Milan J. Math. Vol. 81 (2013) 121–151
DOI 10.1007/s00032-012-0191-x
Published online October 11, 2012

Milan Journal of Mathematics

Non-Archimedean Probability

Vieri Benci, Leon Horsten and Sylvia Wenmackers

Abstract. We propose an alternative approach to probability theory closely related to the framework of *numerosity* theory: non-Archimedean probability (NAP). In our approach, unlike in classical probability theory, all subsets of an infinite sample space are measurable and only the empty set gets assigned probability zero (in other words: the probability functions are regular). We use a non-Archimedean field as the range of the probability function. As a result, the property of countable additivity in Kolmogorov's axiomatization of probability is replaced by a different type of infinite additivity.

Mathematics Subject Classification (2010). 60A05, 03H05.

Keywords. Probability, axioms of Kolmogorov, nonstandard models, fair lottery, non-Archimedean fields.

1. Introduction

Kolmogorov's classical axiomatization embeds probability theory into measure theory: it takes the domain and the range of a probability function to be standard sets and employs the classical concepts of limit and continuity. Kolmogorov starts from an elementary theory of probability "in which we have to deal with only a finite number of events" [16] (p. 1). We will stay close to his axioms for the case of finite sample spaces, but critically investigate his approach in the second chapter of [16] dealing with the case of "an infinite number of random events". There, Kolmogorov introduces an additional axiom, the Axiom of Continuity. Together with the axioms and theorems for the finite case (in particular, the addition theorem, now called 'finite additivity', FA), this leads to the generalized addition theorem, called 'σ-additivity' or 'countable additivity' (CA) in the case where the event space is a Borel field (or σ-algebra, in modern terminology).

Within Kolmogorov's probability theory, it may happen that a non-empty subset of an infinite sample space gets assigned measure zero; this has been called a failure of 'regularity' (see *e.g.* [17, 12]). In particular, for fair lotteries on uncountable domains, such as $\mathbb{R}$ or $2^{\mathbb{N}}$, all countable subsets get assigned probability zero. The observation that an event of probability zero may nevertheless happen, is a well-known hurdle in the interpretation of standard probability theory, which also

Birkhäuser

TABLE 1. Various quantitative probability theories.

Domain: / Range:	Standard	Ideal
$\mathbb{R}$ (Archimedean field)	Kolmogorov	(b) Loeb
Non-Archimedean field	(c) NAP	(a) Nelson

complicates the didactics of the topic [12]. Moreover, some problems—such as a fair lottery on $\mathbb{N}$ or $\mathbb{Q}$—cannot be modeled within Kolmogorov's framework at all. Weakening additivity to finite additivity, as de Finetti advocated [10], solves the latter issue but introduces strange consequences of its own [13]. It has been suggested that using infinitesimals, as available in the framework of nonstandard analysis (NSA), would allow describing regular probability functions on infinite domains [18, 21]. We investigate this claim further.

Multiple alternative approaches to probability theory have been proposed in the literature. We focus here on proposals that involve changing the domain or the range of the probability function to a nonstandard set in the sense of NSA. We can thus distinguish three categories of alternatives, presented in Table 1: the probability function has (a) both a nonstandard domain and a nonstandard range, (b) only a nonstandard domain, or (c) only a nonstandard range.

Alternative (a) is easily obtained in the context of NSA by applying the Transfer Principle (see *e.g.* [6]) to standard Kolmogorovian probability functions on finite domains. An example of this approach was developed by Nelson: in [20], he presented his "Radically elementary probability theory" based on non-Archimedean, hyperfinite sets as the domain and range of the probability function. His framework has the benefit of making probability theory on infinite sample spaces equally simple and straightforward as the corresponding theory on finite sample spaces; the appropriate additivity property is hyperfinite additivity. Nelson's theory is regular. However, this elegant theory does not apply *directly* to our current quest for finding regular probability functions on standard infinite sets, such as a fair lottery on $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, or $2^{\mathbb{N}}$.

Alternative (b) is the dominant line of research in nonstandard measure and integration theory; it is concerned mainly with finding new results in *standard* measure and integration theory [9]. A measure with standard range and nonstandard domain can be obtained in NSA by starting from (a) and applying the standard part function afterwards, which maps a hyperreal measure to the unique nearest real value (see *e.g.* [6]). Probability measures of this type are known as Loeb measures [19]. Although the well-developed theory of Loeb measures has proven fruitful in many applications, it too simply does not address the problem that concerns us here.

To describe regular probability functions on standard infinite domains, we need to investigate the previously unexplored alternative (c). We develop this new approach to probability theory, which we call non-Archimedean probability (NAP) theory, in the course of this paper. We formulate new desiderata (axioms) for a concept of probability that is able to describe the case of a fair lottery on $\mathbb{N}$, as well as other cases where infinite sample spaces are involved. As such, the current article generalizes the solution to the infinite lottery puzzle presented in [23]. Within NAP-theory, the domain of the probability function can be the full powerset of any standard set from applied mathematics, whereas the general range is a non-Archimedean field. We investigate the consistency of the proposed axioms by giving a model for them. We show that our theory can be understood in terms of a novel formalization of the limit and continuity concept (called 'Λ-limit' and 'non-Archimedean continuity', respectively). Kolmogorov's CA is replaced by a different type of infinite additivity.

Observe that we do not start from NSA as our background theory. And yet, it does turn out that our theory depends on the same kind of mathematical structures that power NSA (such as the existence of free ultrafilters). We also do not presuppose in the axioms that probability is a special case of measure. However, our theory does suggest an interesting counting measure, which also applies to infinite sets: for fair lotteries, the probability assigned to an event by NAP-theory is directly proportional to the 'numerosity' of the subset representing that event [4]. We mentioned the failure of regularity as a problem in the didactics of probability theory. Although much of the material presented here is too advanced to be taught at the level of high school students or college freshmen, we do think that an introduction to probability by means of the axioms of our NAP-theory *is* a realistic option. After all, standard probability theory is also taught as a rule-based system, without proving that the rules are consistent by constructing a model for them. There is a philosophical companion article [7] to this article, in which we dissolve philosophical objections against using infinitesimals to model probability on infinite sample spaces.

We regard all mentioned frameworks for probability theory—the four quadrants of Table 1—as mathematically correct theories (*i.e.* internally consistent), with a different scope of applicability. For example, if one wants to describe a fair lottery on the sample space $\mathbb{N}$, NAP-theory is the only option. It also allows one to subsequently conditionalize on any subset of $\mathbb{N}$. This example is very simple, but it shows that NAP can treat problems which do not make sense in standard probability theory, and which are not addressed directly by the existing nonstandard approaches. Exploring the connections between the various theories helps to understand each of them better. We agree with Nelson [20] that the infinitesimal probability values should not be considered as an intermediate step—a method to arrive at the answer—but rather as the final answer to probabilistic problems on infinite sample spaces. Approached as such, NAP-theory is a versatile tool with epistemological advantages over the orthodox framework.

1.1. Some notation

Here we fix the notation used in this paper. We use the phrase ‘fair lottery’ to refer to any uniform probability measure. Furthermore, we set:

- $\mathbb{N} = \{1, 2, 3, \ldots\}$ is the set of positive natural numbers;
- $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$ is the set of natural numbers;
- if A is a set, then $|A|$ will denote its cardinality;
- if A is a set, $\mathcal{P}(A)$ is the set of the parts of A, $\mathcal{P}_{fin}(A)$ is the set of *finite* parts of A, and $\mathcal{P}^0_{fin}(A)$ is defined as $\mathcal{P}_{fin}(A) \setminus \varnothing$;
- for any set A, χ_A will denote the indicator function (or characteristic function) of A, namely

$$\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ \\ 0 & \text{if } \omega \notin A; \end{cases}$$

- if F is an ordered field and $a, b \in F$, then we set

$$[a, b]_F = \{x \in F \mid a \leq x \leq b\}$$
$$[a, b)_F = \{x \in F \mid a \leq x < b\}\,;$$

- if F is an ordered field and $F \supseteq \mathbb{R}$, then F is called a superreal field;
- for any set $\mathcal{D}$, $\mathfrak{F}(\mathcal{D}, \mathbb{R})$ will denote the (real) algebra of functions $u : \mathcal{D} \to \mathbb{R}$ equipped with the following operations: for any $u, v \in \mathfrak{F}(\mathcal{D}, \mathbb{R})$, for any $r \in \mathbb{R}$, and for any $x \in \mathcal{D}$:

$$(u + v)(x) = u(x) + v(x),$$
$$(ru)(x) = ru(x),$$
$$(u \cdot v)(x) = u(x) \cdot v(x);$$

- if F is a non-Archimedean field, then we set

$$x \sim y \Leftrightarrow x - y \text{ is infinitesimal } \Leftrightarrow \forall n \in \mathbb{N},\ |x - y| < \frac{1}{n};$$

  in this case we say that x and y are infinitely close;
- if F is a non-Archimedean superreal field and $\xi \in F$ is bounded, then $st(\xi)$ denotes the unique real number x infinitely close to ξ.

2. Kolmogorov’s probability theory

2.1. Kolmogorov’s axioms

Classical probability theory is based on Kolmogorov’s axioms (KA) [16]. We give an equivalent formulation of KA, using P_{KA} to indicate a probability function that obeys these axioms. The sample space Ω is a set whose elements represent elementary events.

Axioms of Kolmogorov

- (K0) **Domain and range.** The events are the elements of $\mathfrak{A}$, a σ-algebra over Ω,[1] and the probability is a function
$$P_{KA} : \mathfrak{A} \to \mathbb{R}.$$
- (K1) **Positivity.** $\forall A \in \mathfrak{A}$,
$$P_{KA}(A) \geq 0.$$
- (K2) **Normalization.**
$$P_{KA}(\Omega) = 1.$$
- (K3) **Additivity.** $\forall A, B \in \mathfrak{A}$ such that $A \cap B = \varnothing$,
$$P_{KA}(A \cup B) = P_{KA}(A) + P_{KA}(B).$$
- (K4) **Continuity.** *Let*
$$A = \bigcup_{n \in \mathbb{N}} A_n$$
where $\forall n \in \mathbb{N}, A_n \subseteq A_{n+1} \in \mathfrak{A}$*; then*
$$P_{KA}(A) = \sup_{n \in \mathbb{N}} P_{KA}(A_n).$$

We will refer to the triple $(\Omega, \mathfrak{A}, P_{KA})$ as a *Kolmogorov Probability space.*

Remark 1. If the sample space is finite then it is sufficient to define a normalized probability function on the elementary events, namely a function
$$p : \Omega \to \mathbb{R}$$
with
$$\sum_{\omega \in \Omega} p(\omega) = 1.$$
In this case the probability function
$$P_{KA} : \mathcal{P}(\Omega) \to [0, 1]_{\mathbb{R}}$$
is defined by
$$P_{KA}(A) = \sum_{\omega \in A} p(\omega) \tag{1}$$
and KA are trivially satisfied. Unfortunately, eq. (1) cannot be generalized to the infinite case. In fact, if the sample space is $\mathbb{R}$, an infinite sum might yield a result in $[0, 1]_{\mathbb{R}}$ only if $p(\omega) \neq 0$ for at most a denumerable number of $\omega \in A$.[2] In a sense, the Continuity Axiom (K4) replaces eq. (1) for infinite sample spaces.

[1] $\mathfrak{A}$ is a σ-algebra over Ω if and only if $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ such that $\mathfrak{A}$ is closed under complementation, intersection, and countable unions. $\mathfrak{A}$ is called the 'event algebra' or 'event space'.

[2] Similarly, in classical analysis the sum of an uncountable sequence is undefined.

2.2. Problems with Kolmogorov's axioms

Kolmogorov uses $[0,1]_{\mathbb{R}}$ as the range of P_{KA}, which is a subset of an ordered field and thus provides a good structure for adding and multiplying probability values, as well as for comparing them. However, this choice for the range of P_{KA} in combination with the property of Countable Additivity (which is a consequence of Continuity) may lead to problems in cases with infinite sample spaces.

Non-measurable sets in $\mathcal{P}(\Omega)$. A peculiarity of KA is that, in general $\mathfrak{A} \neq \mathcal{P}(\Omega)$. In fact, it is well-known that there are (probability) measures (such as the Lebesgue measure on $[0,1]$) which cannot be defined for all the sets in $\mathcal{P}(\Omega)$. Thus, there are sets in $\mathcal{P}(\Omega)$ which are not events (namely elements of $\mathfrak{A}$) even when they are the union of elementary events in $\mathfrak{A}$.[3]

Interpretation of $P_{KA} = 0$ and $P_{KA} = 1$. In Kolmogorov's approach to probability theory, there are situations such that:

$$P_{KA}(A_j) = 0, \quad \text{with } j \in J \tag{2}$$

and

$$P_{KA}\left(\bigcup_{j \in J} A_j\right) = 1. \tag{3}$$

This situation is very common when J is not denumerable. It looks as if eq. (2) states that each event A_j is impossible, whereas eq. (3) states that one of them will occur certainly. This situation requires further epistemological reflection. Kolmogorov's probability theory works fine as a mathematical theory, but the *direct* interpretation of its language leads to counterintuitive results such as the one just described. An obvious solution is to interpret probability 0 as 'very unlikely' (rather than simply as 'impossible'), and to interpret probability 1 as 'almost surely' (instead of 'absolutely certain'). Yet, there is a philosophical price to be paid to avoid these contradictions: the correspondence between mathematical formulas and reality is now quite vague—just how probable is 'very likely' or 'almost surely'?—and far from intuition.

Fair lottery on $\mathbb{N}$. We may observe that the choice $[0,1]_{\mathbb{R}}$ as the range of the probability function is neither necessary to describe a fair lottery in the finite case, nor sufficient to describe one in the infinite case.

- For a fair finite lottery, the unit interval of $\mathbb{R}$ is not necessary as the range of the probability function: the unit interval of $\mathbb{Q}$ (or, maybe some other denumerable subfield of $\mathbb{R}$) suffices.
- In the case of a fair lottery on $\mathbb{N}$, $[0,1]_{\mathbb{R}}$ is not sufficient as the range: it violates our intuition that the probability of any set of tickets can be obtained by adding the probabilities of all individual tickets.

Let us now focus on the fair lottery on $\mathbb{N}$. In this case, the sample space is $\Omega = \mathbb{N}$ and we expect the domain of the probability function to contain all the

[3] For example, if $\Omega = [0,1]_{\mathbb{R}}$ and P_{KA} is given by the Lebesgue measure, then all the singletons $\{x\}$ are measurable, but there are non-measurable sets; namely the union of events might not be an event.

singletons of $\mathbb{N}$, otherwise there would be 'tickets' (individual, natural numbers) whose probability is undefined. Yet, we expect them to be defined and to be equal in a fair lottery. Indeed, we expect to be able to assign a probability to any possible combination of tickets. This assumption implies that the event algebra should be $\mathcal{P}(\Omega)$. Moreover, we expect to be able to calculate the probability of an arbitrary event by a process of summing over the individual tickets.

This leads us to the following conclusions. First, if we want to have a probability theory which describes a fair lottery on $\mathbb{N}$, assigns a probability to all singletons of $\mathbb{N}$, and follows a generalized additivity rule as well as the Normalization Axiom, the range of the probability function has to be a subset of a non-Archimedean field. In other words, the range has to include infinitesimals. Second, our intuitions regarding infinite concepts are fed by our experience with their finite counterparts. So, if we need to extrapolate the intuitions concerning finite lotteries to infinite ones, we need to introduce a sort of limit-operation which transforms 'extrapolations' into 'limits'. Clearly, this operation cannot be the limit of classical analysis. Since classical limits are implicit in Kolmogorov's Continuity Axiom, this axiom must be revised in our approach.

Motivated by the case study of a fair infinite lottery, at this point we know which elements in Kolmogorov's classical axiomatization we do not accept: the use of $[0,1]_{\mathbb{R}}$ as the range of the probability function and the application of classical limits in the Continuity Axiom. However, we have not specified an alternative to his approach: this is what we present in the next sections.

3. Non-Archimedean Probability

We begin this section by stating the axioms of our new theory of probability. Initially, the meaning of the final axiom that replaces classical continuity by what we call "non-Archimedean continuity" may not be transparent to the reader. Therefore, we discuss the purpose of this axiom further in subsection 3.2. Moreover, the construction presented in section 4.2 may clarify this matter further.

3.1. The axioms of Non-Archimedean Probability

We will denote by $\mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$ the algebra of real functions defined on $\mathcal{P}^0_{fin}(\Omega)$.

Axioms of Non-Archimedean Probability

- (NAP0) **Domain and range.** *The events are* **all** *the elements of* $\mathcal{P}(\Omega)$ *and the probability is a function*
$$P : \mathcal{P}(\Omega) \to \mathcal{R}$$
where $\mathcal{R}$ *is a superreal field.*
- (NAP1) **Positivity.** $\forall A \in \mathcal{P}(\Omega)$,
$$P(A) \geq 0.$$
- (NAP2) **Normalization.** $\forall A \in \mathcal{P}(\Omega)$,
$$P(A) = 1 \Leftrightarrow A = \Omega.$$

- (NAP3) **Additivity.** $\forall A, B \in \mathcal{P}(\Omega)$ such that $A \cap B = \varnothing$,

$$P(A \cup B) = P(A) + P(B).$$

- (NAP4) **Non-Archimedean Continuity.** $\forall A, B \in \mathcal{P}(\Omega)$, with $B \neq \varnothing$, *let* $P(A \mid B)$ *denote the conditional probability, namely*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{4}$$

Then
  - $\forall \lambda \in \mathcal{P}^0_{fin}(\Omega)$, $P(A \mid \lambda) \in \mathbb{R}^+$;
  - *there exists an algebra homomorphism*

$$J : \mathfrak{F}\left(\mathcal{P}^0_{fin}(\Omega), \mathbb{R}\right) \to \mathcal{R}$$

such that $\forall A \in \mathcal{P}(\Omega)$

$$P(A) = J\left(\varphi_A\right)$$

where

$$\varphi_A(\lambda) = P(A \mid \lambda) \quad \textit{for any} \;\; \lambda \in \mathcal{P}^0_{fin}(\Omega).$$ [4]

The triple (Ω, P, J) will be called *NAP-space* (or NAP-theory).

Now we will analyze the first three axioms and the fourth will be analyzed in the next section.

The differences between (K0),...,(K3) and (NAP0),...,(NAP3) derive from (NAP2). As consequence of this, we have that:

Proposition 2. *If* (NAP0),...,(NAP3) *holds, then*

(i) $\forall A \in \mathcal{P}(\Omega)$, $P(A) \in [0,1]_{\mathcal{R}}$.
(ii) $\forall A \in \mathcal{P}(\Omega)$, $P(A) = 0 \Leftrightarrow A = \varnothing$.
(iii) *Moreover, there are sufficient conditions for* $\mathcal{R}$ *to be non-Archimedean, such as:*
  (a) Ω *is countably infinite and the theory is fair, namely* $\forall \omega, \tau \in \Omega$, $P(\{\omega\}) = P(\{\tau\})$*;*
  (b) Ω *is uncountable.*

Proof. Take $A \in \mathcal{P}(\Omega)$ and let $B = \Omega \setminus A$; then, by (NAP2) and (NAP3),

$$P(A) + P(B) = 1;$$

then, since $P(B) \geq 0$, $P(A) \leq 1$ and then (i) holds. Moreover, $P(A) = 0 \Leftrightarrow P(B) = 1$ and hence, by (NAP2), $P(A) = 0 \Leftrightarrow B = \Omega$ and so $P(A) = 0 \Leftrightarrow A = \varnothing$. Now let us prove (iii)(a) and assume that $\forall \omega \in \Omega$, $P(\{\omega\}) = \varepsilon > 0$. Now we argue indirectly. If the field $\mathcal{R}$ is Archimedean, then there exists $n \in \mathbb{N}$ such that $n\varepsilon > 1$; now let A be a subset of Ω containing n elements, then by (NAP3) it follows that

[4] In the remainder of this text, each occurence of λ is to be understood as referring to any $\lambda \in \mathcal{P}^0_{fin}(\Omega)$; $f(\lambda)$ will be used instead of $f(\cdot)$, where f is a function on $\mathcal{P}^0_{fin}(\Omega)$.

$P(A) = nP(\{\omega\}) = n\varepsilon > 1$ and this fact contradicts (NAP2); then $\mathcal{R}$ has to be non-Archimedean. Now let us prove (iii)(b) and for every $n \in \mathbb{N}$ set

$$A_n = \{\omega \in \Omega \mid 1/(n+1) < P(\{\omega\}) \leq 1/n\}.$$

By (NAP3) and (NAP2), it follows that each A_n is finite; actually, it contains at most $n+1$ elements.

Now, again, we argue indirectly and assume that the field $\mathcal{R}$ is Archimedean; in this case there are no infinitesimals and hence

$$\Omega = \bigcup_{n\in\mathbb{N}} A_n$$

and this contradicts the fact that we have assumed Ω to be uncountable. □

Remark 3. In the axioms (NAP0),...,(NAP3), the field $\mathcal{R}$ is not specified. This is not surprising since also in Kolmogorov's theory the same may happen. For example, consider $\Omega = \{a, b\}$, with $P_{KA}(\{a\}) = 1/\sqrt{2}$ and hence $P_{KA}(\{b\}) = 1 - 1/\sqrt{2}$. In this case the natural field is $\mathbb{Q}(\sqrt{2})$. However in Kolmogorov's theory, since there is no need to introduce infinitesimal probabilities, all these fields are contained in $\mathbb{R}$ and hence it is simpler to fix $[0,1]_{\mathbb{R}}$ as the range. We suggest an analogous approach with NAP; this will be done in section 4.5.

3.2. Analysis of the fourth axiom

If A is a bounded subset of a non-Archimedean field then the supremum might not exist; consider for example the set of all infinitesimal numbers. Hence the axiom (K4) cannot hold in a non-Archimedean probability theory. In this section, we will show an equivalent formulation of (K4) which can be compared with (NAP4) and helps to understand the meaning of the latter.

Conditional Probability Principle (CPP). *Let Ω_n be a family of events such that $\Omega_n \subseteq \Omega_{n+1}$ and $\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n$; then, eventually*

$$P_{KA}(\Omega_n) > 0$$

and, for any event A, we have that

$$P_{KA}(A) = \lim_{n\to\infty} P_{KA}(A \mid \Omega_n).$$

The following theorem shows that the Continuity Axiom (K4) is equivalent to (CPP).

Theorem 4. *Suppose that* (K0),...,(K3) hold. (K4) *holds if and only if* (CPP) *holds.*

Proof. Assume (K0),...,(K4) and let Ω_n be as in (CPP). By (K2) and (K4)

$$\sup_{n\in\mathbb{N}} P_{KA}(\Omega_n) = 1$$

and so eventually $P_{KA}(\Omega_n) > 0$. Now take an event A. Since $P_{KA}(A \cap \Omega_n)$ and $P_{KA}(\Omega_n)$ are monotonic sequences, we have that

$$\begin{aligned}\lim_{n\to\infty} P_{KA}(A \mid \Omega_n) &= \lim_{n\to\infty} \frac{P_{KA}(A \cap \Omega_n)}{P_{KA}(\Omega_n)} \\ &= \frac{\sup_{n\in\mathbb{N}} P_{KA}(A \cap \Omega_n)}{\sup_{n\in\mathbb{N}} P_{KA}(\Omega_n)} = \frac{P_{KA}(A)}{P_{KA}(\Omega)} = P_{KA}(A).\end{aligned}$$

Now assume (K0),...,(K3) and (CPP). Take any sequence Ω_n as in (CPP); first we want to show that

$$\lim_{n\to\infty} P_{KA}(\Omega_n) = \sup_{n\in\mathbb{N}} P_{KA}(\Omega_n) = 1. \tag{5}$$

Take $\bar{n}$ such that $P_{KA}(\Omega_{\bar{n}}) > 0$; such a $\bar{n}$ exists since (CPP) holds. Then, using (CPP) again, we have

$$P_{KA}(\Omega_{\bar{n}}) = \lim_{n\to\infty} P_{KA}(\Omega_{\bar{n}} \mid \Omega_n) = \frac{\lim_{n\to\infty} P_{KA}(\Omega_{\bar{n}} \cap \Omega_n)}{\lim_{n\to\infty} P_{KA}(\Omega_n)} = \frac{P_{KA}(\Omega_{\bar{n}})}{\lim_{n\to\infty} P_{KA}(\Omega_n)}.$$

Since $P_{KA}(\Omega_{\bar{n}}) > 0$, eq. (5) follows.

Now let A_n be a sequence as in (K4) and set

$$\Omega_n = (\Omega \setminus A) \cup A_n.$$

Then Ω_n and A satisfies the assumptions of (CPP). So, by eq. (5), we have that

$$\begin{aligned}P_{KA}(A) = \lim_{n\to\infty} P_{KA}(A \mid \Omega_n) &= \frac{\lim_{n\to\infty} P_{KA}(A \cap \Omega_n)}{\lim_{n\to\infty} P_{KA}(\Omega_n)} \\ &= \frac{\lim_{n\to\infty} P_{KA}(A_n)}{1} = \sup_{n\in\mathbb{N}} P_{KA}(A_n).\end{aligned}$$

□

So (CPP) is equivalent to (K4) and it has a form which can be compared with (NAP4). Both (CPP) and (NAP4) imply that the knowledge of the conditional probability relative to a suitable family of sets provides the knowledge of the probability of the event. In the Kolmogorovian case, we have that

$$P_{KA}(A) = \lim_{n\to\infty} P_{KA}(A \mid \Omega_n) \tag{6}$$

and in the NAP case, we have that

$$P(A) = J\left(P(A \mid \cdot)\right). \tag{7}$$

If we compare these two equations, we see that we may think of J as a particular kind of limit; this fact justifies the name *Non-Archimedean Continuity* given to (NAP4). This point will be developed in section 4.4.

3.3. First consequences of the axioms

Define a function

$$p : \Omega \to \mathcal{R}$$

as follows:

$$p(\omega) = P(\{\omega\}).$$

We choose arbitrarily a point $\omega_0 \in \Omega$, and we define the *weight* function as follows:

$$w(\omega) = \frac{p(\omega)}{p(\omega_0)}.$$

Proposition 5. *The function w takes its values in $\mathbb{R}$ and for any finite λ, the following holds:*

$$P(A \mid \lambda) = \frac{\sum_{\omega \in A \cap \lambda} w(\omega)}{\sum_{\omega \in \lambda} w(\omega)}. \tag{8}$$

Proof. Take $\omega \in \Omega$ arbitrarily and set $r = P(\{\omega\} \mid \{\omega, \omega_0\})$. By (NAP4), $r \in \mathbb{R}$ and, by the definition of conditional probability (see eq. (4)), we have that

$$r = \frac{p(\omega)}{p(\omega) + p(\omega_0)} = \frac{w(\omega)}{w(\omega) + 1} < 1$$

and hence

$$w(\omega) = \frac{r}{1 - r} \in \mathbb{R}.$$

Eq. (8) is a trivial consequence of the additivity and the definition of w:

$$P(A \mid \lambda) = \frac{\sum_{\omega \in A \cap \lambda} p(\omega)}{\sum_{\omega \in \lambda} p(\omega)} = \frac{\sum_{\omega \in A \cap \lambda} p(\omega_0) w(\omega)}{\sum_{\omega \in \lambda} p(\omega_0) w(\omega)} = \frac{\sum_{\omega \in A \cap \lambda} w(\omega)}{\sum_{\omega \in \lambda} w(\omega)}. \qquad \square$$

We recall that χ_λ denotes the indicator function of λ.

Lemma 6. $\forall \omega \in \Omega$, *we have*

$$J(\chi_\lambda(\omega)) = 1 \tag{9}$$

and

$$J\left(\sum_{\omega \in \lambda} w(\omega)\right) = \frac{1}{p(\omega_0)}. \tag{10}$$

Proof. We have that

$$\chi_\lambda(\omega)[1 - \chi_\lambda(\omega)] = 0$$
$$\chi_\lambda(\omega) + [1 - \chi_\lambda(\omega)] = 1;$$

then, setting $\xi = J(1 - \chi_\lambda(\omega))$, we have that

$$J(\chi_\lambda(\omega)) \cdot \xi = 0$$
$$J(\chi_\lambda(\omega)) + \xi = 1;$$

then $J(\chi_\lambda(\omega))$ is either 1 or 0. We will show that $J(\chi_\lambda(\omega_0)) = 1$.

By eq. (8), since $w(\omega_0) = 1$, we have that

$$p(\omega_0) = J(P(\{\omega_0\} \mid \lambda)) = J\left(\frac{w(\omega_0)\chi_\lambda(\omega_0)}{\sum_{\omega \in \lambda} w(\omega)}\right) = \frac{J(\chi_\lambda(\omega_0))}{J\left(\sum_{\omega \in \lambda} w(\omega)\right)}.$$

By Prop. 2 (ii), we know that $p(\omega_0) > 0$, and therefore we obtain that $J\left(\chi_\lambda\left(\omega_0\right)\right) = 1$ and $J\left(\sum_{\omega\in\lambda} w\left(\omega\right)\right) = 1/p(\omega_0)$. $\square$

3.4. Infinite sums

The NAP-axioms allow us to generalize the notion of sum to infinite sets in such a way that eq. (1) and eq. (8) hold also for infinite sets.

If $F \subset \Omega$ is a finite set and $x_\omega \in \mathbb{R}$ for $\omega \in \Omega$, using eq. (9), we have that

$$\sum_{\omega\in F} x_\omega = \sum_{\omega\in F}\left[x_\omega J\left(\chi_\lambda\left(\omega\right)\right)\right] = J\left(\sum_{\omega\in F} x_\omega \chi_\lambda\left(\omega\right)\right) = J\left(\sum_{\omega\in F\cap\lambda} x_\omega\right).$$

The last term makes sense also when F is infinite (since $F\cap\lambda$ is finite). Hence, it makes sense to give the following definition:

Definition 7. For any set $A \in \mathcal{P}(\Omega)$ and any function $u : A \to \mathbb{R}$, we set

$$\sum_{\omega\in A} u(\omega) = J\left(\sum_{\omega\in A\cap\lambda} u(\omega)\right).$$

Using this notation, by eq. (10), it follows that

$$\sum_{\omega\in\Omega} w\left(\omega\right) = \frac{1}{p(\omega_0)}$$

and hence we get that

$$P(A) = P(A \mid \Omega) = \frac{\sum_{\omega\in A} w\left(\omega\right)}{\sum_{\omega\in\Omega} w\left(\omega\right)} = \frac{1}{p(\omega_0)}\sum_{\omega\in A} w\left(\omega\right).$$

Moreover, for any set A and B, we have that

$$P(A \mid B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{1}{p(\omega_0)}\sum_{\omega\in A\cap B} w\left(\omega\right)}{\frac{1}{p(\omega_0)}\sum_{\omega\in B} w\left(\omega\right)} = \frac{\sum_{\omega\in A\cap B} w\left(\omega\right)}{\sum_{\omega\in B} w\left(\omega\right)}.$$

This equation extends eq. (8) when λ is infinite. Moreover, taking $B = \Omega$, we get

$$P(A) = \frac{1}{p(\omega_0)}\sum_{\omega\in A} w\left(\omega\right).$$

This equation extends eq. (1) when A is infinite since $p\left(\omega\right) = w\left(\omega\right)p(\omega_0)$.

The next proposition replaces σ-additivity:[5]

Proposition 8. *Let*

$$A = \bigcup_{j\in I} A_j$$

where I is a family of indices of any cardinality and $A_j \cap A_k = \varnothing$ for $j \neq k$; then

$$P(A) = J(\sigma)$$

[5] Because it also holds for non-denumerably infinite sample spaces, this proposition encapsulates what some philosophers have called 'perfect additivity'; see *e.g.* [10] (Vol. 1, p. 118).

where

$$\sigma(\lambda) := \sum_{j \in I} P(A_j \mid \lambda).$$

Proof. Since λ is finite, $P(A_j \mid \lambda)$ can be computed just by making finite sums:

$$\sum_{j \in I} P(A_j \mid \lambda) = \frac{\sum_{j \in I} \sum_{\omega \in A_j \cap \lambda} w(\omega)}{\sum_{\omega \in \lambda} w(\omega)} = \frac{\sum_{\omega \in A \cap \lambda} w(\omega)}{\sum_{\omega \in \lambda} w(\omega)} = P(A \mid \lambda).$$

So we have

$$J(\sigma) = J\left(\sum_{j \in I} P(A_j \mid \lambda)\right) = J(P(A \mid \lambda)) = P(A). \qquad \square$$

4. NAP-spaces and Λ-limits

In this section, we will show how to construct NAP-spaces. In particular, this construction shows that the NAP-axioms are not contradictory. Also, we will introduce the notion of Λ-limit which will be useful in the applications.

4.1. Fine ideals

Before constructing NAP-spaces, we give the definition and some properties of fine ideals.

Definition 9. Fine ideal: An ideal I in the algebra $\mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$ is called fine[6] if it is maximal and if for any $\omega \in \Omega$, $1 - \chi_\lambda(\omega) \in I$.

Proposition 10. *If Ω is an infinite set, then $\mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$ contains a fine ideal.*

Proof. We set

$$I_0 = \left\{\varphi \in \mathfrak{F}\left(\mathcal{P}^0_{fin}(\Omega), \mathbb{R}\right) \mid \exists \lambda_0 \in \mathcal{P}^0_{fin}(\Omega), \forall \lambda \supseteq \lambda_0,\ \varphi(\lambda) = 0\right\}.$$

It is easy to see that I_0 is an ideal; in fact:

- if $\forall \lambda \supseteq \lambda_0,\ \varphi(\lambda) = 0$ and $\forall \lambda \supseteq \mu_0,\ \psi(\lambda) = 0$, then $\forall \lambda \supseteq \lambda_0 \cup \mu_0,\ (\varphi + \psi)(\lambda) = 0$ and hence $\varphi + \psi \in I_0$;
- if $\forall \lambda \supseteq \lambda_0,\ \varphi(\lambda) = 0$, then, $\forall \psi \in \mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$, we have that $\forall \lambda \supseteq \lambda_0,\ \varphi(\lambda) \cdot \psi(\lambda) = 0$ and hence $\varphi \cdot \psi \in I_0$.

Moreover, $1 - \chi_\lambda(\omega) \in I_0$ since $1 - \chi_\lambda(\omega) = 0\ \forall \lambda \supseteq \lambda_0 := \{\omega\}$.

The conclusion follows taking a maximal ideal I containing I_0 which exists by Krull's theorem. $\square$

Proposition 11. *Let (Ω, P, J) be a NAP-space; then $\ker(J)$ is a fine ideal.*

[6] The name *fine* ideal has been chosen since the maximal ultrafilter

$$\mathcal{U} := \{\varphi^{-1}(0) \mid \varphi \in I\}$$

is a *fine* ultrafilter.

Proof. Since $\mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$ is a ring with identity and $\mathcal{R}$ is a field, by elementary algebra it follows that $\ker(J)$ is a maximal ideal and by eq. (9), it follows that it is fine. □

4.2. Construction of NAP-spaces

In the previous section, we have seen that, given a NAP-space (Ω, P, J) (with Ω infinite), it is possible to define the weight function $w : \Omega \to \mathbb{R}^+$ and a fine ideal $I \subset \mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$. In this section, we will show that also the converse is possible, namely that in order to define a NAP-space it is sufficient to assign

- the sample space Ω;
- a weight function $w : \Omega \to \mathbb{R}^+$;
- a fine ideal I in the algebra $\mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$.

The weight function allows to define the conditional probability of an event A with respect to an event $\lambda \in \mathcal{P}^0_{fin}(\Omega)$ according to the formula

$$P(A \mid \lambda) = \frac{\sum_{\omega \in A \cap \lambda} w(\omega)}{\sum_{\omega \in \lambda} w(\omega)}.$$

The fine ideal I allows us to define an ordered field

$$\mathcal{R}_I := \frac{\mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})}{I}$$

and an algebra homomorphism

$$J_I : \mathfrak{F}\left(\mathcal{P}^0_{fin}(\Omega), \mathbb{R}\right) \to \mathcal{R}_I \tag{11}$$

given by the canonical projection, namely

$$J_I(\varphi) = [\varphi]_I \quad \text{where} \quad [\varphi]_I := \varphi + I.$$

The map J_I allows us to define an infinite sum as in Def. 7 and to define the probability function as follows:

$$P_I(A) = \frac{\sum_{\omega \in A} w(\omega)}{\sum_{\omega \in \Omega} w(\omega)}. \tag{12}$$

Thus we have obtained the following theorem:

Theorem 12. *Given* (Ω, w, I), *the triple* (Ω, P_I, J_I) *defined by eq.* (12) *and eq.* (11) *is a NAP-space, namely it satisfies the axioms* (NAP0),…,(NAP4).

Definition 13. (Ω, P_I, J_I) will be called the NAP-space produced by (Ω, w, I) and P_I will be called the NAP-function produced by (Ω, w, I).

4.3. The Λ-property

In section 4.2, we have seen that in order to construct a NAP-space, it is sufficient to assign a triple (Ω, w, I). However, it is not possible to define I explicitly since its existence uses Zorn's lemma and no explicit construction is possible. In any case, it is possible to choose I in such a way that the NAP-theory satisfies some other properties which we would like to include in the model. Some of these properties will be described in section 5 which deals with the applications. In this section, we will give a general strategy to include these properties in the theory.

Definition 14. Directed set: A family of sets $\Lambda \subset \mathcal{P}^0_{fin}(\Omega)$ is called a directed set if

- if $\lambda_1, \lambda_2 \in \Lambda$, then $\exists \mu \in \Lambda$ such that $\lambda_1 \cup \lambda_2 \subset \mu$;
- the union of all the elements of Λ gives Ω.

The notion of directed set allows to enunciate the following property.

Definition 15. Λ-property: Given a directed set Λ, we say that a NAP-space $(\Omega, P, \mathcal{R})$ satisfies the Λ-property: if, given $A, B \in \mathcal{P}(\Omega)$ such that

$$\forall \lambda \in \Lambda,\ P(A \cap \lambda) = P(B \cap \lambda),$$

then

$$P(A) = P(B).$$

Given any directed set $\Lambda \subset \mathcal{P}^0_{fin}(\Omega)$, it is easy to construct a NAP-space which satisfies the Λ-property. Given Λ, we define the ideal

$$I_{0,\Lambda} := \left\{ \varphi \in \mathfrak{F}\left(\mathcal{P}^0_{fin}(\Omega), \mathbb{R}\right) \mid \forall \lambda \in \Lambda,\ \varphi(\lambda) = 0 \right\};$$

by Krull's theorem there exists a maximal ideal $I_\Lambda \supset I_{0,\Lambda}$. It is easy to check that I_Λ is a fine ideal. Then, given a weight function w, we can consider the NAP-space produced by (Ω, w, I_Λ) and we have that:

Theorem 16. *The NAP-space produced by (Ω, w, I_Λ) satisfies the Λ-property.*

Proof. Given $A, B \in \mathcal{P}(\Omega)$ as in def. 15, set

$$\varphi(\lambda) = P(A \cap \lambda) - P(B \cap \lambda);$$

then $\forall \lambda \in \Lambda,\ \varphi(\lambda) = 0,$ and hence $\varphi \in I_{0,\Lambda} \subset I_\Lambda$ and so $J(\varphi(\lambda)) = 0$. Then we have:

$$\begin{aligned} P(A) - P(B) &= J(P(A \mid \lambda)) - J(P(B \mid \lambda)) \\ &= J\left(\frac{P(A \cap \lambda) - P(B \cap \lambda)}{P(\lambda)}\right) = \frac{J(\varphi(\lambda))}{J(P(\lambda))} = 0. \end{aligned}$$

Then the Λ-property holds. $\square$

4.4. The Λ-limit

If we compare eq. (6) and eq. (7), it makes sense to think of J as a particular kind of limit and to write eq. (7) as follows:

$$P(A) = \lim_{\lambda \in \mathcal{P}^0_{fin}(\Omega);\ \lambda \uparrow \Omega} P(A \mid \lambda).$$

More in general, we can define the following limit:

$$J(\varphi) = \lim_{\lambda \in \mathcal{P}^0_{fin}(\Omega);\ \lambda \uparrow \Omega} \varphi(\lambda),$$

where $\varphi \in \mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$. The above limit is determined by the choice of the ideal $I_\Lambda \subset \mathfrak{F}(\mathcal{P}^0_{fin}(\Omega), \mathbb{R})$, and, by Th. 16, it depends on the values that φ assumes on Λ; we can assume that $\varphi \in \mathfrak{F}(\Lambda, \mathbb{R})$. Then, many properties of this limit depend on the choice of Λ; so we will call it Λ-limit, and we will use the following notation

$$J(\varphi) = \lim_{\lambda \in \Lambda} \varphi(\lambda)$$

which is simpler and carries more information. Notice that the Λ-limit, unlike the usual limit, exists for any function $\varphi \in \mathfrak{F}(\Lambda, \mathbb{R})$ and that it takes its values in the non-Archimedean field $\mathcal{R}_{I_\Lambda}$.

The following theorem shows some other properties of the Λ-limit. These properties, except (v), are shared by the usual limit. However, (v) and the fact that the Λ-limit always exists, make this limit quite different from the usual one.

Theorem 17. *Let $\varphi, \psi : \Lambda \to \mathbb{R}$; then:*

(i) *If $\varphi_r(\lambda) = r$ is constant, then*

$$\lim_{\lambda \in \Lambda} \varphi_r(\lambda) = r.$$

(ii)

$$\lim_{\lambda \in \Lambda} \varphi(\lambda) + \lim_{\lambda \in \Lambda} \psi(\lambda) = \lim_{\lambda \in \Lambda} \left(\varphi(\lambda) + \psi(\lambda)\right).$$

(iii)

$$\lim_{\lambda \in \Lambda} \varphi(\lambda) \cdot \lim_{\lambda \in \Lambda} \psi(\lambda) = \lim_{\lambda \in \Lambda} \left(\varphi(\lambda) \cdot \psi(\lambda)\right).$$

(iv) *If $\varphi(\lambda)$ and $\psi(\lambda)$ are eventually equal, namely $\exists \lambda_0 \in \Lambda :\ \forall \lambda \supset \lambda_0,\ \varphi(\lambda) = \psi(\lambda)$, then*

$$\lim_{\lambda \in \Lambda} \varphi(\lambda) = \lim_{\lambda \in \Lambda} \psi(\lambda).$$

(v) *If $\varphi(\lambda)$ and $\psi(\lambda)$ are eventually different, namely $\exists \lambda_0 \in \Lambda :\ \forall \lambda \supset \lambda_0,\ \varphi(\lambda) \neq \psi(\lambda)$, then*

$$\lim_{\lambda \in \Lambda} \varphi(\lambda) \neq \lim_{\lambda \in \Lambda} \psi(\lambda).$$

(vi) *If, for any λ, $\varphi(\lambda)$ has finite range, namely $\varphi(\lambda) \in \{r_1, \dots, r_n\}$, then*

$$\lim_{\lambda \in \Lambda} \varphi(\lambda) = r_j$$

for some $j \in \{1, \dots, n\}$.

Proof. (i) We have that

$$\lim_{\lambda\in\Lambda}\varphi(\lambda)=J(r\cdot 1)=r\cdot J(1)=r\cdot 1=r.$$

(ii) and (iii) are immediate consequences of the fact that J is an homomorphism.

(iv) Suppose that $\forall\lambda\supset\lambda_0,\ \varphi(\lambda)=\psi(\lambda)$; we set $\lambda_0=\{\omega_1,\ldots,\omega_n\}$ and

$$\zeta\left(\lambda\right)=\chi_\lambda(\omega_1)\cdot\ldots\cdot\chi_\lambda(\omega_n).$$

If $\lambda_0\setminus\lambda\neq\varnothing$, then $\zeta\left(\lambda\right)=0$ since some of the $\chi_\lambda(\omega_j)$ vanish; if $\lambda_0\setminus\lambda=\varnothing$, then $\varphi(\lambda)-\psi(\lambda)=0$ by our assumptions; therefore,

$$\zeta\left(\lambda\right)\cdot\left[\varphi(\lambda)-\psi(\lambda)\right]=0. \tag{13}$$

Moreover, by eq. (9), we have that

$$J(\zeta)=J(\chi_\lambda(\omega_1))\cdot\ldots\cdot J(\chi_\lambda(\omega_n))=1$$

and so, by eq. (13),

$$\begin{aligned}\lim_{\lambda\in\Lambda}\varphi(\lambda)-\lim_{\lambda\in\Lambda}\psi(\lambda)&=J\left(\varphi-\psi\right)=J\left(\varphi-\psi\right)\cdot J(\zeta)\\&=J\left(\left[\varphi-\psi\right]\cdot\zeta\right)=J(0)=0.\end{aligned}$$

(v) We set

$$\theta(\lambda)=\begin{cases}1 & \text{if } \lambda_0\setminus\lambda\neq\varnothing\\ \frac{1}{\varphi(\lambda)-\psi(\lambda)} & \text{if } \lambda\supset\lambda_0;\end{cases}$$

then $\forall\lambda\supset\lambda_0$,

$$\left(\varphi(\lambda)-\psi(\lambda)\right)\cdot\theta(\lambda)=1$$

and hence, by (i) and (iv),

$$\begin{aligned}1&=\lim_{\lambda\in\Lambda}\left[\left(\varphi(\lambda)-\psi(\lambda)\right)\cdot\theta(\lambda)\right]\\&=\lim_{\lambda\in\Lambda}\left(\varphi(\lambda)-\psi(\lambda)\right)\cdot\lim_{\lambda\in\Lambda}\theta(\lambda).\end{aligned}$$

From here, it follows that $\lim_{\lambda\in\Lambda}\left(\varphi(\lambda)-\psi(\lambda)\right)\neq 0$ and by (ii) we get that $\lim_{\lambda\in\Lambda}\varphi(\lambda)\neq\lim_{\lambda\in\Lambda}\psi(\lambda)$.

(vi) We have that

$$\left(\varphi(\lambda)-r_1\right)\cdot\ldots\cdot\left(\varphi(\lambda)-r_n\right)=0;$$

then, taking the Λ-limit,

$$\left(\lim_{\lambda\in\Lambda}\varphi(\lambda)-r_1\right)\cdot\ldots\cdot\left(\lim_{\lambda\in\Lambda}\varphi(\lambda)-r_n\right)=0;$$

and hence, there is a j such that

$$\lim_{\lambda\in\Lambda}\varphi(\lambda)-r_j=0$$

and so $\lim_{\lambda\in\Lambda}\varphi(\lambda)=r_j$. □

If we use the notion of Λ-limit, def. (7) becomes more meaningful; in this case in order to define an infinite sum $\sum_{\omega\in A} u(\omega)$, we define the *partial sum* as follows

$$\sum_{\omega\in A\cap\lambda} u(\omega), \quad \text{with } \lambda \in \mathcal{P}^0_{fin}(\Omega)$$

and then, we define the infinite sum as the Λ-limit of the partial sums, namely

$$\sum_{\omega\in A} u(\omega) = \lim_{\lambda\in\Lambda} \sum_{\omega\in A\cap\lambda} u(\omega).$$

Moreover, the notion of Λ-limit provides also a meaningful characterization of the field $\mathcal{R}_{I_\Lambda}$:

$$\mathcal{R}_{I_\Lambda} = \left\{ \lim_{\lambda\in\Lambda} \varphi(\lambda) \mid \varphi \in \mathfrak{F}(\Lambda, \mathbb{R}) \right\}. \tag{14}$$

Concluding, we have obtained the following 'general strategy' for defining NAP-spaces:

General strategy. In the applications, in order to define a NAP-space we will assign

- the sample space Ω;
- a weight function $w : \Omega \to \mathbb{R}^+$;
- a directed set $\Lambda \subseteq \mathcal{P}^0_{fin}(\Omega)$ which provides a notion of Λ-limit and, via eq. (14), the appropriate non-Archimedean field.

4.5. The field $\mathbb{R}^*$

In this section, we will describe a non-Archimedean field $\mathbb{R}^*$ which contains the range of any non-Archimedean probability P one may wish to consider in applied mathematics.

To do this, we assume the existence of uncountable, non-accessible cardinal numbers and, as usual, we will denote the smallest of them by κ. If we assume the existence of κ, then there exists a nonstandard model $\mathbb{R}^*$ of cardinality κ and κ-saturated. This fact implies that it is unique up to isomorphisms. We refer to [14] (p. 195) for details.

Given a NAP-space $(\Omega, P_I, \mathcal{R}_I)$, with Ω infinite, we have that also $\mathcal{R}_I$ is a nonstandard model of $\mathbb{R}$ and if $|\Omega| < \kappa$, we have that $\mathcal{R}_I \subset \mathbb{R}^*$.

So using $\mathbb{R}^*$ and the notion of Λ-limit, axioms (NAP0) and (NAP4) can be reformulated as follows:

- (NAP0)* **Domain and range.** *The events are* **all** *the elements of* $\mathcal{P}(\Omega)$ *and the probability is a function*

  $$P : \mathcal{P}(\Omega) \to \mathbb{R}^*$$

  where $\mathbb{R}^*$ *is the unique* κ*-saturated nonstandard model of* $\mathbb{R}$ *having cardinality* κ*.*
- (NAP4)* **Non-Archimedean Continuity.** *Let* $P(A \mid B)$, $B \neq \varnothing$, *denote the conditional probability, then,* $P(A \mid \lambda) \in \mathbb{R}$ and

  $$P(A) = \lim_{\lambda\in\Lambda} P(A \mid \lambda)$$

for some directed set $\Lambda \subset \mathfrak{F}\left(\mathcal{P}^0_{fin}(\Omega), \mathbb{R}\right)$.

Remark 18. The previous remark shows some relation between NAP-theory and nonstandard analysis. Actually, the relation is deeper than it appears here. In fact, NAP could be constructed within a nonstandard universe based on the notion of Λ-limit (see [8]). The idea to use NSA in probability theory is quite old and we refer to [15] and the references therein; these approaches differ from ours since they use NSA as a tool aimed at finding real-valued probability functions. Another approach to probability related to NSA is due to Nelson [20]; however, recall that also his approach is quite different from ours since it takes the domain of the probability function to be a nonstandard set too.

5. Some applications

5.1. Fair lotteries and numerosities

Definition 19. Fair: If, $\forall \omega_1, \omega_2 \in \Omega$, $p(\omega_1) = p(\omega_2)$, then the probability theory (Ω, P, J) is called *fair*.

If (Ω, P, J) is a fair lottery and Ω is finite, then, if we set $\varepsilon_0 = p(\omega_0) = 1/|\Omega|$, it turns out that, for any set $A \in \mathcal{P}^0_{fin}(\Omega)$,

$$|A| = \frac{P(A)}{\varepsilon_0}.$$

This remark suggests the following definition:

Definition 20. Numerosity: If (Ω, P, J) is a fair lottery, the numerosity of a set $A \in \mathcal{P}(\Omega)$ is defined as follows:

$$\mathfrak{n}(A) = \frac{P(A)}{\varepsilon_0}$$

where ε_0 is the probability of an elementary event in Ω.

In particular if A is finite, we have that $\mathfrak{n}(A) = |A|$. So the numerosity is the generalization to infinite sets of the notion of "number of elements of a set" different from the Cantor theory of infinite sets.

The theory of numerosity has been introduced in [2] and developed in various directions in [11], [3], and [5]. The definition above is an alternative way to introduce a numerosity theory.

We now set

$$\mathbb{Q}^* = \left\{ \lim_{\lambda \in \Lambda} \varphi(\lambda) \mid \forall \lambda \in \Lambda,\ \varphi(\lambda) \in \mathbb{Q} \text{ for some } \Lambda \text{ with } |\Lambda| < \kappa \right\}.$$

We will refer to $\mathbb{Q}^*$ as the field of hyperrational numbers.

Proposition 21. *If* (Ω, P, J) *is a fair lottery, and* $A, B \subseteq \Omega$*, with* $B \neq \varnothing$*, then*

$$P(A \mid B) \in \mathbb{Q}^*.$$

Proof. We have that

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \lim_{\lambda \in \mathcal{P}^0_{fin}(\Omega)} \frac{P(A \cap B \cap \lambda)}{P(B \cap \lambda)} = \lim_{\lambda \in \mathcal{P}^0_{fin}(\Omega)} \frac{|A \cap B \cap \lambda|}{|B \cap \lambda|}.$$

The conclusion follows from the fact that

$$\frac{|A \cap B \cap \lambda|}{|B \cap \lambda|} \in \mathbb{Q}.$$

□

5.2. A fair lottery on $\mathbb{N}$

Let us consider a fair lottery in which exactly one winner is randomly selected from a countably infinite set of tickets. We assume that these tickets are labeled by the (positive) natural numbers. We call such a lottery the "de Finetti lottery".[7] Now let us construct a NAP-space for such a lottery using the general strategy developed in section 4.4.

We take

$$\Omega = \mathbb{N},$$

$$w : \mathbb{N} \to \mathbb{R}^+ \text{ identically equal to one}$$

and

$$\Lambda_{[n]} = \left\{ \lambda_n \in \mathcal{P}^0_{fin}(\Omega) \mid n \in \mathbb{N} \right\} \tag{15}$$

where

$$\lambda_n = \{1, 2, 3, \ldots, n\}.$$

In this case we have that, for every $A \in \mathcal{P}(\Omega)$

$$P(A \mid \lambda_n) = \frac{|A \cap \{1, \ldots, n\}|}{n}$$

and hence

$$P(A) = \lim_{\lambda_n \in \Lambda} \frac{|A \cap \{1, \ldots, n\}|}{n}.$$

Using the notion of numerosity as defined in section 5.1, we set

$$\alpha := \mathfrak{n}(\mathbb{N}) = \frac{1}{p(1)};$$

then the probability of any event $A \in \mathcal{P}(\Omega)$ can be written as follows:

$$P(A) = \frac{\mathfrak{n}(A)}{\alpha}.$$

So, the probability of A is ratio of the "number" of the elementary events in A and the total "number" of elements α (where "number" is understood in terms of numerosity).

One of the properties which our intuition wants to be satisfied by the de Finetti lottery is the 'Asymptotic Density Property', which relates the non-Archimedean probability function P to a classical limit (in as far as the latter exists):

[7]This example has been discussed by multiple philosophers of probability, including de Finetti [10]; at the end of this section, we indicate how our solution relates to his approach. The solution presented here rephrases the one given in [23] within the more general NAP framework.

Definition 22. Asymptotic Density Property. Let $A \in \mathcal{P}(\mathbb{N})$ be a set which has an asymptotic density (or natural density), namely there exists $L \in [0,1]$ such that

$$\lim_{n\to\infty} \frac{|A \cap \{1,\dots,n\}|}{n} = L.$$

We say that the Asymptotic Density Property holds if we have

$$P(A) \sim L. \tag{16}$$

It is easy to check that

Proposition 23. *If P is the NAP-function produced by $\left(\mathbb{N}, 1, I_{\Lambda_{[n]}}\right)$, it satisfies the Asymptotic Density Property.*

Now, we will consider additional properties which would be nice to have and we will show how the choice of Λ works. For example, the probability of extracting an even number seems to be equal to that of extracting an odd number; thus we must have

$$P(\mathbb{E}) = P(\mathbb{O}) \tag{17}$$

and since by (NAP2) and (NAP3), we have

$$P(\mathbb{E}) + P(\mathbb{O}) = 1,$$

it follows that

$$P(\mathbb{E}) = P(\mathbb{O}) = \frac{1}{2}. \tag{18}$$

Now, let us compute for example $P(\mathbb{E})$. We have that

$$P(\mathbb{E}) = \frac{\mathfrak{n}(\mathbb{E})}{\alpha}$$

so we have to compute $\mathfrak{n}(\mathbb{E})$:

$$\begin{aligned}
\mathfrak{n}(\mathbb{E}) &= \lim_{\lambda_n\in\Lambda} |\mathbb{E} \cap \{1,\dots,n\}| \\
&= \lim_{\lambda_n\in\Lambda} \left|\left\{2,4,6,\dots,2\cdot\left[\frac{n}{2}\right]\right\}\right| \\
&= \lim_{\lambda_n\in\Lambda} \left[\frac{n}{2}\right] = \lim_{\lambda_n\in\Lambda} \frac{n}{2} - \lim_{\lambda_n\in\Lambda} c_n
\end{aligned}$$

where $[r]$ denotes the integer part of r and

$$c_n = \begin{cases} 1/2 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even.} \end{cases}$$

Then, by Theorem 17(vi), $\lim_{\lambda_n\in\Lambda} c_n$ is either 0 or 1/2, but this fact cannot be determined since we do not know the ideal I_Λ. However, if we think that this fact is relevant for our model, we can follow the strategy suggested in section 4.3 and make a better choice of Λ. If we choose a smaller Λ, it carries more information.

For example, we can replace the choice of eq. (15) with the following one:

$$\Lambda_{[2m]} := \{\{1,2,\dots,2m\} \mid m \in \mathbb{N}\}.$$

In this case we have that

$$\mathfrak{n}(\mathbb{E}) = \lim_{\lambda_n \in \Lambda_{[2m]}} \left[\frac{n}{2}\right] = \lim_{\lambda_n \in \Lambda_{[2m]}} \frac{n}{2} = \frac{\alpha}{2}.$$

On the other hand, we can choose

$$\Lambda = \Lambda_{[2m-1]} := \{\{1, 2, \ldots, 2m-1\} \mid m \in \mathbb{N}\}$$

and in this case

$$\mathfrak{n}(\mathbb{E}) = \lim_{\lambda \in \Lambda_{[2m-1]}} \left(\left[\frac{n}{2}\right] - \frac{1}{2}\right) = \lim_{\lambda \in \Lambda_{[2m-1]}} \left(\frac{n-1}{2}\right) = \frac{\alpha - 1}{2}.$$

Thus $P(\mathbb{E}) = \frac{1}{2}$ or $\frac{1}{2} - \frac{1}{2\alpha}$ depending on the choice of Λ. Also, it is possible to prove that any choice of $\Lambda \subseteq \Lambda_{[n]}$ gives one of these two possibilities.

Remark 24. The equality

$$P(A) = L \tag{19}$$

cannot replace eq. (16) for all the sets which have an asymptotic density; in fact take two sets A and $B = A \cup F$ where F is a finite set (with $A \cap F = \varnothing$). Then, if L is the asymptotic density of A, then it is also the asymptotic density of B since

$$\begin{aligned} &\lim_{n\to\infty} \frac{|B \cap \{1, \ldots, n\}|}{n} \\ &= \lim_{n\to\infty} \frac{|A \cap \{1, \ldots, n\}|}{n} + \frac{|F \cap \{1, \ldots, n\}|}{n} \\ &= L + 0 = L. \end{aligned}$$

On the other hand, by (NAP3)

$$P(B) = P(A) + P(F)$$

and by (NAP1), $P(F) > 0$ and hence $P(B) \neq P(A)$. Thus, it is not possible that $P(A) = L$ **and** $P(B) = L$.

Even if eq. (19) cannot hold for all the sets, our intuition may suggest that in some cases it should be true and it would be nice if eq. (19) holds for a *distinguished* family of sets. For example, if we have eq. (17) then eq. (18) holds, and hence $P(\mathbb{E})$ and $P(\mathbb{O})$ have the probability equal to their asymptotic density. So, the following question arises naturally: is it possible to have a 'de Finetti probability space' produced by $\{\mathbb{N}, 1, I_\Lambda\}$ such that

$$P(\mathbb{N}_k) = \frac{1}{k}$$

where

$$\mathbb{N}_k = \{k,\ 2k,\ 3k, \ldots, nk, \ldots\}.$$

The answer is yes; it is sufficient to choose

$$\Lambda = \Lambda_{[m!]} := \{\{1, \ldots, m!\} \mid m \in \mathbb{N}\}. \tag{20}$$

In fact, in this case, we have

$$P(\mathbb{N}_k) = \frac{\lim_{\lambda_n \in \Lambda_{[m!]}} \ |\mathbb{N}_k \cap \{1, \ldots, m!\}|}{\alpha}$$
$$= \frac{\lim_{\lambda_n \in \Lambda_{[m!]}} \ \left[\frac{n}{k}\right]}{\alpha} = \frac{\lim_{\lambda_n \in \Lambda_{[m!]}} \ \frac{n}{k}}{\alpha} = \frac{1}{k}.$$

More in general, if $\Lambda = \Lambda_{[m!]}$, we can prove that the sets

$$\mathbb{N}_{k,l} = \{k - l,\ 2k - l,\ 3k - l, \ldots, nk - l, \ldots\}, \text{ with } l \in \{0, \ldots, k - 1\}$$

have probability $1/k$, namely the same probability as their asymptotic density. This generalizes the situation which we have analyzed before where $\mathbb{E} = \mathbb{N}_2$ and $\mathbb{O} = \mathbb{N}_{2,1}$.

Remark 25. Our construction of a non-Archimedean probability P, allows to construct the following *Archimedean* probability function:

$$P_{Arch}(A) = st(P(A))$$

where $st\,(\xi)$ denotes the standard part of ξ, namely the unique standard number infinitely close to ξ. P_{Arch} is defined on **all** the subsets of A, it is finitely additive and it coincides with the asymptotic density when it exists. Although we prefer a theory based on non-Archimedean probabilities, we regard de Finetti's reaction to the infinite lottery puzzle as an equally valid approach.[8] The construction of P_{Arch} shows how the two approaches are connected.

5.3. A fair lottery on $\mathbb{Q}$

A fair lottery on $\mathbb{Q}$, by definition, is a NAP-space produced by $(\mathbb{Q}, 1, I)$ for any arbitrary I; however, as in the case of de Finetti lottery, we are allowed to require some additional properties which appear natural to our intuition and then, we can inquire if they are consistent.

For example, if we have two intervals $[a_0, b_0)_{\mathbb{Q}} \subset [a_1, b_1)_{\mathbb{Q}}$, we expect the conditional probability to satisfy the following formula:

$$P([a_0, b_0)_{\mathbb{Q}} \,|\, [a_1, b_1)_{\mathbb{Q}}) = \frac{b_0 - a_0}{b_1 - a_1}. \tag{21}$$

In fair lotteries, the probability is strictly related to the notion of numerosity and the above formula is equivalent to the following one

$$\mathfrak{n}([a, b)_{\mathbb{Q}}) = (b - a) \cdot \mathfrak{n}([0, 1)_{\mathbb{Q}}) \tag{22}$$

namely that the "*number of elements contained in an interval*" is proportional to its length.

Clearly, eq. (21) follows from eq. (22). In fact,

$$P([a, b)_{\mathbb{Q}}) = \frac{\mathfrak{n}([a, b)_{\mathbb{Q}})}{\mathfrak{n}(\mathbb{Q})} = (b - a) \cdot \frac{\mathfrak{n}([0, 1)_{\mathbb{Q}})}{\mathfrak{n}(\mathbb{Q})}.$$

---

[8] De Finetti was aware that probability could be treated as a non-Archimedean quantity, but rejected this approach as "a useless complication of language", which "leads one to puzzle over 'les infiniment petits' "[10] (Vol. 2, p. 347). He proposed to stay within an Archimedean range, but to relax Kolmogorov's countable additivity to finite additivity.

Then

$$P([a_0,b_0)_{\mathbb{Q}} \,|\, [a_1,b_1)_{\mathbb{Q}}) = \frac{P([a_0,b_0)_{\mathbb{Q}})}{P([a_1,b_1)_{\mathbb{Q}})} = \frac{b_0-a_0}{b_1-a_1}.$$

Also, it is easy to check that eq. (22) follows from eq. (21).

In order to prove that the property of eq. (22) is consistent with a NAP-theory, it is sufficient to find an appropriate family $\Lambda \subset \mathcal{P}^0_{fin}(\mathbb{Q})$.

We will consider the family

$$\Lambda_{\mathbb{Q}} := \{\mu_n \mid n = m!,\ m \in \mathbb{N}\} \tag{23}$$

with

$$\begin{aligned}\mu_n &= \left\{\frac{p}{n} \mid p \in \mathbb{Z},\ \left|\frac{p}{n}\right| \le n\right\} \qquad (24)\\ &= \left\{-n, -\frac{n^2-1}{n}, \ldots, -\frac{1}{n}, 0, \frac{1}{n}, \ldots, \frac{n^2-1}{n},\ n\right\}.\end{aligned}$$

In this case we have that:

Proposition 26. *If P is the NAP-function produced by $(\mathbb{Q},1,I_\Lambda)$ with $\Lambda = \Lambda_{\mathbb{Q}}$ defined by eq.* (23)*, then eq.* (22) *holds.*

Proof. We write a and b as fractions with the same denominator:

$$a = \frac{p_a}{q};\ b = \frac{p_b}{q}.$$

Then, if you take n sufficiently large (*e.g.* $n = m!,\ m \ge \max(a,b,q)$), q divides n and we have that

$$\left|[a,b)_{\mathbb{Q}} \cap \mu_n\right| = \left|\left\{a,\ a+\frac{1}{n},\ a+\frac{2}{n}, \ldots, b-\frac{1}{n}\right\}\right| = (b-a)\cdot n.$$

From here, eq. (22) easily follows. □

Let us compare the NAP-spaces produced by $(\mathbb{N},1,I_{\Lambda_{\mathbb{N}}})$ and $\left(\mathbb{Q},1,I_{\Lambda_{\mathbb{Q}}}\right)$, where we have set

$$\Lambda_{\mathbb{N}} = \{\mu_n \cap \mathbb{N} \mid \mu_n \in \Lambda_{\mathbb{Q}}\}. \tag{25}$$

We want to show that this choice of $\Lambda_{\mathbb{N}}$ and $\Lambda_{\mathbb{Q}}$ makes these two NAP-theories consistent in the sense described below. First of all, notice that $\Lambda_{\mathbb{N}}$ defined by eq. (25) coincides with $\Lambda_{[m!]}$ defined by eq. (20) and that $\lambda_n = \mu_n \cap \mathbb{N}$.

If we denote by $P_{\mathbb{N}}$ and $P_{\mathbb{Q}}$ the respective probabilities and we take $A \subset \mathbb{N} \subset \mathbb{Q}$, we have that

$$P_{\mathbb{N}}(A) = P_{\mathbb{Q}}(A \mid \mathbb{N}). \tag{26}$$

In fact

$$P_{\mathbb{Q}}(A \mid \mathbb{N}) = \frac{\mathfrak{n}\,(A \cap \mathbb{N})}{\mathfrak{n}\,(\mathbb{N})} = \frac{\mathfrak{n}\,(A)}{\mathfrak{n}\,(\mathbb{N})} = P_{\mathbb{N}}(A).$$

Moreover, if we set, as usual, $\alpha := \mathfrak{n}\,(\mathbb{N})$, it is easy to check that

$$\begin{aligned}\mathfrak{n}\left(\mathbb{Q}^+\right) &= \alpha^2\\ \mathfrak{n}\,(\mathbb{Q}) &= 2\alpha^2+1.\end{aligned}$$

Thus, in this model, in a 'lottery with rational numbers' the probability of extracting a positive natural number is

$$P_{\mathbb{Q}}(\mathbb{N}) = \frac{\mathfrak{n}\,(\mathbb{N})}{\mathfrak{n}\,(\mathbb{Q})} = \frac{\alpha}{2\alpha^2+1} = \frac{1-\varepsilon}{2\alpha}$$

where ε is a positive infinitesimal.

5.4. A fair lottery on $\mathbb{R}$

Now let us consider a fair lottery on $\mathbb{R}$, namely a NAP-space produced by $(\mathbb{R}, 1, I_{\Lambda_{\mathbb{R}}})$ and let us examine the other properties which we would like to have.[9] Considering the example of the previous section, we would like to have the analog of eq. (21) just replacing $\mathbb{Q}$ with $\mathbb{R}$. This is not possible. Let us see why not. If in eq. (21) we take $[a_0, b_0)_{\mathbb{R}} = [0,1)_{\mathbb{R}}$ and $[a_1, b_1)_{\mathbb{R}} = \left[0, \sqrt{2}\right)_{\mathbb{R}}$, we would get

$$P\left([0,1)_{\mathbb{R}} \mid [0,\sqrt{2})_{\mathbb{R}}\right) = \frac{1}{\sqrt{2}}$$

and this fact is not possible by Prop. 21 (in fact $\frac{1}{\sqrt{2}} \notin \mathbb{Q}^*$ by Th. 17 (v) and the definition of $\mathbb{Q}^*$). However, we can require a weaker statement, namely that, given two intervals $[a_0, b_0)_{\mathbb{R}} \subset [a_1, b_1)_{\mathbb{R}}$,

$$P\left([a_0, b_0)_{\mathbb{R}} \mid [a_1, b_1)_{\mathbb{R}}\right) \sim \frac{b_0 - a_0}{b_1 - a_1}. \tag{27}$$

Actually, in terms of numerosities, we can require that

$$\forall a, b \in \mathbb{Q},\ \mathfrak{n}([a,b)_{\mathbb{R}}) = (b-a) \cdot \mathfrak{n}([0,1)_{\mathbb{R}} \tag{28}$$

$$\forall a, b \in \mathbb{R},\ \mathfrak{n}([a,b)_{\mathbb{R}}) = [(b-a)+\varepsilon] \cdot \mathfrak{n}([0,1)_{\mathbb{R}} \tag{29}$$

where ε is an infinitesimal which might depend on a and b.

Now we will define $\Lambda_{\mathbb{R}} \subset \mathcal{P}^0_{fin}(\mathbb{R})$ in such a way that eq. (28) and eq. (29) are satisfied; first, we set

$$\Theta = \mathcal{P}_{fin}([0,1]_{\mathbb{R}} \setminus [0,1]_{\mathbb{Q}})$$

namely Θ is the family of the non-empty finite sets of irrational numbers between 0 and 1. Then for any $n \in \mathbb{N}$ and $\theta \in \Theta$, we set

$$\mu_{n,\theta} := \mu_n \cup \left\{ \frac{p+a}{n} \mid \frac{p}{n} \in \mu_n \setminus \{n\},\ a \in \theta \right\}$$

where μ_n is defined by eq. (24). You may think to have "constructed" $\mu_{n,\theta}$ in the following way: you start with the segment $[-n, n]$ and you divide it in n^2 parts of length $1/n$ of the form $\left[\frac{p}{n}, \frac{p+1}{n}\right]$, $p = -n^2, -n^2+1, \ldots, n^2-1$. In each of these parts you put a "rescaled" copy of θ, namely points of the form $\frac{p+a}{n}$ with $a \in \theta$. Thus, the set $\mu_{n,\theta}$ contains n^2+1 rational numbers and $n^2 \cdot |\theta|$ irrational numbers.

---

[9] The problem of a fair lottery on a non-denumerable sample space is usually presented as a fair lottery (or random darts throw) on the unit interval of $\mathbb{R}$ (or a darts board whose perimeter is indexed by this interval), see *e.g.* [12]. The related problem of a random darts throw on the unit square of $\mathbb{R}^2$ is considered in [1].

Then we set

$$\Lambda_{\mathbb{R}} = \{\mu_{n,\theta} \mid n = m!,\ m \in \mathbb{N},\ \theta \in \Theta\}\,. \tag{30}$$

Proposition 27. *If P is the NAP-function produced by $(\mathbb{R}, 1, I_\Lambda)$ with $\Lambda = \Lambda_{\mathbb{R}}$ given by eq.* (30)*, then eq.* (28) *and eq.* (29) *hold.*

Proof. Take an interval $[a, b)$ with $a, b \in \mathbb{Q}$; then, if n is sufficiently large $[a, b) \cap \mu_{n,\theta}$ contains $n\,(b-a)$ rational numbers and $n\,|\theta|\,(b-a)$ irrational numbers so that

$$|[a, b) \cap \mu_{n,\theta}| = n\,(b-a)\,(|\theta| + 1)\,.$$

Then, if we choose $a = 0$ and $b = 1$, we have that

$$\mathfrak{n}([0,1)_{\mathbb{R}}) = \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} |[0,1) \cap \mu_{n,\theta}| = \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} n\,(|\theta| + 1)\,.$$

Then, in general we have that

$$\begin{aligned}\mathfrak{n}([a,b)_{\mathbb{R}}) &= \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} |[a,b) \cap \mu_{n,\theta}| = \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} [n\,(b-a)\,(|\theta|+1)] \\ &= (b-a)\cdot \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} n\,(|\theta|+1) = (b-a)\,\mathfrak{n}([0,1)_{\mathbb{R}}).\end{aligned}$$

Then eq. (28) holds. Now let us prove eq. (29).

If $a \in \mathbb{R}\setminus\mathbb{Q}$ or $b \in \mathbb{R}\setminus\mathbb{Q}$, then $[a, b) \cap \mu_n$ contains at most $n\,(b-a)+1$ rational numbers, in fact, in $[a, b)$ you can fit at most $n\,(b-a)$ intervals of the form $\left[\frac{p}{n}, \frac{p+1}{n}\right]$. Moreover, since you have at most $n\,(b-a)$ intervals of the form $\left[\frac{p}{n}, \frac{p+1}{n}\right]$ and two smaller intervals at the extremes of the form $\left[a, \frac{p_a}{n}\right]$ and $\left[\frac{p_b}{n}, b\right]$ for a suitable choice of p_a and p_b, $[a, b) \cap \mu_{n,\theta}$ contains at most $[n\,(b-a)+2]\,|\theta|$ irrational numbers; then

$$\begin{aligned}|[a,b) \cap \mu_{n,\theta}| &\le n\,(b-a) + 1 + [n\,(b-a)+2]\,|\theta| \\ &= n\,(b-a)\,(|\theta|+1) + 2\,|\theta| + 1 \\ &\le n\left(b - a + \frac{2}{n}\right)(|\theta|+1)\,.\end{aligned}$$

Thus

$$\begin{aligned}\mathfrak{n}([a,b)_{\mathbb{R}}) &= \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} |[a,b) \cap \mu_{n,\theta}| \le \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} \left[n\left(b-a+\frac{2}{n}\right)(|\theta|+1)\right] \\ &\le \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} \left(b-a+\frac{2}{n}\right)\cdot \lim_{\mu_{n,\theta}\in\Lambda_{\mathbb{R}}} n\,(|\theta|+1) \\ &= \left(b-a+\frac{2}{\alpha}\right)\mathfrak{n}([0,1)_{\mathbb{R}}).\end{aligned}$$

Moreover, since $[a,b)$ contains at least $n(b-a)-1$ intervals of type $\left[\frac{p}{n},\frac{p+1}{n}\right]$, arguing in a similar way as before, we have that

$$\begin{aligned}
|[a,b)\cap\mu_{n,\theta}| &\geq n(b-a)+[n(b-a)-1]\,|\theta| \\
&= n(b-a)(|\theta|+1)-|\theta| \\
&\geq n\left(b-a-\frac{1}{n}\right)(|\theta|+1).
\end{aligned}$$

Then,

$$\begin{aligned}
\mathfrak{n}([a,b)_{\mathbb{R}}) = \lim_{\mu_{n,\theta}\uparrow\mathbb{R}} |[a,b)\cap\mu_{n,\theta}| &\geq \lim_{\mu_{n,\theta}\uparrow\mathbb{R}} \left[n\left(b-a-\frac{1}{n}\right)(|\theta|+1)\right] \\
&\geq \lim_{\mu_{n,\theta}\uparrow\mathbb{R}} \left(b-a-\frac{1}{n}\right)\cdot \lim_{\mu_{n,\theta}\uparrow\mathbb{R}} n(|\theta|+1) \\
&= \left(b-a-\frac{1}{\alpha}\right)\mathfrak{n}([0,1)_{\mathbb{R}}).
\end{aligned}$$

Thus eq. (29) follows with

$$|\varepsilon| \leq \frac{2}{\alpha}.$$

□

So, if we have two intervals $[a_0,b_0)_{\mathbb{R}} \subset [a_1,b_1)_{\mathbb{R}}$, using the above proposition, we get eq. (27). Moreover, it is easy to prove that the NAP-space produced by $(\mathbb{R},1,I_{\Lambda_{\mathbb{R}}})$ is consistent with $\left(\mathbb{Q},1,I_{\Lambda_{\mathbb{Q}}}\right)$, namely that the analog of eq. (26) holds: if $A\subset\mathbb{Q}$, then

$$P_{\mathbb{Q}}(A) = P_{\mathbb{R}}(A \mid \mathbb{Q}).$$

5.5. The infinite sequence of coin tosses

Let us consider an infinite sequence of tosses with a fair coin.[10] In the Kolmogorovian framework, the infinite sequence of fair coin tosses is modeled by the triple $(\Omega,\mathfrak{A},\mu)$, where $\Omega=\{H,T\}^{\mathbb{N}}$ is the space of sequences which take values in the set $\{H,T\}$ namely *Heads* and *Tails*. We will denote by $\omega=(\omega_1,\ldots,\omega_n,\ldots)$ the generic sequence.

$\mathfrak{A}$ is the σ-algebra generated by the 'cylindrical sets'. A cylindrical set of codimension n is defined by a n-ple of indices $(i_1,\ldots,i_n)$ and an n-ple of elements in $\{H,T\}$, namely $(t_1,\ldots,t_n)$ where t_k is either H or T.

A cylindrical set of codimension n is defined as follows:

$$C_{(t_1,\ldots,t_n)}^{(i_1,\ldots,i_n)} = \{\omega\in\Omega \mid \omega_{i_k}=t_k\}.$$

From the probabilistic point of view, $C_{(t_1,\ldots,t_n)}^{(i_1,\ldots,i_n)}$ represents the event that that i_k-th coin toss gives t_k for $k=1,\ldots,n$.

[10] This example is used by Williamson in an attempt to refute the possibility of assigning infinitesimal probability values to a particular outcome of such a sequence [24]. As observed by Weintraub, Williamson's argument relies on a relabeling of the individual tosses, which is not compatible with non-Archimedean probabilities [22].

The probability measure on the generic cylindrical set is given by

$$\mu\left(C_{(t_1,\dots,t_n)}^{(i_1,\dots,i_n)}\right) = 2^{-n}. \tag{31}$$

The measure μ can be extended in a unique way to P_{KA} on the algebra $\mathfrak{A}$ (by Carathéodory's theorem).

In this particular model, you can see the problems with the Kolmogorovian approach which we discussed in section 2.2:

- every event $\{\omega\} \in \Omega$ has 0 probability but the union of all these 'seemingly impossible' events has probability 1;
- if F is a finite set and $\{\omega\} \subset F$, then the conditional probability $P_{KA}(\{\omega\} \mid F)$ is not defined; nevertheless, you know that the conditionalizing event is not the empty event, so the conditional probability should be defined: it makes sense and its value is $\frac{1}{|F|}$;
- there are subsets of Ω for which the probability is not defined (namely the non-measurable sets).

Now, we will construct a NAP-space so that we can compare the two different approaches. We need to construct a NAP-function P produced by (Ω, w, I_Λ) which satisfies the following assumptions:

(i) if $F \subset \Omega$ is a finite non-empty set, then

$$P(A \mid F) = \frac{|A \cap F|}{|F|};$$

(ii) eq. (31) holds for P, namely

$$P\left(C_{(t_1,\dots,t_n)}^{(i_1,\dots,i_n)}\right) = 2^{-n}. \tag{32}$$

Experimentally, we can only observe a finite numbers of outcomes: both cylindrical events (*cf.* (ii)) and finite conditional probability (*cf.* (i)) are based on a finite number of observations. In some sense, (i) and (ii) are the 'experimental data' on which to construct the model.

Property (i) implies that we get a fair probability, thus we have to take $w \equiv 1$; so every infinite sequence of coin tosses in $\Omega = \{H, T\}^{\mathbb{N}}$ has probability $1/\mathfrak{n}(\Omega)$. Property (ii) is the analog of eq. (28) in the case of a fair lottery on $\mathbb{R}$.[11] So, we have to choose Λ in such a way that eq. (32) holds.

To do this, we need some other notation; if $b = (b_1, \dots, b_n) \in \{H, T\}^n$ is a finite string and $c = (c_1, \dots, c_n, \dots) \in \{H, T\}^{\mathbb{N}}$ is an infinite sequence, then we set

$$b \circledast c = (b_1, \dots, b_n, c_1, \dots, c_k, \dots)$$

namely, the sequence $b \circledast c$ is obtained by the sequence b followed by the infinite sequence c. Now, if $\sigma \in \mathcal{P}_{fin}^0\left(\{H, T\}^{\mathbb{N}}\right)$ and $n \in \mathbb{N}$, we set

$$\lambda_{n,\sigma} = \{b \circledast c \mid b \in \{H, T\}^n \text{ and } c \in \sigma\}$$

[11] Eq. (32) implies that for any μ-measurable set E, we have that $P(E) \sim \mu(E)$ and this relation is the analog of eq. (27).

and

$$\Lambda_{CT} := \left\{ \lambda_{n,\sigma} \mid \sigma \in \mathcal{P}^0_{fin}\left(\{H,T\}^{\mathbb{N}}\right) \text{ and } n \in \mathbb{N} \right\}.$$

Notice that $\left(\{H,T\}^{\mathbb{N}}, 1, I_{\Lambda_{CT}}\right)$ produces a well-defined NAP-space since Λ_{CT} is a directed set; in fact

$$\lambda_{n_1,\sigma_1} \cup \lambda_{n_2,\sigma_2} \subset \lambda_{\max(n_1,n_2),\sigma_3}$$

for a suitable choice of $\sigma_3 \in \mathcal{P}^0_{fin}\left(\{H,T\}^{\mathbb{N}}\right)$.

Moreover Λ_{CT} is the 'wise choice' as the following theorem shows:

Theorem 28. *If P is the NAP-function produced by $\left(\{H,T\}^{\mathbb{N}}, 1, I_{\Lambda_{CT}}\right)$, then eq.* (32) *holds.*

Proof. We have

$$\mathfrak{n}(\Omega) = \lim_{\lambda_{N,\sigma} \in \Lambda_{CT}} |\lambda_{N,\sigma}| = \lim_{\lambda_{N,\sigma} \in \Lambda_{CT}} \left(2^N \cdot |\sigma|\right). \tag{33}$$

Now consider the cylinder $C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)}$ and take $N = \max\ i_n$. Then, for every σ, we have that

$$\lambda_{N,\sigma} \cap C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)} = \{\omega \in \lambda_{N,\sigma} \mid \omega_{i_k} = t_k\}.$$

Then

$$\left|\lambda_{N,\sigma} \cap C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)}\right| = \frac{|\lambda_{N,\sigma}|}{2^n} = 2^{N-n} \cdot |\sigma|$$

and hence, by eq. (33),

$$\begin{aligned}\mathfrak{n}\left(C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)}\right) &= \lim_{\lambda_{N,\sigma} \in \Lambda_{CT}} \left|\lambda_{N,\sigma} \cap C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)}\right| = \lim_{\lambda_{N,\sigma} \in \Lambda_{CT}} \left(2^{N-n} \cdot |\sigma|\right) \\ &= 2^{-n} \lim_{\lambda_{N,\sigma} \in \Lambda_{CT}} \left(2^N \cdot |\sigma|\right) = 2^{-n} \cdot \mathfrak{n}(\Omega).\end{aligned}$$

Concluding, we have that

$$P\left(C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)}\right) = \frac{\mathfrak{n}\left(C^{(i_1,\ldots,i_n)}_{(t_1,\ldots,t_n)}\right)}{\mathfrak{n}(\Omega)} = 2^{-n}. \qquad \square$$

References

[1] P. Bartha and C. Hitchcock, *The shooting-room paradox and conditionalizing on measurably challenged sets*, Synthese **118** (1999), 403–437.

[2] V. Benci, *I numeri e gli insiemi etichettati*, Conferenze del seminario di matematica dell'Universita' di Bari, vol. 261, Laterza, Bari, Italy, 1995, p. 29.

[3] V. Benci and M. Di Nasso, *Alpha-theory: an elementary axiomatic for nonstandard analysis*, Expositiones Mathematicae **21** (2003), 355–386.

[4] ———, *Numerosities of labelled sets: a new way of counting*, Advances in Mathematics **173** (2003), 50–67.

[5] V. Benci, M. Di Nasso, and M. Forti, *An Aristotelian notion of size*, Annals of Pure and Applied Logic **143** (2006), 43–53.

[6] ———, *The eightfold path to nonstandard analysis*, Nonstandard Methods and Applications in Mathematics (N. J. Cutland, M. Di Nasso, and D. A. Ross, eds.), Lecture Notes in Logic, vol. 25, Association for Symbolic Logic, AK Peters, Wellesley, MA, 2006, pp. 3–44.

[7] V. Benci, L. Horsten, and S. Wenmackers, *Infinitesimal probabilities*, In preparation, 2012.

[8] E. Bottazzi, *Ω-Theory: Mathematics with Infinite and Infinitesimal Numbers*, Master thesis, University of Pavia, Italy, 2012.

[9] N. Cutland, *Nonstandard measure theory and its applications*, Bulletin of the London Mathematical Society **15** (1983), 529–589.

[10] B. de Finetti, *Theory of probability*, Wiley, London, UK, 1974, Translated by: A. Machí and A. Smith.

[11] T. Gilbert and N. Rouche, *Y a-t-il vraiment autant de nombres pairs que de naturels?*, Méthodes et Analyse Non Standard (A. Pétry, ed.), Cahiers du Centre de Logique, vol. 9, Bruylant-Academia, Louvain-la-Neuve, Belgium, 1996, pp. 99–139.

[12] A. Hájek, *What conditional probability could not be*, Synthese **137** (2003), 273–323.

[13] J.B. Kadane, M.J. Schervish, and T. Seidenfeld, *Statistical implications of finitely additive probability*, Bayesian Inference and Decision Techniques (P.K. Goel and A. Zellnder, eds.), Elsevier, Amsterdam, The Netherlands, 1986.

[14] H.J. Keisler, *Foundations of infinitesimal calculus*, University of Wisconsin, Madison, WI, 2011.

[15] H.J. Keisler and S. Fajardo, *Model theory of stochastic processes*, Lecture Notes in Logic, Association for Symbolic Logic, 2002.

[16] A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitrechnung*, Ergebnisse der Mathematik, 1933, Translated by N. Morrison, *Foundations of probability*. Chelsea Publishing Company, 1956 (2nd ed.).

[17] I. Levi, *Coherence, regularity and conditional probability*, Theory and Decision **9** (1978), 1–15.

[18] D. K. Lewis, *A subjectivist's guide to objective chance*, Studies in Inductive Logic and Probability (R. C. Jeffrey, ed.), University of California Press, Berkeley, CA, 1980, pp. 263–293.

[19] P.A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Transactions of the American Mathematical Society **211** (1975), 113–122.

[20] E. Nelson, *Radically elementary probability theory*, Princeton University Press, Princeton, NJ, 1987.

[21] B. Skyrms, *Causal necessity*, Yale University Press, New Haven, CT, 1980.

[22] R. Weintraub, *How probable is an infinite sequence of heads? A reply to Williamson*, Analysis **68** (2008), 247–250.

[23] S. Wenmackers and L. Horsten, *Fair infinite lotteries*, Accepted in *Synthese*, DOI: 10.1007/s11229-010-9836-x, 2010.

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

Vieri Benci
Dipartimento di Matematica Applicata
Università degli Studi di Pisa
Via F. Buonarroti 1/c
56127 Pisa
Italy
and
Department of Mathematics
College of Science
King Saud University
11451 Riyadh
Saudi Arabia
e-mail: benci@dma.unipi.it

Leon Horsten
Department of Philosophy
University of Bristol
43 Woodland Rd
BS81UU Bristol
United Kingdom
e-mail: leon.horsten@bristol.ac.uk

Sylvia Wenmackers
Faculty of Philosophy
University of Groningen
Oude Boteringestraat 52
9712 GL Groningen
The Netherlands
e-mail: s.wenmackers@rug.nl

Received: May 18, 2012.