Classic Probability Revisited (I): Mathematical Models of an Extended Probability Theory

YINGXU WANG
International Institute of Cognitive Informatics and Cognitive Computing (ICIC)
Laboratory for Computational Intelligence, Software Science, and Denotational Mathematics
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

Abstract: - Part I of this paper presents a set of extended mathematical models of probability theory in order to explain the nature, properties, and rules of general probability. It is found that probability is a hyperstructure beyond those of the traditional monotonic and one-dimensional discrete structures. The sample space of probability is not invariant in general cases. Types of vents in the sample space may be refined as joint or disjoint and dependent, independent, or mutually-exclusive. These newly identified properties lead to a three-dimensional dynamic model of probability structures constrained by the type of sample spaces, the relation of events, and the dependency of events. A set of algebraic operators on the mathematical structures of the general probability theory is derived based on the extended mathematical models of probability. It is revealed that the Bayes' law needs to be extended in order to fit more general contexts on variant sample spaces and complex event properties in fundamental probability theories. The revisited probability theory enables a rigorous treatment of 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, and brain informatics.

Key-Words: - Denotational mathematics, probability theory, probability algebra, fuzzy probability, formal inference, cognitive informatics, cognitive computing, computational intelligence, semantic 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 probability theory. Theories of probability in its modern form was unified by Pierre Simon and Marquis de Laplace in the 19^{th} 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 probability estimations and norms in a certain sample space. The main methodology of classic probability theory is an external or 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 merely reduced to {0, 1} immediately after the given event has realized in a certain probability space.

It is recognized that the classic probability theory is cyclically defined among a set of highly coupled operations where only logically conjunctive, disjunctive, and conditional events are considered. This paper presents a 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. Formal properties of general probability are elicited and summarized in Section 4.

Due to its excessive length, this paper is presented in two parts on: i) The mathematical models of general probability; and ii) The algebraic operations of the extended probability theory. This paper is the first part of the revisited probability theory on the mathematical models of general probability in the hyperstructure of probability sample space.

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. Because of the cyclically defined framework of classic probability theory, related literature and textbooks describe the highly coupled probability operations in various incoherent 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.:

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

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.

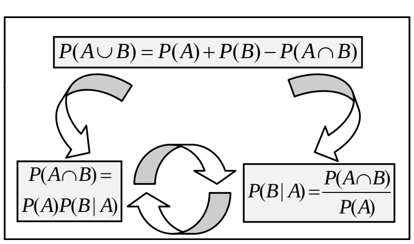

Fig. 1. The highly coupled dependency in classic probability theory

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.:

$$P(A \cup B) \to P(A \cap B) \leftrightarrow P(B \mid A) \quad (2)$$

The highly coupled dependency between key probability operators results in numerous problems in probabilistic reasoning, theorem proving, and applications in 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 analyzed and elaborated in this subsection.

Example 1. Assume a bag has a black ball and a white ball denoted by two events of 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 $|S'| = 1$.

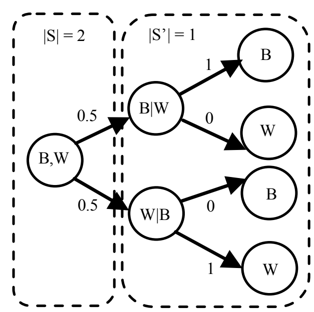

Fig. 2. Conditional probability in variant sample spaces

This is why most illustrations and case studies on classic probability theory request that 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. The *sample space* of probabilistic events as the context of probability is variant in general in a series of probabilistic trails, i.e.:

$$S(E_1, E_2) \neq S(E_2 \mid E_1) \neq S(E_1 \mid E_2) \quad (3)$$

where E_1 and E_2 are two sets of arbitrary events, $S(E_1, E_2)$, $S(E_2 | E_1)$, and $S(E_1 | E_2)$ 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 can be proven by considering the sizes of the sample space under four different scenarios as follows:

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

According to Lemma 2, the basic assumption of classic probability theory on invariant sample spaces is merely a simplified special case of that of the general dynamic sample space. The generic and typical layout of the variant sample space is illustrated in Fig. 2. The finding in Lemma 2 indicates that *Bayes' law of conditional probability* is not generally true in the variant probability space, which will be proven by Corollary 2 in Part II of this paper.

3. Mathematical Models of the General Probability Theory

The mathematical model of the extended probability theory is defined on the universe of discourse of probability, the dynamic sample space, and 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 of probability is a foundation that clarifies the general context and layout of probability theory.

Definition 2. The *set of states,* Ξ, with individual bivalent probabilistic status ξ_i, $\xi_i \in \Xi$, $1 \le i < |\Xi|$, 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.:

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

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

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

The universe of discourse of probability can be modeled based on the three essences as introduced in Definitions 2 to 4.

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

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

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 Extended Probability Theory

On the basis of the universe of discourse of general probability, fundamental properties of events and sample spaces of probability theory are formally analyzed in this subsection.

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:

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

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

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

where ME is a special type of event dependency in which the sets of events never appear simultaneously or concurrently.

It is noteworthy that event dependency is different from 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. The *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 $S \subset \text{Þ}\mathfrak{E}$ in 𝔘, i.e.:

$$\begin{aligned} S &\triangleq \{\overset{n}{\underset{i=1}{R}} E_i \mid E_i \subseteq \text{Þ}S \subset \text{Þ}\mathfrak{E}\} \\ &= \{\overset{n}{\underset{i=1}{R}}\overset{n_i}{\underset{j=1}{R}} e_{ij} \mid e_{ij} \in E_i \subseteq \text{Þ}S \subset \text{Þ}\mathfrak{E}\} \end{aligned} \quad (11)$$

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

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

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

which is invariant, disjoint, and mutually-exclusive.

Example 3. Given a complex 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:

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

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 𝔘, i.e.:

$$|S| \triangleq |\bigcup_{i=1}^{n} E_i| = |\bigcup_{i=1}^{n}\{\overset{n_i}{\underset{j=1}{R}} e_{ij} \mid \overset{n}{\underset{i=1}{R}}\overset{n_i}{\underset{j=1}{R}} e_{ij} \in E_i \subseteq \text{Þ}S\}| \quad (12)$$

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

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

■

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 a Cartesian product $\overset{m}{\underset{i=1}{\times}} S_i$, $S_i \subseteq \text{Þ}\mathfrak{E}$ in 𝔘, i.e.:

$$\begin{aligned} S^m &\triangleq \overset{m}{\underset{i=1}{\times}} S_i \\ &= \{\overset{n_1}{\underset{i_1=1}{R}}\overset{n_2}{\underset{i_2=1}{R}}\cdots\overset{n_m}{\underset{i_m=1}{R}}(e_{i_1i_2\ldots i_m} \mid e_{i_1i_2\ldots i_k\ldots i_m} \in S^m \wedge e_{i_k} \in E_{i_k} \subseteq S_{i_k})\} \end{aligned}$$

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

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

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 1 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 in Part I, 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 combined events $HH, HT, TH, \text{and}\, TT$ as follows:

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

Example 5. Consider the case analyzed 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 due to the roughness between balls in different colors. On the basis of $S_2 = \{\mathop{R}_{i=1}^{5} W = 0.09, \mathop{R}_{i=6}^{10} B = 0.11\}$, 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:

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

where the variant sample space $S_2^{2'}$ encompasses four complex events with modified event probabilities by reflecting the influences between sequential or conditional events on the second 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 $E_1 \rightarrow E_2$ due to the removal or disappearance of an event e_i, $e_i \in E_1$ after the previous trial, i.e.:

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

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

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

■

It is noteworthy that the term of the variant sample space $\widetilde{S}' \neq \widetilde{S}$ as given in Theorem 1 refers to the variable distributions of probabilities for individual events in a series of experiments due to the impact of previous events.

Theorem 1 reveals that *Bayes' law of conditional probability* is not generally true in the variant probability space, which will be proven by Corollary 2 in Part II. This explains why almost all textbooks on classic probability theory request that, in a typical layout, a ball drawn from the bag must be returned after each trail. This treatment has over simplified the general context of probability theory. Many problems and exceptions of classic probability theory stemmed from the overlooking of the principle of variant sample spaces of probability as stated in Theorem 1.

3.3 Mathematical Model of the Extended Probability Theory

Probability is a mathematical structure built on the basis of set theory and the exploration of 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. Given the sample space $S_2 = \{\overset{5}{\underset{i=1}{R}} W = 0.09, \overset{10}{\underset{i=6}{R}} B = 0.11\}$ 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 constrained by the unit size, i.e.:

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

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*)	**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	$S' \times R \times D$		$X \cap Y \neq \varnothing$	$X \to (Y = Y')$

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* demotes 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.: a) *invariant* sample space and *disjoint/*ME-*dependent* events, b) *invariant* sample space and *joint/independent* events, c) *variant* sample space and *disjoint/independent* events, and d) *invariant* sample space and *joint/dependent* events.

It will be demonstrate and proven in Section 2 in Part II that any complex probability can be expressed by an algebraic operation of the primitive single variable probabilities in the theory of general probability.

4. The Framework of the General Probability Theory

On the basis of the mathematical models of general probability and its discourse as defined in preceding subsections, 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 mainly covers a special case of the general probability in the invariant sample space with mainly joint and independent events.

Each probability operator 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 Part II.

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 the singularities of the entire and empty sample spaces, respectively, where $\mathfrak{E} \sqsubseteq \mathfrak{U}$ 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.

Table 2. The Framework of Algebraic Operators of the General Probability Theory

No.	Operator	Definition				
		General	Invariant sample space (Classic) $(S' = S)$		Variant sample space (Revisited) $(S' \neq S)$	
		$(a \in A \subset S \wedge b \in B \subset S')$	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	Conditional	$P(b \mid a) \triangleq P(A \to B)$	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. Properties of General Probability

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

The mathematical model of probability, the framework of 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 operational properties and rules of probability algebra will be derived in Section 3 in Part II. This leads to the revealing and proof of 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.

5. Conclusion

As the first part of the revisited probability theory, a general theory of probability has been rigorously introduced as an extension of the classic probability to deal with complicated dynamic sample spaces as well as complex event relations and dependencies. The revisited probability theory has been formally described as a framework of hyperstructures of dynamic probability and their algebraic operations. Mathematical models and formal operators of the general probability framework 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 cyclic-definition problems in traditional probability theories. It has been proven that Bayes' law may be revisited and validated 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 further extends the general probability theory to fuzzy probability spaces and fuzzy algebraic operations.

Acknowledgment: The author would like to acknowledge the support in part by a discovery fund 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 University 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] Whitworth, W.A. (1959), *Choice and Chance*, Hafner Publishing Co.
[35] Zadeh, L.A. (1965), Sets, *Information and Control*, 8, 338-353.
[36] Zadeh, L.A. (1968), Probability Measures of Events, *Journal of Math. Analysis and Appl.* 23, 421-427.
[37] Zadeh, L.A. (1996), Logic and the Calculi of Rules and Graphs, *Multiple-Valued Logic, 1*, 1-38.
[38] 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.