Frequentist Probability

Aris Spanos*
Virginia Tech, Blacksburg, VA, USA

Abstract

The primary objective of this article is to discuss a model-based frequentist interpretation that identifies the probability of an event A with the *limit* of its relative frequency of occurrence. What differentiates the proposed interpretation from the traditional ones are several key features: (i) events and probabilities are defined in the context of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$, (ii) it is anchored on the strong law of large numbers, (iii) it is justified on empirical grounds by validating the model assumptions vis-á-vis data $\mathbf{x}_0$, (iv) the 'long-run' metaphor can be rendered operational by simple simulation based on $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$, and (v) the link between probability and real-world phenomena is provided by viewing data $\mathbf{x}_0$ as a 'truly typical' realization of the stochastic mechanism defined by $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$. This link constitutes a feature shared with the Kolmogorov complexity algorithmic perspective on probability, which provides a further justification for the proposed frequentist interpretation.

Key words: frequentist interpretation of probability; statistical model; model-based induction; strong law of large numbers; long-run metaphor; Kolmogorov complexity

*This is an earlier version of the paper: Spanos, A. (2017), "Frequentist Probability", Wiley StatsRef: Statistics Reference Online.stat07913/.

1 Introduction

The term *frequentist probability* refers to the frequentist interpretation of probability as opposed to other interpretations such as the classical, the degrees of belief (subjective, logical) and the propensity interpretations; see Refs 1, 2. The different interpretations aim to answer the question: what does probability correspond to in the real world? This correspondence is crucial because it determines the kind of inductive procedures one would follow with a view to "learn from data about phenomena of interest," that is, it determines the nature of the inductive (statistical) inference called for. What distinguishes the frequentist interpretation from the others is the fact that probability is viewed as directly related to observed relative frequencies in real data; an objective feature of the real world. It refers to the limit of the relative frequency of the occurrence of an event by repeating the experiment that brings the particular event about. In this sense frequentist probability has always been associated with the statistical analysis of data.

The frequentist interpretation can be traced back to the *statistical regularities* noted first by Graunt [3] in mortality data and subsequently reasserted in demographic, anthropomorphic, economic and social (crimes, violent deaths, etc.) data by Quetelet [4]; see Ref. 5. Subsequently, several attempts to formalize the frequentist interpretation of probability by Ellis [6], Venn [7] and Von Mises [8] did not fully succeed primarily because the statistical analysis at the time was still in a state of flux.

The formalization of the frequentist interpretation of probability as it relates to the associated statistical approach began in the early 1930s. The current model-based frequentist approach to statistical modeling and inference was pioneered by Fisher [9] and extended by Neyman and Pearson [10]. The mathematical foundation in the form of an axiomatic approach to probability theory was framed by Kolmogorov [11] (*see* cref{stat04803}). The coalescing of these two crucial developments was first enunciated by Cramer [12] (*see* cref{stat01354}). These developments provided a coherent grounding for the frequentist interpretation of probability anchored on stable "long-run" frequencies interpretation [12] that is grounded on the strong law of large numbers (SLLNs) (*see* cref{stat05877}). The measure-theoretic foundation (*see* cref{stat02290}) was instrumental since, as argued by Adams and Guillemin [13] in a book entitled "Measure Theory and Probability":

> "What we hope to convey here is that had the Lebesgue theory of measure not existed, one would be forced to invent it to contend with the paradoxes of large numbers." (p. x)

The key to Fisher's approach was the notion of a prespecified parametric *statistical model* that provides the proper context for assigning probabilities to the relevant events associated with data. The statistical model can be viewed as a stochastic generating mechanism (GM) with prespecified premises that give rise to deductively derived inference propositions. In contrast to deduction, the validity of the inductive inferences requires the soundness of the prespecified premises vis-á-vis the data: the probabilistic assumptions imposed on the data. This is often insufficiently appreciated by the current discussions of the frequentist interpretation of probability.

The primary objective of this article is to articulate a *model-based* frequentist interpretation of probability with a view to contrast it with alternative frequentist interpretations, including von Mises [8] (*see* cref{stat01343}). As argued in Ref. 14, the *model-based* frequentist interpretation of probability is not vulnerable to many widely discussed criticisms, such as: (i) the circularity of its definition, (ii) its reliance on "random samples", and (iii) its inability to assign "single event" probabilities. These are legitimate criticisms of enumerative induction anchored on von Mises's "collective" [15], but they are misplaced when leveled against the model-based frequencies interpretation. This stems from the fact that when enumerative induction and von Mises's collective are viewed from the frequentist model-based perspective, it becomes clear that their implicit notion of a statistical model is highly restrictive, and the SLLN does not hold for von Mises's collective for purely mathematical reasons;

> "Trying to be "precise" by making a *definition* out of the "long-term frequency" idea lands us in real trouble. Measure theory gets us out of the difficulty in a very subtle way discussed in Chapter 4." ([16], p. 25)

Section 2 discusses the mathematical probability associated with [11], as a prelude to the Fisher-Neyman-Pearson model-based frequentist inference as well as the SLLN in Section 3. Section 4 explains the model-based frequentist interpretation of probability and relates that interpretation of Kolmogorov complexity. Section 5 relates the frequentist and propensity interpretations and addresses Humphrey's paradox.

2 Mathematical Probability

2.1 Kolmogorov's Axiomatic Approach

At the dawn of the 20th century, the axiomatic treatment of probability was part of the famous 6th problem, known as *the mathematical treatment of the axioms of physics*, presented by David Hilbert in 1900 at the Paris conference of the International Congress of Mathematicians at the Sorbonne (*see* cref{stat02979}).

The first major breakthrough that led to advanced integration known as *measure theory* was Lebesgue's thesis in 1902. He extended the definition of volume in $\mathbb{R}^n$ to Borel sets to define a measure as a positive countably additive set function. A Borel set is any set in a topological space that is constructed using open sets and the operations of countable union, countable intersection, and relative complement. The next important step was provided by Caratheodory in 1913 who proved that any measure defined on a given ring Q of subsets of a given set Ω can be extended to the σ-field (σ-algebra) generated by Q. In 1915 Frechet pointed that the notion of the σ-field of subsets of an abstract space provides the key for ensuring that the usual definitions and operations of measure theory are well defined. The final piece of the puzzle was given by the Radon-Nikodym theorem in 1930 (*see* cref{stat05944}), which provided necessary and sufficient conditions for a countably additive function defined on sets to be expressed as an integral; see Ref. 17.

The complete picture was framed by Kolmogorov [11]. The foundation of the axiomatic approach to probability is specified by a probability space $(\Omega, \mathcal{F}, \ \mathbb{P}(.))$, where:

1. Ω denotes the set of all possible distinct outcomes.
2. $\mathcal{F}$ denotes a set of subsets of Ω, called *events* of interest, endowed with the mathematical structure of a σ-field, that is, it satisfies the following conditions:

(i) $\Omega{\in}\mathcal{F}$, (ii) if $A{\in}\mathcal{F}$, then $\overline{A}{\in}\mathcal{F}$, (iii) if $A_i{\in}\mathcal{F}$ for $i{=}1, 2, \ldots, n, \ldots$. then $\bigcup_{i=1}^{\infty} A_i{\in}\mathcal{F}$.

3. $\mathbb{P}(.)$: $\mathcal{F} \to [0, 1]$ denotes a set function that satisfies the axioms (*see* cref{stat02819}):

(A1) $\mathbb{P}(\Omega){=}1$, for any outcomes set Ω
(A2) $\mathbb{P}(A) \geq 0$, for any event $A{\in}\mathcal{F}$
(A3) *Countable additivity.* For $A_i{\in}\mathcal{F}, \ i{=}1, 2, \ldots$ such that $A_i \cap A_j{=}\emptyset$ for all $i{\neq}j, \ i, j{=}1, 2, \ldots,$ then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

This formalization places probability squarely into the mathematical field of *measure theory* concerned more broadly with assigning size, length, content, area, volume, so on to sets; see Ref. 18.

The probability space $(\Omega, \mathcal{F}, \ \mathbb{P}(.))$ provides an idealized description of the stochastic mechanism that gives rise to the events of interest and related events $\mathcal{F}$ [which is closed under the set theoretic operations of union ($\cup$), intersection ($\cap$) and complementation ($^{-}$)], with $\mathbb{P}(.)$ assigning probabilities to events in $\mathcal{F}$.

2.2 Random Variables and Statistical Models

An important extension of the initial Kolmogorov formalism based on $(\Omega, \ \mathcal{F}, \ \mathbb{P}(.))$ is the notion of a *random variable* (*see* cref{stat04404}): a real-valued function:

$$X(.)\text{: } \Omega \to \mathbb{R}, \text{ such that } \{X \leq x\}{\in}\mathcal{F} \text{ for all } x{\in}\mathbb{R}$$

That is, $X(.)$ assigns numbers to the elementary events in Ω in such a way so as to preserve the original event structure of interest ($\mathcal{F}$). This extension is important for bridging the gap between the mathematical model $(\Omega, \mathcal{F}, \ \mathbb{P}(.))$ and the observable stochastic phenomena of interest, because observed data usually come in the form of *numbers* on the real line.

The most crucial role of the random variable $X(.)$ is to transform the original abstract probability space $(\Omega, \mathcal{F}, \ \mathbb{P}(.))$ into a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ defined on the real line:

$$(\Omega, \mathcal{F}, \ \mathbb{P}(.)) \ \stackrel{X(.)}{\longrightarrow} \ \mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x}){=}\{f(\mathbf{x}; \boldsymbol{\theta}), \ \boldsymbol{\theta}{\in}\Theta\}, \ \mathbf{x}{\in}\mathbb{R}_X^n \tag{1}$$

Hence, the notion of probability associated with $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ in equation (1) is purely measure-theoretic and follows directly from the axioms A1-A3; see Ref. 19.

The relevant random variable underlying the traditional frequentist interpretation is defined by $\{X{=}1\}{=}A$, $\{X{=}0\}{=}\overline{A}$, where $A{\in}\mathcal{F}$, $\mathbb{P}(A){=}p$ and $\mathbb{P}(\overline{A}){=}1{-}p$, which is a Bernoulli (Ber) distributed random variable. The limiting process associated with the relative frequency interpretation requires "repeating the experiment under identical conditions", which is framed in the form of an indexed sequence of random variables (*see* cref{stat03025}) $\{X_k,\ k\in\mathbb{N}:=(1,2,\ldots n,\ldots)\}$ assumed to be independent and identically distributed (IID), that is, the statistical model is the simple Bernoulli model:

$$\mathcal{M}_\theta(\mathbf{x}): X_k \backsim \mathsf{BerIID}(\theta,\theta(1-\theta)),\ 0\leq\theta\leq 1,\ k\in\mathbb{N} \quad (2)$$

Another widely used statistical model is the simple Normal (N):

$$\mathcal{M}_\theta(\mathbf{x}): X_k \backsim \mathsf{NIID}(\mu,\sigma^2),\ -\infty<\mu<\infty,\ \sigma^2>0,\ k\in\mathbb{N} \quad (3)$$

In general, the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ is viewed as a parameterization of the stochastic process $\{X_k,k\in\mathbb{N}\}$ whose probabilistic structure is chosen so as to render data $\mathbf{x}_0:=(x_1,\ldots,x_n)$ a *typical realization* thereof; see Ref. 20.

3 Model-based Frequentist Interpretation

Can the above Kolmogorov formalism be given an interpretation by assigning a meaning to the primitive term *probability*? The general thesis concerning the relationship between mathematics and empirical modeling adopted in this article has been first articulated by Cramer [12]:

> "The mathematical theory belongs entirely to the conceptual sphere, and deals with purely abstract objects. The theory is, however, designed to form a model of a certain group of phenomena in the physical world, and the abstract objects and propositions of the theory have their counterparts in certain observable things, and relations between things. If the model is to be practically useful, there must be some kind of general agreement between the theoretical propositions and their empirical counterparts." (p. 332)

The proposed frequentist interpretation relates to a specific *objective*: modeling observable phenomena of interest exhibiting chance regularity patterns and referred to as *stochastic*. In relation to this it is important to keep in mind Kolmogorov's [21] distinction between randomness proper and stochastic randomness:

> "... we should have distinguished between randomness proper (as absence of any regularity) and stochastic randomness (which is the subject of probability theory). There emerges the problem of finding reasons for the applicability of the mathematical theory of probability to the phenomena of the real world." (p. 1)

Neyman [22] described the statistical modeling process in three stages:

"There are three distinct steps in this [statistical modeling] process:

(i) Empirical establishment of apparently stable long-run relative frequencies (or frequencies for short) of events judged interesting, as they develop in nature.

(ii) Guessing and then verifying the "chance mechanism", the repeated operations of which produces the observed frequencies. This is a problem of "frequentist probability theory". Occasionally, this step is labeled "model building". Naturally, the guessed chance mechanism is hypothetical.

(iii) Using the hypothetical chance mechanism of the phenomenon studied to deduce rules of adjusting our actions (or decisions) to the observations so as to ensure the highest "measure of success". (Ref. 22, 1977, p. 99)

In (i) Neyman demarcates the domain of statistical modeling to *stochastic phenomena* that exhibit *chance regularities*, in the form of the long-run stability of relative frequencies. In (ii) he provides a clear statement concerning the nature of *specification* and model validation (verification), and in (iii) he brings out the role of ascertainable error probabilities in assessing the optimality of inference procedures.

3.1 The Strong Law of Large Numbers

3.1.1 Frequentist interpretation

The probability of an event A, say, $\mathbb{P}(A)=p$, is directly related to the relative frequency of the occurrence of event A, as defined in the context of $(\Omega, \mathcal{F}, \ \mathbb{P}(.))$.

3.1.2 The long-run metaphor

The frequentist interpretation is articulated in terms of the "long-run" metaphor which states that in a sequence of trials the relative frequency s_n of occurrence of event A will approximate (oscillate around) the value of its true probability $p=\mathbb{P}(A)$. This metaphor, however, carries the seeds of a potential confusion between the stochastic process $\{X_k, \ k\in\mathbb{N}\}$ itself and one its finite realizations $\{x_k\}_{k=1}^n$.

How is such an interpretation formally justified? The simple answer is that it is justified by invoking the SLLN.

Under certain restrictions on the probabilistic structure of the process $\{X_k, \ k\in\mathbb{N}\}$, the most restrictive being that it is IID, one can prove *mathematically* that:

$$\mathbb{P}(\lim_{n\to\infty}(\tfrac{1}{n}\textstyle\sum_{k=1}^n X_k)=p)=1 \tag{4}$$

known as *convergence almost surely* (a.s). It is important to emphasize that the notion of convergence in (4) is probabilistic in the sense that a sequence of random variables $S_n=(\frac{1}{n}\sum_{k=1}^n X_k)$ converges to a constant p *with probability one*; see Ref. 19, p. 505-7.

This result is an instance of the well-known SLLN first proved by Borel [23]. The SLLN asserts that for any IID process $\{X_k, \ k\in\mathbb{N}\}$ and any event $A\in\mathcal{F}$, the relative frequency of occurrence of A converges to $\mathbb{P}(A)$ with probability one. Since the early twentieth century the original IID assumptions have been weakened considerably. For instance, when $\{X_k, \ k\in\mathbb{N}\}$ is a *martingale difference* process (Ref. 24) the result in Equation (4) holds. An analogous (but weaker) result, in the form of *convergence in probability*:

$$\lim_{n\to\infty}\mathbb{P}\left(\left|(\tfrac{1}{n}\textstyle\sum_{k=1}^n X_k)-p\right|<\epsilon\right)=1, \quad \text{for any } \epsilon>0 \tag{5}$$

is associated with the weak law of large numbers (WLLN); see Ref. 19.

In an attempt to delineate what this result asserts more clearly, let us emphasize what it does *not* mean.

In the first place, the result in Equation (4) does *not* involve any claims that the sequence of numbers $\{s_n\}_{n=1}^\infty$, where $s_n=\frac{1}{n}\sum_{k=1}^n x_k$, converges to p in a purely mathematical sense. This conflates the probabilistic convergence in Equation (4) with a mathematical convergence of the sequence $\{s_n\}_{n=1}^{\infty}$. Indeed, this result holds for the process $\{X_k,\ k\in\mathbb{N}\}$ and not the particular realization $\{x_k\}_{k=1}^{n}$. Second, the result in Equation (4) refers only to what happens at the limit $n=\infty$ and says nothing about the behavior of $\frac{1}{n}\sum_{k=1}^{n} X_k$ for a given n. Indeed, the result in Equation (4) provides no information pertaining to the accuracy of $(\frac{1}{n}\sum_{k=1}^{n} x_k)$ as an approximation of $\mathbb{P}(A)$ for a given n. For that one needs a different limit theorem known as the LIL (*see* cref{stat02919}):

$$\mathbb{P}\left(\lim_{n\to\infty}\sup\left[\frac{(S_n-p)/\sqrt{p(1-p)}}{\sqrt{2n\ln(\ln(n))}}\right]=1\right)=1 \tag{6}$$

which quantifies the *rate* of convergence of the process $\{S_n\}_{n=1}^{\infty}$. Most importantly, the SLLN holds for the process $\{X_k,\ k\in\mathbb{N}\}$, provided the latter satisfies certain probabilistic assumptions. Assuming IID is the most restrictive set of assumptions which can be weakened considerably; see Ref. 18. In practice, one can appeal to the SLLN and its various extensions after the invoked assumptions are validated vis-á-vis the data using misspecification testing; see Ref. 25. This validation assesses whether data $\mathbf{x}_0$ constitute a "truly typical" realization of the process $\{X_k,\ k\in\mathbb{N}\}$; see Ref. 19.

4 A Frequentist Interpretation of Probability

The *frequentist interpretation* articulated in this paper identifies the probability of an event A with the (probabilistic) *limit* of the relative frequency of its occurrence, $s_n=\frac{1}{n}\sum_{k=1}^{n} x_k$, in the context of a well defined stochastic mechanism represented by the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$. The proposed frequentist interpretation has five key features:

(i) it revolves around the notion of a statistical model $\mathcal{M}_{\theta}(\mathbf{x})$, broadly viewed to accommodate non-random samples, (ii) it is firmly anchored on the SLLN, (iii) it is justified on empirical, not a priori, grounds, (iv) the "long-run" metaphor can be rendered operational, and (e) the link between the measure theoretic results and real-world phenomena is provided by viewing data $\mathbf{x}_0$ as a "truly typical" realization of the stochastic process $\{X_t,\ t\in\mathbb{N}\}$ underlying $\mathcal{M}_{\theta}(\mathbf{x})$. That is, the validity of the model assumptions secures the meaningfulness of identifying the limit of the relative frequencies $\{s_n\}_{n=1}^{\infty}$ with the probability p by invoking Equation (4). Given that the probabilistic assumptions Bernoulli, IID are testable vis-à-vis data $\mathbf{x}_0$, the frequentist interpretation is justifiable on *empirical*, not on *a priori*, grounds.

4.1 The Pivotal Role of a Statistical Model

The cornerstone of the proposed frequentist interpretation of probability is provided by the notion of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x}),\ \mathbf{x}\in\mathbb{R}_X^n$, which plays a pivotal role in both modeling and inference because, among other things, it specifies the inductive premises of inference, delimits *legitimate* events and data $\mathbf{x}_0$, and assigns probabilities to all legitimate events via $f(\mathbf{x};\boldsymbol{\theta}),\ \mathbf{x}{\in}\mathbb{R}^n_X$; see Ref. 20. Formally, an event A is legitimate if it is defined by a well-behaved (Borel) function of $\mathbf{X}$ (Ref. 18). In addition, $\mathcal{M}_\theta(\mathbf{x})$ also ascertains the optimality of inference procedures in terms of the relevant error probabilities. This is because for any statistic (*see* cref{stat00762}) (estimator, test statistic, and predictor), say $T_n{=}g(X_1,\ldots,X_n)$, its sampling distribution (*see* cref{stat00735}) is derived from $f(\mathbf{x};\boldsymbol{\theta})$ via:

$$F(t;\boldsymbol{\theta}){:=}\mathbb{P}(T_n\leq t;\boldsymbol{\theta}){=}\underbrace{\int\int\cdots\int}_{\{\mathbf{x}:\ g(x_1,\ldots,x_n)\leq t;\ \mathbf{x}\in\mathbb{R}^n_X\}} f(\mathbf{x};\boldsymbol{\theta})dx_1dx_2\cdots dx_n \tag{7}$$

This indicates that in frequentist inference $f(\mathbf{x};\boldsymbol{\theta})$ provides the sole source of relevant probabilities, including error probabilities used to evaluate the reliability of inferential procedures; see Ref. 26.

4.2 Rendering the "Long-run" Metaphor Operational

The long-run metaphor has long been considered the Achilles heel of frequentist induction. Its invocation has encouraged the view that error probabilities are only useful pre-data, and only in the context of Neyman's behavioristic interpretation of tests; see Ref. 27. This issue lies at the heart of the tension between Fisher's significance testing (*see* cref{stat05865}) and the Neyman-Pearson (N-P) hypothesis testing (*see* cref{stat06344}). Indeed, the appositeness of the long-run metaphor has been questioned by Fisher [28].

The error-statistical post-data evaluation of inference, based on the notion of severity, has shed different light on this problem and provided a way to blend the pre data and post-data perspectives harmoniously, by providing a proper interpretation for Fisher's post data p-value; see Ref. 29. This is based on the argument that the key feature of the long-run metaphor in the context of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ is the *repeatability* (in principle) of the idealized data-generating process, and not the *temporal* dimension as often assumed. In light of this, the issue can be laid to rest by using the statistical GM associated with a particular $\mathcal{M}_\theta(\mathbf{x})$ to render the long-run metaphor operational; see Table 1.

Table 1: Simple Normal model

Statistical GM:		$X_k=\mu+u_k,\quad k{\in}\mathbb{N}$	
[1]	Normal:	$X_k\backsim\mathsf{N}(.,.),\ x_k{\in}\mathbb{R}$	$k{\in}\mathbb{N}$.
[2]	Constant mean:	$E(X_k){=}\mu$	$k{\in}\mathbb{N}$.
[3]	Constant variance:	$Var(X_k){=}\sigma^2$	$k{\in}\mathbb{N}$.
[4]	Independence:	$\{X_k,\ k{\in}\mathbb{N}\}$ is an independent process	

Example. In the case of the simple Normal model (Table 1), the statistical GM:

$$X_k=\mu+\sigma\varepsilon_k,\ \varepsilon_k\backsim\mathsf{N}(0,1),\quad k{=}1,2,\ldots,n, \tag{8}$$

where $\varepsilon_k \backsim \mathsf{N}(0,1)$ denotes *pseudo-random* numbers from the standard Normal distribution, can be used to simulate the long-run metaphor by generating artificial realizations using pseudo-random numbers from $\mathsf{N}(0,1)$ in conjunction with (8). Using the simulated data one can construct the empirical counterpart to any probabilistic assignment $\mathbb{P}(A)$, for any legitimate event A – any well-behaved functions of $\mathbf{X}$ – including the sampling distribution of any statistic of interest like $\widehat{\boldsymbol{\theta}}_n{:=}(\overline{X}_n, s^2)$ and $\tau(\mathbf{X}){=}\sqrt{n}(\overline{X}_n{-}\mu_0)/s$, as well as the associated error probabilities of interest, both pre-data and post-data; see Ref. 28.

4.3 Model-based Frequentist inference

In the context of the Fisher-Neyman-Pearson model-based inference, where the data $\mathbf{x}_0{:=}(x_1, x_2, \ldots, x_n)$ are viewed as a typical realization of a stochastic mechanism described by a statistical model, denoted by:

$$\mathcal{M}_\theta(\mathbf{x}){=}\{f(\mathbf{x};\boldsymbol{\theta}),\ \boldsymbol{\theta}{\in}\Theta\},\ \mathbf{x}{\in}\mathbb{R}^n_X$$

the frequentist interpretation of probability is inextricably bound up with what [22], p. 99, called *stable long-run relative frequencies* of events of interest. $\mathcal{M}_\theta(\mathbf{x})$ is assumed to represent an idealized GM that could have given rise to data $\mathbf{x}_0$. The existence of the probabilities described by $f(\mathbf{x};\boldsymbol{\theta})$, $\mathbf{x}{\in}\mathbb{R}^n_X$, depends crucially on being able to demonstrate that one can estimate consistently the *invariant features* of the phenomenon being modeled, as reflected in the constant but unknown parameters $\boldsymbol{\theta}$. That is, if one can prove that there exists a (strongly) consistent estimator (*see* cref{stat05836}) $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, that is,

$$\mathbb{P}(\lim_{n\to\infty}(\widehat{\boldsymbol{\theta}}_n){=}\boldsymbol{\theta}_0){=}1 \tag{9}$$

where $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$, under the assumptions of the model $\mathcal{M}_\theta(\mathbf{x})$, then the frequentist definition of probability is established on empirical grounds. This, by itself, does not ensure the reliability of any inference based on:

$$\mathcal{M}_{\widehat{\boldsymbol{\theta}}_n}(\mathbf{x}){=}\{f(\mathbf{x};\widehat{\boldsymbol{\theta}}_n)\},\ \mathbf{x}{\in}\mathbb{R}^n_X$$

for a give n, but the same assumptions that give rise to Equation (9) can be used to evaluate the relevant error probabilities associated with different forms of inference concerning $\boldsymbol{\theta}_0$. These error probabilities can be used to evaluate the reliability of inferences based on $\mathcal{M}_{\widehat{\boldsymbol{\theta}}_n}(\mathbf{x})$.

What are the distinguishing features of the frequentist approach to statistical inference?

1. The interpretation of probability is the *frequentist*: the relative frequencies associated with the long-run metaphor (in a hypothetical set up) reflect the corresponding probabilities; the formal link comes in the form of the SLLN.

2. Data $\mathbf{x}_0$ are viewed as a *particular realization* of a generic process $\{X_t,\ t{\in}\mathbb{N}{:=}(1,2,\ldots)\}$ that renders $\mathbf{x}_0$ a *truly typical realization* thereof. The "typicality" stipulation is testable vis-á-vis data $\mathbf{x}_0$ using misspecification testing; see Ref. 19, Chapter 15. The statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ constitutes a parameterization of the process $\{X_t,\ t{\in}\mathbb{N}\}$. In

this sense, the chance regularities exhibited by data $\mathbf{x}_0$ constitute the *only relevant statistical information* for selecting the probabilistic structure of $\{X_t,\ t\in\mathbb{N}\}$.

3. *Substantive information* plays a crucial role in influencing the selection of the particular parameterization of the process $\{X_t,\ t\in\mathbb{N}\}$ to specify $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$. The particular parameterization is chosen to enable one to pose the primary questions of interest. Hence, substantive information comes in the form of restrictions on statistical parameters $\boldsymbol{\theta}$, which needs to be tested before imposed.

4. The primary aim of the frequentist approach is to *learn from data* about the "true" statistical data GM:

$$\mathcal{M}_{\boldsymbol{\theta}^*}(\mathbf{x})=\{f(\mathbf{x};\boldsymbol{\theta}^*)\},\ \mathbf{x}\in\mathbb{R}_X^n$$

The expression "$\boldsymbol{\theta}^*$ denotes the true value of $\boldsymbol{\theta}$" is a shorthand for saying that "data $\mathbf{x}_0$ constitute a typical realization of the sample $\mathbf{X}$ with distribution $f(\mathbf{x};\boldsymbol{\theta}^*)$".

In light of the above model-based frequentist interpretation of probability, the *reasoning* underlying frequentist inference procedures comes in two forms:

1. *Factual*: The true state of nature $\boldsymbol{\theta}=\boldsymbol{\theta}^*$, whatever $\boldsymbol{\theta}^*$ happens to be. This is the type of reasoning underlying estimation and prediction.

2. *Hypothetical*: Various hypothetical values of $\boldsymbol{\theta}=\boldsymbol{\theta}_i,\ i=0,1,2,\ldots,$ judiciously chosen from the parameter space Θ, are indirectly compared to $\boldsymbol{\theta}^*$ via the actual data $\mathbf{x}_0$ using test statistics; this is the reasoning underlying statistical testing.

In concluding this section, it is important to emphasize that, by themselves, mathematical results, such as the SLLN in Equation (4) and the LIL in Equation (6), do not suffice to provide an apposite frequentist interpretation that addresses the foundational problems pertaining to frequentist *inductive reasoning*. Statistical induction requires a pertinent link between the mathematical framework and the actual data GM. In error statistics this link takes the form of the *interpretive provisions*:

1. Data $\mathbf{x}_0:=(x_1,x_2,\ldots,x_n)$ is viewed as a "typical" realization of the process $\{X_k,\ k\in\mathbb{N}\}$ specified by the statistical model $\mathcal{M}_\theta(\mathbf{x})$ and

2. The "typicality" of $\mathbf{x}_0$ (e.g. IID) can be assessed using trenchant mispecification testing.

Interestingly enough, the above model-based frequentist interpretation shares these provisions with an algorithmic perspective on probability based on *Kolmogorov complexity*; see Ref. 30.

4.4 Kolmogorov complexity: an algorithmic perspective

Kolmogorov complexity provides a purely *non-probabilistic* rendering to the frequentist interpretation that operationalizes all the relevant measure-theoretic results; see Refs 31, 32, p. vi, 33.

The algorithmic complexity perspective provides a *non-probabilistic* interpretation to infinite realizations of IID processes $\{x_k\}_{k=1}^{\infty}$ by focusing on the *effective computability* and *incompressibility* of its finite initial segment $\mathbf{x}_0:=\{x_k\}_{k=1}^{n}$. A particular finite sequence $\{x_k\}_{k=1}^{n}$ is "algorithmically incompressible" iff the shortest program that will output $\mathbf{x}_0$ and halt is about as long as $\mathbf{x}_0$ itself. Incompressible sequences (strings) turn out to be indistinguishable from typical realizations of IID Bernoulli processes, and vice versa. In this sense, incompressible sequences provide a model of the most basic probabilistic process without any reference to probability theory. In addition, the complexity framework can be used to characterize:

> "random infinite sequences as sequences all of whose initial finite segments pass all effective randomness tests" (Ref. 30, p. 56).

Indeed, these tests rely on non-probabilistic (algorithmic) notions of partial recursive functions and incompressibility. The key to the duality between the stochastic and algorithmic perspectives is provided by Li and Vitanyi, [28, p. 146]:

> "Martin-Löf's [34] important insight that to justify any proposed definition of randomness one has to show that the sequences that are random in the stated sense satisfy the several properties of stochasticity we know from the theory of probability."

That is, the notion of Kolmogorov complexity provides the first successful attempt to operationalize stochastic randomness, by ensuring the compliance of algorithmically incompressible sequences to the above measure theoretic results, including the SLLN Equation (4) and the LIL Equation (6).

In summary, the model-based frequentist and the algorithmic perspective based on Kolmogorov complexity, despite being grounded on entirely different mathematical formulations, share several features and give rise to complementary interpretations of probability that do *not* contravene the measure theoretic results.

5 The Propensity Interpretation of Probability

The propensity interpretation, associated with Peirce and Popper [35, 36], views probability as a propensity (disposition, or tendency) of a real world stochastic mechanism to yield a certain outcome, or a long run relative frequency of such an outcome. The propensity interpretation is invoked to explain why such stochastic mechanisms will generate a given outcome type at a persistent rate.

The propensity interpretation of probability has a clear affinity with the frequentist interpretation in so far as (i) it assumes the presence of a stochastic mechanism and (ii) views probability as a feature of the real world. This affinity has generated confusion in the philosophy of science literature; see Ref. 37. In particular, [38] pointed out that the propensity interpretation associated with real world stochastic mechanisms carries with it a built-in causal connection between different events, say A and B, which renders reversing conditional probabilities (*see* cref{stat02846}) such as $\mathbb{P}(A|B)$ to $\mathbb{P}(B|A)$ meaningless when A is the effect and B is the cause. This is usually interpreted as suggesting that the propensity interpretation of probability does not satisfy the basic rules of mathematical probability.

The Humphreys paradox can be easily explained away when one distinguishes between a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$, and a substantive model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$, where the two are related via certain parameter restrictions $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})=\mathbf{0}$; see Ref. 39. $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ is a purely probabilistic construal that comprises the probabilistic assumptions imposed on the data $\mathbf{x}_0$ and represents a particular parameterization of the stochastic process

$\{X_k,\ k\in\mathbb{N}\}$ underlying $\mathbf{x}_0$. In the context of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ probabilies are generic and consistent with the Kolmogorov axioms. In contrast, $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$ is based on substantive subject matter information, including causal assumptions, and aims to approximate the real-world GM as closely as possible. In the context of $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$ probabilities could and often have causal interpretation assigned to them. However, in empirical modeling one needs to separate the two models, ab initio, with a view to allow the substantive information in $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$ (including causality assumptions) to be tested against the data before being imposed. In this sense, there is no conflict between the frequentist and propensity interpretations of probability, as the former is germane to the statistical $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$, and the latter to the substantive model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$.

6 Summary and Conclusions

The model-based frequentist interpretation perspective identifies the probability of an event A with the *limit* of its relative frequency of occurrence. What differentiates the proposed interpretation from the traditional one are several key features: (i) events and probabilities are defined in the context of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$, broadly viewed to accommodate non-IID samples [40], (ii) it is firmly anchored on the SLLN, (iii) it is justified on empirical grounds by validating the model assumptions vis-á-vis data $\mathbf{x}_0$, (iv) the "long-run" metaphor can be rendered operational by simple simulation based on $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$, and (e) the link between the measure-theoretic results and real-world phenomena is provided by viewing data $\mathbf{x}_0$ as a "truly typical" realization of the stochastic process $\{X_t,\ t\in\mathbb{N}\}$ underlying $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$. This interpretative link constitutes a feature that the proposed interpretation shares with an algorithmic perspective based on the notion of *Kolmogorov complexity.* The latter provides a purely *non probabilistic* (algorithmic) rendering that interprets without transgressing the measure theoretic results.

References

1. Childers, T. (2013) *Philosophy and Probability*, Oxford University Press, Oxford.

2. Hájek, A. (2012) Interpretations of probability, in The Stanford Encyclopedia of Philosophy, (ed. E.N. Zalta,
URL = http://plato.stanford.edu/archives/win2012/entries/probability-interpret/ (accessed 22 August 2016).

3. Graunt, J. (1662) *Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality*, Martin, London.

4. Quetelet, L.A. (1849) *Letters on the Theory of Probabilities*, O.G. Downes, (Trans.), Layton, London.

5. Hacking, I. (1965) *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*, Cambridge University Press, Cambridge.

6. Ellis, R.L. (1849) On the foundations of the theory of probability. *Trans. Camb. Philos. Soc.* VIII, 1-6.

7. Venn, J. (1866) *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to its Application to Moral and Social Science*, Macmillan, London.

8. von Mises, R. (1928/1957) *Probability, Statistics and Truth*, 2nd edn, Dover Publication, New York.

9. Fisher, R.A. (1922) On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. London, Ser. A* **222**, 309–368.

10. Neyman, J. and Pearson, E.S. (1933) On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. London, Ser. A* **231**, 289–337.

11. Kolmogorov, A.N. (1933) *Foundations of the Theory of Probability*, 2nd English edn, Chelsea Publishing Company, New York.

12. Cramer, H. (1946) *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ.

13. Adams, M. and Guillemin, V. (1996) *Measure Theory and Probability*, Birkhauser, Boston, MA.

14. Spanos, A. (2013) A frequentist interpretation of probability for model-based inductive inference. *Synthese* **190**, 1555–1585.

15. Salmon, W.C. (1967) *The Foundations of Scientific Inference*, University of Pittsburgh Press, Pittsburgh, PA.

16. Williams, D. (2001) *Weighing the Odds: A Course in Probability and Statistics*, Cambridge University Press, Cambridge.

17. Doob, J.L. (1996) The development of rigor in mathematical probability (1900–1950). *Am. Math. Mon.* **103**, 586-595.

18. Billingsley, P. (2003) *Probability and Measure*, 4th edn, John Wiley & Sons, Inc., New York.

19. Spanos, A. (1999) *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press, Cambridge.

20. Spanos, A. (2006) Where do statistical models come from? Revisiting the problem of specification, in *Optimality: The Second Erich L. Lehmann Symposium, Lecture Notes-Monograph Series*, vol. 49 (ed. J. Rojo), Institute of Mathematical Statistics, Beachwood, OH, pp. 98–119.

21. Kolmogorov, A.N. (1983) On logical foundations of probability theory, in *Probability Theory and Mathematical Statistics* (eds K. Ito and J.V. Prokhorov), Springer-Verlag, New York, pp. 1–5.

22. Neyman, J. (1977) Frequentist probability and frequentist statistics. *Synthese* **36**, 97–131.

23. Borel, E. (1909) Sur les probabilites et leurs applications arithmetiques. *Rend. Circ. Mat. Palermo* **26**, 247–271.

24. Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Applications*, Academic Press, New York.

25. Mayo, D.G. and Spanos, A. (2004) Methodology in practice: statistical misspecification testing. *Philos. Sci.* **71**, 1007–1025.

26. Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*, Chapman & Hall, London.

27. Mayo, D.G. (1996) *Error and the Growth of Experimental Knowledge*, The University of Chicago Press, Chicago.

28. Fisher, R.A. (1955) Statistical methods and scientific induction. *J. R. Stat. Soc., Ser. B* **17**, 69–78.

29. Mayo, D.G. and Spanos, A. (2006) Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *Br. J. Philos. Sci.* **57**, 323–357.

30. Li, M. and Vitanyi, P. (2008) *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd edn, Springer-Verlag, New York.

31. Kolmogorov, A.N. (1963) On tables of Random numbers. *Sankhya, Indian J. Stat. A* **25**, 369–376.

32. Chaitin, G.J. (2001) Exploring Randomness, Springer-Verlag, New York.

33. Nies, A. (2009) *Computability and Randomness*, Oxford University Press, Oxford.

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

35. Miller, R.W. (1975) Propensity: popper or peirce? *Br. J. Philos. Sci.* **26**, 123–132.

36. Burch, R.R. (2014) Charles Sanders Peirce, The Stanford Encyclopedia of Philosophy, Winter 2014 edn, (ed. E.N. Zalta),

URL = http://plato.stanford.edu/archives/win2014/entries/peirce/ (accessed 22 August 2016).

37. Gillies, D. (2000) *Philosophical Theories of Probability*, Routledge, London.

38. Humphreys, P. (1985) Why propensities cannot be probabilities. *Philos. Rev.* **94**, 557–570.

39. Spanos, A. (2006) Revisiting the omitted variables argument: substantive vs. statistical adequacy. *J. Econ. Methodol.* **13**, 179–218.

40. Doob, J.L. (1953) *Stochastic Processes*, John Wiley & Sons, Inc., New York.