Hypothetical Frequencies as Approximations

Jer Steeger[1] ●

Abstract

Hájek (Erkenntnis 70(2):211–235, 2009) argues that probabilities cannot be the limits of relative frequencies in counterfactual infinite sequences. I argue for a different understanding of these limits, drawing on Norton's (Philos Sci 79(2):207–232, 2012) distinction between approximations (inexact descriptions of a target) and idealizations (separate models that bear analogies to the target). Then, I adapt Hájek's arguments to this new context. These arguments provide excellent reasons not to use hypothetical frequencies as idealizations, but no reason not to use them as approximations.

1 Introduction

> *This account of probability will be my target—15 times over. Why so many arguments? Is this an exercise in overkill? Or worse, is it an exercise in* under-*kill, my deployment of so many arguments betraying a lack of faith that any one actually does the job? On the contrary, as in* Murder on the Orient Express*, I think that many of the blows may well be fatal on their own (although in the book, the victim received only twelve of them).*
>
> —Alan Hájek, "Fifteen Arguments against Hypothetical Frequentism"

If you are partial to frequentism, it pays to keep your guard up. It sometimes feels like philosophers of probability have been sharpening their knives for the past several decades, making increasingly pointed jabs at the view. Hájek (1997, 2009) collects thirty of their most cogent and popular arguments, eloquently refining earlier criticisms due to Fine (1973), Jeffrey (1992), and others while presenting wholly new ones. Perusing these arguments, the average scientist might rush to insist that frequentists use "approximations" or "idealizations" that Hájek is taking far too seriously. But among philosophers of probability, a rough consensus has emerged

---

✉ Jer Steeger
  jsteeger@uw.edu

1 Department of Philosophy, University of Washington, Savery Hall, Room 361, Seattle, WA 98195, USA

⚛ Springer

that this response cannot save the view as a satisfactory *analysis* of the concept—an explication of the term "probabilities" that tells us what the things it refers to *really are* (La Caze, 2016; Rowbottom, 2015).

Very roughly, Hájek argues that frequentists face a fork in the road and crash in the middle. On one horn of the supposed dilemma, frequentists could say that probabilities *really are* the finite frequencies that we study empirically. He directs fifteen of his arguments against this analysis, which he calls *finite frequentism* (FF):

(FF)   The probability of an attribute *A* in a finite reference class *B* is the relative frequency of actual occurrences of *A* within *B*.

One key argument is that FF fails to recover the full richness of the probability calculus. Repeatedly measuring finite frequencies, Hájek argues, is a good method for "approximating" irrational probabilities—but these frequencies do not recover irrational probabilities on their own. Hence Hájek's slogan, "Finite frequentism: good methodology, bad analysis" (1997, p. 226). So, on the other horn of the supposed dilemma, frequentists could recover the probability calculus by saying probabilities *really are* limiting frequencies. Hájek (2009) thus points his remaining fifteen arguments at the view he calls *hypothetical frequentism* (HF):

(HF)   The probability of an attribute *A* in a reference class *B* is *p* if and only if the limit of the relative frequency of occurrences of *A* within *B* would be *p* if *B* were infinite.

He argues that it is not clear how one might view hypothetical frequencies as "idealizations" *and* assert the truth of HF. So long as the frequentist believes HF, he thinks, that frequentist must believe in bizarre and unempirical counterfactuals involving infinite sequences. The roads to a frequentist analysis, then, seem closed for good. La Caze (2016) offers one semi-recent endorsement of this moral, maintaining that any attempt to save HF from one of Hájek's arguments—including the appeal to idealizations—leaves the view "exposed" to others among the remaining fourteen.

But are FF and HF the only roads to a frequentist analysis, or even the most desirable? A reassessment seems warranted. On a von Mises-styled approach, limiting frequencies are meant to account for the objective probabilities appearing in scientific theories, with statistical mechanics serving as a paradigm. And lately, philosophers of physics have become increasingly more attuned to the subtleties of using infinite limits in various physical theories (Earman, 2019; Fletcher et al., 2019; Palacios & Valente, 2021; Valente, 2019). Of particular importance to this literature is Norton's (2012) distinction between limits used as mere approximations and those used to construct idealizations. Norton first introduces a general distinction between an *approximation*, an inexact description of a target system, and an *idealization*, a separate (real or fictitious) system that bears some crucial analogy to the target. He then argues that scientists often use infinite limits to give approximate descriptions of finite targets—rather than the other way around, as Hájek's approach to finite frequencies seems to suggest. Moreover,

Norton argues that several approximating limits in physics do not support idealized models. The idealization in question might fail to be well-defined, or it might exhibit pathologies that are deeply misleading about the target. Plausibly, the frequentist's infinite limits are of precisely this latter sort: they are excellent approximations that fail to support idealizations.

If that is right, two things about the structure of Hájek's arguments start to look a little strange. First, following Norton, we can use HF's limits to give inexact descriptions of the finite relative frequencies of FF—and so the supposed dilemma turns out to be a false one. Second, HF places undue emphasis on the limit system. If it is right that we ought to use the frequentist's limits as mere approximations, then we should actively *ignore* the infinite system and focus instead on the limiting value of the property *p*. This sort of thinking might prompt one to take a rough pass at a different biconditional, one that addresses both of these issues:

(AF)   The probability of an attribute *A* in a finite reference class *B* is *p* if and only if [a limit of a certain sort] yields a value of *p* that approximates the relative frequency of actual occurrences of *A* within *B*.

AF stands for *approximation-first frequentism*; the bit in square brackets is yet to be refined. AF offers a new kind of frequentist analysis of probability. Like FF and HF, AF is an explicit definition that tells us what probabilities *really are*. AF, however, does not say that probabilities are things in the world. It says that probabilities are nothing more than approximate descriptions. There are roughly-stable relative frequencies out in the world, and probability is the language that we use to (partially) describe them. Since folks adopt this sort of attitude towards many other limits in science, AF seems well worth considering.

This attitude also casts Hájek's canonical arguments against HF in a new light. Rather than sending frequentism to the grave, they pinpoint precisely why one should not use the frequentist's limits as idealizations. Consider, for example, Hájek's argument that the value of the limit depends on the ordering of the infinite sequence, whereas a relative frequency in a finite sequence clearly does not. Here, the infinite system (an infinite *B*) yields a pathology (*p* of *A* depends on ordering) that misleads one about the nature of the target (a finite *B*). That is a good reason to be wary of the infinite system! The approximation-first frequentist takes this warning to heart by suggesting that only the limiting property *p* should matter for the analysis. AF similarly recasts Hájek's remaining fourteen concerns about HF (and one more popular one about the Law of Large Numbers). These knives become foils, exposing the pitfalls of treating limiting frequencies as robust idealizations rather than mere approximations.

In Sect. 2, I review Norton's distinction between approximations and idealizations. Then I consider several approximating limits in scientific theories, and I argue that formulations analogous to HF make poor sense of them. Parallel reasoning motivates one to adopt something like AF. Section 3 uses von Mises's frequentism to give a more precise statement of AF, and it shows how AF explains

Kolmogorov's axioms as approximately true laws of nature. Section 4 adapts Hájek's fifteen arguments to this new context to demonstrate the importance of interpreting limiting frequencies as mere approximations. Section 5 concludes.

2 Approximations and Idealizations

As flagged above, a von Mises-styled approach to frequentism has a narrow aim: it seeks only to analyze the objective probabilities that appear in science.[1] Such an analysis will fall short of Hájek's aim to provide a *univocal* analysis—an analysis that accounts for *all* uses of probability, from the weight of evidence in the courthouse to my confidence that it will rain tomorrow to the chance that a fair coin lands heads. The sort of frequentist that I have in mind only seeks to analyze probabilities of the last sort, probabilities describing roughly-stable relative frequencies of attributes in finite empirical trials of a given type. On this tack, frequentism looks more or less like any standard scientific theory. It posits some universal generalizations (i.e., Kolmogorov's laws) governing regularities in the behavior (i.e., roughly-stable relative frequencies) of empirical systems (i.e., finite sequences of trials of a given type). So if the limits in the frequentist's theory seem puzzling, it makes good sense to look to the limits in that theory's siblings.

How should the modern philosopher of science interpret these limits? It is tempting to take them at face value, in line with something like van Fraassen's semantic view of theories. Recall that van Fraassen (1980) takes his semantic view to be a suitable response to the failure of logical positivism to recast theories in a strictly empirical language. Instead of purging science of metaphysical terms, the thought goes, we should take theories at their word and sort out the metaphysics later. And since theories largely talk in terms of *models* of the world, the semantic view suggests (very roughly) that we identify the content of a theory with its models. On this view, it would seem that when a physicist says they take the number of particles in their model to infinity, we should take them to be positing a *different model*. We get something like the following picture:

$$\text{target system} \xrightarrow{\text{take a limit of a parameter}} \text{limit system}$$

---

[1] I do not claim that frequentism is the only viable or even the best analysis of objective probabilities in science. I just argue that one version of it (AF) is a very good one. In fact, I take AF to be compatible with various other accounts of objective probabilities, including propensity accounts. For further discussion of this point, see my response to Hájek's (12) in Sect. 4.3 and footnote 15.

Regarding whether this argument addresses Hájek's concerns, note that Hájek readily acknowledges that frequentism is "at best an analysis of *objective* probability" but he maintains that "it cannot even be that" (1997, p. 209, emphasis his). Thus, one might read Hájek as challenging that any version of frequentism can succeed by its own lights. In turn, I seek to defend that a particular version of frequentism, AF, can succeed by its own lights. For more on the sense in which AF offers an analysis of objective probability, see footnote 8.

But Norton ([2012](#)) cogently argues that we should be a bit more subtle. As it turns out, we can often precisely and usefully define the limit of a property of a finite target system without introducing any new model. The limit property serves, first and foremost, as an *approximation*, a partial description of the finite target. In other words, Norton suggests that the following picture really ought to come first:

$$\text{property of target system} \xrightarrow{\text{take a limit of a parameter}} \text{limit property}$$

But then this limit property *might or might not* usefully hold of some limit system. In other words, one might or might not be able to read the limit property as a literal description of an *idealization*, some entirely novel (and possibly fictional) system that bears crucial analogies with the target.[2]

Norton helpfully illustrates the difference by considering a simple geometric property: the surface-area-to-volume ratio of a shape like a sphere or an ellipsoid. Take, to start, a sphere of radius r in $\mathbb{R}^3$. Its surface area is $4\pi r^2$, and its volume is $4\pi r^3/3$, yielding a ratio of $3/r$. One can precisely define the limit of this property: as we take the value of r to infinity, the ratio converges to zero. This limit property works wonderfully as an approximate description of the ratio's value for large-but-finite spheres. But *there is no infinite sphere!* A bit more precisely: while we can treat each of our finite spheres as sets of points in $\mathbb{R}^3$, the $r = \infty$ case does not straightforwardly correspond to any particular subset of that space. The limit system does not exist, but the limit property does—and it provides a useful approximation, to boot.

Now imagine elongating a unit sphere into an ellipsoid with a semi-major axis of length a in $\mathbb{R}^3$. In the infinite limit, this ellipsoid becomes an infinitely-long cylinder—a new system that might or might not serve as a useful idealization. The ellipsoid's volume is $4\pi a/3$, and as it elongates, its surface area gets arbitrarily close to $\pi^2 a$. Thus, in the infinite limit, the ratio of surface area to volume is given by $3\pi/4$—and this limiting value is an excellent approximation of the ratio for ellipsoids with large a. But can we reason the other way, from the infinite cylinder back to an approximate description of the ellipsoid? Certainly not! The surface-area-to-volume ratio of the infinite cylinder is ambiguous. We could obtain the infinite cylinder by elongating a finite cylinder. And if we take this process to define the

---

[2] The literature on approximations and idealizations in science is broad and rich, and it explores many interesting distinctions beyond what Norton has in mind here. On the one hand, Frigg and Hartmann ([2020](#)) introduce a general distinction between Aristotelian idealizations (models that strip details away) and Galilean idealizations (models that deliberately distort); Weisberg ([2007](#)) differentiates three versions of the latter, while McMullin ([1985](#)) identifies at least six. Potochnik ([2017](#), [2020](#)) cogently argues that these sorts of distinctions ought to center scientists' many diverse and intertwining aims (such as fidelity to causal structure, pedagogical clarity, and computational simplicity). On the other hand, Rueger and Sharp ([1998](#)) restrict the notion of approximation to quantitative closeness—and while Norton's notion is a broader one, this idea plays a crucial role in his account. I do not want to police the various uses of the terms "approximation" and "idealization" here and elsewhere; I think each of the above accounts has its virtues. I only mean to adopt Norton's definitions as a convention, as his core distinction turns out to be particularly useful for thinking about frequentism.

surface-area-to-volume ratio, we get a value of 2 instead! The limit system, in this case, is too impoverished to serve as a good guide to finite behavior. However, this failure of the limit system is not a reason to abandon the use of the limit property (a surface-area-to-volume ratio of $3\pi/4$) as an expedient description of arbitrarily-large ellipsoids.

What happens if we pursue an HF-style approach to, say, the large-ellipsoid limit? We might get something like the following biconditional:

(SV)   The surface-area-to-volume ratio of an ellipsoid with large *a* is close to *S/V* if and only if this ratio would equal *S/V* if *a* were infinite.

Given the ambiguity of the limit system, it is not clear how to evaluate the truth of this claim. Specifically, this claim does not tell us *how* to take the large *a* limit. So, for all SV tells us, we might expect the approximate value of the ratio to be 2— and this value is a poor match for the ratio *S/V* of actual large ellipsoids. We can even make the sense in which it is a "poor match" precise: specify (as a conventional choice) an error tolerance $\epsilon > 0$, such that if *S/V* is within ϵ of our limit, we deem the limit to be *empirically correct*. For a not-too-conservative choice of ϵ (e.g., $\epsilon < .1$) and a not-too-large choice of *a* (e.g., $a > 10$), the elongated-ellipsoid limit is empirically correct and the elongated-cylinder limit is not.

One might try to wave this point away if it only concerned toy examples like the sphere and the ellipsoid. Scientists, however, widely exploit these sorts of approximations. For example, Norton argues that the same thing happens in statistical mechanics when taking the Boltzmann-Grad limit. This limit generates the Boltzmann equation, which approximately governs the time evolution of the distribution of particles in an ideal gas. This approximation describes a gas consisting of *N* hard spheres of diameter *d*, under two assumptions: (a) that the density of particles is low enough to treat them independently and (b) that a typical particle undergoes on the order of one collision per unit of time. Another way of putting (b) is that the mean free path of a particle, given approximately by $\lambda = 1/(2\pi N d^2)$, should be of order one. The Boltzmann equation holds precisely in the Boltzmann-Grad limit, where we take *N* to infinity and *d* to zero in such a way that Nd^2 remains a constant of order one (Lanford, 1981, p. 72). However, it would be fatal to conclude that this limit yields a well-behaved idealization.

Again, imagine an HF-like approach to this limit. We might get something like:

(BG) The density of particles in a macroscopic gas is ρ if and only if the particle density of a suitable *N*-particle gas would be ρ if *N* were infinite (and were the diameter *d* of the particles zero).

In the derivation of Boltzmann's equation, we assume that *d* is non-zero to determine the state of particles after collisions. However, if *d* were truly equal to zero, then there would be no preferred plane of collision—and so we could not determine the post-collision state (Norton, 2012, p. 219). Thus, it is unclear how to evaluate the truth of BG. A literal infinity of point particles is a poor guide to the behavior of

finite systems. Only the behavior of the approximating limit matters for our description of the physics, for our account of the laws governing the gas's dynamical evolution.

Palacios and Valente (2021) provide another particularly elegant example: the low-velocity limit of special relativity. As is well-known, one can recover the laws of classical mechanics by taking the limit of relativistic laws as v/c approaches zero (where c is the speed of light). Take, for example, momentum: a relativistic particle has a momentum $p = \gamma m_0 v$, where m_0 is the particle's invariant mass, v is its velocity (in a given inertial frame), and γ is the Lorentz factor,

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} = 1 + \frac{1}{2}\left(\frac{v}{c}\right)^2 + \frac{3}{8}\left(\frac{v}{c}\right)^4 + \frac{5}{16}\left(\frac{v}{c}\right)^6 + \cdots,$$

where the second equality gives the first few terms of its Taylor series expansion. As v/c approaches zero, γ approaches one, and the momentum of the particle is given by its classical definition, $p = m_0 v$. So the conservation of relativistic momentum (in the absence of a net external force) explains the conservation of classical momentum when an object's velocity is very small compared to c. Again, on an HF-style approach, we might try to capture this recovery of classical momentum with the following claim:

(SR) The momentum of a classical particle is p if and only if the limit of the relativistic momentum would be p if v/c were zero.

It is true that $p = mv$ holds exactly when $v = 0$. But $v = 0$ is strictly false of any moving target. And classical mechanics is useful for far more than describing stationary objects! Appeal to the approximating limit resolves this issue. For the velocities of everyday objects (like bikes or cars or trains), the second term of the series expansion of γ only makes a difference of about one in one hundred billion—and the corrections offered by the following terms only get smaller from there. In this way, special relativity suggests we can use a very small ϵ to assert the empirical correctness of classical laws of physics for everyday objects.

We could multiply examples indefinitely—and they need not be drawn from the physical sciences, either! Strevens (2019) cogently argues that another example comes from a standard justification of deterministic models in population genetics. Following Strevens, we can use a straightforward model to illustrate the point. Consider just two generations of a population B of monoploid organisms, i.e., organisms containing just one set of chromosomes. Now suppose that each organism must possess one of two alleles expressing variants of a given trait, A or A', the former of which provides an adaptive advantage over the latter. Our model's input is the proportion of the parent generation with A, and its output is a distribution predicting how prevalent A will be in the next generation. Fix the size of the population at an even number N for both generations. Next, suppose that our model maps a parent generation with half A-organisms and half A'-organisms to the Gaussian approximating the binomial distribution for p with a sample size N. As N approaches infinity,

the standard deviation of the Gaussian approaches zero. The proportion of next-generation organisms with *A* is thereby "fixed" as we approach the limit: if the mean of the Gaussian for a half-*A*, half-*A'* parent generation is $p = 3/4$, then three-fourths of the next generation will have trait *A* with probability one. Thus, it seems that we can replace our probabilistic model with a deterministic one for large enough populations.[3] But suppose we gloss this explanation in the HF-like way:

(GD) The proportion of the organisms that possesses trait *A* in the next generation of *B* will be *p* if and only if that proportion would be *p* if *B* were infinite.

Once again, the limit system has a fatal flaw. The notion of "the proportion of a population with a given trait" only makes good sense for finite populations! As Abrams (2006) notes, a close analog of Hájek's problem of ordering applies directly to this case: the value of *p* depends on the order of the sequence of members of the population. Strevens infers from this fact that infinite-population models cannot serve as idealizations in Norton's sense. He argues for a weaker sense in which these models provide idealizations, but he concedes that a merely-approximating limit solves the paradox.[4] Such a limit says that the actual proportion of *A*s will be at most ϵ away from *p* for any finite-but-large population *B*.

The frequentism case seems to be one more example of a merely-approximating limit. It is the goal of the next section to substantiate this claim, but here is a quick and rough sketch of the idea. Our target system is now some collection Σ of attributes A_1, A_2, and so on, associated with a reference class *B* describing some empirical setup. Suppose we have *N* repeated trials or runs of this setup, and we record which attributes appear in which trials. We specify one way of taking the limit of these attributes' relative frequencies as $N \to \infty$. If we have chosen the right way, then the resulting value should approximate the *actual* frequencies (for a not-too-liberal choice of ϵ). One can then derive that these approximate values must obey

---

[3] One might roughly identify the variance of the Gaussian in this model with genetic drift. Then, one could say that something like the above argument demonstrates why genetic drift is eliminated in large populations, isolating effects due to other evolutionary forces. Gillespie's (1998, §2.2) textbook presentation, for example, seems to adopt this view. Nevertheless, this sort of causalist interpretation is not required for the infinite-population idealization to pose an issue. The issue remains as long as we use the deterministic model to explain actual genetic data, even if we only take this model to posit brute statistical facts. For an overview of the controversies surrounding causalist versus statisticalist interpretations and the precise definition of genetic drift, see Pence (2021).

[4] Strevens argues that deterministic models are *asymptotic idealizations*, useful models *extrapolated* from the finite stochastic models discussed above, but not uniquely fixed by them in the $N \to \infty$ limit. This view preserves, e.g., Gillespie's talk of deterministic models as "infinite population models," if only as a conventional choice of language. I find myself agreeing with what Strevens calls his "Nortonizing voice"—a voice which, in short, suggests that we should not be wedded to such conventions (2019, p. 1727). But Strevens also gives good reasons to preserve infinity-talk in this particular context, and I do not want to take a hard stance on this strand of the debate.

Abrams (2006), writing before Norton's work on idealization, also argues for a different resolution of the paradox: he claims that problem of ordering is not applicable because (a) the infinite-population limit uses a prior notion of probability, and (b) the limit depends only on *N* rather than any particular specification of population members. Notably, (b) anticipates my response to Hájek's (3), below (Sect. 4.1).

Kolmogorov's laws. Of course, HF makes poor sense of this explanation—just as SV, BG, SR, and GD make poor sense of their respective explanations.

In sum, it seems that limits of relative frequencies explain laws that approximately govern empirical systems, just like the Boltzmann-Grad limit, the low-velocity limit, and the infinite-population limit. If one takes an HF-style approach to these latter limits, we get accounts that make poor sense of well-regarded scientific explanations. Parallel reasoning motivates us to take an approximation-first approach to frequentism. But what, exactly, should this approach look like?

3 Approximation-First Frequentism

This section takes a stab at formally stating an approximation-first approach to frequentism. I slightly modify von Mises's definition of a collective in order to sharpen my statement of AF (Sect. 3.1). Then I demonstrate how the resulting definition of probability explains Kolmogorov's axioms as approximately true laws of nature (Sect. 3.2).

3.1 Getting Rid of Randomness

Von Mises provides an excellent starting point for our journey to an approximation-first view. He famously asserts that the term "probability" only makes sense when defined in terms a *collective*, an infinite sequence of attributes satisfying two formal criteria. One of these states that the relative frequency of each attribute converges to a fixed value. The other asserts that this value cannot change whenever one selects a subsequence using some fixed rule or place selection. Many have criticized von Mises on the grounds of the latter condition, which aims to fix a sense in which the sequence of outcomes is random. In this vein, it is perspicuous to view collectives as attempted—but failed—idealizations. To no avail, Von Mises tried to cook up an analogy between the randomness of chancy events and some property of his limit systems. Still, we might yet extract a fruitful *approximation* from his recipe.

To do so, let us unpack his formal definition. We start with a set X of arbitrary size, the elements of which comprise all theoretically possible outcomes or attributes of a statistical experiment. Then we fix an algebra Σ of subsets of X that represent coarser-grained attributes, ones that we might discern in practice. So, for example, if $A = \{x_1, x_2\}$, then x_1 and x_2 refer to the different states of affairs that might underlie the observed attribute A. The requirement that Σ is an algebra allows us to talk about attributes using the familiar language of classical propositional logic. To wit: Σ must contain the empty set and X itself, corresponding to falsity and truth, respectively; it must be closed under complements, corresponding to the "not" connective of classical logic; and it must be closed under finite unions and intersections, corresponding to the "or" and "and" connectives, respectively.

Suppose, now, that $\mathcal{K}$ is an infinite sequence of elements of X. We let f_N denote the function returning the relative frequency of an attribute in the first N outcomes of $\mathcal{K}$. If x_1 and x_2 occur five times in the first seven elements of $\mathcal{K}$, for instance,

then $f_7(A) = \frac{5}{7}$. And if x_2 occurs just three of those five times and $B = \{x_2, x_3\}$, then $f_7(A \cap B) = \frac{3}{7}$. And so on. In the infinite limit, these frequencies will become our probabilities. Thus, the convergence requirement ensures that we can make sense of probability talk for any logically-expressible attribute.

The randomness requirement is where the trouble comes in. To motivate it, note that some convergent $\mathcal{K}$ have a glaring defect when viewed as idealizations: the outcomes in $\mathcal{K}$ might follow some predictable rule, while actual chancy events never do. A sequence that is intuitively *not* random helps to illustrate this point. Consider, for instance, the sequence of coin tosses that lands heads-tails-heads-tails forever:

$$\text{HT HT HT HT} \ldots$$

This sequence has a limiting frequency of one-half for heads. But gamblers can easily exploit it by betting heads on every other toss. That is a far cry from what gamblers can do with *actual* coins. Note, however, that we can easily define a rule that selects every odd-numbered toss, and this rule picks out a subsequence where the limiting frequency of heads is one. It seems, then, that we might be able to explain the predictability of this sequence as follows: the limiting frequencies change under an *admissible* place selection, one that is easily specifiable like the odd-number rule. It will not do to allow for *any* place selection. Consider, for example, a rule that depends on the outcome x_n: the rule that selects all heads outcomes wherever they appear, creating a sequence with a limiting frequency of one for heads regardless of that limit's original value. A hypothetical gambler might exploit this rule if they could access some oracle who knew all the outcomes in advance. *Actual* gamblers, however, are not so lucky! Thus, it seems that one might characterize admissible rules as those rules that real gamblers could exploit, in some sense or another. Prima facie, such a characterization *might* provide a fruitful analogy between infinite sequences and finite sequences. This analogy would not do away with Hájek's problem of ordering; the limiting frequencies of infinite sequences cannot be invariant under re-orderings like the relative frequencies of finite ones. Nevertheless, the former sequences might still be *random* like the latter ones, in the sense that gamblers could not exploit either.

Let us formalize this intuition. We say that a place selection is some formal rule that picks out a subsequence of $\mathcal{K}$ to create a new infinite sequence, one that might or might not yield new limiting frequencies for attributes like A. Let γ be the set of all and only the admissible rules, each of which specifies just one subsequence of $\mathcal{K}$. We now have the tools for a mathematically precise statement of von Mises's frequentism.

Von Mises's frequentism. A collective $\mathcal{K}$ is an infinite sequence of elements of X satisfying the following criteria:

Convergence. For any $A \in \Sigma$, the limit of $f_N(A)$ as $N \to \infty$ exists.

Randomness. Each rule in γ specifies a subsequence of $\mathcal{K}$ that preserves these limiting frequencies.

The probability p of $A \in \Sigma$ relative to a collective $\mathcal{K}$ is the limiting value of the relative frequency of A in $\mathcal{K}$,

$$p(A) := \lim_{N \to \infty} f_N(A).$$

Note that this definition splits neatly into two halves. On the one hand, the condition of convergence specifies a limiting property, the limit of the finite frequencies of our target system. On the other, the condition of randomness pinpoints a property of the limit system that, prima facie, might bear a fruitful analogy with the target.

However, while von Mises had high hopes for this analogy, most philosophers of probability agree that its outlook is rather dim. In Diaconis and Skyrms' (2017) retelling, Per Martin-Löf nailed the coffin shut. Martin-Löf (1969) pinpoints an unavoidable difficulty in using collectives to explain randomness: so long as we can enumerate all of the admissible place selections, it is impossible to rule out sequences that a gambler can exploit. A bit more precisely, he cites the following result of Ville (1939): as long as γ is countable, there is a collective $\mathcal{K}$ for the fair coin where, after some finite initial segment, the relative frequency of heads approaches its limiting value of one-half from above. A gambler can exploit this sequence. As long as they always bet heads, they are sure to eventually end up with a net gain, at which point they can walk away. If we want collectives to rule out these sorts of sequences, we need to allow for an *uncountable* set of admissible place selections. This fact rules out Church's (1940) proposal to define admissible place selections using computable functions, which are countable in number. Coming up with a principled characterization of admissibility now seems quite difficult, indeed.

Martin-Löf takes this issue to reveal a deep defect in von Mises's program. However, it is worth stressing that these authors have different core concerns. Von Mises seeks to explain as much about probability as he can—including probability functions' conformity to Kolmogorov's rules—in terms of facts about relative frequencies. Martin-Löf, on the contrary, chiefly aims to explicate the essential properties of Bernoulli sequences, infinite sequences of independent and identically-distributed two-valued random variables.[5] Perhaps accordingly, Martin-Löf is happy to take Kolmogorov's rules as an axiomatic definition of probability, one that does not require explanation in terms of frequencies. He then constructs his own immensely fruitful explanation of the randomness with Kolmogorovian probabilities already in tow.[6] As Diaconis and Skyrms note, this account succeeds in characterizing the

---

[5] While Martin-Löf (1966) presents his definition of a random sequence as a competitor to von Mises's definition of a collective, he arguably elides this difference between their projects' aims. He seems to identify collectives with Bernoulli sequences; he writes, for example, "Bernoulli sequences (Kollektivs, in the terminology of von Mises) [...]" (Martin-Löf 1966, p. 618). But von Mises seems to have a more flexible conception of what a collective can be, as evidenced by his discussion of "chance" (1981, p. 29).

[6] Very roughly, Martin-Löf considers the probability of infinite sequences based on the binomial distribution, e.g., the probability of a sequence of outcomes for the tosses of a fair coin. He then approximates a definitively non-random infinite sequence by a nested series of ever-more improbable ones. So, for example, if U_1 contains all sequences starting with "heads, ..." U_2 might contain all sequences starting with "heads, tails, ...," and so on, as long as the probability of U_n is no greater than $(\frac{1}{2})^n$. Such a sequence constitutes a Martin-Löf test; if a sequence lies in the intersection of all the U_n, it fails the test. Contrarily, we say that a sequence that passes every such test is Martin-Löf random. One can precisely define a gam-

randomness of an infinite sequence, but it does not offer a von Mises-style analysis of probability (2017, pp. 163–164). It uses Kolmogorov's axioms to explain a property of an infinite system, rather than the other way around.

However, the von Mises-style frequentist should not throw in the towel just yet. Von Mises himself suggests a live option (and one that is, in fact, compatible with Martin-Löf's strategy): we can remove the demand that our *definition* of probability explicate randomness on its own. He notes that we can still make sense of a probabilistic concept when an infinite sequence of outcomes satisfies convergence but fails to satisfy randomness (although he prefers to use the term "chance" in this context; 1981, p. 29).[7] Let us call such a sequence a *chance collective*. A chance collective allows for a weakened version of von Mises's frequentism.

Von Mises's frequentism (weakened form). A chance collective $\mathcal{K}$ is an infinite sequence of elements of *X* satisfying the following criterion:

Convergence. For any $A \in \Sigma$, the limit of $f_N(A)$ as $N \to \infty$ exists.
The probability p of $A \in \Sigma$ relative to a chance collective $\mathcal{K}$ is the limiting value of the relative frequency of A in $\mathcal{K}$,

$$p(A) := \lim_{N \to \infty} f_N(A).$$

Conveniently, this weakening leaves us with *just* the approximating limit from von Mises's original definition. So we are free to interpret chance collectives as *mere approximations*, ways of precisely specifying a limit property that partially describes a finite target.

This weakened frequentism embraces the "pathological" sequences described above. Both the repeating heads-tails-heads-tails sequence and the converging-from-above sequence are viable chance collectives. So one might object that my weakened proposal does nothing to address the problem that Martin-Löf poses. I agree! Chance collectives lack a property of randomness, and so frequentists cannot use them to explain why gamblers cannot exploit chancy events. Still, I do not think that frequentists need to give such an explanation, as long as they are upfront about the fact that they are only giving *partial* descriptions of their targets. Frequentists should view an account of randomness as icing on the cake, something nice to have but not a tragedy to lose. At first, it might seem a shame that our limit systems possess an additional disanalogy with our finite targets. However, given the problem of ordering, we were skeptical of these limit systems from the jump. What is crucial is that our limit *properties* (the limiting frequencies) remain explanatorily powerful: they suffice to explain all of Kolmogorov's rules (as I will argue in the next section).

Footnote 6 (continued)

bling procedure, a computable martingale, that cannot win for a sequence if and only if that sequence is Martin-Löf random. For more details, see Martin-Löf (1966) and Diaconis and Skyrms (2017).

[7] Martin-Löf (1969) notes that the basic idea of getting rid of randomness also has precedent in the work of Tornier (1936). However, Tornier couches this idea in a different mathematical formalism, one invoking matrices of outcomes rather than sequences.

Given the prevalence of mere approximations elsewhere in science, this much might be enough.

This move is made all the more appealing by the work of Abraham Wald, which, as even Martin-Löf is happy to admit, "did away with all purely mathematical objections" against von Mises's collectives (1969, p. 30). Wald's key result goes as follows: so long as Σ and γ are countable, there exists, for any finitely additive probability measure p on Σ, a *continuum* of collectives where the limiting frequencies are identical to the probabilities (1938, Theorem III, pp. 84–85). As a special case, let γ be the empty set (which is countable), so that every chance collective is a collective by the lights of γ. Thus, Wald's result immediately implies that there exists a continuum of *chance* collectives recovering any finitely additive p on any countable Σ. That means that we can use a chance collective to explain nearly any probability function we care to use in the sciences. It turns out that we can use a simple and widespread assumption from physics to explain most of the rest—namely, the assumption that the properties of empirical systems ought to be, in some relevant sense, continuous.

I will make good on that last promise in the next section. First, I want to address an issue with the above statement of the weakened form of von Mises's frequentism: right now, the role played by approximation is left wholly implicit. It would be better to be direct about the relevant features of our target system. Let us do so with the reference class *B*. *B* ought to specify the empirical setup that produces the stable frequencies, and it ought to specify our desired error tolerance ϵ. We can use these facts about *B* to determine a chance collective, $\mathcal{K}_B$, that suitably represents it. As such, we now have one way of filling in the square brackets in the earlier statement of AF to obtain a mathematically precise and philosophically transparent definition[8] of probability:

(AF) The probability of an attribute *A* in a finite reference class *B* is *p* if and only if, in the chance collective $\mathcal{K}_B$ that represents *B*, the limiting frequency of *A*,

$$p(A) := \lim_{N \to \infty} f_N(A),$$

yields a value of $p(A) = p$ that approximates the relative frequency of actual occurrences of *A* within *B*.

Just as before, we say that the limit approximates the relative frequency when it converges to a value that is a good match for the empirical data—that is, when it is empirically correct for a not-too-liberal choice of ϵ.

I am leaving the matter of what it means for $\mathcal{K}_B$ to represent B deliberately vague, as many factors might justifiably play a role. B might include details about the physical laws governing the system in question, laws which might favor one choice of limiting values over another. Or B might contain empirical facts about past stable frequencies. Or it might contain both. In any case, what is crucial is that $\mathcal{K}_B$ is appropriate for the *empirical setup* and the *error tolerance* that B specifies.

To see how this works in practice, consider typical tosses of a fair coin. On AF, the hypothesis that the coin is "fair" means that the observed frequency of heads in a sequence of flips will differ from one-half by no more than some specified error tolerance, ϵ. In this case, Σ might be the Boolean algebra generated by the propositions H for "the coin lands heads" and T for "the coin lands tails," and B might include the claims that the coin is evenly-weighted and that agents toss it in the usual way. B should also specify our error tolerance; say, $\epsilon = .05$. In this case, the symmetry of the coin motivates our choice of a $\mathcal{K}_B$ with symmetrical limiting frequencies, one-half for heads and one-half for tails. We could, instead, include past tosses of the coin in B and use this fact to fix $\mathcal{K}_B$ instead. Either way, $\mathcal{K}_B$ approximates the data well. Nevertheless, changes in the empirical setup might require a different choice. If the coin is weighted unevenly or a pathological agent "tosses" the coin by dropping it from a short height, we ought to specify a different reference class B' and assign it a new collective. Changes in our error tolerance might require a new collective, too. Chance collectives are objective because they are grounded in the properties of empirical systems, but they are free to grow and change along with our theories and measurements of these systems.

Conveniently, typical coin tosses also offer a paradigm of this growth process. Consider a slightly different Σ for these tosses: the Boolean algebra generated by the propositions S for "the coin lands with the same side up as that which we flipped it from" and D for "the coin lands on the other side." Let B be the same as before. One might think that it would suffice to take the old collective $\mathcal{K}_B$ and replace all the Hs with Ss and all the Ts with Ds. But with regards to S and D, it turns out that the usual way that agents toss fair coins is ever-so-slightly biased. Diaconis et al. (2007) argue via both physical analysis and extensive empirical trials that such flips tend to land on the same side on which they started, with a probability of about .51. Suppose, then, that we were to decrease our error tolerance to $\epsilon < .01$ and that the number of coin flips were to be suitably large (in a sense that I will make precise in the next section). Then the strategy of substituting the propositions in $\mathcal{K}_B$ would no longer do. We would need a new collective yielding a limiting frequency for S that is closer to .51.

In any of these cases, the bound ϵ introduces some unavoidable ambiguity in our choice of chance collective. A $\mathcal{K}'_B$ with limiting values of .5001 for heads and .4999 for tails, for example, will do just as well as $\mathcal{K}_B$ to define probabilities for B, according to AF. So, to recover talk of *unique* probabilities, we fix by convention *just one* suitable chance collective for any given reference class B (hence the reference to *the* collective that represents B in AF). Of these, any of the continuously many $\mathcal{K}_B$ with the same limiting values will do. Nevertheless, note well that convention plays a minimal role in this account. The empirical setup and the error tolerance do the bulk of the work.

AF is quite the denuded frequentist definition of probability, to be sure! The sense in which it says "objective probabilities" are "objective"—namely, that the appropriate limiting values depend on the empirical setup—is thin. The error tolerance introduces a hefty subjective element to it, too. Moreover, AF simply *abandons* the goal of explaining the randomness of chance events. Can it do any explanatory work at all? In the next section, I argue that AF, as meager as its commitments are, provides a strikingly rich explanation of Kolmogorov's axioms as approximate laws of nature governing regularities in finite sequences.

3.2 Kolmogorov's Axioms as (Approximate) Laws of Nature

As discussed above, the $v/c \to 0$ limit of special relativity explains why the conservation of classical momentum holds as an approximate law of nature. In short, when v/c is small enough, this law yields predictions that are close enough to actual values to be considered empirically correct. That is precisely the sort of explanation that AF gives for Kolmogorov's axioms, as I will now detail.

But first, a caveat. As Chakravartty (2007, 2010) notes, scientific realists have formed a longstanding (albeit rough) consensus that strict truth is too much to demand from scientific theories—but they still tend to disagree about what it means for a theory to be "approximately true." There are several well-developed theories of approximate truth on offer (see, e.g., Popper, 1972; Oddie, 1986; Aronson, 1990). However, Chakravartty is skeptical that any of them capture the qualitative nature of the sort of approximate truth yielded by approximations and idealizations. Palacios and Valente (2021) agree, and they suggest a minimal criterion for approximately true theories; namely, just that the theory is empirically correct in the sense detailed Sect. 2. That is, they take a theory to be approximately true when that theory predicts real values for properties of a system within a tolerable ϵ of their empirically-measured values. I consider this condition a good starting point for someone aiming to develop an account of approximate truth suitable for the scientific realist. In other words: whatever else it means for Kolmogorov's laws to be approximately true, it should at least mean that they are empirically correct.

Let us briefly review these laws, starting with what Kolmogorov (1956) calls the "elementary" theory. This theory contains all the laws that any probability function *must* satisfy, and collectives derive these laws exactly. We can state the first three of these as follows, for a countable event algebra Σ defined on a sample space X and a function $p : \Sigma \to \mathbb{R}$.

Non-negativity. The probability of any $A_i \in \Sigma$ is greater than or equal to zero;

$$p(A_i) \geq 0.$$

Normalization. The probability of one of the possible states in X occurring is one;

$$p(X) = 1.$$

Finite additivity. Let I be a totally-ordered and finite set of indices. For any subset $\{A_i\}_{i \in I} \subset \Sigma$ where $A_i \subseteq X \setminus A_j$ for all $i \neq j$,

$$p\left(\bigcup_i A_i\right) = \sum_i p(A_i).$$

Like a host of other scientific theories (and especially physical theories), Kolmogorov's theory also has a system-subsystem relation. The relation here is just conditional probability, the definition of which allows for modifications to the algebra of events. For instance, suppose we wish to restrict our attention to only those attributes in Σ that are present when A_i is present. Formally, these are elements A_j such that $A_j \cap A_i = A_j$. These elements form a subalgebra, $\downarrow A_i$. Via the usual definition of conditional probability, the state p on Σ yields the following state on the subsystem $\downarrow A_2$, supposing that $p(A_2)$ is non-zero:

Conditional probability. When $p(A_2) > 0$, the probability of $A_1 \in \Sigma$ conditional on $A_2 \in \Sigma$ is equal to the joint probability of A_1 and A_2 over the probability of A_2:

$$\text{if } p(A_2) > 0, \quad p(A_1 \mid A_2) = \frac{p(A_1 \cap A_2)}{p(A_2)}.$$

If $p(A_1 \mid A_2) = p(A_1)$, then we say that A_1 is *independent* of A_2.[9] These four rules exhaust Kolmogorov's formal description of the core, elementary theory. He goes on to strengthen the finite additivity for infinite cases but stresses that he "*arbitrarily*" restricts his attention to functions that satisfy it (1956, p. 15, emphasis Kolmogorov's). It is not hard to give a better justification for this move (as Kolmogorov is probably well aware; more on that in a bit). But I think Kolomogorov is quite right to stress, here, that the above four rules are the only hard and fast requirements on probability.

AF provides a straightforward and exact derivation of these rules. Given that Σ is countable, there are (continuously) many chance collectives $\mathcal{K}_B$ to choose from; let us pick any one of these. Note first that each $f_N(A_i)$ must be a rational, non-negative number. If an infinite sequence of rational, non-negative numbers converges, it must converge to a non-negative real number. So non-negativity is satisfied. Similarly, the

---

[9] Note well that Kolmogorov's definition of conditional probability in his elementary theory does not allow one to conditionalize on probability-zero events! Fitelson and Hájek (2017) criticize the definition on these grounds. That being said, nothing requires the approximation-first frequentist to stop here. As Meehan (2021) notes, Kolmogorov generalizes the definition when he extends his theory to infinite algebras Σ, and the general definition recovers the elementary one as a special case. Contra Fitelson and Hájek, Meehan cogently argues that we have good reasons to adopt Kolmogorov's general definition of conditional probability. It would be interesting to see if AF can naturally justify this definition, like how it justifies countable additivity below. But I leave this matter for future work.

relative frequency of *some* state obtaining, $f_N(X)$, must be one for every N. An infinite sequence of ones converges to one, yielding normalization.

Finite additivity follows from the continuity of addition. Let A_1 and A_2 be incompatible attributes; formally, this means that $A_1 \cap A_2 = \emptyset$. The relative frequency of A_i is the number of occurrences of A_i over N (for $i = 1, 2$). But whenever A_1 occurs in $\mathcal{K}_B$, A_2 cannot. So the relative frequency that one or the other occurs in the initial N trials of $\mathcal{K}_B$ is just the frequency of the first added to that of the latter: $f_N(A_1 \cup A_2) = f_N(A_1) + f_N(A_2)$. Generalizing, if we have a finite set $\{A_i\}_{i \in I}$ of incompatible attributes whose disjunction is defined, then

$$p\left(\bigcup_i A_i\right) = \lim_{N \to \infty}\left(\sum_i f_N(A_i)\right) = \sum_i \left(\lim_{N \to \infty} f_N(A_i)\right) = \sum_i p(A_i).$$

The second equality follows from the continuity of addition, while the first and third follow from AF. So finite additivity is assured.

We obtain the definition of conditional probability by looking at appropriate subsequences of $\mathcal{K}_B$. To start, suppose we want to describe the relative frequencies of attributes under the assumption that A_2 must occur. Let us restrict our attention to the case where the limiting frequency of A_2 is greater than zero. In this case, the selection rule picking out events in $\mathcal{K}_B$ where A_2 occurs yields a subsequence, and this subsequence is a chance collective for $\downarrow A_2$. Pick any $A_1 \in \Sigma$, and note that the limiting frequency of A_1 in this subsequence is equal to the limit of the fraction of A_2-events in $\mathcal{K}_B$ that also yield A_1. In other words:

$$\text{if } \lim_{N \to \infty} f_N(A_2) > 0, \quad p(A_1 \mid A_2) = \lim_{N \to \infty} \frac{f_N(A_1 \cap A_2)}{f_N(A_2)},$$

where $p(A_1 \mid A_2)$ refers to the limiting frequency of A_1 in the subsequence of A_2-events in $\mathcal{K}_B$. Then, due to the continuity of quotients, we have

$$\text{if } p(A_2) > 0, \quad p(A_1 \mid A_2) = \lim_{N \to \infty} \frac{f_N(A_1 \cap A_2)}{f_N(A_2)} = \frac{\lim_{N \to \infty} f_N(A_1 \cap A_2)}{\lim_{N \to \infty} f_N(A_2)} = \frac{p(A_1 \cap A_2)}{p(A_2)}.$$

That completes our derivation of Kolmogorov's laws for countable algebras Σ. So long as the empirical data verifies that the finite frequency of each $f_N(A)$ is tolerably close to the limit value $p(A)$, then these limits serve as reasonable approximations of our target. Kolmogorov's core axioms then come out as approximately true (i.e., at least empirically correct) laws governing finite relative frequencies.

That might suffice, but I think AF also lends Kolmogorov's strengthening of additivity a compelling justification. To recover this strengthening, allow the I in our statement of finite additivity to be countably infinite. Notably, Kolmogorov himself does not introduce this new condition—countable additivity—as a brute posit. Rather, he *derives* it from an additional continuity assumption (1956, pp. 15–16). Selecting for continuity makes good sense on AF. AF directs us to find a chance collective that represents B, and this choice can (and should) be sensitive

to common modeling practices. One of these takes models of empirical systems to obey something like Leibniz's Law of Continuity—which, in slogan-form, states that nature makes no leaps.

On the current pragmatic approach, it helps to think of this "law" as an empirically-motivated constraint on models of physical systems. Our physical theories describe many phenomena—from the transfer of heat to the generation of electricity to the motion of pool balls—as the gradual change of continuous physical quantities. In particular, such theories usually assign a system some state q in a space of states $\mathcal{M}$, and they represent an observable quantity with some continuous function f of q. Mathematically, f being continuous means that for any convergent sequence of q_js in $\mathcal{M}$,

$$f\left(\lim_{j\to\infty} q_j\right) = \lim_{j\to\infty} f\left(q_j\right).$$

There are many ways to motivate this condition. For one, differential equations elegantly and concisely describe the sorts of gradual changes mentioned above, and f must be continuous if it is differentiable. Still, not all changes are gradual. Physicists often use discontinuous functions to describe phase transitions, like the spontaneous magnetization of iron in an external magnetic field. In short, the continuity constraint is very helpful for describing some systems and less helpful for describing others.

Here is how this story plays out for probability: if we think that our probabilistic system makes no leaps, then attributes that grow ever-closer ought to have ever-closer relative frequencies. In other words, our probabilities should obey an appropriate analog of the above equation for f. But what does it mean for attributes to be "ever-closer"? One useful notion comes from set theory. Take an infinite sequence of sets A_i in Σ such that $\bigcup_i A_i = A$, where we let $I = \mathbb{N}$. One can rewrite this sequence as a sequence of nested sets, A_1, $A_1 \cup A_2$, $A_1 \cup A_2 \cup A_3$, and so on. We can define the limit of this sequence as the set of elements contained in all but a *finite* number of sets in this sequence. Then we have

$$\lim_{j\to\infty} \bigcup_{i=1}^{j} A_i := \left\{ x \ \middle| \ \exists j' \, \forall j > j', \, x \in \bigcup_{i=1}^{j} A_i \right\} = \bigcup_i A_i.$$

Now we can choose to require that our approximate descriptions of relative frequencies respect this notion of continuity. In other words, we can require that p is continuous with regards to the set-theoretic notion of convergence defined above,

$$p\left(\lim_{j\to\infty} \bigcup_{i=1}^{j} A_i\right) = \lim_{j\to\infty} p\left(\bigcup_{i=1}^{j} A_i\right).$$

Note that the expression on the right-hand side defines the infinite sum $\sum_i p(A_i)$. So this expression is identical to the requirement of countable additivity.[10] If a probability function is countably additive, then it is finitely additive, and so there are $\mathcal{K}_B$s that yield it. Countable additivity thus amounts to the condition that we restrict our attention to approximations that make no leaps—a restriction that might make good sense for some systems and less sense for others.

This much still leaves out some rather important probability functions, including the Gaussian distribution that we used to justify deterministic models in population genetics. Typically, when we define probability density functions like the Gaussian directly on our sample space X, we let them fix the probabilities of subsets of X that we take to be "measurable." We usually designate the Borel sets as such. Recall that the algebra of Borel sets is the σ-algebra closure of the collection of open sets of X, i.e., its closure under complements and countable unions and intersections. In general, the algebra of Borel sets of X is uncountable. However, as long X is second-countable (like, for example, $\mathbb{R}^n$), it has some countable basis $\mathcal{B}$ for its topology. Let $\Sigma(\mathcal{B})$ denote its algebraic closure, i.e., its closure under complements and finite unions and intersections. Now let $\sigma(\Sigma(\mathcal{B}))$ denote the σ-algebra completion of *that* algebra. $\sigma(\Sigma(\mathcal{B}))$ is the algebra of Borel sets for X. Now note that there is a chance collective recovering any countably additive probability measure on $\Sigma(\mathcal{B})$. By Carathéodory's extension theorem, every such measure has a unique extension to one on $\sigma(\Sigma(\mathcal{B}))$. That suffices to recover the standard way that we use probability density functions on $\mathbb{R}^n$. In short, AF recovers countably-additive measures on the Borel sets of X as the continuous extensions of its continuous approximations.[11]

---

[10] Kolmogorov (1956) derives countable additivity from a slightly different (but mathematically equivalent) continuity assumption; his condition asserts that the probabilities assigned to a nested sequence of sets of decreasing size ought to converge to zero. Interestingly, Elliot (2020) reads Jaynes (2003) as giving an argument very similar to one that I present here. Jaynes, however, thinks that countable additivity is a necessary condition, and Elliot is quite right to stress that the continuity is not *mathematically* necessary. However, I only mean to assert that the condition jibes well with common modeling practices.

Moreover, I only mean to assert that continuity is a particularly perspicuous path to countable additivity for *frequentism*. The issue, of course, arises anew in the Bayesian context. de Finetti (1974) is famously skeptical of applying countable additivity in this context, but many have given excellent reasons to reevaluate his arguments; see (among others) Seidenfeld and Schervish (1983), Williamson (1999), Howson (2008), and Easwaran (2013) for further discussion.

[11] This basic strategy is originally due to Wald (1938). Note, however, that Wald uses the stronger condition of complete additivity ("Totaladditivität") that allows I to be uncountably infinite. Moreover, he does not appeal explicitly to continuity to justify it. Perhaps accordingly, Martin-Löf asserts that Wald's strategy is "unmotivated" and "clearly unsatisfactory" without further comment (1969, p. 34).

van Fraassen (1977) offers a more interesting critique, which runs as follows. To start, note that the Gaussian assigns probability zero to every Borel set of $\mathbb{R}$ containing countably many points. But then note any collective $\mathcal{K}$ defined on X comprises just such a Borel set. With this idea in mind, van Fraassen asserts that because the probability function "assigns zero to every countable point-set, it is not identifiable with any relative frequency" (1977, p. 138). Of course, $\mathcal{K}$ is not included in the algebra $\Sigma(\mathcal{B})$. Still, van Fraassen might mean to suggest that it is arbitrary to ignore the fact that the collective gives the Borel set $\mathcal{K}$ a natural or intuitive limiting value of one. In my view, however, this suggestion amounts to a different proposal for extending the limiting frequencies defined by a continuous collective. Van Fraassen runs a similar argument against an example due to Reichenbach. That argument yields a different undesirable generalization of the core, countable-Σ collectives. In both cases, van Frassen suggests that the lack of a unique extension deprives limiting frequencies of their ability to represent probabilities. I view this lack of uniqueness as a resource: it allows the frequentist to accommodate various modeling assump-

As a bonus, we can use our definition of probability to help determine when N is large enough for a very small ϵ. The idea is to exploit the recursive structure of repeatable trials. A trial is simply a test of the propositions in Σ, and we have much flexibility as to what we put in Σ. Σ might be the Boolean algebra generated by a single proposition A. Consider, then, N independent tests of Σ. We can treat each test as a new proposition: A_1 or its negation for the first test, A_2 or its negation for the second, and so on. Let Σ' be the Boolean algebra generated by *those* propositions. Now pick a chance collective $\mathcal{K}_B$ for Σ with a limiting frequency of p for A. Under the assumption that the limiting frequencies for $A_1, A_2, \ldots, A_N$ in Σ' are independent and identical to p, we can derive what they must be from Kolmogorov's rules. As it turns out, they must agree with the binomial distribution. Thus, we can find limiting values for seeing m instances of A in the N tests. Of course, the actual value of m/N for any given trial of Σ' will deviate from p. But a good measure of such deviations is given by what we usually call the standard deviation,

$$ \sigma = \sqrt{\frac{p(1-p)}{N}}, $$

as it turns out that the binomial distribution specifies a limiting value of about .98 for m/N to fall within 3σ of p. As N goes to infinity, σ approaches zero. So we can use our theory to assess how large N needs to be to justify a small ϵ. In particular, the above expression supports Diaconis et al.'s (2007) claim that we need roughly a quarter of a million typical coin tosses to reveal the dynamical bias they predict.

I take AF to be a ruthlessly pragmatic approach to frequentism. It says barely anything about the world beyond what the empiricist holds sacred: those roughly-stable finite frequencies that we observe *really are there*, and they *really are well-described* by infinite limits. One might reasonably say more. In particular, one might think that limiting frequencies being "close enough" to actual frequencies in large-enough sequences of type B implies that the limiting frequency is, in some sense, *typical* of B. Hubert (2021) makes this intuition rigorous with an account that he calls *typicality frequentism*. By stipulating a typicality measure on the state space of some physical theory with in-principle infinitely-repeatable measurements, Hubert defines the probability of some state as the limiting value that its relative frequencies typically approach. I like this analysis, but I want to stress that a simpler and weaker one—approximation-first frequentism—is available. AF does not require any antecedent physical theory or definition of typicality.

Moreover, I think that AF is particularly well-suited to address Hájek's arguments against HF. These arguments serve as excellent foils for the approximation-first view. They help pinpoint the dangers of treating limiting frequencies as robust idealizations.

Footnote 11 (continued)

tions, including (but not limited to) continuity. I elaborate on this idea in my discussion of Hájek's (13), below (Sect. 4.2).

4 Fifteen Foils for the Approximation-First View

Hájek's arguments against HF can be roughly split into three main themes: giving up on empiricism, technical difficulties, and issues with the explanation. There is some overlap amongst these, but I group five arguments into each category (to respect symmetry). I will discuss how AF recasts each group of arguments in turn (Sects. 4.1–4.3) and close by considering a classic worry about "circular" appeals to the Law of Large Numbers (Sect. 4.4).

4.1 Giving Up on Empiricism?

The giving-up-on-empiricism theme canvasses Hájek's first four arguments and the ninth: (1) "An abandonment of empiricism," (2) "The counterfactuals appealed to are utterly bizarre," (3) "There is no fact of what the hypothetical sequences look like," (4) "The problem of ordering," and (9) "Subsequences can be found converging to whatever value you like."

(1) and (3) target higher-level reasons that one might take hypothetical frequencies to depart from general empiricist commitments. As to (1): Hájek argues that hypothetical frequencies are "unknowable in the strongest sense"—observers would have to live forever to see them or otherwise measure events in increasingly-shorter time intervals in some "Zeno-like" nightmare (2009, p. 214). So, Hájek claims, a particular sort of empiricist should balk. For example, a logical empiricist in the style of Carnap ought to note that there is no reasonable observable vocabulary that contains such entities.

That might well be so. Nevertheless, I think that scientific realists of the sort described above comprise the best modern audience for frequentism. These realists have broken definitively with the logical empiricism of Carnap. They have accepted the need for approximate truth; they recognize the importance of idealizations and approximations. And they are more than happy to assert that other limiting values, such as instantaneous velocities, can be (at least) approximately true descriptions of empirical systems.

Interestingly, Hájek also argues that there is a relevant difference between limiting frequency and limiting velocity: one can know the value of the latter (but not the former) "well enough," he claims. But how so? An observer measures velocity with rods and clocks, recording finite distances traversed in finite time (and dividing the former by the latter). Perhaps an idealized observer might render these intervals ever-shorter. But now Zeno's ghost haunts us again—and in any case, ideal observers are not actual observers. Actual observers seem better off sticking with the approximation approach: they should treat the limiting value in question as an approximation of a finite target. Then they know (at least) when that value is empirically correct. This last claim is also true of limiting frequencies.

Hájek offers another reason why one might not know limiting frequencies: he subscribes to the view that knowledge is factive, and he argues (3) that there is no fact of the matter about what the hypothetical sequences look like. As Jeffrey (1992)

puts it, asking whether a coin lands heads on the tenth toss is like asking about the mass of the tenth planet. In either case, there is no fact of the matter. Hájek claims that the situation is even worse than this. He thinks that "it is consistent with the concept of mass" to answer the planet question, but giving a fact about the tenth toss denies the chanciness of the system.

This criticism, (3), is a fine one, as far as it goes—but the approximation-first frequentist already takes it to heart. One chance collective yielding the empirically correct limiting frequencies is as good as any other, and there are (continuously) many sequences to choose from. In particular, one can find a $\mathcal{K}_B$ and a $\mathcal{K}'_B$ that both yield the same limiting frequencies but give different outcomes for the tenth toss. One only picks one of these to represent B to precisely and consistently define our approximating limit. This choice of representation is a convention, and it does not grant us knowledge of the outcome of the tenth toss (or any other) in advance. It only grants us knowledge of the limiting frequencies, qua approximate descriptions of finite frequencies. Hájek might want to deny that one can know these latter facts. However, while this sort of response might hold water against HF, it begs the question against AF.

Hájek's (2) and (4) target more specific features of empirical practice that ostensibly do not jibe with hypothetical frequencies. In (2), Hájek argues that the frequentist posits counterfactuals that are just too bizarre. In his words:

> Consider the radium atom's decay. We are supposed to imagine infinitely many radium atoms: that is, a world in which there is an infinite amount of matter (and not just the 10^{80} or so atoms that populate the actual universe, according to a recent census). Consider the coin toss. We are supposed to imagine infinitely many results of tossing the coin: that is, a world in which coins are 'immortal' [...] In short, we are supposed to imagine utterly bizarre worlds. (2009, p. 216)

It is hard to take this concern very seriously in light of the above examples of approximating limits in science. One does not ever intend to apply the Boltzmann equation to a gas of, say, a centillion particles—a gas with far more than 10^{80} particles, by many orders of magnitude! Nonetheless, the Boltzmann-Grad limit alludes to such systems on its way to an exact derivation of ideal-gas dynamics. In both this case and the case of AF, the limiting property is true—approximately—of large systems whose size lies below a given threshold.

Let us turn to (4), the problem of ordering. As flagged in the introduction, Hájek's (4) is a paradigm case of an idealization distorting its target. The value of a limiting frequency depends on the order of events, while the value of a finite relative frequency never does. This fact locates a necessary dis-analogy between the idealization and the target. It provides an excellent reason to think that infinite sequences cannot serve as idealizations of chancy events. But it gives no reason not to use these sequences as approximations.

Hájek appeals to (9), the fact that subsequences of outcomes can be found where (non-trivial) relative frequencies converge to whatever value you like, to provide another example that he means to support his arguments (3) and (4). If (4) makes one wary of treating collectives as idealizations, then (9) should seal the deal: it is

precisely the problem of defining admissible place selections. As discussed above (Sect. 3.1), this issue panned out rather poorly for von Mises! However, the response to this point is the same as for (3) and (4). Any suitable sequence of events will do on the approximation-first view, whether that sequence comes from a subset of some other model or not. Pathological subsequences are a problem for the idealization, not the approximation.

4.2 Technical Difficulties?

A good chunk of Hájek's arguments fall under the theme of technical difficulties. These arguments mainly arise from HF being a particular sort of modal claim. These issues disappear on the approximation-first view, as the limit merely describes one actual, finite target. They include (5) "The limit may not exist," (6) "The limit may equal the wrong value," (8) "The limit might exist when it should not," (13) "Limiting relative frequency violates countable additivity," and (14) "The domain of limiting relative frequencies is not a field."

For (5), Hájek notes that one can specify infinite sequences of outcomes where the limit of a relative frequency does not exist. Then he asserts that a fair coin might yield the outcome

$$\text{HT HHTT HHHHTTTT HHHHHHHHTTTTTTTT} \dots$$

of 2^n heads followed by 2^n tails. The relative frequency of heads does not converge but instead oscillates endlessly. Nonetheless, Hájek asserts that this lack of convergence does not contradict his claim that the coin is fair (and so has probability one-half). He also claims that no principle rules out this sequence. On the one hand, this claim seems a bit strange, as the same sort of counterfactual that Hájek labels as "bizarre" in (2)—one in which a coin is "immortal"—is now meant to be "no more impossible than your favorite 'well-mixed' sequence" (2009, p. 220). On the other hand, Hájek might mean to say that the truth of *any* modal claim about the above sequence of tosses is far from obvious, and so one's theory of probability should not speak one way or the other on the matter.

In either case, this worry is strictly one for HF. By appealing to a chance collective, AF ensures that the limits exist, nullifying the technical worry raised by (5). More to the point, AF does not, itself, make any modal claims: it asserts that the infinite limit is an approximately true description of one actual finite target (of a specific sort). While one might say more than this given a fully fleshed-out theory of approximate truth, it is not clear that one needs to. Possible worlds need not ever enter the frequentist's picture. On its own, AF is strictly silent on Hájek's "well-mixed" infinite sequence, nullifying the philosophical worry raised by (5).

Similarly, Hájek claims in (6) that a fair coin might land heads forever. In (14), he constructs a sequence where the relative frequencies converge for A and B but not for $A \cap B$ and suggests that this sequence might also be metaphysically possible. The same response applies to both: AF gives a mere approximation of an actual system. It does not itself take a stance for or against these modal claims.

(8) pursues the reverse strategy of (5), asserting that a limiting frequency might exist when it should not. Hájek gives three tentative examples of cases that might yield undesirable convergent frequencies: whether a dart hits a non-measurable subset of the dartboard; free acts like raising my hand on a Tuesday; and non-projectible predicates like the grueness of a gemstone. Here, again, Hájek is suggesting that these convergences might be actual modal possibilities. Whether they are or not might matter to HF, but it makes no difference to AF, which describes finite systems. Let us see, however, if these examples hold more weight if one takes them to specify undesirable *finite* frequencies rather than convergent ones. On this pass, the first example might still pose a problem. A scientist *could* use non-measurable sets to represent features of an actual dartboard, and they *could* take actual dart-throws to land inside or outside such sets definitively. If the relative frequency of "inside" outcomes is stable, it might be hard to build a good chance collective that approximates it. However, this choice of representation is, at the very least, questionable—measurable sets are called "measurable" for a reason! They provide an abundance of excellent candidates for representing regions of a dartboard where one can definitively say that the dart is in or out. Take, for example, all of the closed and bounded regions of $\mathbb{R}^2$; these are all measurable, and they seem to provide all of the descriptions we could need for regions of the dartboard that one can measure with a ruler. The other two examples are easier to handle. If scientists want to start recording my Tuesday hand-raises or my daily grueness when I am sick, they are free to do so. If the results yield stable frequencies (which is admittedly unlikely), then they are free to model them with chance collectives. But that is a feature, not a bug. It demonstrates that the view can accommodate nearly any reference class B that one deems worthwhile when doing science.

In (13), Hájek notes that limiting frequencies are not guaranteed to satisfy countable additivity.[12] He asserts that this fact is a "serious blow" against the frequentist "who holds sacred Kolmogorov's full set of axioms of probability" (Hájek 2012, p. 229). As discussed above (Sect. 3.2), frequentists can easily secure countable additivity when they want by choosing to enforce the Law of Continuity. Perhaps more to the point, it is not clear that scientists *should* hold sacred the full set of probability axioms. Kolmogorov himself was hesitant to do so. More importantly, merely-finitely-additive probabilities afford interesting representational options! Halvorson (2001) provides one notable example, using such probabilities to explicate continuous physical quantities in both classical and quantum mechanics. In particular, his account shows how merely-finitely-additive probabilities can represent particles with sharp position-values in *both* theories.[13] Frequentism, then, accommodates

---

[12] This criticism echoes earlier arguments made by van Fraassen (1977); for more on those, see footnote 11.

[13] Halvorson pursues a roughly operational approach to continuous physical quantities, which goes as follows. Typically, we take empirical measurements to pin down the values of such quantities to within some fixed interval on the real number line. However, we do not think that we can tell the difference between regions $[a, b]$ and $[a, b)$. One way to capture this intuition is to note that these sets differ only by the singleton set $\{b\}$, which is of Lebesgue measure zero. Halvorson, then, considers the algebra $\mathcal{B}(\mathbb{R})/\mathcal{N}$ of equivalence classes of Borel sets that differ only by sets of Lebesgue measure zero, and he defines probability functions directly on that algebra. This account allows for a sense in which a quantum particle might be said to have a precise position. Pick some point on the real number line, λ; for any equivalence class A containing an open neighborhood of λ, let $p_\lambda(A) = 1$. It turns out that p_λ extends to a probability function that gives a sure-fire response (either zero or one) for every equivalence class. However,

various representational aims in various contexts: the frequentist-scientist is free to decide that a precise position value is more important than continuity or vice versa. Once again, AF shows how an ostensible bug is, in fact, a feature.

4.3 Issues with the Explanation?

The remaining third of Hájek's arguments aim to find issues with the nature of the explanation that frequentism offers. However, these arguments are recast by AF's narrow explanatory scope. They include (7) "The order of explanation is back-to-front," (10) "Necessarily single-case events," (11) "Uncountably many events," (12) "Exchangeability, and independent, identically distributed trials," and (15) "No infinitesimal probabilities."

In (7), Hájek claims that von Mises explains the "regularities in the behavior" of actual sequences of outcomes by positing that they are initial segments of collectives (Hájek 2009, p. 225). This explanation is wrong, Hájek argues, because probability ought to explain long-run behavior, not the other way around. Note, however, that AF does not seek to *explain* the regularity of finite sequences. The data might or might not be regular, and our theory of the system might or might not posit exploitable symmetries. So there might or might not be a suitable limiting frequency. If the data *is* regular, then frequentism explains why it conforms (approximately) to Kolmogorov's laws. Hájek might argue for a different understanding of these laws. However, the understanding of Kolmogorov's axioms as approximate laws of nature is the one that AF is after, and long-run behavior suffices to explain them as such.

What of (10), the infamous problem of necessarily single-case events? On the one hand, it is tempting to point out that the sorts of empirical scenarios we set out to analyze rarely, if ever, concern such events, so we can safely ignore them. On the other hand, the approximation-first frequentist can helpfully say more: they can give a natural account of correct, objective probabilities for necessarily single-case events in terms of counterfactual (or, if need be, counter-possible) claims about *finite* sequences of them. Take Hájek's case of a "fair coin" that is only ever tossed once, after which it is destroyed. What could justify the claim that the fair-coin chance collective is better suited than any other to represent this case? A belief in a counterfactual might do so. One might reasonably believe that were the soon-to-be-destroyed coin instead flipped a large-but-finite number of times, the limiting frequencies of the fair-coin chance collective would approximate the outcomes well. Indeed, one uses similar reasoning when deriving the thermodynamic properties of a given gas with a statistical ensemble. There is only ever one gas with one microstate! All the other gases in the ensemble are counterfactual. Nevertheless, this finite collection of

Footnote 13 (continued)

this probability function is only finitely additive (Halvorson 2001, §2). We can recover this function by suitably extending a chance collective $\mathcal{K}_B$ consisting of an infinite sequence of λs and defined on a countable family of open neighborhoods of λ.

hypothetical gases justifies one's beliefs about the original gas's macrostate.[14] Just so with the ascription of even odds to the annihilated coin.

Hájek poses a converse problem in (11) by claiming that there might be interesting cases with uncountably many events. His example is the probability that some spacetime point $\mathbf{x} \in \mathbb{R}^4$ is such that a given physical field $\mathbf{F}$ has a particular magnitude $|\mathbf{F}(\mathbf{x})| = s$ at that point. Hájek seems to suggest that what ought to play the role of the collective, in this case, is something like a set of propositions $|\mathbf{F}(\mathbf{x})| = s$ or $|\mathbf{F}(\mathbf{x})| \neq s$ for each $\mathbf{x} \in \mathbb{R}^4$. The problem is that this set is uncountable. However, it is hard for me to imagine how to measure the magnitude of a field at a precise spacetime point. There is a similar scenario that seems more plausible: namely, checking whether the flux of $\mathbf{F}$ falls within some range Δs in a certain Borel set X of $\mathbb{R}^4$ (using, say, a Gauss-Meter in the case that $\mathbf{F}$ is a magnetic field). As I discuss above (Sect. 3.2), AF can easily handle a case like this one by assigning a (continuous) collective to some countable basis and extending the resulting probability function to all the Borel sets. Still, Hájek suggests that nothing favors one countable sequence of spacetime points (or, by extension, one countable sequence of spacetime regions) as being more representative than any other. For AF, this concern returns us to the question of when a chance collective represents a reference class. The matter of which objective factors ought to fix this representation is subtle. Given the examples above, however, it does not strike me as particularly mysterious.

In (12), Hájek argues that HF gets the explanandum wrong because not all trials are independent and identically distributed. He imagines a dart-thrower repeatedly aiming for the bulls-eye; for each throw, there is a probability that the dart-thrower hits it. Plausibly, the trials are not independent; a string of hits might give the thrower a boost of confidence or psyche them out. Nor are they identically distributed, as the thrower presumably gets better with experience. On AF, the predictable response to this scenario is that one ought to account for factors that change the applicable limit in the reference class B—e.g., by imagining a hypothetical sequence with the thrower's skill "frozen" at a given time. In a sense, this argument re-frames the "reference class problem" that Hájek (1997) poses against finite frequentism. There, he suggests that our "freezing" response amounts to radical eliminativism about unconditional probabilities, as it makes all probabilities tacitly conditional on some reference class B and "science seems to abound with statements of unconditional probability" (1997, p. 215). But I think that this argument reifies a bit of technical ephemera. The unconditional probability statements that appear in science *are* usually understood to be tacitly conditional to some reference class. When

---

[14] At the very least, Gibbs (1902) seems to think about ensembles this way. He instructs his reader to "imagine a great number of independent systems, identical in nature [macrostate], but differing in phase [microstate]" (1902, p. 5). Regarding the use of continuous distributions to describe such ensembles, he writes: "In strictness, a finite number of systems cannot be distributed continuously in phase. But by increasing indefinitely the number of systems, we may approximate to a continuous law of distribution, such as is here described. To avoid tedious circumlocution, language like the above may be allowed, although wanting in precision of expression, when the sense in which it is to be taken appears sufficiently clear" (1902, p. 5). While Gibbs's use of the term "approximation" is similar Hájek's, it seems clear that he intends his target system to be a counterfactual and finite collection of gases.

one assigns an unconditional probability directly to a projection on Hilbert space in quantum mechanics or a function on a phase space in classical mechanics, one does not mean that it is not relevant that we are, say, working in a non-relativistic regime. Scientists omit such claims from their probability statements because all situations in a given domain of study satisfy them—it is a matter of expediency, not metaphysical scripture. In a somewhat similar vein, Hájek also argues that the "freezing" move amounts to switching out frequencies for propensities, and frequencies thereby become an "idle wheel" (2009, p. 229). However, the wheel is *not* idle, even if propensities lurk around the corner.[15] However one cashes out approximate truth, AF still uses hypothetical frequencies to explain Kolmogorov's rules as approximate laws of nature.

I have saved the best for last: (15), the argument that hypothetical frequencies cannot yield infinitesimal probabilities, in a sense made precise by nonstandard analysis. On the one hand, I only intend AF to analyze scientists' probability talk as it currently stands. By and large, that talk occurs in $\mathbb{R}$-valued language. On the other hand, nothing stops this talk from changing! Moreover, as noted above, I do not think that frequentists should hold technical ephemera sacred. They need not read unconditional probabilities literally or impose countable additivity. Hájek's "parting offer" of a hyperfinite version of frequentism is an exercise that fully aligns with this attitude.[16] Indeed, it might find purchase on the nonstandard approach to quantum mechanics recently illustrated by Barrett and Goldbring (2021).[17]

4.4 A Vicious Circle?

While it is not one of his official fifteen, Hájek also tackles an infamous argument having to do with the Law of Large Numbers (LLN). According to the strong version of this law, the relative frequency of an attribute in a series of independent and identically distributed trials converges to that attribute's probability with probability one. As Hájek notes, some frequentists argue that this law supports their analysis by showing a sense in which non-converging sequences are pathological.

If one attempts to *define* probability as the number obtained in the LLN limit, then one faces a vicious circle. However, as Hájek notes, LLN can also yield a *virtuous* circle for the frequentist. To wit, LLN shows a sense in which frequentism is self-consistent. It says that an infinite sequence of frequencies converges to the correct value with the highest probability that it can—which means that an infinite

---

[15] This discussion points to how one might combine a propensity account with AF. To wit, propensities can serve as the truth-makers of the approximately true partial descriptions that AF's chance collectives provide. Such propensities might provide a causal explanation of the sort of empirical pattern that AF explicates. However, I will leave the further development of this view (and its alternatives) for future work. Many thanks to an anonymous referee for raising this point.

[16] The idea of using nonstandard analysis in the foundations of probability theory dates back at least to Nelson (1987), although he did not work in an explicit frequentist context.

[17] In particular, it might be interesting to combine Hájek's proposal with Barrett and Goldbring's nonstandard version of Everett's (1957) theorem on the convergence of relative frequencies in typical branches. This version of the theorem follows quickly from results due to Raab (2004).

sequence of infinite sequences of frequencies converges to the correct value with the highest probability that it can, and so on. We have already seen one way that this circle is virtuous. Just this sort of recursion guides our choice of ϵ, via our appeal to standard deviations from limiting values. Moreover, while probability theory cannot *prove* that convergence obtains, it does not predict any deviations from the correct value, either.

The approximation-first view gets off the boat here; a virtuous circle is good enough. Like most mathematicians did for Peano arithmetic after Gödel, we do not fret too much that our theory cannot strictly prove its own consistency. Hájek suggests that some frequentists are after something more, some further metaphysical "comfort" (1997, p. 223). That might well be true. However, as flagged above, the approximation-first view is ruthlessly pragmatic. It leaves metaphysically-hungry frequentists out in the cold. I think, though, that that might be the price we have to pay to get a perspicuous view of frequentism on the table—one that makes good sense of its origins in and commitments to empirical practice.

5 Discussion

Starting from a careful reading of von Mises, I have pared away extraneous metaphysical and mathematical concerns to arrive at what I think is the core of hypothetical frequentism: the approximation-first view, AF. Hájek's detailed tear-down of HF provides fifteen valuable foils for this view. They help pinpoint why one should use limiting frequencies as mere approximations rather than idealizations.

That's the main moral that I want to draw from this discussion. It might also be worthwhile to briefly discuss a couple of morals that I *don't* want to draw. First, I am not trying to feed the frequentist-fanatic. I know that philosophers and physicists have clashed in a particularly abrasive way over frequentism in the past, and will plausibly continue to do so in the future. The reader might be more or less acutely aware of the following caricature: the stodgy old physicist who wields frequentism as their philosopher-bashing-baton, waving it threateningly at any mention of subjective Bayesianism. No one should aspire to this caricature. I hope to have stressed that frequentism, even perspicuously interpreted, has a very narrow scope. Many areas of scientific practice explicitly invoke subjective notions of probability, and these areas either stand to benefit or are already benefiting from Bayesian techniques. But one can use Bayesian subjective probability hand-in-hand with frequentist objective probability. So it is still worth getting clear on the best way to use frequentism.

Second, I recognize that my approach to this topic is putting me in serious running for the title of World's Most Boring Cop. The thought goes: obviously, frequentism is not dead; obviously, Hájek did not kill it; obviously, the view will never die. Virtually every physicist and statistician appeals and will continue to appeal to frequentist explanations of Kolmogorov's rules (in more-or-less exactly the way recited in Sect. 3) when they teach their undergraduates, and nothing that Hájek or any other philosopher does is going to change that fact. That's not the point of the fifteen arguments. Their point is to interrogate whether limiting frequencies can do

hefty metaphysical work. And I charge in on my pragmatic horse with my rote and minimal desiderata to serve and protect the status quo, missing the deeper picture.

I do not want to dismiss the importance of this picture or the virtues it offers. As the old Bishop Butler quote goes,

> Probability is the Very Guide of Life.

And sketching a rich ontology with detailed counterfactuals is one way that a theory can provide guidance. But it is just one way. A theory can distend and distort, approximate and idealize, and still provide a helpful picture of the world. By setting the record straight on frequentism, by bringing our understanding of it closer to our understanding of similar scientific explanations, I think that we can help probability serve as just a bit better of a guide.

References

Abrams, M. (2006). Infinite populations and counterfactual frequencies in evolutionary theory. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 37*(2), 256–268.

Aronson, J. L. (1990). Verisimilitude and type hierarchies. *Philosophical Topics, 18*(2), 5–28.

Barrett, J. A. & Goldbring, I. (2021). Everettian mechanics with hyperfinitely many worlds. http://philsci-archive.pitt.edu/19419/.

Carnap, R. (1947). *Meaning and necessity: A study in semantics and modal logic*. The University of Chicago Press.

Chakravartty, A. (2007). *A metaphysics for scientific realism*. Cambridge University Press.

Chakravartty, A. (2010). Truth and representation in science: Two inspirations from art. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention: Representation in art and science* (pp. 33–50). Springer.

Church, A. (1940). On the concept of a random sequence. *Bulletin of the American Mathematical Society, 5*(2), 130–135.

de Finetti, B. (1974). *Theory of probability: A critical introductory treatment* (Vol. 1). Wiley.

Diaconis, P., Holmes, S., & Montgomery, R. (2007). Dynamical bias in the coin toss. *SIAM Review, 49*(2), 211–235.

Diaconis, P., & Skyrms, B. (2017). *Ten great ideas about chance*. Princeton University Press.

Earman, J. (2019). The role of idealizations in the Aharonov-Bohm effect. *Synthese, 196*(5), 1991–2019.

Easwaran, K. (2013). Why countable additivity? *Thought: A Journal of Philosophy, 2*(1), 53–61.

Elliot, C. (2020). E. T. Jaynes's solution to the problem of countable additivity. *Erkenntnis*.

Everett, H. (1957). "Relative state" formulation of quantum mechanics. *Reviews of Modern Physics, 29*(3), 454–462.

Fine, T. L. (1973). *Theories of probability: An examination of foundations*. Academic Press.

Fitelson, B., & Hájek, A. (2017). Declarations of independence. *Synthese, 194*(10), 3979–3995.

Fletcher, S. C., Palacios, P., Ruetsche, L., & Shech, E. (2019). Infinite idealizations in science: An introduction. *Synthese, 196*(5), 1657–1669.

Frigg, R., & Hartmann, S. (2020). Models in Science. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. Spring 2020 edition.

Gibbs, J. W. (1902). *Elementary principles in statistical mechanics: Developed with special reference to the rational foundations of thermodynamics*. Yale University Press.

Gillespie, J. H. (1998). *Population genetics: A concise guide*. The Johns Hopkins University Press.

Gupta, A. (2021). Definitions. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. Winter 2021 edition.

Hájek, A. (1997). "Mises redux''–redux: Fifteen arguments against finite frequentism. *Erkenntnis, 45*, 209–227.

Hájek, A. (2009). Fifteen arguments against hypothetical frequentism. *Erkenntnis, 70*(2), 211–235.

Hájek, A. (2012). Interpretations of probability. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. The Metaphysics Research Lab (Winter 2012 edition).

Hájek, A. (2019). Interpretations of Probability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University (Fall 2019 edition).

Halvorson, H. (2001). On the nature of continuous physical quantities in classical and quantum mechanics. *Journal of Philosophical Logic, 30*(1), 27–50.

Howson, C. (2008). De Finetti, countable additivity, consistency and coherence. *The British Journal for the Philosophy of Science, 59*(1), 1–23.

Hubert, M. (2021). Reviving frequentism. *Synthese, 199*(1–2), 5255–5284.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press.

Jeffrey, R. C. (1992). *Probability and the art of judgment*. Cambridge University Press.

Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd ed.). Chelsea Publishing Company.

La Caze, A. (2016). Frequentism. In A. Hájek & C. Hitchcock (Eds.), *The Oxford handbook of probability and philosophy*. Oxford University Press.

Martin-Löf, P. (1966). The definition of random sequences. *Information and Control, 9*, 602–619.

Martin-Löf, P. (1969). The literature on von Mises' Kollektivs revisited. *Theoria, 35*(1), 12–37.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A, 16*(3), 247–273.

Meehan, A. (2021). You say you want a revolution: Two notions of probabilistic independence. *Philosophical Studies, 178*(10), 3319–3351.

Nelson, E. (1987). *Radically elementary probability theory*. Princeton University Press.

Norton, J. D. (2012). Approximation and idealization: Why the difference matters. *Philosophy of Science, 79*(2), 207–232.

Oddie, G. (1986). *Likeness to truth*. Reidel.

Palacios, P., & Valente, G. (2021). The paradox of infinite limits: a realist response. In T. D. Lyons & P. Vickers (Eds.), *Contemporary scientific realism: The challenge from the history of science, book section 14* (pp. 284–312). Oxford University Press.

Pence, C. H. (2021). *The causal structure of natural selection*. Cambridge University Press.

Popper, K. R. (1972). *Conjectures and refutations: The growth of scientific knowledge* (4th ed.). Routledge & Kegan Paul.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Potochnik, A. (2020). Idealization and many aims. *Philosophy of Science, 87*(5), 933–943.

Quine, W. V. O. (2013). *Word and object*. The MIT Press.

Raab, A. (2004). An approach to nonstandard quantum mechanics. *Journal of Mathematical Physics, 45*(12), 4791–4809.

Rowbottom, D. P. (2015). *Probability*. Cambridge University Press.

Rueger, A., & Sharp, D. (1998). Idealization and stability: a perspective from nonlinear dynamics. *Poznán Studies in the Philosophy of the Sciences and the Humanities, 63*, 201–216.

Seidenfeld, T., & Schervish, M. J. (1983). A conflict between finite additivity and avoiding Dutch book. *Philosophy of Science, 50*(3), 398–412.

Strevens, M. (2019). The structure of asymptotic idealization. *Synthese, 196*(5), 1713–1731.

Tornier, E. (1936). Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie. Leipzig

Valente, G. (2019). On the paradox of reversible processes in thermodynamics. *Synthese, 196*(5), 1761–1781.

van Fraassen, B. C. (1977). Relative frequencies. *Synthese, 34*, 133–166.

van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon Press.

Ville, J. (1939). *Étude critique de la notion de collectif. Monographies des Probabilités*. Gauthier-Villars.

von Mises, R. (1981). *Probability, statistics, and truth* (2nd ed.). Dover.

Wald, A. (1938). Die widerspruchsfreiheit des kollektivbegriffes. *Actualités Scientifiques et Industrielles, 735*, 79–99.

Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy, 104*(12), 639–659.

Williamson, J. (1999). Countable additivity and subjective probability. *The British Journal for the Philosophy of Science, 50*(3), 401–416.

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