



Sets and Probability

Leon Horsten¹ · Hazel Brickhill²

Received: 23 April 2020 / Accepted: 9 February 2023
© The Author(s) 2023

Abstract

In this article the idea of random variables over the set theoretic universe is investigated. We explore what it can mean for a random set to have a specific probability of belonging to an antecedently given class of sets, or, in other words, to have a specific probability of having a given set-theoretic property.

1 Introduction

Probabilistic notions have been applied to mathematical objects and concepts. Probabilistic concepts have been applied in the theory of random graphs, random closed sets, random Banach spaces, random series, random fields, random vectors, etcetera. The aim of the present article is to apply a notion of probability not to any specific restricted class of mathematical objects, but to the mathematical universe as a whole. More in particular, we wish to explicate *what it could mean for a property A of sets to have a probability of being true of a set y in the set theoretic universe V* . Properties are identified with their extensions, so that A ranges over all proper and improper classes in V .

The aim is to develop a theory of the probability of events of the form $A(\tau)$, where A is a class and the variable τ is a *random variable*. Intuitively, then, ‘ $\Pr(A(\tau))$ ’ refers to the probability that the random variable τ takes a value that has the property A .

The authors are grateful to Philip Welch for helpful suggestions and conversations.

Funding Open Access funding enabled and organized by Projekt DEAL.

✉ Leon Horsten
Leon.Horsten@uni-konstanz.de
Hazel Brickhill
Hazel.Brickhill@uni-konstanz.de

¹ Postfach 21 Fachbereich Philosophie, Universität Konstanz, 78457 Konstanz, Deutschland

² Fachbereich Mathematik und Statistik, Universität Konstanz, 78457 Konstanz, Deutschland

Formally, a random variable is a function from an outcome space (also called a state space) Ω to another space E , which is often a measurable space (Blitzstein and Hwang 2015, p. 92). The outcome space Ω is associated with a probability function. The outcome space Ω represents the states that the random variable can be in; the space E represents the values associated with the states that one is interested in.

A familiar simple example is the throwing of a fair coin, where the outcome space is {state of Heads, state of Tails}, the space it maps into is $\{0, 1\}$, and the associated probability function is such that the probability of the coin landing Heads is the same as it landing Tails. In this situation (as in many others), the outcome space can be *identified* with the measurable space. This is also true for the case that we are interested in, where the state space must be large enough to have different states associated with the random variables taking different sets as value, but need not be larger. So we will take the outcome space Ω to be V , and we will take the space E that it maps into also to be V .

Without invoking fixed sets of postulates, intuitions about probability have occasionally been used in set theory, for instance to motivate new basic principles (Freiling, 1986).¹ The article (Freiling, 1986) also provides an instance of an application of the notion of *random variable* to a class of mathematical entities (at the real numbers of the unit interval); so does (Scott, 1967) (random reals). In the light of all this, it is natural to wonder what we might require from probability functions associated with random variables *on* V .

Surely it would be unreasonable to insist on there being one *uniquely correct* probability function that yields the probability of a random variable taking a value in a given class of sets. At any rate, finding such a “uniquely correct” or “objective” probability function is not widely regarded as a viable research objective. On the other hand, for our functions to have any hope of meriting the label *probability function*, the usual rules for finitely manipulating probabilities must apply. Thus they have to satisfy the calculation rules for *finitely additive probability functions*. So all of the probability functions that we shall consider are *finitely additive probability functions* in the sense of Kolmogorov, except for his stipulation that probabilities are measured by elements of the real $[0, 1]$ interval.

From the outset we impose three additional constraints on the class of probability functions that we are interested in:²

1. *Totality*. The probability functions are defined on all classes.
2. *Uniformity*. All singleton events are given the same probability.
3. *Regularity*. All singleton events are given non-zero probability.

For familiar reasons these extra constraints are not compatible of infinite sample spaces *if* the values of probabilities lie in the real $[0, 1]$ interval. Consider the fair lottery on \mathbb{N} . By the extra constraints, there is some $r \in]0, 1[$ such that for every ticket n , the probability of n being the winning ticket is r . But then, by Finite

¹ Such attempts are mostly regarded as unsuccessful (Hamkins, 2015).

² For a discussion of the first two of these constraints in the context of non-Archimedean probability theory, see (Benci et al., 2018); for a discussion of the third constraint, see (Wenmackers & Horsten, 2013).

Additivity and the Archimedean property of \mathbb{R} , there is a natural number k such that for every collection S of k tickets, the probability of the winning ticket belonging to S is greater than 1. This means that the desired probability functions will have been *non-Archimedean*.

As will be explained in detail in Sect. 2, unlike Kolmogorov probability functions, these non-Archimedean probability functions depend on a choice of ultrafilter on the “small” subsets of the sample space.³ The resulting probability functions will also turn out not satisfy σ -additivity. Nonetheless we will see that they instead satisfy a generalised infinite additivity rule.⁴

Totality means that there are no *non-measurable classes*. This makes the probabilities that we will consider in a sense maximally informative, which we take to be a desirable feature. Regularity means that our probability functions will be very *fine-grained*, since they will *always* distinguish between impossible and contingent events. We take this, too, to be a desirable feature.

Lastly, we turn to Uniformity. (Freiling, 1986) considers the abstract possibility of *randomly* throwing darts at the real number line. In this paper, we consider the even more abstract possibility of randomly throwing darts at the set theoretic universe. When considering a dart randomly thrown at \mathbb{R} , it would be unnatural to consider probability functions that take the probability of the dart landing on the number π^{-1} to be 0.9, say. Indeed, if the event of the randomly thrown dart landing on a given real number receives a probability at all (which it will, in the setting that we shall be considering), then the randomness assumption causes us to expect the probability of it landing on one real number to be the same as the probability of it landing on some other real number.⁵ In the same vein, in our setting the Uniformity property says that the probability of one given set being randomly picked from the set theoretic universe is the same as the probability that some other set is randomly picked. In view of this, we take Uniformity to be desirable for the kinds of abstract processes that we want to model.

Uniformity is commonly seen as a *symmetry property* for probability functions.⁶ Pruss rightly stresses, however, that Uniformity does not even come close to exhausting the content of the condition fairness and symmetry for probabilistic scenarios: “A lottery with our radically skewed probabilities that nonetheless treat[s] all individual integers as equiprobable does not intuitively appear to be fair. Thus, [even] weak translation invariance plus equiprobability of singletons does not appear to be sufficient to capture our intuitions of fairness and symmetry” (Pruss 2021, p. 8521).

In this article, we investigate *additional constraints* that a finitely additive, total, uniform, and regular probability function might satisfy. In particular, we will consider extra requirements that, *given* Totality and Regularity and Uniformity, from a

³ This sensitivity of the properties of such probability functions to choice of ultrafilter has been the subject of discussion in the literature: see for instance (Kremer, 2014).

⁴ This generalised infinite additivity rule also follows from (Benci et al 2013, Sect. 3.4, Proposition 8.)

⁵ Similarly, the attempt to model a fair lottery on the natural numbers motivates an assumption of uniformity in (Wenmackers & Horsten, 2013) .

⁶ One (of many) example(s) that can be adduced to illustrate this is (North, 2010).

pre-theoretic perspective represent *possible* or even natural probabilistic scenarios.⁷ For instance, we may ask if such a probability function might in addition satisfy the principle that if any two sets A and B have the same cardinality, then a random variable on V will be as likely to take a value in A as in B ; or one may ask whether such a probability function might in addition satisfy the principle that the probability of a random variable taking the value of an ordinal is infinitesimally close to 0. We will see that some such additional requirements are jointly satisfiable, others not.

The arguments that establish the satisfiability of combinations of such additional constraints turn on a careful selection of ultrafilters on “small” subsets of the state space. A key objective of the present article is to explore the sensitivity of properties of random variables on V on the choice of ultrafilter. We aim to chart, to some extent, the relationship between and trade-offs that come with various technical choices in the definition of random variables on the set theoretic universe. This strikes us as a worthy foundational enterprise.

The project in which we are engaging in this article is related to the work in (Benci et al., 2007), (Benci et al., 2006). The aim of these articles is to construct a theory of sizes for mathematical universes inspired by the *Euclidean principle* that the size of the whole is larger than the sizes of its proper parts. Now there is of course a familiar theory of size—Cantor’s theory of *cardinality*,—which does *not* satisfy this Euclidean principle. But Benci and his co-authors prefer the Euclidean theory. In the abstract of (Benci et al., 2006), they write that they maintain the Euclidean principle in exchange for giving up half of Cantor’s equinumerosity principle for sets (Benci et al 2006, p. 43).⁸ So Benci and his co-authors propose their Euclidean theory of size as a rival to Cantor’s theory.

There is a close relationship between the mathematical techniques that are used in the present article and the mathematical techniques that are used in (Benci et al., 2007), (Benci et al., 2006). But ideologically, our standpoints are very dissimilar. Whereas Benci and his co-authors from (Benci et al., 2007) and (Benci et al., 2006) reject Cantor’s theory of cardinality in favour of the Euclidean principle, we fully embrace it. Nonetheless, the probability functions that will be constructed satisfy the *probabilistic* Euclidean principle that the probability of an event is *strictly greater* than the probability of each of its proper sub-events. This is in contrast with Kolmogorov probability, where it frequently happens that the probability of an event is *equally great* as the probability of one or more of its proper sub-events.

Moreover, what we shall mean by ‘mathematical universe’ is not the same as what is meant in (Benci et al., 2007) by the term. The authors of (Benci et al., 2007) impose mainly algebraic constraints on what counts as a mathematical universe (Benci et al 2007, Introduction). We, in contrast, take the term ‘mathematical universe’ in the set theoretical sense: *the* mathematical universe is the arena in which all of mathematics can in principle be carried out. Naively, you may take there to be one preferred set theoretic universe: V . But if you are uncomfortable with taking

⁷ Benci et al call the *desideratum* that a probability theory allows as many seemingly coherent probabilistic scenarios the requirement of *weak Laplacianism*: see (Benci et al 2018, p. 510–511).

⁸ This does not prevent them from making use of Cantor’s theory of cardinals in their articles, as an anonymous referee rightly pointed out to us.

V as given, then you might want to take a mathematical universe to be a rank V_α that constitutes a model of most or perhaps even all of the standard principles of set theory, such as some strongly inaccessible rank. Indeed, we will see that for random variables defined on *any* large set S , the general idea of equipping them with a probability function will be the same as that for random variables on V .

We will discuss two ways of generating non-Archimedean probability functions for random variables on V . In Sect. 2 a *simple* way of generating such probability functions (the *finite snapshot approach*) will be described. In Sect. 3 we go on to discuss how global properties of these probability functions can be made to hold by imposing constraints on the process of generating such functions (choice of ultrafilters). In Sect. 4, a theoretically more satisfying but also more complicated way of generating non-Archimedean probability functions for random variables on V is discussed (the *bootstrapping* method).

2 The Finite Snapshot Approach

A random variable τ on V is a function from states to the outcome space, i.e., an element of ${}^V V$. So there are *many* random variables on V . Our aim is to associate a notion of probability with elements of ${}^V V$ that meet the minimal constraints (Totality, Uniformity and Regularity) that were described in Sect. 1.

In fact, we want to give precise meaning to conditional probability statements of the form

$$\Pr(\sigma \in A \mid \tau \in B),$$

where $\sigma, \tau \in {}^V V$ and $A, B \subseteq V$. We will see that it will in fact be sufficient for our purposes to give meaning to *unconditional* probability statements of the form $\Pr(\sigma \in A)$. Given that random variables are elements of ${}^V V$, it is clear that the symbol ‘ \in ’ in ‘ $\Pr(\sigma \in A)$ ’ does not have its usual literal meaning of membership (in a class), as we will see shortly. Informally, ‘ $\Pr(\sigma \in A)$ ’ should be read as: ‘the probability that the random variable σ is in a state where it has as its value some element of A ’.

Our fundamental problem amounts to giving meaning to expressions of the form $\Pr(\sigma \in A)$. Such probability measures will be determined by a choice of a fine ultrafilter⁹ on the collection $[V]^{<\omega}$ of *finite subsets* of the state space.¹⁰

The starting point is a fine ultrafilter \mathcal{U} on $[V]^{<\omega}$. This fine ultrafilter \mathcal{U} defines a non-Archimedean field $\mathcal{F}_\mathcal{U}$ in the following way.

For any two functions $f, g : [V]^{<\omega} \rightarrow \mathbb{Q}$ we define:

⁹ A *filter* on a set S is a non-empty collection F of subsets of S that is closed under finite intersections and arbitrary supersets. \mathcal{U} is an *ultrafilter* on a set S if \mathcal{U} is a filter on S such that for every $A \subseteq S$, either $A \in \mathcal{U}$, or $S \setminus A \in \mathcal{U}$. An ultrafilter \mathcal{U} on a set $[S]^{<\omega}$ (i.e. the collection of all finite subsets of S) is *fine* if for every $a \in S$, the collection A of all finite subsets of S that include a as an element, belongs to \mathcal{U} .

¹⁰ What follows is an adaptation of the approach of (Brickhill et al 2018, Sect. 2). The name ‘finite snapshot approach’ is drawn from a passage in that article: “ $\Pr(A)$ is conceived of as the limit of the relative frequency of A ’s on *finite snapshots* of the sample space” (Brickhill 2018, p. 524).

Definition 1

$$f \approx_{\mathcal{U}} g \leftrightarrow \{T \in [V]^{<\omega} : f(T) = g(T)\} \in \mathcal{U}.$$

In words: two functions are equivalent under \mathcal{U} if they coincide on ultrafilter-many states.

The relation $\approx_{\mathcal{U}}$ is an equivalence relation, so we can take equivalence classes for which we then have

$$[f]_{\mathcal{U}} = [g]_{\mathcal{U}} \Leftrightarrow f \approx_{\mathcal{U}} g.$$

Moreover, it is again a routine exercise to verify that the $[f]_{\mathcal{U}}$'s form a hyper-rational field $\mathcal{F}_{\mathcal{U}}$.

Now suppose $A \subseteq V$ and $\theta \in {}^V V$. Then we define the function $f_{\theta \in A} : [V]^{<\omega} \rightarrow \mathbb{Q}$ as follows:

Definition 2 For every non-empty $T \in [V]^{<\omega}$:

$$f_{\theta \in A}(T) \equiv \frac{|\{s \in T : \theta(s) \in A\}|}{|T|}.$$

In words: for every *finite* set of states T , $f_{\theta \in A}(T)$ is the ratio between the number of states s in T for which $\theta(s) \in A$ and the number of states in T . In this sense, $f_{\theta \in A}(T)$ is the probability of $\theta \in A$ on a *finite snapshot of states*, i.e., $\Pr(\theta \in A \mid \iota \in T)$, where ι is the identity random variable and the probability is given by the ratio formula.

Similarly, we define the function $f_{\theta \in A \wedge \nu \in B}$ as follows:

Definition 3 For every non-empty $T \in [V]^{<\omega}$:

$$f_{\theta \in A \wedge \nu \in B}(T) \equiv \frac{|\{s \in T : \theta(s) \in A \text{ and } \nu(s) \in B\}|}{|T|}.$$

Now we are ready to define the probability of $\theta \in A$, relative to a fine (and therefore free¹¹) ultrafilter \mathcal{U} on $[V]^{<\omega}$:

Definition 4

$$\Pr_{\mathcal{U}}(\theta \in A) \equiv [f_{\theta \in A}]_{\mathcal{U}}.$$

Similarly, we define $\Pr_{\mathcal{U}}(\theta \in A \wedge \nu \in B)$ as $[f_{\theta \in A \wedge \nu \in B}]_{\mathcal{U}}$. Thus we have constructed a probability function $\Pr_{\mathcal{U}}$ that takes its values in the *hyperrational field* $\mathcal{F}_{\mathcal{U}}$, i.e., $\mathcal{F}_{\mathcal{U}}$ contains infinitesimal elements. Such probability functions are sometimes called NAP functions (**N**on-**A**rchimedean **P**robability functions).

Conditional probability can then be expressed in terms of unconditional probability:

$$\Pr_{\mathcal{U}}(\theta \in A \mid \nu \in B) \equiv \frac{c_{\mathcal{U}}(\theta \in A \wedge \nu \in B)}{\Pr_{\mathcal{U}}(\nu \in B)}.$$

Definition 5

Thus we have given a recipe for the construction of probability measures $\Pr_{\mathcal{U}}$ on V that is *mathematically coherent*. Nonetheless, since ultrafilters on $[V]^{<\omega}$ are

¹¹ An ultrafilter \mathcal{U} on a set S is *free* if the intersection of all elements of \mathcal{U} is empty.

hyperclasses (i.e., entities that contain proper classes as elements), it follows from definition 4 that the resulting probability measures are hyperclasses, too.¹² It is of course a difficult *philosophical* question whether classes and/or hyperclasses exist, and, if they do, what their nature is. We will not go into this question here, but merely reiterate our earlier observation¹³ that those who are sceptical about classes may take V to be some strongly inaccessible rank.

3 Constraints

From Sect. 1 we know that the aim is not to arrive at a *unique* (correct) probability function on V . But we did insist from the outset on our probability functions satisfying three global constraints: totality, uniformity, and regularity. It will be shown that these properties are always guaranteed to hold.

There are further global conditions on probability functions on V that seem from a pre-theoretic point of view attractive, and that are not guaranteed to hold without further work. These global constraints will be explored in what follows. We will show that many of them can be forced to hold by imposing constraints on the ultrafilters from which the probability functions are generated.

3.1 Elementary Properties

The definition of $\Pr_{\mathcal{U}}$ is relative to an initial choice of the fine ultrafilter \mathcal{U} . The properties of $\Pr_{\mathcal{U}}$ depend on \mathcal{U} . Nonetheless, certain basic properties of $\Pr_{\mathcal{U}}$ can be easily seen to hold regardless of which fine ultrafilter \mathcal{U} is chosen. For instance, it is easy to see that $\Pr_{\mathcal{U}}$ is always a total finitely additive probability function (Benci et al 2013, Sect. 4).

Now we define the notion of a *bijective random variable*:

Definition 6 A random variable θ is said to be a *bijective random variable* if for any set x , there is exactly one element u of the state space such that $\theta(u) = x$.

In words: a bijective random variable is a random variable that takes every value exactly once. This simply means that bijective random variables have no built-in bias towards taking as their value any particular set.

In terms of the notion of random variable (on V), we define the notions of *regularity* and *uniformity*:

Definition 7 (regularity) A probability function $\Pr_{\mathcal{U}}$ is regular if for every bijective random variable θ and for every $x \in V$, $\Pr_{\mathcal{U}}(\theta = x) > 0$.

¹² It is unavoidable that probability measures on V of this kind are larger than set-size: see (Pruss, 2013).

¹³ See p. 4.

Definition 8 (*uniformity*) A probability function $\Pr_{\mathcal{U}}$ is uniform if for every bijective random variable θ and for all $x, y \in V$:

$$\Pr_{\mathcal{U}}(\theta = x) = \Pr_{\mathcal{U}}(\theta = y).$$

Proposition 1 For every fine ultrafilter \mathcal{U} :

1. $\Pr_{\mathcal{U}}$ is regular;
2. $\Pr_{\mathcal{U}}$ is uniform.

Proof These properties are proved as propositions 2.5 and 2.6 in (Brickhill 2018, p. 525–526), respectively. \square

Here we note that it is the fineness of the ultrafilter that guarantees the Regularity and Uniformity of the resulting probability function. So our requirement of fineness on ultrafilters is motivated by our desire to obtain Regularity and Uniformity.¹⁴

The Euclidean probability is formally defined as follows:

Definition 9 (*Euclidean*) A probability function $\Pr_{\mathcal{U}}$ is Euclidean if for every bijective random variable θ and all $A, B \subseteq V$:

$$A \subsetneq B \Rightarrow \Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B).$$

Then we have:

Proposition 2 For every fine ultrafilter \mathcal{U} , the probability function $\Pr_{\mathcal{U}}$ is Euclidean.

Proof By finite additivity and regularity. \square

Now we turn to infinite additivity. Countable additivity means that the probability of the union of a countable family of disjoint sets is the *infinite sum* of the probabilities of the elements of the family, where the notion of infinite sum is spelled out in terms of the classical notion of limit. In the present setting, the probability $\Pr_{\mathcal{U}}$ of the union of *any* family of disjoint sets is also the infinite sum of the probabilities of the elements of the family (Benci et al 2013, Sect. 3.4). But now the notion of infinite sum is spelled out in terms of the generalised notion of limit based on the ultrafilter \mathcal{U} . More precisely, the new notion of infinite sum is defined as follows. Suppose we are given a family $\{q_i : i \in S\}$ of rational numbers, and $I \subseteq S$. Then consider the function $f : [S]^{<\omega} \rightarrow \mathbb{Q}$ given by

$$f(T) = \sum_{i \in I \cap T} q_i.$$

¹⁴ The Uniformity property can be relaxed by assigning *weights* to states (see (Brickhill et al 2018, section 2.4). For certain purposes, such as establishing analogues of the representation theorems that are proved in (Brickhill & Horsten, 2018), this is useful and desirable. But, for the reasons given on p. 3, we will not consider such “non-random” probability functions here.

This function can be seen as giving the value of the infinite sum on all *finite parts* (“snapshots”) of the index set. So we identify the infinite sum of the family $\{q_i : i \in I\}$ of rational numbers with the generalised limit of f according to the ultrafilter \mathcal{U} :

$$\sum_{i \in I}^* q_i \equiv [f]_{\mathcal{U}}.$$

Definition 10

Using this notion of infinite sum,¹⁵ we can express the probability of the union of a disjoint family of sets as the sum of the probabilities of the members of that family:

Proposition 3 *For any index set I , if $A = \bigcup_{i \in I} A_i$, with $A_i \cap A_j = \emptyset$ for all $i, j \in I$, then for every random variable τ :*

$$\Pr_{\mathcal{U}}(\tau \in A) = \sum_{i \in I}^* \Pr_{\mathcal{U}}(\tau \in A_i).$$

In sum, $\Pr_{\mathcal{U}}$ has a natural infinite additivity property that is sometimes called *perfect additivity*.

Proposition 4 *For every fine ultrafilter \mathcal{U} , the probability function $\Pr_{\mathcal{U}}$ is perfectly additive.*

Proof This proposition is proved as proposition 8 in (Benci et al 2013, p. 132–133). □

3.2 Symmetry Principles

The Euclidean-ness of $\Pr_{\mathcal{U}}$ has implications for *symmetry principles*. We have already seen that our probability functions satisfy Uniformity, which is naturally regarded as a symmetry principle that holds for our probability functions. However, we will now recount how the Euclidean-ness of $\Pr_{\mathcal{U}}$ entails that certain other symmetry principles fail. The results below complement other results in the literature that indicate that in the presence of Uniformity, it is difficult to satisfy further symmetry principles.¹⁶

Proposition 5 *For every fine ultrafilter \mathcal{U} , the probability function $\Pr_{\mathcal{U}}$ is not invariant under all permutations of V .*

Proof We concentrate on \mathbb{N} as it is canonically represented in V (by means of the Zermelo ordinals, for instance). Define a permutation π of V as follows:

¹⁵ This infinite sum gives rise to a generalised notion of integral: see (Eskew, 2021).

¹⁶ See (Benci et al., 2007), (Benci et al., 2013), (Benci et al., 2018), (Pruss, 2021).

- $\pi(x) = x$ for $x \in V \setminus \mathbb{N}$; Otherwise:
- $\pi(x) = x + 2$ for x even;
- $\pi(1) = 0$;
- $\pi(x) = x - 2$ for x odd and > 1 .

Let $A \equiv \{0, 2, 4, \dots\}$, and let θ be a bijective random variable. Then $\pi(A) \subsetneq A$. Therefore, by the Euclidean principle, $\Pr_{\mathcal{U}}(\theta \in \pi(A)) < \Pr_{\mathcal{U}}(\theta \in A)$. \square

This of course entails that there are bijective random variables θ, θ' such that for some $A \subseteq V$,

$$\Pr_{\mathcal{U}}(\theta \in A) \neq \Pr_{\mathcal{U}}(\theta' \in A).$$

One popular global constraint on probability measures is *translation-invariance*. The Lebesgue measure has this property, and Banach limits seem to occupy a privileged position in the class of generalised limits at least in part because they are translation-invariant. In our context, translation-invariance does not make obvious sense. For a random class A , it is not clear what ' $A + \alpha$ ' (where α is a number) *means*. But a clear interpretation of 'adding an ordinal number' can of course be given if A is a collection of ordinals:

Definition 11 For A any collection of ordinals:

$$A \oplus \alpha \equiv \{\beta : \exists \gamma \in A \text{ such that } \beta = \gamma + \alpha\}.$$

Then for A to be translation-invariant means that for all ordinals α and for every θ ,

$$\Pr_{\mathcal{U}}(\theta \in A) = \Pr_{\mathcal{U}}(\theta \in A \oplus \alpha).$$

However, even if we consider non-Archimedean measures (of the kind that we have been describing) on ordinals, translation-invariance conflicts with the Euclidean Property of our generalised probability functions. In particular, there is no NAP probability function $\Pr_{\mathcal{U}}$ on any infinite cardinal κ such that there is even one ordinal α with $0 < \alpha < \kappa$ and

$$\Pr_{\mathcal{U}}(\theta \in \kappa) = \Pr_{\mathcal{U}}(\theta \in \kappa \oplus \alpha).$$

The reason is simple. We have $\kappa \oplus \alpha = \kappa \setminus \alpha \subsetneq \kappa$, so if we had $\Pr_{\mathcal{U}}(\theta \in \kappa) = \Pr_{\mathcal{U}}(\theta \in \kappa \oplus \alpha)$, then we would contradict the Euclidean principle.

As this example shows, such translations are not necessarily one to one so we may not want full invariance in general. In (Benci et al 2007, section 1.3), Benci, Forti, and Di Nasso explore a *restricted* notion of translation-invariance of NAP-like measures on ordinals. We do not pursue this theme further here, but only pause to note that there are other reasonable-looking principles that are hard to satisfy. In the context of their theory of *numerosities*, Benci, Forti, and Di Nasso consider a principle that in the present context would take the following form:

Definition 12 (*Difference Principle*)

$$\forall A, B \in V : \Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B) \Rightarrow \\ \exists C \in V : \Pr_{\mathcal{U}}(\theta \in B) = \Pr_{\mathcal{U}}(\theta \in A) + \Pr_{\mathcal{U}}(\theta \in C).$$

On countable sample spaces, the difference principle can be made to hold by building $\Pr_{\mathcal{U}}$ from a *selective* ultrafilter (Benci & Di Nasso, 2003).¹⁷ But the existence of selective ultrafilters is independent of ZFC.¹⁸ As far as we know, it is an open question whether the difference principle can be consistently made to hold for NAP probability functions on uncountable sample spaces.

3.3 Probability and Cardinality

In this subsection we investigate the relation between our notion of generalised probability on the one hand, and the familiar notion of cardinality on the other hand.

3.3.1 Hume’s Principle for Probability

One might naively wonder whether the following probabilistic analogue of Hume’s Principle for cardinality can hold:

Definition 13 (*Hume’s principle for probability*) For all $A, B \in V$:

$$|A| = |B| \Rightarrow \Pr_{\mathcal{U}}(\tau \in A) = \Pr_{\mathcal{U}}(\tau \in B).$$

But the probability functions $\Pr_{\mathcal{U}}$ that we have been considering cannot satisfy Hume’s principle for probability, as its failure is an immediate consequence of Proposition 5: invariance under permutations and Hume’s principle for probability are mathematically equivalent. However, this was only to be expected. After all, we do not expect Kolmogorov probability (on infinite spaces) to satisfy any such principle.

3.3.2 Superregularity

The hyper-rational field $\mathcal{F}_{\mathcal{U}}$ in which the probability functions $\Pr_{\mathcal{U}}$ take their values contain *infinitesimal numbers*—this is what makes it non-Archimedean. We will write $\Pr_{\mathcal{U}}(\sigma \in A) \approx 0$ if $\Pr_{\mathcal{U}}(\sigma \in A) < n^{-1}$ for each $n \in \mathbb{N}$. And we will write $\Pr_{\mathcal{U}}(\sigma \in A) \ll \Pr_{\mathcal{U}}(\tau \in B)$ if

$$\frac{\Pr_{\mathcal{U}}(\sigma \in A)}{\Pr_{\mathcal{U}}(\tau \in B)} \approx 0.$$

We have seen that $\Pr_{\mathcal{U}}$ cannot satisfy Hume’s principle for probability. But, at least at first sight, it seems that it would be reasonable to ask for:

¹⁷ An ultrafilter \mathcal{U} on \mathbb{N} is *selective* if for every partition of $[\mathbb{N}]^2$ into two pieces, there is a set $H \in \mathcal{U}$ homogeneous for the partition.

¹⁸ Their existence can be proved from the Continuum Hypothesis: see (Booth, 1970).

$$|A| < |B| \Rightarrow \Pr_{\mathcal{U}}(\delta \in A) < \Pr_{\mathcal{U}}(\delta \in B).$$

Indeed, if in addition $|B| \geq \omega$, then we might even seek to demand

$$|A| < |B| \Rightarrow \Pr_{\mathcal{U}}(\sigma \in A) \ll \Pr_{\mathcal{U}}(\sigma \in B).$$

Further, this may be expected to hold if B is a proper class but A is a set. The result is a size constraint which is a strengthening of the requirement of regularity:

Definition 14 (Superregularity)

$$\omega \leq |A| < |B| \leq |V| \Rightarrow \Pr_{\mathcal{U}}(\theta \in A) \ll \Pr_{\mathcal{U}}(\theta \in B).$$

Note that if A is finite and B is infinite then the consequent holds automatically.

By a suitable restriction on admissible ultrafilters \mathcal{U} , superregularity can indeed be made to hold:

Theorem 1 *There are fine ultrafilters \mathcal{U} such that $\Pr_{\mathcal{U}}$ is superregular.*

Proof If $A, B \in V$ such that $\omega \leq |A| < |B|$ are given, then we have $\Pr_{\mathcal{U}}(\theta \in A) \ll \Pr_{\mathcal{U}}(\theta \in B)$ if and only if for each $n \in \mathbb{N}$,

$$\{D \in [V]^{<\omega} : \frac{\Pr(\theta \in A \mid \theta \in D)}{\Pr(\theta \in B \mid \theta \in D)} \leq n^{-1}\} \in \mathcal{U}.$$

The aim is to build an ultrafilter \mathcal{U} for which this holds.

For any $n \in \mathbb{N}$, define

$$C_{AB}^n \equiv \{D \in [V]^{<\omega} : \frac{\Pr(\theta \in A \mid \theta \in D)}{\Pr(\theta \in B \mid \theta \in D)} \leq n^{-1}\}.$$

Moreover, let

$$A_x \equiv \{D \in [V]^{<\omega} : x \in D\}.$$

Define also

$$\mathcal{F} \equiv \{C_{AB}^n : n \in \mathbb{N}, |A| < |B|\} \cup \{A_x : x \in V\}.$$

We want to prove that \mathcal{F} has the finite intersection property. Therefore take any $x_1, \dots, x_k \in V$, and any $\langle A_1, B_1, n_1 \rangle, \dots, \langle A_l, B_l, n_l \rangle$ such that $|A_j| < |B_j|$ and $n_j \in \mathbb{N}$ for $j \leq l$. Assume for the construction that $|A_1| \leq |A_2| \leq \dots \leq |A_l|$. For every finite D , if $\{x_1, \dots, x_k\} \subseteq D$, then $D \in \bigcap_{i \leq k} A_{x_i}$. So setting $n = \max\{n_j : j < l\}$ we will extend $\{x_1, \dots, x_k\}$ to a set in $C_{A_j B_j}^n$, and hence $C_{A_j B_j}^{n_j}$, for each $j \leq l$. Set $F_0 = \{x_1, \dots, x_k\}$ and $a_0 = |F_0 \cap A_1|$. As B_1 is infinite and of larger cardinality than A_1 we add $n \cdot a_0$ elements of $B_1 \setminus A_1$ to F_0 , yielding a finite set F_1 . Now set $a_1 = |F_1 \cap A_2|$, and add $n \cdot a_1$ elements of $B_2 \setminus (A_1 \cup A_2)$ to F_1 to give F_2 . Note we can find these elements of B_2 as $|B_2| > |A_2| \geq |A_1|$. Continuing in this manner, set $F = F_l$. Then we have ensured that for all $j \leq l$

$$\frac{\Pr(\theta \in A_j \mid \theta \in F)}{\Pr(\theta \in B_j \mid \theta \in F)} \leq n^{-1},$$

and so we have $F \in C_{A_j B_j}^n$, and since $D \subseteq F$, we also have $F \in \bigcap_{i \leq k} A_{x_i}$.

So \mathcal{F} indeed has the finite intersection property, whereby it can be extended to a filter and then further to an ultrafilter \mathcal{U} . By design, then, the resulting probability function $\Pr_{\mathcal{U}}$ is super-regular. □

Once again, Hume’s Principle for probability cannot hold for the notion of probability that we are investigating. But this leaves open the question whether the *converse* of Hume’s Principle for probability can be made to hold. This is called *Cantor’s Principle* in (Benci et al., 2007), where the authors investigate it in the context of their Euclidean theory of size:

Definition 15 (Cantor’s Principle)

$$\text{If } \Pr_{\mathcal{U}}(\theta \in A) = \Pr_{\mathcal{U}}(\theta \in B), \text{ then } |A| = |B|.$$

Benci, Forti, and Di Nasso prove that ‘Cantor’s Principle’ can be made to hold (Benci et al 2007, Sect. 3.2). It is also clear that Cantor’s Principle (for A, B such that $|A|, |B| \geq \omega$) follows from super-regularity.

3.3.3 The Power Set Principle

The question whether

$$\forall A, B \in V : |A| < |B| \Rightarrow |\mathcal{P}(A)| < |\mathcal{P}(B)|$$

is true, is independent of the axioms of set theory. (Of course the principle is true if the Generalised Continuum Hypothesis holds.) Like the cardinality operator, our generalised probability functions are measures of some kind. One might wonder what should follow from $\Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B)$. In particular, given that $\Pr_{\mathcal{U}}$ is intended to be a *fine-grained* quantitative possibility measure, perhaps probability should be expected to co-vary with the power set operation in some fairly direct manner. In other words, it is natural to ask if the following principle can be made to hold:

Definition 16 (*Power Set Condition*)

$$\forall A, B \in V : \Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B) \Leftrightarrow \Pr_{\mathcal{U}}(\theta \in \mathcal{P}(A)) < \Pr_{\mathcal{U}}(\theta \in \mathcal{P}(B)).$$

It turns out that the power set condition can indeed be satisfied:

Theorem 2 *There are fine ultrafilters \mathcal{U} such that $\Pr_{\mathcal{U}}$ satisfies the power set condition.*

The argument for this is somewhat more involved.

We aim to prove Theorem 2 by building the probability function up from an ultrafilter \mathcal{U} which is based on a pre-filter $\mathcal{C} \subseteq \mathcal{P}([V]^{<\omega})$.¹⁹

The class \mathcal{C} is built up in stages, and in such a way that it eventually witnesses the truth of the power set condition for all $A, B \in V$.

Stage 0

The class \mathcal{C}_0 consists of all

$$A_x \equiv \{a \in [V]^{<\omega} : x \in a\},$$

for $x \in V$. This is to ensure that the ultrafilter that will be built from \mathcal{C} is fine. We know that \mathcal{C}_0 has the finite intersection property.

Limit stages

For limit stages λ , we simply set $\mathcal{C}_\lambda \equiv \bigcup_{\beta < \lambda} \mathcal{C}_\beta$.

Successor stages

Given fineness, we may, and will, ignore the elements of V_ω . At stage $\alpha > \omega$, where α is a successor ordinal, we consider the sets of $V_\alpha \setminus V_{\alpha-1}$ and ensure that the power set condition eventually holds for all these sets and their power sets, by adding families of finite sets to $\mathcal{C}_{\alpha-1}$ in such a way that the finite intersection property is preserved.

As an illustrative and indeed representative example we do the case where $\alpha = \omega + 1$.

Let there be given an enumeration $\{A_1, B_1\}, \dots, \{A_\beta, B_\beta\}, \dots$ of the pairs of elements of $V_{\omega+1} \setminus V_\omega$.

For the induction, we assume that, by having added appropriate sets of finite sets to \mathcal{C}_0 , the power set condition holds for $\{A_1, B_1\}, \dots, \{A_\beta, B_\beta\}$ and their power sets, and that in the process the finite intersection property has been preserved. The aim is now to extend this so that it also holds for $\{A_{\beta+1}, B_{\beta+1}\}$. In other words, we have constructed \mathcal{C}_1^β , and we want to obtain $\mathcal{C}_1^{\beta+1}$, where $\mathcal{C}_1^0 \equiv \mathcal{C}_0$.

Definition 17

$$\mathcal{C}_{A < B} \equiv \left\{ D \in [V]^{<\omega} : \frac{|A \cap D|}{|D|} < \frac{|B \cap D|}{|D|} \right\}.$$

Definition 18

$$\mathcal{C}_{A \geq B} \equiv \left\{ D \in [V]^{<\omega} : \frac{|A \cap D|}{|D|} \geq \frac{|B \cap D|}{|D|} \right\}.$$

Claim

Either $\mathcal{C}_1^\beta \cup \{C_{A_\beta < B_\beta}\}$ has the finite intersection property, or $\mathcal{C}_1^\beta \cup \{C_{A_\beta \geq B_\beta}\}$ has the finite intersection property (or both).

Proof Suppose not. Then there is a finite intersection F of elements of \mathcal{C}_1^β such that $F \cap C_{A_\beta < B_\beta} = \emptyset$, and there is a finite intersection F' of elements of \mathcal{C}_1^β such that $F' \cap C_{A_\beta \geq B_\beta} = \emptyset$. But then $(F \cap F') \cap C_{A_\beta < B_\beta} = \emptyset$ and $(F \cap F') \cap C_{A_\beta \geq B_\beta} = \emptyset$. But

¹⁹ A pre-filter is a non-empty set with the finite intersection property.

$C_{A_\beta < B_\beta} \cup C_{A_\beta \geq B_\beta} = [V]^{<\omega}$. So then $(F \cap F') = \emptyset$. But this contradicts the inductive assumption that C^β has the finite intersection property.

Thus define $C_1^{\beta+1}$ to be $C_1^\beta \cup \{C_{A_\beta < B_\beta}\}$ if this has the finite intersection property, or $C_1^\beta \cup \{C_{A_\beta \geq B_\beta}\}$ otherwise, and by the claim, $C_1^{\beta+1}$ has the finite intersection property. Now setting $C_1^- \equiv \bigcup_\beta C_1^\beta$, we may conclude that C_1^- has the finite intersection property.

At this point we must extend C_1^- by adding to C_1^- :

- every set of the form $C_{\mathcal{P}(A) < \mathcal{P}(B)}$ such that $C_{A < B} \in C_1^-$;
- every set of the form $C_{\mathcal{P}(A) \geq \mathcal{P}(B)}$ such that $C_{A \geq B} \in C_1^-$.

Call the resulting set C_1 . Our aim is to prove that C_1 has the finite intersection property.

Consider an arbitrary non-empty finite family $\mathcal{F} \subseteq C_1$. Without loss of generality we may assume that the ‘judgements’ in \mathcal{F} of the form $C_{\mathcal{P}(A) < \mathcal{P}(B)}$ or $C_{\mathcal{P}(A) \geq \mathcal{P}(B)}$, taken together, describe a finite total pre-ordering relation R on some set $\{\mathcal{P}(A_1), \dots, \mathcal{P}(A_k)\}$. Further, we may also assume that for sets A and B from $V_{\omega+1} \setminus V_\omega$, $C_{\mathcal{P}(A) < \mathcal{P}(B)} \in \mathcal{F}$ if and only if $C_{A < B} \in \mathcal{F}$, and $C_{\mathcal{P}(A) \geq \mathcal{P}(B)}$ iff $C_{A \geq B} \in \mathcal{F}$. Thus \mathcal{F} contains witnesses for all the relevant judgements we may be interested in.

Let $\mathcal{F}^- = \mathcal{F} \cap C_1^-$, so \mathcal{F}^- consists only of judgements about sets in $V_{\omega+1} \setminus V_\omega$. Then we know from the foregoing that $\bigcap \mathcal{F}^- \neq \emptyset$. So take some $F^- \in \bigcap \mathcal{F}^-$. Our plan is inductively to extend F^- , using the pre-order R , to a finite set $F \in \bigcap \mathcal{F}$.

We will add to F^- elements that ensure that the constraints of R are satisfied. Moreover, by choosing the elements to be added to F^- from $V_{\omega+1} \setminus V_\omega$,²⁰ we ensure that the constraints imposed by \mathcal{F}^- remain satisfied. As a result, F will satisfy all constraints from \mathcal{F} , so $\bigcap \mathcal{F} \neq \emptyset$ and hence C_1 has the finite intersection property.

As an example, suppose that R says that

$$|\mathcal{P}(A_1)| < |\mathcal{P}(A_2)| < |\mathcal{P}(A_3)| = |\mathcal{P}(A_4)|.$$

- (1) We start by ensuring that $|\mathcal{P}(A_1)| < |\mathcal{P}(A_2)|$ is satisfied.

Suppose that F^- already contains n elements of $\mathcal{P}(A_1)$. Since $C_{A_1 < A_2} \in \mathcal{F}$, there must be an element $x^- \in A_2 \setminus A_1$. This implies that there are infinitely many infinite sets x in $\mathcal{P}(A_2) \setminus \mathcal{P}(A_1)$ such that $x^- \in x$: we add $n + 1$ such elements to F^- , and call the resulting finite set F_1^- .

- (2) We proceed in similar fashion to ensure that $|\mathcal{P}(A_2)| < |\mathcal{P}(A_3)|$ is satisfied:

Suppose that F_1^- already contains m elements from $\mathcal{P}(A_2)$, observing that it may be the case that $m > n + 1$, for there may already be a finite number of elements of $\mathcal{P}(A_2)$ in F^- . Since $C_{A_2 < A_3} \in \mathcal{F}$, there must be an element $y_1^- \in A_3 \setminus A_2$, and since $C_{A_1 < A_3} \in \mathcal{F}$, there must be an element $y_2^- \in A_3 \setminus A_1$. So there are infinitely many infinite sets y in $\mathcal{P}(A_3)$ such that $y_1^-, y_2^- \in y$: add $m + 1$ such elements to F_1^- , and call the resulting set F_2^- .

²⁰ For later stages we will take these sets from $V_{\alpha+1} \setminus V_\alpha$, i.e. sets of rank α .

- (3) Now suppose that there are m_1 elements of $\mathcal{P}(A_3)$ in F_2^- , and m_2 elements of $\mathcal{P}(A_4)$ in F_2^- . Moreover, suppose that $m_2 < m_1$. (The case where $m_1 < m_2$ is similar.) Since $C_{A_3 \geq A_4}, C_{A_4 \geq A_3} \in \mathcal{F}$, but also $A_3 \neq A_4$, there must be some $x_1 \in A_3 \setminus A_4$ and some $x_2 \in A_4 \setminus A_3$. Moreover, since $C_{A_1 < A_4}, C_{A_2 < A_4} \in \mathcal{F}$, there are elements $x_3 \in A_4 \setminus A_1, x_4 \in A_4 \setminus A_2$. So $\mathcal{P}(A_4)$ contains infinitely many infinite sets x such that $\{x_2, x_3, x_4\} \subset x$. Similarly, $\mathcal{P}(A_3)$ contains infinitely many infinite sets x that are outside $\mathcal{P}(A_1), \mathcal{P}(A_2), \mathcal{P}(A_4)$. So we add a sufficient number of such elements to F_2^- so that there are an equal number p of “witnesses” for $\mathcal{P}(A_3)$ as for $\mathcal{P}(A_4)$ but where p is larger than the number of witnesses for $\mathcal{P}(A_2)$. Call the resulting set F_3^- .
- (4) To conclude, we set $F \equiv F_3^-$. It is clear that $F \in \bigcap \mathcal{F}$.

This procedure of extending F^- easily generalises to any finite total pre-ordering on $\{\mathcal{P}(A_1), \dots, \mathcal{P}(A_k)\}$. Thus we have shown that \mathcal{C}_1 has the finite intersection property.

This procedure for extending \mathcal{C}_0 to \mathcal{C}_1 while preserving the finite intersection property also works for larger successor ordinals: at level $V_{\alpha+1}$ (stage $\beta + 1$ with $\alpha = \omega + \beta$) we can extend the corresponding F^- using subsets of rank α . As we have said above, at limit stages we can simply take unions. Ultimately we set $\mathcal{C} \equiv \bigcup_{\alpha \in On} \mathcal{C}_\alpha$.

The class \mathcal{C} will then have the finite intersection property, so it can be extended to a filter and then to an ultrafilter \mathcal{U} . The probability function based on \mathcal{U} will make the power set condition true for all $A, B \in V$, and this concludes the proof of theorem 2. □

With a minimal amount of extra work, our proof can be seen to show something slightly stronger: for all A, B with $|A|, |B| \geq \omega$, we have

$$\Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B) \Leftrightarrow \Pr_{\mathcal{U}}(\theta \in \mathcal{P}(A)) \ll \Pr_{\mathcal{U}}(\theta \in \mathcal{P}(B)).$$

The reason is that in enlarging the set F^- we always have infinitely many elements to choose from.

For any probability measure $\Pr_{\mathcal{U}}$ that satisfies power set condition we also have that $\forall A, B \in V, \forall n \in \omega$:

$$\Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B) \Leftrightarrow \Pr_{\mathcal{U}}(\theta \in \mathcal{P}^n(A)) < \Pr_{\mathcal{U}}(\theta \in \mathcal{P}^n(B))$$

where $\mathcal{P}^n(A) = \mathcal{P}(\mathcal{P}(\dots \mathcal{P}(A) \dots))$. An easy argument shows this cannot extend to infinite applications of the power set operation: if $B = \mathcal{P}(A)$ then $\mathcal{P}^\omega(B) = \mathcal{P}^\omega(A)$.

One might wonder whether the motivations behind the power set condition should not also support imposing the following *restricted power set condition* on $\Pr_{\mathcal{U}}$:²¹

Question 1 Are there probability measures such that

²¹ This question was suggested to us by Philip Welch.

$$\forall A, B \in V : \Pr_{\mathcal{U}}(\theta \in A) < \Pr_{\mathcal{U}}(\theta \in B) \Leftrightarrow \Pr_{\mathcal{U}}(\theta \in [A]^{<\omega}) < \Pr_{\mathcal{U}}(\theta \in [B]^{<\omega})?$$

3.4 The Ordinals

For $\alpha \geq \omega$, in each level $V_{\alpha+1} \setminus V_{\alpha}$ of the iterative hierarchy one finds only one ordinal, but infinitely many sets that are not ordinals. This might lead one to believe that a probability function on V should satisfy

$$\Pr_{\mathcal{U}}(\theta \in \text{On}) \approx 0,$$

where ‘On’ is the class of ordinals.

Just as it seems reasonable to require that the probability of choosing an even natural number from the set of natural numbers must be equal to or infinitesimally close to $\frac{1}{2}$ (see Wenmackers et al 2013, section 6.2), it seems reasonable to require that

$$\Pr_{\mathcal{U}}(\theta \in \text{Even} \mid \theta \in \text{On}) \approx \frac{1}{2},$$

where ‘Even’ is the class of even ordinals, which is defined in the obvious way.

Moreover, between any two limit ordinals there are infinitely many successor ordinals, so one might expect

$$\Pr_{\mathcal{U}}(\theta \in \text{Lim} \mid \theta \in \text{On}) \approx 0,$$

where ‘Lim’ is the class of limit ordinals.

We will sketch how probability functions can be constructed that meet these expectations. Indeed, we will see that there are probability functions that meet these ‘ordinal expectations’ and in addition meet the size constraint of super-regularity.

Theorem 3 *There is a super-regular probability function $\Pr_{\mathcal{U}}$ such that:*

1. $\Pr_{\mathcal{U}}(\theta \in \text{On}) \approx 0$;
2. $\Pr_{\mathcal{U}}(\theta \in \text{Even} \mid \theta \in \text{On}) \approx 2^{-1}$;
3. $\Pr_{\mathcal{U}}(\theta \in \text{Lim} \mid \theta \in \text{On}) \approx 0$.

Proof As before, the aim is to choose wisely the ultrafilter \mathcal{U} on which $\Pr_{\mathcal{U}}$ is based. We want \mathcal{U} to be such that for all $k, l, m \in \mathbb{N}$:

- $\frac{\Pr_{\mathcal{U}}(\theta \in A)}{\Pr_{\mathcal{U}}(\theta \in B)} \leq k^{-1}$ if $\omega \leq |A| < |B|$;
- $\Pr_{\mathcal{U}}(\theta \in \text{Even} \mid \theta \in \text{On}) - \Pr_{\mathcal{U}}(\theta \in \text{Odd} \mid \theta \in \text{On}) \leq l^{-1}$ and $\Pr_{\mathcal{U}}(\theta \in \text{Lim} \mid \theta \in \text{On}) \leq l^{-1}$;
- $\Pr_{\mathcal{U}}(\theta \in \text{On}) \leq m^{-1}$.

Now we define:

- $A_x \equiv \{D \in [V]^{<\omega} : x \in D\}$;
- $C_{AB}^k \equiv \{D \in [V]^{<\omega} : \frac{\Pr[A|D]}{\Pr[B|D]} \leq k^{-1}\}$;
- $I^l \equiv \{D \in [V]^{<\omega} : \forall \alpha \in D \cap On \exists \beta (\alpha \in [\beta, \beta + l] \subseteq D)\}$;
- $W^m \equiv \{D \in [V]^{<\omega} : \Pr[On \mid D] \leq m^{-1}\}$.

And now we set:

$$\mathcal{F}_0 \equiv \{A_x, C_{AB}^k, I^l, W^m : x \in V, k, l, m \in \mathbb{N} \text{ and } \omega \leq |A| < |B|\}$$

The inclusions of the sets A_x in \mathcal{F}_0 will ensure the resulting filter is *fine*, the inclusion of the sets C_{AB}^k ensures super-regularity, the inclusion of the sets W^m ensures property 1. of the theorem holds. The inclusion of the sets I^l will give us property 3. as the ratio of limit ordinals to successor ordinals in any element of I^l is less than or equal to l^{-1} , but these sets also give us property 2. as any $D \in I^l$ can have at most one extra odd or even ordinal for every l elements it contains, so the ratio of odd to even ordinals must tend towards 1.

Claim: \mathcal{F}_0 has the finite intersection property.

Take any finite $F \subset \mathcal{F}_0$. Let some x_1, \dots, x_n be the indices corresponding to the A_x type elements of F . Now $\bigcap_{i \leq n} I^{l_i} = I^{l_{max}}$ where $l_{max} = \max\{l_i : i < n\}$, and similarly for $\bigcap_{i \leq n} W^{m_i}$, so as before in Theorem 1, it suffices to consider the maximum values of k, l, m represented as indices of, respectively, the $C, I,$ and W type sets in F .

- (1) $A \in \bigcap_{i \leq n} A_{x_i} \Leftrightarrow \{x_1, \dots, x_n\} \subseteq A$. So we start with the finite set $A_0 \equiv \{x_1, \dots, x_n\}$, and will extend it.
- (2) Again we concentrate on one pair $\langle A, B \rangle$ such that $\omega \leq |A| < |B|$; we leave out further cases as they are similar. There are arbitrarily large finite subsets $C \subseteq B$ that are l -isolated from elements of A , meaning that each ordinal in C is more than l ordinals removed from any ordinal in A . We choose any such $C \subseteq B$ that is of size at least $k \cdot n$, and we set $A_1 \equiv A_0 \cup C$.
- (3) Now we extend A_1 to ensure that all ordinal intervals are of length $\geq l$: for each $\alpha \in A_1$, we add $\alpha + 1, \dots, \alpha + l$. Call the resulting finite collection A_2 . Note that by our choice of l -isolated elements in (2), for any $\alpha \in A_1 \setminus A_0$, none of $\alpha + 1, \dots, \alpha + l$ are elements of A , and thus the ration of elements in A to elements in B remains below $1 : k$.
- (4) Let $|A_2| = j$. Then we add $j \cdot m$ elements of $V \setminus (A \cup B \cup On)$ to A_2 and call the resulting set A_3 .

It is now routine to verify that $A_3 \in \bigcap_{i \leq n} A_{x_i} \cap C_{AB}^k \cap I^l \cap W^m$. The case including further sets $C_{A'B'}$ is similar, thus the claim is verified. So \mathcal{F}_0 indeed has the finite intersection property, whereby it can be extended to a filter and then further to an ultrafilter \mathcal{U} . By design, the resulting probability function $\Pr_{\mathcal{U}}$ has the required properties. □

4 The bootstrapping approach

The probability $\Pr_{\mathcal{U}}(\theta \in A)$ is obtained by ‘summing up’ the probabilities $\Pr(\theta \in A \mid \iota \in S)$, where ι is the identity random variable, for all ‘small’ parts S of V ; such $\Pr(\theta \in A \mid \iota \in S)$ are seen as *approximations* of $\Pr_{\mathcal{U}}(\theta \in A)$.

In the finite snapshot approach, ‘small’ in this context means ‘finite’. But from a conceptual point of view, ‘finite’ might be taken to be *too* small as far as the test sets (or snapshots) are concerned. Compared to V , *all* sets —and not just the finite sets— are small. So to determine $\Pr_{\mathcal{U}}(\theta \in A)$, we should take the ‘limit’ of the values $\Pr(\theta \in A \mid \iota \in S)$, where S is a set of any size. Then if S is infinite, $\Pr(\theta \in A \mid \iota \in S)$ cannot just be taken to be given by the ratio formula but needs to be *defined*.

In the approach to which we now turn (the *bootstrapping approach*), a probability $\Pr_{\mathcal{U}}(\theta \in A)$ is determined by the probabilities $\Pr_{\mathcal{U}}(\theta \in A \mid \iota \in S)$, where $\Pr_{\mathcal{U}}(\theta \in A \mid \iota \in S)$, for S a large set, is then in turn determined by probabilities $\Pr_{\mathcal{U}}(\theta \in A \mid \iota \in S')$ for S' being smaller ‘snapshots’ than S , and so on, until we reach the finite snapshots and can appeal to the probability functions that were discussed in the previous sections. Thus the bootstrapping account can be seen as a *generalisation* of the finite snapshot approach.

4.1 The Rough Idea

In general terms, this is how we will proceed:

- (1) By the construction from the previous section, a fine ultrafilter on $[S]^{<\omega}$ yields a notion of probability on all sets $S \in V$ with $|S| < \omega_1$. In other words, this yields a suitable notion of probability, call it \Pr^S , for every *countable* set S .
- (2) The notion of \Pr^S for all $S \in V$ with $|S| < \omega_2$ is determined using the notion of probability on *countable* sets: the probability of A on such an S is determined by the class of probabilities of A on the countable ‘snapshots’ of S . Using these countable probability functions, a fine ultrafilter on $[S]^{<\omega_1}$ gives us a notion of probability on sets S with $|S| < \omega_2$.

Again the resulting functions \Pr^S are essentially NAP-functions as defined in (Benci et al., 2013). They are total, regular, etc.

⋮
 (β) A fine ultrafilter on $[S]^{<\omega_\alpha}$, together with probability functions \Pr^S for all S such that $|S| < \omega_\alpha$, yields a notion of probability on all sets S with $|S| < \omega_{\alpha+1}$.

⋮
 Limit stages of course do not present a problem. So by transfinite recursion on cardinality this yields for *every* set S a notion \Pr^S of probability on S .

Then a fine ultrafilter \mathcal{U} on $V = [V]^{<Card}$ yields, using the general notion \Pr^S for $S \in V$, a notion \Pr^V that is a total (class) function from properties A and random variables θ to values $\Pr^V(\theta \in A)$ in a non-Archimedean class field. This probability function again satisfies the principles of the theory NAP in (Benci et al., 2013).

For this construction, we need suitable (fine) ultrafilters on increasingly larger and larger sets, and a fine ultrafilter \mathcal{U} on $[V]^{<Card}$. But we will see that all the set ultrafilters used in the construction can be *uniformly* obtained as *restrictions* to sets S of the given fine ultrafilter on $[V]^{<Card}$. So Pr^V is determined by *one* initial choice of \mathcal{U} , whereby Pr^V can be seen as the ‘limit’ of its set-restrictions Pr^S , where the functions Pr^S can in turn be seen as ‘limits’ of restrictions to *their* small subsets. This uniform construction has the advantage that the resulting probability functions are all *coherent*, in the sense that for a set T , $\text{Pr}^S(\theta \in A \mid \iota \in T)$ is (up to a canonical embedding, as is explained in detail in the first part of the proof of Proposition 7 below) the same for all $S \supseteq T$ and hence also for V .

Now it is time to look at details of the construction.

4.2 Details 1: Restrictions of Fine Ultrafilters

Since our construction involves ultrafilters on sets $[S]^{<\kappa}$ with $\kappa > \omega$, we make the following definition, which accords with the usual definition of fineness on $[S]^{<\omega}$.

Definition 19 For any infinite cardinal κ , an ultrafilter \mathcal{U} on $[S]^{<\kappa}$ is fine iff for every $x \in S$:

$$\{T \in [S]^{<\kappa} : x \in T\} \in \mathcal{U}.$$

Moreover, an ultrafilter \mathcal{U} on $[V]^{<Card}$ is fine iff for every $x \in V$:

$$\{T \in [V]^{<Card} : x \in T\} \in \mathcal{U}.$$

We first show that appropriate restrictions of ultrafilters to smaller sets can be obtained in a uniform fashion.

Definition 20 Suppose $S \in V$, $|S| = \kappa$, and \mathcal{U} a fine ultrafilter on $[S]^{<\kappa}$, and $S' \subseteq S$ with $|S'| = \alpha < \kappa$. Then we define the restriction $\mathcal{U}_{S'}$ of \mathcal{U} to S' as follows.

For any $X \in \mathcal{P}([S]^{<\kappa})$, let

$$X_{S'} \equiv \{y \mid \exists z \in X : y = z \cap S' \text{ and } |y| < \alpha\}.$$

Then $\mathcal{U}_{S'} \equiv \{X_{S'} \mid X \in \mathcal{U}\}$.

Proposition 6 For any $S \in V$ with $|S| = \kappa$, there are fine ultrafilters \mathcal{U} on $[S]^{<\kappa}$ that restrict to a fine ultrafilter on every $S' \subseteq S$ with $|S'| = \alpha$, and $\omega \leq \alpha < \kappa$.

Further, such ultrafilters are coherent in that if $T \subset S'$ with $\omega \leq |T| < |S'|$, then $(\mathcal{U}_{S'})_T = \mathcal{U}_T$.

Proof We build the ultrafilter from a pre-filter \mathcal{F}_0 , which can then be extended to a filter and then to an ultrafilter.

For each $x \in S$, let

$$A_x \equiv \{X \in [S]^{<\kappa} : x \in X\}.$$

And let for each S' with $|S'| = \alpha < \kappa$ and $S' \subseteq S$:

$$R^{S'} \equiv \{X \in [S]^{<\kappa} : X \cap S' \in [S']^{<\alpha}\}.$$

Now set

$$\mathcal{F}_0 \equiv \{A_x : x \in S\} \cup \{R^{S'} : S' \subseteq S \text{ and } |S'| < \kappa\}.$$

It is easy to see that \mathcal{F}_0 has the finite intersection property and so can be extended to an ultrafilter \mathcal{U} . And by design, \mathcal{U} is fine.

Clearly $\mathcal{U}_{S'} \subseteq \mathcal{P}([S']^{<\alpha})$. We must check the fine ultrafilter properties for $\mathcal{U}_{S'}$:

- (1) *Fine*. This follows from the fact that \mathcal{U} is fine: for $x \in S'$ this is witnessed by $(A_x)_{S'}$.
- (2) *Finite intersection*. Let $X, Y \in \mathcal{U}_{S'}$. Then there are $\bar{X}, \bar{Y} \in \mathcal{U}$ such that $X = \bar{X}_{S'}$ and $Y = \bar{Y}_{S'}$. By the finite intersection property of \mathcal{U} , we know that $\bar{X} \cap \bar{Y} \in \mathcal{U}$. But $X \cap Y \supseteq (\bar{X} \cap \bar{Y})_{S'}$. So $X \cap Y \in \mathcal{U}_{S'}$.
- (3) *Ultra*. Take any $X \subseteq [S']^{<\alpha}$, and let $X^c \equiv [S']^{<\alpha} \setminus X$. Let $\bar{X} \equiv \{x \in [S]^{<\kappa} \mid x \cap S' \in X\}$ and let $\bar{X}^c \equiv \{x \in [S]^{<\kappa} \mid x \cap S' \notin X\}$. Then $\bar{X}^c = [S]^{<\kappa} \setminus \bar{X}$. By the ultra property for \mathcal{U} , we have $\bar{X} \in \mathcal{U}$ or $\bar{X}^c \in \mathcal{U}$. But $X = \bar{X}_{S'}$ and $X^c = \bar{X}^c_{S'}$. So $X \in \mathcal{U}_{S'}$ or $X^c \in \mathcal{U}_{S'}$.
- (4) *Non-principality*. This is implied by fineness.
- (5) *Empty set property*: We have to show that $\emptyset \notin \mathcal{U}_{S'}$. It suffices to show that for each $X \in \mathcal{U}$, $X_{S'} \neq \emptyset$. Since $R^{S'} \in \mathcal{U}$, $X \cap R^{S'} \neq \emptyset$. But for any set x in this intersection, $x \cap S' \in [S']^{<\alpha}$. So $x \cap S' \in X_{S'} \neq \emptyset$.

For coherence, take $T \subset S' \subset S$ with $|T| < |S'| < |S|$ and let $X \in \mathcal{U}$. As $R^{S'} \in \mathcal{U}$, and restrictions of subsets are subsets of restrictions, it is enough to show that $((X \cap R^{S'})_{S'})_T = (X \cap R^{S'})_T$. Now $((X \cap R^{S'})_{S'})_T = \{y \mid \exists z \in X \cap R^{S'} : y = z \cap T, |y| < |T| \text{ and } |z \cap S'| < |S'|\}$, but by definition, for any $z \in R^{S'}$ we have $|z \cap S'| < |S'|$. Thus $((X \cap R^{S'})_{S'})_T = \{y \mid \exists z \in X \cap R^{S'} : y = z \cap T \text{ and } |y| < |T|\} = (X \cap R^{S'})_T$. \square

It can then be seen that this property must also hold for fine ultrafilters on $[V]^{<Card}$:

Consequence 1 There are fine ultrafilters \mathcal{U} on $[V]^{<Card}$, such that for every set S with $|S| = \alpha$, \mathcal{U}_S is a fine ultrafilter on $[S]^{<\alpha}$ and the coherence property holds.

Proof By the same reasoning as in the previous proposition. \square

4.3 Details 2: Defining Probability Functions

Now we show how for every set, a probability function on that set can be defined. The same procedure can then be used to define a probability function on V , and these probability functions are *coherent*.

The key is to spell out what is involved in the β -th step of the recursive procedure for defining probabilities on sets:

(β) A fine ultrafilter \mathcal{U} on $[S]^{<\omega_\beta}$ (with $\omega_\beta = |S|$), together with probability functions \Pr^T for all T such that $|T| < \omega_\beta$, yields a notion of probability \Pr^S on S .

As in Sect. 2, we define a function $f_{\theta \in A}$ such that for all $T \in [S]^{<\omega_\beta}$:

$$f_{\theta \in A}(T) \equiv \Pr^T(\theta \in A \mid \iota \in T) \equiv \Pr^T(\{s \in T : \theta(s) \in A\}).$$

Similarly, we define a function $f_{\theta \in A \wedge \nu \in B}$ such that for all $T \in [S]^{<\omega_\beta}$:

$$f_{\theta \in A \wedge \nu \in B}(T) \equiv \Pr^T(\theta \in A \wedge \nu \in B \mid \iota \in T) \equiv \Pr^T(\{s \in T : \theta(s) \in A \wedge \nu(s) \in B\}).$$

Then $\Pr^S(\theta \in A)$ is defined as $[f_{\theta \in A}]_{\mathcal{U}}$, and $\Pr^S(\theta \in A \mid \nu \in B)$ is defined as

$$\frac{[f_{\theta \in A \wedge \nu \in B}]_{\mathcal{U}}}{[f_{\nu \in B}]_{\mathcal{U}}}.$$

This function \Pr^S will then be an NAP probability function in the sense of (Benci et al., 2013).

Now in an exactly similar way, we define a class probability function $\Pr_{\mathcal{U}}^+$ on V , using the probability functions on ‘small’ classes (i.e., sets) and ultrafilters on ‘small’ classes which (given proposition 6) we can now assume to have been defined on the basis of an ultrafilter \mathcal{U} on $[V]^{<Card}$ with which we start. The function $\Pr_{\mathcal{U}}^+$ is total, regular, and uniform for the same reasons as why its ‘smaller cousin’ $\Pr_{\mathcal{U}}$ has these properties.

We now check coherence. We will do this only for probabilities of the identity random variable rather than random variables in general, as although coherence holds for all random variables, it is more technical to prove. Below we use $\Pr(A)$ to denote $\Pr(\iota \in A)$, $\Pr(A|T)$ for $\Pr(\iota \in A \mid \iota \in T)$, and $f_A(T)$ for $f_{\iota \in A}(T)$, where ι is the identity random variable.

Proposition 7 For any class A and sets $T \subset S$ with $|T| < |S|$ we have²²

$$\Pr^T(A) = \Pr^S(A|T).$$

Proof We show by induction on $|T|$ that that the above holds for all $S \supset T$ with $|S| > |T|$. This is trivial for finite T as on both sides we are just using the ratio formula. For infinite T , strictly speaking, the range of \Pr^T may be a non-Archimedean field that is different from the range of \Pr^S , but there is a natural embedding of the

²² Coherence for an *arbitrary* random variable would be:

$$\Pr^T(\theta \in A) = \Pr^S(\theta \in A \mid \iota \in T).$$

former into the latter defined by $i([f]_{\mathcal{U}_T}) = [\bar{f}]_{\mathcal{U}_S}$ where for $X \in S^{<|S|}$, $\bar{f}(X) = f(X \cap T)$ where the latter is defined and zero otherwise. The embedding is well-defined as $\{X \in S^{<|S|} : |X \cap T| < |T|\} = (R^T)_S \in \mathcal{U}_S$, so if $g \in [f]_{\mathcal{U}_T}$, so g agrees with f on a set Y in \mathcal{U}_T , then \bar{f} agrees with \bar{g} on $(R^T)_S \cap \bar{Y}$, where \bar{Y} is any set in \mathcal{U}_S with $\bar{Y}_T = Y$. Such a \bar{Y} exists by cohercnce.

For a given infinite T we assume the property holds for any T' with $|T'| < |T|$, and show it holds for T . Let S be arbitrary with $|S| > |T|$.

Using the embedding we have $i(\text{Pr}^T(A)) = i([f_A]_{\mathcal{U}_T}) = [\bar{f}_A]_{\mathcal{U}_S}$. Now for $X \in (R^T)_S (\in \mathcal{U}_S)$ we have:

$$\bar{f}_A(X) = f_A(X \cap T) = f_{A \cap T}(X \cap T) = \text{Pr}^{X \cap T}(A \cap T).$$

As $X \in (R^T)_S$ we have $|X \cap T| < |T|$ so by our inductive hypothesis

$$\text{Pr}^{X \cap T}(A \cap T) = \text{Pr}^X(A \cap T|T) = \frac{f_{A \cap T}(X)}{f_T(X)}.$$

But by definition, $[\frac{f_{A \cap T}}{f_T}]_{\mathcal{U}_S} = \text{Pr}^S(A|T)$, so $[\bar{f}_A]_{\mathcal{U}_S} = \text{Pr}^S(A|T)$ and we're done. □

4.4 Comparison of the Finite Snapshot Approach and the Bootstrapping Approach

In our definition of the probability of a set theoretic property, the probability $\text{Pr}_{\mathcal{U}}^+(\theta \in A)$ of θ having the property A is determined by the probabilities $\text{Pr}^S(\theta \in A)$ of A on large ‘snapshots’ S , where a probability $\text{Pr}^S(\theta \in A)$ (for S a large set) is then in turn determined by the probabilities $\text{Pr}^{S'}(\theta \in A)$ for S' being smaller ‘snapshots’ than S , and so on. Conceptually, the definition in Sect. 4.3 is superior to the simpler definition suggested from Sect. 2: we want to take the behaviour of the property on as many and as large ‘snapshots’ as possible into account.

It is not straightforward to compare the simple and the more involved definition: the simple method is based on an ultrafilter on $[V]^{<\omega}$ whereas the more involved method is based on an ultrafilter on $V = [V]^{<Card}$.

The obvious suggestion is to base the comparison on the relation between a probability function determined by an ultrafilter \mathcal{U} on $[V]^{<Card}$ and its *restriction*²³ to $[V]^{<\omega}$ defined as $\mathcal{U} \upharpoonright \omega = \{X \cap [V]^{<\omega} \mid X \in \mathcal{U}\}$. But:

Proposition 8 *Not all ultrafilters on $[V]^{<Card}$ restrict to ultrafilters on to $[V]^{<\omega}$.*

Proof Consider $\mathcal{A} \cup \overline{[V]^{<\omega}}$, where \mathcal{A} is the collection of sets of the form $\overline{A_x}$ for various x as used throughout the paper (guaranteeing fine-ness) and $[V]^{<\omega}$ is the relative complement of $[V]^{<\omega}$ in $[V]^{<Card}$. Then $\mathcal{A} \cup [V]^{<\omega}$ has the finite intersection

²³ This is a different notion of restriction to that defined in the previous section as here we are only restricting *Card*, while the underlying class V remains the same.

property and so can be extended to a fine ultrafilter \mathcal{U} on $[V]^{<Card}$. But $\emptyset \in \mathcal{U} \upharpoonright \omega$. So \mathcal{U} does not restrict to an ultrafilter on $[V]^{<\omega}$. \square

On the other hand, every fine ultrafilter on $[V]^{<Card}$ restricting to an ultrafilter on $[V]^{<\omega}$ essentially is an ultrafilter on $[V]^{<\omega}$:

Proposition 9 *Suppose \mathcal{U} is a fine ultrafilter on $[V]^{<Card}$ restricting to an ultrafilter $\mathcal{U} \upharpoonright \omega$ on $[V]^{<\omega}$. Then $[V]^{<\omega} \in \mathcal{U}$.*

Proof Since \mathcal{U} is ultra, we have $[V]^{<\omega} \in \mathcal{U}$ or $\overline{[V]^{<\omega}} \in \mathcal{U}$. But if $\overline{[V]^{<\omega}} \in \mathcal{U}$, then $\emptyset \in \mathcal{U} \upharpoonright \omega$, so that \mathcal{U} does not restrict, contradicting the assumption. So $[V]^{<\omega} \in \mathcal{U}$. \square

This means that the probability functions on V generated by the bootstrapping method cannot be reduced to ‘simple’ probability functions on V that were discussed in the previous section.

In sum, in the preceding sections we have explored two methods for modelling, by means of non-Archimedean probability functions, the properties of random variables ranging over the set theoretic universe. Concerning the finite snapshot method, we found that many of the probabilistic properties that seem intuitively plausible can be satisfied. The bootstrapping method is more satisfying from a conceptual point of view. But we have only been able to show that the resulting probability functions satisfy minimal requirements. Much work on the bootstrapping method therefore remains to be done.

5 Concluding Remarks

The real numbers are in a sense close to our physical world: we routinely use them to model physical phenomena and processes. For that reason, throwing a dart randomly at the real unit interval appears to be probabilistically meaningful,²⁴ even though, due to our finite powers of discrimination, this scenario cannot be experimentally realised.²⁵ In contrast with this, throwing a dart randomly at V seems much further removed from what probability theory is about. Therefore one might wonder whether the notion of random variable on V is a probabilistically meaningful concept at all.²⁶

In response to this, one might argue that if the abstract notions of random graph and random space are probabilistically meaningful, then why is the notion of random set not equally meaningful? But even those who are, for the reasons given in the previous paragraph, sceptical about attempts to apply probabilistic notions to the

²⁴ This is witnessed, for instance, by (Freiling, 1986) and the reactions in the literature to that article.

²⁵ For a discussion about whether infinite fair lotteries, of which the scenario under consideration is an instance, can be physically realised, see (Norton, 2018) and (Norton, 2020).

²⁶ Thanks to an anonymous referee for voicing this concern.

mathematical universe as a whole, might find implications of our results for generalised probability functions on the real numbers meaningful, as we shall now argue.

Consider the scenario of throwing a dart randomly at the real unit interval from the perspective of Kolmogorov probability. From that perspective, this is a probabilistically coherent scenario. It is then easy to see that not only all point events must have probability 0, but it must also be the case (because of σ -additivity) that for each *countably infinite* set $A \subseteq [0, 1]$, the probability that the random dart lands on a number in A must be 0.

If we consider this scenario from the perspective of the framework of the present article, the picture changes. Already according to the notion of probability described in the framework of (Benci et al., 2013), the dart landing in any given countably infinite subset of $[0, 1]$ is “more probable” than it landing in any finite subset of the unit interval.

This seems natural, but the intuition takes us further: shouldn't it be the case that in general, for *any* sets $A, B \subseteq [0, 1]$ with $|A| < |B|$, the probability that the dart lands in B is higher than the probability that it lands in A ? Theorem 1 shows that there exist generalised probability functions that satisfy this condition, so that in the framework of (Benci et al., 2013) this intuition can be fulfilled. This goes beyond the results in (Benci et al., 2013), where this was only shown for A finite: Theorem 1 implies that any uncountable subset can be given higher probability than any countable subset. The phenomenon becomes even more dramatic when the Continuum Hypothesis fails:²⁷ we may require, for *any* two infinite sets $A, B \subseteq [0, 1]$ such that $|A| < |B| < |\mathbb{R}|$ (and there *will* then be such pairs of sets), the probability that the dart lands in B is higher than the probability that the dart lands in A .

This can then be seen as a natural strengthening of the phenomenon (i.e., the above for A finite) that was already known to hold in the framework of (Benci et al., 2013). Since we are still only talking about the real unit interval, this is a result that has concrete probabilistic content. The bootstrapping method can also be applied to generalised probability on the real interval $[0, 1]$, so the conceptual refinement gained is equally relevant to this more concrete situation.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors have no financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

²⁷ Of course we do not at present know whether the Continuum Hypothesis is true or false.

References

- Benci, V., & Di Nasso, M. (2003). Numerosities of labelled sets. A new way of counting. *Advances in Mathematics*, *173*, 50–67.
- Benci, V., Di Mauro, Nasso, & Forti, M. (2006). An Aristotelian notion of size. *Annals of Pure and Applied Logic*, *143*, 43–53.
- Benci, V., Di Mauro, Nasso, & Forti, M. (2007). An Euclidean measure of size for mathematical universes. *Logique et Analyse*, *50*, 43–62.
- Benci, V., Horsten, H., & Wenmackers, S. (2013). Non-Archimedean probability. *Milan Journal of Mathematics*, *81*, 121–151.
- Benci, V., Horsten, L., & Wenmackers, S. (2018). Infinitesimal probabilities. *British Journal for the Philosophy of Science*, *69*, 509–552.
- Brickhill, H., & Horsten, L. (2018). Triangulating non-Archimedean probability. *Review of Symbolic Logic*, *11*, 519–546.
- Blitzstein, J., & Hwang, J. (2015). *Introduction to probability*. Boca Raton: CRC Press.
- Booth, D. (1970). Ultrafilters on a countable set. *Annals of Mathematical Logic*, *2*, 1–24.
- Eskeu, M. (2021). *Integration via ultrafilters*. <http://arxiv.org/abs/2004.09103>.
- Freiling, C. (1986). Axioms of infinity. Throwing darts at the real number line. *Journal of Symbolic Logic*, *51*, 190–200.
- Hamkins, J. (2015). Is the dream solution to the continuum hypothesis attainable? *Notre Dame Journal of Formal Logic*, *56*, 135–145.
- Horsten, L. *The Metaphysics and Mathematics of Arbitrary Objects*. Cambridge University Press, forthcoming.
- Kremer, P. (2014). Indeterminacy of fair infinite lotteries. *Synthese*, *191*, 1757–1760.
- North, J. (2010). An empirical approach to symmetry and probability. *Studies in History and Philosophy of Modern Physics*, *41*, 27–40.
- Norton, J. (2018). How to build an infinite lottery machine. *European Journal for Philosophy of Science*, *8*, 71–95.
- Norton, J. (2020). How NOT to build an infinite lottery machine. *Studies in History and Philosophy of Science*, *82*, 1–8.
- Pruss, A. (2013). Probability, regularity, and cardinality. *Philosophy of Science*, *80*, 231–240.
- Pruss, A. (2021). Non-classical probabilities invariant under symmetries. *Synthese*, *199*, 8507–8532.
- Scott, D. (1967). A proof of the independence of the continuum hypothesis. *Mathematical Systems Theory*, *1*, 89–111.
- Wenmackers, S., & Horsten, L. (2013). Fair infinite lotteries. *Synthese*, *190*, 37–61.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.