Defining Probability for Science: A New Paradigm

William A. Gardner

Abstract

Traditional axiomatically defined probability is, by intention, completely non-empirical. This does not serve empirical science as well as two alternative 100%-empirical probabilities: one for applications involving populations, especially the life sciences, and one for applications not involving populations, but involving single records of empirical time-series data. This article presents an in-depth mathematical and pragmatic analysis and comparison of these probability definitions. Logical reasoning based on this careful and thorough critique leads to two related disruptive proposals: (1) a paradigm shift in the area within the field of Mathematical Statistics that serves the sciences and engineering as well and (2) an adjustment to college and university curricula that cover probability and statistics for the sciences and engineering. Part of the critique consists of proving two key justifications for stochastic processes to be mathematically false.

Summary

In empirical work in science involving time series analysis based on time-average statistics derived from available time series of empirical data, any probabilistic analysis of the statistics must be as realistic as possible. Yet, abstractions inherent in the orthodox definition of probability take us away from empiricism. The orthodox definition of probability used throughout the sciences (and engineering) is *maximally abstract* and includes a hypothesized abstract population, regardless of relevance to applications. Upon careful review of this definition and consideration of its historical development, it becomes apparent that the originators of this definition were not strongly influenced by the needs of empirical science. Mathematician's objective of defining "the real probability", which would not exhibit the variability seen with empirical probability, ultimately led to a completely abstract or unrealistic definition for use in empirical science. Motivated by this observation, this article proposes an alternative definition of probability for single time-series records, with no population of time series, and provides a thorough comparative analysis between the orthodox definition and what is appropriately called the *maximally empirical* definition of probability—a definition that differs from both orthodox probability and orthodox so-called *empirical probability* (which still uses orthodox abstract probability). This cogent assessment is telling and leads directly to the conclusion that a paradigm shift in science and in the field of mathematical statistics that provides science with its tools for performing probabilistic analysis of statistics is long overdue. In addition, the formula for creating an analogous maximally empirical probability theory for populations of time series, where nonstationarity is of interest, is provided and is even more straightforward and is again distinct from the orthodox sound-alike empirical

probability. Together, these two theories provide for both sciences not involving populations and the life sciences which typically do involve populations.

Outline

I. **Background:**

-Statistical Time Series Analysis from Mathematics and Statistical Signal Processing from Engineering are synergistic

-Definition of Target field of endeavor

II. **Introduction:**

-Example of Target problem

-Empirical Probability—Two meanings

-Classical Population Probability and FOT-Probability are distinct empirical probabilities

III. **Point-by-Point Comparison:**

1) *No operational impact of FOT-probability on parametric statistical inference, but conceptually superior*
2) *FOT-probability is less precise but more accurate than the stochastic process*
3) *Stochastic Processes are 100% subjective and have no known link to real data*
4) *FOT-probabilistic analysis of time-average statistics is 100% objective and empirical*
5) *Low-reliability FOT-probabilistic analysis due to limited time-series data is better than undefined reliability of stochastic processes—optimizing for limited data*
6) *FOT-probability theory/method maximally respects empiricism in science, and the Stochastic-Process theory/method is ambivalent to empiricism in science*
7) *FOT-probability theory/method is a success story for Finitism*
8) *FOT-probability is 100% Objective and empirical; The stochastic process is 100% subjective and axiomatic*

IV. **Thesis:**

-Mathematics vs. Science

-Proposed Paradigm Shift

V. **Additional Details of Analysis:**

-Abstract Probability

-Empirical Probability

-Non-parametric Statistics

VI. **Present Situation and a Way Forward**

-Striking departure of the Stochastic Process from FOT-Probability for analysis

I. BACKGROUND

Statistical Time Series Analysis is an older term from science and formal statistics for a broad field of study that has much in common with the newer field born in engineering and called *Statistical Signal Processing*. These two fields have been cross fertilizing each other for some decades now. A recent example of what statistical signal processing has to offer to time series analysis for an ancient problem in the physics of the solar system, given in [1], concerns the topic of hidden periodicities and addresses an application to irregular periodicity detection in the 200-year-long daily record of Sunspot numbers. This new *Fraction-of-Time Probabilistic Method* of data-model fitting led to the discovery of a new 132-year cycle in this time series which, in turn, has led to hypotheses about galactic influence on our star, the Sun. This new discovery from a 200-year-old record of time series data speaks to the power of the methodology originating in engineering. As a metric for how old this is, the term *scientist* was first put into use a little under a decade less than 200 years ago, although the term *science* dates back to the 14th century. The present article aims to apply decades of work in engineering to the broad topic of probabilistic analysis of time-average statistics derived from single records of empirical time series data arising from scientific and engineering experimentation and investigation when no population of time series is available and unnecessarily abstract models are undesirable.

- ***This singular topic of interest in the primary portion of this article, which deals with non-population time-series data, is referred to as the target to emphasize that everything said in the main part of this article is intended to apply exclusively to this specific yet broad topic; namely, the probabilistic analysis of time-average statistics derived from a single record of time-series data. This is important in order to avoid unhelpful irrelevant complaints and protests from aficionados of stochastic processes The article briefly steps outside these target bounds in Appendix II.***

The objective in this application of engineering work is to question the scientific appropriateness of the long-standing orthodox definition of probability, born in mathematics, that pervades all of science and engineering, by conducting a deep comparative analysis, specifically relevant to this topic of the differences between this old standard and a newer alternative definition that arose in engineering. The conclusion reached is a bold recommendation for a paradigm shift from the orthodox definition to the alternative specifically for use in the targeted topic. In the larger scheme of things, the part of the study of probability having to do with *interpretations of probability* dwarfs the treatment in this article and its relatively narrow focus. Treading far more lightly in the comprehensive overview, entitled *Interpretations of Probability* provided by Stanford's Metaphysics Research Lab in the Department of Philosophy, which is accessible here https://plato.stanford.edu/archives/win2012/entries/probability-interpret/, it still take 50% more verbiage there than the generous allotment in this article, and the concluding words there are: "It should be clear from the foregoing that there is still much work to be done regarding the interpretations of probability."

From time to time, authors take a critical look at probability and its uses in science. The present article has the same topical focus as two recent perspective articles and an editorial in *Nature*, as summarized below: the questioning of our pervasive use of orthodox probability in our work. But the present article differs from these previous items in that it has substantially broader scope within the specific topic, and substantially deeper analysis, and this is also true with respect to all known past treatments of this specific topic that the Author has been able to find in the literature. This article is based on the Author's 40 years of diligent research on this topic.

Detailed theoretical treatment of the argument in favor of empirical time-average probability over the orthodox abstract purely axiomatic definition of probability based on the concept of random draws from an abstract population, as well as extensive practical demonstrations, could easily occupy multiple lengthy research papers and book chapters and has been so addressed, as evidenced by the references cited in this article. But the succinct treatment of this fundamental issue in this review article is intended to capture the essence of all this work in a manner that is accessible to the broad audience comprised of scientists and engineers who use probability in their work in time series analysis of empirical data.

Because of the widespread unawareness of the competitor to orthodox probability, which competitor is comprised of a fully developed theory that is strongly analogous to orthodox probability theory, it is the Author's considered opinion that a concise overview, with the aims of that which he has been writing about for decades, is long overdue and of potentially significant importance for all scientists and engineers practicing probabilistic analysis of statistics derived from empirical time series data.

Regarding the first of two recent articles in *Nature* [2], a comment thereon in [3] states "probability itself is an expression of uncertainty, but the calculations behind the percentages can seem objective. The author of that article, David Spiegelhalter, argues that this is impossible" and, to quote him, Spiegelhalter says, "My argument is that any practical use of probability involves subjective judgements." In contrast to this perspective, this article explains that involvement of subjective judgements is not present when *empirical probability* is used properly. Spiegelhalter also states in his article [2] that "All of statistics and much of science depends on probability—an astonishing achievement, considering no one's really sure what it is." Again, in contrast, this article explains precisely what empirical probability is and what orthodox population probability is, though the latter is admittedly abstract with no solid links to real data or, therefore, to empirical science.

Regarding the second of two recent articles in *Nature* [4], the author Peters argues that the ergodic hypothesis has been abused, especially in economics. The present article agrees with this without restriction to economics, but the content of this article does not support Peters' use of the term "ergodicity economics", the reason being that Peters' work is centered on non-ergodic behavior in economics, that is, on time-average statistical behavior over time for individuals in a population in contrast to highly distinct ensemble-average statistical behavior over the population. Peters' recommendations for change in econometrics from using ensemble averages to using time averages has nothing to do with ergodicity. Peters is simply recommending the use of Fraction-of-Time (FOT)-probability instead of population probability. This practice in other fields of statistical analysis, especially engineering, has been promoted in quite a few books and journal research papers since the seminal 1987 book [5].Yet, it remains an unorthodox practice, and this is the motive for this article's objective of exposing the long-overlooked advantages of FOT-probability, especially for probabilistic analysis of time-average statistics derived from single time-series records, with no reference to real or hypothesized imaginary populations of time series.

II. INTRODUCTION

The next section provides a simple point-by-point comparison of (1) the Stochastic Process model based on axiomatically defined abstract population probability and (2a) the alternative empirical finite-time Fraction-of-Time Probability model for statistical time series analysis and, separately, (2b) the non-empirical limit FOT-probability (in which the length of a hypothetical time-series record grows without bound) for statistical inference (parameter estimation and decision making) using mathematical models. To make the points of comparison there more concrete, we start here in the present section with an example of probabilistic analysis of time-average statistics obtained from a time series of data using FOT-probability, and we explain how this analysis would change if we were to use the orthodox method based on the population-probability of the stochastic process model.

Example: An example of a time-average-based statistic from a time series is the empirical variance of a data record $x(t)$ for some range W of time, t. For specificity, let $x(t)$ represent any erratic time series of measurements on a physical system in a physics experiment for which the empirical variance is defined by

$$\widehat{\sigma^2} = \frac{1}{W}\int_0^W [x(t) - \widehat{m}(x)]^2 \, dt$$

where $\widehat{m}(x)$ is the empirical mean defined by

$$\widehat{m}(x) = \frac{1}{W}\int_0^W x(t) \, dt$$

(An alternative discrete-time model is completely analogous.)

Assuming the data used above on the time interval $[0, W]$ is only 10% of the entire available data record on the larger time interval $[0, V]$, we can partition the data into 10 contiguous subsegments and calculate 10 independent measurements of empirical variance, indexed by n:

$$\widehat{\sigma^2}(n) = \frac{1}{W}\int_{nW}^{(n+1)W} [x(t) - \widehat{m_n}(x)]^2 \, dt, \quad n = 0,1,2,3,\ldots,9$$

As a metric for quantifying how reliable any one of these variance measurements is, we can calculate the percentage *P* of the 10 blocks of data of length *W* for which the deviation from the average of the 10 variances exceeds, say, 5% of that average:

$$P = \frac{1}{10}\sum_{n=0}^{9} U\left[\left|\widehat{\sigma^2}(n) - \widehat{m}\left(\widehat{\sigma^2}\right)\right| - 0.05\widehat{m}\left(\widehat{\sigma^2}\right)\right]$$

where *U* is the unit step function that is 0 when its argument is less than zero and one otherwise.

This quantity *P* is the fraction of discrete time *n* for which the empirical variance exceeds the specified error bound. This *Fraction of Time* (FOT) of occurrence of an event has all the properties of the standard axiomatic definition of probability and is referred to as *FOT-Probability*. It is an example of an element of a complete theory of finite-time probability (either discrete or continuous time) that can be used for probabilistic analysis of time series data [5],[6]-[9],[11]-[15].

In contrast to the above empirical method based entirely on real data, the classical method in mathematical statistics, misleadingly called *empirical probability*, which pervades science and engineering today, is to not use the data to measure the percentage (or probability) *P* over time, but rather to posit an abstract population-probability distribution function (model) for the statistic interpreted as a random variable over a (non-existent) population, and then calculate the abstract probability of this hypothetical random variable deviating from its mean by more than 5% of this mean. This would correspond to replacing the uniform average over the 10 real time samples in *P*, in the above formulation, with a probability-weighted average over all possible values of the random variable model of the statistic, the *expected value* of this random variable, or, equivalently, replacing the uniform average with the hypothetical ensemble or population average over infinitely many random samples of this random variable. Knowing that the average of time samples will, under mild conditions on a mathematical model of the time series, converge to the limit FOT-probability-weighted average, why would one create an abstract probability model for an imaginary population, instead of using the data at hand? Unfortunately, as an editor of Nature points out [16], ". . . we now instinctively calculate expectation values with the implicit belief that they reflect what happens over time." which requires a property called *ergodicity*, but ". . . proving ergodicity mathematically is generally very hard: in fact, for any system that finds itself out of equilibrium it is safe to assume it is non-ergodic." Clarification: in the classical usage of the term *empirical probability* mentioned above, it is the probabilistic parameter that is empirical, but the quantity used in mathematical statistics to replace *P* is the abstract non-empirical probability. This distinguishes common practice today from the alternative proposed here.

To complement this example, we briefly consider an alternative example which is a dual of the first example, where the statistic is an empirical variance and the probabilistic metric for this statistic is an empirical probability concerning the variance. In the dual problem, the statistic is a probability and the probabilistic metric is a variance. As an example, the estimated probability of the occurrence of an event, obtained by averaging over time the (0,1)-valued indicator of this event's occurrence in the time series of data, could be probabilistically evaluated using the estimated variance of this estimate. Reminder: the estimates in the FOT-probability theory are estimates of the limit-FOT-probability quantity, not the analogous quantity from an abstract population-probability model—a stochastic process.) Needless to say, in the target topic being addressed in this article, the statistic can be whatever the experimenter specifies and the empirical probabilistic metric also can be whatever the experimenter specifies, and both are defined in terms of, and computed directly from, the available empirical data.

The probabilistic concepts used in this example are admittedly unsophisticated and even trivial, yet a formal and powerful probability theory and associated mechanics of probability manipulation and computation has been built upon this FOT-probability concept, and it is highly analogous to the orthodox purely abstract axiomatically defined probability, including key theorems like the *Fundamental Theorem of Expectation* [5], that we all have been indoctrinated with, typically without any mention of the existence of this empirical alternative. This situation suggests that pragmatic users of probability in the sciences and engineering may be figuring this out on their own to the best of their ability and using some version of FOT probability, despite its apparent complete absence from their training in university coursework and the statistical tools developed by mathematical statisticians.

To seek an answer to the question posed in the first example above, the article includes further below a general point-by-point comparison of (1) the orthodox method of probabilistic analysis based on the abstract stochastic process model that hypothesizes a population of time series that does not exist and (2) the unorthodox method of probabilistic analysis based on the empirical FOT-Probability model that uses nothing more than time averages of available real data. But first, a few more words are merited to dispel the likely misinterpretation of the FOT-probability theories as essentially the same as the classical method well known as *empirical probability*.

Empirical Probability --Two Meanings: *Empirical Probability* is a very well-known concept and technique presented in many books. In its most common form, it is viewed as the method of computing relative frequencies of occurrence of events from a population of experimental trials to produce an estimate of an unknown abstract population probability. Most of the relatively few treatments for extending this concept/method to time series data assume an abstract discrete-time stochastic process model with independent identically distributed (i.i.d.) time samples. This makes the hypothetical time samples from the stochastic process mathematically identical to statistically independent samples from a population; so, the concept/method applied to time series data is then identical to that developed for samples from a population (for which there is no time series). We can group with the quantities addressed in this concept/method statistics that are not only empirical event-probabilities but also are empirical moments formed as averages of powers of samples of random variables, as estimates of abstract population moments, such as means and variances, and averages of other functions.

There appears to be only sparse consideration of this classical empirical probability concept for time-series data for which it is not assumed that these are sequences of samples of abstract i.i.d. random variables—a seriously limiting assumption. The reason would appear to be that this situation of statistical dependence among time sample cannot be made equivalent to having statistically independent samples from an abstract population—a requirement dictated by orthodoxy! It follows that, although the Author refers herein to finite-time FOT-probability as an empirical probability (because this is a natural use of the broad meanings of these two terms), the concept, theory, and method of FOT-probability are distinct from what we might call classical empirical probability. Consequently, the powerful theory of FOT-probability, upon which the Author wrote a substantial book in the mid-1980s, is distinct from classical empirical probability and its minimal known extensions to time series data. Having no need for an assumption of i.i.d. time samples, for the sole purpose of respecting orthodoxy, FOT-probability applies generally to stationary time series and various forms of cyclostationary time series, which include statistical cyclicities (for which stochastic-process counterparts can include generalizations that incorporate a nonstationary component, described using the modifier *asymptotically-mean*–see Item 8 in Section VIII). Another distinction is that the theory of FOT-probability makes no assumption of an abstract population or associated population-probability distribution. Thus, it is not a part of the abstract theory of stochastic processes and consequently has no need for the abstract Birkhoff ergodic theorem.

Despite the broad applicability of FOT-probability theory to empirical probabilistic analysis of time series data, and the absence of any need to assume an abstract population or assume an abstract stochastic process model, this theory has been largely ignored for 38 years by all except for a few of the Author's colleagues and their co-authors, who we might refer to as disciples. This odd situation is addressed in this article.

The FOT-probability referred to above involves only finite-length discrete- or continuous-time time series and makes no assumptions or use of stochastic processes; this makes it a *100% empirical theory.* There also is another version of the theory based on a hypothetical infinitely long time series for which the limit as averaging time approaches infinity is taken. It is fairly well known from the past, dating back to N. Wiener's *Generalized Harmonic Analysis* that that such infinite time averaging can be mathematically shown to get rid of all random effect in the time series. For example, the global temporal mean-square deviation of a sliding finite-time (local) temporal mean value about its infinite-time average can be shown to converge to zero as the finite averaging time grows without bound. This is analogous to the expectation operation getting rid of all random effects in a random variable. In fact, the infinite-time average is an expectation operator in the limit-FOT-probability theory. Interestingly, the 100% empirical finite-time average can also be mathematically shown to be an expectation operator satisfying the *Fundamental Theorem of Expectation* [5], though there is no argument that the finite-time expectation operator gets rid of all random effects. It is mentioned here in passing that the limit FOT-probability is a valid candidate in mathematicians' pursuit of "the real probability" discussed in Section VII on history.

The preceding remarks are all mathematically proven in the seminal treatment [5] and re-proven in 6]-[15].

In conclusion:

- ***FOT-probability theory is distinct from the classical concept of empirical probability. And, to the Author's knowledge, FOT-probability theory is not a part of the traditional field of* Mathematical Statistics*. It is therefore referred to herein as an* unorthodox theory*.***

The limit-FOT-probability theory complements the 100% empirical finite-time FOT-probability theory and has completely distinct uses, as explained here and elsewhere in this article. In particular, this theory is a dual to stochastic process theory and can replace that more abstract theory for purposes of fitting models to time series data or, in other words, it is a tool for parametric statistics and is every bit as viable for statistical inference and decision making, such as the Bayesian theory of minimum risk inference and the theory of maximum-likelihood parameter estimation and hypothesis testing, as the classical stochastic-process theory, as explained in Section III.

III. POINT-BY-POINT COMPARISON

A detailed comparison of (1), the *orthodox* stochastic process model of a time series for probabilistic analysis of empirical statistics derived from time-average measurements on the time series and (2), the *unorthodox* empirical FOT-probability model of a time series for this same purpose, is presented below.

1) *No operational impact of FOT-probability on parametric statistical inference, but conceptually superior*

No modeling capability or applicability of the Bayesian Statisitcal Inference and Decision Theory/Method is forfeited by using infinite-time FOT-probability CDF models for single time series instead of abstract population models (stochastic processes). In fact, one can obtain a parametric formulaic FOT-probability model for a time series directly from a formulaic stochastic process model for any time series in *statistical constant or cyclic or multi-cyclic equilibrium,* whose stochastic-process counterpart can include all asymptotically mean-almost-cyclostationary time series (which include all stationary, asymptotically mean-stationary, cyclostationary or almost-cyclostationary, and asymptotically mean-cyclostationary or asymptotically mean-almost-cyclostationary processes) [7].

The procedure simply requires that the specification of all abstract CDFs and moments in the stochastic process model be conceptually replaced with FOT-CDFs and FOT-moments, and all time-invariant random parameters be replaced with unknown time-invariant parameters. By doing this, and assuming formulas for the parameter-conditional PDF of the observable random variables can be derived from the model, the Bayesian method for statistical inference and decision is immediately applicable. The only other adjustment to the model that is required arises in connection with Bayesian Learning, which is represented by the evolution of a prior probability into a sequence of posterior probabilities that result from sequentially adding more and more observations. The adjustment is that the prior probability for any random parameter in the stochastic-process model, which is often subjective, be replaced with a model consistent with the reinterpretation of all random parameters as unknown parameters: this consistent model for prior probabilities is a uniform distribution over some user-specified admissible set of parameter values. Then in the FOT-model, the relationship between the posterior and prior probabilities is as follows:

$$P\left[a \,\middle|\, \{x(t): t \in V\} = \boldsymbol{y}\right] = \frac{f_{\boldsymbol{x}}(\boldsymbol{y} \mid a)}{\frac{1}{|A|} \int_A f_{\boldsymbol{x}}(\boldsymbol{y} \mid a) \mathrm{d}a} P[a]$$

where the prior = $1/|A|$ for *a* inside *A* and = 0 for *a* outside *A*, where *A* is the admissible region in parameter space for the parameter *a*. In this probability relation, *f* represents a joint limit FOT PDF model for the vector ***x*** of elements given by the set from the hypothetical time series, and the vector argument ***y***, of *f* , represents the actual vector of observed values of the data segment in the set of time points V as indicated in the left hand side of the equation. When V is a continuous time interval, this formula is just symbolic, because the number of time points in an interval of the real line is uncountably infinite, requiring an infinite dimensional integral. Only when V is a finite discrete set of time points does this symbolic formula become a quantity that can, in principle, be calculated.

The definition of joint FOT-PDF is the joint derivative with respect to all variables in the argument vector $\boldsymbol{y}$ of the joint cumulative probability distribution (CDF) for the set of time samples of the time series. Conceptually, this CDF is the limit joint FOT-CDF defined by [5]

$$F_{\boldsymbol{x}}(\boldsymbol{y}) = \left\langle \prod_{k=1}^{K} U\left[y_k - x\left(t + t_k\right)\right] \right\rangle$$

where the angle brackets represent the limit of the average over time t as the averaging time approaches infinity, and U is the unit step function that is 1 for all positive values of its argument and 0 otherwise. However, in the above probability relation, the PDFs are conditional PDFs and this is addressed as follows.

A conditional FOT-PDF can either be posited, as is done in many applications of stochastic process models, or it can be mathematically derived from a formulaic model for the infinitely long time series. The above key probability relation generalizes in an obvious manner to vector-valued parameters. So, the concept of conditional probability does apply here, but in a different sense than for stochastic processes; for an FOT PDF, the vertical-bar symbol for conditioning on a simply means the PDF depends on a parameter with unknown value, and a is a hypothetical value for that parameter. Thus, the mechanics of Bayesian inference are the same for limit FOT-probability models as they are for stochastic processes. But the necessity of interpreting the PDFs in terms of a non-existent population of time series is absent and, with it, common confusion that is not effectively addressed by the ergodic theorem (see Item 3) below) also is absent.

2) FOT-probability is less precise but more accurate than stochastic process

FOT-Probabilistic analysis of empirical time-average statistics is based entirely on a *less precise* but potentially *more accurate* empirical probability model (the descriptors *precise* and *accurate* are used somewhat loosely here), in comparison with the stochastic process used in orthodox probabilistic analysis or the analogous limit FOT-CDF of a hypothetical infinitely long time series: In place of typically standard posited CDF models for abstract stochastic processes (abstract population probability models despite the absence of a population associated with the data), the alternative uses exact FOT-probability measured (typically directly) in terms of finite time averages of specified functions of the actual single-record of time-series data. Because the amount of real data must be finite, the model precision is finite--some random effects remain despite the time averaging. This is seen by the variation in such FOT-probabilities computed over multiple distinct time intervals. However, because there is generally no known relationship between empirical data and a stochastic process model, the stochastic process model's accuracy is unknown, so its precision is not of much practical use. Moreover, the limit-FOT probability, like the stochastic process has perfect precision.

3) The Stochastic Process is 100% subjective and has no known link to real data

All that can be done in the orthodox stochastic-process approach to probabilistic analysis of time-average statistics, in an attempt to establish a link with real data, is to invoke the ergodic hypothesis, which is often unverifiable, to claim the abstract subjective probabilities or expected values used are asymptotes of time-averages as averaging time grows. However, these are hypothetical time averages that are made on hypothetical sample paths from the hypothetical ensemble (population) of the abstract subjective stochastic process, which *has no known link to real data*. This point exposes what seems to be a common misconception: The ergodic hypothesis gets us no closer to empirical reality. (This misconception is likely a result of the influence of orthodoxy on thought by technical specialists with dominant left-brain activity, in the absence of sufficient balancing right-brain activity [17])

4) *FOT-probabilistic analysis of time-average statistics is 100% objective and empirical*

In contrast to the undesirable situation in the setting of empirical science in Item 3), the method in Item 2) is 100% empirical and uses only objective probabilities calculated from the data.

5) *Low-reliability FOT-probabilistic analysis due to limited time-series data is better than undefined reliability of stochastic processes—optimizing for limited data*

In applications for which the length of the available times series of data is relatively limited, the empirical reliability of the FOT-probability model (CDFs or moments) is limited, and the degree of limitation cannot be quantified (because there is no more data available for this purpose if it is all used to compute the statistic). More generally, the empirical reliability of the time-average statistics of primary interest, calculated for each of a set of contiguous subsegments using an intra-segment average of some measurement function of the time-series data, is *empirically quantified* using the empirical CDFs or moments obtained from inter-segment averages of a performance-metric function over a set of subsegments, as illustrated in the example. It is acknowledged here that this is an undesirable but unavoidable situation in which estimates are used to evaluate estimates. The result of excessively limited data identifies the only drawback of the alternative FOT-probability approach: It may produce a low-reliability model (and, in fact, no model, if all the data is used to compute one instance of the statistic) due to the amount of data being too small. Yet, the subjective probability model, the stochastic process, has completely undefined reliability, unless a further layer of abstraction is introduced (similarly to what is done in the proof of the Law of Large Numbers); specifically, the introduction of the doubly stochastic process concept can be used, whereby its CDFs and moments are random variables with known means, variances, etc. This is again an undesirable situation, in that it is far removed from empirical realism and therefore inappropriate in empirical science. Furthermore, even when the available data is insufficient to obtain reliable probabilistic analysis of measured statistics using FOT-probabilities or moments, the statistics themselves might well be sufficiently reliable for the experimentalist's purposes. This is determined by the subsegment length, not the number of subsegments. Finally, as a consolation prize for adopting the FOT-probability model when the amount of available data is insufficient to obtain reliable FOT-CDFs for probabilistic analysis of the finite time-average statistics, one may posit a model as presently done when stochastic processes are used, but there's no need to assume this is a population probability, which is an integral part of using stochastic process models. Also, there is always the possibility of experimenting with the partition of the available data, by which the number of subsegments is traded off with their length, which means the reliability of the statistics is traded off with the reliability of the probabilistic model for evaluating the reliability of the statistics. Taken to the limit, one can use a maximal number of subsegments to obtain maximal reliability for an estimated CDF (or other probabilistic parameter to be used on the measured statistic), which forces a minimal segment length used to obtain a minimal-reliability statistic and then, as a second step, use all the data to recompute the statistic, and adjust the measured already-computed reliability of the initial statistic to reflect the factor by which the record length used for the statistic has been increased. For example, if the variance of the statistic is the probabilistic model for quantifying the reliability of the statistic, then the factor by which the segment length for computing the statistic has been increased can be used to divide the value of the measured variance computed for the statistic that uses the minimal amount of data. This factor is larger than the correct adjustment to the extent that the subsegments are not statistically independent. This process can be optimized by overlapping data segments by 50%, which doubles the number of subsegments at the cost of reducing the degree of their statistical independence. But it also reduces the correct value of the factor for adjusting the variance from the number of subsegments to the effective number of statistically independent subsegments. Classical work in power spectral density estimation has shown that 50% overlap tends to be optimal relative to larger or smaller percentages. (The approach described here reflects a level of pragmatism not often shown in highly orthodox methodology. It is suggested this is a reflection of the orthodoxy represented by complete reliance on the abstract stochastic process model.)

6) *FOT-probability theory/method maximally respects empiricism in science, and the Stochastic Process theory/method is ambivalent to empiricism in science*

If the time series data is obtained from a science experiment, in which the objective is to gain an understanding of a system or phenomenon, in which case accurate and precise models are scarce, then the obviously strong departure from realism of the stochastic process, that is described above, is inconsistent with the required empiricism in science. In contrast, the unorthodox alternative is 100% empirical. In those applications where limitations on the length of time series data is of concern, there is apparently no evidence that giving up altogether on empiricism and adopting the stochastic process approach is going to lead to better empirical science.

7) *FOT-probability theory/method is a success story for Finitism*

The argument here favoring the unorthodox option is not unrelated to the concept that forms the basis for Finitism: “If it’s infinite, it’s not science”. There are two sources of infinity in the orthodox approach to probabilistic analysis of empirical time-average statistics: 1) the assumed population (sample space) for the posited axiomatically defined stochastic process is typically infinite for non-trivial experiments, 2) the ergodic theorem provides a link between time averages and expected values for only infinite averaging time and, even then, the averages are hypothetically made on members of a hypothetical infinite population bearing no connection to the single record of real data being time averaged by the empiricist. Thus, the orthodox stochastic process model is rejected by Finitism but, unlike many other items rejected by Finitism, there is an obvious finite replacement that is clearly pragmatically superior.

8) *FOT-probability is 100% objective and empirical; The stochastic process is 100% subjective and axiomatic*

This article touches on epistemology of science as it relates to current science methodology. It addresses the strange situation in the field of time series analysis of the dominance of a theory based on an abstract axiomatic definition of probability over a strongly analogous theory based on a non-axiomatic empirical definition of probability in terms of time averages of measurements made on real times series data. This dominance of stochastic processes is an egregious violation of parsimony, which is a cornerstone of science. The violation consists of a model based on the assumed mathematical existence of a physically non-existent entity in the specific science topic targeted in this article: a population of time series. The violation also consists of the key role played by the *Ergodic Hypothesis* because this hypothesis is often not validated (validation can be a substantial mathematical challenge and is, consequently, often avoided), and this theorem is all about infinities: infinite time averaging and infinite ensemble averaging in the empirical-like interpretation of *Expected Value* (from the Law of Large Numbers). Moreover, the ergodic hypothesis does not establish what is needed for the stochastic process model to be scientific in the sense of relating to the available data. That is, there is no link between the statistics calculated from empirical data, that are to be probabilistically analyzed, and the expected values or probabilities of functions of stand-in counterpart statistics calculated from the abstract model. Rather, this theorem establishes only equality, in an infinite limit, of time averages of hypothetical sample paths from the stochastic process to mathematical expected values from the stochastic process, neither of which have any link to the available data.

- ***The orthodox probabilistic model, being 100% axiomatic, takes us as far from empiricism as possible. By adopting the purely abstract axiomatic definition of probability, MATHEMATICIANS CLOSED THE DOOR ON ANY FORM OF REALISM AND ANY CONNECTION TO EMPIRICAL DATA, AS THOROUGHLY EXPOSED IN THIS ARTICLE. NEITHER ERGODICITY NOR THE LAW OF LARGE NUMBERS CAN OPEN THIS DOOR.***

IV. THESIS

In this article, many distinct practical advantages of the unorthodox empirical theory of probability over the dominant abstract theory are exposed (above) and subjected to a rational discussion. It is submitted here, as a thesis, that:

- ***A reasonable conclusion to draw is that there is a need for a paradigm shift in education in the topic of probability and in the practice of probabilistic analysis of time series statistics in the field of mathematical statistics.***

Further support for this thesis is provided below and addresses the less-than-ideal implementation of the *Scientific Method*. The passage of 38-years since the seminal publication of a book on the FOT-probability paradigm with no appreciable movement in the direction established by the Author's proposal for a paradigm shift from the orthodox stochastic process model to the alternative, reflects the difficulty of, metaphorically speaking, releasing our grip on a rock. Numerous examples of theoretical and methodological breakthroughs resulting from use of this paradigm shift are cited in Sec. VIII and further support the suggestion that collective behavior and science by consensus may be responsible for the absence of a paradigm shift, which absence is illogical, otherwise non-scientific, and suggestive of human error.

In this article, the dilemma, concerning (1) the dominant use of the stochastic process in science and (2) the non-scientific character of the stochastic process, is immediately resolved by replacing the stochastic process model of probability with the long-overlooked Fraction-of-Time (FOT) probability model. Mathematicians have argued that the stochastic process is mathematically superior to the FOT-probability model, in large part, because the probability measure of event sets in the sample space and the sample space itself defining the stochastic process are axiomatically willed to possess the properties of *sigma additivity* of sets and *sigma linearity* of the expectation operator (which properties concern infinite unions of sets and infinite linear combinations of random variables), whereas the theory of FOT-probability is based on the *Lebesgue Measure* on finite intervals of time and the *Relative Lebesgue Measure* on the entire real line of sets of time points. The former measure is and the latter measure is not sigma additive and the induced expectation operator is and is not sigma linear, respectively, and this reflects the character of the entities of interest; it is not a result that is simply willed by axiom as it is for the stochastic process. The concern about these sigma properties is a mathematical concern but not a mathematical showstopper—see below. This fact suggests that the historical abandonment of FOT-probability was due to a *Red Herring* (being thrown onto the path towards the theoretical development of FOT-probabiity eventually produced by the Author). Since the case of finite-time averaging is the only case that is possible in real science, the choice of the FOT-probability definition over the stochastic-process definition gives up nothing from the standpoint of science. It does give up something from the mathematical standpoint of developing and proving mathematical theorems about abstract models of time series because finite-time FOT-probability is not as viable as the stochastic process for such purposes. The first reason is that edge effects associated with the ends of the finite-time interval used for computing statistics occur. For example, the input/output relation for the correlogram and periodogram statistics for filters, which is exact in the limit as the data segment length approaches infinity, is only approximate for finite-time averages [5,p.46]. The second reason is that the infinite-time version of FOT-probability does not possess the aforementioned sigma properties possessed by the stochastic process (as axiomatically dictated). Nevertheless, this is not a showstopper when one encounters an infinite union or an infinite sum. For example, when one encounters in mathematical manipulation of a time series the FOT-expected value of an infinite sum, evaluation must proceed by proper interpretation of the infinite sum as a limit of finite sums, and use of standard finite mathematical manipulation until the point is reached where the limit can be unambiguously evaluated [thereby avoiding just blindly interchanging expectation and infinite summation as done with a stochastic process, for which this is guaranteed to be valid by the axiomatically willed stochastic-process property of sigma linearity—provided that one's stochastic process is truly a Kolmogorov process. (Does anyone ever check?)

Consequently, there is no absence of sigma additivity and sigma linearity of the finite-time FOT-probability model and, so, such absence is not a valid reason for outright rejection as a mathematically viable tool for the scientific purposes targeted in this article and it is also not a valid reason for outright rejection of the limit FOT-probability model for parametric statistics, which is addressed in Sec. III. In the case of interest in empirical probabilistic analysis of finite-time statistics, these sigma properties are either irrelevant or satisfied and, in the *case* of statistical inference, such as parameter estimation for limit FOT-probability models, the absence of the sigma properties can, in principle, be accommodated through more careful mathematical manipulation as described above. And, in practice, when this does not resolve the issue, the likely reason is that a contrived model has been chosen.

In summary, sigma properties are mathematics; they are not science. Their absence in the case for which limit FOT-probabilities are used can likely be managed and the mentioned approximations that may arise with the finite-time model can often be characterized as a minor distraction in empirical science. So, let us dig further into this troublesome matter of the dominance of the abstract model over the empirical model for the purpose of fully exposing the present existence of a sound logical argument for initiating a paradigm shift in which we replace the comfortable role the theory of the stochastic process presently plays in probabilistic analysis of empirical time series data, as the only probabilistic tool considered, with an equally comfortable role the theory of the FOT-probability model is fully qualified to play.

It can be seen from the straightforward example given in Sec. II that empirical probabilistic analysis of real finite-time statistics is practically viable. The sometimes-erudite argumentation offered in this article is primarily intended to defend against mathematicians' complaints that apparently led to the universal adoption of the stochastic process mid-last-century as the only viable option.

- ***The elucidation here of mathematicians' complaints and the reasons given here for the impotence of these complaints is intended to remove the mystery that empiricists may perceive and may be unduly concerned about: mysteries that may have left scientists of the past feeling impotent to resist the movement away from empirical probability and toward the abstract axiomatic definition of probability put forth by A. N. Kolmogorov, which was uniformly celebrated by mathematicians who, generally speaking, may not have much interest in empiricism—the core of science***.

V. ADDITIONAL DETAILS OF ANALYSIS

Human nature's resistance to conceptual change—one might call it mental inertia—is a common challenge to assimilating new ways of thinking. Case in point: early-to-mid-20th Century mathematicians became captivated by a particular way of thinking about probability, which had long been only an intuitive concept, and consequently moved away from another way of thinking already underway, but nascent, that is more natural, more intuitive—yet still mathematical—and more in tune with empiricism in science and engineering.

- ***Mathematicians' abstract axiomatic probability has become orthodox, and the empirical alternative,*** **FOT-Probability*****, has all but vanished from formal recognition in academia and our research journals.***

To review the focus of this article before proceeding with deeper analysis, the particular area of study that is of concern is again defined by a desire to perform probabilistic analysis of data as a key conceptual tool in science and engineering. The particular object for probabilistic analysis considered here is pervasive throughout these fields: statistics obtained by time-averaging measurement functions of observations comprised of single records of time series data originating from physical phenomena with no interest in populations.

As practiced today, the topic of Item1 in Sec. III is known as *parametric statistics* because it is based on use of an abstract probabilistic model, the *stochastic process*, of the empirical data, which model typically contains parameters with unknown values to be adjusted to fit the data in one way or another. The alternative topic of primary interest in this article is referred to as *non-parametric statistics.* Yet, as practiced today and discussed in Section III, non-parametric statistics in general *still uses mathematicians' preferred definition of abstract probability for quantifying performance of non-parametric statistics* rather than using the alternative, also viable (for the use of interest addressed herein) definition of empirical FOT-probability. A key observation here is that populations of actual times series data are not part of the area of study focused on in this article, though there is another branch of statistics that specifically addresses data from real populations of physical entities and, as discussed in Section VI, is apparently amenable to an analog of FOT-probability theory and methodology called *Fraction-of-Population Probability* theory and methodology, which goes well beyond todays so-called *empirical probability* as explained in Sec. II. This immediately calls into question mathematicians' formulation of probability based on the axiom of existence of a population of time series.

- ***The axioms that fully define mathematicians' preferred concept of probability render orthodox probability a completely abstract concept, free of any relationship to reality. Most of us have, over time, come to accept this abstract definition as the only option. But it is not—not across the broad range of uses of probability concepts and especially not in the identified target application area that naturally motivates interest in an alternative definition of probability.***

Before proceeding, it is mentioned that the images included here are a sampling of human pillars of science and its methodology, from our past, who studied epistemology and championed empirical methodology. A goal of this article is to honor these giants for the wisdom of their insights into science.




Francis Bacon 1561-1626 John L. Lock 1632-1704 George Berkeley 1685-1753 David Hume 1711- 1776

Abstract probability: The population-based probability model of time series data is called the *stochastic process*. This model is today ubiquitous in the field of mathematical statistics, including specifically the part of this field that is targeted in this article. To illustrate the degree of abstraction in this probability model, we need to briefly dive into key characteristics of the model that arise in applications to empirical time-series analysis. In the identified target application area, heavy reliance is made on two fundamental results in the mathematical theory of probability: the Birkhoff Ergodicity Theorem and the Law of Large Numbers.

- ***Ergodicity of a stochastic process model tells us that infinite time averages (unimplementable with real data ~ UWRD) on UWRD sample paths from the model equal UWRD expected values from the model. It also tells us that finite-time averages of UWRD samples paths from the model approximate both these UWRD model characteristics.***

- ***The Law of Large Numbers (LLN) tells us that UWRD infinite ensemble averages over UWRD sample paths (members of the hypothetical population) from a stochastic process model equal UWRD expected values from the model. It also tells us that finite-ensemble averages of UWRD sample paths from the model approximate UWRD expected values from the model.***

- ***Neither this fundamental theorem nor law says a word about the real data one works with in real science and investigative engineering!***

Any relationship to real data depends on how well the stochastic process model fits the real data, which is a problem about which the ergodicity theorem and the LLN offer no help and for which no generally applicable or universal methodology exists.

The challenge of fitting a stochastic process model to real data, a problem in the field of parametric statistics, would seem to be off point for the target application area and specific objective here, which is a problem in non-parametric statistics not involving populations. Nevertheless, this topic is briefly delved into here to put FOT-probability into even broader perspective.

With the stochastic process tool, the opportunities for getting misleading results from models that do not represent the data with fidelity are common for the simple reasons mentioned above. This should not be surprising given the tenuous link with real data such as (1) the absence of populations, (2) the gap between real data and the individual sample paths of the model, which are only surrogates, (3) the challenging mathematics of validating the ergodic hypothesis for specific models, which challenge has recently been taken to new heights in the new cycloergodicity theorems [8] relevant to data exhibiting statistical cyclicity which is pervasive in natural science and some fields of engineering as well, (4) the limited intuitive explanations of the mechanism(s) that are responsible for the ergodicity theorems' validity, and (5) the hidden abstraction of the qualifier "with probability equal to 1" in ergodicity theorems and the related LLN.

Empirical probability: Motivated by these five difficult issues, we consider the alternative in which the probability of an event of interest occurring in a given time series is defined to be the Fraction-of-Time (FOT) the event occurs over time in the time series data. In contrast to the use of the abstract axiomatic stochastic-process tool, every numerical result one produces with the alternative FOT-probability calculations comes exclusively from the data; the statistics one chooses to calculate from the data and the FOT-probabilistic quantities one chooses to evaluate the statistics both come from the data, consisting of a single time series record, exclusively. The example given in Section II nicely illustrates FOT-probability and its difference from the axiomatic definition of probability.

Unlike the stochastic process, the finite-time version of this alternative is a non-population, non-abstract, non-axiomatic probability that is 100% empirical [5],[7] and is referred to herein as *empirical probability* and is therefore well suited to non-parametric statistics which, for time series analysis, is taken *in this article* to mean statistical analysis not involving stochastic process models (possibly containing unknown parameters). For clarity, the modified term *strict-sense non-parametric*, which refers to both the statistics and the probabilistic analysis of these statistics as being non-parametric, could be used, to distinguish it from the *traditional empirical probability* which, as described in Section II, is very-much parametric when it comes to probabilistic analysis of statistics.

The FOT-probability theory, in its original form produces time invariant models for both finite-time data and infinite-time data, such as empirical cumulative probability distributions. Such models are said to be *stationary*. The Author has generalized the theory of FOT-probability to produce periodically and almost periodically time variant probabilistic models, referred to as *cyclostationary* and *almost cyclostationary*, for applications to data exhibiting statistical cyclicity. Other forms of nonstationarity of the data, excluding minor variations on these, are not generally accommodated by FOT-probability but can be for special cases, most notably slow nonstationarity, also called *local stationarity* cf. [8]. The time varying probability model of a generally nonstationary stochastic process has absolutely no link to real non-population times series data. So, there is no further discussion here of this irrelevant topic. Worthy of note in passing, the restriction to stationary models rules out long-term empirical-probability measurements on any life form, including animals (and humans) and, to a lesser extent in some cases, plants, because life forms age. So, applications to the biological sciences or natural science is limited and this explains why population probability dominates statistical analysis in this broad field of science, where populations do actually exist. Yet, as mentioned in Section VI (and addressed in Appendix II), an analog 100% empirical probability theory can apparently be built for actual non-stationary populations of time series, by calculating relative frequencies of occurrence of events over a finite population of time series for events depending on any selected sets of time points.

One can refer to some or all of the time-average statistics and their FOT-probabilities as *statistical or probabilistic data modeling* but there's no possibility here that these models do not fit the data exactly. Every ingredient in any probabilistic analysis of statistics using FOT-probability is determined by the data and only the data and no approximations need be made. An empirical model can never lie about the data (although users of empirical models can, even if unintentionally), which certainly cannot be said for stochastic process models specifically because of their tenuous link with real data.

- ***The term and associated concept of* probabilistic analysis of time series statistics *in use today, even if the adjective empirical is inserted, is invariably based entirely on the standard abstract population probability model, as if the existence of both the strongly analogous totally empirical FOT-probability theory and the theory of the limit FOT-Probability are completely unknown.***

- ***Orthodox* empirical probability *is not what its name implies: although the statistics that measure probability are empirical, the analysis of these statistics is today invariably done using the abstract non-empirical stochastic process. The one exception to this extreme situation is the genuine 100% empirical methodology used in publications by the Author and his colleagues.***

The tempting interpretation of the fact that half the references cited in this article are authored or co-authored by the Author as an indication of limited utility of FOT-probability puts the cart before the horse, which has experimentally been shown to be ineffective for moving forward. A much more feasible argument for the unpopularity of FOT-probability in the literature is given in Section VII, History, and is the motivation for writing this article and being so explicit and sometimes redundant about all points to be made.

As explained in Sec. III, even for parametric methods of statistical inference and decision making, the FOT-probability approach is based on empirical data and only empirical data or an idealization thereof defined as a limit of a finite-time single-data-record model, which is typically posited on the basis of the modeler's knowledge of the physical system under study. This limit model can, in some instances or in principle, be derived from principles of physics. The operational viability of statistical inference and decision making based on FOT-probability is amply demonstrated in the seminal book [5] and amply supported by other books (e.g., [9],[10]) and numerous journal publications in the ensuing 38 years since publication of [5] (cf. bibliographies and reference lists in [9],[11],[12] and more recently [13]-[15],[18]). In addition, it is briefly shown in Section III that there is an FOT-probability-based theory of Bayes Minimum-Risk Statisitcal Inference and Decision and of Maximum-Likelihood Parameter

Estimation and Hypothesis Testing that is dual to this part of the classical theory based on the stochastic process.

In summary, if one chooses to do parametric statistical analysis, data modeling is by definition required and there seems to be no generally applicable or universal methodology using either stochastic processes or FOT-probability. So, the choice between stochastic process models and FOT-probability is perhaps almost equivocal for this purpose. Nevertheless, for the identified target application area, it is methodology for *non-parametric statistics* that is of prime interest.

Non-parametric statistics: The target application area is a part of the field of non-parametric statistics. FOT-probability is understandably called *empirical probability* by the Author in reference to this field. Despite a significant body of knowledge developed for non-parametric statistics in time series analysis, there does not appear to be a general development of a theory of FOT-probability analogous to the theory of stochastic processes, which is pervasive. Even the topic named *empirical probability theory* in classical mathematical statistics uses stochastic process models to mimic the analysis of empirical time-average statistics but actually analyzes surrogate sample-path statistics from the abstract population as if they were real empirical data. The mathematical theory of statistics has simply not been able to break free of reliance on abstract stochastic processes for the purpose of any and all probabilistic analysis of statistics derived from real data. (We might consider the statistical methods collectively called *bootstrapping* to be an exception to this broad statement, and this is briefly returned to in Sec. VI) As illustrated here, the word *probability* today has a single meaning around the world: it is based on an assumed population, and it is totally abstract, being fully defined in terms of axioms unrelated to reality.

The preceding remarks suggest:

- ***The stochastic process, specifically for non-parametric probabilistic analysis of time-average statistics from empirical single-record time-series data, shouldn't even be in the running for the most competitive probabilistic tool to be used in the race toward knowledge acquisition that we call science* or investigative engineering.**

Moreover:

- **For parametric statistics, the unnecessary stochastic process departure from realism impedes conceptualization of real-world problems in the broad application area identified herein and their potential solutions. Multiple quotations supporting this theme from past giants in the field of time series analysis are provided in the reference in Item 8 in Section VII.**

The seeking of a high-fidelity stochastic-process model of a real natural (not man-made) system or phenomenon and the translation back and forth between such an abstract model and reality that is required by science and (though to a lesser extent) engineering is difficult, subject to error, and inherently flawed [8],[14],[15],[18] as rationally explained in these first five sections.

VI. PRESENT SITUATION AND A WAY FORWARD

The common practice in time-series data processing of doing probabilistic analysis using the stochastic process and then, at the end, translating results involving expected values from the abstract model to real data by simply either deleting the expectation operation or replacing it with a finite time average and then replacing the abstract sample path with real data is often unjustified because the ergodic hypothesis has not been validated for the model and is more generally unjustified because ergodicity links expected values only to time averages on abstract sample paths of the stochastic process, NOT to the real data the user would like to analyze.

- ***There is nothing empirical about probabilistic analysis based on the stochastic process, which is a striking departure from analysis based on FOT-probability.***

- ***Given that the stochastic process tool is presently taught at essentially all colleges and universities offering probability instruction for science and engineering, to the complete and utter exclusion from relevant curricula of FOT-probability, the above indisputable facts dictate that consideration of a paradigm shift in both education and the practice of mathematical statistics in science and investigative engineering is merited.***

This does not mean wholesale replacement of instruction in stochastic processes, which have many valuable applications outside of the application area focused on in this article. It simply means augmentation of existing material being taught to create an awareness of an important, but presently ignored, conceptual framework and a basic understanding of what this alternative offers relative to present orthodoxy.

- ***This corrective action is absolutely crucial for the benefit of empirical science and, considering that science without empiricism is not science at all, it is absolutely crucial for all science*.**

Nevertheless, there are no panaceas in the real world. The two and (to my knowledge) only relative drawbacks of FOT-probability *for the identified areas of study* are (1) it offers no counterpart to the particularly abstract generally nonstationary stochastic process (unless there are copious amounts of data enabling independent FOT-probability analysis in different time regimes), and (2) the user of FOT-probability must have a sufficiently long record of time series data in order to obtain a desired level of reliability.

Regarding Item (1). The classical theory and method of empirical probability, which applies to populations, might be at least partially extendable and generalizable to provide an empirical counterpart to nonstationary stochastic processes. But more generally promising is the apparent fact that many of the arguments establishing superiority of FOT-probability theory and method relative to stochastic process theory and method apparently apply equally to a *Fraction-of-Population* analog called FOP-probability theory and method. This would then include natural science in the paradigm shift called for. To be clear, the FOP-probability theory and method referred to here is distinct from classical empirical probability, because, besides the statistics being based on finite population averages, so is the probabilistic model for evaluating the reliability of the statistics. This avoids the subjectivity of present-day probabilistic analysis of statistics in life sciences. In addition, there’s no need to make the very restrictive assumption that the time series are i.i.d sequences in some imagined probability space as in classical empirical probability. However, this substantially increases the computations needed, because, with no knowledge of stationarity, the FOP-probability for any event involving any finite set of time points must be recomputed for every time shift of interest of that set of points. It also should be emphasized here that, unlike past tentative and incomplete work on FOT-probability, which was limited to stationary times series, the complete theory proved by the author includes time series exhibiting cyclostationarity with single and multiple incommensurate periods of statisitcal cyclicity.

Regarding Item (2). To avoid the issue of item (2) above, the analyst (a) must have enough data to compute useful (sufficiently reliable) statistics and (b) must have enough additional data to compute multiple samples (from subsequent segments of the time series) of these statistics from which FOT-probability theory can be used to compute useful (sufficiently reliable) measures of reliability of the statistics. In practice, the methodology here is typically executed by partitioning the available time series record into subsegments, each subsegment of which is used to compute statistics, as illustrated in the example in Section II. Then time-average measurements of functions of these statistics are computed by averaging over the segment time index for the purpose of quantifying reliability, as also illustrated in the example. Details regarding this partitioning process on how to minimize the impact of limited time-series data are presented in Section III. In addition, an alternative that may have the potential benefit of reducing requirements on data-record length is methodology referred to as *bootstrapping* [19]. In the simplest of terms, applied to time-series analysis, bootstrapping amounts to replacing the deterministic partition of time series data discussed above by a random selection of subsets of time points, with replacement. Bootstrapping has the appearance of "getting something for nothing", which is at odds with the maxim "there are no free lunches". The example of the multi-segment methodology given in Section II applies to stationary time series but can be generalized for cyclostationary time series for single and multiple incommensurate periods of cyclicity, cf. [5. Chaps. 5,15], [13] As a final remark, it should be mentioned here that in exactly the cases for which a time series record is not sufficiently long, the utility of ergodicity for stochastic processes is lost, because ergodicity's asymptotic results are then even more irrelevant than explained in Section V w.r.t. the apparently ignored distinction between time averages of sample paths of the stochastic process and time averages of the time series data.

The concerns here and remedies suggested also apply to FOP-probability when population size is limited. In life sciences where huge populations exist, this is not a problem in principle, but it is an economic issue, and there appears to be a tradeoff between more reliable 100% empirical studies and the economic cost of gathering data.

Though corroboration of the likelihood that FOT-probability is actually in use in the practice of science, despite apparently being a relative rarity in research journal publications in science, would require an extensive formal study, it presently appears as though this practice is not receiving support through education in academia. As a remedy, one simple but valuable addition to curricula that includes stochastic processes is a simple supplement exposing students to the theory of FOT probability, and revelation of its strong analogy with, but still distinct advantages over (in appropriate applications), the theory of stochastic processes that are stationary or exhibit some form of cyclostationarity. This would give them an option following graduation for choosing, in each application they face in the future, which of two conceptually distinct yet mathematically analogous competing theories to use. Similarly, for life sciences curricula where populations and nonstationarity are of critical importance, the FOP-probability analog of FOT-probability theory should be developed and taught. The relatively mature theory of FOT-probability for cyclostationary time series provides a template for development of an analogous theory of FOP-probability. The differences between these two analogs are 1) FOP performs averages over the population index instead of time and it does this over all time shifts within a range of interest to capture nonstationarity, whereas for FOT and cyclostationarity, only time shifts throughout one period are needed.

VII. HISTORY

A study of the history of mathematization of the concept of probability provides telling insight into how we got ourselves into the present predicament in which the FOT-probability option is apparently ignored [12]. It is said in the historical review of probability in [20] that the *frequentist view,* based on interpretation of probability as the relative frequency of occurrence of an event, has problems:

- ***A mathematician's explanation of why the stochastic process was adopted [20]:* "It is of course impossible to actually perform an infinity of repetitions of a random experiment to determine the probability of an event. But if only a finite number of repetitions of the process are performed, different relative frequencies will appear in different series of trials. If these relative frequencies are to define the probability, the probability will be slightly different every time it is measured. But the real probability should be the same every time. If we acknowledge the fact that we only can measure a probability with some error of measurement attached, we still get into problems as the error of measurement can only be expressed as a probability, the very concept we are trying to define. This renders even the frequency definition circular."**

From this Author's perspective, the key concept of "the real probability" that drove mathematicians to eventually define probability axiomatically, thereby relegating the relative-frequency concept to a position of inferior status, **was the source** of so much dispute in the early 20th Century, as reviewed in [20, Sec. Frequentism]. Recognizing this, it should be asked: *What justification is there for belief in "the real probability" of anything non-real, like the stochastic process?*

- ***This Author submits that the issue of* ideal probability*, touched on in [20], has not been adequately dealt with. It is argued here that in some instances events are defined so broadly that they do not qualify as real events, so seeking real probabilities is misguided, and in those cases for which events are clearly real, real probabilities often do exist.***

The connection here to the theme of this article is that FOT-probability is an uncommon type of relative frequency: it is the relative frequency of occurrence of an event of interest over time in a record of real time series data. It is not a relative frequency of occurrence of an event over a set of statistically independent trials of a hypothetical experiment, which is the relative frequency that was of primary interest to probabilists. Nevertheless, the concern that an FOT-probability measurement will not be the same if it is repeated over subsequent segments of the time series exists, as an analog of the problem that bothered probabilists in the past.

- ***The Author simply sidesteps the concern that the relative-frequency definition of probability was seen in the past as circular, by proposing that available time series data be used, via FOT-probabilities, to probabilistically evaluate time-average statistics which can themselves be FOT-probabilities, as illustrated in the example in Section II. The fact that these probabilities are not free of variability, does not mean they are not real probabilities, as explained below. "The real probabilities" mathematicians envision can only exist mathematically, which is as unrealistic (totally non-empirical) as one can get. Moreover, for real events, properly defined empirical probability is indeed real as shown here.***

In empirical work, *reality* concerning a phenomenon or input-output system being investigated consists of two things: 1) whatever pre-data-measurement concrete facts are known to be true about that which is being investigated and 2) whatever data in the form of observations/measurements of behavior made on that which is being investigated is available. *For example, if one has no a-priori facts but does have a single record of time-series data, then that data IS the reality. Any event that does occur at any of multiple instants during the time span of the record is a real event and, if one selects time instants at which to look for this event at random, uniformly over the record, the fraction-of-time (FOT) that the event occurs is the real probability of that real event*. If one then records more data from the same phenomenon or system, the reality changes and so we expect the probability of a real event to change. Rather than considering new data to be a new reality ending up with a new real probability that differs from the first one computed, giving rise to concern about variability (as suggested in [20]), the new reality should be considered to be the concatenation of the two data records. One can then compute a new real probability and rest assured that, although it differs from the first real probability, it *should* be considered to be more reliable. Moreover, given still more data, a 100% empirical measure of reliability can be computed. Several mathematical properties of this reliability measure support this "should" statement. To mention just one, given the axiom that the FOT-probability of occurrence of the event converges as the data record length grows without bound, it can be shown that for any desired level of reliability measured in any of several allowable ways, there exists a data record that is long enough to achieve that level of reliability. The theory and method of limit FOT-probability is a counterpart (a dual) to the more abstract stochastic process model that past mathematicians chose as the basis for their definition of true probability of the sort that has zero variability. The reason it is said here that the stochastic process is more abstract is because, although they both leave reality behind by taking a limit as a sample size approaches infinity, to obtain zero-variability probabilities, the stochastic process also axiomatically posits a population which has no counterpart in reality for the topic being addressed in this article. This is an egregious further conceptual departure from reality. Even more significant is the fact that the proposed finite-time FOT-probability does not depart from reality at all, because it does not take the aforementioned limit. (Since this entire article is focused on probability computed from empirical data, no further consideration is given here to the "concrete facts" referred to above in defining *reality*.)

- ***The mathematicians' concern about circularity was created by their pursuit of "the real probability". Once we define what* reality *is and distinguish between real and unreal events, the issue vanishes, and the use of empirical probabilities to evaluate the reliability of other empirical probabilities can be seen to be sound, and the only limitation on how effective this is is determined by the amount of data available. This is discussed in Section VI, The Way Forward.***

- ***For completeness here, the reader is reminded that, given a mathematical model of a time series, the limit of the finite-time FOT-probability can be taken and this produces one example of the mathematician's concept of "real probability". As discussed above and in Section II, all random effects vanish in the limit of the FOT-probability. And this is achieved without hypothesizing an abstract infinite population of time series. This enables users of FOT-probability to engage in some of the types of mathematical analysis that the stochastic process is touted for, as discussed in Sec. III.***

Before moving on in the discussion, the opportunity is seized here to illustrate the duality of orthodox probability and FOT-probability. The following calculation involving only fractions of time illustrates the conventional probability rule for a mixture of FOT-probabilities. We begin with the following definitions: m_1 is the number of occurrences of event *A* within data record y_1 which contains n_1 time points, $P\left(y_1 \,\middle|\, mixture\ \{y_1, y_2\}\right) = n_1/(n_1 + n_2)$ = the FOT-probability that the randomly selected time instant falls in the y_1 segment, given the mixture of records y_1 and y_2, $P\left(A \,\middle|\, y_1\right) =$ the FOT-probability that event *A* occurs at a randomly selected time instant, given the single record y_1 from the mixture, $P\left(A \,\middle|\, mixture\ \{y_1, y_2\}\right) =$ the FOT-probability that the event *A occurs at a randomly selected time instant,* given the mixture. Given these definitions, we have

$$P\left(A \,\middle|\, mixture\ \{y_1, y_2\}\right) \underset{def}{=} (m_1 + m_2) / (n_1 + n_2)$$

$$= [m_1 / n_1]\, [n_1 / (n_1 + n_2)] + [m_2 / n_2]\, [n_2 / (n_1 + n_2)]$$

$$\underset{def}{=} P\left(A \,\middle|\, y_1\right) P\left(y_1 \,\middle|\, mixture\{y_1, y_2\}\right) + P\left(A \,\middle|\, y_2\right) P\left(y_2 \,\middle|\, mixture\{y_1, y_2\}\right).$$

Although the theory and method of Bayes' statistical inference can be used with limit FOT-probabilities, as explained in Sec. III, it cannot be used with finite-time FOT-probability, because the FOT-PDFs of events of interest, given specific event outcomes, are not available from typical measurements, because this requires an unusual level of control over the phenomenon being investigated. Nevertheless, an alternative to the concept of *Bayesian Learning* is available and can easily be derived from the above expression. Using new but obvious notation, it can be shown that the FOT-probability of an event *A*, given *K* segments of time series data can be recursively computed from the FOT-probability of event *A*, given *K-1* segments and an update term as follows:

$$P\left(A \,\middle|\, \{y_k\}_1^K\right) = P\left(A \,\middle|\, \{y_k\}_1^{K-1}\right) P\left(\{y_k\}_1^{K-1} \,\middle|\, \{y_k\}_1^K\right) + P\left(A \,\middle|\, y_K\right) P\left(y_K \,\middle|\, \{y_k\}_1^K\right)$$

For applications in which the FOT-probability of an event *A* is expected to change due to actual changes in the phenomenon under investigation, as data continues to come along, the above recursion can be modified to give more relative weight to the most recent data. To achieve this, in the RHS of the above recursion, modify the 2nd factor in each of the two terms by replacing n_K with Fn_K, where F is the factor by which you wish to increase the *effective length* of the most recent data segment, e.g. the 2nd factor in the second term of the recursion is replaced with $Fn_K / (Fn_K + n_{K-1} + \ldots + n_1)$ and a similar replacement is done in the second factor of the first term.

As another example of the possibilities presented by FOT-probability, it is briefly mentioned here that one can transform a single-record stationary time series (or single- or multi-period cyclostationary time series) with finite memory into a near-Markov time series, by performing linear least-squared-error prediction to subtract out dependence of present on past that goes back farther than one unit in time. The errors between the present value and the past values used in this procedure are obtained by sliding the segment of data spanning the present and *M* most recent data values backward one unit of time for each data set over which the squared linear prediction errors will be summed before minimization w.r.t. the coefficients in the linear combination of *M-1* past values excluding the most recent past value, The desired transformed time series is the series of prediction errors.

It seems that mathematicians' emphasis on the superior mathematical viability (due entirely to Kolmogorov's 1933 axioms [21], which abandoned all realism) of stochastic processes for formulating and proving theorems had an outsized impact on users of probability in science and engineering, especially in the identified target application area. There appears to have been negligible push-back on this shift away from earlier use in physics and engineering of the FOT-probability concept as reflected in Norbert Wiener's 1930 treatise on Generalized Harmonic Analysis [22]. It is conceivable that Brillinger's mid-1970s book on time series analysis [23] is the (or one of the) "smoking gun(s)" in the temporary but long-lasting (1960's to present) death of the FOT-probability alternative because of the position he took, namely, to *explicitly* choose stochastic process theory over FOT-probability theory, while claiming the two theories (the limit FOT probability model and the stochastic process model) are equivalent for stationary processes. This equivalence is mathematically disproved in Appendix II and the differences are substantive. Several other classic books on stochastic processes as well might have provided considerable impetus for forgetting all about the FOT-probability concept, which is what happened. The authors of these books are mathematicians. Textbooks that followed in engineering and science appear to have simply followed suit, with primarily one salient exception: the Author's mid-1980s textbook [5] and on the order of 50 subsequent solid research papers, book chapters, and books, all authored by the Author and his colleagues (recall the preceding reference in Sec. V to the advisability of not putting the cart before the horse).

Although mathematics is its own master, historically much of mathematics has been driven by science. The opposite should not be happening: mathematics should not drive science except in those rare circumstances for which solid new insights into difficult practical problems suddenly arise from a stroke of mathematical genius.

- ***THE PRESENT UBIQUITOUS USE OF THE STOCHASTIC PROCESS IN THE FIELD IDENTIFIED HEREIN HAS THE UNMISTAKABLE APPEARANCE OF MATHEMATICS DRIVING SCIENCE AND THE RESULT, IN THE AUTHOR'S OPINION, IS A DEGRADATION OF EMPIRICISM IN SCIENCE AND INVESTIGATIVE ENGINEERING, WHICH IS NOT A DESIRABLE OUTCOME.***

VIII. DEMONSTRATIONS

Due to limited space, non-trivial demonstrations of the many benefits of FOT probability over population probability that are identified herein cannot be included. In place of such demonstrations, readers are referred to the recent discovery of a *new long-term cycle* in Sunspot series data based exclusively on FOT-probabilistic analysis [1],[18]. This is a topic that has been studied for at least two centuries.

- ***The recent discovery at this late date of something never before seen in the 200-year-old Sunspot series suggests there may indeed be conceptual advantages of the FOT-Probability theory and methodology.***

This is salient evidence, though certainly not proof, of potential pragmatic benefits of the essentially unknown FOT-probability over the orthodox alternative and various ad hoc methods. But, more generally:

- ***The Author suggests that the absence of awareness of FOT-probability theory and consequent total dependence on population probability in the absence of populations could be responsible for subtle delusion or even deceit and, without deep probing, invisible and therefore insidious effects on thinking. The Author is not the first to express this concern [15].***

Reflecting on commentary throughout this article, it seems there has been an all-out acceptance of the population probability concept illogically in applications *where there is no population*, resulting in blindness to FOT-probability. As discussed in Sec. VII, mathematicians' motivation for preferring population probability was their apparent belief that it was the only choice that produced what is called *the true probability* (a probability without variability); however, limit FOT-probability also offers this absence of variability. The complete takeover by the stochastic process has the appearance of subtle indoctrination, but who or what would be responsible for such indoctrination? The Author can offer only the conjecture that potentially unintentional influence of mathematicians in the 1960s -1970s era of stochastic-process book-writing may have indoctrinated readers because of the absence of any counterbalancing teachings in the alternative nascent FOT-probability theory. As an example of the blindness mentioned here, readers are referred to a published debate on this probability topic, which has been reproduced at the educational website [7, p.3] specifically because it explicitly illustrates this blindness and illogical behavior.

To conclude this article and to complement the implication of the highly unexpected scientific discovery in the Sunspot series example given in Sec. I, the following nine examples are given:

- ***A list of nine major breakthroughs in the two subfields identified below is presented and supporting documentation is cited:***

  (1) parametric statistics as used in communication systems design and analysis, based on limit FOT-probability models, and

  (2) non-parametrical statistics as used for statistical inference algorithm design and implementation, based on finite-time FOT-probability calculations

Seven of the following nine major breakthroughs are a direct result of the Author's first two breakthroughs: introduction and single-handed comprehensive development of two major independent but complimentary wide-reaching theories: Fraction-of-Time Probability and Cyclostationarity, the first of which is hopefully about to motivate a major paradigm shift in the definition of probability for time series analysis in science and the second of which now permeates all fields of science and engineering. The joint exploitation of these two theories has directly led to breakthroughs 3 - 9:

1) Full Development of basic Fraction-of-Time Probability for time series (including signals) exhibiting stationarity and pioneering multi-period cyclostationarity. This generalized N.

Wiener's *Generalized Harmonic Analysis* theory: (1) from 2nd-order to nth-order, 2) from wide-sense to strict-sense, and 3) from stationary time series to cyclostationary and almost cyclostationary time series—see Item 2) [5],[7, p.9.1]-[10]

2) Full seminal development of basic Multi-Period Cyclostationarity theory (a comprehensive set of 20 independent theorems defining much of the theory of cyclostationarity), illustrated by a plethora of signal modeling examples and algorithms invention for statistical analysis and statistical inference and decision making--see Items 3) – 9) [5]-[7, p.9.1]-[10]
3) Comprehensive communication signals modeling (cyclostationarity characteristics, especially spectral correlation density functions) including first derivation of the cumulant as the solution to a problem, which isn't even probabilistic, and is the basis for breakthroughs 4 - 6) [5,chap.12], [24],[25]
4) Separation of spectrally overlapping signals with revolutionary Frequency-Shift Filtering techniques based on the Author's original theory of Cyclic Wiener Filtering, generalizing N. Wieners theory of non-causal Wiener Filtering from stationary signals to multi-period cyclostationary signals [5],[26]
5) Classification of spectrally overlapping communication signals demonstrating innovative capability widely recognized as crucial to cognitive radio and signal interception for national security [27]
6) Blind Communication Channel Equalization using only 2nd-order statistics—a new capability previously unrecognized because of use of stationary signal models [28], [29, Chap. 3]
7) Nonlinear system identification for system input signals of opportunity or under the control of the experimentalist; using cyclostationarity to get over the hurdle for which N. Wiener's MIT group did not find success in 20 years [5, p.11.7]
8) Probability modeling in science specifically for time series analysis of time-average statistics—a major paradigm shift, in the making, toward true empiricism in science [30]
9) Parameter estimation with radically new Method of Moments—first substantive breakthrough in MoM theory and methodology in over a century [31]

The technical claims made in this article are backed up with the publications by Gardner and Napolitano and their coauthors in the reference list.

- ***A reasonable conclusion to draw is that there is a need for a paradigm shift in education in the topic of probability and in the practice of probabilistic analysis of time series statistics in the field of mathematical statistics.***

Appendix I: Fraction-of-Population (FOP) Probability for Both Statistics and Their Probabilistic Analysis

In a completely analogous fashion, FOP-Probability can be defined for empirical populations of times-series data just as done for FOT-Probability for single empirical time series by simply averaging over the population member index for each of all time translations of interest instead of averaging over the time index. In general, this produces a nonstationary CDF. By partitioning the population member index set Ω into a superset of subsets, statistics can be measured over each of the subsets, and then probabilistic functions (e.g., for quantifying statistic reliability) can be computed by averaging over the index for the superset.

Given a finite set of time series $\{x(t, \omega);\ t \in W,\ \omega \in \Omega\}$ over a time interval *W*, with population sample index set Ω, the joint empirical CDF for *K* time points computed from this data is defined by

$$F_{\boldsymbol{x}}(\boldsymbol{y}, t) = \left\langle \prod_{k=1}^{K} U\left[y_k - x\left(t + t_k, \omega\right)\right] \right\rangle$$

where the angle brackets here mean average over the sample index ω instead of the time index *t* as in Sec. III. By doing this for each value of time translation *t* over some set of interest, a generally nonstationary CDF is obtained. Just as for the FOT-CDF, the FOP-CDF satisfies the *Fundamental Theorem of Expectation* [5]. Actually, for the FOT-CDF, it is a Fundamental Theorem of Time Averaging [5], and for the FOP-CDF, it is a Fundamental Theorem of Population Averaging (FTPA).

The traditional probability functions, including joint moments and cumulants for multiple time points can be calculated directly from a population average, or indirectly from the CDF using the FTPA.

Appendix II Proof of Inequivalence of the Kolmogorov Ergodic Stochastic Process Model and the FOT-Probability Model

In order to prove that the Kolmogorov ergodic stationary stochastic process model is not equivalent to the FOT-probability model (contradicting Brillinger [23]), this appendix presents a stochastic process model that IS equivalent and then shows that this process is not the same as a standard Kolmogorov process with the same CDFs. This material is taken from [15].

Definition of stationary FOT-stochastic process

Def.S1: The Sample Space of the Stationary FOT-Stochastic Process is comprised of all the time translates of a single relatively measurable discrete- or continuous-time sample path (persistent real-valued function of a real variable), x, subject to the constraint that replications are disallowed (no two sample paths can be identical):

$\Omega_d = \{\{x_{n-\omega};\ n \in \mathrm{Z}\};\ \omega \in \mathrm{Z}\}$,

$\Omega_c = \{\{x(t-\omega);\ t \in \mathrm{R}\};\ \omega \in \mathrm{R}\}$

Def.S2: The probability of any relatively measurable subset of elements from the sample space index set R or Z, called an event, is the value of the relative measure of that set.

Def. S3: The FOT-CDF of any relatively measurable discrete- or continuous-time function, $f[x](t)$ or $f[x](n)$, which is jointly relatively measurable, for m real-valued time points $\{t_1, t_2, t_3, \ldots, t_m\}$ or m integer-valued time points $\{n_1, n_2, n_3, \ldots, n_m\}$, respectively, of the Stationary FOT-Stochastic Process $x(t)$ or x_n is the relative measure of the event set

$E_c(m) \triangleq \{\omega \in \mathrm{R};\ f[x](t_1-\omega) \le \xi_1,\ f[x](t_2-\omega) \le \xi_2,\ \ldots,\ f[x](t_m-\omega) \le \xi_m\}$

or

$E_d(m) \triangleq \{\omega \in \mathrm{Z};\ f[x](n_1-\omega) \le \xi_1,\ f[x](n_2-\omega) \le \xi_2,\ \ldots,\ f[x](n_m-\omega) \le \xi_m\}$

for all real-valued m-tuples $\{\xi_1, \xi_2, \xi_3, \ldots, \xi_m\}$.

It follows from Def.S3 that the1st-order FOT-CDF for a continuous time stationary FOT-stochastic process is given explicitly by the formula

$$Fx(\xi) \triangleq \mu_R(\{t \in \mathrm{R}: x(t) \le \xi\})$$

$$= \lim_{U\to\infty} (1/U)\, \mu(\{t \in [\, t_0-U/2,\, t_0+U/2];\, x(t) \le \xi\})$$

$$= \lim_{U\to\infty} (1/U) \int_{t_0-U/2}^{t_0+U/2} \mathrm{u}(\xi - x(t))\, \mathrm{d}t$$

for all real ξ, and similarly for higher-order FOT-CDFs, where μ_R is the Relative Lebesgue Measure and μ is the Lebesgue Measure. For discrete time, the FOT-CDF is given by

$$Fx(\xi) \triangleq \mu_R\, (\{n \in \mathrm{Z};\, x_n \le \xi\})$$

$$= \lim_{N\to\infty} [1/(2N+1)]\, \#(\{n \in [n_0-N,\, n_0+N];\, x_n \le \xi\})$$

$$= \lim_{N\to\infty} [1/(2N+1)] \sum_{n_0-N}^{n_0+N} \mathrm{u}(\xi - x_n)$$

where # is the counting measure. As another example, for m = 2, we have the 2nd-order FOT-CDF

$$Fx(\xi_1,\xi_2) \triangleq \mu_R(\{t \in \mathrm{R};\, x(t+t_1) \le \xi_1,\, x(t+t_2) \le \xi_2\})$$

$$= \lim_{U\to\infty} (1/U)\, \mu(\{t \in [\, t_0-U/2,\, t_0+U/2];\, x\,(t+t_1) \le \xi_1,\, x(t+t_2) \le \xi_2\})$$

$$= \lim_{U\to\infty} (1/U) \int_{t_0-U/2}^{t_0+U/2} \mathrm{u}(\xi_1 - x(t+t_1))\, \mathrm{u}(\xi_2 - x(t+t_2))\, \mathrm{d}t$$

for all real ξ. Note: The constraint in Def.S1 that disallows replications in the sample space also disallows constant signals, which are a degenerate case of stationary signals.

The probability of the entire sample space of the Stationary FOT-Stochastic Process is equal to 1, meaning every experimental outcome for this model is one of the members of the sample space. That is, for a discrete sample space $\Omega_d\,(N)$ with a finite number N of translates, the probability of each translate is $1/N$ and since these translates are mutually exclusive events, the probability of the entire set of N translates is the sum over N probabilities, each equal to $1/N$, which sum equals 1. In the limit, as the number of translates N included in the sample space approaches infinity, we get the result that the probability of each sample path is 0 and the probability of the total sample space Ω_d is 1. Similarly, for a continuous sample space, the probability of each sample path is 0, because the relative measure of a single point on the real line is 0, and the probability of the total sample space Ω_c is 1, because the relative measure of the entire real line is 1.

For this FOT-stochastic process, any one of the translates, $\{x(t-\omega);\, t \in \mathrm{R}\}$ for any particular $\omega \in \mathrm{R}$ or $\{x_{n-\omega};\, n \in \mathrm{Z}\}$ for any particular $\omega \in \mathrm{Z}$, can be taken as the Sample Space Generator. In practice, the sample space generator would be taken to be the single observed signal, conceptually extended from the finite observation interval to the real line, or to the integers; and when a formulaic specification of the process is made, the sample space generator would be obtained from the formula for any specified set of random samples of the random functions in the formula. So, given a specification of one sample path, we have a specification of the entire sample space.

Stationary FOT Ergodic Theorem:

1. Every Stationary FOT-Stochastic Process is Strongly Ergodic, by construction, meaning the infinite time averages of relatively measurable functions of the process exist and are independent of the particular sample paths selected and are equal to the expected values of those functions obtained using the FOT-CDF or FOT-PDF.

2. Every Finite-Ensemble Average of every function of a Stationary FOT-Stochastic Process is identical to a Finite-Time Average of that function.

The validity of this theorem follows directly from the Definitions. It is noted here that ensemble averages are typically conceived of as being performed on randomly selected ensemble members, which do not occur in any ordered fashion. In contrast, time averages are typically performed on time-ordered time samples or time translates. Item 2 in this theorem does not assume any ordering. However, when one approaches the question of convergence of time averages as the length of averaging time approaches infinity, time ordering is desirable and typically assumed (e.g., as in a Riemann integral), but no such ordering can be assumed for random selection of ensemble members. To avoid the technical details involved here (which are of no pragmatic interest), Item 2 addresses only finite averages and, like Item 1, states a fact that is obvious from the construction of the sample space.

Relation to Wold's Isomorphism

Wold introduced an isomorphism in 1948 [32], which is referred to here in its extended form that accommodates continuous time processes, between (1) the sample space of a stochastic process, defined to consist of the collection of all time translates of a single time function, including that time function itself, and (2) this single time function. This isomorphism establishes a distance-preserving relationship between the stochastic process, with its definition of squared distance as the ensemble-averaged squared difference between two processes, and a single sample-path of that stochastic process, with its definition of squared distance as the time-average of the squared difference between two sample-paths. This mapping between the metric space of a stochastic process and the metric space of a single sample path therefore preserves distance and is consequently an isomorphism. The above sample space is identical to that in Def. S1 for a Stationary FOT-Stochastic Process. By complementing this sample space with an FOT-Probability measure satisfying Defs. S2 and S3, we obtain a Stationary FOT-Stochastic Process. Wold did not take this step, and– according to the Author's literature search– apparently did not pursue the conceptual path taken in the present article.

Comparison of Kolmogorov and FOT-stochastic process models (the magic hand)

To illustrate how simple the sample space of a stationary FOT-stochastic process is, compared with one of the simplest examples of the sample space of a Kolmogorov process, consider an infinite sequence of statistically independent finite-alphabet real-valued equally probable symbols, with alphabet size K. The Kolmogorov sample space for a finite sequence of length N contains K^N distinct sequences and the probability of each is $(1/K)^N$. The probability of the entire sample space is the sum of the probabilities of the K^N mutually exclusive and exhaustive sample paths, each having probability

$(1/K)^N$, which sum equals 1. In the limit, as the sequence length approaches infinity, we get the result that the probability of each sample path is 0 and the probability of the total sample space is 1.

This sample space includes as a strict subset the entire FOT sample space generated from any one of the Kolmogorov sample paths. The Kolmogorov probability of this FOT sample space is the limit, as N approaches infinity, of $N(1/K)^N$. Therefore, the Kolmogorov probability of the entire FOT sample space is 0. This is a result of the fact that the sample space represents a single signal—a single infinite sequence of $K-$ary symbols, not all possible infinite sequences of $K-$ary symbols. The Kolmogorov sample space apparently contains not only the FOT sample space of all translates of one infinite sequence but also contains the FOT sample spaces of all translates of every possible infinite sequence of $K-$ary symbols. Despite the huge difference in the sizes of these two sample spaces, as N approaches infinity, it is interesting to note that the FOT probability of a subsegment comprised of a specific sequence of length N occurring over the lifetime of the function is $(1/K)^N$, and this is the same as the probability of selecting a sample path from the corresponding Kolmogorov stochastic process that possesses a particular subsegment of length N comprised of this specific sequence. Because the time position in a stationary time series or a stationary stochastic process is of no probabilistic consequence, the difference in sizes of these sample spaces appears to be of no consequence unless one is interested in studying populations of time series. As a reminder, the Birkhoff ergodic theorem guarantees that the time average of every sample path in this immense sample space equals w.p.1 the expected value and this equals w.p.1 every ensemble average. This mysterious result is not necessary in practice; it is not a prerequisite for having a probability theory for time-series analysis. The much simpler FOT-stochastic process will do for types of applications of interest in this article, for which populations of signals are not of primary interest, and this means that the entire stochastic process concept can be discarded for these types of applications and replaced with a single signal and its FOT-probability model. Sample spaces are then irrelevant. The cost of abandoning the Kolmogorov stochastic process model is that the FOT-Probability measure is in general not sigma-additive, and the corresponding FOT-expectation operation is not in general sigma-linear. However, the utility of these sigma properties exists only when performing calculations involving infinitely many subsets of the sample space or sums of infinitely many functions of the process. Moreover, to benefit from these properties, one must verify that a specified probability measure does indeed exhibit these assumed properties. This is rarely done in practice, except when well-known probability measures, like the Gaussian, which have already been verified, are adopted. But there are no models for manmade communications signals in use that are Gaussian and the same is apparently true for models of naturally occurring biomedical signals, and signals of many other origins. If there is not a large number of independent samples of random variables added together to form a random variable to be modeled, there is generally no reason to expect that random variable to be Gaussian.

Another way to compare these two models of stochastic processes is as follows. Consider, as an example, a Bernoulli sequence with parameter p = 0.3. This is a sequence of statistically independent binary random variables with values of 0 and 1 having probabilities of 0.3 and 0.7, respectively. A sample path for the Kolmogorov model is denoted by $x(n,\omega)$, where n is integer-valued and ω also needs only take on a countable infinity of values, and can therefore be taken to be integer valued. The values this function of two integer variables can take on are 0 and 1. The specification of the actual infinitely large 2-dim array of 0's and 1's is such that every possible sequence of 0's and 1's is included once and only once. So, the specification of the sample space is simply exhaustive. But there is a specification of a probability measure for this function of ω for subsets of values of n. The measure tells us the limit, as the number of randomly selected values of ω approaches infinity, of the relative frequency of sets of 0's and 1's at these subsets of discrete time points that will occur as outcomes. This probability measure is like a *magic hand* that guides the selection of experimental outcomes so that at each time point 1's are selected in 70% of the experimental outcomes and 0's are selected in 30% of the outcomes. And, for example, the pair of adjacent outcomes of 0 followed by 1 are selected in (0.3)(0.7) = 21% of the outcomes. There is an inherent abstractness here, which I call a *magic hand*. It cannot in general be made concrete or given a concrete interpretation. And it is not a property of the sample space. It is simply a specified rule regarding the randomly selected outcomes of an experiment. It should be clarified here that the strong law of large numbers establishes that averages over ensembles of random samples converge to expected values w.p.1 not because of replication in the sample space (which is not allowed), but rather because of the magic hand. Replications of entire sample paths occurring with non-zero probability are disallowed in the Kolmogorov model, as they are in the FOT model; however, for any finite set of time samples, the same finite set of sample path values can occur in infinitely many distinct sample paths all of which differ in at least some of the values at other time points. But the numbers of these partial replicas are determined by nothing more than combinatorics. In contrast, the relative frequency of occurrence in random samples of sets of process values at subsets of time points is determined by only the magic hand. This fact is often not recognized in the literature. For example, even the classic book by Middleton [33, Section 1.3, pp. 26–27] includes invalid attempts at explaining the convergence of ensemble averages to expected values in terms of replications of sample paths in the sample space. Similarly, for the sample space defining the FOT-stochastic process (e.g., continuous time), replications like $\{x(t-\omega_1); t \in R\} = \{x(t-\omega_2); t \in R\}$, $\omega_1 \neq \omega_2$, are disallowed (Def. S1) because they do not produce what we think of as random functions since they imply $x(t)$ is simply periodic with period = $| |\omega_1 - \omega_2| |$. In contrast to the Kolmogorov sample space for the Bernoulli process, a sample path for the corresponding FOT-stochastic process is given by (with some abuse of notation) $\{x(n,\omega); n,\omega \in Z\} = \{x(n-\omega); n,\omega \in Z\}$ and this function $x(n)$ takes on values of 0 and 1. Given a single sample path $x(n)$ on the integers, we have a full but non-exhaustive specification of $x(n,\omega)$ throughout the entire sample space (2 dim array). Because of this, there is no need for a magic hand. We can derive the probability measure by simply calculating (in principle, at least) the limit of the relative frequencies of 1's in $x(n)$. Any statistical dependence of these binary variables in the sequence also can (in principle, at least) be calculated from joint FOT-probabilities. Work on designing sequences that exhibit specified relative frequencies can be found in the literature. The above discussion illustrates that the details and level of abstraction of the Kolmogorov stochastic process model are often not observed in applied work in statistical signal processing. Consequently, there is little pragmatic justification for continuing to hang onto the baggage (abstraction) that comes with this standard model when populations of signals are not of primary concern, when we have the much simpler and more concrete alternative, the FOT-Probability model for single signals.

In [15], an analog of the above creation of an FOT stationary stochastic process model is presented for an FOT cyclostationary stochastic process model, and a model for an FOT multi-cycle cyclostationary process (for incommensurate cycle frequencies) is defined in terms of these first two models.

Acknowledgement

This contribution to knowledge is dedicated to my wife, Nancy S. Gardner, 15 August 1946 – 26 March 2024, in remembrance of a lifetime of smiles, love, companionship, and support of my work summarized in this article.

References

[1] W. A. Gardner, Long-term solar cycle discovery and galactic influence on the Sun, *Solar Physics*, submitted, 2025

[2] *Spiegelhalter, D. Why probability doesn't exist (but it is useful to act as if it does). Nature* **636**, 560-563 (2024) https://doi.org/10.1038/d41586-024-04096-5

[3] *Flowing Data* https://flowingdata.com/2024/12/19/probability-does-not-exist/#:~:text=Probability%20itself%20is%20an%20expression%20of%20uncertainty%2C%20but,any%20practical%20use%20of%20probability%20involves%20subjective%20judgements.

[4] Peters, O. The ergodicity problem in economics. *Nature Physics,* **15** (12), 1216–1221 (2019) https://doi.org/10.1038/s41567-019-0732-0 .

[5] W. A. Gardner, Statistical Spectral Analysis: A Non-Probabilistic Theory, (Prentice-Hall, Englewood Cliffs, NJ,1987). (Author Comments: *SIAM Review*, **36**, No. 3, 528-530,1994, https://epubs.siam.org/doi/epdf/10.1137/1036137)

[6] W. A. Gardner and W. A. Brown. Fraction-of-time probability for time-series that exhibit cyclostationarity, *Signal Processing*, **23,** No. 3, 273-292 (1991)

[7] W.A. Gardner, https://cyclostationarity.com/

[8] W. A. Gardner, On Cycloergodicity, *Signal Processing (EURAIP)*, July 2025

[9] W.A. Gardner, Ed., Cyclostationarity in Communications and Signal Processing (IEEE Press, Piscataway, NJ, sponsored by *IEEE Communications Society*, 1994)

[10] A. Napolitano, Cyclostationary Processes and Time Series: Theory, Applications, and Generalizations. (Academic Press, 2019)

[11] W. A. Gardner, A. Napolitano, and L. Paura, Cyclostationarity: Half a century of research, *Signal Processing (EURASIP)* , **86,** 639-697 (2006)

[12] J. Leskow and A. Napolitano, "Foundations of the functional approach for signal analysis", *Signal Processing (EURASIP)*, **86**, n.12, 3796-3825 (2006)

[13] D. Dehay, J. Leskow, and A. Napolitano, "Time average estimation in the fraction-of-time probability framework", *Signal Processing (EURASIP)*, **153**, 275-290 (2018) doi:10.1016/j.sigpro.2018.07.005

[14] A. Napolitano and W.A. Gardner, Fraction of time probability: advancing beyond the need for stationarity and ergodicity assumptions, *IEEE Access*, **10**, 34591-34612 (2022) doi:10.1109/ACCESS.2022.3162620

[15] W. A. Gardner, Transitioning away from stochastic process models, *J. Sound and Vibration* **565** (2023) doi.org/10.1016/j.jsv.2023.117871

[16] Editor. Time to move beyond average thinking. *Nature Physics.* **15**, (12), 1207 (2019) https://doi.org/10.1038/s41567-019-0758-3

[17] I. McGilchrist, The Master and His Emissary: The Divided Brain and the Western World, New Haven and London: (Yale University Press, 2010) ISBN: 978-0-300-16892-1

[18] Napolitano and W.A. Gardner, Discovering and modeling hidden periodicities in science data, *J. Advances Signals Processing,* https://doi.org/10.1186/s13634-025-01207-w, (2025)

[19] J. L Horowitz, "Bootstrap methods in econometrics"*. Annual Review of Economics.* ***11*** *193-224* (2019) arXiv:1809.04016. doi:10.1146/annurev-economics-080218-025651

[20] https://en.wikipedia.org/wiki/Probability_interpretations

[21] A. N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd ed., English. Dover Publications, Inc. Mineola New York

[22] N. Wiener, Generalized harmonic analysis, *Acta Mathematika*, **55**, 117-258 (1930)

[23] D. R. Brillinger, *Time Series: Data Analysis and Theory,* Reprint, SIAM. (Philadelphia, 2001)

[24] W. A. Gardner and C. M. Spooner, The cumulant theory of cyclostationary time series, Part I: Foundation, *IEEE Transactions on Signal Processing*, **42**, No. 12, 3387-3408 (1994)

[25] C. M. Spooner and W. A. Gardner, The cumulant theory of cyclostationary time series, Part II: Development and applications, IEEE Transactions on Signal Processing, **42**, No. 12, 3409-3429 (1994)

[26] W. A. Gardner, Cyclic Wiener filtering: Theory and Method, *IEEE Transactions on Communications*, **41**, pp. 151-163 (1993)

[27] https://cyclostationary.blog/my-papers/

[28] S. V. Schell and W. A. Gardner, Spatio-temporal filtering and equalization for cyclostationary signals, Chapter in Control and Dynamic Systems, edited by C. T. Leondes, Academic Press, **64**, 1-85, 1994

[29] W. A. Gardner. Cyclostationarity in Communications and Signal Processing, (Edited book with invited and reviewed contributions), Piscataway, NJ (IEEE Press, sponsored by *IEEE Communications Society*,1994)

[30] W. A. Gardner, On cycloergodicity, *Signal Processing* (EURASIP), **238** (2026) 110186

[31] W. A. Gardner, A radically different method of moments, *EURASIP Journal on Advances in Signal Processing*, *EURASIP Journal on Advances in Signal Processing*, **17** (2025) https://doi.org/10.1186/s13634-025-01219-6

[32] H. O. A Wold, On prediction in stationary time series, *Ann. Math. Stat*. **19**, 558-567 (1948)

[33] D. Middleton, An Introduction to Statistical Communication Theory, Reissued by (Wiley-IEEE Press, 1996) Print ISBN: 9780780311787,htt p://refhub.elsevier.com/S0022-460X(23)00320-6/sb13