Transitioning Away from Stochastic Process Models

William A. Gardner
8 January 2022

Abstract

The standard theoretical foundation for statistical signal processing is presently the discrete-time and continuous-time Kolmogorov stochastic process models for persistent signals and especially, but not exclusively, strongly ergodic and cycloergodic Kolmogorov stochastic process models. After a brief discussion exposing drawbacks of these generic models for many applications in statistical signal processing, particularly those involving empirical data, an alternative stochastic process model is proposed for statistically stationary signals and a complementary model for statistically cyclostationary signals also is proposed. For these alternative models, defined first in terms of a specification of their samples spaces, cumulative probability distribution functions (CDFs) or, equivalently, probability density functions (PDFs) are derived from Fraction-of-Time (FOT) probability calculations on a single member of the sample space, and then shown to be valid CDFs over the entire sample space of the process. If all such finite-dimensional CDFs are specified, then this corresponds to a complete probabilistic model for the alternative stochastic process. The motivating difference between Kolmogorov's model and this alternative model is that the alternative is derived from empirical data, at least in principle. It is not posited in an abstract manner that typically leads to a number of conceptually confusing and usually unanswerable questions about the behavior of the sample paths. These alternative models are then complemented with another model for poly-cyclostationary signals that exhibit multiple incommensurate periods of cyclostationarity. The conceptual and practical advantages of these three types of alternative models are discussed in some detail, and it is shown that the entire framework of stochastic processes, with its non-empirical abstraction, can be altogether circumvented by using FOT-Probability models for single signals, without any reference to stochastic processes. These single-signal models are identical to the alternative stochastic process models introduced here, but they do away with the unnecessary sample space because it is redundant. These most elegant of models provide all the same tools for statistical analysis—including CDFs, PDFs, temporal moments and cumulants, spectral moments and cumulants, and so on—but without any reference to stochastic processes. In the final analysis, it is recommended that the alternative stochastic process models introduced here be used primarily as a pedagogical tool that helps in understanding the circumstances under which stochastic process models are unnecessary for statistical signal processing and probabilistic analysis involving stationary, cyclostationary, and poly-cyclostationary signals. These circumstances are, simply stated, any situation in which stochastic processes are appropriate provided that only ergodic or cycloergodic or multi-cycle generalizations thereof are of interest. The general situation for which stochastic processes are actually required, rather than avoidable, as a mathematical basis for statistical processing and analysis is that for which the lack of ergodicity is an essential characteristic. This is typically those situations for which ensembles of signals are an essential ingredient. Nevertheless, when a stochastic process model is non-ergodic but is *conditionally ergodic*—meaning conditioned on knowledge of some finite set of parameters of the signal model, the data PDF is ergodic—and when this conditioning can be either experimentally

implemented or mathematically enforced in a data model, then the conditional FOT-PDFs can be measured or calculated and used in the same manner as PDFs for traditional stochastic processes. This enables the incorporation of FOT-Likelihood functions in the FOT-Probability theory.

1. Introduction

The statements of theoretical results and discussion of practical ramifications provided in this article are written for statistical signal processing engineers and like-minded time-series analysts, which may include physicists and other specialists in the physical sciences, and other fields where statistical analysis of empirical time-series is of interest. It is felt that mathematical proofs at any higher level of rigor than that which is presented herein would be distracting and are not included for this reason and others. Because the specific reasoning given in this article is not at odds with the day-to-day reasoning generally used by the intended audience, little of value would be added for this audience if a more mathematically rigorous presentation were provided. The preference acted on here is especially appropriate since the whole point of the effort leading to these new models is to show practitioners that the substantial abstractions and unmet challenges of trying to verify strong ergodicity or cycloergodicity of traditional stochastic process models are in the great majority of applications nothing more than distractions from the reality of empirical data and its processing and analysis and the more elegant theory that is identified here and is based on *Fraction-of-Time (FOT) Probability* for single signals.

Perhaps the most important reason for not getting distracted by rigor is that these new models are intended for only the pedagogical purpose of providing a conceptual transition from stochastic process models to FOT-Probability models of single signals and demonstrating that stochastic process models are often an unnecessary abstraction: they forfeit parsimony and mathematical elegance relative to the alternative single-signal models with fraction-of-time probability calculated directly from the single signal.

The three-decade history from the 1930s through the 1950s of time-average statistical theory of time series is traced in [1] but the first approach to more comprehensive *Fraction-of-Time Probabilistic Modeling* of signals seems not to have been introduced until the concise publications of Brennan [2] and Hofstetter [3] in the 1960s. This approach was later developed independently[1] and more comprehensively, including extension/generalization from stationarity to cyclostationarity, with in-depth application to the theory of statistical spectral analysis by Gardner in 1987 [4] (see also in [5]).

The time-average approach was the starting point for the use of statistical time-series analysis in physics but has been largely ignored for well over half a century by many college instructors and criticized by some mathematicians for supposedly being non-rigorous. However, it has recently been shown by Leśkow and Napolitano to have a rigorous basis in measure theory, using mathematical tools dating back to the work of Kac and Steinhaus in 1938 [6]. This basis for measure-theoretic rigor underlying Fraction-of-Time Probability Theory was apparently lost track of in the shadow of Kolmogorov's contributions earlier the same decade. But, well over half a century later, it was uncovered by Leśkow and Napolitano in 2006 [7], where a more complete list of early (1920s to 1940s) contributors to time-average statistical theory is given (see also [8] by Napolitano).

To counter the appearance of avoiding technical detail that may be important in comparing the two approaches to stochastic process modeling discussed in this paper, a glimpse into such details is provided in this paragraph and here and there in following sections. The Relative Measure used in [7] for the mathematical foundation of FOT-probability model is not sigma additive (probabilities of infinite unions of nested event sets do not all converge), but in Kolmogorov's stochastic process probability model, sigma additivity of the proposed probability measure *is only assumed* by virtue of Axiom VI [9]. So, this axiom does not guarantee that, for any particular stochastic process model one adopts, the probability measure will in fact be sigma additive. Kolmogorov simply removes the mathematically undesirable general lack of sigma additivity of measures by axiomatically removing from consideration all probability measures that are not sigma additive. But how often do we encounter practitioners seeking to determine if the probability measure for some stochastic process model they have adopted is sigma additive or even just seeking to explicitly describe the probability measure for their adopted model? This is a very rare event. For the Fraction-of-Time Probability Theory discussed herein, an alternative restrictive assumption is required: the undesirable general lack of relative measurability of functions of time series is avoided by removing from consideration all time series and functions of those time series that are not relatively measurable. Such prohibited time series can be constructed, but they also can be considered anomalous. These restrictions are discussed further in [7].

Outline of Paper

1. Introduction
2. Brief Historical Remarks
3. Results
   - 3.1 Kolmogorov's model of a stochastic Process
     - Birkhoff's Ergodic Theorem for discrete Time
     - Birkhoff's Ergodic Theorem for Continuous Time
   - 3.2 The Measure Theory of FOT-Probability
   - 3.3 Definition of *Stationary FOT-Stochastic Process*
     - Stationary FOT Ergodic Theorem
     - Relation to Wold's Isomorphism
   - 3.4 Comparison of Kolmogorov and FOT stochastic Process Models
   - 3.5 Definition of *Cyclostationary FOT-Stochastic Process*
     - Cyclostationary FOT Cycloergodic Theorem
   - 3.6 The FOT-Probability Model for Almost Cyclostationary Processes
   - 3.7 Cycloergodicity for Multiple Incommensurate Periods
     - How to Generalize Birkhoff's Ergodic Theorem
   - 3.8 Purely Empirical FOT-Probability Models for Regular Cyclicity
   - 3.9 Purely Empirical FOT-Probability Models for Irregular Cyclicity
4. Discussion of Results
5. References

---

[1] In the early 1980s, as I was writing the textbooks [4] and [5], I discovered earlier work [2] and [3] as a result of discussions with Professor Thomas Kailath of Stanford University. I added to the Introduction in my book draft citations of this work from two decades earlier. As discussed in the present article and in more depth at the UC Davis website [22], earlier work on time-average theory, including [2] and [3], appears to have been largely forgotten as the *stochastic process bandwagon* trend developed (this colorful characterization was passed on to me and my colleagues at UC Davis in 1987 by the late coding theorist and cryptographer Professor James L. Massey of ETH Zurich).

2. Brief Historical Remarks

To put this proposed evolutionary step in larger perspective, some stages of signal modeling that this community has passed through over the last century are briefly summarized. Time-series analysis goes back more than a century, but the time of R. A. Fisher one century ago seems to be a turning point when broader theoretical frameworks began to be formulated. This includes most notably Fisher's Principle of Maximum Likelihood, which is among the most commonly used optimization criteria for designing statistical inference and decision rules—algorithms—in use today within the statistical signal processing community. This includes both signal-parameter estimation and signal detection and classification. Predating Fisher by two centuries was Thomas Bayes, who gave birth to the theory of Minimum-Risk Statistical Inference and Decision (which addresses the same or similar signal parameter estimation and signal detection and classification problems that Maximum-Likelihood addresses, but with the added axiom that prior probabilities [prior to experimentation including observation or data collection] are assumed to exist). More recently, just preceding the middle of last century, Norbert Wiener used his developing statistical theory of single time functions (signals) to derive what we now call the Wiener Filter and related linear time-invariant signal processors, using a time-average counterpart of the Bayes Minimum-Risk design criterion, where risk was specified to be expected squared error, reformulated as time-averaged squared error. This was the continuous-time counterpart of Carl Friedrich Gauss's discrete least-squares optimization criterion used two centuries ago. Wiener's time-average theory and its applications to the nascent field of statistical communication theory was given a boost in visibility and further developed in 1960 with the publication of a book by one of Wiener's previous students at M.I.T, Yuk Wing Lee [10]. That same year, David Middleton's landmark book *An Introduction to Statistical Communication Theory* was published. In contrast to Lee's book, Middleton's ~~used~~ was solidly based on the theory of stochastic processes. It has been said to cover a panoramic view unmatched by any other publication in the field [11]. This book was likely instrumental in cementing the place of the stochastic process in statistical signal processing. Middleton states in his preface "The mathematical exposition is for the most part heuristic". Although he does favor obtaining autocorrelation functions from signal models using time-averaging, he then takes an expected value to obtain an ensemble autocorrelation. Because of this approach, he misses the fact that some of his signal models are cyclostationary, not stationary. Nevertheless, he does note that, in general, his approach produces stationary autocorrelations for nonstationary processes. This precedes more theoretical work decades later on what are called *asymptotically-mean stationary processes*, which includes as special cases cyclostationary and almost cyclostationary processes. Middleton, however, does not adopt the Kolmogorov model for stochastic processes. He uses heuristics instead.

Contemporaneously with Wiener in the 1930s and 1940s, Kolmogorov introduced the now-standard theory of the stochastic process as a probabilistic model for time-series. Also contemporaneous was the establishment of *Information Theory* by its originators, Harry Nyquist, Ralph Hartley, and Claude Shannon during the 1920s – 1940s. The landmark event *establishing* the discipline of information theory and bringing it to immediate worldwide attention was the publication of Claude E. Shannon's classic paper "A Mathematical Theory of Communication" in the Bell System Technical Journal in 1948. This theory is strongly probabilistic. From 1960 forward, Wiener's time-average approach quickly faded into the background, and Kolmogorov's expected-value approach grew into the standard we use today. It is conceivable that this was in large part a result of the boom that information theory initiated and possibly also a result of the mathematical rigor of Kolmogorov's book on the theory of stochastic processes. Interestingly, though, information theory involving signals is valid for time-average probabilities, not just ensemble-average probabilities, as discussed further on in this paper.

What has for almost a century been referred to as *statistical time-series analysis* has increasingly come to be relabeled *statistical signal processing*, perhaps because of the lead electrical engineers have taken in developing the technology used for implementation. This field of study, born within the field of electrical engineering, was originally based in large part on what is called *statistical communication theory*, which arose out of the work of Wiener and his contemporaries but was reformulated in terms of expected values and stochastic process models. This theory is more probabilistic than it is statistical, yet it is called a statistical theory by the authors of classic books on the subject, written starting in the 1950s-1960s, particularly Middleton's book. Middleton is, however, precise in his distinction between statistical and probabilistic quantities. But, over time, the language has become less precise. Today, the terms *signal* and *time series* are often used interchangeably by more broadly educated practitioners, with some preference given to *time series* by statisticians and preference given to *signals* by electrical engineers. The primary difference between time-series analysis and signal processing is that, prior to the communications technology revolution, the term signal was not yet being used for essentially any time-record of data. In communication theory, the stochastic process model of signals was adopted because a key concept was to design inference-making algorithms that optimized expected performance (minimized *expected cost*, which is the definition of Bayesian Risk). That is, performance was to be optimized over the ensemble of all sample paths of a stochastic process model of a type of signal of interest. For example, in telecommunications, the Wiener filter—according to modern theory—was the solution to minimum-mean-squared-error estimation of a transmitted signal, given a corrupted version of that signal obtained from a remote receiver. Thus, the statistical averaging of interest, performed by the expectation operation, was performed for example over all speech to be telecommunicated (referring back to the early days of Bell Telephone Laboratories), as well as all noise corrupting the transmitted signal. This included all speaker physiologies, all languages, and all accents. Standardized fixed ensemble-statistics computed empirically and expected values were used for designing channel filters and equalizers, which themselves were fixed or manually adjustable. But, as technology progressed, fixed optimum solutions began to be replaced with adaptive solutions that automatically optimized performance for each and every single signal. This required working with statistics obtained from time averaging single signals, not ensemble averaging multiple signals. This gave impetus to preferring ergodic stochastic process models for signals because then solutions implemented with algorithms that computed and used time-average statistics gave good approximations to the ensemble-averages dealt with in the mathematical models used for deriving the algorithms, and this rendered the stochastic process theory, in which electrical engineers had been indoctrinated, adequate for these. But despite ergodic theory, users did not know how to test their mathematical signal models for strong ergodicity. Birkhoff's ergodic theorem provided the ergodicity condition only in terms of the abstract mathematical probability measure defined (generically specified) in terms of a function of arbitrary subsets in a sigma field—the mathematical sample space—which also was defined (generically specified) in terms of sample paths often having no explicit description, e.g., interfering signals known only by their power spectral densities. So, the ergodicity condition was rarely able to be tested. Empirical data was of no use for this purpose because the condition involves only the abstract probability measure; it's a property of the mathematical model, not the empirical data. Practitioners often just evoked the Ergodic Hypothesis and typically left it untested. This is discussed early on by Middleton and remained the status quo up to and including today. But, once ergodicity was invoked, the stochastic process model was, in principle, no longer the most appropriate model, as explained in this paper and its references. With time-averages of primary concern, ensemble averages became, in principle but often unknowingly, irrelevant, and the abstraction of stochastic processes became unnecessary and nothing more than a distraction—something not recognized by most users. Although Middleton uses time averages, especially for calculating autocorrelation functions and associated quantities, before he takes the expected value, he does not appear to comment on the broader concept of FOT-probability.

Although 35 years have passed since a comprehensive development of an alternative probability theory for random signals that is based entirely on time averages was published in textbook form [4], this alternative theory has been largely ignored by all but a small minority of users of stochastic processes. For instructors of courses on statistical signal processing, teaching this alternative requires an introductory textbook, since the only textbook available [4] is written for advanced students. Similarly, a 2nd book (not a textbook with exercises) treating this alternative theory that appeared just two years ago is written for experts or at least mathematically mature readers. This stagnation in statistical signal processing pedagogy in universities occurred even though this simpler more transparent theory was proven in [4] to be analogous and actually *operationally equivalent* in many ways to the probability theory based on abstract and, one might even say, mysterious ergodic stochastic process models and, with regard to calculations, *yields the same results* in many cases (when Axiom VI is unnecessary). It is hoped that the pedagogical approach taken in this paper, whereby alternative stochastic process models are introduced as a conceptual transition from Kolmogorov's abstract stochastic process to concrete FOT-Probability models for single signals will spark interest in universities in developing new introductory courses based on the time-average theory of signals. Some of the many practical advantages of doing so are discussed in this article.

To be especially clear at the outset about limitations of FOT-Probability Theory, the particularly important area of statistical inference and decision-making on the basis of time-series observations is briefly discussed. Generally speaking, FOT-Probability models are well matched to what might be loosely called *non-parametric inference and decision*, for which no use is made of assumed functional forms of Cumulative Distribution Functions (CDFs) of the data with or without known, unknown, or random parameters of the functional form; the only CDF used is that measured from the observed time-series data. The complementary area of statistical inference and decision-making denoted with the adjective *parametric* partitions into two general types, one of which is accommodated by FOT-Probability models and the other of which is not.

The type of parametric statistical inference and decision making that is not accommodated by FOT-probability theory is that which is based on non-ergodic stochastic process models and some ergodic models for which probability functions, including CDFs or possibly just some moments, for the data conditioned on knowledge of some model-parameter values and/or hypotheses are needed but cannot be measured or calculated from a model for the observed data. Such cases can arise in Maximum-Likelihood Methods and Bayesian Minimum-Risk Methods of inference and decision making. If such parameters are modeled as random variables, the data must be considered to have arisen from a non-ergodic process since observation of one record of data cannot be used to learn about the influence of other values of the parameters that did not occur in the record of data. For example, if received data consists of signal plus noise under one hypothesis and noise only under an alternative hypothesis, the stochastic process model for the data that is not conditioned on a specific hypothesis cannot be ergodic.

In contrast to these parametric methods based on non-ergodic models, there is a type of parametric inference and decision making that is based on formulaic data models (sample-path models) in which the values of some parameters are unknown but are not treated as random variables. These are stochastic process models that are known only partially. For such models, one can in principle use the expectation operation to mathematically calculate the dependence of theoretical probability functions, such as moments, on the unknown parameters and determine equations interrelating multiple instances of these functions (different moments); these equations generally involve the unknown parameters. The approach consists of solving these equations, when possible, for the unknown parameters and then substituting empirical measurements of moments in place of the expected-value moments. This is called the *Method of Moments* for inferring parameter values.

Popular sample-path models used in the Method of Moments are autoregressive (AR), moving average (MA), and ARMA models and their periodic and poly-periodic generalizations. All such parametric methods are accommodated by the theory of FOT-moments associated with FOT-probabilities, for which the expected values in the Method of Moments used to derive from the data model interrelating equations are replaced with limits of time averages, and the empirical counterparts that were used to replace expected values in the solution are still finite-time averages, and they are now used to replace the limits of time averages. A survey of FOT parametric statistical spectral analysis is available in [4]; see also [8], [16], [17].

3. Results

3.1 Kolmogorov's model of a stochastic Process

We are interested here in discussing alternatives to both the discrete- and continuous-time versions of Kolmogorov's 1933 definition [8] of a stochastic process consisting of a sample space (the set of all sample paths, or signal realizations), a sigma field of subsets (events) in the sample space with a sigma algebra, and a probability measure on the event sets. These "sigma" requirements, meaning "convergence requirements for countably infinitely many operations", derive from Kolmogorov's Axiom VI in his definition of a stochastic process. In practice, the specification of a particular probability measure is rarely carried out because this is a difficult mathematical challenge for which there is no recipe. Sometimes practitioners will specify some lower order CDFs or Probability Density Functions (PDFs) as a half-hearted attempt. In the special case of a Gaussian process, the specification of the 2^{nd}-order CDF or PDF is all that is needed to derive from it all orders of CDFs and PDFs. Once all orders are specified, one can invoke the Kolmogorov Extension Theorem to conclude that the measure for the sample space has been effectively, if not explicitly, specified.

Because the probability measure for a stochastic process is rarely specified in practice, Axiom VI can only rarely be tested. Consequently, it is common practice to simply assume Axiom VI is satisfied by the selected model and proceed to use the consequences of that axiom in performing calculations involving infinite sums—not a particularly justifiable approach.

In other cases, practitioners will construct a formulaic model of a stochastic process as some combination of specified deterministic functions and some random variables. For example, essentially all digital communications signal models are specified in this manner. This typically provides no insight into the probability measure for the process but does often enable the practitioner to calculate some moments and/or cumulants and, much less frequently, some CDFs or PDFs. In a number of cases for which statistical inference using the stochastic process model is of interest, it suffices to calculate only the PDF for the observed data, conditioned on knowledge of the random variables in the model that are to be estimated, or conditioned on hypotheses to be tested. This can be adequate for deriving maximum-likelihood inference rules and in some cases minimum-Bayes-Risk inference rules.

In summary, it is a rare occasion when Kolmogorov's model of a stochastic process is able to be specified and used for time-series analysis, aka statistical signal processing. A particularly egregious consequence of this common practice is having to assume an adopted and possibly only partially specified model is strongly ergodic. This assumption—when valid—enables one to accurately approximate expected values calculated from the model using time averages on sufficiently long finite segments of a single realization of the signal being modeled. Without actually knowing that the model used for calculating expected values is ergodic, such time averages may or may not be accurate approximations. In the frequently encountered cases for which the expected values to be approximated are not calculated, the practitioner has no idea of whether or not the calculated time averages are useful approximations.

The above less-than-desirable situation concerning the use of Kolmogorov's stochastic process model has been tolerated for nearly a century now. Evidently, we've "gotten by" despite the unsavory facts summarized above. Nevertheless, there do exist alternative approaches to modeling signals for purposes of statistical inference and analysis. The purpose of this paper is to present such a model—the FOT-Probability model of a single signal—and explain how it relates to Kolmogorov's model and how much easier it is to use in practice in a more justifiable manner for applications in statistical signal processing. It should however be mentioned here that the FOT-Probability model can be used for statistical inference and decision-making involving likelihood functions only when such likelihood functions can be measured or calculated as conditional FOT-PDFs. This is further discussed in Section 4.

For the purpose at hand, let $T_t(A)$ denote the time-translation set-operator that shifts, by any real number $t \in R$, typically representing time, all sample paths in an event set A, and let $T_n(B)$ denote the discrete-time counterpart for any integer $n \in Z$. Following are the two ergodic theorems that are assumed to apply in many applications:

Birkhoff's Ergodic Theorem for Discrete Time (BET--DT)

> Consider a discrete-time Kolmogorov stochastic process with integer-valued time, satisfying Kolmogorov's 6 defining axioms [9], for which all event sets E that are translation-invariant, $T_n\{E\} = E$ for all integers n, have probabilities of either $P(E) = 0$ or $P(E) = 1$. By Birkhoff's 1931

Ergodic Theorem [12], this stochastic process is *ergodic w.p.1*, and is also referred to as *strongly ergodic.* Birkhoff's *ergodicity condition* here is not only sufficient but is also necessary for discrete-time-averages of functions of the stochastic process to converge to the corresponding expected values, as the averaging time approaches infinity.

Birkhoff's Ergodic Theorem for Continuous Time (BET--CT)

Consider a continuous-time Kolmogorov stochastic process, satisfying Kolmogorov's 6 defining axioms [9], for which all event sets E that are translation-invariant, $T_t\{E\} = E$ for all real t, have probabilities of either $P(E) = 0$ or $P(E) = 1$. By Birkhoff's 1931 Ergodic Theorem [9], extended from discrete- to continuous-time (e.g., page 1 of [13]), this stochastic process is *ergodic w.p.1*, and is also referred to as *strongly ergodic.* Birkhoff's *ergodicity condition* here is not only sufficient but is also necessary for continuous-time-averages of functions of the stochastic process to converge to the corresponding expected values as the averaging time approaches infinity.

These theorems require an additional axiom, here labeled *Axiom VII*, or they require a proof of a proposition. Without this Axiom VII or a proof, these theorems are not applicable in the way they have been applied for many years. This needed axiom or proof guarantees that the limits of the time averages of interest in practice exist. If they do exist, then the relevant ergodic theorem establishes that they equal w.p.1 the corresponding expected values. For discrete time, this proposition has been proved at least in some cases such as for finite-alphabet processes. As per my knowledge, it may or may not have been proved for continuous time. The propositions can be stated as follows: For a Kolmogorov discrete-time (continuous-time) process, the samples paths of well-behaved functions of the process are relatively measurable, as defined below.

One example of a sufficient condition for existence of the continuous-time average, which has been assumed in the early work on ergodic theorems, like Birkhoff's work (cf. [1]) is that the function of time is any well-behaved function of the positions of the particles of a dynamical system described by differential equations for which the sum of kinetic energies of all the particles in the system is time invariant. Unfortunately, this is typically not an appropriate model for the manmade signals used in communication systems.

3.2 The Measure Theory of FOT-Probability

The material in this subsection is taken from [7], also cf. [8, Chap. 2]. Let us consider the set $A \in \mathcal{B}_{\mathbb{R}}$, where $\mathcal{B}_{\mathbb{R}}$ is the σ-field of the Borel subsets and let μ be the Lebesgue measure on the real line R . The *relative measure* of A is defined by Kac and Steinhaus [6] as follows

$$\mu_R(A) \triangleq \lim_{T\to\infty} \frac{1}{T} \mu\left(A \cap \left[t_0 - T/2, t_0 + T/2\right]\right)$$

provided that the limit exists. In such a case, the limit does not depend on t_0 and the set A is said to be *relatively measurable* (RM).

Let $x(t)$ be a Lebesgue measurable function on the real line. The function $x(t)$ is said to be relatively measurable [6] if the set $\{t \in R : x(t) \le \xi\}$ is RM for every $\xi \in R - N_0$, where N_0 is at most a countable set of points. Each RM function $x(t)$ generates the function

$$F_x(\xi) \triangleq \mu_R(\{t \in R : x(t) \le \xi\})$$
$$= \lim_{T\to\infty} \frac{1}{T} \mu\left(\left\{t \in \left[t_0 - T/2, t_0 + T/2\right] : x(t) \le \xi\right\}\right)$$
$$= \lim_{T\to\infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} \mathrm{u}(\xi - x(t))\,\mathrm{d}t$$

at all points ξ where the limit exists. In this equation, $\mathrm{u}(\xi)$ denotes the unit step function: $\mathrm{u}(\xi)=1$ for $\xi \ge 0$ and $\mathrm{u}(\xi)=0$ for $\xi < 0$.

The function $F_x(\xi)$ has all the properties of a valid cumulative distribution function (CDF), except for the right-continuity property (at points of discontinuity). It represents the *fraction-of-time* (FOT) that the function $x(t)$ is below the threshold ξ, as illustrated in Fig. 1. For this reason, $F_x(\xi)$ is referred to as *the FOT-distribution* of the function $x(t)$.

Since the relative measure of every finite set is zero, the relative measure of every finite-energy or transient function $x(t)$ has the trivial distribution function $F_x(\xi) = \mathrm{u}(\xi)$. Only finite-average-power or persistent functions, such as almost periodic functions, can have a non-trivial FOT-distribution.

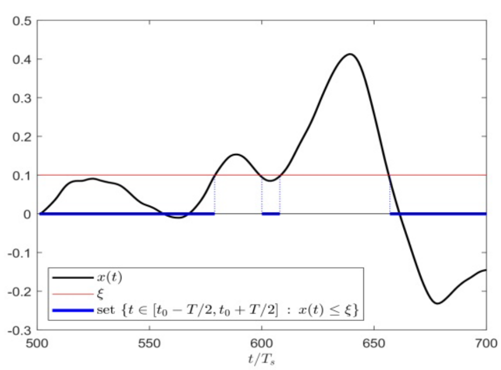

Fig. 1 The measure of the set $\left\{t \in \left[t_0 - T/2, t_0 + T/2\right] : x(t) \le \xi\right\}$ *(blue thick line) normalized by the total time* T *is the fraction of time that the function* $x(t)$ *is below the threshold* ξ *as* t *ranges over* $\left[t_0 - T/2, t_0 + T/2\right]$.

If $x(t)$ is a relatively measurable and not necessarily bounded persistent function and $g(\cdot)$ is a well-behaved function, then the following Fundamental Theorem of Time Average [4] can be verified [6, Theorem 3.2]

$$\lim_{T\to\infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} g(x(t))\,\mathrm{d}t = \int_R g(\xi)\,\mathrm{d}F_x(\xi)$$

where the first integral in the left member is in the Lebesgue sense and does not depend on t_0, and the integral in the right member is in the Riemann-Stieltjes sense. When $F_x(\xi)$ is differentiable, its derivative, denoted by $f_x(\xi)$, is the probability density function, and $dF_x(\xi)$ can be replaced in the right member with $f_x(\xi)\mathrm{d}x$, in which case the integral is in the Riemann sense.

From this theorem, it follows that *the infinite-time average is the expectation operator for the FOT-distribution* $F_x(\xi)$ and for every bounded $x(t)$ we have

$$\langle x(t)\rangle_t \equiv \lim_{T\to\infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} x(t)\,\mathrm{d}t = \int_R \xi\,\mathrm{d}F_x(\xi)$$

The analogy between FOT-probability and Kolmogorov probability [4], [19] is evident.

For a 1st-order strict-sense stationary process $X(t)$ with distribution $F_X(\xi) \triangleq P[X(t) \le \xi]$, the stochastic counterpart of the above time-average definition of the distribution is

$$F_X(\xi) = \mathrm{E}\{\mathrm{u}(\xi - X(t))\}$$

where $\mathrm{E}\{\cdot\}$ is the expected value operation, which equals the limit ensemble average operation, and which replaces the time average operation used in the FOT-probability approach. Similarly, the Kolmogorov counterpart of the *Fundamental Theorem of Time Average* is the following *Fundamental Theorem of Expectation*

$$\mathrm{E}\{g(X(t))\} = \int_R g(\xi)\,\mathrm{d}F_X(\xi)\,.$$

A necessary and sufficient condition for the relative measurability of a function is not known. However, if $x(t)$ is a bounded function, the existence of the time average

$$\lim_{T\to\infty}\frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} x^p(t)\,\mathrm{d}t .$$

for every positive integer p is a necessary condition for the relative measurability of $x(t)$. In addition, it follows from the Fundamental Theorem of Time Average that, if $x(t)$ is continuous and bounded and the left-hand side of the equation

$$\lim_{T\to\infty}\frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} x^p(t)\,\mathrm{d}t = \int_R \xi^p\,\mathrm{d}F_x(\xi)$$

exists for every positive integer p, then $x(t)$ is relatively measurable, and the above equation is valid.

As a final remark, it is noted that the absence of right-continuity of the FOT-distribution is not important in applications where integrals in $\mathrm{d}F_x(\xi)$ are of interest. For stochastic probability the right-continuity of the distribution is a consequence of the assumed σ-additivity of the probability measure P.

The preceding theory has a completely analogous discrete-time counterpart, which can be obtained by simply replacing integrals over continuous time with sums over discrete time [8, Chap. 2]. The same terminology is used. For example, the relative measure of a finite set A is defined by

$$\mu_R(A) \triangleq \lim_{N\to\infty}\frac{1}{2N+1}\#\left(A\cap\left[n_0-N/,n_0+N\right]\right)$$

where $\#(A)$ is the counting measure of the finite set A, which equals the number of elements in A.

We can now proceed with the definition of the stationary FOT-stochastic process. As above, $x(t)$ represents a persistent relatively measurable real-valued function of time defined over the entire real line and x_n represents a persistent relatively measurable real-valued sequence indexed by discrete time over the entire set of integers.

Multiple functions are said to be *Jointly Relatively Measurable* if they each are relatively measurable, meaning there FOT-CDFs exist, and their joint FOT-CDFs exist.

3.3 Definition of *Stationary FOT-Stochastic Process*

Axiom 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, x, subject to the constraint that replications are disallowed (no two sample paths can be identical):

$$\Omega_d = \{\{x_{n-\omega}; n\in Z\}; \omega\in Z\},$$
$$\Omega_c = \{\{x(t-\omega); t\in R\}; \omega\in R\}$$

Axiom 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.

Axiom 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 $\{t_1, t_2, t_3, ..., t_m\}$ or m integer-valued time points $\{n_1, n_2, n_3, ..., n_m\}$, respectively, of the *Stationary FOT-Stochastic Process* $x(t)$ or x_n is the relative measure of the event set

$$E_m^c \triangleq \{\omega \in R; f[x](t_1-\omega) \le \xi_1, \\ f[x](t_2-\omega) \le \xi_2, ..., f[x](t_m-\omega) \le \xi_m\}$$

or

$$E_m^d \triangleq \{\omega \in Z; f[x]_{n_1-\omega} \le \xi_1, \\ f[x]_{n_2-\omega} \le \xi_2, ..., f[x]_{n_m-\omega} \le \xi_m\}$$

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

It follows from Axiom S3 that the 1st order FOT-CDF for a continuous-time stationary FOT process is given explicitly by the formula

$$F_x(\xi) \triangleq \mu_R(\{t \in \mathbb{R} : x(t) \le \xi\})$$
$$= \lim_{U\to\infty} \frac{1}{U} \mu\left(\left\{t \in \left[t_0 - U/2, t_0 + U/2\right] : x(t) \le \xi\right\}\right)$$
$$= \lim_{U\to\infty} \frac{1}{U} \int_{t_0-U/2}^{t_0+U/2} \mathrm{u}(\xi - x(t))\mathrm{d}t$$

for all real ξ, and similarly for higher-order FOT-CDFs; and, for discrete-time, the FOT-CDF is given by

$$F_x(\xi) \triangleq \mu_R\left(\{n \in Z : x_n \le \xi\}\right)$$
$$= \lim_{N\to\infty} \frac{1}{2N+1} \#\left(\left\{n \in \left[n_0 - N, n_0 + N\right] : x_n \le \xi\right\}\right)$$
$$= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=n_0-N}^{n_0+N} \mathrm{u}\left(\xi - x_n\right)$$

As an example, for $m = 2$,we have the 2nd order FOT-CDF

$$F_x(\xi_1, \xi_2) \triangleq \mu_R(\{t \in R : x(t+t_1) \le \xi_1, \\ x(t+t_2) \le \xi_2\})$$
$$= \lim_{U\to\infty} \frac{1}{U} \mu(\{t \in R : x(t+t_1) \le \xi_1,\ x(t+t_2) \le \xi_2\})$$
$$= \lim_{U\to\infty} \frac{1}{U} \int_{t_o-U/2}^{t_o+U/2} \mathrm{u}\left(\xi_1 - x(t+t_1)\right) \mathrm{u}\left(\xi_2 - x(t+t_2)\right) \mathrm{d}t$$

for all real ξ. Note: The constraint in Axiom S1 that disallows replications in the sample space also disallows constant signals, which are a degenerate case of stationary signals. A viable alternative is to remove this constraint.

The probability of the entire sample space of the *Stationary FOT-Stochastic Process* is equal to 1, meaning every experimental outcome is one of the members of the sample space. That is, for a discrete sample space Ω_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, $\{x(t-\omega) : t \in R\}$ for any particular $\omega \in R$ or $\{x_{n-\omega} : n \in Z\}$ for any particular $\omega \in Z$, 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 variables in the formula. So, given a specification of one sample path, we have a specification of the entire sample space. Here are some examples that are commonly encountered in communications.

Example 1: Binary Amplitude Modulated Pulse Train Signal

$$x_1(t) = \sum_{k=-\infty}^{+\infty} a_k p_1(t - kT_1)$$

where $\{a_k\}$ is an FOT-stationary sequence of ± 1 values having equal FOT-probabilities and some specified continuous FOT power spectral density function and $p_1(t)$ is an absolutely integrable pulse of essentially arbitrary shape

Example 2: Amplitude Modulated Sine Wave Carrier Signal

$$x_2(t) = a_2(t)\cos(2\pi f_2 t + \theta_2)$$

where $a_2(t)$ is an FOT stationary Gaussian signal with some specified continuous FOT power spectral density function

Example 3: Amplitude-Shift Keyed Sine Wave Carrier Signal

$$x_3(t) = \sum_{k=-\infty}^{+\infty} a_k p_3(t - kT_3)\cos(2\pi f_3 t + \theta_3)$$

where $\{a_k\}$ is an FOT-stationary sequence of ± 1 values having equal FOT-probabilities and some specified continuous FOT power spectral density function and $p_3(t)$ is an absolutely integrable pulse of essentially arbitrary shape

Example 4: Phase Modulated Sine Wave Carrier Signal

$$x_4(t) = a_4 \cos(2\pi f_4 t + \theta_4(t))$$

where a_4 is a constant, $\theta_4(t)$ is an FOT-stationary signal with uniform FOT-PDF on $[0, 2\pi]$ and some specified FOT power spectral density function

Example 5: Multiplexed Signal with two statistically independent components

$$x_5(t) = x_2(t) + x_4(t)$$

There are numerous examples of calculations of FOT probabilistic parameters for formulaic specifications like those in the above examples; the first extensive catalog appeared in the book [4] and this was recently supplemented with additional examples in the book [8]. The great majority of these are calculations of cyclic autocorrelations and cyclic spectra (spectral correlation functions), but there are also some examples of calculations of higher-order moments and cumulants, both temporal and spectral types, cf. [23]. Calculations of cumulative FOT-probability distribution functions are less common. The reason is undoubtedly a result of the effort required. It is more practically feasible to use computer simulations to numerically evaluate FOT-CDFs, and likely there are results available in the literature.

Stationary FOT Ergodic Theorem:

a) 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.
b) 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 defining Axioms. 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 b) 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 b) addresses only finite averages and, like Item a), states a fact that is obvious from the construction of the sample space.

Relation to Wold's Isomorphism

Wold introduced an isomorphism in 1948 [14], 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 Axiom 1 in the definition of a Stationary FOT-Stochastic Process. By complementing this sample space with an FOT-Probability measure satisfying Axioms S2 and S3, we obtain a Stationary FOT-Stochastic Process.

3.4 Comparison of Kolmogorov and FOT-stochastic Process Models

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 for 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, this 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.

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 described earlier in this paper and further in the Results section, 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 not sigma-additive, and the corresponding FOT-expectation operation is not 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 signal.

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 need 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 2-dim array of 0's and 1's is such that every possible sequence of 1's and 0'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 t. The measure tells us the limit, as the number of randomly selected values of ω approaches infinity, of the relative frequency of 0's and 1's that will occur as outcomes. This probability measure is like a magic hand that guides the selection of experimental outcomes so that 1's are selected in 0.7% of the experiments and 0's are selected in 0.3% of the outcomes. And, for example, the pair of adjacent outcomes of 0 followed by 1 are selected in (0.7)(0.3) = 0.21% 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 should be clarified here that the strong law of large numbers [9] establishes that averages over ensembles of random samples converge to expected values w.p.1 *not because of replication* in the sample space, 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. 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 [11, Sec. 1.3] includes 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 (Axiom 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 early literature (cf. references in [22]).

The above discussion illustrates that the details and level of abstraction of the Kolmogorov stochastic process model are often not observed in applied theoretical work in statistical signal processing. Consequently, there is little pragmatic justification for continuing to hang on to the baggage (abstraction) that comes with this standard model, when we have the much simpler and more concrete alternative, the FOT-probability model for single signals.

3.5 Definition of *Cyclostationary FOT-Stochastic Process*

Axiom CS1: The *Sample Space* of the *Cyclostationary FOT-Stochastic Process with Period T* is comprised of all the time translates, by integer multiples of the period, of a single relatively measurable discrete- or continuous-time sample path, x, subject to the constraint that replications are disallowed (no two sample paths can be identical):

$$\Omega_d = \{\{x_{n-\omega T}; n \in Z\}; \omega \in Z\},$$
$$\Omega_c = \{\{x(t-\omega T); t \in R\}; \omega \in Z\}$$

The period T can be any real number for continuous-time processes but must be an integer for discrete-time processes.

Axiom CS2: 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.

Axiom CS3: 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 $\{t_1, t_2, t_3, ..., t_m\}$ or m integer-valued time points $\{n_1, n_2, n_3, ..., n_m\}$, of the *Cyclostationary FOT-Stochastic Process* $x(t)$ or x_n, with *Period* T, is the relative measure of the event set

$$E_m^c \triangleq \{\omega \in Z; f[x](t_1 - \omega T) \le \xi_1, f[x](t_2 - \omega T) \le \xi_2, ..., f[x](t_m - \omega T) \le \xi_m\}$$

or

$$E_m^d \triangleq \{\omega \in Z; f[x]_{n_1-\omega} \le \xi_1, f[x]_{n_2-\omega} \le \xi_2, ..., f[x]_{n_m-\omega} \le \xi_m\}$$

for all real-valued m-tuples $\{\xi_1, \xi_2, \xi_3, ..., \xi_m\}$, and all these FOT-CDFs are periodic functions of time: If $\{t_1, t_2, t_3, ..., t_m\}$ is replaced with $\{t_1+T, t_2+T, t_3+T, ..., t_m+T\}$

or, if $\{n_1, n_2, n_3, ..., n_m\}$ is replaced with $\{n_1 + T, n_2 + T, n_3 + T, ..., n_m + T\}$, the FOT-CDF remains unchanged.

It follows from Axiom CS3 that the first-order FOT-CDF for a continuous-time cyclostationary FOT process is given explicitly by the formula

$$
\begin{aligned}
F_{x,T}(\xi,t) &\triangleq \mu_R(\{n \in Z : x(t-nT) \le \xi\}) \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \#(\{n \in [n_0 - N, n_0 + N] : \\
&\qquad\qquad x(t-nT) \le \xi\}) \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n = n_0 - N}^{n_0+N} \mathrm{u}\left(\xi - x(t-nT)\right)
\end{aligned}
$$

for all real t and ξ, and similarly for higher-order FOT-CDFs; and the first order FOT-CDF for a discrete-time FOT process is given explicitly by the formula

$$
\begin{aligned}
F_{x,T}(\xi,n) &\triangleq \mu_R(\{n \in Z : x_{k-nT} \le \xi\}) \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \#(\{n \in [n_0 - N, n_0 + N] : x_{k-nT} \le \xi\}) \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n = n_0 - N}^{n_0+N} \mathrm{u}\left(\xi - x_{k-nT}\right)
\end{aligned}
$$

for all real ξ and all integer n. In contrast to the periodicity of these FOT-CDFs, the FOT-CDFs for a stationary FOT-stochastic process remain unchanged for *all* real-valued or integer-valued T. They are periodic with every period and are therefore time-invariant.

Note: The constraint in Axiom CS1 that disallows replications in the sample space also disallows periodic signals, which are a degenerate case of cyclostationary signals. A viable alternative is to remove this constraint.

For this FOT-stochastic process, any one of the translates, $\{x(t-\omega T) : t \in R\}$ for any particular $\omega \in Z$ or $\{x_{n-\omega T} : n \in Z\}$ for any particular $\omega \in Z$, can be taken as the *Sample Space Generator*. Observe that, whereas the sample space for the stationary FOT process is uncountably infinite for continuous time, it is only countable infinite for the continuous-time cyclostationary FOT process.

Although not immediately obvious, a single sample-space generator (a single signal) can, in general, generate a stationary FOT process or a cyclostationary FOT process with any one of multiple incommensurate periods. If the single signal exhibits no cyclostationarity, all the FOT-CDFs will be time-invariant and identical. If the single signal exhibits only one period, then its cyclostationary FOT-CDF will be periodic, not time-invariant and it will therefore be distinct from the stationary FOT-CDF. And if the single signal exhibits two incommensurate periods, the sample space generator can generate a time invariant FOT-CDF and two distinct periodic FOT-CDFs, by using different sets of translation amounts. And so on. For the five example signal models specified above, we have the following results for the distinct FOT-CDFs that can be produced from each signal.

Example 1: $x_1(t)$ has stationary FOT-CDF and one cyclostationary FOT-CDF with period $T = T_1$

Example 2: $x_2(t)$ has stationary FOT-CDF and one cyclostationary FOT-CDF with period $T = 1/2f_2$

Example 3: $x_3(t)$ has stationary FOT-CDF and multiple cyclostationary FOT-CDFs with periods $T^{(j)} = 1/(2f_3 + j/T_3)$ for possibly all integers j, assuming that f_3 and $1/T_3$ are incommensurate

Example 4: $x_4(t)$ has stationary FOT-CDF and one cyclostationary FOT-CDF with period $T = 1/2f_4$

Example 5: $x_5(t)$ has stationary FOT-CDF and multiple cyclostationary FOT-CDFs with periods $T^{(j)} = 1/(nf_2 + mf_3)$ for possibly all pairs of integers (n, m) (except those for which $(n_2, m_2) = (kn_1, km_1)$ for any integer k) if f_2 and f_3 are incommensurate; otherwise just one cyclostationary FOT-CDF with period $T = 1/nf_2 = 1/mf_3$ for the smallest pair of integers n, m for which this equality holds.

Cyclostationary FOT Cycloergodic Theorem:

a) Every Cyclostationary FOT-Stochastic Process is *Strongly Cycloergodic*, by construction, meaning the infinite time averages, with cyclostationarity period T, of relatively measurable functions of the process exist and are independent of the particular sample paths selected and are equal to the time-periodic expected values of those functions obtained using the FOT-CDF or FOT-PDF.
b) Every Finite-Ensemble Average of every function of a Cyclostationary FOT-Stochastic Process is identical to a Finite-Time Periodic Average of that function.

The validity of this theorem follows directly from the defining Axioms. 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 b) 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, 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 b) addresses only finite averages and, like Item a), states a fact that is obvious from the construction of the sample space.

3.6 The FOT-Probability Model for Almost Cyclostationary Processes

The approach taken here consists of starting with the *concept* of a standard CDF defined by the expected value of an indicator function as normally done for Kolmogorov processes, and then using the well-known Fourier decomposition of an almost periodic function into a sum of sinusoidal components, with one component for each of all the sine-wave frequencies exhibited by the process. However, there is no need to assume a Kolmogorov stochastic process model, in particular; this would introduce major unnecessary abstraction and complexity. We consider the CDF

$$\begin{aligned} F_X(\xi, t) &\triangleq E\{\mathrm{u}(\xi - X(t))\} \\ &= \sum_{\alpha} F_X^{\alpha}(\xi) \exp[i2\pi\alpha t] + \Theta_X(t) \end{aligned}$$

where the expectation operation $E\{.\}$ is nothing more than a notion and where

$$\begin{aligned} F_X^{\alpha}(\xi) &= \lim_{U\to\infty} \frac{1}{U} \int_{-U/2}^{U/2} F_X(\xi, t) \exp[-i2\pi\alpha t] dt \\ &= \lim_{U\to\infty} \frac{1}{U} \int_{-U/2}^{U/2} E\{\mathrm{u}(\xi - X(t))\} \exp[-i2\pi\alpha t] dt \end{aligned}$$

and $\Theta_X(t)$ represents any non-cyclic non-stationarity that might be present in the process model, meaning

$$\lim_{U\to\infty} \frac{1}{U} \int_{-U/2}^{U/2} \Theta_X(t) \exp[-i2\pi\alpha t] dt = 0$$

for all α. For our purposes, we assume $\Theta_X(t) \equiv 0$. Then, as a final step, the unspecified expectation operation is removed to obtain the following definition in terms of a single sample path:

$$F_x^{\alpha}(\xi) = \lim_{U\to\infty}\frac{1}{U}\int_{-U/2}^{U/2} \mathrm{u}(\xi - x(t))\exp[-i2\pi\alpha t]dt .$$

The almost periodically time-varying CDF is then given by

$$F_x(\xi,t) \triangleq \sum_{\alpha} F_x^{\alpha}(\xi)\exp[i2\pi\alpha t] .$$

The reader must be careful to distinguish between capital X, which represents a stochastic process, and lower case x, which represents a single sample path, or simply a single signal, such as one specified formulaically, as in Examples 1 – 5.

It has been shown, in several distinct ways [4], [5], [8], that the above definition of an FOT-CDF is indeed a valid cumulative probability distribution function. The Fourier components $F_x^{\alpha}(\xi)$ of the FOT-CDF are called the *cyclic CDFs* [4], although they are not actually CDFs for $\alpha \neq 0$; they are a generalization to complex-valued distribution functions. But this doesn't matter if their combination in the formula for $F_x(\xi,t)$ is indeed a valid FOT-CDF. In fact, this is true for various choices of values of α to include in the Fourier series formula. This includes the choice of only $\alpha = 0$, and it includes the choice of all harmonics of $\alpha = 1/T$ for each incommensurate period T chosen, although the periods chosen are arbitrary. For every non-zero value of α chosen, its negative must also be chosen, but this is apparently not sufficient. It is conceivable that not all harmonics for each chosen period must be included. This would certainly be true if the cyclic CDFs for some of the harmonics were zero; but this may not be possible and is considered unlikely because of the infinity of step discontinuities in the function $\mathrm{u}(\xi - x(t))$. Research on this topic is ongoing. In any case, the above-listed allowable choices illustrate that a multiplicity of FOT-CDF models for a single signal is possible.

Even though the formula for calculating the cyclic FOT-probability models does not directly use actual FOT calculations—it uses the modified relative measure that includes sinusoidal weighting—it is identical to an actual FOT calculation, according to the *synchronized averaging identity* [4]:

$$\begin{aligned}
F_x(\xi,t) &\triangleq \sum_{\alpha} F_x^{\alpha}(\xi)\exp[i2\pi\alpha t] \\
&= F_x^{0}(\xi) \\
&\quad + \sum_{j\in Z}\left\{\sum_{k\in Z}\left(F_x^{k/T_j}(\xi)\exp[i2\pi(k/T_j)t] - F_x^{0}(\xi)\right)\right\} \\
&= F_x^{0}(\xi) + \sum_{j\in Z}\left\{F_{x,T_j}(\xi,t) - F_x^{0}(\xi)\right\}
\end{aligned}$$

where $F_x^{0}(\xi)$ (defined in Sec 3.3) and $F_{x,T_j}(\xi,t)$ (defined in Sec 3.5) are FOT-stationary and FOT-cyclostationary CDFs, respectively. In this formula for the FOT-CDF of an almost cyclostationary time series, $\{T_j\}$ represents the set of periods of the possibly countably infinite number of periodic components of the almost periodic FOT-CDF. When this set is finite, the time series is said to be *Poly-Cyclostationary*.

This completes the specification of an *Almost Cyclostationary FOT-Probability Model* for single signals. It is very flexible, involves no weird stochastic process abstractions like the *magic hand* referred to above, and is compatible with both the commonly used formulaic signal models and purely empirical data. That is, when all one has is empirical data—a single signal with possibly no information upon which to base a formulaic signal model, the Fourier components in the CDF formula can still be calculated directly from the data by using the above formula with the limit operation removed (as justified in subsequent sections). This still produces a valid CDF from the Fourier series formula. In this case, there is no formulaic signal model from which to identify the periods whose harmonics frequencies are to be used for α; however, there may be information from the physical source of the data that can be used to hypothesize cycle frequency values. Otherwise, an exhaustive search over all feasible values of α must be performed using a significance test of some sort on each calculated cyclic CDF.

3.7 Cycloergodicity for Multiple Incommensurate Periods

Many communications signals with sample paths specified formulaically exhibit cyclostationarity with multiple incommensurate periods (they are *poly-cyclostationary* or *almost cyclostationary*, but not purely cyclostationary or purely stationary) and, as shown by Boyles and Gardner in 1983 [15], they can be tested for what is here called *Sinusoidal Ergodicity* (SE). This means some such processes can exhibit the strong sinusoidal ergodic properties required to support the commonly assumed convergence of estimates of sinusoidal components (which are typically called *cyclic components*) of their almost-periodically time-varying probabilistic parameters, such as cyclic autocorrelations and cyclic spectral densities (also called spectral correlation functions). However, *these processes cannot be included in the traditional ergodic theory stemming from Birkhoff's work or its extension to the cycloergodic theory of cyclostationary processes of Boyes and Gardner.* This is mathematically proved in [15] and illustrated with the example of a Bernoulli process with a periodically time-varying probability of success having its period incommensurate with the sampling-time increment. What has essentially invariably been done since the introduction of almost cyclostationary processes in 1978 [26] is to specify such processes in a formulaic manner (e.g., Examples 3 and 5 above) and to then invoke a strong cycloergodic hypothesis, sometimes based on the demonstration of a much weaker form of cycloergodicity, such as cycloergodicity in the mean square sense. But we are now going to go beyond this.

The sample spaces for the cyclostationary FOT-stochastic processes reveal why *there cannot exist a single FOT-stochastic process with more-than-one incommensurate period*: A single sample space cannot consist of only translates of one period if it also consists of only translates of another incommensurate period. What one must therefore do with the FOT model introduced in Section 3.6 is to introduce a unique sample space for each and every incommensurate period of cyclostationarity of interest for a single record of data or a single formulaic model. However, this is just a conceptual aid. For operational purposes, all one needs is the formula for almost cyclostationary CDFs given in Section 3.6 and the method presented in Section 3.5 for calculating cyclostationary CDFs for each period. This calculation can be empirical, using a record of observed data, or it can be performed mathematically using a formulaic specification of the time series. This, in turn, provides insight into how to generalize Birkhoff's ergodic theorem to accommodate almost cyclo-stationary processes of the Kolmogorov type, as explained next.

But first, let us sum up the situation for formulaic FOT-Probability models for almost cyclostationary time series. The deterministic periodicity with multiple periods combined in a sample-path formula, such as those in Examples 1 – 5, with stationary FOT time-series components, provides the basis for constructing the CDFs or PDFs from FOT calculations using the time-series model. Nonlinear functions of a time series whose sample-path formula contains multiple periodicities contain in general not only harmonics not originally present, of the fundamental frequencies originally present, but also linear combinations with integer-valued coefficients, of all these harmonics. Consequently, in constructing the CDFs for such a time series, it must be assumed at the outset that the CDFs will contain sinusoidally time-varying components with all these various mixed frequencies.

How to Generalize Birkhoff's Ergodic Theorem for Continuous-Time Almost Cyclostationary Kolmogorov Stochastic Processes

The content of this section does not contribute to the primary objective of this article, but it does follow easily from the concepts introduced in the previous section and it does provide a genuine generalization of ergodic theory of stationary and cyclostationary processes to poly-cyclostationary and almost cyclostationary Kolmogorov stochastic processes. *Strong Cycloergodic theory of Kolmogorov stochastic Processes*, which extends and generalizes existing ergodic theory, is developed in (Boyles and Gardner 1983), where it is shown that sinusoidal and periodic components of time-varying probabilistic parameters can be consistently estimated w.p.1 from time averages on one sample path. It is also established that a strong theory of cycloergodicity inclusive enough to cover all applications of practical interest had, at that time, not yet be shown to exist. Moreover, it is shown that such a theory cannot presuppose the existence of a dominating stationary measure, as does the theory presented therein. Nevertheless, it would appear that it can be argued that because a continuous-time cyclostationary process can be characterized as a discrete-time vector-valued (or function-valued) stationary process, Birkhoff's Ergodic Theorem (Birkhoff 1931) for scalar-valued discrete-time stationary processes, if generalized to vector-valued processes, leads to a completely analogous cycloergodic theorem for continuous-time cyclostationary processes. The vector (or function), at any discrete time equal to an integer multiple of the period of cyclostationarity, consists of the infinite set of process values over the period between that discrete time and the previous discrete time.

Furthermore, it is shown in (Gray 2009, Chap. 7, and refs. therein) that Birkhoff's ergodic theorem has been extended from stationary to asymptotically-mean stationary (AMS) discrete-time processes. This extension guarantees the existence of consistent estimators for the discrete-time averages of time-varying probabilistic parameters, such as probability density functions. Because almost-cyclostationary (ACS) discrete-time processes are AMS, this extended theorem applies to discrete-time ACS processes (and the same might well be true for continuous-time ACS processes after discrete-time sampling) but it does not apply directly to estimation of the sinusoidal and periodic components of almost-periodically time-varying probabilistic parameters.

Nevertheless, (Gray 2009, Chap. 7) does discuss ergodicity of N-stationary discrete-time processes, which are N-dimensional vector-valued representations for discrete-time cyclostationary processes with period N. Furthermore, the discrete-time infinite-dimensional vector-valued process described above that represents a continuous-time scalar-valued process is AMS if that continuous-time process is ACS (which includes, as special cases, poly-cyclostationary, cyclostationary, and stationary processes).

Consequently, for any selected period of a continuous-time ACS process, one can form a discrete time vector-valued AMS process as explained above. Then the time average of a probabilistic parameter of this vector-valued process will equal the periodic component of probabilistic parameter of the original ACS process. In this way any periodic component for any real-valued period T of the almost periodically time-varying probabilistic parameters of the original scalar-valued continuous-time ACS process can be guaranteed to be consistently estimable by applying the proposed ergodic theorem to the infinite-dimensional vector-valued discrete-time AMS process.

It follows that the discrete-time AMS version of the Birkhoff ergodic theorem can be extended / generalized to accommodate cycloergodicity for continuous-time ACS processes by requiring that the ergodicity condition for discrete-time AMS processes be satisfied by the vector-valued representation for each and every period T of the continuous-time process. In addition, there appears to be a partially cycloergodic version of this proposed theorem that requires the ergodicity condition for some but not all periods be satisfied.

This leaves one class of ACS processes for which a cycloergodic theorem remains to be proposed, and this is the class of discrete-time processes having measures that possess non-zero sinusoidal components with sine-wave frequencies that are incommensurate with the time-sampling rate. Some such processes do indeed allow for consistent estimation of such sinusoidal components, but others do not. A necessary and sufficient condition for consistent estimation has apparently not yet been proposed but the Author expects one to be discovered by following ideas in the present paper

3.8 Purely Empirical FOT-Probability Models for Regular Cyclicity

We can obtain finite-data probability models by using the FOT-CDF formula in Section 3.6, but without taking the limit as the averaging time approaches infinity, and still get CDFs that are exactly constant (using only $\alpha = 0$) or periodic (using $\alpha = 0$ and $\alpha = kT \text{ and } \alpha = -kT$ for $k = 1,\ 2,\ ...,K$) or poly-periodic (using $\alpha = 0$ and $\alpha = kT \text{ and } \alpha = -kT$ for $k = 1,\ 2,\ ...,K$ and any finite set of incommensurate real-valued periods T). We consider here only finite numbers of cycle frequencies since calculation involving an infinite number cannot be purely empirical. However, it appears from recent unpublished work by Napolitano and coworkers that omission of cycle frequencies for which the cyclic components are not identically zero renders the formula for the CDF only approximate. Such approximations do not necessarily retain the characteristic properties of valid CDFs.

Nevertheless, it is expected that the approach with finite K can produce accurate approximations with sufficiently large but finite values of K. Moreover, by recognizing that the above CDF formula can be used for all time t, even though it is calculated from only a finite interval U of time, we see that these purely empirical models are not just approximately constant or periodic or poly-periodic, they are exactly so.

More generally, the program of calculation for any probabilistic parameters, such as joint moments, using a finite segment of data $x(t)$, is that everywhere the data occurs, in the infinite-interval formula for the parameter of interest [4], for some function of the data that is of interest, such as a lag product, the time support of that data is windowed to the finite observation interval, just like what is done in the conventional correlogram & cyclic correlogram, and periodogram & cyclic periodogram [4]. Then the time-invariant Fourier coefficient of the sinusoidal component, with frequency α, of the function of the time series over the finite observation window is extracted and multiplied by $\exp[i2\pi\alpha t]$ (with t extending over the reals) in the usual manner, but without the limit as integration time approaches infinity. These components when added together for all detected or selected cycle frequencies comprise an almost periodic function over all time and, when restricted to the finite time support of the function of the data, comprise an approximation to that function. The approximation is not a least-squares fit because the sinewave components are not mutually orthogonal except over the entire real line unless their frequencies are commensurate. It also does not equal the limit almost periodic component, but it would hypothetically converge to it as the observation time approaches infinity, provided that the function is relatively measurable. But the theory does not use the limit together with conditions for or assumptions of convergence of the probability of events. It simply uses the finite time statistics (approximate Fourier components) that are artificially extended over all time. These extracted almost periodic representations can be used just as they are used in the limit theory and can be calculated from either a finite-time record of $x(t)$ or an explicit mathematical model of $x(t)$. The data windowing used does not affect the theoretical equality of these two calculations provided that the data record is producible from the mathematical model, except for the difference between the values of the random elements in the mathematical model and the actual values of those elements in the record of data, such as the amplitude sequence in an amplitude modulated periodic pulse-train signal. The link here, which replaces the ergodic theorem, is the assumption that the single data record is indeed a segment of one translate of a single time series and that the functions of this time series that are of interest are relatively measurable. This then enables a standard type of argument that agreement between the two methods of calculation (finite-time and infinite-time averaging) can be made as close as desired by using a long-enough finite-segment of data [7].

All the usual tools still apply. For example, the proof of the central limit theorem for FOT-probability [24] is applicable to the theory for finite records by simply arguing that for any arbitrarily small error, epsilon, in equality between the limit quantity (Gaussian distribution) and the measured quantity, one can in principle chose a finite record length that is long enough to achieve an error size not exceeding epsilon.

There's nothing here of any technical sophistication. The novelty is in recognizing that finite-time FOT models that are precisely stationary or poly-cyclostationary can be constructed from a finite record of data, and these models can be used for all the usual probability calculations to within some finite precision determined by the length of the data segment and particular cycle frequencies used. The sensitivity of the precision to the numbers of harmonics of each fundamental frequency that are used increases as the degree of nonlinearity of the function of the data increases. A lag product, for example, has a low degree of nonlinearity, but the step discontinuities of the indicator function used to calculate CDFs results in a high degree of nonlinearity. Much research remains to be done on methods for calculating approximate poly-cyclostationary CDFs

In the Fourier-coefficient formulas for the function (of the data) of interest, the time-shifted finite segments of data will force the integrand to be zero outside of a subinterval defined by the intersection of the time-translated finite-segment support intervals and the integration interval. Assuming all time-shifts of interest are much smaller than the segment length, this approach is acceptable. But it will window the n-dim space of n time shifts. Assuming desired spectral resolution width in any spectral parameters (PSD, SCF, etc.) is larger than the reciprocal of the smallest value, $U - \max\left\{\left|t_i - t_j\right|\right\}$, for data-segment length U, the achieved spectral resolution can be acceptable. Ideally, we'd like this smallest value to be much larger than the coherence length of

$x(t)$ (here meant to be the time separation between time samples that is just large enough to result in negligible statistical dependence) to ensure statistical reliability.

A refinement that should moderately improve reliability and reduce bias is to truncate the integration interval involving time-shifts $\{t_i\}$ to the closest integer multiple of $1/\alpha$ that does not exceed $U - \max\{|t_i - t_j|\}$.

3.9 Purely Empirical FOT-Probability Models for Irregular Cyclicity

Cyclicity is ubiquitous in scientific data, but for many if not most natural sources of data, the cyclicity is irregular: the period of cyclic time-variation itself changes with time, slowly in some applications and rapidly in others. One approach to accommodating this is to restrict cyclostationary modeling to data segments that are short enough for the period to be treated as if it were constant. A more general and less restrictive approach is to hypothesize that the irregularity results from a time-warping of an otherwise regular cyclicity. This is true for some irregularly cyclic data sources and not true for others, such as rotating machine vibrations with time-varying rotational speed as explained in [16]. Fortunately, there is a middle ground of natural sources of data for which the irregular cyclicity—though irregularly fluctuating too rapidly to treat as regular—is due to time warping of otherwise regular cyclicity and the rate of variation of the warping function is slow enough to be tracked. A broadly applicable approach to doing this is introduced in [16] and is based on the concept of *property-restoral adaptation.*

Methodology and algorithms for such adaptation are presented therein for restoral of regular cyclicity. The adaptation process produces both a time-dewarped version of the original data, which is more nearly cyclostationary, and explicitly identifies the dewarping function. In some applications, identification of the warping function inherent in the data, by inverting the identified dewarping function, is the end goal for this time-series analysis; in other cases, further time-series analysis that exploits the restored cyclostationarity is the end goal. In this latter case, by preprocessing data that exhibits irregular cyclicity to restore cyclostationarity enables the user to go on to construct cyclostationary FOT-probability models. These models can be used directly for some applications and can be time-warped to obtain irregularly cyclic probability models. A generally applicable rule of thumb for predicting how well this methodology can perform is described in [16] in terms of a comparison between (1) what can be called the *coherence time* (or *statistical dependence time*) of the data or the *data memory length* and (2) the *constancy time* (reciprocal of some measure of the rate of time variation) of the warping function. Best performance is expected when (2) exceeds (1) by a factor much larger than unity. This is akin to the well-known concept of *local stationarity* but generalized to *local cyclostationarity* and also the more esoteric concept of *local ergodicity* generalized to *local cycloergodicity.* But fortunately, such abstractions are avoided when using FOT-probability models. Complementary work on property-restoral de-warping has been conducted in [25].

4. Discussion of Results

We have known for nearly a century that Birkhoff's Ergodic Theorem, extended from discrete-time to include continuous-time, provides a condition on the sample space and probability measure of Kolmogorov's generic stochastic process model that makes time-averages of measurements on (functions of) the process converge, with probability equal to 1 (w.p.1), to expected values of those measurements. And, we also have known all this time that Kolmogorov's Law of Large Numbers proves that ensemble averages converge to expected values w.p.1. However, practitioners using these results are generally unable to understand, with any level of intuition, why these equalities between fundamentally different entities are valid.

In contrast, the alternative and greatly simplified stochastic process models presented in this paper are transparent. It is obvious why time averages equal ensemble averages, because the sample space consists of all time-translated versions of a single signal, and it is obvious why these both equal expected values.

In applications where we are interested in only ergodic processes, there does not appear to be any pragmatic reason for adopting the complicated abstract Kolmogorov model instead of the simpler more concrete alternative model. In fact, once we've accepted the alternative model as sufficient for our purposes, we can take the next step of recognizing that this alternative model is identical to the entity comprised of a single signal and its Fraction-of-Time (FOT) Probabilities which are derived directly from this single signal. The conclusion is that sample spaces and stochastic processes are unnecessary unless non-ergodic models are the entities of interest.

This is a situation where a pragmatic person would ask "what's the point of teaching students of statistical signal processing about the strongly ergodic Kolmogorov stochastic process model as a tool for problem solving, with its unnecessary abstraction and its ergodic hypothesis which can almost never be tested in practice, when the model of a single time series (a persistent function of time), together with the concrete time-average operation is operationally equivalent? If we hold to the principle of scientific parsimony and we value mathematical elegance and we act logically and rationally, shouldn't we terminate this nearly-one-century-long practice immediately? It is relevant here that it has been said: *If elegance in science is just an attractive attribute, then elegance is not a necessary goal but simply something to be admired when it happens. However, if elegance is a requisite feature of good science, then the characteristics defining elegance deserve the same attention given to scientific rigor.*

To be sure the ramifications of what is stated above are understood by the reader, it is also stated explicitly here, and shown in [4] (see also [5] and [17, Chapt 1]) that the temporal-expectation (time-average) operation behaves just like the stochastic-expectation operation and produces all probabilistic quantities we are familiar with: cumulative probability distribution functions, probability density functions, moments, characteristic functions, cumulants, etc. For example, both operations obey a Fundamental Theorem of Expectation. It's just that, for temporal expectation, the term *probability* means Fraction of Time (FOT) of occurrence of an event at all sets of times, with specified time-separations, over the temporal lifetime of the time series, instead of fraction of repeated experiments (each producing a time-series) over which an event occurs at a particular set of times.

There are two exceptions to this equivalence, and they are the sigma linearity property of expectation and the relative measurability property of single time functions; these properties are simply dictated by the creators of these two theories: the first by Kolmogorov's Axiom VI and the second by the Kac-Steinhaus Axiom of Relative Measurability. Axiom VI may or may not be satisfied by a stochastic process model that some practitioner specifies. And relative measurability is not necessarily satisfied by all the time-series models practitioners may specify. For example, the samples paths of a strongly ergodic stochastic process are not necessarily relatively measurable; so, this property must be assumed for the strongly ergodic stochastic process (call it Axiom VII) for continuous time if the limits of time averages in the Birkhoff ergodic theorem are to exist. Although there's no question that sigma additivity of probability measures and sigma linearity of expectation are useful mathematically, users can rarely verify that the models they use actually exhibit these properties. Nice mathematical properties for both stochastic processes and single time functions come at a cost of restricted applicability. This is the nature of models, especially those involving infinity. It is not necessarily a basis for arguing the superiority or inferiority of the ensemble-average theory over the time-average theory. More in-depth analysis of this topic is provided in [1]. But it is important to mention here that just because the use of the relative measure (time-averaging operation) does not generally enable the user to interchange the limit in the time-averaging integral with the summation over a countable infinity of additive terms does not mean that one cannot proceed with the calculation. The interchange of operations must be executed before the limit is taken. In some cases, this is required only for the limit that defines the time average; in other cases, it may be required also for the limit that defines the infinite summation.

A comprehensive theory and methodology of FOT-Probability and statistical spectral analysis is presented in the 35-year-old book [4]. In addition, this book extends/generalizes the theory from stationary time series to cyclostationary, poly-cyclostationary, and almost cyclostationary time series, which provide higher fidelity models of many time series encountered in engineering and the sciences, as evidenced by the many new signal processing algorithms it has engendered over the last 35 years. The similar-vintage book [18] provides the theory of the stochastic-process counterparts of these extended/generalized time-series theories. A much more recent and more comprehensive book on both these stochastic-process and time-series models is also available [8] and is recommended. This latter book is encyclopedic and is the most scholarly treatment of cyclostationarity available.

For example, some continuous-time functions for which averages over discrete times exist may not be relatively measurable on the real line and therefore may not be averageable over all real time. This requires the addition of a 7^{th} axiom to Kolmogorov's stochastic process model to accommodate Birkhoff's ergodic theorem for continuous time averages. As another example, the Channel Coding Theorem of Information Theory cannot be based on FOT-probability because it is formulated in terms of a non-ergodic stochastic process: The stochastic-process output from any and every random channel except for a random time-delay, is non-ergodic, regardless of whether or not the channel input is ergodic. (The random-delay exception is not allowed for cycloergodicity.) For example, Middleton's classic models of non-Gaussian noise are non-ergodic, because these noise models depend on random time-invariant parameters such as the random number of noise sources seen by the receiver and their random locations relative to the receiver (see, for example, [19], and references therein).

As another example, the theories of maximum-likelihood parameter estimation and hypothesis testing are based on the likelihood function, which is the PDF of the observed data, conditioned on each specific hypothesis and/or hypothetical parameter value. Also, Bayesian minimum-risk parameter estimation and hypothesis testing inference rules can be expressed in terms of likelihood functions. Consequently, these theories and methodologies can only be based on FOT-Probability if conditional FOT-Probabilities and/or PDFs can be experimentally measured or mathematically calculated from mathematical sample-path models of the data. Frequently this can indeed be done as demonstrated with many examples in [4], [8], [17]. However, it cannot always be done.

The example of the channel coding theorem provides a segue to the question "what does the FOT-Probability perspective presented in this paper on signal modeling leave out that Kolmogorov stochastic processes encompass?" The answer is that *the theory of non-ergodic stochastic processes does not have an FOT-Probability counterpart.* Non-ergodic models of signals do have their uses. Specifically, when important conditions of an experiment change from one trial of the experiment to another, the impact revealed in an ensemble average of these changes cannot be determined from a time average on the time series from a single experimental trial. A common example in signal processing is the speech signal. In bandpass analog modulated signals such as AM, PM, and FM sinewave carriers and baseband (lowpass) pulse-modulated analog signals, such as PAM, PPM, and PWM, modulated by possibly quantized but not digitally encoded speech, there is no physical reason to assume that an ensemble of speakers will produce speech records that are simply time-translated versions of each other. Therefore, none of these modulated signals for multiple speakers can be expected to be simply time-translated versions of each other. The character of speech differs from one speaker to another due to physiological, language, and accent differences. So, an ergodic stochastic process model is inappropriate; but this does not mean the use of time-average statistics for a speaker is inappropriate. Consequently, FOT-probabilistic models for single speakers are appropriate and can be used in algorithms for signal processing that will be used for multiple speakers as explained next. But it is first mentioned that non-ergodic stochastic process models are also used for non-Gaussian noise modeling, not just signal modeling [19].

If one wants to design a speech processing algorithm that provides optimum performance averaged over all speakers in a group, a non-ergodic stochastic process model for the speech is appropriate. However, if one wants to design a *data-adaptive* algorithm that provides optimum performance for each and every speaker, then an FOT-Probability model is the appropriate conceptual tool, and the speech statistics required by the algorithm will be learned and adapted to for each individual signal. There is no place for an ergodic stochastic process model for a non-adaptive algorithm design for multiple speakers. The same remarks apply for applications involving communications channels that introduce noise that is collectively modeled in terms of multiple noise sources, random in number, and with multiple locations, random in their coordinates, relative to the receiver [19].

The same type of situation arises for many forms of information-modulated pulse and carrier signals, whether the information is discrete-time, continuous-time, analog, or digital. If there are different sources of information, such as telemetry for a variety of measurement types, different types of files of information exchanged between computers, etc., then there will be situations where one member from an ensemble of signals cannot be expected to be simply a time-translated version of some other member. So ergodic stochastic process models will typically be of poor fidelity, but FOT-Probabilistic models can be of high fidelity *for each individual signal* and are therefore the most appropriate for analysis and performance prediction for adaptive signal processing algorithms.

Finite-time time-average statistics are ubiquitous in statistical signal processing algorithms, and such algorithms are typically implemented with DSP software and/or hardware, which greatly facilitates adaptivity. The potential for considerably higher fidelity of the FOT-Probability models and the fact that these models, using idealized infinite-time averages follow all the same rules for finite mathematical manipulation as do stochastic process models, should encourage DSP algorithm designers to use FOT-Probability models in place of the traditional stochastic process models. And it is important to note that, as discussed in this paper, the Fundamental Theorem of Time Averaging applies to not only limits of time-average statistics but also finite-time averages: it applies to completely empirical quantities! Yet, there is a caveat: For the models derived from finite-time averages, some properties of the expectation and infinite-time-average models are only approximated. This appears to be more of an issue with poly-cyclostationary models, less so with cyclostationary models, and even less so with stationary models. This is due, at least in part, to the loss of the exact orthogonality of the harmonics of 1) a periodic function on a finite interval that is not an integer multiple of the period, and 2) a poly-periodic function on all finite intervals, and also due to the loss of exact statistical independence of random time series on all finite intervals. Consequently, the accuracy of these approximations becomes an important issue.

For readers who have been indoctrinated in stochastic process theory, the question that should be popping up at this point is: "where does the concept of *mean-square (m.s.) ergodicity and ergodicity in probability (weak ergodicity)*, as distinct from *ergodicity w.p.1* or *strong ergodicity*, arise in these considerations?". Typical engineering textbooks, such as the popular book by A. Papoulis [20], do not treat strong ergodicity. The fact of the matter is that m.s. and weak ergodicity and their extension/generalization to m.s. and *weak cyclo-ergodicity* introduced by Boyles and Gardner [15] (see also [8]) is of some use in practice. But it must be realized that these forms of ergodicity are *much* weaker than strong ergodicity. For example, m.s. ergodicity guarantees that the squared difference between a time average and an ensemble average (both possibly modified for cyclostationarity) does go to zero in the limit as averaging time approaches infinity, but *only* on average over the typically infinite ensemble. Therefore, this difference need not go to zero for many members of the ensemble. And these members need not be exotic as may those that may be present but are ignored by using the w.p.1 modifier. One might think that because squared error cannot be negative, the average squared error can be zero only if every individual error is zero. But this is not true because we are considering infinitely many errors. For continuous-time averages, a countably infinite number of errors can be non-zero while the average is still equal to zero. Although less commonly known, the average over all time can be zero even if the error at an uncountably infinite number of times is non-zero. The error can be non-zero throughout any finite interval, while the average error over all time is zero. Such are the vagaries of infinity. Consequently, signal processing engineers designing algorithms based on a theory of expected performance using a model that is only m.s. or weakly ergodic can be surprised by the occurrence of sample paths for which time averages differ greatly from the ensemble averages used in the design. If the algorithm is data-adaptive, it will use its own time-averages, and the resultant difference will be between performance predicted and performance achieved when ensemble average statistics and even the time-average statistics used for design and test are different from those arising in operation; and this difference might be either a pleasant or disappointing surprise. But, if the algorithm has been only theoretically optimized using ensemble-averages, and is not data-adaptive, the surprise can be expected to be disappointing for many signal realizations. Examples of fixed optimum vs. adaptive algorithms, as referred to here, are fixed Wiener filters vs. adaptive filters using least-mean-squares (LMS) and recursive least squares (RLS) algorithms; and also, for parameter estimators, detectors, and classifiers, as well as filters, there are fixed optimized implementations and there are adaptive implementations such as property restoral (PR) algorithms including modulus-restoral and cyclostationarity-restoral algorithms [21].

As another example, the direction in which communications technology has been moving for the last four or five decades is from primarily fixed and slowly adjustable channel equalizers (e.g., mid-century Bell Telephone Laboratories) to rapidly adaptive channel equalizers. This is a consequence, in part, of the evolution from hard-wired telephone channels for analog signals to wireless channels (e.g., for digital cellular communications signals) carrying not only voice but also data. This evolution has made non-ergodic stationary stochastic-process signal models as well as non-ergodic noise models less relevant and single-signal and single-noise FOT-Probabilistic models more relevant, as discussed above. The problem, however, is that engineering education in theoretical tools, unlike education in technology—which is typically at the forefront—has not kept up with this evolution. It is stuck teaching stochastic processes now much as it did five decades ago—except for a shift from mostly continuous-time signals to mostly discrete-time signals—even though the more relevant theory of FOT-Probability models was made available 35 years ago [4], [5].

The difference between the terms *statistical* and *probabilistic* are pointed out here for further clarity. Probabilities and probabilistic parameters, such as means, variances, correlations, probability densities, etc., defined in terms of mathematical expectation calculated from mathematic models, are theoretical or mathematical constructs. They come from within our heads through our imagination or as solutions to mathematical equations. In contrast, averages of empirical measurements, such as estimates of these theoretical quantities, are statistics. They can be obtained from finite ensemble averages derived from repeated experimentation or from finite-time averages performed on a single time series of measurements. This difference is very often ignored in the terminology chosen by users of these tools. This can cause the same type of confusion as that resulting from use of theoretical stochastic process models for implementations based on time-averages from single time series. Because stochastic processes are mathematical entities, no actual single signal can ever be considered to be ergodic or non-ergodic. It is a real statistic, not an imaginary probability model. For example, the Statistical Theory of Communication and Information Theory are both primarily probabilistic theories, but they do deal with statistics to some extent. When the focus is on statistics in communications, the traditional name for these theories is appropriate, but many if not most books on this subject focus on probabilities. In contrast, turbulence studies are especially interested in ensembles, for example, all aircraft of a specified design in all operational environments, or even a single aircraft in all operational environments. Here the ensemble in the definition of a stochastic process can be real, not just imagined. Yet, the stochastic process models used in turbulence studies are not real, only the finite ensembles of actual measured turbulence—the statistics—are real. The example set in Middleton's classic book [11], of being consistently clear about this distinction, has not been as diligently followed as would behoove the statistical signal processing community. It is my belief that the all-too-common lack of distinction between probabilities and statistics is a clear reflection of the confusion caused, at least in part, by the abstraction of the stochastic process model that engineers are indoctrinated in.

The entire subject of this article is but one example of a philosophical challenge of great practical import which we face every day in every endeavor: distinguishing between models of reality that our brains create and the real thing—reality itself. People generally act on the basis of their models of reality for better or for worse. The effectiveness of interpersonal communication is dictated by the models in terms of which the communicators think. Further discussion of the impact, of the challenge to better match models with reality, on the conduct of science is available on page 7 of this University of California website [22].

Acknowledgment

The Author acknowledges the important role that his collaborative work with Professor Antonio Napolitano on writing the paper [1] played in the development of the Author's thoughts that are conveyed in the present paper.

5. References

1. Napolitano, A., and W. A. Gardner. 2022. "Fraction-of-Time Probability: Advancing Beyond the Need for Stationarity and Ergodicity Assumptions." *IEEE Access*. April 2022
2. Brennan, D. G,1961. "Probability Theory in Communication System Engineering." Chapter 2 in *Communication System Theory*. Ed. E. J. Baghdady. New York: McGraw Hill
3. Hofstetter, E. M. 1964. "Random Processes." In *The Mathematics of Physics and Chemistry*, edited by H. Margenau and G. M. Murphy. Vol. II. Princeton, NJ: D. Van Nostrand Co.
4. Gardner, W. A. 1987. *Statistical Spectral Analysis: A Nonprobabilistic Theory*. In. Englewood Cliffs, NJ: Prentice-Hall.
5. Gardner, W. A., and W. A. Brown. 1991. "Fraction-of-Time Probability for Time-Series That Exhibit Cyclostationarity." *Signal Processing* 23: 273–92.
6. Kac, M., and H. Steinhaus. 1938. "Sur Les Foncions Indépendantes IV." *Studia Mathematica* 7: 1–15.
7. Leśkow, J., and A. Napolitano. 2006. "Foundations of the Functional Approach for Signal Analysis." *Signal Processing* 86 (12): 3796–3825.
8. Napolitano, A. 2019. *Cyclostationary Processes and Time Series: Theory, Applications, and Generalizations*. Elsevier. https://doi.org/10.1016/C2017-0-04240-4.
9. Kolmogorov, A. N. 1933. *Foundations of the Theory of Probability*. 1956, Translated to English, 2nd ed. Chelsea Publishing Co.
10. Y. W. 1960. *Statistical Theory of Communication*. New York: Wiley.
11. Middleton, D. 1960. *An Introduction to Statistical Communication Theory*. Reissued by Wiley-IEEE Press, 1996. Online ISBN: 9780470544112, Print ISBN: 9780780311787.
12. Birkhoff, George D. 1931. "Proof of the Ergodic Theorem." *Proceedings of the National Academy of Sciences* 17 (12): 656–60. https://doi.org/10.1073/pnas.17.2.656.
13. Bergelson, V., A. Leibman, C.G. Moreira. 2011. "From discrete- to continuous-time ergodic theorems." *Ergodic Theory and Dynamical Systems*, Cambridge University Press, Volume 32: 2 , April 2012 , pp. 383 – 426.https://www.cambridge.org/core/journals/ergodic-theory-and-dynamical-systems/article/abs/from-discrete-to-continuoustime-ergodic-theorems/E5CCE8CBDB2FA4445BE2FCFC0A08C7C2
14. Wold, H. O. A. 1948. "On Prediction in Stationary Time Series." *Ann. Math. Statist.* 19: 558–67.

15. Boyles, R. A., and W. A. Gardner. 1983. "Cycloergodic Properties of Discrete Parameter Non-Stationary Stochastic Processes." *IEEE Transactions on Information Theory* IT-29: 105–14.
16. Gardner, W. A. 2018. "Statistically Inferred Time Warping: Extending the Cyclostationarity Paradigm from Regular to Irregular Statistical Cyclicity in Scientific Data." *EURASIP Journal on Advances in Signal Processing*, (1): 59. https://doi.org/10.1186/s13634-018-0564-6.
17. Gardner, W. A., ed. 1994. *Cyclostationarity in Communications and Signal Processing*, IEEE Press.
18. Gardner, W. A.1985. *Introduction to Random Processes with Applications to Signals and Systems*. In. New York: Macmillan.
19. Middleton, D. 1979. "Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance." IEEE Transactions on Electromagnetic Compatibility EMC-21: 3, 209-220.
20. Papoulis, A. 1965.*Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Kogakusha, Tokyo, 9th edition, ISBN 0-07-119981-0
21. Schell, S. V., and W. A. Gardner. 1995 "Programmable Canonical Correlation Analysis: A Flexible Framework for Blind Adaptive Spatial Filtering." *IEEE Transactions on Signal Processing*, 43: 12, pp. 2898-2908.
22. Gardner, W. A. "Cyclostationarity.com." 2018. https://cyclostationarity.com.
23. Spooner, C.M. https://cyclostationary.blog/
24. Dehay, D., J. Leśkow, and A. Napolitano. 2013. "Central Limit Theorem in the Functional Approach." *IEEE Transactions on Signal Processing* 61 (16): 4025–37. https://doi.org/10.1109/TSP.2013.2266324.
25. Napolitano, A., 2017. "Time-Warped Almost-Cyclostationary Signals: Characterization and Statistical Function Measurements." *IEEE Transactions on Signal Processing* 65 (20): 5526–41. https://doi.org/10.1109/TSP.2017.2728499.
26. Gardner, W.A. 1978. "Stationarizable Random Processes. *IEEE Transactions on Information Theory*, 24 (1): 8-22.