KOLMOGOROV ON THE ROLE OF RANDOMNESS IN PROBABILITY THEORY

CHRISTOPHER P. PORTER

ABSTRACT. In this article, I discuss the extent to which Kolmogorov drew upon von Mises' work in addressing the problem as to why probability is applicable to events in the real world, which I refer to as the *applicability problem*. In particular, I highlight the role of randomness in Kolmogorov's account, and I argue that this role differs significantly from the role that randomness plays in von Mises' account.

1. INTRODUCTION

Kolmogorov's fundamental contribution to the development of algorithmic randomness is well-known. In particular, his definition of randomness for finite strings, given in terms of what is nowadays referred to as *Kolmogorov complexity*, has proven to be a highly fruitful notion in theoretical computer science. However, Kolmogorov's primary motivation in formulating his definition was rooted in the foundations of probability theory. For in Kolmogorov's view, in the foundations of probability, "[t]he problem is to describe the reasons why mathematical probability theory can be applied at all to phenomena of the real world" ([Kol83], pg. 1) Let us call this problem the *applicability problem.*

In a number of passages, we find Kolmogorov making claims such as:

(i) "In establishing the premises necessary for the applicability of the theory of probability to the world of actual events, the author has used, in large measure, the work of R. v. Mises" ([Kol56a], pg. 3, fn. 4).

(ii) "[T]he basis of the applicability of the mathematical theory of probability to random events of the real world is the *frequency approach to probability* in one form or another, which was so strongly advocated by von Mises ([Kol63], pgs. 176-177); and

(iii) "Following von Mises, it is often claimed that the basis of the applications of probability theory is the hypothesis of stability of frequencies. In a form closer to practice, this notion is also accepted in my well-known book on the foundations of probability theory" ([Kol83], pg. 213).

Clearly, then, Kolmogorov's approach to the applicability problem is based in some way on the frequency theory of Richard von Mises, according to which the probability of an event is the limiting relative frequency of that event in an infinite sequence of trials. Thus we can ask:

$\mathcal{Q}_1$: To what extent did Kolmogorov draw upon von Mises' frequency theory to address the applicability problem?

Elsewhere Kolmogorov claims that another ingredient that is necessary to address the applicability problem is a definition of randomness, for instance, writing in one of his articles

---

that a certain notion of partial randomness should be ruled out as not being "the type of 'randomness' that is needed to make the application of probability theory possible" ([Kol83], pg. 213). Given that von Mises' account includes a definition of randomness, one might further ask:

$\mathcal{Q}_2$: Is the definition of randomness that Kolmogorov took to be necessary to address the applicability problem also drawn from the work of von Mises?

My goal in this article is thus two-fold: (i) to answer $\mathcal{Q}_1$ by outlining the ways in which Kolmogorov drew upon von Mises' work in addressing the applicability problem, and (ii) to answer $\mathcal{Q}_2$ by arguing that although Kolmogorov's initial definition of randomness was inspired by von Mises' definition, there is a significant difference between the *role* that randomness plays in the account of von Mises and in that of Kolmogorov.

2. Von Mises' Frequentist Definition of Probability

Before we can address $\mathcal{Q}_1$ and $\mathcal{Q}_2$, our first order of business is to consider von Mises's frequency approach to probability, which is commonly referred to as *hypothetical frequentism*. In particular, my goal in this section is to identify the role that randomness plays in von Mises' account. To this end, we'll have to consider a number of details of von Mises' definition in order to clearly identify this role.

In 1919, von Mises presented what he considered to be a "scientifically adequate" definition of probability in his paper "Grundlagen der Wahrscheinlichkeitsrechnung", [vM19]. For von Mises, a scientifically adequate definition of probability is only applicable in those cases in which we "have a practically unlimited sequence of uniform observations", and thus single-case probabilities, assigned to events that do not occur in a sequence of trials, fall outside of the scope of von Mises' account ([vM81], pg. 11). This restriction of probability to those events that occur a potentially unlimited number of times was enforced by von Mises' definition of probability in terms of what he called *collectives*. Roughly, a collective is (i) an infinite sequence $X = x_0 x_1 x_2 \ldots$ consisting of what von Mises refers to as *attributes* (where each of the attributes x_i belongs to a fixed collection A) that (ii) satisfies certain conditions of randomness (which will be specified shortly). Thus, given an attribute $a \in A$, we can only determine the probability of a insofar as it is a member of some collective.[1]

Given a collective X made up of the attributes from A (which I'll write as $X \in A^\omega$), we determine the probability of an attribute a in X by computing the relative frequency of a in X, where the relative frequency of an attribute in a collective is defined as follows. Let $X = x_0 x_1 x_2 \ldots$ be a collective in which the attributes $a_0, a_1, \ldots, a_n, \ldots$ occur, where the collection of attributes can be either finite or infinite; that is, let X be an infinite sequence of elements from the set $A = \{a_0, a_1, \ldots, a_n, \ldots\}$. Then the relative frequency of the attribute a_j in the initial segment of X of length n is the value

$$\frac{\#\{i < n : x_i = a_j\}}{n}.$$

---

[1]It's not altogether clear what von Mises takes an attribute to be: in some cases, he speaks of attributes as observations or events, but in other cases, attributes seem to function as labels that we affix to given events. The most reasonable approach is to consider events as attribute-instances; strictly speaking, we don't consider the probability of a given attribute, but rather the probability that an event will instantiate that attribute. I am indebted to Antony Eagle for this latter point.

To compute the probability of the attribute a_j in the collective X, we compute the limit of the relative frequency of a_j in the initial segments of X. That is, the probability of the attribute a_j in the collective X is

$$
(1) \qquad p_j = \lim_{n \to \infty} \frac{\#\{i < n : x_i = a_j\}}{n},
$$

which we'll refer to as *the limiting relative frequency of a_j in X*. In this way, we get a probability distribution on the collection $\{a_0, a_1, \ldots, a_n, \ldots\}$.

Thus for von Mises, probability is relative frequency. This much of von Mises' account was not novel, for as he notes the frequency theory of probability had already appeared in previous works, most notably in John Venn's 1866 *The Logic of Chance* [Ven66]. What *is* novel about von Mises' frequentist account of probability is the role that randomness plays in his account. Specifically, von Mises required that collectives satisfy a condition of randomness, which is enforced by the second of his two axioms of collectives. Let us thus turn to von Mises' two axioms of collectives.

As motivation for von Mises' first axiom of collectives, note that there are some infinite sequences of events consisting of attributes for which the limit in (1) above does not exist. It thus follows that the probabilities of the attributes in those sequences are not defined. Thus, for an infinite sequence to count as a collective, it must satisfy von Mises' first axiom of collectives:

(VM_1) If $X = x_0 x_1 x_2 \ldots \in A^\omega$ is a collective, then for each $a_j \in A$,

$$
\lim_{n \to \infty} \frac{\#\{i < n : x_i = a_j\}}{n}
$$

exists.

According to von Mises, sequences satisfying (VM_1) still may not be appropriate for defining probabilities; such sequences must also satisfy a requirement of randomness. That is, the "limiting values must remain the same in all partial sequences which may be selected from the original one in an arbitrary way" ([vM81], pg. 25).[2] The partial sequences that we select from a given sequence are to be extracted by what von Mises calls *place selections*. Informally, a place selection is a method of extracting a subsequence from a given sequence. More precisely, it is a function $S : A^{<\omega} \to \{0, 1\}$ (where $A^{<\omega}$ is the collection of finite strings of members from A) that takes as input a finite initial segment of a sequence X and outputs a 0 or 1, indicating whether or not we should include the next value of X in the selected subsequence, which I'll write as X_S. For instance, we can define a place selection that selects every odd-indexed place of the sequence X, or one that selects every place that is preceded by two consecutive a's for some fixed $a \in A$.

Now, given a place selection S and a sequence $X = x_0 x_1 x_2 \ldots \in A^\omega$, once we apply S to X to extract a subsequence $X_S = x_{\ell_0} x_{\ell_1} x_{\ell_2} \ldots$, we can check to see if the limiting relative frequencies of the attributes of X_S are equal to those in X. Suppose that $X \in A^\omega$ satisfies

---

[2]Von Mises' use of the word 'arbitrary' here is misleading, since it looks as if he is requiring that the limiting values remain the same in subsequences selected from the original sequence in *any way whatsoever*. But this can't be what von Mises has in mind, for as will be discussed shortly, this requirement would result in a highly defective definition of randomness. Let us set aside this point for now, and fill out the rest of the details of von Mises' condition of randomness.

($\mathsf{VM_1}$), so that for each $a_j \in A$,

$$\lim_{n \to \infty} \frac{\#\{i < n : x_i = a_j\}}{n} = p_j.$$

Then the limiting relative frequencies of the selected subsequence $X_S = x_{\ell_0} x_{\ell_1} x_{\ell_2} \dots$ are the same as those of X if for each $a_j \in A$,

$$\lim_{n \to \infty} \frac{\#\{i < n : x_{\ell_i} = a_j\}}{n} = p_j.$$

In this case, we will write $\mathsf{relfreq}(X) = \mathsf{relfreq}(X_S)$ and we will say that the limiting relative frequencies of X are *invariant under the place selection S*.

There is one last step before we formulate the condition of randomness referred to above: we need to isolate a specific collection of place selections, referred to by von Mises as the *admissible* place selections. According to von Mises, a place selection is admissible only if the choice to select an attribute in a sequence does not depend on the value of the attribute, but only on the value of the previous attributes in the sequence, or the index of the given attribute in the sequence, or both.[3] Then, von Mises' second axiom of collectives, the so-called condition of randomness, can be formulated as follows:

($\mathsf{VM_2}$) If $X \in A^\omega$ is a collective, then for every admissible place selection S,

$$\mathsf{relfreq}(X) = \mathsf{relfreq}(X_S)$$

whenever X_S is an infinite sequence.[4]

Any sequence satisfying ($\mathsf{VM_1}$) and ($\mathsf{VM_2}$) is thus a collective; further, we can define the probability of the occurrence of an attribute *in a given collective* to be the limiting relative frequency of the occurrence of the attribute in that collective.

Now, there are two questions to be addressed before we can move on to Kolmogorov's account. First, is there a way to formulate von Mises' definition without an appeal to the nebulous notion of admissibility? Second, why did von Mises require that collectives satisfy the condition of randomness to begin with?

Concerning the first point, von Mises failed to provide a more precise account of admissibility, leaving his definition vulnerable to criticism. More specifically, it was claimed that von Mises' definition was inconsistent, a claim made by a number of von Mises' contemporaries, most notably Erich Kamke,[5] who argued that for any given sequence X, one could always

---

[3]The observant reader will note that I did not provide a sufficient condition for the admissibility of a place selection, as von Mises never provides any such condition. For the moment, let us set this concern aside and suppose that we have identified some collection of place selections as admissible.

[4]This last qualification is necessary since for some place selections S, X_S is finite.

[5]Kamke included criticisms of von Mises' axioms in his 1933 article "Über neuere Begründungen der Wahrscheinlichkeitsrechnung" [Kam33] and also in his book *Einführung in die Wahrscheinlichkeitstheorie* [Kam32].

find a place selection S such that the limiting relative frequencies of X are not invariant under S.[6]

Von Mises' eventual response to this criticism was inspired by a result proved by Abraham Wald, which Wald characterized as a proof of the consistency of collectives. What Wald showed was that for any countable collection $\mathscr{C}$ of place selections, there are continuum many sequences that are invariant under the place selections in $\mathscr{C}$.[7] In light of this result, von Mises dropped the condition of admissibility from his second axiom and instead defined collectives in a relative sense: For a countable collection $\mathscr{C}$ of place selections, a sequence X is a collective relative to $\mathscr{C}$ if the limiting relative frequencies of the attributes in X exist and X is invariant under the place selections is $\mathscr{C}$.

But which countable collection $\mathscr{C}$ of place selections should we use to define collectives? According to the later version of von Mises' theory, as laid out in his textbook *Mathematical Theory of Probability and Statistics*, for each problem in the probability calculus that we are attempting to solve, there is some countable collection $\mathscr{C}$ of place selections such that we can solve the given problem by means of collectives that are invariant under the place selections in $\mathscr{C}$. Thus, von Mises writes,

> We obtain a concrete idea of the set $[\mathscr{C}]$ of place selections which are supposed not to change the frequency limits if we visualize $[\mathscr{C}]$, for example, as follows: in $[\mathscr{C}]$ are contained all those place selections which present themselves in a particular problem under consideration ([vM64], pg. 12).

Thus, the later version of von Mises' theory avoids the inconsistency objection raised by his contemporaries. However, it's far from clear how it is that place selections can "present themselves in a particular problem under consideration". But this point is in fact related to the second question we raised above: What role does the condition of randomness play in von Mises' theory?

To answer this question, we need to consider how von Mises employed collectives to solve problems of the probability calculus. According to von Mises, "In a problem of the probability calculus, the data as well as the results are probabilities", a view summarized by the pithy phrase "the beginning and the end of each problem must be probabilities" ([vM81], pg. 32). Thus, for von Mises, a problem of the probability calculus begins with what is called an *initial collective*, to which we apply certain collective-preserving operations,[8] eventually resulting in what is called a *final collective*, from which we can extract the solution to the problem in question. Moreover, by means of these operations on collectives, we can show that certain laws of the probability calculus, such as the addition rule, the product rule, and the

---

[6]Kamke argued as follows: Suppose we are given a collective X made up of the attributes in A, and suppose further that there are attributes $a, b \in A$ that occur infinitely often in X. Then if p_a and p_b are the limiting relative frequencies of a and b, respectively, then it follows that either $p_a \neq 1$ or $p_b \neq 1$ (or both). For the sake of argument, let us assume that $p_a \neq 1$. If we let $n_1, n_2, \ldots$ be a sequence of natural numbers such that $X(n) = a$ if and only if $n = n_i$ for some $i \in \omega$ (that is, the sequence $(n_i)_{i \in \omega}$ is the set of positions at which X has the value a), then we can define a place selection S such that S only selects elements for the subsequence at the indices $(n_i)_{i \in \omega}$. Thus when given the collective X, S extracts the subsequence $X_S = aaaaaaaaaa\ldots$, and thus ($\mathsf{VM_2}$) is not satisfied, as the limiting relative frequency of a in the selected subsequence X_S is 1 and not p_a, the limiting relative frequency of a in the original sequence X.

[7]See [Wal38] for details.

[8]Von Mises defines four operations on collectives which play a fundamental role in his theory. For an overview of these operations, see, for instance, [vM81], pgs. 38-49.

complementation rule, are satisfied by the probabilities under consideration, *as long as the collectives are sufficiently random.*[9] Thus, the randomness of collectives is necessary to ensure that certain laws of the probability calculus are satisfied by the limiting relative frequencies in collectives, which in turn allows one to solve problems of the probability calculus. Thus, in von Mises' view, the place selections that "present themselves in a particular problem under consideration" are those place selections that ensure that we can carry out the calculations needed to solve the given problem.

There is one last point worth mentioning. An alternative to dropping the condition of admissibility from von Mises' definition (and thus working with a relative notion of collectives) is to provide a precise account of admissibility. This is the approach pursued by Alonzo Church, who in his 1940 paper "On the Concept of a Random Sequence" [Chu40] suggested that we take the admissible place selections to be the effectively calculable, and thus computable, place selections. This is an important development, for as I will discuss in Section 5, Kolmogorov took his 1963 definition of randomness to be a generalization of Church's definition, which itself was a modification of von Mises' original 1919 definition.

Now that we've discussed von Mises' account of probability and the role that randomness plays in that account, we now turn to determining the ways in which Kolmogorov drew upon von Mises' frequency theory to address the applicability problem.

3. The Influence of von Mises on Kolmogorov

Perhaps the most detailed expression of the relationship between Kolmogorov's views on the applicability problem and those of von Mises are found in Kolmogorov's 1963 paper "On Tables of Random Numbers". In the preface of the Russian translation of "Tables", which originally appeared in English in *Sankhyā, the Indian Journal of Statistics*,[10] Kolmogorov remarks that his discussion in the article "reflects a certain stage of my attempts to give meaning to the frequency interpretation of probability due to R. von Mises" ([Kol63], pg. 176).

Given what we've seen of von Mises' account, this remark is somewhat surprising. For in 1933 Kolmogorov published his monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung*[11], in which he provides an axiomatization of the notion of probability that takes probability to be primitive, an approach at odds with von Mises' account of probability, which, as we saw, took as primitive the notion of collective, and then defined probabilities in terms of collectives.[12]

---

[9]As an illustration of this point, Michiel van Lambalgen has shown the following. Let $X \in 2^\omega$ be a sequence generated by the tosses of a fair coin. Then if (i) the limiting relative frequency of tails (represented by the attribute 1) in X is $\frac{1}{2}$ and (ii) if the product rule holds for two consecutive tosses, then X is invariant under the following three place selections:

    (S_1) if n is odd, choose the nth value of the sequence;

    (S_2) if n is even, choose the nth value of the sequence; and

    (S_3) if n is even and the $(n-1)$st value of the sequence is 1, choose the nth value of the sequence.

See [Lam87], pgs. 36-37, for the proof of this result.

[10]The Russian version of the article was published in *Semiotika i Informatika* in 1982, and then, oddly enough, was later translated *back* into English in the third volume of Kolmogorov's collected works.

[11]The *Grundbegriffe* was translated into English in 1950 as *Foundations of the Theory of Probability* [Kol56a]. Henceforth, I will quote from the second edition English translation, published in 1956.

[12]Moreover, von Mises writes,

Kolmogorov was certainly aware of this difference, but it seemed to have little bearing on why he took interest in von Mises' frequentism. For Kolmogorov acknowledges at the beginning of "Tables" that even though his axiomatization of probability addressed "most of the difficulties in constructing the mathematical apparatus appropriate for numerous applications of probabilistic methods", the result was that "the problem of finding the causes of the applicability of mathematical probability theory was felt by many researchers to be of secondary importance" ([Kol63], 176).[13] As we'll see shortly, Kolmogorov even briefly takes up this issue in the *Grundbegriffe*, while also acknowledging the influence of von Mises.

Kolmogorov thus invokes von Mises' frequency theory not in order to make applications of probability theory, but rather to explain why probability is applicable in the first place. But Kolmogorov does not accept von Mises' frequentism *in toto*. Despite agreeing with von Mises that the frequentist approach, in some form, should serve the basis of the applicability of probability, Kolmogorov identifies some important differences. In particular, Kolmogorov states that two of his "long-held beliefs" were

(1) that von Mises' hypothetical frequentism cannot serve as the basis for the applicability of probability, since von Mises' approach considers infinite sequences of trials, yet in practice we only deal with finitely many trials; and

(2) the finite frequentist approach to probability, according to which the probability of an event is the relative frequency of that event in a sufficiently long, but finite, sequence of trials, cannot be formulated in a "rigorous formal purely mathematical way" ([Kol63], pg. 177).

It was only after Kolmogorov developed a definition of randomness for finite strings, as introduced in "Tables", that he came to abandon (2).[14] In Kolmogorov's view, with such a definition in hand, one could develop a rigorous finite frequency theory, which could in turn serve as the basis of the applicability of probability.

So what exactly did Kolmogorov glean from von Mises in responding to the applicability problem? To answer this question, we'll have to delve into Kolmogorov's 1933 *Grundbegriffe*. In so doing, we will encounter the raw materials that will allow us to determine what Kolmogorov took an application of probability to be and what was in his view the basis of the applicability of probability.

3.1. **Kolmogorov's *Grundbegriffe*.** As is widely known, Kolmogorov provided his axiomatization of probability in his 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung*, which

---

> According to our point of view, such a system of axioms [as Kolmogorov's] cannot take the place of our attempt to clarify and delimit the concept of probability. This becomes evident if we think of the simple case of the die or the coin where the above-indicated mathematical difficulties do not exist or rather where their solution is immediate without drawing on the mathematical theory of sets ([vM81], pg. 99).

For more details, see pgs. 98-99 of [vM81].

[13]In the original version of "Tables" as published in *Sankhyā*, the word "causes" is translated as "basis".

[14]The definition of randomness that Kolmogorov introduces in "Tables", while not the standard definition given in terms of Kolmogorov complexity, is a precursor to the complexity-based definition, which was officially published in 1965 in his paper "Three Approaches to the Definition of the Notion of the Amount of Information". However, Kolmogorov explicitly states the connection between the earlier definition and the later one, writing, "The complete development of this approach [i.e., the approach to defining random finite strings] involves introducing a measure of complexity of algorithms. I intend to discuss this question in another article" ([Kol63], pg. 177).

quickly gained widespread acceptance after its publication and soon became the foundation of measure-theoretic probability.

For our purposes, what is of interest in the *Grundbegriffe* is Kolmogorov's discussion of applications of probability in §2 of Chapter I, entitled "The Relation to Experimental Data", for this is the section in which he acknowledges the influence of von Mises (as mentioned above): In a footnote at the start of the section, Kolmgorov writes, "In establishing the premises necessary for the applicability of the theory of probability to the world of actual events, the author has used, in large measure, the work of R. v. Mises" ([Kol56a], pg. 3, fn. 4) As we will see, however, the way in which Kolmogorov draws upon von Mises' work is not clear from the discussion in §2; to determine von Mises' influence, we will have to look elsewhere.

Kolmogorov begins §2 by listing the steps one must take when "[w]e apply the theory of probability to the actual world of experiments" ([Kol56a], pg. 3). From this passage, we find that for Kolmogorov, an application of probability requires (at least) the following four components:

(C_1) an isolation of a set of repeatable conditions $\mathfrak{S}$;

(C_2) the identification of a field of probability $\mathfrak{F}$, i.e., the collection of possible events (closed under disjunction, conjunction, and complementation) that can occur when the conditions in $\mathfrak{S}$ are repeated;

(C_3) an assignment of non-negative real numbers to the elementary events in $\mathfrak{F}$ in such a way that satisfies Kolmogorov's axioms of probability; and

(C_4) the derivation of the probability of some event $A \in \mathfrak{F}$ by means of the axioms of probability and the probabilities of the elementary events.[15]

Given these components, there are a number of ways one might cash out 'an explanation of the applicability of probability'. For instance, one might consider an explanation of the applicability of probability to be an explanation as to why satisfying each of (C_1)-(C_4) suffices to determine the probabilities of the non-elementary events of the field of probability in question. But even if such an explanation were available, this would still not explain how, for instance, we assign probabilities to the elementary events in $\mathfrak{F}$, as in (C_3).

Moreover, Kolmogorov appears to have something even more specific in mind. In particular, Kolmogorov wanted to ascertain what was necessary, beyond the axioms of probability, to apply probability to problems concerning events in the real world. More concisely, Kolmogorov's concern was to determine what belongs in the second box below (which I'll call the *missing ingredient*):

$$\boxed{\text{Axioms of Probability}} \quad + \quad \boxed{\qquad ? \qquad} \quad \Longrightarrow \quad \boxed{\text{Applications of Probability}}$$

---

[15]As an illustration, Kolmogorov provides a helpful example. Let $\mathfrak{S}$, the set of repeatable conditions, be the tossing of a coin twice. Then the set of elementary events is $E = \{HH, HT, TH, TT\}$, and the field of probability consists of all subsets of E. Moreover, if A denotes the event of a repetition, i.e. $A = \{HH, TT\}$, then we can determine the probability of A from the probabilities of TT and HH by means of Kolmogorov's axioms. But how do we determine the probabilities of the elementary events? I will take up this question shortly.

The answer to the question as to what Kolmogorov took the missing ingredient to be is found in a letter Kolmogorov wrote to Fréchet in 1939.[16] Kolmogorov writes,

> You [Fréchet] are also right to attribute to me [...] the opinion that the formal axiomatization should be accompanied by an analysis of its real meaning. Such an analysis is given, perhaps too briefly, in the section "The relation to the world of experience" in my book. Here I insist on the view, expressed by Mr. von Mises himself [...] that "collectives" are <u>finite</u> (though very large) in real practice ([SV05], pg. 68).

Kolmogorov continues by describing three different approaches to probability, which he refers to as theories of probability:

> **Theory A:** A theory of probability given in terms of sufficiently long, finite collectives, where we work with only approximately stable frequencies.
> **Theory B:** A theory of probability given in terms of infinite collectives, i.e., von Mises' approach.
> **Theory C:** An axiomatic theory, such as the one given in the *Grundbegriffe*.

Concerning theory A, Kolmogorov concisely expresses the second of his two long-held views from "Tables", namely that "[t]his theory uses ideas that cannot be defined in a purely formal (i.e., mathematical) manner, but it is the only one to reflect experience truthfully" ([SV05], pg. 68). Concerning theory B, Kolmogorov expresses his first long-held view: "[T]his theory can be developed in a purely formal way", but "its relation to experience cannot have any different nature than for any other axiomatic theory" ([SV05], pg. 68). Lastly, concerning Theory C, an axiomatic theory of probability, Kolmogorov writes, "Its practical value can be deduced directly from the 'approximate' theory A without appealing to theory B. This is the procedure that seems simplest to me ([SV05], pg. 68).

Thus, in opting for a combination of theory A and theory C, Kolmogorov lays his cards on the table: The finite frequentist approach is the missing ingredient, as the "practical value" of the axiomatization of probability "can be deduced from directly" from such a theory. Moreover, by means of such a theory, we have a way of assigning probabilities to the elementary events in a given field of probability $\mathfrak{F}$; the probabilities assigned to the elementary events are given by the relative frequencies of these events in sufficiently long trials. Thus, we have:

$$\boxed{\text{Axioms of Probability}} \; + \; \boxed{\text{Finite Frequentism}} \; \implies \; \boxed{\text{Applications of Probability}}$$

There is, however, a problem with the above analysis, which we consider in the next section.

4. The Role of Randomness in Kolmogorov's Account

The problem with the analysis I've provided thus far is this: Kolmogorov certainly appears to draw the "premises necessary for the applicability of the theory of probability to the world of actual events" from the frequency theory of von Mises, yet on von Mises' account,

---

[16]This letter, part of the Fréchet papers in the archives of the Academy of Sciences in Paris, is translated from the original French in [SV05].

we cannot determine probabilities from just any sequence for which the limiting relative frequencies exist. Rather, as we've seen, on von Mises' account, the satisfaction of a requirement of randomness is also necessary.

Kolmogorov also held that a requirement of randomness must factor in to an explanation of the applicability of probability, as mentioned in the Introduction. Thus, we can now sharpen our question $\mathcal{Q}_2$ as follows: How does Kolmogorov's requirement of randomness feature in his explanation of the applicability of probability? Moreover, does it play the same role in Kolmogorov's finite frequentism as it does in von Mises' hypothetical frequentism?

To answer the first question, we now turn to Kolmogorov's 1956 survey article, "The Theory of Probability", which is contained in the second volume of *Mathematics: Its Content, Method, and Meaning*, a three-volume collection of surveys of different areas of mathematics written by a number of leading Russian mathematicians.

4.1. **Kolmogorov's survey "The Theory of Probability".** Kolmogorov begins his survey by identifying the events to which we assign probabilities: the random events. He writes,

> The simplest laws of natural science are those that state the conditions under which some event of interest to us will either certainly occur or certainly not occur; i.e., these conditions may be expressed in one of the the following two forms:
> (1) If a complex (i.e., a set or collection) of conditions S is realized, then event A certainly occurs;
> (2) If a complex of conditions S is realized, then event A cannot occur.
> In the first case the event A, with respect to the complex of conditions S, is called a "certain" or "necessary" event, and in the second an "impossible" event. ([Kol56b], pg. 229)[17]

But there is a third category of events: events A that under a fixed complex of conditions S sometimes occur and sometimes do not occur. According to Kolmogorov, these events are random with respect to the conditions S.

In discussing how we determine the probability of such random events, Kolmogorov is explicit about his frequentist approach to probability, writing,

> The assertation that the event A occurs under conditions S with a definite probability $P(A|S) = p$ amounts to saying that in a sufficiently long series of tests (i.e., realizations of the complex of conditions S), the frequencies [. . .] of the occurrence of the event A [. . .] will be approximately identical to one another and close to p ([Kol56b], pgs. 230-231).

This much is familiar, given our earlier discussion of Kolmogorov's account. But Kolmogorov goes on to point out two instances of what he considers to be "unavoidable" vagueness in his definition. First, there is vagueness in the 'closeness' of the observed frequency of an event A and the actual probability of A: just how close is close enough? In particular, Kolmogorov

---

[17]Kolmogorov offers the following example. If we take the complex of conditions S to be the atmospheric pressure, with temperature t between 0 degrees and 100 degrees Celsius, and we consider the events
 A_1, that water occurs in liquid state,
 A_2, that water occurs in a gaseous state, and
 A_3, water occurs in a solid state,
it follows that A_1 is certain while A_2 and A_3 are impossible.

acknowledges that he doesn't specify how long a sequence of experiments must be "in order to exhibit beyond all doubt the existence of the suppose [sic] stability" ([Kol56b], pg. 248).

The second instance of unavoidable vagueness is more interesting. In Kolmogorov's words,

> More fundamental, however, is the second lack of clearness concealed in our formulations; it concerns the manner of forming the sequences of trials in which we are to examine the stability of the frequency of occurrence of the event A. [¶] As stated earlier, we are led to statistical and probabilistic methods of investigation in those cases in which an exact specific prediction of the course of events is impossible. But if we wish to create in some artificial way a sequence of events that will be, as far as possible, purely random, then we must take special care that there shall be no methods available for determining in advance those cases in which A is likely to occur with more than normal frequency" ([Kol56b], pg, 248).

Here we see for first time in Kolmogorov's works an explicit connection between random events and random sequences composed of such events, where the random sequences have some sort of invariance reminiscent of von Mises' definition: in order to determine the probability of an event A, A must occur in a sequence of events with the property that "there shall be no methods for determining in advance those cases in which A is likely to occur with more than normal frequency."[18] To illustrate this point, Kolmogorov provides an illustrative example:

> Such precautions are taken [i.e. that no methods of determination are available], for example, in the organization of government lotteries. If in a given lottery there are to be M winning tickets in a drawing of N tickets, then the probability of winning for an individual ticket is equal to $p = M/N$. This means that in whatever manner we select, in advance of the drawing, a sufficiently large set of n tickets, we can be practically certain that the ratio μ/n of the number μ of winning tickets in the chosen set to the whole number n of tickets in this set will be close to p. For example, people who prefer tickets labeled with an even number will not have any systematic advantage over those who prefer tickets labeled with odd numbers, and in exactly the same way there will be no advantage in proceeding on the principle, say, that it is always better to buy tickets with numbers having exactly three prime factors, or tickets whose numbers are close to those that were winners in the preceding lottery, etc. ([Kol56b], pg. 248).

Now, the vagueness which Kolmogorov takes to be unavoidable is in these 'methods': which methods are we to employ in determining in advance the cases in which the frequency of A differs from the frequency of A in the original sequence? There appears to be no single correct answer to this question. Instead, we define the randomness of an event only once we've fixed our methods for extracted subsequences from the original sequence. Again, this should remind us of von Mises' definition. In particular, this account is quite similar to the later version of von Mises' definition of probability, which considers collectives that are relative to some fixed countable collection of place selections. Kolmogorov then concludes,

---

[18]The 'normal' here is not meant in a technical sense, as in 'normal distribution', but rather, the 'normal frequency' of A is simply the frequency of A in the given sequence. Kolmogorov is worried here that there might be a subsequence of the given sequence in which A occurs more frequently than in the original sequence.

> Thus, to say that an event A is 'random' or 'stochastic' and to assign it a definite probability $p = P(A|S)$ is possible *only when we have already determined the class of allowable ways of setting up the series of experiments.* The nature of this class will be assumed to be included in the conditions S ([Kol56b], pg. 249, emphasis mine).

That is, we can only be justified in assigning a probability to a random event if the event occurs in some finite sequence of events that has the property that the event has the same relative frequency in every subsequence selected from the original sequence by means of the methods we have determined *in advance.* Moreover, we build this conditions into our conditions S, so that when we consider the probability of A relative to S, this probability will be relative to the relevant methods of determination.

In sum, by 1956 Kolmogorov had arrived at the view that we can only assign probabilities to random events, which themselves must occur in a random sequence of events. However, Kolmogorov admits that this account is beset with vagueness, given that, at that time, he had yet to find a reasonable formal definition of random finite sequences.

Having traced Kolmogorov's thought on the applicability problem from the *Grundbegriffe* in 1933 to "The Theory of Probability" in 1956, let us now return to "Tables of Random Numbers" to tie up some loose ends.

5. "Tables of Random Numbers" Revisited

Back in Section 2, I mentioned that Kolmogorov eventually abandoned his long-held belief that the finite frequentist approach to probability could not be formulated in a rigorous, purely mathematical way once he produced a formal definition of randomness. With this definition in hand, Kolmogorov could finally develop the finitist frequentist approach the contours of which he laid out in "A Theory of Probability". Thus, in "Tables" he writes,

> Concerning the second [long-held belief], I have come to the conclusion that the notion of random distribution can be introduced in a purely formal way, namely: it can be shown in sufficiently large collections the distribution of a certain property may be such that the frequency of its appearance will be aproximately [sic] the same for all sufficiently large samples, if only the *law is sufficiently simple* ([Kol63], pg. 177, emphasis in the original).

To decipher what Kolmogorov means by a law that is "sufficiently simple", we need to lay out several details of his definition of randomness.[19]

Drawing upon Church's modification of von Mises' definition of admissible place selections, according to which the admissible place selections are the computable place selections, Kolmogorov defines a string to be random if the relative frequencies of the values of the string are "roughly" invariant under selection by what he calls *admissible selection algorithms.* These admissible selection algorithms behave much like place selections, except that they can proceed *non-monotonically*; that is, instead of extracting a substring along the initial segments of a given string, an admissible selection algorithm can select values in any

---

[19]To be clear, this is the definition Kolmogorov offers in "On Tables of Random Numbers", not his complexity-based definition that he lays out in 1965 in "Three Approaches to the Definition of the Notion of Amount of Information", [Kol65], although they are related.

computably specified order,[20] just as long as it doesn't revisit any location in the process of extracting a substring (which Kolmogorov refers to as a 'sample') from the given string.

To define a random finite string, we have to fix a number of parameters: $N, n \in \omega$ and $p, \epsilon \in [0, 1]$. For a fixed length N, we define a finite binary string T (which Kolmogorov refers to as a table) to be (n, ϵ, p)-random with respect to a collection of admissible selection algorithms $\mathcal{R}_N$ if for any sample A of length at least n selected from T by means of an admissible selection algorithm, the relative frequency of 1s in A is within ϵ of the value p.

The key result of "Tables" is the following theorem, which shows that under certain natural conditions on the size of the collection of admissible selection algorithms in question, (n, ϵ, p)-random tables exists.

Theorem. If $\#(\mathcal{R}_N) \leq \frac{1}{2}e^{2n\epsilon^2(1-\epsilon)}$, then for $p \in [0, 1]$, there is a table T of size N that is (n, ϵ, p)-random with respect to $\mathcal{R}_N$.

According to Kolmogorov, the meaning of this theorem becomes more clear if we consider the result in terms of $\log_2 \#(\mathcal{R}_N)$. This value, according to Kolmogorv, gives "the amount of information needed to indicate a specific element R from $\mathcal{R}_N$" ([Kol63], pg. 179-180). Kolmogorov continues, "Clearly in the case of a large $[\log_2 \#(\mathcal{R}_N)]$, the system $\mathcal{R}_N$ must contain algorithms whose specification (and not only whose actual execution) is complicated (requires no less that $[\log_2 \#(\mathcal{R}_N)]$ binary digits)" ([Kol63], pg. 180). Thus the theorem implies that

$$\log_2 \#(\mathcal{R}_N) \leq 2(\log_2 e)n\epsilon^2(1 - \epsilon) - 1$$

is sufficient for the existence of a (n, ϵ, p)-random table of length N. Thus, in a very rough sense, the smaller value of $\log_2 \#(\mathcal{R}_N)$, the simpler are the admissible selection algorithms contained in $\mathcal{R}_N$. Thus, we can always find random strings was long as the collection $\mathcal{R}_N$ is sufficiently small, or contains "sufficiently simple laws".[21]

There are important similarities between Kolmogorov's definition of random finite strings and von Mises' definition of random infinite sequences. In fact, Kolmogorov makes an even stronger claim, writing, "I believe that this construction correctly reflects the von Mises approach in all its generality" ([Kol63], pg. 179). Thus, it appears that the answer to $\mathcal{Q}_2$, whether the definition of randomness that Kolmogorov took to be necessary to address the applicability problem is drawn from the work of von Mises, is a resounding "Yes". In the closing section, however, I will draw upon the discussion above to conclude that despite this formal similarity, the role that the definition of randomness plays in Kolmogorov's account is significantly different from the role it plays in von Mises' account.

6. On The Different Roles of Randomness

As we have seen, Kolmogorov's definition of random finite strings is formally similar to von Mises' definition of random infinite sequences, as both are defined in terms of the invariance of relative frequencies under place selections. But this should not cause us to conclude that

---

[20]The order in which the admissible selection algorithm selects can also depend on the previous locations considered, and the value of the string at those location, and so these rules are adaptive.

[21]The reader familiar with Kolmogorov's complexity-based definition of random finite strings should recognize that Kolmogorov is only a few steps from this more general, and more elegant, definition of randomness. Unfortunately, it is beyond the scope of the article to discuss this second definition.

these definitions play the same role in the theories of Kolmogorov and of von Mises'. Let us briefly review the role of randomness in both von MIses' and Kolmogorov's theories.

As discussed in Section 2, according to von Mises, even though we only consider probabilities that are defined in sequences satisfying the condition of randomness, this condition guarantees that we'll be able to carry out the requisite calculations in the process of solving problems of the probability calculus. For as we discussed earlier, on von Mises' approach, a problem in the probability calculus begins with a given initial collective and ends with a final collective, from which we can extract the solution to our problem. Moreover, since von Mises took collectives as primitive, he could not appeal to Kolmogorov's axioms of probability to guarantee that the probabilities that arise in the process of solving the problem in question satisfy certain laws of probability. Instead, by means of the collective-preserving operations on collectives, we can guarantee that the relevant laws of probability hold, as long as our collectives are sufficiently random, i.e., invariant under the appropriate place selections.

In contrast, on Kolmogorov's approach, invariance under selection by admissible selection algorithms guarantees that we correctly assign probabilities to the elementary events in our field of probability. Given a sufficiently large collection of objects, we would like to determine the probability of the occurrence of some property of that holds of certain objects in the collection. In line with the frequency theory, this probability is the relative frequency of the occurrence of the property, but only in those collections in which this relative frequency is invariant under all sufficiently large samples extracted by a sufficiently simple law. That is, we can determine the probability of the property in question only if the property occurs in a sufficiently long, random sequence of events. Thus, on Kolmogorov's account, a definition of randomness for finite strings is necessary to determine the probabilities that are assigned to the elementary events of a probability field $\mathfrak{F}$: we can only guarantee that we've correctly assigned probability to the elementary events if they occur in a sufficiently long random sequence, for then the relative frequency of each of these events will be invariant among samples taken from this sequence. Once these probabilities are determine, we can thus appeal to the axioms of probability to determine the probabilities of other events in the field $\mathfrak{F}$.

In sum, the role of randomness in von Mises' account differs significantly from the role of randomness in Kolmgorov's account: On von Mises' account, randomness secures the passage from initial probabilities to final probabilities, while on Kolmogorov's account, randomness secures the determination of initial probabilities.

References

[Chu40] Alonzo Church. On the concept of a random sequence. *Bull. Amer. Math. Soc.*, 46:130–135, 1940. 6

[Kam32] Erich Kamke. *Einführung in die Wahrscheinlichkeitstheorie.* S. Hirzel, Leipzig, 1932. 4

[Kam33] Erich Kamke. Über neuere Begründungen der Wahrscheinlichkeitsrechnung. *Jahresbericht Deutsche Mathematiker Vereinigung*, 42:14–27, 1933. 4

[Kol56a] A. N. Kolmogorov. *Foundations of the theory of probability.* Chelsea Publishing Co., New York, 1956. Translation edited by Nathan Morrison, with an added bibliography by A. T. Bharucha-Reid. 1, 6, 8

[Kol56b] A. N. Kolmogorov. The theory of probability. In A.D. Aleksandrov, A. N. Kolmogorov, and M.A. Lavrent'ev, editors, *Mathematics: Its Content, Methods, and Meaning*, volume 2, pages 229–264. The MIT Press, 1956. 10, 11, 12

[Kol63] A. N. Kolmogorov. On tables of random numbers. *Sankhyā Ser. A*, pages 176–183, 1963. Reprinted in Selected works of A. N. Kolmogorov. Vol. III: Information Theory and the Theory of Algorithms, Shiryayev, A.N. Kluwer Academic Publishers Group. 1, 6, 7, 12, 13

[Kol65] A. N. Kolmogorov. Three approaches to the definition of the notion of amount of information. *Problemy Peredači Informacii*, pages 184–193, 1965. Reprinted in Selected works of A. N. Kolmogorov. Vol. III: Information Theory and the Theory of Algorithms, Shiryayev, A.N. Kluwer Academic Publishers Group. 12

[Kol83] A. N. Kolmogorov. On logical foundations of probability theory. In *Probability theory and mathematical statistics (Tbilisi, 1982)*, volume 1021 of *Lecture Notes in Math.*, pages 1–5. Springer, Berlin, 1983. 1, 2

[Lam87] Michiel Van Lambalgen. *Random Sequences.* PhD thesis, University of Amsterdam, 1987. 6

[SV05] Glenn Shafer and Vladimir Vovk. The origins and contributions of Kolmogorov's *Grundbegriffe.* Working Paper 4, probabilityandfinance.com, 2005. 9

[Ven66] John Venn. *The logic of chance: an essay on the foundations and province of the theory of probability, with especial reference to its logical bearings and its application to moral and social science, and to statistics.* Macmillan and Co, London and Cambridge, 1866. 3

[vM19] Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.*, 5(1-2):52–99, 1919. 2

[vM64] Richard von Mises. *Mathematical theory of probability and statistics.* Edited and Complemented by Hilda Geiringer. Academic Press, New York, 1964. 5

[vM81] Richard von Mises. *Probability, statistics and truth.* Dover Publications Inc., New York, English edition, 1981. 2, 3, 5, 7

[Wal38] Abraham Wald. Die Widerspruchsfreiheit des Kollektivbegriffes. In *Colloque consacré au calcul des probabilités*, volume 735 of *Actualités Scientifiques et Industrielles*, pages 79–99. Hermann, 1938. 5