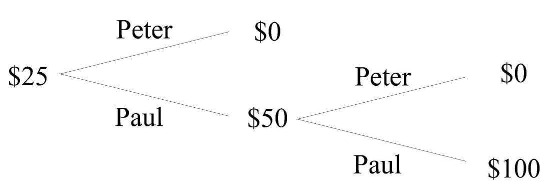
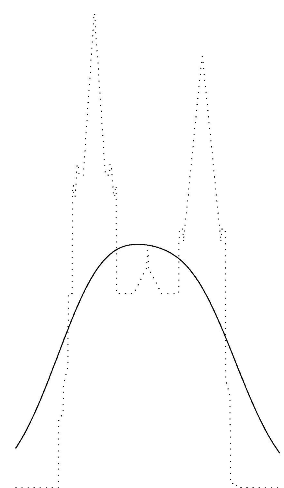
The non-mathematical section of Kolmogorov's *Grundbegriffe*

Vladimir Vovk





The Game-Theoretic Probability and Finance Project

Working Paper #69

First posted August 18, 2026. Last revised August 24, 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, and the possibility of observing a prespecified event of probability zero in a sequence of experiments.

A version of this paper is published as an arXiv technical report: http://arxiv.org/abs/2608.17793.

Contents

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., [6]), but it also includes a two-page section about the philosophy of probability. The title of this section, [8, 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 [20]. 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 Kolmogorov calls Principle A and Principle B, and the possibility of observing a prespecified event of probability zero [8, Sect. I.2, Remark II]. Both unusual features are related to the possibility of combining several experiments satisfying Kolmogorov's conditions into one that still satisfies them, or some of them. This will be covered in Sects. 4 and 5. The main finding of Sect. 5 is that a natural interpretation of Remark II in [8, Sect. I.2] is consistent with Kolmogorov's Axioms I–V and inconsistent with Axiom VI (that of continuity, equivalent to countable additivity in the presence of the other axioms).

Remark 1. In this note I translate Kolmogorov's [8] "Versuch" as "experiment" ("trial" is a strong competitor). "Experiment" is used consistently in the English translation of the *Verhältnis* in [16, Sect. 5.2.1], while the other translations listed in the bibliographic entry for [8] are less consistent in this section (in Sect. I.5 on independence Kolmogorov introduces a completely different notion of "Versuch"). Morrison's English translation uses both "experiment" and "test", and the Russian translations use both "опыт" and "испытание".

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

An important source of information about the *Grundbegriffe* is the volume [18] of correspondence between Kolmogorov and Aleksandrov (the title of the volume is a line from Kolmogorov's poem devoted to Aleksandrov [17, preface, p. 14]). Kolmogorov started writing the *Grundbegriffe* on 26 October 1932 (letter to Aleksandrov of this date, [18]) 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 the translation was updated in 1974 by another of his students, Albert Shiryaev. The second Russian edition modernized Kolmogorov's notation, and Shiryaev also extended the last three sections [8, 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 [19]).

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 [16, Sect. 5.2.1]; I will use it when quoting (directly and indirectly) from the *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. Axioms I–V introduce a basic set function P and require it to be a finitely additive probability measure, while Axiom VI adds the requirement of P's countable additivity.

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 lecture notes (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 [18]. The whole of Chapter I was to be largely identical (довольно точно будет совпадать) to the Dnipro lecture notes. Kolmogorov had been visiting Dnipro twice a year since Autumn 1931 [13, p. 43], in Spring and Autumn, and this continued, perhaps less regularly, until 1940 [14, p. 113]. His last visit before starting the *Grundbegriffe* ended on 26 September 1932 (his letter to Aleksandrov of 24 September announces his departure), and he visited Dnipro again during the *Grundbegriffe* production, lecturing on 4–10 March 1933 mainly on probability theory (his 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 [16, 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.

[1]Dnipro was, and still is, Ukraine's fourth largest city (after Kharkiv, which was the capital at the time, Kyiv, and Odesa). It has been known under many names, including Katerynoslav, Dnipropetrovsk, and their Russian variations (Yekaterinoslav 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 has been the official name.

In describing the way the theory of probability is applied to the real world, Kolmogorov talks about a system 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 $P(A)$ (necessarily $P(A) \in [0,1]$) 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 $P(A)$.

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 2. 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*. Let me take "unlimited repetition" to mean that the number of experiments is potentially infinite. Fixing the number of repetitions in advance would then contradict the requirement of unlimited repetition, and even fixing an everywhere finite stopping time would contradict it (the finiteness of the number of events considered at each step combined with Kőnig's lemma makes the stopping time bounded). For the purposes of Principles A and B, however, the numbers of experiments appear to be fixed; this is definitely true for B (one experiment) 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, [8, Sect. I.1]. 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 [8, 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 [16, Sect. 5.2.2]. This will be further discussed in Sect. 4.

The final part of the *Verhältnis* contains two remarks. The main part of 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 [22] for a review). The very end of Remark I will be discussed in the next section. In general, Remark I appears uncontroversial.

Remark II starts by observing that while the axioms of the previous section [8, Sect. I.1] imply that the impossible event $\emptyset$ has probability zero, the converse implication is not true. The rest of Remark II claims much more [16, 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 the mention of 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 experiment has zero probability, observing it in a sequence of experiments, finite or infinite, should also have zero probability, no matter whether the experiments 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. 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., [16, 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.

Kolmogorov does not hesitate to apply Principle A to a combination of experiments in the last sentence of Remark I: "it in no way follows from Principle A that m/n will differ only a little from $P(A)$ in every one of a very large number of series of experiments, where each series consists of n experiments" [16, Sect. 5.2.1]. He freely contemplates repetitions of the combined experiment thus assuming A-combinability.

B-combinability looks to me even simpler and more natural. It allows us to derive 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 experiment. Kolmogorov's proof of the law of large numbers in the *Grundbegriffe* [8, the part of Sect. VI.3 present in all editions] implies that the relative frequency of occurrences of A in n experiments deviates from $P(A)$ by more than a given constant $\epsilon > 0$ with P^n probability at most

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

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

There are several possible explanations for Kolmogorov having both Principle A and Principle B in his book. The most basic explanation is that the experiments 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 P^n; the component experiments do not have to be independent, let alone IID. This was Shiryaev's implicit suggestion in [19, bottom of p. 324], and it was discussed more explicitly in [16, Sect. 5.2].

The third explanation is that the accuracy of zero probability diminishes as we increase the number of experiments. As mentioned in [16, 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 [2] and [16, 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 [4], but it may have been important for Kolmogorov, with the extra bonus of justifying the first five axioms. The original title of the *Verhältnis* was "Empirische Begründung der Axiome" (Kolmogorov's letter of 26 October 1932 in [18]; essentially this is still the name of a subsection, "Empirische Deduktion der Axiome"), which makes the presence of Principle A especially desirable. Another role of Principle A might have been serving as a pedagogical device, originally 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 previous section. On the one hand, the number of experiments in Remark II is potentially infinite, as discussed in Remark 2; on the other hand, we do not need any assumptions on the dependence structure of the sequence of experiments (remember that the notion of independence, the topic of [8, Sect. I.5], has not even been introduced yet at the point where Principles A and B are discussed 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? Let us assume that it is B-combinable.

If the system of conditions $\mathfrak{S}$ is repeated a given finite number n of times, Remark II is not applicable, since the five axioms introduced in [8, Sect. I.1] imply

$$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i) = 0 \tag{1}$$

regardless of whether or not the experiments are dependent, where A_i is the event (in the combined experiment) of observing the event A in the ith experiment. The conclusion is also true when we fix in advance an everywhere finite stopping time, as discussed in Remark 2. Even if the sequence of experiments is infinitely long, Remark II still fails under countable additivity, which allows us to set $n := \infty$ in (1).

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 experiments with an infinite sequence; this seems inevitable in mathematical considerations, and potential infinity will enter via our interest exclusively in events being settled after finitely many experiments. 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 nth 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 experiments 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 [7] (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 a finitely additive probability measure defined on all subsets of Ω. □

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 experiments 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 [8, Sect. II.1], makes the appearance of A practically impossible, without it we can make the appearance of A practically certain. Let me 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 experiments; 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 that Remark II is not bought at the price of exotic dependence: the experiments 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 $1/\tau$.

It is interesting to contrast the incompatibility between Remark II and Kolmogorov's Axiom VI, which is equivalent to countable additivity and allows us to set $n := \infty$ in (1), with his contention that "it is almost impossible to elucidate" the empirical meaning of countable additivity and that requiring countable additivity is an arbitrary (willkürlich, произвольный) choice [8, Sect. II.1]. This choice still has empirical implications: accepting the new axiom forces us to reject the postulated probability model when encountering a prespecified event of probability zero in a sequence of experiments.

Of course, we cannot know whether invoking finite additivity is the right way to understand Remark II in the *Verhältnis*. If it is, the existence of P satisfying the condition in Proposition 3 should have been obvious to Kolmogorov as it easily follows from the Hahn–Banach theorem (see Sect. A.2 in the appendix). The theorem and its various corollaries are covered in detail in Banach's book [1, starting from Sect. II.2], and this book was a subject of correspondence between Kolmogorov and Aleksandrov in October 1932 [18, Kolmogorov's letter of 15 October and Aleksandrov's letter of 21 October], the month when Kolmogorov started his work on the *Grundbegriffe*.

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 [21] and Shafer and Vovk [16].

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 [10] and [11]. But this is a different story.

Acknowledgments

Many thanks to Glenn Shafer for numerous discussions of the *Grundbegriffe*, including [8, Sect. I.2, including Remark II], and for his comments on this note. I am grateful to Sasha Shen for discussions of [8, 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. Thanks to Jan von Plato for his attention to my work. The Bodleian Library team have been extraordinarily helpful.

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

References

[1] Stefan Banach. *Théorie des opérations linéaires*. Hafner, New York, 1932. Reviewed by Kolmogorov [9]. The Polish (much shorter) original: *Teoria operacji*, vol. 1, *Operacje liniowe*, wyd. Kasy im. Mianowskiego, Warszawa, 1931. Ukrainian translation: Курс функціонального аналізу: (Лінійні операції), Радянська школа, Київ, 1948.

[2] Maurice S. Bartlett. Probability in logic, mathematics and science. *Dialectica*, 3:104–113, 1949.

[3] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, London, 1983.

[4] Hasok Chang. Operationalism. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Autumn 2021 edition, 2021.

[5] Robert Goldblatt. *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis*. Springer, New York, 1998.

[6] 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.

[7] 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.

[8] 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.

[9] Andrei N. Kolmogorov. Review of S. Banach, *Théorie des opérations linéaires*. *Zentralblatt für Mathematik und ihre Grenzgebiete*, 5:209, 1933. Zbl 0005.20901.

[10] 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".

[11] 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.

[12] Péter Komjáth and Vilmos Totik. Ultrafilters. *American Mathematical Monthly*, 115:33–44, 2008.

[13] Sergei M. Nikol'skii. Aleksandrov and Kolmogorov in Dnepropetrovsk. *Russian Mathematical Surveys*, 38(4):41–55, 1983.

[14] Sergei M. Nikol'skii. In memory of A. N. Kolmogorov. In *Kolmogorov in Perspective*, pages 109–116. American Mathematical Society, 2000.

[15] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001.

[16] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006.

[17] 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.

[18] 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, Этих 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.

[19] 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* [8].

[20] Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik (in German, Probability Calculus and its Application 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*.

[21] Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*. Cambridge University Press, Cambridge, 1994.

[22] 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 [7] 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 $\mathbb{N}$. 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 kth position. Define P to be the finitely additive measure concentrated on

$\{e_k : k \in \mathbb{N}\}$ such that, for any $E \subseteq \Omega$,

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

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 the position k of the single 1 is outside 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 $\overline{\mathbb{R}}$ 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

$$p_n := |\{k \le n : e_k \in E\}| / n \tag{3}$$

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: $(a_1, a_2, \dots) < (b_1, b_2, \dots)$ means that $a_k < b_k$ 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 termwise). Such p' is known as the *standard part* of p. This completes the definition of P described informally in the previous paragraph; 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 [12]. For a short self-contained description of ultrafilters and hyperreals, see the first subsection of Sect. 11.5 in the book [15], which applies them in the foundations of probability. For a detailed exposition of non-standard analysis (including the hyperreals) based on ultrafilters, see [5]. See [7, Theorem 3] for an alternative definition, not using ultrafilters or non-standard analysis, of limiting relative frequency.

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 a finitely additive probability measure on all subsets.

Theorem 5 ([7], 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 $\mathcal{C}$,*

$$\left(\sum_{i=1}^{c} 1_{C_i} \le \sum_{j=1}^{c'} 1_{C'_j}\right) \Longrightarrow \left(\sum_{i=1}^{c} P(C_i) \le \sum_{j=1}^{c'} P(C'_j)\right), \tag{4}$$

where 1_A stands for the indicator function of A.

For the direction that we need in the proof of Proposition 3, namely, the "if" statement, [7] simply refers to [3, Theorem 3.2.10]. Alternatively, the "if" statement can be easily deduced from the Hahn–Banach theorem: the condition (4) allows us to extend P to a linear functional on the linear space of finite linear combinations of the indicator functions of members of $\mathcal{C}$; this linear functional is bounded above by the supremum sublinear functional $f \mapsto \sup f$; and finally, the Hahn–Banach theorem allows us to extend this linear functional to all bounded functions so that it is still bounded above by the supremum functional. The resulting extension will take nonnegative values on nonnegative functions f (since for such f the supremum of $-f$ is nonpositive), and we can set P to its values on the indicator functions.

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_k for sufficiently large k.

A.3 Proof of Proposition 4

(a) Let E be determined by the experiments 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 $C \subseteq E^{\mathrm{c}}$ 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.