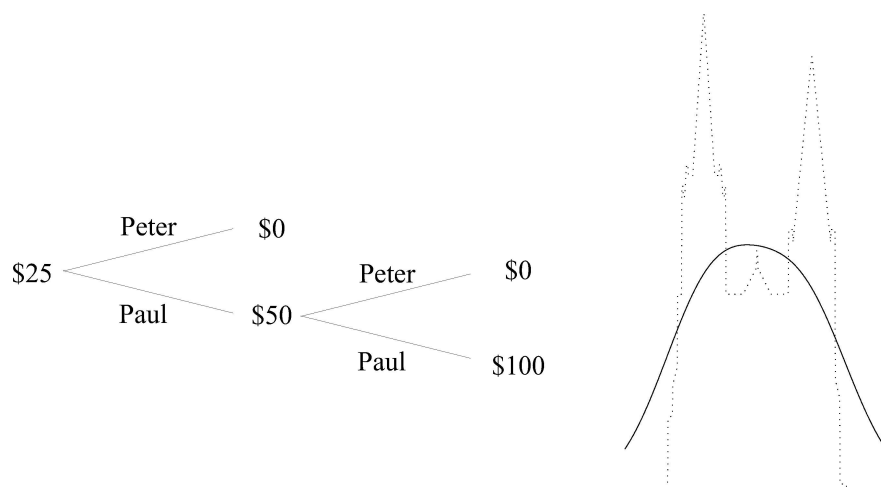


The non-mathematical section of Kolmogorov's *Grundbegriffe*

Vladimir Vovk



The Game-Theoretic Probability and Finance Project

Working Paper #69

August 18, 2026

Project web site:

<http://www.probabilityandfinance.com>

Abstract

Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung* introduced the standard measure-theoretic formalization of probability but included only a two-page section about the relation of the mathematical theory of probability to the world of experience. This note discusses this brief non-mathematical section concentrating on what I regard as its two most controversial features: having two bridges between the mathematical theory and the world of experience, instead of the standard one; the possibility of observing a prespecified event of probability zero in a sequence of trials.

Contents

1	Introduction	1
2	Some history of the <i>Grundbegriffe</i> and the <i>Verhältnis</i>	1
3	Summary of the <i>Verhältnis</i>	2
4	Why two principles?	4
5	A finitely additive measure consistent with Remark II	5
6	Conclusion	7
	References	8
A	Some details	9

1 Introduction

Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung*, first published in German in 1933 and introducing the standard measure-theoretic framework for probability, is often praised for its philosophical neutrality (see, e.g., [5]), but it also included a two-page section about the philosophy of probability. The title of this section, [7, Sect. I.2], is “Das Verhältnis zur Erfahrungswelt”, and Kolmogorov's footnote to this title makes it clear that it is a variation on von Mises's section title “Das Verhältnis der Theorie zur Erfahrungswelt” in his book [19]. In this note I will refer to Kolmogorov's section as the *Verhältnis*.

The main topic of this note is unusual features of the *Verhältnis*: having two bridges connecting measure-theoretic probability with reality, which he calls Principle A and Principle B, and the possibility of observing a prespecified event of probability zero [7, Sect. I.2, Remark II]. Both unusual features are related to the possibility of combining several experiments satisfying Kolmogorov's conditions, or some of them, into one that still satisfies those conditions. This will be covered in Sects. 4 and 5. The main finding of Sect. 5 is that Remark II in [7, Sect. I.2] is consistent with Kolmogorov's Axioms I–V and inconsistent with Axiom VI (that of continuity, equivalent to σ -additivity in the presence of the other axioms).

2 Some history of the *Grundbegriffe* and the *Verhältnis*

An important source of information about *Grundbegriffe* is the volume [16] of correspondence between Kolmogorov and Aleksandrov (the title of the volume is a line from Kolmogorov's poem devoted to Aleksandrov [15, preface, p. 14]). Kolmogorov started writing it on 26 October 1932 (letter to Aleksandrov of this date, [16]) and planned to work on it until January 1933 (letter of 8 November). The preface is dated “Easter 1933” (which was 16 April 1933, both in Russia and in the West).

The *Grundbegriffe* was translated into Russian in 1936 by Kolmogorov's student Grigory Bavli and then updated in 1974 by his other student, Albert Shiryaev. The second Russian edition modernized Kolmogorov's notation, and Shiryaev also extended the last three sections [7, Sects. VI.3–VI.5] by adding proofs (following Kolmogorov's contemporary papers). It seems very likely that Kolmogorov saw and approved the Russian translations. The third Russian edition appeared in 1998, after Kolmogorov's death in 1987; it seems identical to the second edition of 1974 apart from a new one-page preface by Prokhorov and Shiryaev and Shiryaev's essay on the history of probability (later translated into English as [17]).

An English translation of the *Grundbegriffe* was published only in 1950, although in his preface the translator, Nathan Morrison, mentions an earlier English translation. Morrison's translation used both the German original and the first Russian translation. Another English translation of the *Verhältnis*

appeared as [14, Sect. 5.2.1]; I will use it when quoting (directly and indirectly) from *Verhältnis* and will use Morrison’s translation when quoting from the rest of the *Grundbegriffe*.

One difference between the three Russian editions and the German original is that the former merge the latter’s Axioms I and II into one axiom, Axiom I. The numbering of axioms used in this note is the one in the German original (and the English translation); there are five axioms (I–V) preceding the *Verhältnis* and one axiom (Axiom VI of continuity) following it.

The earliest known draft of Chapter I of the *Grundbegriffe*, including the *Verhältnis*, was prepared by Kolmogorov for his lectures in Dnipro.¹ The *Verhältnis* was copied from the conspectus (конспект) of Kolmogorov’s Dnipro lectures and greatly expanded, according to Kolmogorov’s plan that he shared with Aleksandrov in the same letter of 26 October 1932 [16]. The whole of Chapter I was to be largely identical (довольно точно будет совпадать) to the Dnipro conspectus. Kolmogorov had been visiting Dnipro twice a year since Autumn 1931 [11, p. 43], in Spring and Autumn, and this continued, perhaps less regularly, until 1940 [12, p. 113]. His last visit before starting *Grundbegriffe* ended on 26 September 1932 (his letter to Aleksandrov of 24 September announces his departure), and he visited it again during the *Grundbegriffe* production, lecturing on 4–10 March 1933 mainly on probability theory (letter of 4 March).

3 Summary of the *Verhältnis*

When discussing the *Verhältnis*, I will use the notation used in the second and third Russian editions (and the English translation in [14, Sect. 5.2.1]).

The *Verhältnis* consists of three parts: a general scheme of the applications of probability, including Principles A and B; a frequentist derivation of the measure-theoretic axioms introduced by this point (not including the axiom of continuity); Remarks I and II.

In describing the way the theory of probability is applied to the real world Kolmogorov talks about a complex of conditions $\mathfrak{S}$ which allows an unlimited (unbeschränkt, неограниченный) number of repetitions. The theory of probability is applicable when we may assume that each event A that does or does not occur under conditions $\mathfrak{S}$ is assigned a number $P(A) \in [0, 1]$ with the following properties:

Principle A One can be practically certain that, if $\mathfrak{S}$ is repeated a large number of times, n , and the event A occurs m times, then m/n will differ only slightly from $P(A)$.

¹Dnipro was, and still is, Ukraine’s fourth largest city (after Kharkiv, the capital at the time, Kyiv, and Odesa). It has been known under many names, including Katerynoslav, Dnipropetrovsk, and their Russian variations (Yekaterynoslav and Dnepropetrovsk). Kolmogorov and Aleksandrov refer to it as Dnepropetrovsk, but even in the Soviet times the city was widely known as Dnipro (or Dnepr in Russian). Since 2016 Dnipro is the official name.

Principle B If $P(A)$ is very small, then one can be practically certain that the event A will not occur on a single realization of $\mathfrak{S}$.

What is implicit in Principles A and B is that the event A should be prespecified, i.e., chosen prior to the experiment.

Remark 1. The title of Chapter I (of which the *Verhältnis* forms a part) is “Elementary theory of probability”, and at the beginning of the chapter Kolmogorov explains that elementary theory of probability is defined as part of the theory in which we have to deal with probabilities of only a finite number of events. It appears that this restriction is applicable only to the mathematical theory and not to its informal motivation in the *Verhältnis*. Unlimited repetition means that the number of trials is potentially infinite. Fixing the number of repetitions in advance would contradict this requirement, and even fixing a finite stopping time would contradict it (by the finiteness of the number of events considered at each step combined with König’s lemma). For the purposes of Principles A and B, however, the numbers of trials appear to be fixed; this is definitely true for B (one trial) and likely true for A (n chosen in advance).

The next part derives informally from Principle A the axioms of finitely additive probability introduced in the previous section. For example, if events A_1 and A_2 are disjoint, the frequencies m , m_1 , and m_2 of occurrences of $A_1 \cup A_2$, A_1 , and A_2 , respectively, in n repetitions will satisfy

$$\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n},$$

and so “it appears appropriate to set” $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. This is the informal derivation of Axiom V in [7, Sect. I.1]; analogous arguments work for all other axioms.

Having both Principles A and B is unusual, and Kolmogorov’s predecessors often assumed only Principle B, sometimes referred to as Cournot’s principle. Once we assume Principle B, Principle A seems to follow by the law of large numbers, a mathematical statement (details will be given below). Such statements were made, e.g., on numerous occasions by Borel and Lévy [14, Sect. 5.2.2]. This will be further discussed in Sect. 4.

The final part of the *Verhältnis* contains two remarks. Remark I says that if two assertions are both practically certain, then their conjunction is also practically certain, though with a little lower degree of certainty; however, if the number of assertions is very large, we cannot draw any conclusion whatsoever about the practical certainty of their conjunction from the practical certainty of each of them individually. This is a version of the requirement that the event A in Principle B should be prespecified, and it prevents what philosophers call the lottery paradox (see [21] for a review). Remark I appears uncontroversial.

Remark II starts by observing that while the axioms of the previous section imply that the impossible event $\emptyset$ has probability zero, the converse implication is not true. The rest of Remark II claims much more [14, Sect. 5.2.1]:

By Principle B, the event A ’s having probability zero implies only that it is practically impossible that it will happen on a particular

unrepeated realization of the conditions $\mathfrak{S}$. This by no means implies that the event A will not appear in the course of a sufficiently long series of experiments. When $P(A) = 0$ and n is very large, we can only say, by Principle A, that the quotient m/n will be very small—it might, for example, be equal to $1/n$.

In the second Russian edition the wording of Remark II became less emphatic, with the Russian equivalent of “by no means” (никоим образом) removed. It is interesting that Kolmogorov and Shiryaev also dropped mentioning the ratio $1/n$ as an example of non-zero m/n .

Remark II appears to contradict the modern intuition of probability. If a prespecified event A pertaining to one trial has zero probability, observing it in a sequence of trials, finite or infinite, should also have zero probability, no matter whether the trials are independent or not. This will be further discussed in Sect. 5.

4 Why two principles?

Can we really derive Principle A from Principle B and a mathematical statement? Doing so requires applying mathematical probability theory to a combination of n experiments rather than one experiment. Strictly speaking, Kolmogorov’s conditions $\mathfrak{S}$ are for one experiment, not for a sequence of experiments. Are the experiments in Principle A combinable in such a way that Kolmogorov’s empirical picture is applicable to the combined experiment? (See, e.g., [14, Sect. 5.2.2].)

In our current context, there are two basic kinds of combinability: under A-combinability, we are allowed to apply Principle A to the combination, and under B-combinability, we are allowed to apply Principle B. We can also talk about full combinability, where the combination is subject to both principles.

B-combinability looks to me the simplest and most natural kind. This is what we need for the derivation of Principle A from Principle B under the additional assumption that the combined experiment is described by the probability measure P^n , where P is the probability measure (perhaps only finitely additive at this point in Kolmogorov’s exposition) describing one trial. Kolmogorov’s proof of the law of large numbers in the *Grundbegriffe* [7, Sect. VI.3, the part present in all editions] implies that the relative frequency of occurrences of A in n trials deviates from $P(A)$ by more than a given constant $\epsilon > 0$ with P^n probability at most

$$\frac{P(A)(1 - P(A))}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}. \quad (1)$$

For a small ϵ and large $n\epsilon^2$, we can derive Principle A from Principle B.

Remark 2. A-combinability requires repeating experiments that already consist of repeated experiments. This might look complicated but appear necessary under a frequentist interpretation of the law of large numbers, e.g., in the form of (1). In fact, von Mises takes an opposite approach (see, e.g., [18, p. 114 of the

English translation]). Instead of repeating infinitely many times an experiment repeated finitely many times, von Mises starts from the combination, a single infinite sequence, which is then cut into an infinite sequence of finite sequences of length n (using his operations of selection and combination). We can apply the same approach in the case of repeating finitely many times an experiment repeated finitely many times; this would be more consistent with Kolmogorov’s philosophy and even easier.

There are several possible explanations for Kolmogorov having both Principle A and Principle B in his book. The most basic explanation is that the trials in Kolmogorov’s picture are not even B-combinable. However, this answer seems to limit severely the applicability of probability theory, which runs against typical modern intuitions of probability.

Another explanation is that, even if Kolmogorov’s experiments are B-combinable, the probability measure (at this point of the book finitely additive) does not have to be a product measure; the component experiments do not have to be independent or, more generally, do not have to be IID. This was Shiryaev’s implicit suggestion in his [17, bottom of p. 324], and it was discussed more explicitly in [14, Sect. 5.2].

The third explanation is that the accuracy of zero probability diminishes as we increase the number of trials. As mentioned in [14, Sect. 2.2.3], there was no gulf between zero probability and merely small probability in the writings of Borel and his French colleagues; for them, the vanishingly small was merely an idealization of the very small. However, extending this to Kolmogorov seems to run counter to everything we know about his rigour.

To me, the most convincing explanation is that, even if we regard Principle A as derivable from Principle B, we still need the former to give an operational interpretation of probability. Principle A gives us a way of measuring probabilities in principle (ignoring the unknown accuracy of such measurements) and allows us to motivate Kolmogorov’s axioms (see also [1] and [14, Sect. 5.2.2]). In principle, from the modern philosophical point of view there is no need to have an operational interpretation of theoretical notions [3], but it may have been important for Kolmogorov, and the extra bonus is the justification of the first five axioms that it provides. The original title of the *Verhältnis* was “Empirische Begründung der Axiome” (Kolmogorov’s letter of 26 October 1932 in [16]; essentially this is still the name of a subsection, “Empirische Deduktion der Axiome”), which makes the presence of Principle A especially desirable. Perhaps having Principle A separately was also a pedagogical device employed by Kolmogorov in teaching his Dnipro students.

5 A finitely additive measure consistent with Remark II

The assumption of B-combinability appears as natural in the context of Remark II as in the context of deriving Principle A from Principle B in the pre-

vious section. On one hand, the number of trials in Remark II is potentially infinite, as discussed in Remark 1, but on the other hand, we do not need any assumptions on the dependence structure of the sequence of trials (remember that the notion of independence, the topic of [7, Sect. I.5], is not even introduced yet when discussing Principles A and B in the *Grundbegriffe*). If the potentially infinite sequence of experiments in Remark II is not B-combinable, the resulting theory appears very restrictive; why shouldn't it be applicable to a sequence of experiments?

If the complex of conditions $\mathfrak{S}$ is repeated a given finite number n of times, Remark II is not applicable, since the five axioms introduced in Sect. I.1 imply

$$P(A_1 \cup \dots \cup A_n) \leq \sum_{i=1}^n P(A_i) = 0$$

regardless of whether or not the trials are dependent, where A_i is the event (in the combined experiment) of observing the event A in the i th trial. The conclusion is also true when we fix in advance a finite stopping time, as discussed in Remark 1. Even if the sequence of trials is infinitely long, Remark II still fails under countable additivity. Perhaps this demonstrates Kolmogorov's contention in Chapter II that "it is almost impossible to elucidate" the empirical meaning of countable additivity.

Let us check that Remark II is compatible with the first five axioms, those that do have a clear empirical meaning according to Kolmogorov. First we replace a "potentially infinite" sequence of trials with an infinite sequence; this seems inevitable in mathematical considerations, and potential infinity will enter via our interest exclusively in events that are settled after finitely many trials. We need infinite sequences only because we do not have an everywhere finite stopping time chosen in advance.

Let us encode the appearance of the event A of probability zero as 1 and its failure to appear as 0. This makes our sample space $\Omega := \{0, 1\}^\infty$; define $X_n(\omega)$, $\omega \in \Omega$, as the n th element of ω . Let $\mathbb{N} := \{1, 2, \dots\}$. The following simple corollary of known results is a possible mathematical counterpart of Remark II with the individual trials regarded as B-combinable.

Proposition 3. *There exists a finitely additive probability measure P defined on all subsets of Ω such that $P(X_n = 1) = 0$ for each $n \in \mathbb{N}$ whereas $P(\exists n \in \mathbb{N} : X_n = 1) = 1$.*

Proof sketch. This follows immediately from Theorem 1 in [6] (stated in the appendix as Theorem 5), which gives a simple necessary and sufficient condition for the existence of an extension of a function defined on some subsets of Ω to all subsets of Ω . $\square$

In the appendix we will see what such a probability measure P may look like. It is impossible to really construct P , since its existence depends on the axiom of choice (or something similarly nonconstructive), but we can easily construct it when given a simple mathematical object (a free ultrafilter on $\mathbb{N}$).

When observing the trials sequentially, we can stop when the A of Remark II happens (by Principle B this is practically certain under P), and then the ratio m/n will be $1/n$. While Axiom VI, introduced later, in Sect. II.1, makes the appearance of A practically impossible, without it we can make the appearance of A inevitable. Let us state formally this and other properties of P .

Proposition 4. *Let P be any finitely additive probability measure as in Proposition 3 and let $\tau(\omega) := \min\{n : \omega_n = 1\}$ (with $\min \emptyset := \infty$). Then:*

- (a) *P agrees with the point mass at $(0, 0, \dots)$ on every event determined by finitely many trials; in particular, $X_1, X_2, \dots$ are independent (in the finitely additive sense) under P .*
- (b) *$P(\tau > n) = 1$ for every $n \in \mathbb{N}$, whereas $P(\tau < \infty) = 1$.*

The proof is spelled out in detail in the appendix. Part (a) says the model is not buying Remark II at the price of exotic dependence: the trials may be taken independent, and every prediction about a fixed finite horizon is exactly the one the classical model makes. Part (b) gives Kolmogorov's $1/n$: at the stopping time τ , which is P -almost surely finite but larger than every fixed n with probability one, the observed frequency is exactly $1/\tau$.

6 Conclusion

This note reflects my attempts to reconcile Kolmogorov's thinking with modern views, and it pays particular attention to the aspects of the *Verhältnis* that have looked puzzling to me. For a broader perspective, the reader may consult von Plato [20] and Shafer and Vovk [14].

In his preface to the second Russian edition of the *Grundbegriffe* Kolmogorov points out a fundamental limitation of the *Verhältnis* and draws the reader's attention to the new approach to the relation of probability theory to reality sketched in his papers [8] and [9]. But this is a different story.

Acknowledgments

Many thanks to Glenn Shafer for numerous discussions of the *Grundbegriffe* including [7, Sect. I.2, including Remark II] and to Sasha Shen for a more recent discussion of [7, Sect. I.2, Remark II]. In this note I used Shen's comparison of the wordings of Remark II in the *Verhältnis* in different Russian editions.

I acknowledge the use of Claude Opus 5 in exploring proof ideas, which I reviewed carefully, and for checking the note. I take full responsibility for this note's claims and statements, including mathematical statements and their proofs.

References

- [1] Maurice S. Bartlett. Probability in logic, mathematics and science. *Dialectica*, 3:104–113, 1949.
- [2] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, New York, 1983.
- [3] Hasok Chang. Operationalism. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Autumn 2021 edition, 2021.
- [4] Robert Goldblatt. *Lectures on the Hyperreals: An Introduction to Non-standard Analysis*. Springer, New York, 1998.
- [5] Alan Hájek. Interpretations of probability. In Edward N. Zalta and Uri Nodelman, editors, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2023 edition, 2023.
- [6] Joseph B. Kadane and Anthony O’Hagan. Using finitely additive probability: uniform distributions on the natural numbers. *Journal of the American Statistical Association*, 90:626–631, 1995.
- [7] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. Russian translation: Основные понятия теории вероятностей. ОНТИ, Moscow, 1936. English translation: *Foundations of the Theory of Probability*. Chelsea, New York, 1950. Second English edition: Chelsea, New York, 1956. Second Russian edition: Наука, Moscow, 1974. Third Russian edition: ФАЗИС, Moscow, 1998.
- [8] Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–7, 1965. Russian original: Три подхода к определению понятия “количество информации”.
- [9] Andrei N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, IT-14:662–664, 1968. Russian version: К логическим основам теории информации и теории вероятностей, published in Проблемы передачи информации, 1969.
- [10] Péter Komjáth and Vilmos Totik. Ultrafilters. *American Mathematical Monthly*, 115:33–44, 2008.
- [11] Sergei M. Nikol’skii. Aleksandrov and Kolmogorov in Dnepropetrovsk. *Russian Mathematical Surveys*, 38(4):41–55, 1983.
- [12] Sergei M. Nikol’skii. In memory of A. N. Kolmogorov. In *Kolmogorov in Perspective*, pages 109–116. American Mathematical Society, 2000.
- [13] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It’s Only a Game!* Wiley, New York, 2001.

- [14] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov’s *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006.
- [15] Albert N. Shiryaev, editor. *Truth is a blessing: Biobibliography (in Russian, Истина-благо: Биобиблиография)*, volume 1 of *Колмогоров: юбилейное издание в трех книгах*. Физматлит, Moscow, 2003.
- [16] Albert N. Shiryaev, editor. *The braid of these running lines...: Selected passages from the correspondence between A. N. Kolmogorov and P. S. Aleksandrov (in Russian, Этух строк бегущих тесьма...: Избранные места из переписки А. Н. Колмогорова и П. С. Александрова)*, volume 2 of *Колмогоров: юбилейное издание в трех книгах*. Физматлит, Moscow, 2003.
- [17] Albert N. Shiryaev. Development of mathematical theory of probability: Historical review. In *Albert N. Shiryaev, Probability-2*, pages 313–331. Springer, New York, 2019. A version of this essay was published as “Математическая теория вероятностей. Очерк истории становления” in the third Russian edition of Kolmogorov’s *Grundbegriffe* [7].
- [18] Richard von Mises. *Wahrscheinlichkeit, Statistik und Wahrheit*. Springer, Berlin, 1928. English translation: *Probability, Statistics and Truth*. William Hodge, London, 1939.
- [19] Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik (in German, Probabilities and their Applications in Statistics and Theoretical Physics)*, volume 1 of *Vorlesungen aus dem Gebiete der angewandten Mathematik (Lectures on Applied Mathematics)*. Franz Deuticke, Leipzig and Vienna, 1931. This book was the only volume in the series *Lectures on Applied Mathematics*.
- [20] Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*. Cambridge University Press, Cambridge, 1998.
- [21] Gregory Wheeler. A review of the lottery paradox. In Gregory Wheeler and William Harper, editors, *Probability and Inference: Essays in Honour of Henry E. Kyburg Jr.*, pages 1–31. College Publications, London, 2007.

A Some details

In this appendix I will give a more intuitive argument for Proposition 3 that is also more explicit (it is constructive modulo a given free ultrafilter $\mathfrak{U}$ on $\mathbb{N}$, although such a $\mathfrak{U}$ cannot be constructed itself, as mentioned earlier). In order to make the exposition self-contained, its second section states Theorem 1 of [6] and gives the details of the derivation of Proposition 3. Finally, a detailed proof of Proposition 4 is also provided.

A.1 An explicit P in Proposition 3

The notion of a *free ultrafilter* (on $\mathbb{N}$) is very intuitive, but let us start from a formal definition. We say that a family $\mathfrak{U}$ of subsets of $\mathbb{N}$ is a *filter* if $\emptyset \notin \mathfrak{U}$, $\mathbb{N} \in \mathfrak{U}$, a superset of any element of $\mathfrak{U}$ is in $\mathfrak{U}$, and the intersection of any two elements in $\mathfrak{U}$ is in $\mathfrak{U}$. A filter $\mathfrak{U}$ is an *ultrafilter* if for every $A \subseteq \mathbb{N}$, either A or its complement is in $\mathfrak{U}$. An ultrafilter $\mathfrak{U}$ is *free* if it does not contain any singleton.

Fix a free ultrafilter $\mathfrak{U}$ on $\mathbb{N} := \{1, 2, \dots\}$. Let us say that a set $A \subseteq \mathbb{N}$ is *big* if $A \in \mathfrak{U}$; otherwise, it is *small*. The definition of a free ultrafilter requires that every singleton be a small set, subsets of small sets be small, finite unions of small sets be small, and the complements of small sets be big and vice versa.

Let e_k , $k \in \mathbb{N}$, be the sequence in $\Omega = \{0, 1\}^\infty$ that has a 1 only at the k th position. Define P to be the finitely additive measure concentrated on $\{e_k : k \in \mathbb{N}\}$ such that, for any $E \subseteq \Omega$,

$$P(E) := \begin{cases} 1 & \text{if } \{k : e_k \in E\} \in \mathfrak{U} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

In other words, $P(E) = 1$ if E contains many e_k (in the sense of $\{k : e_k \in E\}$ being big) and $P(E) = 0$ if E contains few e_k (in the sense of $\{k : e_k \in E\}$ being small). The intuition behind P is that it expresses the belief that the sequence of outcomes will be one of the e_k (i.e., will contain only one 1) and that we believe that $k \notin A$ for a prespecified small $A \subseteq \mathbb{N}$ while we believe that $k \in A$ for a prespecified big $A \subseteq \mathbb{N}$. It is obvious that P satisfies the conditions of Proposition 3.

The finitely additive probability measure (2) is not very interesting in that it takes only two values, 0 or 1. To make its range $[0, 1]$, we can modify (2) using the notion of the “standard part” of a sequence of real numbers; before defining it formally, let me explain how it can be used in constructing P . The standard part of a sequence is its generalized limit; it always exists in $\mathbb{R}$ and coincides with the limit when it exists. We define $P(E)$ as the standard part of the sequence $(p_1, p_2, \dots)$, where

$$p_n := |\{k \leq n : e_k \in E\}| / n \quad (3)$$

is the relative frequency of $e_k \in E$ over the first n steps. We can say that $P(E)$ is the limiting relative frequency of $e_k \in E$, $k \in \mathbb{N}$.

Let us now define P formally. A *hyperreal (number)* is defined to be a sequence of real numbers, where hyperreal numbers $(a_1, a_2, \dots)$ and $(b_1, b_2, \dots)$ are regarded as equivalent if $a_k = b_k$ for many k (formally, a hyperreal number can be also defined as an equivalence class). The linear order on the hyperreals is defined in the same way: $(a_1, a_2, \dots) \leq (b_1, b_2, \dots)$ means that $a_k \leq b_k$ for many k . Real numbers can be embedded into the hyperreals as $a \in \mathbb{R} \mapsto (a, a, \dots)$. It is clear that for any hyperreal $p = (p_1, p_2, \dots)$ with all $p_k \in [0, 1]$ (such as (3)) there exists a unique real number $p' \in [0, 1]$ such that, for any real $\epsilon > 0$,

$|p - p'| < \epsilon$ (where the operations $|\cdot|$ and $-$ are defined for hyperreals term-wise). Such p' is known as the *standard part* of p . This completes the definition of P ; checking that it satisfies the conditions of Proposition 3 is straightforward.

An excellent source about ultrafilters (and Banach limits, which can be used in place of taking the standard part) is [10]. For a short self-contained description of ultrafilters and hyperreals, see the first subsection of Sect. 11.5 in the book [13], which applies them in the foundations of probability. For a detailed exposition of non-standard analysis (including the hyperreals) based on ultrafilters, see [4].

A.2 Details of the deduction of Proposition 3 from a general criterion

Proposition 3 follows from the following necessary and sufficient condition for the possibility of extending a given function on subsets of Ω to all subsets.

Theorem 5 ([6], Theorem 1). *Let $\mathcal{C}$ be any collection of subsets of a set Ω with $\Omega \in \mathcal{C}$, and let P be a nonnegative real-valued function on $\mathcal{C}$ with $P(\Omega) = 1$. Then P can be extended to a finitely additive probability measure on all subsets of Ω if and only if, for all finite lists $C_1, \dots, C_c$ and $C'_1, \dots, C'_{c'}$ of members of $\mathcal{C}$,*

$$\left(\sum_{i=1}^c 1_{C_i} \leq \sum_{j=1}^{c'} 1_{C'_j} \right) \implies \left(\sum_{i=1}^c P(C_i) \leq \sum_{j=1}^{c'} P(C'_j) \right), \quad (4)$$

where 1_A stands for the indicator function of A .

The direction that we need for Proposition 3 (namely, the “if” direction) is actually stated already in [2, Theorem 3.2.10].

Let us now deduce Proposition 3 from Theorem 5. Let $\mathcal{C}$ consist of the sets Ω , $Z := \{(0, 0, \dots)\}$, and $\{X_n = 1\}$, $n \in \mathbb{N}$. It suffices to check (4), where $P(\Omega) := 1$, $P(Z) := 0$, and as before, $P(\{X_n = 1\}) = 0$ for all $n \in \mathbb{N}$. Arguing indirectly, suppose (4) is violated for some lists. Without loss of generality, we can assume that all C'_j are either Z or of the form $\{X_n = 1\}$ and that $c = 1$ and $C_1 = \Omega$. We arrive at a contradiction by noticing that the antecedent of (4) is false at e_i for sufficiently large i .

A.3 Proof of Proposition 4

(a) Let E be determined by the trials in a finite set $F \subseteq \mathbb{N}$, and let $C := \{\omega : \omega_n = 0 \text{ for all } n \in F\}$. The complement of C is $\bigcup_{n \in F} \{X_n = 1\}$, a finite union of events of probability zero, so $P(C) = 1$. If $(0, 0, \dots) \in E$ then $C \subseteq E$ and $P(E) = 1$; otherwise $C \subseteq E^c$ and $P(E) = 0$. Independence follows since every finite-dimensional cylinder is such an E , and its probability, 1 or 0 according as it requires no 1 or some 1, agrees with the product of the marginals.

(b) $\{\tau > n\} = \{X_1 = \dots = X_n = 0\}$ has probability 1 by (a), and $\{\tau < \infty\} = \{\exists n \in \mathbb{N} : X_n = 1\}$ has probability 1 by assumption.