Infinitesimal Probabilities Based on Grossone

C. S. Calude[1] and M. Dumitrescu[2]
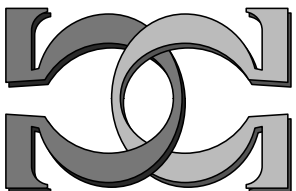
[1]University of Auckland, New Zealand
[2]University of Bucharest, Romania
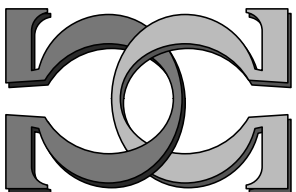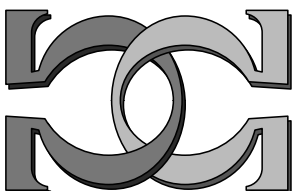
Infinitesimal Probabilities Based on Grossone

Cristian S. Calude*   Monica Dumitrescu†

October 17, 2019

Abstract

In finite probability theory the only probability-zero event is the impossible one, but in standard Kolmogorov probability theory probability-zero events occur all the time. Prominent logicians, probability experts and philosophers of probability, including Carnap, Kemeny, Shimony, Savage, De Finetti, Jeffrey, have successfully argued that a sound probability should be regular, that is, only the impossible event should have zero probability. This intuition is shared by physicists too. Totality is another desideratum which means that every event should be assigned a probability. Regularity and totality are achievable in rigorous mathematical terms even for infinite events via hyper-reals valued probabilities. While the mathematics of these theories is not objectionable, some philosophical arguments purport to show that infinitesimal probabilities are inherently problematic.

In this paper we present a simpler and natural construction – based on Sergeyev's calculus with Grossone (in a formalism inspired by Lolli) enriched with infinitesimals – of a regular, total, finitely additive, uniformly distributed probability on infinite sets of positive integers. These probability spaces – which are inspired by and parallel the construction of classical probability – will be briefly studied. In this framework De Finetti fair lottery has the natural solution and Williamson's objections against infinitesimal probabilities are mathematically refuted.

1 Introduction

The gap between finite and infinite probabilities has been highlighted by probability-zero events since the Kolmogorov axiomatic has been proposed, see [2].

The standard probability theory is a mathematically rigorous, consistent theory with an immense range of applications which demonstrate its usefulness in modelling many situations in the physical world [12]. However, there are infinite probabilistic scenarios that cannot be described in a satisfactory way in terms of this theory. One example is the De Finetti fair lottery in which draws are taken from the infinite set $\{1, 2, 3, \dots\}$ with equal non-zero probability.

In this paper we use Grossone [15–18] in a formalism inspired by Lolli [9] to construct and study a class of infinitesimal probabilities. We prove that these probabilities satisfy the properties regularity, totality, perfect additivity and weak Laplacean-ity discussed in [2], provide the natural solution to De Finetti infinite fair lottery, Williamson's objections [21] against infinitesimal probabilities are mathematically refuted.

In what follows $\mathbb{N}$ is the set of positive integers, $\mathbb{R}$ is the set of reals, the power set of X is denoted by $\mathcal{P}(X)$ and the number of elements of a finite set is indicated by #.

In Sections 2 and 3 we present random events, classical probabilities, Kolmogorov axioms and natural probability desiderata. Section 4 is a brief informal presentation of Grossone which is followed by Lolli's formalism in Section 5. De Finetti's infinite lottery is discussed in Section 6. The main Section 7 is devoted to Grossone-like uniform probability spaces. The last two Sections 8 and 9 are dedicated to non-uniform Grossone probabilities and conclusions.

---

*School of Computer Science, University of Auckland, Auckland, New Zealand.

†Faculty of Mathematics and Computer Science, Bucharest University, Romania.

2 Random events and classical probabilities

Informally, *a random experiment* $\mathcal{E}$ (process/variable) is an experiment that is not and cannot be made exact and, consequently, its outcomes cannot be predicted. Carrying out a random experiment is called a *trial*. An outcome obtained as a result of a trial is called an *elementary event*. The set of all possible outcomes of a random experiment is called the *sample space* and is denoted by Ω.

The subsets of Ω are called *events* and the family of events is denoted by $\mathcal{P}(\Omega)$. For example, rolling two dice is a random experiment $\mathcal{E}$; the sample space is the set $\Omega = \{(i,j) \mid 1 \leq i, j \leq 6\}$. This experiment has 6^2 elementary events $\{(i,j)\}$. The pair $(1,6)$ is one possible outcome of a trial. The event "the sum of the two dice is equal to 3" is $A = \{(1,2),(2,1)\}$. Events can be combined with usual set-theoretical operations of union, intersection and complement. The empty set, denoted by $\emptyset$, is the *impossible* event. To conclude, to a random experiment $\mathcal{E}$ we associate the mathematical object $(\Omega, \mathcal{P}(\Omega))$.

Let us consider a random experiment $\mathcal{E}$ and an event $A \in \mathcal{P}(\Omega)$. If we perform n trials of $\mathcal{E}$, then the ratio between the number occurrences of A and n is called the "frequency of A in n trials", and is denoted by $f_n(A)$. It is an empirical fact that, when n increases, $f_n(A)$ tends to "stabilise" to a number called "the probability" of A, which is denoted by $P(A)$. If instead of an arbitrary A we consider the elementary events, performing n trials will produce *a frequency distribution* for elementary events; this distribution will approach *a probability distribution* as the number of trials increases. Since $f_n(A \cup B) = f_n(A) + f_n(B)$ whenever $A \cap B = \emptyset$, it is natural to assume that $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$ and $P(\Omega) = 1$.

In the example of rolling two dice we have 6^2 elementary events. If the dice are "fair", the probability of any elementary event is $P((i,j)) = 1/36$.

When Ω is a finite set and $A \in \mathcal{P}(\Omega)$, then the probability defined by $P(A) = \frac{\#(A)}{\#(\Omega)}$ is called the *classical probability*. For every event $A \in \mathcal{P}(\Omega)$, $P(A) = 0$ if and only if $A = \emptyset$. See more in Nelson [11].

3 Kolmogorov axioms and probability desiderata

Kolmogorov axioms are the standard framework for probability theory. The sample space, that is, the set of (atomic) outcomes is denoted by Ω. A *measurable space* is a couple $(\Omega, \mathcal{B}(\Omega))$ where the *event space* is a Borel field of subsets of Ω, $\mathcal{B}(\Omega) \subseteq \mathcal{P}(\Omega)$. A *probability space* is a triple $(\Omega, \mathcal{B}(\Omega), \Pr)$, where $(\Omega, \mathcal{B}(\Omega))$ is a measurable space and $\Pr\colon \mathcal{B}(\Omega) \longrightarrow [0,1]$ is a probability measure, that is, $\Pr$ satisfies the following two conditions: a) the probability of a countable union of mutually-exclusive sets in $\mathcal{B}(\Omega)$ is equal to the countable sum of the probabilities of each of these sets, and b) $\Pr(\Omega) = 1$. In case condition a) is satisfied only for finitely-many mutually-exclusive sets, the probability space will be called *finitely-additive*.

Kolmogorov axioms for infinite sample spaces do not respect any of the following – arguably natural – desiderata [2]:

1. *regularity*: the probability of every possible event, that is, any non-empty subset of Ω, should be strictly larger than that of the impossible event[1],

2. *totality*: every element of the event space should be assigned a probability value,

3. *perfect additivity*: the probability of an arbitrary union of mutually disjoint events should be equal to the sum of the probabilities of the separate events, where "sum" has to be appropriately defined,

---

[1]This is a special case of the Euclidean principle requiring that any set should have a strictly larger probability than each of its strict subsets. It is easy to see that every regular, total and finitely-additive probability is Euclidean: indeed, for all $X, Y \subseteq \Omega$, if $X \subset Y$, then because $Y = X \cup (Y \setminus X)$ and $Y \setminus X \neq \emptyset$, then by finite additivity of the mutually disjoints events X and Y, we have: $\Pr(Y) = \Pr(X) + \Pr(Y \setminus X) > \Pr(X)$ as $\Pr(Y \setminus X) > 0$.

4. *weak Laplacean-ity*: the probability theory should allow for a uniform probability distribution on the sample space.[2]

De Finetti fair lottery, in which draws are taken from the infinite set $\{1, 2, 3, \dots\}$ with equal probability, is a classical example in which Kolmogorov axioms fail the above desiderata [2, 20].

The above desiderata are achievable in rigorous mathematical terms even for infinite events and one way to do it is to use hyper-reals valued probabilities, see for example NAP [1, 2].

In what follows we will present a simpler and natural construction of infinitesimal probabilities based on Grossone inspired from the classical probability which satisfies all above desiderata.

4 An intuitive glance at Grossone

The set of positive integers (shortly, *numbers* in what follows)

$$\mathbb{N} = \{1, 2, 3, \dots\} \tag{1}$$

can be represented by different numeral systems. The most common ones use a positional numeral system with bases 10 or 2. In some sense these systems "reveal" the same information about the set (1). The system proposed by Sergeyev [15–18] gives more information about (1) by using a formalism supporting *Aristotles Principle: "the whole is greater than its parts"*.[3] To this aim a new infinite number ① (called *Grossone*) is introduced: informally, ① *is the number of elements of the set of numbers* (1).

The infinite number ① can be used to calculate the number of elements of various infinite sets of numbers. For example, the number of elements of the set $\mathbb{N} \setminus \{10\}$ is $① - 1$; the sets of even and odd numbers have the same number of elements, namely $\frac{①}{2}$; the set $\{1\} \cup \{2n \mid n \in \mathbb{N}\}$ has $① + 1$ elements. The elements $①, ① + 1, ① - 1, \frac{①}{2}$ and many others like

$$\cdots, \frac{①}{2} - 2, \frac{①}{2} - 1, \frac{①}{2}, \frac{①}{2} + 1, \frac{①}{2} + 2, \cdots, ① - 2, ① - 1, ①$$

can be thought as infinite numbers.

Accordingly, we can "magnify" $\mathbb{N}$ to the set

$$\mathbb{N}^\dagger = \left\{ 1, 2, \cdots, \frac{①}{2} - 2, \frac{①}{2} - 1, \frac{①}{2}, \frac{①}{2} + 1, \frac{①}{2} + 2, \cdots, ① - 2, ① - 1, ① \right\}. \tag{2}$$

We may think of the sets (1) and (2) as the results of two different experiments performed on the set of positive integers.[4] The set $\mathbb{N}^\dagger$ will be used to "measure" the sizes of different, finite or infinite, subsets of $\mathbb{N}$.

---

[2]A more general, but less formal, form is to ask that the probability theory should allow for a mathematical representation of any probabilistic situation that is conceptually possible.

[3]Aristotles Principle is not satisfied by set theory. For other mathematical approaches to the notion of "size" of sets based on Aristotles Principle see [3, 6].

[4]According to [15–18], (1) and (2) represent the same set of numbers. In (1) we can "see" only the finite numbers, but in (2) we can "see" both finite and infinite numbers.

5   Grossone calculus

We introduce now a calculus with ① following the axiomatic theory of Grossone developed and studied by Lolli [9]. Our approach differs two fold from Lolli's: a) we do no restrict ourselves to the predicative second order logic, that is, quantifications of sets are allowed, b) while Lolli uses $\mathbb{N}$ for analysing $\mathbb{N}^\dagger$, we proceed the other way round: we study $\mathbb{N}$ with $\mathbb{N}^\dagger$. The notation $n \in \mathbb{N}$ means that n is a finite element in $\mathbb{N}^\dagger$.

The following infinite list of rules **R1** and **R2** defines the arithmetic structure of $\mathbb{N}^\dagger$.

R1. Infinity: For every $n \in \mathbb{N}$, $n < ①$.

R2. Divisibility: For every $n \in \mathbb{N}$, $\frac{①}{n}$ is an infinite number.

As proved in [9], the rules **R1** and **R2** justify all infinite numbers in (2), operations with ① like $n + ①, ① - n, \frac{①-n}{k}$ and identities like $0 \cdot ① = ① \cdot 0 = 0, ① - ① = 0, \frac{①}{①} = 1, 1^① = 1, 0^① = 0$. Is ① prime? The answer is negative: in fact ① is "maximally composite" as by **R2** for every finite $n \in \mathbb{N}$, $\frac{①}{n} \in \mathbb{N}^\dagger$, hence n divides ①.

The ①-measure $\mu \colon \mathcal{P}(\mathbb{N}) \to \mathbb{N}^\dagger \cup \{0\}$ is introduced using the following two axioms[5]:

$A_\mu 1$. For every $x \in \mathbb{N}$, $\mu(\{x\}) = 1$.

$A_\mu 2$. For every sets $X, Y \subseteq \mathbb{N}$, if $X \cap Y = \emptyset$, then $\mu(X \cup Y) = \mu(X) + \mu(Y)$.

In conjunction with **R2** we introduce a partition $(\mathbb{N}_{i,n})_{1 \leq i \leq n}$ of $\mathbb{N}$: for every $n \in \mathbb{N}$ and $1 \leq i \leq n$,

$$\mathbb{N}_{i,n} = \{i, n+i, 2n+i, \dots\}.$$

For every $x \in \mathbb{N}^\dagger$, $\prec_x$ is the initial segment of numbers less than or equal to x, $\prec_x = \{y \in \mathbb{N} \mid y \leq x\}$. If $x \in \mathbb{N}$, then $\prec_x$ is finite, but if x is an infinite number, then the set $\prec_x$ is infinite; for example, $\prec_① = \mathbb{N}$. A set X is called bounded if there exists an x such that $X \subseteq \prec_x$.

The following two theorems have been proved in [9].

Theorem 1. *The following facts can be proved from **R1**, **R2**, $A_\mu 1$ and $A_\mu 2$:*

1. *$\mu(\emptyset) = 0$.*

2. *For every bounded sets $X, Y \subseteq \mathbb{N}$, if $X \subseteq Y$, then $\mu(X) \leq \mu(Y)$.*

3. *For every $x \in \mathbb{N}$, $\mu(\prec_x) = x$; in particular, $\mu(\prec_①) = ①$.*

4. *Every proper subset of $\mathbb{N}$ has a measure x with $x < ①$.*

Theorem 2. *From **R1**, **R2**, $A_\mu 1$ and $A_\mu 2$ for every $n \in \mathbb{N}$ and $1 \leq i \leq n$, we have $\mu(\mathbb{N}_{i,n}) = \frac{①}{n}$.*

From Theorem 2 the ①-measure of each $\mathbb{N}_{i,n}$ is n times less than the ①-measure of the whole set $\mathbb{N}$. This validates to the intuition that the 'number' of even numbers is the same as the 'number' of odd numbers and half of the 'number' of all numbers .

Corollary 3. *From **R1**, **R2**, $A_\mu 1$ and $A_\mu 2$ we have*

---

[5]The name "measure" will be justified by Theorem 1.

1. $\mu(\mathbb{N}) = ①$,

2. $\mu(\mathbb{N}_{1,2}) = \mu(\mathbb{N}_{2,2}) = \frac{①}{2}$,

3. $\mu(\mathbb{N}_{1,3})) = \mu(\mathbb{N}_{2,3}) = \mu(\mathbb{N}_{3,3}) = \frac{①}{3}$,

4. $\mu(\mathbb{N} \cup \{0\}) = ① + 1$,

5. $\mu(\mathbb{N} \setminus \{3, 5, 10, 23, 114\}) = ① - 5$,

6. $\mu(\{x \in \mathbb{N} \mid x = n^2, n \in \mathbb{N}\}) = \lfloor \sqrt{①} \rfloor$.

6 De Finetti infinite fair lottery

By a finite lottery we mean a process that assigns exactly one winner among a finite set of tickets (contained in an urn) in a fair way, that is, each ticket can be won with the same probability. If the tickets are drawn from the label set $\{1, 2, 3, \ldots, n\}$, then the probability that each i, $1 \leq i \leq n$, is a winner is $1/n$. In particular, the probability of a combination of tickets contains the winner is the sum of the individual probabilities of the tickets. This process can be modelled by a uniform probability space which fulfils the Kolmogorov's axioms.

What about an infinite lottery in which tickets are drawn from the label infinite set $\{1, 2, 3, \ldots\}$. Can it be fair? "Classically" this is impossible because informally the uniform distribution will require that each number has probability $1/\infty = 0$; for a formal argument see [4], [19, Ch 11].

Using the Grossone framework it was informally argued in [14] that the infinite fair lottery $\mathbb{N}$ – which has $①$ numbers – is possible and the probability of each number to be a winner is $1/①$, a positive infinitesimal. This intuition will be made rigorous in Theorem 4.

7 Grossone-like uniform probability spaces

In this section we construct and study Grossone-like uniform probability spaces. The approach is inspired by and parallels the construction of classical probability.[6]

Next we introduce the infinitesimals based on $①$. A real number x is *infinitesimal* if $|x| < r$ for every positive real r (see more in [8]).

R3. Infinitesimals: There exists $①^{-1}$ with $①^{-1} \cdot ① = ① \cdot ①^{-1} = 1$.

From **R1** and **R3**, $①^{-1} > 0$ and for every $n \in \mathbb{N}$, $①^{-1} < \frac{1}{n}$, so $①^{-1}$ is an infinitesimal less than 1.

In what follows we will assume **R1**, **R2**, **R3**, $\mathbf{A}_\mu \mathbf{1}$ and $\mathbf{A}_\mu \mathbf{2}$. The probabilities defined in this section will have also infinitesimal values drawn from the set

$$\mathbb{I} = \left\{ \frac{\mu(A)}{\mu(\Omega)} \mid \emptyset \neq A \subset \Omega \subseteq \mathbb{N}, \Omega \text{ infinite} \right\}, \tag{3}$$

(see **R3**). Hence all probabilities will be defined on $\mathcal{P}(\Omega)$ and will take values in the extended unit interval $[0, 1]^\dagger = [0, 1] \cup \mathbb{I}$.

In this framework Williamson's arguments against infinitesimal probability are mathematically refuted.

---

[6]In Section 9.1 [18] an informal discussion of a probability on a set with more than $①$ elements is presented.

7.1 Grossone probability space

Let us consider the sample space $\mathbb{N}$ and the elementary events of the random experiment $\mathcal{E}$ consisting in picking at random a number from $\mathbb{N}$. The *Grossone uniform probability space*

$$\mathbb{N}(①) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathrm{Pr})$$

is defined by assigning to every event $A \in \mathcal{P}(\mathbb{N})$ the probability

$$\mathrm{Pr}(A) = \frac{\mu(A)}{\mu(\mathbb{N})} = \frac{\mu(A)}{①}. \tag{4}$$

Theorem 4. *The probability space $\mathbb{N}(①)$ is regular, total, finitely additive and uniform.*

Proof. In view of Theorem 1 the function $\mathrm{Pr} \colon \mathcal{P}(\mathbb{N}) \to [0,1]^{\dagger}$ defined by (4) satisfies the properties of a finite-additive probability. Indeed, $\mathrm{Pr}(A) \geq 0$ for all $A \in \mathcal{P}(\mathbb{N})$, $\mathrm{Pr}(\mathbb{N}) = 1$ and for all sets $A, B \in \mathcal{P}(\mathbb{N})$ such that $A \cap B = \emptyset$ we have $\mathrm{Pr}(A \cup B) = \mathrm{Pr}(A) + \mathrm{Pr}(B)$ by $\mathbf{A}_\mu\mathbf{2}$.

By Theorem 1 and (4), every event has a probability and for every $A \in \mathcal{P}(\mathbb{N})$, $\mathrm{Pr}(A) = 0$ iff $A = \emptyset$.

Finally, $\mathbb{N}(①)$ is uniform because, by $\mathbf{A}_\mu\mathbf{1}$, every elementary event $\{n\}$ has the same chance to occur:

$$\mathrm{Pr}(\{n\}) = \frac{\mu(\{n\})}{①} = \frac{1}{①}. \tag{5}$$

$\square$

Theorem 4 shows that the Grossone probability space $\mathbb{N}(①)$ satisfies all four probability desiderata in Section 3. In contrast with the standard probability theory, where there is no infinite fair lottery, the probability space $\mathbb{N}(①) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathrm{Pr})$ is an adequate model for the fair lottery on $\mathbb{N}$. In this probability space we can calculate the probability of every event, in particular, the probability of each event in Corollary 3. For example, in $\mathbb{N}(①)$ the probability of picking at random an arbitrary odd number is

$$\mathrm{Pr}(\mathbb{N}_{1,2}) = \frac{\frac{①}{2}}{①} = \frac{1}{2}. \tag{6}$$

7.2 Grossone-like probability spaces

Let $\Omega \subset \mathbb{N}$ be infinite. The probability $\mathrm{Pr}_\Omega : \mathcal{P}(\Omega) \to [0,1]^{\dagger}$ is defined for every $n \in \Omega$ by

$$\mathrm{Pr}_\Omega(\{n\}) = \frac{1}{\mu(\Omega)}.$$

As a consequence of Theorem 4 we get

Corollary 5. *The probability space $\Omega(①) = (\Omega, \mathcal{P}(\Omega), \mathrm{Pr}_\Omega)$ is regular, total, finitely additive and uniform.*

The probability space $\Omega(①)$ will be called a *Grossone-like probability space.*[7] It is clear that by (5) and the assumption $\Omega \subset \mathbb{N}$, for every $n \in \Omega$, $\mathrm{Pr}(\{n\}) < \mathrm{Pr}_\Omega(\{n\})$.

Grossone-like probability spaces can be defined not only on infinite sets of numbers, but also on other infinite discrete sets, like the sets of integers or rationals.

---

[7]In case Ω is contextually clear Pr_Ω will be simply written Pr.

7.3   Williamson's arguments

Williamson's argument [21] purports to show that infinitesimal probabilities are inherently problematic; for more philosophical arguments refuting Williamson's conclusion see [2, 13]. In this section we refute mathematically Williamson's conclusion.

Consider an urn containing tickets labelled with the elements of $\mathbb{N}$ and a mechanism to implement a fair lottery for the tickets in the urn. We consider two cases.

In case (1), all tickets are in the urn and the probability of winning each arbitrary single ticket n in this lottery is $\Pr_1(n)$, a probability which can be infinitesimal, but not zero.

In case (2), one ticket k is removed from the urn prior to the drawing of the winning ticket. Accordingly, the urn contains one ticket less, so the probability of winning each remaining ticket $m \in \mathbb{N} \setminus \{k\}$ is (after renormalisation):

$$\Pr_2(m) = \frac{\Pr_1(m)}{1 - \Pr_1(m)}. \tag{7}$$

However, in isolation case (2) looks exactly as before the removal of the ticket k, which is case (1). Why? Because of a simple set-theoretic argument: *both urns have the same set-theoretical cardinality.* Accordingly, for every $m \in \mathbb{N} \setminus \{k\}$:

$$\Pr_2(m) = \Pr_1(m). \tag{8}$$

From (7) and (8) we deduce that for every $m \in \mathbb{N} \setminus \{k\}$, $\Pr_1(m) = \Pr_2(m) = 0$, a contradiction. This is interpreted by Williamson [21] as an inherently problematic situation arising in any attempt to use infinitesimal probabilities, irrespective of their mathematical constructions.

The reason case (1) and case (2) *look exactly the same* is because the urns have the same set-theoretical cardinality, that is, because informally they have the same number of elements. The cardinality of a finite set is indeed the same as the number of elements of the set; however, *this is not true for infinite sets.*

Here is the resolution of Williamson's contradiction using the Grossone probability space. In case (1) the probability space is

$$\mathbb{N}(①) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \Pr_1)$$

in which the winning probability of every single ticket labelled by $n \in \mathbb{N}$ is

$$\Pr_1(\{n\}) = \frac{1}{①}.$$

In case (2) when k has been removed from the urn, the correct probability space is the Grossone-like space

$$(\mathbb{N} \setminus \{k\})(① - 1) = (\mathbb{N} \setminus \{k\}, \mathcal{P}(\mathbb{N} \setminus \{k\}), \Pr_2).$$

In this space the winning probability of every single ticket labelled by $m \in \mathbb{N} \setminus \{k\}$ is obviously larger than the probability in the space $\mathbb{N}(①)$:

$$0 < \Pr_2(\{m\}) = \frac{1}{① - 1} > \frac{1}{①} = \Pr_1(\{m\}).$$

A similar analysis works if instead of removing one ticket from the urn we remove infinitely many tickets, say, all tickets labelled with even numbers. In this example, the first probability space remains unchanged, but the second probability space is the Grossone-like space

$$\mathbb{N}_{2,2}(①) = (\mathbb{N}_{2,2}, \mathcal{P}(\mathbb{N}_{2,2}), \mathrm{Pr}_3).$$

In this space the winning probability of every single ticket labelled by $m \in \mathbb{N}_{2,2}$ is larger than the probability in $\mathbb{N}(①)$:

$$\mathrm{Pr}_3(\{m\}) = \frac{1}{\frac{①}{2}} > \frac{1}{①} = \mathrm{Pr}_1(\{m\}) > 0.$$

7.4   Independence

Consider a Grossone-like probability space $(\Omega, \mathcal{P}(\Omega), \mathrm{Pr}_\Omega)$. Two events A and B are *independent* if $\mathrm{Pr}_\Omega(A \cap B) = \mathrm{Pr}_\Omega(A)\mathrm{Pr}_\Omega(B)$. A finite set of events is mutually independent if every event is independent of any intersection of the other events.

Example 6. *Consider the Grossone-like model for De Finetti infinite fair lottery $\mathbb{N}(①)$ and the following two events: A: the winner is smaller or equal to 10, and B: the winner is an odd number. We have:*

$$\mathrm{Pr}(A) = \mathrm{Pr}(\prec_{10}) = \frac{10}{①}; \ \mathrm{Pr}(B) = \mathrm{Pr}(\mathbb{N}_{1,2}) = \frac{1}{2}, \ \mathrm{Pr}(A \cap B) = \mathrm{Pr}(\{1,3,5,7,9\}) = \frac{5}{①},$$

hence, $\mathrm{Pr}(A \cap B) = \mathrm{Pr}(A)\mathrm{Pr}(B)$ so the events A and B are independent.

On the other hand, the events "the winner is an even number" and "the winner is an odd number" are not independent because $\mathrm{Pr}(\mathbb{N}_{1,2}) = \mathrm{Pr}(\mathbb{N}_{2,2}) = \frac{1}{2} \neq 0 = \mathrm{Pr}(\mathbb{N}_{1,2} \cap \mathbb{N}_{2,2}).$

7.5   Conditional probabilities

Consider a Grossone-like probability space $(\Omega, \mathcal{P}(\Omega), \mathrm{Pr}_\Omega)$. The conditional probability of an event A assuming that the event B with $\mathrm{Pr}_\Omega(B) > 0$ has occurred is defined as

$$\mathrm{Pr}_\Omega(A \mid B) = \frac{\mathrm{Pr}_\Omega(A \cap B)}{\mathrm{Pr}_\Omega(B)}.$$

Note that two events A and B are independent iff $\mathrm{Pr}_\Omega(B) = 0$ or $\mathrm{Pr}_\Omega(A \mid B) = \mathrm{Pr}_\Omega(A).$

In Example 6 we have

$$\mathrm{Pr}(\prec_{10} \mid \mathbb{N}_{1,2}) = \frac{\mathrm{Pr}(\prec_{10} \cap \mathbb{N}_{1,2})}{\mathrm{Pr}(\mathbb{N}_{1,2})} = \frac{\frac{5}{①}}{\frac{1}{2}} = \frac{10}{①} = \mathrm{Pr}(\prec_{10}).$$

For $A \subseteq \Omega$ we denote $\bar{A} = \Omega \setminus A$ and $A \to B = \bar{A} \cup B.$

Example 7 (Negation elimination). *For every $A \in \mathcal{P}(\Omega)$ such that $A \neq \emptyset$ ($\bar{A} \neq \emptyset$) and for every $B \in \mathcal{P}(\Omega)$ we have: $\mathrm{Pr}_\Omega(\bar{A} \to B \mid A) = 1$ ($\mathrm{Pr}_\Omega(A \to B \mid \bar{A}) = 1$).*

Proof. We have: $\mathrm{Pr}_\Omega(\bar{A} \to B \mid A) = \mathrm{Pr}_\Omega(A \cup B \mid A) = \frac{\mathrm{Pr}_\Omega(A \cap (A \cup B))}{\mathrm{Pr}_\Omega(A)} = 1$ and similarly $\mathrm{Pr}_\Omega(A \to B \mid \overline{A}) = 1.$ $\square$

Lemma 8 (Law of Total Probability). *If $(B_i)_{1 \le i \le n}$ is a partition of Ω, then for every event $A \ne \emptyset$ and all $1 \le i \le n$ we have:*

$$\mathrm{Pr}_\Omega(A) = \sum_{j=1}^n \mathrm{Pr}_\Omega(B_j)\,\mathrm{Pr}_\Omega(A \mid B_j) = \sum_{j=1}^n \mathrm{Pr}_\Omega(B_j)\,\mathrm{Pr}_{B_j}(A).$$

Proof. We have

$$A = \bigcup_{j=1}^n (A \cap B_j) \text{ and } (A \cap B_j) \cap (A \cap B_i) = \emptyset, \text{ for every } j \ne i.$$

Then, by $\mathbf{A}_\mu\mathbf{2}$ and the definition of conditional probability, we have

$$\mathrm{Pr}_\Omega(A) = \sum_{j=1}^n \mathrm{Pr}_\Omega(A \cap B_j) = \sum_{j=1}^n \mathrm{Pr}_\Omega(B_j)\,\mathrm{Pr}_\Omega(A \mid B_j).$$

$\square$

Theorem 9 (Bayes Theorem). *If $(B_i)_{1 \le i \le n}$ is a partition of Ω, then for every event $A \ne \emptyset$ and all $1 \le i \le n$ we have:*

$$\mathrm{Pr}_\Omega(B_i \mid A) = \frac{\mathrm{Pr}_\Omega(A \mid B_i)\mathrm{Pr}_\Omega(B_i)}{\sum_{j=1}^n \mathrm{Pr}_\Omega(A \mid B_j)\mathrm{Pr}_\Omega(B_j)}. \tag{9}$$

Proof. The property follows from Lemma 8 and the equalities

$$\mathrm{Pr}_\Omega(B_i \mid A) = \frac{\mathrm{Pr}_\Omega(B_i \cap A)}{\mathrm{Pr}_\Omega(A)} = \frac{\mathrm{Pr}_\Omega(B_i)\,\mathrm{Pr}_\Omega(A \mid B_i)}{\mathrm{Pr}_\Omega(A)}.$$

$\square$

Bayes Theorem plays a very important role when we gain an extra knowledge achieved through an auxiliary experiment. Assume we are interested in the outcomes $B_1, \ldots, B_n$ of a random experiment which form a partition of the sample space Ω. Their *prior probabilities* are $\{\mathrm{Pr}_\Omega(B_i), 1 \le i \le n\}$.

We then perform an auxiliary random experiment and denote by A its outcome. Using Bayes formula (9), we can calculate the *posterior probabilities* of $B_1, \ldots, B_n$, that is $\{\mathrm{Pr}_\Omega(B_i \mid A), 1 \le i \le n\}$.

Example 10. *Consider an infinite, fair lottery with tickets, drawn from $\mathbb{N}$, written in red and blue in such a way that both colours are used. We can formulate the following hypotheses about the lottery:*

$$H_i: \text{``The lottery contains } i \text{ red tickets and } (\text{①} - i) \text{ blue tickets''}, 1 \le i \le \text{①} - 1,$$

and, accordingly, model the lottery with the sets

$$\mathbb{N}_{(i,red)} = \{1,2,3,\cdots \mid i \text{ tickets are red}, \text{①} - i \text{ tickets are blue}\}, i \ge 1, i \le \text{①} - 1.$$

It is seen that

$$\mathbb{N}_{(i,red)} = \mathbb{RED}_i \cup \mathbb{BLUE}_i, \mu(\mathbb{RED}_i) = i, \mu(\mathbb{BLUE}_i) = \text{①} - i, 1 \le i \le \text{①} - 1.$$

A drawing of the lottery will produce a (winner) ticket in one of the sets $\mathbb{N}_{(i,red)}$, a process which can be formally described by the following Grossone-like space $(\Omega, \mathcal{P}(\Omega), \mathrm{Pr}_\Omega)$ in which the sample space is

$$\Omega = \bigcup_{i=1}^{①-1} \mathbb{N}_{(i,red)} = \left(\bigcup_{i=1}^{①-1} \mathbb{RED}_i \right) \bigcup \left(\bigcup_{i=1}^{①-1} \mathbb{BLUE}_i \right),$$

and the probability is $\Pr_\Omega(\{x\}) = \frac{1}{①(①-1)}$, for all $x \in \Omega$.

The event that the hypothesis H_i is true is $\mathbb{N}_{(i,red)}$; for simplicity we write $B_i = N_{(i,red)}$.

A priori, the likelihood of the hypothesis H_i is given by the prior probability

$$\Pr_\Omega(B_i) = \frac{\mu(B_i)}{①(①-1)} = \frac{①}{①(①-1)} = \frac{1}{①-1}.$$

The auxiliary experiment now consists in drawing one ticket which should give more information about the events B_i. Assume the winner's ticket is red, which means that the event $\left(\bigcup_{i=1}^{①-1} \mathbb{RED}_i \right) \in \mathcal{P}(\Omega)$ has occurred. Using the simpler notation $A = \bigcup_{i=1}^{①-1} \mathbb{RED}_i$ we have:

$$\Pr_\Omega(A) = \frac{\mu(A)}{①(①-1)} = \frac{1}{①(①-1)} \sum_{i=1}^{①-1} \mu(\mathbb{RED}_i) = \frac{1}{①(①-1)} \sum_{i=1}^{①-1} i = \frac{1}{①(①-1)} \cdot \frac{(①-1)①}{2} = \frac{1}{2}.$$

On the other hand, if the hypothesis H_i is true we have

$$\Pr_\Omega(A \mid B_i) = \frac{\Pr_\Omega(A \cap B_i)}{\Pr_\Omega(B_i)} = \frac{\frac{\mu(A \cap B_i)}{①(①-1)}}{\Pr_\Omega(B_i)} = \frac{\frac{\mu(\mathbb{RED}_i)}{①(①-1)}}{\Pr_\Omega(B_i)} = \frac{i \cdot (①-1)}{①(①-1)} = \frac{i}{①}. \tag{10}$$

As the events $\{B_i \mid 1 \le i \le ①-1\}$ form a partition of Ω, we can now use Bayes formula to calculate, for every $1 \le i \le ①-1$, the posterior probabilities:

$$\Pr_\Omega(B_i \mid A) = \frac{\Pr_\Omega(A \mid B_i) \cdot \Pr_\Omega(B_i)}{\Pr_\Omega(A)} = \frac{\frac{i}{①} \cdot \frac{1}{①-1}}{\frac{1}{2}} = \frac{2i}{①(①-1)}. \tag{11}$$

First note that $\max_{1 \le i \le ①-1} \Pr_\Omega(B_i \mid A) = \Pr_\Omega\left(B_{①-1} \mid A\right) = \frac{2}{①}$. Second, for some $1 \le i, j \le ①-1$, $\Pr_\Omega(B_i) < \Pr_\Omega(B_i \mid A)$ and $\Pr_\Omega(B_j) > \Pr_\Omega(B_j \mid A)$. To locate a "switch" threshold we use (10) and (11) and get the inequalities

$$\frac{2i}{①(①-1)} < \frac{1}{①-1}, \quad \frac{2(i+1)}{①(①-1)} \ge \frac{1}{①-1},$$

which have the solutions $\frac{①}{2} - 1 \le i < \frac{①}{2}$. Then, a "switch" threshold is the Grossone number $i = \frac{①}{2} - 1$.

7.6 Random variables

Using (3), let $\mathbb{R}^\dagger = \mathbb{R} \cup \mathbb{I}$. A real random variable on the Grossone-like probability space $(\Omega, \mathcal{P}(\Omega), \Pr)$ is a function $X: \Omega \to \mathbb{R}^\dagger$ with the property that for every measurable set $B \in \mathcal{B}(\mathbb{R}^\dagger)$ we have $X^{-1}(B) \in \mathcal{P}(\Omega)$. A random

variable X generates a probability space $\left(\mathbb{R}^*, \mathcal{B}\left(\mathbb{R}^\dagger\right), P_X\right)$, in which the probability P_X – called the *probability distribution of X* – is defined as follows:

$$P_X(B) = \Pr\left(X^{-1}(B)\right), \ B \in \mathcal{B}\left(\mathbb{R}^\dagger\right).$$

The distribution of the discrete random variable X can be described by $P_X(x), x \in X(\Omega)$.   For example, if we consider the Grossone probability space $\mathbb{N}(①) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \Pr)$ and the random variable $X \colon \mathbb{N} \to \mathbb{N}, X(n) = n$, then the probability distribution of X is

$$P_X(\{n\}) = \Pr(\{X = n\}) = \frac{1}{①},$$

that is, the *Grossone-like uniform distribution.*

Using Grossone calculus one can calculate the moments and variance of the Grossone-like uniform distribution. For example,

$$E(X) = \sum_{n \in \mathbb{N}} n \cdot \frac{1}{①} = \frac{1}{①} \cdot \frac{①(① + 1)}{2} = \frac{① + 1}{2},$$

$$E\left(X^2\right) = \sum_{n \in \mathbb{N}} n^2 \cdot \frac{1}{①} = \frac{1}{①} \cdot \frac{①(① + 1)(2① + 1)}{6} = \frac{(① + 1)(2① + 1)}{6},$$

$$Var(X) = E\left(X^2\right) - E^2(X) = \frac{(① + 1)(2① + 1)}{6} - \frac{(① + 1)^2}{4} = \frac{(①)^2 - 1}{12}.$$

8   A non-uniform Grossone-like probability

In this section we discuss a non-uniform Grossone-like probability.

The *binomial experiment* with parameters $n \in \mathbb{N}$ and rational $p \in (0, 1)$ is a sequence of n independent experiments, each asking a yes–no question with the outcome being: "success" (with probability p) or "failure" (with probability $q = 1 - p$). If we consider $n = ①$, then the discrete probability distribution of the number of successes in this experiment can be modelled by the Grossone-like probability space $(\mathbb{N} \cup \{0\}, \mathcal{P}(\mathbb{N} \cup \{0\}), \mathrm{P}_{\mathrm{Binomial},p})$ where

$$\mathrm{P}_{\mathrm{Binomial},p}(\{k\}) = \begin{pmatrix} ① \\ k \end{pmatrix} \cdot p^k (1 - p)^{① - k}, \ k \in \mathbb{N} \cup \{0\}.$$

Using Grossone calculus we have

$$\sum_{k=0}^{①} \begin{pmatrix} ① \\ k \end{pmatrix} \cdot p^k (1 - p)^{① - k} = (p + (1 - p))^{①} = 1^{①} = 1.$$

9 Conclusion

In this paper we have constructed and studied a class of infinitesimal probabilities on infinite sets of positive integers based on the notion of Grossone size. More precisely, we have used the "magnified" set $\mathbb{N}^\dagger$ to measure the sizes of different, finite or infinite, subsets of $\mathbb{N}$.

These probabilities have natural properties which are not satisfied in the standard Kolmogorov probability theory on infinite sample sets, like regularity, totality, uniform distributivity and weak Laplacean-ity. In this framework De Finetti fair lottery has the natural solution and Williamson's objections against infinitesimal probabilities are mathematically refuted.

An advantage of the proposed construction over other infinitesimal probabilities studied in the literature [2] comes from its naturalness and simplicity: the construction is operationally similar to classical (finite) probability.

It will be interesting to test the utility of these probabilities in areas in which their properties not shared by the standard probability theory matter: a prime example will be (quantum) physics (see for example [5, 7, 10]).

Acknowledgement

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] V. Benci, L. Horsten, and S. Wenmackers. Non-arhimedian probability. *Milan Journal of Mathematics*, 81:121–151, 2013.

[2] V. Benci, L. Horsten, and S. Wenmackers. Infinitesimal probabilities. *Brit. J. Phil. Sci.*, 69:509–552, 2018.

[3] V. Benci, M. D. Nasso, and M. Forti. An aristotelian notion of size. *Annals of Pure and Applied Logic*, 143(1):43 – 53, 2006.

[4] B. De Finetti. *Theory of Probability*. John Wiley & Sons, London-New York-Sydney-Toronto, 1975.

[5] D. Ellerman. On classical finite probability theory as a quantum probability calculus. http://arxiv.org/abs/1502.01048, February 2015.

[6] T. Gilbert and N. Rouche. Y a-t-il vraiment autant de nombres pairs que de naturels? *Cahiers du Centre de Logique*, 9:99–139, 1996.

[7] A. Khrennikov. Quantum bayesianism as the basis of general theory of decision-making. *Philos Trans A Math Phys Eng Sci.*, 374(2068):20150245, 2016.

[8] G. Lolli. Infinitesimals and infinites in the history of mathematics: A brief survey. *Applied Mathematics and Computation*, 218(16):7979–7988, 2012.

[9] G. Lolli. Metamathematical investigations on the theory of grossone. *Applied Mathematics and Computation*, 255:3–14, 2015.

[10] N. D. Mermin. Physics: Qbism puts the scientist back into science. *Nature*, 507(421–423), 2014.

[11] E. Nelson. *Radically Elementarry Probability Theory*. Princeton University Press, Princeton, NJ, 1987.

[12] P. Olofsson. *Probability, Statistics, and Stochastic Processes*. Wiley Interscience, New York, 2005.

[13] M. W. Parker. Symmetry arguments against regular probability: A reply to recent objections. *European Journal for Philosophy of Science*, 9(8):1–21, 2018.

[14] D. Rizza. Numerical methods for infinite decision-making processes. *International Journal of Unconventional Computing*, 14(2):139–158, 2019.

[15] Y. D. Sergeyev. *Arithmetic of Infinity*. Edizioni Orizzonti Meridionali, 2003.

[16] Y. D. Sergeyev. A new applied approach for executing computations with infinite and infinitesimal quantities. *Informatica*, 19(4):567–596, 2008.

[17] Y. D. Sergeyev. Computations with grossone-based infinities. In C. S. Calude and M. J. Dinneen, editors, *Unconventional Computation and Natural Computation - 14th International Conference, UCNC 2015, Auckland, New Zealand, August 30 - September 3, 2015, Proceedings*, volume 9252 of *Lecture Notes in Computer Science*, pages 89–106. Springer, 2015.

[18] Y. D. Sergeyev. Numerical infinities and infinitesimals: Methodology, applications, and repercussions on two Hilbert problems. *EMS Surveys in Mathematical Sciences*, 4(2):219–320, 2017.

[19] J. Truss. *Foundations of Mathematical Analysis*. Oxford University Press, Oxford, 1997.

[20] S. Wenmackers and L. Horsten. Fair infinite lotteries. *Synthese*, 190:37–61, 2013.

[21] T. Williamson. How probable is an infinite sequence of heads? *Analysis*, 67:173–180, 2007.