(B)$	$A \subseteq B \subset S$
6	$P(\overline{a}) = 1 - P(a)$	$P(\overline{e}) + P(e) \equiv 1,\ e \in E \subseteq S \subset \mathfrak{E}$

5. Formal Operators of Probability Algebra

The theoretical framework of general probability and the mathematical structure of probability algebra are formally presented in Sections 3 and 4. On the basis of the unified mathematical model of the conditional probability, a set of six probability operators is identified as those of conditional, multiplication, division, addition, subtraction, and complement operations in probability algebra. Each probability operator is formally defined and elaborated in the following subsections towards an algebraic framework of the theory of general probability.

Because the universe of discourse of general probability is constrained by three control factors in the Cartesian product $S \times R \times D$ as defined in Table 1, the contexts of probability algebra can be classified into four categories known as: *i*) Invariant sample space and disjoint/ME-dependent events $S \times \overline{R} \times D$; *ii*) Invariant sample space and joint/independent events $S \times R \times \overline{D}$; *iii*) Variant sample space and disjoint/independent events $S' \times \overline{R} \times \overline{D}$; and *iv*) Invariant sample space and joint/dependent events $S' \times R \times D$.

5.1 The Conditional Operator on Consecutive Probabilities

The conditional operation of consecutive probabilities deals with coupled influences between related events in both invariant and variant sample spaces. Because conditional probability forms the foundation for all other operators in the algebraic system of the general probability theory, it must be rigorously analyzed in order to avoid the dilemma of the cyclic definition as in classic probability theory.

Theorem 2. The *conditional operator on consecutive probabilities* of an event *b* influenced by that of a preceding event *a* in the sample space *S* in $\mathfrak{U}$, $P(b \mid a)$, is determined by a ratio between the changed sizes of sets of succeeding events *B'* in the variant sample space *S'* given $a \in A \subset S$, and $b \in B \subset S'$, i.e.:

$$
\begin{aligned}
&P(b \mid a) \triangleq P(A \to B) \\
&= \begin{cases}
i)\ \text{Invariant } S, \text{unrelated } \overline{R}, \text{ and ME-dependent } D\text{: } S \times \overline{R} \times D \\
\quad 0 \\
ii)\ \text{Invariant } S, \text{ related } R, \text{ and independent } \overline{D}\text{: } S \times R \times \overline{D} \\
\quad P(b) \\
ii)\ \text{Variant } S', \text{ unrelated } \overline{R}, \text{ and independent } \overline{D}\text{: } S' \times \overline{R} \times \overline{D} \\
\quad P'(b) = \dfrac{P(b)}{1 - P(a_i)},\ P(a_i) = \dfrac{P(a)}{|A|} \\
iv)\ \text{Variant } S', \text{ related } R \text{ and dependent } D\text{: } S' \times R \times D \\
\quad P''(b) = \dfrac{P(b) - P(b_i)}{1 - P(b_i)},\ P(b_i) = \dfrac{P(b)}{|B|}
\end{cases}
\end{aligned} \tag{19}
$$

Proof. Theorem 2 can be proven in each of the four contexts as defined in Table 1 according to Definition 10 as follows:

$$
\begin{aligned}
&\forall a, b, a \in A \subset S, \text{ and b} \in B \subset S', \\
&P(b \mid a) = P(A \to B) = \frac{|B'|}{|S \setminus a|}, a \in A \wedge b \in B \\
&= \begin{cases}
i)\ \forall S' = S \wedge A \cap B = \varnothing \wedge A \to (B = \varnothing, \text{ME}), \\
\quad \dfrac{|\varnothing|}{|S|} = 0 \\
ii)\ \forall S' = S \wedge A \cap B \neq \varnothing \wedge A \not\to B, \\
\quad \dfrac{|B|}{|S|} = P(b) \\
iii)\ \forall S' \neq S \wedge A \cap B = \varnothing \wedge A \not\to B, \\
\quad \dfrac{|B|}{|S|-1} = \dfrac{P(b)}{1 - P(a_i)} = P'(b),\ P(a_i) = \dfrac{P(a)}{|A|} \\
iv)\ \forall S \neq S \wedge A \cap B \neq \varnothing \wedge A \to B', \\
\quad \dfrac{|B|-1}{|S|-1} = \dfrac{P(b) - P(b_i)}{1 - P(b_i)} = P''(b),\ P(b_i) = \dfrac{P(b)}{|B|}
\end{cases}
\end{aligned} \tag{20}
$$

■

It is noteworthy in Theorem 2 that *S* is variant in general as constrained by Theorem 1 because of the coupling of the conditional events *a* and *b*. In other words, the classic probability in an invariant sample space is only a special case of that of the general variant context.

Example 8. In the *invariant* sample space $S_1 = \{H = 0.68,\ T = 0.32\}$ as modeled in Example 2, the events *head* (H) and *tail* (T) are mutually exclusive in a single toss of the coin. That is, both events cannot happen simultaneously. Once a head is observed, tail will certainly not appear in the same trail, and vice versa. This is a typical context of *mutually exclusive* ($H \cap T = \varnothing$ or *disjoint*), and *dependent* ($T \mid H = \varnothing$ or $H \mid T = \varnothing$) events of conditional probability according to Theorem 2(i) where $P(T \mid H) = 0,\ if\ H \cap T = \varnothing$ and $H \to (T = \varnothing)$.

Theorem 2(i) indicates that a pair of mutually exclusive events X and Y are dependent because $X \cap Y = \varnothing \wedge X \to (Y = \varnothing) \Rightarrow P(Y \mid X) = 0$ under the interactive influence between the ME events.

Example 9. Given a bag containing five black balls (B) and five white balls (W) in $S_2 = \{\underset{i=1}{\overset{5}{R}} P(b_i \mid b_i \in B) = 0.11,\ \underset{i=6}{\overset{10}{R}} P(w_i \mid w_i \in W) = 0.09\}$ as modeled in Example 3. Assume the ball drawn from the bag will be returned to the bag before the next trial, i.e., $S_2^{'} = S_2$, it is a case of invariant sample space, related and independent events of conditional probability according to Theorem 2(ii) as follows:

$$\begin{aligned} P(W \mid B) &= P(W) = 0.45 \\ P(B \mid W) &= P(B) = 0.55 \\ P(W \mid W) &= P(W) = 0.45 \\ P(B \mid B) &= P(B) = 0.55 \end{aligned}$$

Example 10. Reconsider Example 9 in $S_2^{'} = \{\underset{i=1}{\overset{5}{R}} P(b_i \mid b_i \in B) = 0.11,\ \underset{i=6}{\overset{10}{R}} P(w_i \mid w_i \in W) = 0.09\}$ where the ball drawn from the bag will not be returned, i.e., $S_2^{'} \neq S_2$, it becomes a case of *variant* sample space, disjoint/independent or joint/dependent events of conditional probability according to Theorem 2(iii) or 2(iv), respectively, as follows:

$$\begin{aligned} P(W \mid B) &= P'(W) = \frac{P(W)}{1 - P(b_i)},\ P(b_i) = \frac{P(B)}{|B|} = 0.55 / 5 = 0.11 \\ &= \frac{0.45}{1 - 0.11} = \frac{0.45}{0.89} = 0.51 \\ P(B \mid W) &= P'(B) = \frac{P(B)}{1 - P(w_i)},\ P(w_i) = \frac{P(W)}{|W|} = 0.45 / 5 = 0.09 \\ &= \frac{0.55}{1 - 0.09} = \frac{0.55}{0.91} = 0.60 \\ P(B \mid B) &= P''(B) = \frac{P(B) - P(b_i)}{1 - P(b_i)},\ P(b_i) = 0.11 \\ &= \frac{0.55 - 0.11}{1 - 0.11} = \frac{0.44}{0.89} = 0.49 \\ P(W \mid W) &= P''(W) = \frac{P(W) - P(w_i)}{1 - P(w_i)},\ P(w_i) = 0.09 \\ &= \frac{0.45 - 0.09}{1 - 0.09} = \frac{0.36}{0.91} = 0.40 \end{aligned}$$

Contrasting the results obtained in Examples 9 and 10, it is noteworthy that the conditional probabilities in Contexts (iii) and (iv) according to Theorem 2 have increased or decreased, respectively, due to the size shrinkages of sample spaces and/or the number of events as a result of the conditional coupling. The changes between the variant ($S_2^{'}$) and invariant (S_2) sample spaces can be rigorously analyzed as follows:

$$\begin{cases} P'(W \mid B) - P(W \mid B) = 0.51 - 0.45 = 0.06 \\ P'(B \mid B) - P(B \mid B) = 0.49 - 0.55 = -0.06 \end{cases}$$
$$\begin{cases} P'(B \mid W) - P(B \mid W) = 0.60 - 0.55 = 0.05 \\ P'(W \mid W) - P(W \mid W) = 0.40 - 0.45 = -0.05 \end{cases}$$

The results indicate that conditional probabilities in the variant and invariant sample spaces may be significantly different due to the increment or decrement of coupled event influences.

5.2 The Complement Operator in the Context of Probability Space

Theorem 3. The *complement of probability of an event* $a \in A \subset S$ in $\mathfrak{U}$, $P(\overline{a})$, is determined by the probability of all events in S excluding only that of a, i.e.:

$$P(\overline{a}) \triangleq 1 - P(a) \tag{21}$$

Proof. Theorem 3 can be proven according to Definition 10 as follows:

$$\begin{aligned} &\forall a \in A \subset S \text{ and } \overline{a} \in \overline{A} \subset S, \\ &P(\overline{a}) = P(\overline{A} \mid a \in A \subset S \wedge \overline{a} \in \overline{A} \subset S) \\ &\quad = \frac{|\overline{A}|}{|S|} = \frac{|S \setminus A|}{|S|} = \frac{|S|}{|S|} - \frac{|A|}{|S|} \\ &\quad = 1 - P(a) \end{aligned} \tag{22}$$

■

Example 11. On the basis of Example 4, the complement of probability in the sample space $S_1^2 = \{HH = 0.46,\ HT = 0.22,\ TH = 0.22,\ TT = 0.10\}$ can be determined according to Theorem 3 as follows:

$$\begin{aligned} P(\overline{HH}) &= 1 - P(HH) = 1 - 0.46 = 0.54 \\ P(\overline{TH}) &= 1 - P(TH) = 1 - 0.22 = 0.78 \end{aligned}$$

Corollary 3. A *double complement* of the general probability of an event $a \in A \subset S$ in $\mathfrak{U}, P(\overline{\overline{a}})$, results in an involution to the same probability, i.e.:

$$P(\overline{\overline{a}}) = 1 - P(\overline{a}) = 1 - (1 - P(a)) = P(a) \tag{23}$$

5.3 The Multiplication Operator on Disjunctive Probabilities

Theorem 4. The *multiplication of probabilities* of disjunctive events a and b in the sample space S in $\mathfrak{U}$, $P(a \times b)$, is determined by the product of the probabilities of $P(a)$ and $P(b|a)$ given $a \in A \subset S$ and $b \in B \subset S'$, i.e.:

$$P(a \times b) \triangleq P(A \wedge B) = P(a)P(b \mid a)$$
$$= \begin{cases} i) \text{ Invariant } S, \text{unrelated } \overline{R}, \text{ and ME-dependent } D\text{: } S \times \overline{R} \times D \\ \quad 0 \\ ii) \text{ Invariant } S, \text{ related } R, \text{ and independent } \overline{D}\text{: } S \times R \times \overline{D} \\ \quad P(a)P(b) \\ iii) \text{ Variant } S', \text{ unrelated } \overline{R}, \text{ and independent } \overline{D}\text{: } S' \times \overline{R} \times \overline{D} \\ \quad P(a)P'(b) = P(a)\dfrac{P(b)}{1-P(a_i)}, \ \ P(a_i) = \dfrac{P(a)}{|A|} \\ iv) \text{ Variant } S', \text{ related } R \text{ and dependent } D\text{: } S' \times R \times D \\ \quad P(a)P''(b) = P(a)\dfrac{P(b)-P(b_i)}{1-P(b_i)}, \ \ P(b_i) = \dfrac{P(b)}{|B|} \end{cases} \tag{24}$$

Proof. Theorem 4 can be proven according to Theorem 2 and Definition 10 as follows:

$$\forall a, b, a \in A \subset S, \text{ and b} \in B \subset S',$$
$$P(a \times b) = P(A \wedge B) = \frac{|A \cap B|}{|S|} = \frac{|A|}{|S|} \bullet \frac{|B \setminus a_i|}{|S \setminus a_i|}, a_i \in A$$
$$= P(a)P(b \mid a)$$
$$= \begin{cases} i) \ \ \forall S' = S \wedge A \cap B = \varnothing \wedge A \to (B = \varnothing, \text{ME}), \\ \quad 0 \\ ii) \ \ \forall S' = S \wedge A \cap B \neq \varnothing \wedge A \not\to B, \\ \quad P(a)P(b) \\ iii) \ \forall S' \neq S \wedge A \cap B = \varnothing \wedge A \not\to B, \\ \quad P(a)\dfrac{P(b)}{1-P(a_i)} = P(a)P'(b), \ \ P(a_i) = \dfrac{P(a)}{|A|} \\ iv) \ \forall S \neq S \wedge A \cap B \neq \varnothing \wedge A \to B', \\ \quad P(a)\dfrac{P(b)-P(b_i)}{1-P(b_i)} = P(a)P''(b), \ P(b_i) = \dfrac{P(b)}{|B|} \end{cases} \tag{25}$$

■

Example 12. Given an *invariant* sample space $S_1 = \{H = 0.68, \ T = 0.32\}$ as modeled in Example 2, i.e., $S_1^{'} = S_1$, the following disjunctive probabilities for two consecutive tosses of the uneven coin can be derived by probability multiplication according to Theorem 4(ii):

$$P(H \times T) = P(H)P(T) = 0.68 \bullet 0.32 = 0.22$$
$$P(H \times H) = P(H)P(H) = 0.68 \bullet 0.68 = 0.46$$
$$P(T \times H) = P(T)P(H) = 0.32 \bullet 0.68 = 0.22$$
$$P(T \times T) = P(T)P(T) = 0.32 \bullet 0.32 = 0.10$$

Example 13. Given a *variant* sample space $S_2 = \{\overset{5}{\underset{i=1}{R}} P(b_i \mid b_i \in B) = 0.11, \ \overset{10}{\underset{i=6}{R}} P(w_i \mid w_i \in W) = 0.09\}$ as modeled in Example 3, i.e., $S_2^{'} \neq S_2$, the following probability multiplications for two consecutive draws of the uneven balls in the bag can be obtained according to Theorem 4(iii) or 4(iv), respectively:

$$P(B \times W) = P(B)P'(W) = \frac{P(B)P(W)}{1-P(w_i)}, \ P(w_i) = 0.09$$
$$= \frac{0.68 \bullet 0.32}{1-0.09} = \frac{0.22}{0.91} = 0.24$$
$$P(W \times B) = P(W)P'(B) = \frac{P(W)P(B)}{1-P(b_i)}, \ P(b_i) = 0.11$$
$$= \frac{0.32 \bullet 0.68}{1-0.11} = \frac{0.22}{0.89} = 0.25$$
$$P(B \times B) = P(B)P''(B) = P(B)\frac{P(B)-P(b_i)}{1-P(b_i)}, \ P(b_i) = 0.11$$
$$= \frac{0.68(0.68-0.11)}{1-0.11} = \frac{0.39}{0.89} = 0.44$$
$$P(W \times W) = P(W)P''(W) = P(W)\frac{P(W)-P(w_i)}{1-P(bw)}, \ P(w_i) = 0.09$$
$$= \frac{0.32(0.32-0.09)}{1-0.09} = \frac{0.07}{0.91} = 0.08$$

Corollary 4. The *revisited Bayes' law of probability* can be rigorously derived based on Theorem 4 as follows:

$$\forall a, b, a \in A \subset S, \ b \in B \subset S', \text{ and } S' = S,$$
$$P(a \times b) = P(a)P(b \mid a), \ A \cap B \neq \varnothing \wedge A \not\to B$$
$$= P(a)P(b)$$
$$= P(b)P(a \mid b)$$
$$= P(b \times a)$$
$$\Rightarrow \begin{cases} \dfrac{P(b \mid a)}{P(b)} = \dfrac{P(a \mid b)}{P(a)}, \ \textit{iff} \ \ S' = S \wedge A \cap B \neq \varnothing \wedge A \not\to B \wedge B \not\to A \\ \dfrac{P(b \mid a)}{P(b)} \neq \dfrac{P(a \mid b)}{P(a)}, \ \text{Otherwise} \end{cases} \tag{26}$$

Corollary 4 and Theorem 4 indicate that Bayes' law in classic probability theory is a special case of general probability multiplication, which may only hold *iff* $S' = S \wedge A \cap B \neq \varnothing \wedge B \mid A = B$, i.e., when the conditions for invariant sample space and related but independent events are satisfied.

5.4 The Division Operator on Composite Probabilities

The algebraic operation of probability division is an inverse operation of probability multiplication, which was not defined in traditional probability theory.

Theorem 5. The *division of probability* of an event b by that of another event a in the sample space S in $\mathfrak{U}$, $P(b/a)$, is determined by the ratio of their probabilities where $a \in A \subset S$ and $b \in B \subset S'$ i.e.:

$$P(b/a) \triangleq P(\frac{|B|}{|A|}) = \frac{P(b)}{P(a)},\ 0 < P(a) \geq P(b)$$
$$= \begin{cases} i) \ \text{Invariant } S, \text{unrelated } \overline{R}, \text{ and ME dependent } D\text{: } S \times \overline{R} \times D \\ \quad 0 \\ ii) \ \text{Invariant } S, \text{ related } R, \text{ and independent } \overline{D}\text{: } S \times R \times \overline{D} \\ \quad \frac{P(b)}{P(a)} \\ iii) \ \text{Variant } S', \text{ unrelated } \overline{R}, \text{ and independent } \overline{D}\text{: } S' \times \overline{R} \times \overline{D} \\ \quad \frac{P'(b)}{P(a)} \\ iv) \ \text{Variant } S', \text{ related } R \text{ and dependent } D\text{: } S' \times R \times D \\ \quad \frac{P''(b)}{P(a)} \end{cases} \tag{27}$$

Proof. Theorem 5 can be proven according to Theorems 2 and 4 as well as Definition 10 as follows:

$$\forall a, b, a \in A \subset S, \text{ and } b \in B \subset S',$$
$$P(b/a) = P(|B|/|A|) = \frac{|B|}{|A|}$$
$$= \frac{|B|/|S|}{|A|/|S|},\ 0 < |A| \geq |B|$$
$$= \frac{P(b)}{P(a)},\ 0 < P(a) \geq P(b)$$
$$= \begin{cases} i) \ \forall S' = S \wedge A \cap B = \varnothing \wedge A \to (B = \varnothing, \text{ME}), \\ \quad \frac{|B|/|S|}{|A|/|S|} = \frac{P(b)}{P(a)} = \frac{0}{P(a)} = 0,\ A \to (B = \varnothing) \\ ii) \ \forall S' = S \wedge A \cap B \neq \varnothing \wedge A \not\to B, \\ \quad \frac{|B|/|S|}{|A|/|S|} = \frac{P(b)}{P(a)} \\ iii) \ \forall S' \neq S \wedge A \cap B = \varnothing \wedge A \not\to B, \\ \quad \frac{|B|/|S'|}{|A|/|S|} = \frac{P'(b)}{P(a)} = \frac{1}{P(a)} \frac{P(b)}{1 - P(a_i)},\ a_i \in A \\ iv) \ \forall S' \neq S \wedge A \cap B \neq \varnothing \wedge A \to B', \\ \quad \frac{|B'|/|S'|}{|A|/|S|} = \frac{P''(b)}{P(a)} = \frac{1}{P(a)} \frac{P(b) - P(b_i)}{1 - P(b_i)},\ b_i \in B \end{cases} \tag{28}$$

■

Example 14. In the *invariant* sample space $S_1 = \{H = 0.68,\ T = 0.32\}$ as modeled in Example 2, the events *head* (H) and *tail* (T) are mutually exclusive in a *single* toss of the unfair coin. Therefore, the following probability divisions of unrelated events can be obtained according to Theorem 5(i), respectively:

$$P(H/T) = 0$$
$$P(T/H) = 0$$

Example 15. Redo Example 14 with none-mutually-exclusive events in $S_1 = \{H = 0.68,\ T = 0.32\}$, the following probability divisions between those of two consecutive tosses and the first toss can be obtained according to Theorem 5(ii), respectively, as follows:

$$P(HT/H) = \frac{P(HT)}{P(H)} = \frac{0.22}{0.68} = 0.32$$
$$P(TH/T) = \frac{P(TH)}{P(T)} = \frac{0.22}{0.32} = 0.69$$
$$P(HH/H) = \frac{P(HH)}{P(H)} = \frac{0.46}{0.68} = 0.68$$
$$P(TT/T) = \frac{P(TT)}{P(T)} = \frac{0.10}{0.32} = 0.31$$

It is noteworthy that, according to Theorem 5(ii), the event of the divisor must not be mutually exclusive to that of the dividend. Otherwise, Theorem 5(i) should be applied such as in the cases of $P(HH/T) = 0, P(TT/H) = 0, P(HT/T) = 0,$ and $P(TH/H) = 0$ in the given context.

Example 16. Given a *variant* sample space $S_2^{2'} = \{BW = 0.28,\ WB = 0.27,\ BB = 0.27,\ WW = 0.18\}$ as modeled in Example 5, i.e., $S_2^{2'} \neq S_2^2$, the following probability divisions between two draws of the uneven balls in the bag can be obtained according to Theorem 5(iii) or 5(iv), respectively:

$$P(BW/B) = \frac{P'(BW)}{P(B)} = \frac{0.28}{0.55} = 0.51$$
$$P(WB/W) = \frac{P'(WB)}{P(W)} = \frac{0.27}{0.45} = 0.60$$
$$P(BB/B) = \frac{P''(BB)}{P(B)} = \frac{0.27}{0.55} = 0.49$$
$$P(WW/W) = \frac{P''(WW)}{P(W)} = \frac{0.18}{0.45} = 0.40$$

The results obtained in Example 16 can be verified by applying the multiplication rules given in Eq. 24(iii) and 24(iv) as shown in the following example. This approach is particularly useful when the product probability is unknown.

Example 17. Redo Example 16 in $S_2^{2'} = \{BW = 0.28, WB = 0.27, BB = 0.27, WW = 0.18\}$ according to Eq. 24(iii) and 24(iv) obtaining the same results as follows:

$$P(BW/B) = P'(W) = \frac{P(W)}{1-P(b_i)} = \frac{0.45}{1-0.11} = \frac{0.45}{0.89} = 0.51$$
$$P(WB/W) = P'(B) = \frac{P(B)}{1-P(w_i)} = \frac{0.55}{1-0.09} = \frac{0.55}{0.91} = 0.60$$
$$P(BB/B) = P''(B) = \frac{P(B)-P(w_i)}{1-P(b_i)} = \frac{0.55-0.11}{1-0.11} = \frac{0.44}{0.89} = 0.49$$
$$P(WW/W) = P''(W) = \frac{P(W)-P(w_i)}{1-P(w_i)} = \frac{0.45-0.09}{1-0.09} = \frac{0.36}{0.91} = 0.40$$

In probability theory, it is often interested in predicating the odds of random outcomes about the ratio of the probabilities of an event's success and failure.

Definition 11. An *odd*, $\Theta(e)$, is a ratio between probabilities of an event *e* and its complement, or that of its success s_e and failure f_e, i.e.:

$$\forall e, s_e, f_e \in E \subseteq S, \quad \Theta(e) \triangleq \frac{P(e)}{1-P(e)} = \frac{P(e)}{P(\bar{e})} = \frac{P(s_e)}{P(f_e)} \tag{29}$$

It is noteworthy that the range of odds is a nonnegative real number, i.e., $\Theta(e) \geq 0$, which may be great than 1.0 according to Definition 11.

5.5 The Addition Operator on Conjunctive Probabilities

Theorem 6. The *addition of probabilities* of two conjunctive events *a* or *b* in the sample space *S* in $\mathfrak{U}$, $P(a+b)$, is determined by the sum of the probabilities of $P(a)$ and $P(b)$ excluding that of the intersection $P(a \times b)$ given $a \in A \subset S$ and $b \in B \subset S'$, i.e.:

$$P(a+b) = P(A \vee B) = P(a) + P(b) - P(a)P(b\,|\,a)$$
$$= \begin{cases} i) \ \text{Invariant } S, \text{unrelated } \overline{R}, \text{ and ME-dependent } D\text{: } S \times \overline{R} \times D \\ \quad P(a) + P(b) \\ ii) \ \text{Invariant } S, \text{ related } R, \text{ and independent } \overline{D}\text{: } S \times R \times \overline{D} \\ \quad P(a) + P(b) - P(a)P(b) \\ iii) \ \text{Variant } S', \text{ unrelated } \overline{R}, \text{ and independent } \overline{D}\text{: } S' \times \overline{R} \times \overline{D} \\ \quad P(a) + P(b) - P(a)P'(b) \\ iv) \ \text{Variant } S', \text{ related } R \text{ and dependent } D\text{: } S' \times R \times D \\ \quad P(a) + P(b) - P(a)P''(b) \end{cases} \tag{30}$$

Proof. Theorem 6 can be proven according to Theorem 2 and Definition 10 as follows:

$$\forall a, b, a \in A \subset S, \text{ and b} \in B \subset S',$$
$$P(a+b) = P(A \vee B) = \frac{|A \cup B|}{|S|}$$
$$= \frac{|A|}{|S|} + \frac{|B|}{|S|} - \frac{|A \cap B|}{|S|}$$
$$= P(a) + P(b) - P(a)P(b\,|\,a)$$
$$= \begin{cases} i) \ \forall S' = S \wedge A \cap B = \varnothing \wedge A \to (B = \varnothing, \text{ME}), \\ \quad P(a) + P(b) \\ ii) \ \forall S' = S \wedge A \cap B \neq \varnothing \wedge A \not\to B, \\ \quad P(a) + P(b) - P(a)P(b) \\ iii) \ \forall S' \neq S \wedge A \cap B = \varnothing \wedge A \not\to B, \\ \quad P(a) + P(b) - \dfrac{P(a)P(b)}{1-P(a_i)}, \ P(a_i) = \dfrac{P(a)}{|A|} \\ \quad = P(a) + P(b) - P(a)P'(b) \\ iv) \ \forall S \neq S \wedge A \cap B \neq \varnothing \wedge A \to B', \\ \quad P(a) + P(b) - P(a)\dfrac{P(b)-P(b_i)}{1-P(b_i)}, \ P(b_i) = \dfrac{P(b)}{|B|} \\ \quad = P(a) + P(b) - P(a)P''(b) \end{cases} \tag{31}$$

■

Example 18. Reuse the individual probabilities obtained in Example 2 in the invariant sample space $S_1 = \{H = 0.68, T = 0.32\}$. The following additions of conjunctive probabilities for expecting some mixed head and tail of an unfair coin in two tosses can be derived according to Theorem 6(i):

$$P(HT + TH) = P(HT) + P(TH) - P(HT)P(TH)$$
$$= 0.24 + 0.24 - 0 = 0.48$$
$$P(HH + TT) = P(HH) + P(TT) - P(HH)P(TT)$$
$$= 0.36 + 0.16 - 0 = 0.52$$

Example 19. Suppose a system encompasses two components C_1 and C_2 with estimated failure rates as $F_1 = 0.7$ and $F_2 = 0.3$, respectively, in an invariant sample space. The conjunctive probabilities for a system failure of either C_1 or C_2 can be determined according to Theorem 6(ii) as follows:

$$P(F_1 + F_2) = P(F_1) + P(F_2) - P(F_1)P(F_2)$$
$$= 0.7 + 0.3 - 0.7 \bullet 0.3$$
$$= 1.0 - 0.21 = 0.79$$

Example 20. Consider the *variant* sample space $S_2^{2'} = \{BW = 0.28, WB = 0.27, BB = 0.27, WW = 0.18\}$ as modeled in Example 5 where no ball will be returned into the bag after a draw. The following probability additions between two conjunctive draws of the uneven balls in the bag can be obtained according to Theorem 6(iii) or 6(iv), respectively:

$$P(B+W)=P(B)+P(W)-\frac{P(B)P(W)}{1-P(w_i)}$$
$$=0.55+0.45-\frac{0.55\bullet 0.45}{1-0.11}=1-\frac{0.25}{0.89}=0.72$$
$$P(W+B)=P(W)+P(B)-\frac{P(W)P(B)}{1-P(w_i)}$$
$$=1-\frac{0.45\bullet 0.55}{1-0.09}=1-\frac{0.25}{0.91}=0.73$$
$$P(B+B)=P(B)+P(B)-\frac{P(B)(P(B)-P(b_i))}{1-P(b_i)}$$
$$=1.1-\frac{0.55(0.55-0.11)}{1-0.11}=1.1-\frac{0.24}{0.89}=0.83$$
$$P(W+W)=P(W)+P(W)-\frac{P(W)(P(W)-P(w_i))}{1-P(w_i)}$$
$$=0.9-\frac{0.45(0.45-0.09)}{1-0.09}=0.9-\frac{0.16}{0.91}=0.72$$

5.6 The Subtraction Operator on Decompositive Probabilities

The algebraic operation of probability subtraction is an inverse operation of probability addition, which was not defined in traditional probability theory.

Theorem 7. The *subtraction of related probability* of an event b from that of a in the sample spaces S in $\mathfrak{U}$, $P(a-b)$, is determined by the probability of event a excluding that of b given $a \in A \subset S$ and $b \in B \subset S'$, i.e.:

$$P(a-b) \triangleq P(a)-P(a)P(b\,|\,a)$$
$$=\begin{cases} i) & \text{Invariant } S, \text{unrelated } \overline{R}, \text{ and ME dependent } D\text{: } S\times\overline{R}\times D \\ & P(a) \\ ii) & \text{Invariant } S, \text{ related } R, \text{ and independent } \overline{D}\text{: } S\times R\times\overline{D} \\ & P(a)-P(a)P(b)=P(a)P(\overline{b}) \\ iii) & \text{Variant } S', \text{unrelated } \overline{R}, \text{ and independent } \overline{D}\text{: } S'\times\overline{R}\times\overline{D} \\ & P(a)-P(a)P'(b)=P(a)P'(\overline{b}) \\ iv) & \text{Variant } S', \text{ related } R \text{ and dependent } D\text{: } S'\times R\times D \\ & P(a)-P(a)P''(b)=P(a)P''(\overline{b}) \end{cases}$$

where $P(\overline{b})=1-P(b)$. (32)

Proof. Theorem 7 can be proven according to Theorem 2 and Definition 10 as follows:

$\forall a,b, a\in A\subset S,$ and $b\in B\subset S',$

$$P(a-b)=P(A\setminus B)=\frac{|A\setminus B|}{|S|}$$
$$=\frac{|A|}{|S|}-\frac{|A\cap B|}{|S|}$$
$$=P(a)-P(b\,|\,a)$$

$$=\begin{cases} i) & \forall S'=S\wedge A\cap B=\varnothing\wedge A\to(B=\varnothing,\text{ME}), \\ & P(a) \\ ii) & \forall S'=S\wedge A\cap B\neq\varnothing\wedge A\not\to B, \\ & P(a)-P(a)P(b)=P(a)(1-P(b))=P(a)P(\overline{b}) \\ iii) & \forall S'\neq S\wedge A\cap B=\varnothing\wedge A\not\to B, \\ & P(a)-\frac{P(a)P(b)}{1-P(a_i)},\ a_i\in A \\ & \quad =P(a)(1-\frac{P(a)P(b)}{1-P(a_i)})=P(a)(1-P'(b)) \\ & \quad =P(a)P'(\overline{b}) \\ iv) & \forall S'\neq S\wedge A\cap B\neq\varnothing\wedge A\to B, \\ & P(a)-P(a)\frac{P(b)-P(b_i)}{1-P(b_i)},\ b_i\in B \\ & \quad =P(a)(1-\frac{P(b)-P(b_i)}{1-P(b_i)})=P(a)(1-P''(\tilde{b})) \\ & \quad =P(a)P''(\overline{b}) \end{cases} \quad (33)$$

■

Example 21. Given the invariant sample space $S_1=\{H=0.68, T=0.32\}$ as modeled in Example 2, the following probability subtraction operations on the unfair coin can be derived according to Theorem 7(i) and 7(ii), respectively:

$$P(H-T)=P(H)-P(HT)=0.68-0=0.68 \quad // ME$$
$$P(T-H)=P(T)-P(TH)=0.32-0=0.32 \quad // ME$$
$$P(S_1-H)=P(S_1)P(\overline{H})=1\bullet(1-0.68)=0.32 \quad // H\subset S_1$$
$$P(S_1-T)=P(S_1)P(\overline{T})=1\bullet(1-0.32)=0.68 \quad // T\subset S_1$$
$$P(H-H)=P(\varnothing)=0$$
$$P(T-T)=P(\varnothing)=0$$

Example 22. Consider the *variant* sample spaces $S_2=\{\underset{i=1}{\overset{5}{R}}P(b_i\,|\,b_i\in B)=0.11,\ \underset{i=6}{\overset{10}{R}}P(w_i\,|\,w_i\in W)=0.09\}$ and $S_2^{2'}=\{BW=0.28, WB=0.27, BB=0.27, WW=0.18\}$, respectively, as modeled in Examples 3 and 5. The following probability subtraction operations on the uneven balls in the bag can be solved according to Theorem 7(iii), respectively:

$$P(B-W)=P(B)P'(\overline{W})=P(B)(1-\frac{P(W)}{1-P(b_i)})$$
$$=0.55(1-\frac{0.45}{1-0.11})=0.55\bullet 0.49=0.27$$
$$P(W-B)=P(W)P'(\overline{B})=P(W)(1-\frac{P(B)}{1-P(w_i)})$$
$$=0.45(1-\frac{0.55}{1-0.09})=0.45\bullet 0.40=0.18$$
$$P(BB-W)=P(BB)P''(\overline{W})=P(BB)(1-\frac{P(W)}{1-P(b_i)})$$
$$=0.27(1-\frac{0.45}{1-0.11})=0.27\bullet 0.49=0.13$$
$$P(WW-B)=P(WW)P''(\overline{B})=P(WW)(1-\frac{P(B)}{1-P(w_i)})$$
$$=0.18(1-\frac{0.55}{1-0.09})=0.18\bullet 0.40=0.07$$

Example 23. Given the same layout as that of Example 22 in a variant sample space with both related and dependent events, the following probability subtraction operations on the uneven balls in the bag can be solved according to Theorem 7(iv), respectively:

$$P(BW - B) = P(BW)P''(\overline{B}) = P(BW)(1 - \frac{P(B) - P(b_i)}{1 - P(b_i)})$$
$$= 0.28(1 - \frac{0.55 - 0.11}{1 - 0.11}) = 0.28 \bullet 0.51 = 0.14$$
$$P(WB - W) = P(WB)P''(\overline{W}) = P(WB)(1 - \frac{P(W) - P(w_i)}{1 - P(w_i)})$$
$$= 0.27(1 - \frac{0.45 - 0.09}{1 - 0.09}) = 0.27 \bullet 0.60 = 0.16$$

$$P(BB - B) = P(BB)P''(\overline{B}) = P(BB)(1 - \frac{P(B) - P(b_i)}{1 - P(b_i)})$$
$$= 0.27(1 - \frac{0.55 - 0.11}{1 - 0.11}) = 0.27 \bullet 0.51 = 0.14$$
$$P(WW - W) = P(WW)P'(\overline{W}) = P(WW)(1 - \frac{P(W) - P(w_i)}{1 - P(w_i)})$$
$$= 0.18(1 - \frac{0.45 - 0.09}{1 - 0.09} = 0.18 \bullet 0.60 = 0.11$$

Corollary 5. The *complement of probability* on an event $a \in A \subset S$ in $\mathfrak{U}$, $P(\overline{a})$, is a special case of probability subtraction, i.e.:

$$P(\overline{a}) = 1 - P(\overline{a}) \\ = P(S) - P(a) = P(S - a),\ a \in E \subseteq S \quad (34)$$

Table 4. Formal Rules of Probability Algebra

No.	Rule	Invariant sample space ($S' = S$)		Variant sample space ($S' \neq S$)
		Unrelated events ($A \cap B = \varnothing$)	**Related events** ($A \cap B \neq \varnothing$)	
1	Commutative	$P(b \mid a) \neq P(a \mid b)$	$\neq$	$\neq$
		$P(a \times b) = P(b \times a)$	$=$	
		$P(a / b) \neq P(b / a)$	$\neq$	
		$P(a + b) = P(b + a)$	$=$	
		$P(a - b) \neq P(b - a)$	$\neq$	
2	Associative	$P(a \mid (b \mid c)) \neq P((a \mid b) \mid c)$	$\neq$	$\neq$
		$P(a \times (b \times c)) = P((a \times b) \times c)$	$=$	
		$P(a / (b / c)) \neq P((a / b) / c)$	$\neq$	
		$P(a + (b + c)) = P((a + b) + c)$	$=$	
		$P(a - (b - c)) \neq P((a - b) - c)$	$\neq$	
3	Distributive	$P(a \times (b + c)) = P((a \times b) + (a \times c))$	$=$	$\neq$
		$P(a \times (b - c)) = P((a \times b) - (a \times c))$	$=$	
		$P((b + c) / a) = P((b / a) + (c / a)),\ P(a) > 0$	$=$	
		$P((b - c) / a) = P((b / a) - (c / a)),\ P(a) > 0$	$=$	
4	Transitive	$P(a) = P(b) \wedge P(b) = P(c) \Rightarrow P(a) = P(c)$	$=$	$=$
5	Complement	$P(\overline{a}) = 1 - P(\overline{a})$, $P(\overline{a}) + P(\overline{a}) = 1$ $P(S) = 1$, $P(\varnothing) = 0$ $P(\overline{S}) = 0$, $P(\overline{\varnothing}) = 1$		
6	Involution	$P(\overline{\overline{a}}) = P(a)$		
7	Idempotent	$P(a \times a) = P(a)$, $P(a + a) = P(a)$ $P(a / a) = 1$, $P(a - a) = 0$		
8	Identity	$P(a \times S) = P(a)$, $P(a \times \varnothing) = 0$ $P(a / S) = P(a)$, $P(S / a) = P^{-1}(a)$ $P(a + S) = 1$, $P(a + \varnothing) = P(a)$ $P(S - a) = P(S) - P(a)$, $P(\varnothing / a) = 0$ $P(a - \varnothing) = P(a)$, $P(\varnothing - a) = 0$		

6. Formal Properties and Rules of the General Probability Theory

The mathematical model of general probability, the framework of the revisited probability theory, and the formal operators of probability algebra enable rigorous analyses of the nature, properties, and rules of probabilities as well as their algebraic operations. A set of 36 algebraic properties and rules of probability algebra is summarized in Table 4.

Basic rules of probability algebra in the universe of discourse of probability $\mathfrak{U}$ can be expressed in categories of the *commutative, associative, distributive, transitive, complement, involution, idempotent,* and *identity* rules. It is noteworthy that it is unnecessary that each of the probability operators obeys all the general algebraic rules. Each algebraic rule on probability multiplication, division, addition, subtraction, conditional, and complement operations can be proven by applying specific definitions and arithmetic principles. The algebraic rules of the probability theory may be applied to derive and simply complex probability operations in formal probability manipulations and uncertainty reasoning by both humans and cognitive systems. The framework of the revisited probability theory reveals that classic probability theory is a special case or subsystem of the revisited probability theory in terms of both mathematical models and probability operations.

7. Conclusion

A revisited theory of probability and a mathematical structure of probability algebra have been rigorously introduced as an extension of the classic probability theory in order to deal with complicated dynamic sample spaces as well as complex event relations and dependencies. The general probability theory has been formally described as a framework of hyperstructures of dynamic probability and their algebraic operations. Mathematical models and formal operators of probability algebra have enabled rigorous analyses of the nature, properties, and rules of probability theories and their algebraic operations.

It has been found that the conditional probability played a centric role in the framework of probability theories in order to solve the highly coupled problems of cyclic definitions in traditional probability theories. It has been proven that Bayes' law may be revisited based on the properties of the variant sample spaces as revealed in this paper. This work has also led to a theory of fuzzy probability that extends the general probability theory onto fuzzy probability spaces and fuzzy algebraic operations.

Acknowledgment: The author would like to acknowledge the support in part by a discovery grant of the Natural Sciences and Engineering Research Council of Canada (NSERC). The author would like to thank the anonymous reviewers for their valuable suggestions and comments on the previous version of this paper.

References:

[1] Artin, M. (1991), *Algebra*, Prentice-Hall, Inc., Englewood Cliffs, USA.

[2] Bender, E.A. (1996), *Mathematical Methods in Artificial Intelligence*, Los Alamitos, CA: IEEE CS Press.

[3] BISC (2013), *Internal Communications*, University of California, Berkeley, CA.

[4] Cantor, G. (1874), On a Characteristic Property of all Real Algebraic Numbers, *J. Reine Angew. Math.*, 77, 258- 262.

[5] di Finetti, B. (1970), *Theory of Probability*, John Wiley & Sons, Inc., NY.

[6] Gowers, T. ed. (2008), *The Princeton Companion to Mathematics*, Princeton University Press, NJ.

[7] Hacking, I. (1975), *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Inductions and Statistical Inference*, Cambridge Univ. Press, Cambridge, UK.

[8] Johnson, R.A. and G.K. Bhattacharyya (1996), *Statistics: Principles and Methods*, 3rd ed., John Wiley & Sons, Inc., NY.

[9] Kolmogorov, A.N. (1933), *Foundations of the Theory of Probability*, 2nd English ed. (1956), Chelsea Publishing Co., NY.

[10] Lipschutz, S. and M. Lipson (1997), *Theory and Problems of Discrete Mathematics*, 2nd ed., McGraw-Hill, NY.

[11] Mosteller, F. (1987), *Fifty Challenging Problems in Probability*, Dover Publishing, Inc., NY.

[12] Novak, V., I. Perfilieva, and J. Mockor (1999), Mathematical Principles of Logic, Kluwer, Boston/Dordrecht.

[13] Pedrycz, W. and F. Gomide (1998), Introduction to Sets, Cambridge, MA: MIT Press.

[14] Potter, M. (2004), *Set Theory and Its Philosophy: A Critical Introduction,* Oxford University Press.

[15] Ross, T.J. (1995), *Logic with Engineering Applications*, McGraw-Hill Co., 1995.

[16] Todhunter, I. (1865), *A History of the Theory of Probability from the Time of Pascal to that of Laplace*, Reprinted 1965, Chelsea Publishing Co., NY.

[17] Venn, J. (1888), *The Logic of Chance*, 3rd ed., Reprinted 1962, Chelsea Publishing Co., NY.

[18] Wang, Y. (2007), *Software Engineering Foundations: A Software Science Perspective*, CRC

Series in Software Engineering, Vol. II, Auerbach Publications, NY, USA.
[19] Wang, Y. (2008a), On Contemporary Denotational Mathematics for Computational Intelligence, *Trans. Computational Science*, 2, 6-29.
[20] Wang, Y. (2008b), On the Big-R Notation for Describing Iterative and Recursive Behaviors, *International Journal of Cognitive Informatics and Natural Intelligence*, 2(1), Jan., 17-28.
[21] Wang, Y. (2011), Inference Algebra (IA): A Denotational Mathematics for Cognitive Computing and Machine Reasoning (I), *International Journal of Cognitive Informatics and Natural Intelligence*, 5(4), 61-82.
[22] Wang, Y. (2012a), In Search of Denotational Mathematics: Novel Mathematical Means for Contemporary Intelligence, Brain, and Knowledge Sciences, *Journal of Advanced Mathematics and Applications*, 1(1), 4-25.
[23] Wang, Y. (2012b), Formal Rules for Causal Analyses and Inferences, *International Journal on Software Science and Computational Intelligence*, 4(4), 70-86.
[24] Wang, Y. (2012c), Inference Algebra (IA): A Denotational Mathematics for Cognitive Computing and Machine Reasoning (II), *International Journal of Cognitive Informatics and Natural Intelligence*, 6(1), 21-46.
[25] Wang, Y. (2014a), The Theory of Arithmetic in the Extended Domain of Numbers, *Journal of Advanced Mathematics and Applications,* 3(2), 165-175.
[26] Wang, Y. (2014b), Causal Inferences based on Semantics of Concepts in Cognitive Computing, *WSEAS Transactions on Computers*, 13, 430-441.
[27] Wang, Y. (2014c), Towards a Theory of Probability for Cognitive Computing, Proc. 13th IEEE International *Conference on Cognitive Informatics and Cognitive Computing* (ICCI*CC 2014), London, UK, IEEE CS Press, Aug., pp. 21-29.
[28] Wang, Y. (2014d), Keynote: Latest Advances in Neuroinformatics and Systems, *Proceedings of 2014 International Conference on Neural Networks and Systems,* Venice, Italy, March, pp. 14-15.
[29] Wang, Y. (2015a), Towards the Abstract System Theory of System Science for Cognitive and Intelligent Systems, *Springer Journal of Complex and Intelligent Systems,* June, 1(3), 11-32.
[30] Wang, Y. (2015b), Formal Cognitive Models of Data, Information, Knowledge, and Intelligence, *WSEAS Transactions on Computers*, 14, 770-781.
[31] Wang, Y. (2015c), Keynote: Cognitive Robotics and Mathematical Engineering, *14th IEEE International Conference on Cognitive Informatics and Cognitive Computing* (ICCI*CC 2015), Tsinghua University, Beijing, China, IEEE CS Press, July, pp. 4-6.
[32] Wang, Y. (2015d), Towards a Logical Algebra (FLA) for Formal Inferences in Cognitive Computing and Cognitive Robotics, *14th IEEE International Conference on Cognitive Informatics and Cognitive Computing* (ICCI*CC 2015), Tsinghua University, Beijing, China, IEEE CS Press, July, pp. 24-32.
[33] Wang, Y. (2015e), Fuzzy Probability Algebra (FPA): A Theory of Fuzzy Probability for Fuzzy Inference and Computational Intelligence, *Journal of Advanced Mathematics and Applications*, 4(1), 38-55.
[34] Wang, Y. (2016a), On Cognitive Foundations and Mathematical Theories of Knowledge Science, International Journal of Cognitive Informatics and Natural Intelligence, 10(2), 1-24.
[35] Wang, Y. (2016b), Big Data Algebra (BDA): A Denotational Mathematics for Big Data Science and Engineering, Journal of Advanced Mathematics and Applications, 5(1), in press.
[36] Whitworth, W.A. (1959), *Choice and Chance*, Hafner Publishing Co.
[37] Zadeh, L.A. (1965), Sets, *Information and Control*, 8, 338-353.
[38] Zadeh, L.A. (1968), Probability Measures of Events, *Journal of Math. Analysis and Appl.* 23, 421-427.
[39] Zadeh, L.A. (1996), Logic and the Calculi of Rules and Graphs, *Multiple-Valued Logic, 1,* 1-38.
[40] Zadeh, L.A. (2002), Toward a Perception-based Theory of Probabilistic Reasoning with Imprecise Probabilities**,** *Journal of Statistical Planning and Inference*, Elsevier Science, 105, 233-264.