Fraction-of-Time Probability: Advancing Beyond the Need for Stationarity and Ergodicity Assumptions

ANTONIO NAPOLITANO[ID]**1, (Fellow, IEEE), AND WILLIAM A. GARDNER2, (Life Fellow, IEEE)**
[1]Department of Engineering, University of Napoli "Parthenope," 80143 Napoli, Italy
[2]Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95616, USA

Corresponding author: Antonio Napolitano (antonio.napolitano@uniparthenope.it)

ABSTRACT Time series arising from measurements in many fields of physics, engineering, chemistry, biology, and econometrics, are commonly modeled as sample paths from an ensemble which, together with a probability measure, is called a stochastic process. Stationarity and ergodicity assumptions about this model are generally made for analytical convenience and mathematical tractability of the model. In this article, it is shown that a dichotomy, which can be very misleading in practice, exists between the properties of a stochastic process and those of its individual sample paths. This dichotomy can be eliminated by adopting the fraction-of-time (FOT) probability approach reviewed in this article for which a probabilistic model is constructed from a single time series without introducing the abstraction of the stochastic process. Two FOT-probability models are reviewed. The first considers probabilistic functions that do not depend on time and employs the relative measure on the real line as a probability measure and the time average as an expectation operator. Such time series are called stationary signals. The second considers periodic, poly-periodic, and almost periodic probabilistic functions and employs the operator that extracts the finite-strength additive sine-wave components of its argument as an expectation operator. This latter model is appropriate for describing time series originating from phenomena involving a combination of periodic and random phenomena. Such time series are called cyclostationary, poly-cyclostationary, and almost cyclostationary signals. The FOT-probability alternative provides a means for circumventing two standard but undesirable practices: (1) Adopting the Kolmogorov stochastic process model by using its Axiom VI without being able to verify its validity for the specific application and (2) Assuming Birkhoff's ergodicity condition holds without being able to verify its validity for the specific application.

INDEX TERMS Fraction-of-time probability, cyclostationarity, ergodicity, cycloergodicity, stochastic processes, time series.

FOREWORD

The *Periodogram* is a well-known statistic comprised of the squared magnitude of the Fourier transform of a finite segment of data, normalized by the segment length, T. One of several equivalent standard definitions of the power spectral density function of a stationary stochastic process is that it is the limit as T approaches infinity of the expected value of the periodogram. A convergence issue arising in this definition

that many statistical signal processing engineers are at least aware of is that one cannot interchange the order of expectation and limit. The reason is that, before expectation, the limit of the periodogram does not exist. In addition, the variance of the periodogram does not converge to zero as T approaches infinity.

These fairly well-known mathematical facts are responsible for the practical guideline that one cannot obtain a reliable power spectrum estimate from the periodogram, no matter how long the segment length is, unless that periodogram is either 1) frequency smoothed (averaged over frequency

---

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott[ID].

throughout every frequency sub-band of width much larger than the reciprocal of the length of the data segment, or 2) time smoothed (averaging of a sliding periodogram over an interval much longer than the segment length).

So, what may appear to some to be an esoteric mathematical curiosity–the invalidity of interchanging the order of the abstract expectation operation and abstract limit operation–is, in fact, the *reason why* the practical guidelines for measurement of reliable power spectra are what they are, as summarized above. This important fact of statistical spectral analysis, which extends and generalizes to cross-spectrum analysis and spectral-correlation analysis of cyclostationary processes as I have shown in this article [42] becomes transparent only when studied within the fraction-of-time-probability framework as first explained in considerable detail in my advanced-level textbook [37]. By comparison, attempted explanations in other textbooks, in terms of stochastic processes, cannot be said to be transparent. The reason for this is that *properties of ensemble averages of sample paths are typically different than properties of individual sample paths*, which is a theme of this article. This difference does not vanish as commonly thought if the stochastic process is ergodic.

Engineers often deal mathematically with expected values and deal in practice with single sample paths or their time averages, and they often tend to freely interchange these quantities by replacing expected values with finite-time averages and, in some cases, simply deleting the expectation operation in deriving algorithms. Sometimes this is theoretically justifiable, and sometimes it is not. What percentage of practicing statistical signal processing engineers know when and why this interchange is not justifiable? Whatever small percentage this might be, it is likely far larger than the percentage of those who are aware of the fact that this frequently used interchange is not only sometimes unjustifiable but is also avoidable by adopting a little-known analytical tool that replaces the unnecessarily abstract stochastic process tool with a less abstract alternative. The issue here is not resolved by simply using stochastic process models that are ergodic.

The purpose of this tutorial is to teach engineers about this alternative tool and provide a deep understanding of how it relates to the stochastic process.

Many alternative titles for this article were considered; a number of alternatives to the choice made had the advantage of being attention getters and, for that reason, I mention a few of them here: ''Why Ergodic Stochastic Processes are Inappropriate Models in Many Empirical Studies'', ''The Untold Truth About Stochastic Processes: They Tell Us Nothing About Sample Path Behavior'', and ''How to Avoid Making Standard Unverifiable Assumptions in the Use of Stochastic Process Models''.

–William A. Gardner

I. INTRODUCTION AND HISTORICAL PERSPECTIVE

In many fields of physics, engineering, chemistry, biology, and econometrics, randomness in time-series measurements and observations on phenomena being studied has typically been modeled by resorting to the abstract concept of a stochastic process. That is, an empirical time series is modeled as a representative (sample path or realization) of an ensemble of time series with ''similar characteristics'', together with a probability measure defined on the set of ensemble members, namely, a stochastic process. A desirable property of the model is that the probabilistic functions comprising the stochastic process be estimable by measurements made on the single empirical time series.

That is, the ''similar characteristics'' of the sample paths should be such that the properties and, in particular, the probabilistic functions of the whole ensemble such as mean, covariance, and amplitude distribution, can be inferred by measurements made on any one of the sample paths, with the exception of a set of sample paths that occurs with zero probability. This desirable property is called *ergodicity* in classical stochastic process theory. This provides the all-important tie for signal processing applications between the model and that which can be measured. However, it is shown in this review article that this tie is not as strong a tie as one would like for empirical signal processing purposes.

The ergodicity concept was first treated mathematically in 1931 by Birkhoff [14] and in 1932 by von Neumann [105] with reference to dynamical systems. They established conditions under which averaging, at a single time instant, a function of the variables of what was called the *phase space* across an ensemble of different copies of the same system is equivalent to averaging over time for a single system. See [87] for a discussion on the slightly different points of view of Birkhoff and von Neumann.

Subsequently, ergodicity has played a key theoretical role in multiple fields [3], [18], and has been the subject of much thought and discussion. As examples, see the recent treatments in the fields of economics [26], [19], [93], [94], atomic physics [65], and condensed matter systems [83]. Theoretical research on ergodicity has continued for nearly a century; see, for example, the more recent work by Boyles and Gardner [16], Shields [98], Katznelson and Weiss [63], and Gray [47], and the latest work by the Authors reported in this review. For a comprehensive bibliography on ergodicity, see [69].

The adoption of the stochastic process model requires a substantial abstraction, the hypothesized mathematical existence of the ensemble which, in various circumstances, creates serious conceptual problems [45, page 3.4]. In fact, for many applications in the various fields of science and engineering, there is only one record of real data; there is no ensemble of statistically independent random samples of data records. For stochastic processes, mixing (a type of asymptotic independence of time-samples with increasing separation in time) assumptions are typically used as sufficient

conditions for ergodicity. But such conditions often cannot be verified for the adopted mathematical models; generally, they are simply assumed to hold on the basis of little more than faith. It is explained in this review that properties of probabilistic functions defined in terms of ensemble averages in most cases do not correspond to similar properties of analogous functions defined in terms of time averages on single sample paths, even if the stochastic process is ergodic. Consequently, there is a dichotomy between the properties of a stochastic process–the model–and the properties of its individual sample paths–the physical signals. The idea that averaging over an ensemble does not necessarily correspond to averaging over time–the lack of ergodicity of some stochastic process models–is receiving increasing attention in various fields of study [28], [83], [84]. Furthermore, it is explained in the following sections that the critical nature of some assumptions made in the classical probability theory of stochastic processes was already pointed out by Kolmogorov in his seminal work 90 years ago [66, p. 15]. An attempt to describe randomness without introducing an abstract sample space and probability was made by Kolmogorov himself by introducing the concept of complexity for a single random sequence [67]. Such an alternative approach, however, did not enjoy the same success probability theory did. See [21], [80], and [104] for development and discussions.

In this paper, more concrete models based on properties of time averages of individual signals, instead of averages over hypothetical ensembles of signals, are reviewed and new results on the concrete benefits of this alternative are reported. This approach is based on what is called *fraction-of-time (FOT) probability*, which was first introduced in earnest in 1987, with a comprehensive book [37] devoted exclusively to this theme by Gardner, and the duality between these two alternative models was a theme of the earlier book [34]. But, important advances in the theoretical underpinnings of the FOT-probability theory of signals have been made quite recently.

In this approach, starting from a single function of time, a valid distribution function and all other familiar probabilistic parameters such as means, moments, and cumulants are constructed. That is, formulas for calculating these functions/parameters directly from time series are developed. The approach can be put in a rigorous measure-theoretic framework built on the conceptual foundation of *relative measure* introduced by Kac and Steinhaus in the 1930s [60]–[62]. Developments were presented during the 1940s and 1950s in [1], [100]–[102]. More recently, in the 1980s there was [30] and, in the mid-2000s, the concept was revisited by Leśkow and one of the Authors, Napolitano in [72]. Very recent (unpublished) work by the other Author, Gardner, on generalizations of ergodicity to cycloergodicity that ameliorate some of the drawbacks of the standard assumptions of stationarity and ergodicity are included in this review.

The FOT probability approach is methodological, it does not necessarily provide new signal processing recipes.

However, in the paper it is shown that the adoption of this approach allows the experimenter to avoid dichotomies between properties of the stochastic process model and those of its sample paths. The dichotomies illustrated in the paper regard issues frequently encountered in signal processing applications: The expectation of a digitally modulated signal; the linear transformation of a Gaussian signal; and the concept of statistical independence. In addition, in the paper it is shown that some properties of the probability measure and the expectation operation render these entities more amenable to some mathematical operations than do the temporal counterparts of these properties for the sample paths. In contrast, the properties of the relative measure and the infinite time average are exactly the properties of the functions at the hand of the experimenter, the only abstraction being the concept of an infinite observation interval. (Infinite-time averages of single records of data are somewhat loosely said to be "at hand"; the justification for this is the fact that finite-time averages of single time records–the items that are truly "at hand"– can closely approximate infinite-time averages as explained in Sec. V-B; this close relationship between an empirical quantity and a corresponding theoretical quantity is to be contrasted with the major distinction made in this paper between the theoretical expected value (an ensemble property) and both finite-time and infinite-time averages (sample-path properties).)

A precursor of the spectral analysis of a single time series is due to Einstein in his work reported in 1914 [29] (see [36] for Gardner's introduction to and technical commentary on this work and the comments of Yaglom in [113]).

The FOT approach has its roots in Wiener's work on generalized harmonic analysis (GHA) for single time series dating back to 1930 [106] [107, Chap. 4], [108, pp. 37-45]. Wiener's GHA approach was followed by some of his students [70, Chap. 2] and a few authors: Bochner, Hartman, Jessen, Kershner, and Wintner [15], [50], [59], [64], [109], [111], Haviland [52], Bass and his students [4]–[6], [7, part 5, chap. 1-3], [8]–[10], [12], [54], Furstenberg [31], Brennan [17], Hofstetter [55], Benedetto [11, Sec. 2.9], Mäkilä [75], [76], and Nobel [91], [92]. Results have been presented in spectral analysis [20], [53], [95], and for the prediction of individual sequences [86]. Ziv developed a coding theory for individual sequences [118]. The definition of jointly relatively measurable functions [72], and the issues of quantile prediction [71], central limit theorem [24], and time average estimation [25] were addressed by Dehay, Leśkow, and Napolitano.

Roughly speaking, in the above-mentioned papers, an underlying stationary model is assumed. In fact, the FOT expectation operator is the infinite time average and the probabilistic functions derived therefrom are time invariant.

The extension of the FOT approach and GHA to periodic phenomena was made by Gardner in the mid 1980s [37, Part II], [43]. Periodic phenomena are generated by the

interaction of periodic mechanisms and random phenomena. The results are processes that are not periodic but whose probabilistic functions vary periodically with time. These signals have been referred to as *cyclostationary* or *poly-cyclostationary* if, respectively, only one periodicity or a finite number of incommensurate periodicities are present in the probabilistic functions. If the probabilistic functions are almost-periodic functions of time (which can generally be represented by infinite numbers of incommensurate periods), the signals have been named *almost-cyclostationary* [37, Part II] and provide an alternative to *almost-cyclostationary* stochastic processes first introduced in the late 1970s [33].

The extension of the FOT approach and GHA presented in [37, Part II], [43] is of great importance in practice due to the ubiquity of science data generated by periodic phenomena. In communications, radar, sonar, and telemetry, periodicities in the statistical functions arise from the modulation by random data of sinusoidal carriers or periodic pulse trains [37, Chap. 12], [90, Chap. 7]. In vibro-acoustic signals collected from mechanical machinery, periodicities in the statistics are due to rotations of gears, belts, and bearings [90, Sec. 10.6]. In econometrics, the weekly opening and closing of markets and the seasonal supply and demand of products give rise to periodicities in the statistical functions of prices and exchange rates [90, Sec. 10.7]. In radio astronomy, periodicities are due to the revolution and rotation of planets and pulsation of stars; and in human biological signals, by the heart pulsation and other biological rhythms. Periodicities are present in genome sequences, diffusion processes of molecular dynamics, and signals encountered in neuroscience (see [90, Chap. 10] and references therein, and also the 2018 results of an internet search producing 135,000 published research papers on cyclostationarity across essentially all fields of science and engineering [46]).

The extension of FOT and GHA to periodic phenomena made in [37, Part II], [43] is not obvious since Gardner's periodically or almost-periodically time-variant distribution is constructed from a single function of time by recognizing the non-obvious fact that the operator that extracts all the finite-strength additive sine-wave components of its argument is a valid expectation operator.

The reader is informed that the very brief use in the paper of a very few mathematical terms, like *σ-field* and *Borel subsets*, can be ignored without loss of comprehension. And the term *Lebesgue measure* can be thought of as nothing more than a formalization of standard measures of length, area, and volume in Euclidean space. Finally, the differences between the more familiar Riemann integral and the Lebesgue and Riemann-Stieltjes integrals can be ignored, or standard definitions of these integrals can be looked up on the Web, but adequate comprehension for the non-mathematician does not require this.

The paper is organized as follows. In Section II, motivation for introducing FOT probability is presented. In Section III, the relative measure–the analog of the classic probability measure–and its properties are reviewed. In Section IV, the

extension/generalization of the FOT-probability theory to almost-periodic phenomena is treated. In Section V, new insight into the mostly unstudied topic of cycloergodicity is provided and the task of statistical function estimation in the FOT approach is addressed. Examples of the dichotomy existing between the properties of a stochastic process and those of its sample paths are presented in Section VI. Conclusions are drawn in Section VII.

NOTATION

$\mu(\cdot)$	Lebesgue measure
$\mu_R(\cdot)$	relative measure (see (1))
$\langle \cdot \rangle_t$	infinite-time average (see (7))
$\mathbf{1}_A(\cdot)$	indicator of the set A
$x(t)$	time series (small letters)
$X(t)$	stochastic process (capital letters)
$F_x(\cdot)$	fraction-of-time distribution function (see (4a))
$F_X(\cdot)$	(stochastic) probability distribution function
$u(\cdot)$	unit-step function
$\delta(\cdot)$	Dirac delta
$E\{\cdot\}$	ensemble average
$E^{\{\alpha\}}\{\cdot\}$	almost-periodic component extraction operator (see Sec. IV)

ABBREVIATIONS

ACS	almost cyclostationary
AMS	asymptotically-mean stationary
AP	almost periodic
CLT	central limit theorem
FOT	fraction-of-time
GHA	generalized harmonic analysis
PAM	pulse-amplitude modulated
RM	relatively measurable

II. MOTIVATION

The relative measure μ_R and the infinite-time average are the fraction-of-time (FOT) counterparts of the probability measure P and the ensemble average, respectively [62], [72].

Due to the differences between the relative measure, μ_R, on the relatively measurable sets (which are a subset of the σ-field of Borel subsets of the real line) and the probability measure, P, on the σ-field of Borel subsets of a probability sample space, mathematical properties holding for stochastic processes do not necessarily have counterparts that hold for functions of time representing sample paths of these stochastic processes.

The key differences include:
- The class of the *P*-measurable sets is closed under union and intersection; the class of the relatively measurable sets is not.
- *P* is a σ-additive (additivity of countably infinite numbers of terms) measure; μ_R is not.
- Expectation is σ-linear (linearity of an operator applied to a linear combination of a countably infinite number of terms); infinite-time average is not.

- Joint *P*-measurability of sample spaces is typically assumed but cannot be verified; joint relative measurability is a property of functions that can be verified.
- The assumed σ-additivity property of the probability measure is typically unverifiable and restricts the admissible sample spaces of time functions in ways that the relative measure does not.
- The relative measure is applied to the single time function at hand, and functions of this time function, with no restrictions other than its assumed existence. The fact that the relative measure cannot be guaranteed to be σ-additive is a reflection of the reality of the time function at hand, not a deficiency.

These differences clearly show that the mathematical properties of the relative measure render it less amenable to mathematical study than do those of the probability measure *P*. This, however, does not constitute an obstacle to using the FOT approach for signal analysis but, rather, as explained in this paper, provides motivation for using this approach instead of the classical stochastic-process approach based on *P*. In fact, the σ-additivity of the probability measure and σ-linearity of the expectation provide mathematically desirable tractability. But, as explained below, they give rise to a dichotomy between the stochastic process properties and the properties of concrete individual sample paths of the stochastic process–the entities of primary interest to practitioners in many applications. In contrast, such dichotomies do not arise in the FOT probability approach. In addition, the adoption of the FOT approach overcomes all problems arising from the need to check sufficient conditions for validating assumptions for ergodicity–problems which occur frequently in time-series analysis applications.

The proposal to adopt the FOT-probability alternative to the Kolmogorov formulation of a stochastic process is by no means as outrageous as some may think. In fact, there is a long history of discontent with Kolmogorov's model, as discussed at length in [74].

III. FRACTION-OF-TIME PROBABILITY VERSUS CLASSICAL PROBABILITY

A. RELATIVE MEASURE
Let us consider the set $A \in \mathcal{B}_{\mathbb{R}}$, where $\mathcal{B}_{\mathbb{R}}$ is the σ-field of the Borel subsets of the real numbers and let μ be the Lebesgue measure on the real line $\mathbb{R}$. The *relative measure* of *A* is defined by [62]

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

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).

For example, the set

$$A \triangleq \bigcup_{n \in \mathbb{Z}} (n, n + 1/3) \quad (2)$$

where $\mathbb{Z}$ is the set of positive and negative integers, is RM and $\mu_R(A) = 1/3$. The set

$$A \triangleq \bigcup_{n \in \mathbb{N}} \left(-10^{2n}, -10^{2n-1}\right) \cup \left(10^{2n-1}, 10^{2n}\right) \quad (3)$$

is not RM since $\frac{1}{T} \mu(A \cap [-T/2, T/2])$ oscillates between 0 and 1 in a continuous piecewise linear manner as $T \to \infty$.

The relative measure is the Lebesgue measure normalized so that the relative measure of the real line is equal to 1, that is, $\mu_R(\mathbb{R}) = 1$. Note that such a normalization is obtained by a limit operation (as $T \to \infty$) since the Lebesgue measure of the real line is infinite. Therefore, only Lebesgue-measurable sets with infinite Lebesgue measure can have a finite relative measure. For a set of points on the real line to have non-zero relative measure, it must contain subintervals of non-zero Lebesgue measure that cannot be contained within any finite interval. As defined above, the relative measure of such a set is the limit of the ratio of 1) the total Lebesgue measure of those points in the set contained in a contiguous window of width *T* to 2) the window width, as *T* approaches infinity. In other words, it is the fraction of the real line that the set occupies.

The normalization $\mu_R(\mathbb{R}) = 1$ makes the relative measure of subsets of the real line a counterpart of the probability measure *P* defined on sets belonging to the σ-field $\mathcal{F}$ of the sample space Ω. For the probability measure, however, the normalization $P(\Omega) = 1$ is obtained without a limit operation. That is, the sample space, before the measure normalization, is *assumed* to have finite measure. Thus, the normalized probability measure of a set is obtained by the ratio of the original un-normalized measure and the assumed-to-be-finite measure of the whole sample space.

Such a subtle property of the sample space of the classical probability measure has been surfaced by only a few authors (see e.g., Halmos [48] and [49, p. 31]) even if the criticality of considering an infinite sample space (although with finite measure) was already surfaced by Kolmogorov in his fundamental work on the theory of probability where he addressed the necessity to introduce Axiom VI (Axiom of Continuity) which cannot be derived from Axioms I–V. Citing [66, p. 15]: ''For infinite fields, on the other hand, the Axiom of Continuity, VI, [is] proved to be independent of Axioms I–V. Since the new axiom is essential for infinite fields of probability only, *it is almost impossible to elucidate its empirical meaning*, as has been done, for example, in the case of Axioms I–V in § 2 of the first chapter. For, in describing any observable random process we can obtain only finite fields of probability. *Infinite fields of probability occur only as idealized models of real random processes. We limit ourselves, arbitrarily, to only those models that satisfy Axiom VI.* This limitation has been found *expedient* in researches of the most different sort.''

The normalization of the relative measure obtained by a limit operation results in μ_R having mathematical properties that render it less amenable to mathematical analysis than do those of the probability measure *P*, as explained subsequently.

The class $\mathcal{C}$ of the RM sets is not closed under union and intersection. That is, there exist $A, B \in \mathcal{C}$ such that $A \cap B \notin \mathcal{C}$ [72, Fact 2.1] and there exist $A, B \in \mathcal{C}$ such that $A \cup B \notin \mathcal{C}$ [72, Fact 2.2]. As an immediate consequence of [72, Fact 2.1], we have that not all Lebesgue-measurable sets (with infinite measure) are relatively measurable. In addition, non-RM sets are not so rare or exotic (see (3)) as non-Lebesgue-measurable sets [68, Par. 27, problem 7].

B. ABSENCE OF σ-ADDITIVITY OF THE RELATIVE MEASURE

Since the Lebesgue measure is additive, the relative measure is also additive. That is, if $A, B \in \mathcal{C}$, and $A \cap B = \emptyset$, then $\mu_R(A \cup B) = \mu_R(A) + \mu_R(B)$ [72, Fact 2.4]. However, the relative measure is not σ-additive. In fact, following the Kac example [60, p. 46], if $A_i = (i, i+1)$, $i \in \mathbb{Z}$, then $\mu_R\left(\bigcup_{i \in \mathbb{Z}} A_i\right) = \mu_R(\mathbb{R}) = 1$ but $\sum_{i \in \mathbb{Z}} \mu_R(A_i) = 0$ since $\mu_R(A_i) = 0$.

The absence of the σ-additivity property makes μ_R less attractive from a mathematical point of view than the probability measure P whose σ-additivity is a consequence of Axiom VI [66, p. 15]. The criticality of such an Axiom (and its consequences) has already been discussed in previous paragraphs. The importance of σ-additivity for ergodicity, and in particular in the Poincaré's proof of the Recurrence Theorem, is surfaced in [69, p. 17], [82, p. 80].

From the above considerations, it is clear that a probabilistic model built for a persistent single function of time starting from the relative measure will have mathematical properties less amenable to mathematical analysis than will the classical probabilistic model of a stochastic process which is built starting from the probability measure P. As explained below, this fact, which could initially appear to be a weakness of μ_R in comparison to P, constitutes, instead, a *motivation* to adopt the FOT approach for signal analysis instead of the classical stochastic-process approach.

C. RELATIVELY MEASURABLE FUNCTIONS

Let $x(t)$ be a real-valued Lebesgue measurable function. The function $x(t)$ is said to be *relatively measurable* if the set $\{t \in \mathbb{R} : x(t) \leqslant \xi\}$ is RM for every $\xi \in \mathbb{R} - \Xi_0$, where Ξ_0 is at most a countable set of points. Each RM function $x(t)$ generates a function

$$F_x(\xi) \triangleq \mu_R(\{t \in \mathbb{R} : x(t) \leqslant \xi\}) \tag{4a}$$

$$= \lim_{T \to \infty} \frac{1}{T} \mu(\{t \in [t_0 - T/2, t_0 + T/2] : x(t) \leqslant \xi\}) \tag{4b}$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_{t_0 - T/2}^{t_0 + T/2} u(\xi - x(t)) \, dt \tag{4c}$$

at all points ξ where the limit exists. In (4c), $u(\xi)$ denotes the unit step function, that is, $u(\xi) = 1$ for $\xi \geqslant 0$ and $u(\xi) = 0$ for $\xi < 0$.

The function $F_x(\xi)$ is non decreasing and with values in $[0, 1]$. Thus, it has all the properties of a valid distribu-
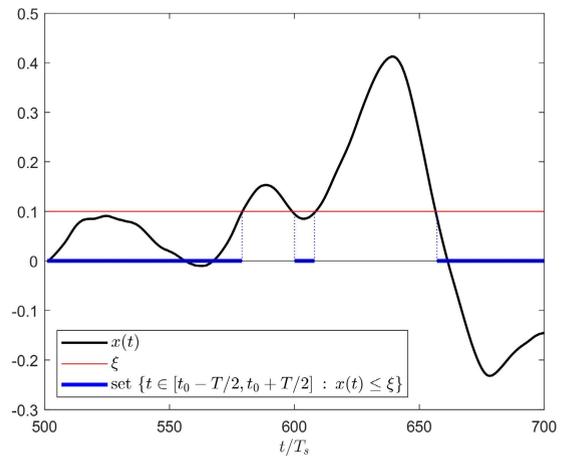
FIGURE 1. The measure of the set $\{t \in [t_0 - T/2, t_0 + T/2] : x(t) \leqslant \xi\}$ (blue thick line) normalized to T is the fraction of time that the function $x(t)$ is below the threshold ξ when t ranges in $[t_0 - T/2, t_0 + T/2]$.

tion function, except for the right-continuity property (at the discontinuity points). It represents the *fraction-of-time* (FOT) that the function $x(t)$ is below the threshold ξ (Fig. 1) [34], [37], [55]. For this reason, $F_x(\xi)$ is referred to as the FOT distribution of the function $x(t)$.

Since the relative measure of finite sets is zero, every finite-energy or transient function $x(t)$ has the trivial distribution function $F_x(\xi) = u(\xi)$. Only finite-average-power or persistent functions can have a non-trivial FOT distribution. The FOT distribution of almost-periodic functions is studied in [64], [109], [111].

Sets that are non Lebesgue measurable are difficult to be visualized and are very rarely, if ever, encountered in applications. In contrast, non-RM sets are not rare or exotic (see (3)) and non-RM functions also can easily be constructed. In fact, let $\mathbf{1}_A(t)$ be the indicator of the set A, that, is $\mathbf{1}_A(t) = 1$ if $t \in A$ and $\mathbf{1}_A(t) = 0$ if $t \notin A$. The function $x(t) = \mathbf{1}_A(t)$ is RM if and only if A is a RM set and it follows that

$$F_x(\xi) = \begin{cases} 0 & \xi < 0 \\ 1 - \mu_R(A) & 0 \leqslant \xi < 1 \\ 1 & \xi \geqslant 1 \end{cases} \tag{5}$$

This fact has been exploited in [73] to design modulation formats for which statistical functions cannot be measured by time averages, for the purpose of obtaining secure communications.

Let $x(t)$ be a relatively measurable, not necessarily bounded function and let $g(\cdot)$ be continuous, bounded, and such that for any $\ell \in \mathbb{R}$, the equation $g(\xi) = \ell$ admits at most a finite number of solutions for ξ belonging to any finite interval. The following *Fundamental Theorem of Expectation* [72, Theorem 3.2] holds

$$\lim_{T \to \infty} \frac{1}{T} \int_{t_0 - T/2}^{t_0 + T/2} g(x(t)) \, dt = \int_{\mathbb{R}} g(\xi) \, dF_x(\xi) \tag{6}$$

where the first integral is in the Lebesgue sense and does not depend on t_0 and the second one is in the Riemann-Stieltjes sense (see [72, pp. 3816-3817] for details).

From (6) 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

$$\langle x(t) \rangle_t \equiv \lim_{T \to \infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} x(t) \, \mathrm{d}t = \int_{\mathbb{R}} \xi \, \mathrm{d}F_x(\xi) \quad (7)$$

The analogy of the FOT approach with the classical stochastic-process approach [34, Sec. 8.6], [40], is evident. On the nexus between the two approaches see also [31] and [81]. For a 1st-order strict-sense stationary process $X(t)$ with distribution $F_X(\xi) \triangleq P[X(t) \leqslant \xi]$, the stochastic counterpart of (4c) is

$$F_X(\xi) = \mathrm{E}\{\mathrm{u}(\xi - X(t))\} \quad (8)$$

where $\mathrm{E}\{\cdot\}$ is the ensemble average and the stochastic counterparts of (6) and (7) are

$$\mathrm{E}\{g(X(t))\} = \int_{\mathbb{R}} g(\xi) \, \mathrm{d}F_X(\xi) \quad (9)$$

and

$$\mathrm{E}\{X(t)\} = \int_{\mathbb{R}} \xi \, \mathrm{d}F_X(\xi) \quad (10)$$

respectively.

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, accounting for the Fundamental Theorem of Expectation, if $x(t)$ is continuous and bounded and the left-hand side of

$$\lim_{T \to \infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} x^p(t) \, \mathrm{d}t = \int_{\mathbb{R}} \xi^p \, \mathrm{d}F_x(\xi) \quad (11)$$

exists for every positive integer p, then $x(t)$ is relatively measurable, and equality (11) holds [110].

Finally, note 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 (see (6) and (7)). In fact, in the discontinuity points, only the amplitude $F_x(\xi+) - F_x(\xi-)$ of the jump and not the value of $F_x(\xi)$ influences the value of the integral. In the stochastic approach the right-continuity of the distribution is a consequence of the σ-additivity of the probability measure P.

D. JOINTLY RELATIVELY MEASURABLE FUNCTIONS

By building counterexamples, it can easily be shown that the class of the RM functions is not closed under addition and multiplication [72, Theorem 3.5]. Thus, the class of the RM functions is not a function space. Such a results is in accordance with the fact that the class of finite-average-power functions is not a linear vector space [4], [76], and [79].

Note that, in contrast, the linear combination of two stochastic processes is still a stochastic process, provided that the two sample spaces are *assumed* to be jointly measurable. This, however, is not an innocuous assumption. Consequently, even if there exists an analogy of results between the FOT and stochastic approaches (compare (7) and (10)), the stochastic process model for a single realization at hand should be used carefully, since properties of the stochastic process do not necessarily correspond to analogous properties of the function of time at hand. Such a deep difference between properties of stochastic processes and properties of functions constitutes a strong *motivation* for adopting the FOT approach for signal analysis. In fact, the experimenter is interested in actual properties of the time series at hand rather than better properties of an abstract model.

The uniformly almost-periodic functions constitute a linear vector space [22, Chap. 1]. However, such a class of function, albeit broad and useful for approximating every function on a finite observation interval, does not provide a suitable model for most signals encountered in practice. An example is the class of communications signals. Several attempts have been made to consider sub-classes of the class of the finite-average-power functions in order to obtain linear vector spaces [5], [76], [103], [115]–[117]. None of these approaches involve FOT distributions or allow one to construct a (non-stochastic) probabilistic model.

Alternatively, in the FOT approach, the joint characterization of two (or more) functions is made by introducing the concept of joint relative measurability of functions [72]. In particular, it can be shown that the sum and product of jointly RM functions is in turn a RM function. Thus, for such a function, an FOT probabilistic model can be constructed. The joint relative measurability is an analytical property of functions and is therefore easier to verify than is the analogous property in the stochastic process framework, that is, the joint measurability of sample spaces. The latter property, in fact, cannot easily be verified in practice since, generally, the sample spaces are not specified.

Two Lebesgue measurable functions $x(t)$ and $y(t)$ are said to be jointly RM [72, Definition 4.1] if the limit

$$
\begin{aligned}
F_{xy}&(\xi_1, \xi_2) \\
&\triangleq \mu_R(\{t \in \mathbb{R} : x(t) \leqslant \xi_1\} \\
&\quad \cap \{t \in \mathbb{R} : y(t) \leqslant \xi_2\}) \quad (12a) \\
&= \lim_{T \to \infty} \frac{1}{T} \mu(\{t \in [t_0 - T/2, t_0 + T/2] : \\
&\quad x(t) \leqslant \xi_1, \ y(t) \leqslant \xi_2\}) \quad (12b)
\end{aligned}
$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_{t_0 - T/2}^{t_0 + T/2} u(\xi_1 - x(t)) \, u(\xi_2 - y(t)) \, dt \quad (12c)$$

exists for all $(\xi_1, \xi_2) \in \mathbb{R}^2 - \Xi_0$, where Ξ_0 is at most a countable set of straight lines in $\mathbb{R}^2$. The function F_{xy} has all the properties of a bivariate joint distribution function with the exception of right continuity at the discontinuity points. Such a definition extends naturally to $n > 2$ functions.

Let $x(t)$ and $y(t)$ be jointly RM functions. Then each function is RM [72, Theorem 4.1]. In addition, the sum $x(t) + y(t)$ and the product $x(t) y(t)$ are RM, provided that at least one of the functions is bounded [72, Theorem 4.2].

An extension to the multivariate case of the fundamental theorem of expectation (see (6)) can be derived [72, Theorem 4.5]. In particular, if $x(t)$ and $y(t)$ are bounded functions and $x(t + \tau)$ and $y(t)$ are jointly RM for every τ, then the temporal cross-correlation function [106] of x and y is given by

$$R_{xy}(\tau) \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{t_0 - T/2}^{t_0 + T/2} x(t + \tau) \, y(t) \, dt \quad (13a)$$

$$= \int_{\mathbb{R}^2} \xi_1 \, \xi_2 \, dF_{xy}(\xi_1, \xi_2; \tau) \quad (13b)$$

where

$$F_{xy}(\xi_1, \xi_2; \tau)$$
$$\triangleq \mu_R \left(\{ t \in \mathbb{R} \, : \, x(t + \tau) \leqslant \xi_1, y(t) \leqslant \xi_2 \} \right). \quad (14)$$

E. ABSENCE OF σ-LINEARITY OF THE INFINITE-TIME AVERAGE

As a consequence of the absence of σ-additivity of the relative measure μ_R, although the corresponding expectation operator is linear, it is not σ-linear.

The infinite-time average of the linear combination of a finite number of jointly RM (not necessarily bounded) functions of time is equal to the linear combination of the time averages [24, Theorem 2.7]. However, this result cannot be extended to the case of a countable infinity of functions of time.

For example, accounting for the identity

$$\cos^2(t) = \sum_{k=-\infty}^{+\infty} \cos^2(t) \mathbf{1}_{[k,k+1)}(t) \quad (15)$$

we have

$$\frac{1}{2} = \left\langle \cos^2(t) \right\rangle_t = \left\langle \sum_{k=-\infty}^{+\infty} \cos^2(t) \mathbf{1}_{[k,k+1)}(t) \right\rangle_t$$

$$\neq \sum_{k=-\infty}^{+\infty} \left\langle \cos^2(t) \mathbf{1}_{[k,k+1)}(t) \right\rangle_t = 0 \quad (16)$$

since $\left\langle \cos^2(t) \mathbf{1}_{[k,k+1)}(t) \right\rangle_t = 0$.

This result is different from the corresponding one in the stochastic approach where the expectation operator is σ-linear, provided that the underlying infinite series of random variables is absolutely convergent [66].

As for the absence of σ-additivity of the relative measure μ_R, the absence of σ-linearity of the corresponding
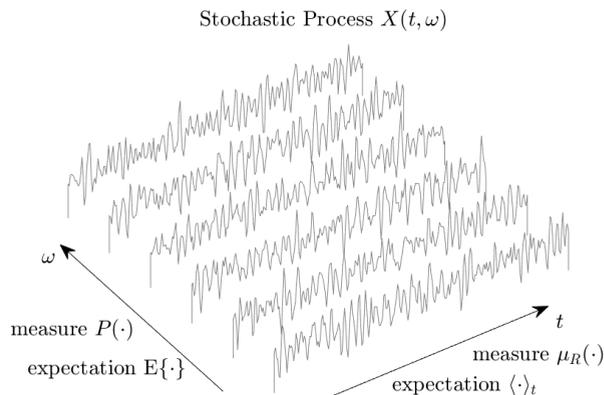
FIGURE 2. Along $\omega \in \Omega$, the probability measure $P(\cdot)$ is σ-additive and the expectation operator, i.e., the ensemble average $E\{\cdot\}$, is σ-linear. Along $t \in \mathbb{R}$, the relative measure $\mu_R(\cdot)$ is not σ-additive and the expectation operator, i.e., the infinite-time average $\langle \cdot \rangle_t$, is not σ-linear.

expectation operator, the infinite-time average, could initially appear to be a weakness of μ_R in comparison with P; however, it actually constitutes, a *motivation* to adopt the FOT approach for signal analysis instead of the classical stochastic-process approach. In fact, the realizations of a stochastic process do not necessarily exhibit properties analogous to those of the stochastic process (Fig. 2).

The absence of σ-linearity of the expectation operator in the FOT approach is illustrated by a suitable example also with reference to the most general case, the almost-periodic component extraction operator, in Section VI-A.

In Appendix I, with reference to the signal considered in Section VI-A, it is shown that the absence of σ-linearity does not necessarily prevent one from performing calculations of infinite time averages; it simply prevents one from using the short cut of interchanging infinite time average and infinite summation operations.

F. CONDITIONAL RELATIVE MEASURABILITY AND INDEPENDENCE

Let A and B be Lebesgue measurable sets and $\{B_n\}$ be an arbitrary increasing sequence of Lebesgue measurable subsets of B with $0 < \lim_{n \to \infty} \mu(B_n)/n < \infty$, $\cup_{n \in \mathbb{N}} B_n = B$. The conditional relative measure of the set A given B is defined by [72, Def. 5.1]

$$\mu_R(A|B) \triangleq \lim_{n \to \infty} \frac{\mu(A \cap B_n)}{\mu(B_n)} \quad (17)$$

provided that the limit exists. In such a case, it is independent of the particular choice of the set sequence $\{B_n\}$ [72, Theorem 5.1].

Let the sets A and B be such that $\mu_R(A|B)$ exists and A is RM. The sets A and B are said to be *independent* if and only if $\mu_R(A|B) = \mu_R(A)$ [72, Def. 5.3].

Consider the definitions $A(\xi_1) \triangleq \{ t \in \mathbb{R} : x(t) \leqslant \xi_1 \}$, with $x(t)$ RM, and $B(\xi_2) \triangleq \{ t \in \mathbb{R} : y(t) \leqslant \xi_2 \}$, with $y(t)$ Lebesgue measurable. Assume that $\forall (\xi_1, \xi_2) \in \mathbb{R}^2 - \Xi_0$, where Ξ_0 is at most a countable set of lines, $\mu_R(A(\xi_1)|B(\xi_2))$ exists. The

functions $x(t)$ and $y(t)$ are *defined* to be *independent* if and only if $\mu_R(A(\xi_1)|B(\xi_2)) = \mu_R(A(\xi_1)) \; \forall(\xi_1, \xi_2) \in \mathbb{R}^2 - \Xi_0$.

Let $x(t)$ and $y(t)$ be jointly RM. In [72, Theorem 5.2] it is *proved* that the functions $x(t)$ and $y(t)$ are independent if and only if, $\forall(\xi_1, \xi_2) \in \mathbb{R}^2$ except at most a countable set of straight lines, we have the equality

$$F_{xy}(\xi_1, \xi_2) = F_x(\xi_1) \, F_y(\xi_2). \tag{18}$$

As an example, two sine waves with incommensurate periods are independent [60], [62].

If $x(t)$ and $y(t)$ are independent, then, for every ξ_1 and ξ_2,

$$
\begin{aligned}
&\mu_R(A(\xi_1)|B(\xi_2)) \\
&\triangleq \lim_{n \to \infty} \frac{\mu(A(\xi_1) \cap B_n(\xi_2))}{\mu(B_n(\xi_2))} \\
&\equiv \lim_{n \to \infty} \frac{\mu(A(\xi_1) \cap [-n/2, n/2])}{n} \\
&\triangleq \mu_R(A(\xi_1)).
\end{aligned}
\tag{19}
$$

That is, the normalization of $\mu(A(\xi_1))$ to obtain a relative measure can be made by considering either subsets $B_n(\xi_2)$ of the set $B(\xi_2)$ built from $y(t)$ or subsets $B_n = [-n/2, n/2]$ (not depending on $y(t)$). Moreover, the subsets $B_n(\xi_2)$, can be arbitrary [72, Theorem 5.1], provided that they satisfy the conditions preceding (17). In other words, the function $y(t)$ from which the normalizing sets $B_n(\xi_2)$ are constructed, has no influence on the relative measure $\mu_R(A(\xi_1)|B(\xi_2))$ and such a relative measure equals $\mu_R(A(\xi_1))$. This result is in agreement with the intuitive concept of independence of two functions or signals in the sense that they have no link with each other.

The intuitive interpretation of the definition of independence in the FOT probability framework has no counterpart in the stochastic process framework where independence of processes is *defined* as the factorization of the joint distribution function into the product of the marginal ones [13], [27], [66]. In contrast, in the FOT probability framework such a factorization is proved to be true as a consequence of the intuitive definition in terms of conditional relative measurability [72, Theorem 5.2].

Note that, in classical probability theory, several authors define independence of events A and B by the condition $P(A|B) = P(A)$, which is equivalent to the condition $P(A, B) = P(A) \, P(B)$ due to the definition $P(A|B) \triangleq P(A, B)/P(B)$ of conditional probability. However, such a definition of independence, unlike the definition in the FOT approach, does not involve the sequence of subsets B_n arbitrarily constructed starting from B (provided that they satisfy the conditions preceding (17)). Hence, the definition of independence in the classical probability theory does not lead to an intuitive interpretation of independence as does that in the FOT approach.

The concept of independent functions has been considered in [60] and [62], [100], and [101], where equation (18) is taken to be the *definition of independence* and, consequently, no link with an intuitive concept of independence is established. A similar point of view is adopted in [77], [78]. In [51],

a condition equivalent to the joint relative measurability is considered before defining independence as the factorization of the joint distribution into the product of the marginals.

G. CENTRAL LIMIT THEOREM

If a sequence of independent zero-mean time series $\{\varphi_k(t)\}_{k \in \mathbb{N}}$ satisfies some mild regularity assumptions, then the FOT distributions of the functions

$$x_n(t) \triangleq \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varphi_k(t) \tag{20}$$

in this sequence approach a zero-mean normal distribution as $n \to \infty$ [24, Theorem 3.5]. That is, we have the FOT Central Limit Theorem (CLT)

$$
\begin{aligned}
&\lim_{n \to \infty} \mu_R (\{t \in \mathbb{R} \,:\, a < x_n(t) \leqslant b\}) \\
&\quad = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{\xi^2}{2\sigma^2}} \, \mathrm{d}\xi
\end{aligned}
\tag{21}
$$

with σ^2 equal to the average-over-k value of the FOT variances of the functions $\varphi_k(t)$.

The proof of this FOT CLT is based on the Taylor series expansion of the characteristic functions of $x_n(t)$ similarly to the proof of the CLT in the classical stochastic approach. The proof in the FOT approach, however, is more challenging due to the presence of the limit operation in the definition of the relative measure μ_R, which limit is not present in the definition of probability measure P.

The exploitation of the FOT CLT theorem allows one to overcome some difficulties that arise in the classical stochastic-process approach in the derivation of a widely adopted model for a communication channel [24, Section VI]. In the stochastic approach, a multipath Doppler channel is shown to introduce normally distributed gains under mild assumptions on the input signal and the channel characteristics [32, Chap. 9.2-9.5, pp. 334-359], [96, Chap. 14-1, pp. 759-762], The justification, however, is only heuristic. In fact, in all the justifications, the input/output relationship of the channel is described in terms of deterministic signals and systems. Then, a stochastic model, whose statistical behavior should reproduce the time behavior of the deterministic model, is heuristically constructed.

In contrast, in the FOT approach in [24, Section VI], it is shown that the multipath Doppler channel introduces a time-varying gain which is RM with normal FOT distribution when the length of the observation interval approaches infinity and the number of paths approaches infinity. The order of these two limit operations cannot be interchanged. Moreover, even for a moderate number of paths, the distribution is almost normal, provided that the observation interval is sufficiently large.

H. WOLD's ISOMORPHISM

In [112], Wold constructs a discrete-time stochastic process whose sample paths are time-shifted versions of a single

time series, and he thereby establishes an isomorphism whose continuous-time counterpart is

$$X(t, s) = x(t - s) \qquad t \in \mathbb{R}, \ s \in \mathbb{R}. \tag{22}$$

In such a case $\Omega = \mathbb{R}$. Thus, in order to have $P(\Omega) = 1$ we must have $P(\cdot) \equiv \mu_{\mathbb{R}}(\cdot)$. This is an example of probability space for which measure normalization involves a limit operation. This fact, however, is not in agreement with the fundamental assumption that the probability space has finite measure (even before normalization), which assumption is made in the classical construction of a probability space [48, Sec. 5]. Therefore, Wold's isomorphism (for discrete time or extended to continuous time) is not compatible with the classical definition of a stochastic process.

Kolmogorov, in [66, p. 15], says that Axiom VI is necessary when the field is infinite. He does not point out the difference, at least in [66], between the case of infinite measure before normalization (e.g., the sample space is the real line) and the case of finite measure before normalization (e.g., the sample space is any finite interval). If the sample space has infinite measure before normalization, then the normalization $P(\Omega) = 1$ is obtained by a limit operation as for the relative measure. In such a case, the probability measure is not σ-additive. Adding Axiom VI adds the σ-additivity property, and this is generally allowable when the sample space is not specified and also allowable when it is specified and does not contradict this axiom. But, of course, Axiom VI is not allowable when the sample space is specified and contradicts the axiom. For example, the sample space specified in Wold's isomorphism requires that the probability measure be normalized through a limiting process, and this normalization does not allow the measure to be σ-additive.

Wold's isomorphism is extended in [37], [43] to cyclostationary time series. In such a case, a stochastic process is constructed such that its sample paths are time-shifted versions of a single time series with time shifts that are integer multiples of the period of cyclostationarity T_0. That is,

$$X(t, s) = x(t - s) \quad t \in \mathbb{R}, \ s = nT_0, \ n \in \mathbb{Z}. \tag{23}$$

In [56], the details of Wold's isomorphism between cyclostationary stochastic sequences and cyclostationary numerical sequences are presented. It is shown how Hilbert-space representations of cyclostationary stochastic sequences are interpreted in the case of numerical cyclostationary sequences.

IV. ALMOST-PERIODIC PHENOMENA

For every time series $x(t)$ for which the sinusoidally weighted time average

$$x_\eta \triangleq \left\langle x(t) e^{-j2\pi\eta t} \right\rangle_t \tag{24}$$

exists $\forall \eta \in \mathbb{R}$, there exists the decomposition

$$x(t) = \sum_{\eta \in E_1} x_\eta e^{j2\pi\eta t} + x_r(t) \tag{25}$$

where E_1 is the countable set of frequencies η such that $x_\eta \neq 0$ and the residual term $x_r(t)$ does not contain any finite-strength additive sine-wave component

$$\left\langle x_r(t) e^{-j2\pi\eta t} \right\rangle_t = 0 \quad \forall \eta \in \mathbb{R}. \tag{26}$$

The periodic component with period T_0 contained in the time series $x(t)$ can be extracted by the synchronized averaging theorem [34], [37]

$$x_{T_0}(t) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(t - nT_0) \tag{27a}$$

$$= \sum_{k=-\infty}^{+\infty} x_{k/T_0} e^{j2\pi(k/T_0)t} \tag{27b}$$

provided that the Fourier series is absolutely convergent. The periodic component extraction operator is generalized by the almost-periodic (AP) component extraction operator $\mathrm{E}^{\{\alpha\}}\{\cdot\}$, that is, the operator that extracts all the finite-strength additive sine-wave components of the function in its argument. If the AP component is poly-periodic with incommensurate periods $T_1, \ldots, T_P$, then

$$\mathrm{E}^{\{\alpha\}}\{x(t)\} = \langle x(t) \rangle_t + \sum_{p=1}^{P} \left[x_{T_p}(t) - \langle x(t) \rangle_t \right] \tag{28}$$

where each periodic component $x_{T_p}(t)$ is given by (27b) with T_0 replaced by T_p.

From decomposition (25) we have

$$\mathrm{E}^{\{\alpha\}}\{x(t)\} = \mathrm{E}^{\{\alpha\}} \left\{ \sum_{\eta \in E_1} x_\eta e^{j2\pi\eta t} + x_r(t) \right\} \tag{29a}$$

$$= \sum_{\eta \in E_1} x_\eta e^{j2\pi\eta t}. \tag{29b}$$

In [43] it is shown that, under mild assumptions on the real valued time series $x(t)$, for every fixed t the function of ξ

$$F_{x(t)}^{\{\alpha\}}(\xi) \triangleq \mathrm{E}^{\{\alpha\}} \{\mathrm{u}(\xi - x(t))\} \tag{30}$$

is a valid cumulative distribution function except for the right-continuity property (at the discontinuity points) (Fig. 3). For a further proof see also [90, Chap. 2].

The cumulative distribution function is almost-periodic by construction. It can be expressed by its Fourier series

$$F_{x(t)}^{\{\alpha\}}(\xi) = \sum_{\gamma \in \Gamma_1} F_x^\gamma(\xi) e^{j2\pi\gamma t} \tag{31}$$

where Γ_1 is a countable set and the complex-valued Fourier coefficients are (Fig. 4)

$$F_x^\gamma(\xi) \triangleq \left\langle \mathrm{u}(\xi - x(t)) e^{-j2\pi\gamma t} \right\rangle_t \tag{32}$$

For $\gamma = 0$ the coefficient $F_x^0(\xi)$ is coincident with the FOT distribution (4c). In addition, $E_1 \subseteq \Gamma_1$ and

$$x_\eta = \int_{\mathbb{R}} \xi \, \mathrm{d}F_x^\eta(\xi). \tag{33}$$
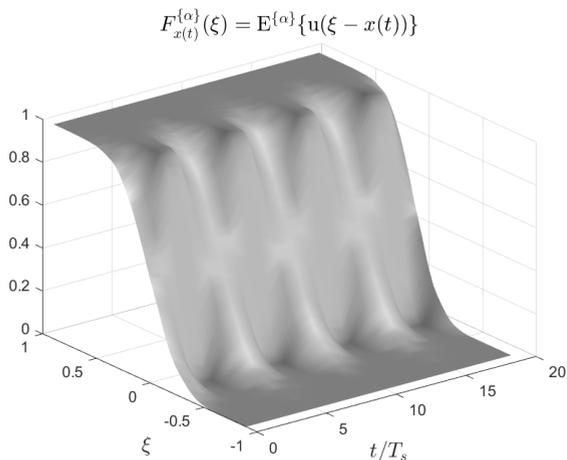
FIGURE 3. Periodically time-variant cumulative distribution function for a double-sideband amplitude-modulated time series.

$$F_{x(t)}^{\{\alpha\}}(\xi) = \mathrm{E}^{\{\alpha\}}\{\mathrm{u}(\xi - x(t))\}$$

FIGURE 5. Magnitude of the Fourier coefficients of the periodically time-variant probability density function for a double-sideband amplitude-modulated time series.

$$f_x^{\gamma}(\xi) = \left\langle \delta(\xi - x(t))\, e^{-j2\pi\gamma t} \right\rangle_t$$

FIGURE 4. Magnitude of the Fourier coefficients of the periodically time-variant cumulative distribution function for a double-sideband amplitude-modulated time series.

$$F_x^{\gamma}(\xi) = \left\langle \mathrm{u}(\xi - x(t))\, e^{-j2\pi\gamma t} \right\rangle_t$$

Moreover, for a well-behaved function $g(\cdot)$, the following *fundamental theorem of expectation* can be proved [43]:

$$\mathrm{E}^{\{\alpha\}}\{g(x(t))\} = \int_{\mathbb{R}} g(\xi)\, \mathrm{d}F_{x(t)}^{\{\alpha\}}(\xi) \tag{34}$$

Equation (34) is the FOT counterpart of the fundamental theorem of expectation in the stochastic approach (9) with $F_X(\xi)$ therein replaced by a time-variant $F_X(\xi; t)$.

From (30) and (34) it follows that $\mathrm{E}^{\{\alpha\}}\{\cdot\}$ is the expectation operator corresponding to the almost-periodically time-variant distribution $F_{x(t)}^{\{\alpha\}}(\xi)$. In addition, the almost-periodic probability density function (pdf) of a time series $x(t)$ can be formally written as

$$f_{x(t)}^{\{\alpha\}}(\xi) \triangleq \frac{\mathrm{d}}{\mathrm{d}\xi} F_{x(t)}^{\{\alpha\}}(\xi) \tag{35a}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\xi} \mathrm{E}^{\{\alpha\}}\{\mathrm{u}(\xi - x(t))\} \tag{35b}$$

$$= \mathrm{E}^{\{\alpha\}}\{\delta(\xi - x(t))\} \tag{35c}$$

where $\delta(\cdot)$ denotes Dirac's delta. It can be expressed by its Fourier series

$$f_{x(t)}^{\{\alpha\}}(\xi) = \sum_{\gamma \in \Gamma_1} f_x^{\gamma}(\xi)\, e^{j2\pi\gamma t} \tag{36}$$

whose convergence is in the sense of generalized functions [114] and the complex-valued Fourier coefficients are given by (Fig. 5)

$$f_x^{\gamma}(\xi) = \frac{\mathrm{d}}{\mathrm{d}\xi} F_x^{\gamma}(\xi) = \left\langle \delta(\xi - x(t))\, e^{-j2\pi\gamma t} \right\rangle_t. \tag{37}$$

The formal expressions (35c)–(37) allow to prove (34) by formal manipulation of Dirac's delta [43], [90, Sec. 2.3].

In [37, Part II], [43], [44], an extension of the FOT approach from stationary time series to time series that exhibit cyclostationarity is constructed; this class of time series is comprised of all those that are cyclostationary, or polycyclostationary, or almost cyclostationary. Such time series, by definition, have time-variant FOT distribution functions that are periodic, polyperiodic, or almost periodic.

The result that (30) is a valid distribution function is not obvious and opens a new and wider perspective on the FOT approach. Specifically, periodically, poly-periodically, or almost-periodically time-variant probabilistic functions can be constructed starting from the unique observed time series at the hand of the experimenter. This completely circumvents the typically unjustifiable assumption that there exists a stochastic process model that is stationary and ergodic and for which the time series at hand is a sample path.

Furthermore, periodic phenomena–those obtained from combinations of periodic or almost-periodic mechanisms and random phenomena–can now be modeled without incurring the dichotomy between abstract properties of stochastic processes and concrete properties of individual sample paths [41]. Examples of this dichotomy are given in Section VI.

For an AP function $x_{\mathrm{ap}}(t)$, we have

$$\mathrm{E}^{\{\alpha\}}\{x_{\mathrm{ap}}(t)\} = x_{\mathrm{ap}}(t). \tag{38}$$

That is, the AP component extraction operator, applied to an AP function, produces the AP function itself. In addition, according to (35c), for an AP function, the time-varying probability density function above the (ξ, t) plane is an impulse fence

$$\begin{aligned} f_{x_{\mathrm{ap}}(t)}^{\{\alpha\}}(\xi) &= \mathrm{E}^{\{\alpha\}}\left\{\delta(\xi - x_{\mathrm{ap}}(t))\right\} \\ &= \delta(\xi - x_{\mathrm{ap}}(t)). \end{aligned} \tag{39}$$

Since the AP component extraction operator is the expectation operator in the AP FOT probability framework, it follows from (38) that the AP functions are the *deterministic* signals in the AP FOT probability framework. All other signals are the *random* signals. Note that *the term "random" here is not intended to be synonymous with "stochastic".* In fact, the adjective *stochastic* is adopted, as usual, when an ensemble of realizations or sample paths exists, whereas the adjective *random* is used in reference to a single function of time.

In other words, decomposition (25) can be interpreted as the decomposition of a generic random signal $x(t)$ into its deterministic (that is, AP) component $x_{\mathrm{ap}}(t)$ and a residual component $x_{\mathrm{r}}(t)$

$$x(t) = x_{\mathrm{ap}}(t) + x_{\mathrm{r}}(t). \tag{40}$$

A detailed decomposition for a class of functions that suitably models communication signals is presented in the Appendix of [9].

Decomposition (40) is the FOT counterpart of the classical decomposition of a stochastic process into the sum of its mean value and a zero-mean (centered) stochastic process.

Starting from decomposition (40) an analogous decomposition can be determined for the impulse-response function of a linear time-variant system [58].

It is worth illustrating with a concrete example that a single time series can admit more than one FOT probability model, depending on which cycle frequencies, if any, are to be recognized. Here we consider a signal that admits two distinct FOT probabilistic descriptions. The time series

$$c(t) = \cos(2\pi f_0 t + \phi_0) \tag{41}$$

is random in the stationary FOT framework and has FOT density (Fig. 6)

$$f_c(\xi) = \frac{1}{\pi} \frac{1}{\sqrt{1 - \xi^2}} \operatorname{rect}(\xi/2) \tag{42}$$

whereas it is deterministic in the almost-periodic FOT framework and, according to (39), has FOT density (Fig. 7)

$$f_{c(t)}^{\{\alpha\}}(\xi) = \delta(\xi - \cos(2\pi f_0 t + \phi_0)). \tag{43}$$

The important consequences of making a specific choice among the three options of stationary, cyclostationary, and almost-cyclostationary models for a time series is thoroughly discussed in [37, Chap. 15], [41, Chap. 1], and [88, Sec. 6.4].
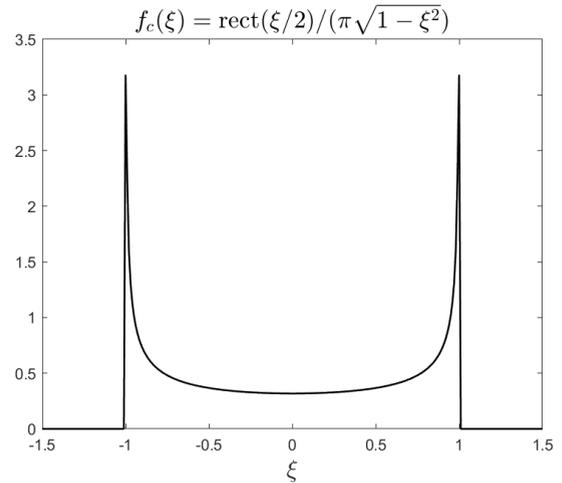
FIGURE 6. Stationary FOT pdf of the time series $c(t) = \cos(2\pi f_0 t + \phi_0)$.
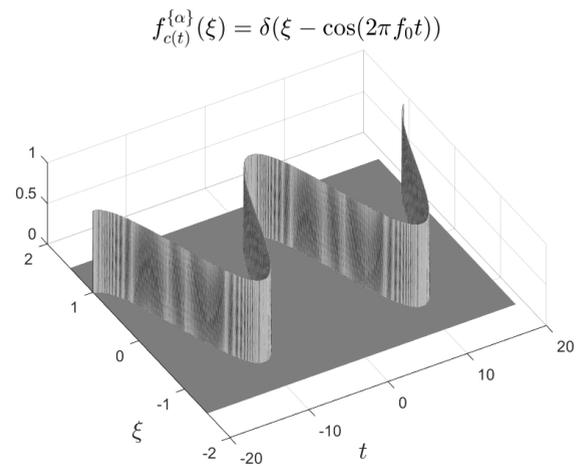
FIGURE 7. Periodically time-variant FOT pdf of the time series $c(t) = \cos(2\pi f_0 t + \phi_0)$ with $\phi_0 = 0$.

In Section VI-B, a presented example shows how a proper choice of the FOT model enables one to avoid stochastic-process discrepancies existing between properties of a stochastic process model and those of its sample paths.

By adopting the almost-periodic component extraction operator as an expectation operator, a valid second-order almost-periodically time-variant joint distribution

$$F_{x(t+\tau)\,x(t)}^{\{\alpha\}}(\xi_1, \xi_2) \triangleq \mathrm{E}^{\{\alpha\}}\left\{\mathrm{u}(\xi_1 - x(t+\tau))\,\mathrm{u}(\xi_2 - x(t))\right\} \tag{44}$$

and a valid almost-periodically time-variant autocorrelation function

$$\mathrm{E}^{\{\alpha\}}\{x(t+\tau)\,x(t)\} \tag{45a}$$

$$\triangleq \int_{\mathbb{R}^2} \xi_1\,\xi_2\,\mathrm{d}F_{x(t+\tau)\,x(t)}^{\{\alpha\}}(\xi_1, \xi_2) \tag{45b}$$

$$= \sum_{\alpha \in A} R_x^{\alpha}(\tau)\,e^{j2\pi\alpha t} \tag{45c}$$

can be defined [37], [43], where A is a countable set of (possibly incommensurate) *cycle frequencies* α and

the Fourier coefficients

$$R_x^\alpha(\tau) \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} x(t+\tau) \, x(t) \, e^{-j2\pi\alpha t} \, \mathrm{d}t \quad (46)$$

are the *cyclic autocorrelation functions*.

The finite-time Fourier transform

$$X_{\frac{1}{\Delta f}}(t,f) \triangleq \int_{t-1/2\Delta f}^{t+1/2\Delta f} x(s) \, e^{-j2\pi fs} \, \mathrm{d}s \quad (47)$$

at a particular frequency value f is the spectral component of the time series $x(t)$ at frequency f with approximate finite-bandwidth Δf. Thus, the *cyclic spectrum*

$$S_x^\alpha(f) \triangleq \lim_{\Delta f \to 0} \lim_{T \to \infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} \Delta f \, X_{\frac{1}{\Delta f}}(t,f)$$
$$\cdot X_{\frac{1}{\Delta f}}^*(t,f-\alpha) \, \mathrm{d}t \quad (48)$$

represents the temporal correlation of two spectral components separated by the quantity α when the bandwidth Δf becomes infinitesimal. For an almost-cyclostationary signal, $S_x^\alpha(f)$ is non zero for only $\alpha \in A$. That is, only spectral components that are separated by quantities equal to one of the cycle frequencies are correlated. In addition, the following *Gardner relation* [90, pp. 10, 20, 56, 57, 139], originally introduced in [35], [37], holds.

$$S_x^\alpha(f) = \int_{\mathbb{R}} R_x^\alpha(\tau) \, e^{-j2\pi f\tau} \, \mathrm{d}\tau. \quad (49)$$

For $\alpha = 0$ it reduces to the Wiener relation between the power spectrum and the time-average autocorrelation function. For this reason, Gardner originally referred to (49) as *Cyclic Wiener Relation*.

The extension of Wiener's GHA to time series that exhibit cyclostationarity can also be achieved by defining cyclostationarity as the property of time series that enables the regeneration of additive finite-strength sine wave components from hidden periodicities in the time series by using homogeneous quadratic time-invariant transformations [35], [37], [39].

Further approaches for signal analysis based on single functions of time are possible by considering as deterministic signals the polynomial phase signals [31], [85].

V. STATISTICAL FUNCTION ESTIMATION
A. CYCLOERGODICITY
Cycloergodic theory, which extends and generalizes existing ergodic theory, is developed in [16], where it is shown that sinusoidal and periodic components of time-varying probabilistic parameters can be consistently estimated from time averages on one sample path. It is also established that a strict-sense 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 [14] 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 [47, 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, [47, 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. In addition, there is a partially cycloergodic version of this proposed theorem that satisfies the ergodicity condition for some but not all periods.

This leaves one class of ACS processes for which there is so far no known cycloergodic theorem, 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 is not presently known.

B. FOT INTERPRETATION OF THE TASK OF STATISTICAL FUNCTION ESTIMATION

In the stochastic approach, the estimator of the expected value of a wide-sense stationary stochastic process is the time average of one sample path of the process over the finite time interval $[t - T/2, t + T/2]$ with center t and width T. Similarly, in the FOT approach, the estimator of the infinite-time average of a time series is the finite-time average of this time series over $[t - T/2, t + T/2]$.

In the stochastic approach, t is fixed (and typically assumed to be 0 or $T/2$). The estimate is a random variable, that is, it depends on the sample point $\omega \in \Omega$ which determines the sample path or realization of the process. The variability of the estimate reflects its dependence on the sample path used for the estimation. Under appropriate mixing and stationarity assumptions, the estimate converges in some probabilistic sense as $T \to \infty$ to the expected value of the process. In the FOT approach, the variability of the estimate reflects its dependence on t, the central point of the observation interval, when t ranges over a wider temporal interval, say $[-Z/2, Z/2]$, with $Z \gg T$ [37, Chap. 15], [88, Sec. 6.3.5].

In the FOT approach, asymptotic characterization of the convergence of the estimator is expressed in terms of a double limit as $Z \to \infty$ and $T \to \infty$, provided that $Z/T \to \infty$. The limit in Z produces the average-over-t behavior of the estimate, such as the average error (bias) and the average squared deviation of the estimate about its average value (the variance), where t is the central point of the observation interval. The limit in T produces an estimate that uses the whole time series for estimation.

Let $z(t)$ be a RM time series obtained as a frequency-shifted second- or higher-order lag product of another time series $x(t)$. That is,

$$z(t) \triangleq \prod_{k=1}^{K} x(t + \tau_k) \, e^{-j2\pi\alpha t}. \tag{50}$$

Second- and higher-order cyclic moments of $x(t)$ are infinite-time averages of $z(t)$ [44], [99].

The finite-time average

$$\widehat{m}_z^{(T)}(t) \triangleq \frac{1}{T} \int_{t-T/2}^{t+T/2} z(u) \, du \tag{51}$$

is the estimator of the infinite-time average

$$m_z \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} z(u) \, du. \tag{52}$$

For T sufficiently large (much greater than the longest period of cyclostationarity of $z(t)$) the function $t \mapsto \widehat{m}_z^{(T)}(t)$

is approximately (asymptotically exactly) wide-sense stationary in the FOT sense. That is, $\widehat{m}_z^{(T)}(t)$ and its homogeneous nonlinear transformations do not contain any substantive (or, in the limit, non-zero) additive finite-strength sine-wave component with nonzero frequency. Thus, the FOT expectation operator of interest for asymptotic properties of the estimator is the infinite-time average. The performance of the estimator can be expressed in terms of FOT bias and variance [37, Chap. 15]

$$\text{bias}\left\{\widehat{m}_z^{(T)}(t)\right\} \simeq \left\langle \widehat{m}_z^{(T)}(t) \right\rangle_t - m_z = 0 \tag{53}$$

$$\text{var}\left\{\widehat{m}_z^{(T)}(t)\right\} \simeq \left\langle \left|\widehat{m}_z^{(T)}(t) - m_z\right|^2 \right\rangle_t \tag{54}$$

where the two approximations become exact equalities in the limit as $T \to \infty$. Equation (53) shows that in the FOT approach the estimator $\widehat{m}_z^{(T)}(t)$ is unbiased. Moreover, its rate of convergence to zero is arbitrarily fast for every bounded RM function (see [25, Theorem 3.2] and consequences).

Assuming summability of second- and fourth-order temporal cumulants of $z(t)$, the estimator is mean-square consistent in the FOT sense, that is,

$$\lim_{T \to \infty} \left\langle \left|\widehat{m}_z^{(T)}(t) - m_z\right|^2 \right\rangle_t$$

$$= \lim_{T \to \infty} \left[\text{var}\left\{\widehat{m}_z^{(T)}(t)\right\} + \left|\text{bias}\left\{\widehat{m}_z^{(T)}(t)\right\}\right|^2 \right] = 0 \tag{55}$$

In addition, under further temporal cumulant summability assumptions, the function

$$t \mapsto \sqrt{T} \left[\widehat{m}_z^{(T)}(t) - m_z\right] \tag{56}$$

has a normal FOT distribution as $T \to \infty$ [25].

A cyclic spectrum estimator is the time-smoothed cyclic periodogram, which is the right-hand side of (48) with finite T and Δf. The estimate is accurate (small FOT bias) provided that Δf is sufficiently small, and it is reliable (small variance) provided that the smoothing product $T \Delta f$ is sufficiently large [37, Chap. 13]. Moreover, this estimator is asymptotically equivalent to the frequency-smoothed cyclic periodogram [42] which, for $\alpha = 0$, reduces to the frequency-smoothed periodogram for power-spectrum estimation.

VI. EXAMPLES OF THE DICHOTOMY BETWEEN PROPERTIES OF A STOCHASTIC PROCESS AND THOSE OF ITS SAMPLE PATHS

In this section, three examples of the dichotomy between properties of a stochastic process and those of its sample paths are illustrated. Specifically, properties that are valid for stochastic processes are shown to be not valid for their sample paths. Such drawbacks of the stochastic-process based models constitute a strong motivation for the adoption of the FOT approach.

A. EXAMPLE 1 OF DICHOTOMY: EXPECTATION OF A PAM TIME SERIES

Let us consider a pulse-amplitude modulated (PAM) stochastic process

$$\xi(t) \triangleq \sum_{n=-\infty}^{+\infty} \widetilde{\xi}_n \, q(t - nT_0) \tag{57}$$

where $\{\widetilde{\xi}_n\}_{n \in \mathbb{Z}}$ is a sequence of random variables with equiprobable values in a finite alphabet and the pulse $q(t) \in L^1(\mathbb{R})$.

Under the assumption

$$\sup_t \sum_{n=-\infty}^{+\infty} \mathrm{E}\left\{|\widetilde{\xi}_n|\right\} \, |q(t - nT_0)| < \infty \tag{58}$$

the σ-linearity of the statistical expectation operator $\mathrm{E}\{\cdot\}$, validates the interchange of the infinite summation and expectation operations, and the expected value of $\xi(t)$ is given by

$$\mathrm{E}\{\xi(t)\} = \sum_{n=-\infty}^{+\infty} \mathrm{E}\left\{\widetilde{\xi}_n\right\} \, q(t - nT_0). \tag{59}$$

Such an interchange of expectation operator and infinite summation is not allowed on the single sample paths (after having replaced the ensemble mean by the infinite time average). Thus, it is not allowed in the FOT approach.

Let us consider the PAM time series

$$x(t) = \sum_{n=-\infty}^{+\infty} \widetilde{x}_n \, q(t - nT_0) \tag{60}$$

with $\widetilde{x}_n$ a discrete-time time series. Even if

$$\sup_t \sum_{n=-\infty}^{+\infty} |\widetilde{x}_n \, q(t - nT_0)| < \infty \tag{61}$$

for the time series $x(t)$ in general we have

$$\langle x(t) \rangle_t \neq \sum_{n=-\infty}^{+\infty} \widetilde{x}_n \, \langle q(t - nT_0) \rangle_t \equiv 0 \tag{62}$$

$$\mathrm{E}^{\{\alpha\}}\{x(t)\} \neq \sum_{n=-\infty}^{+\infty} \widetilde{x}_n \, \mathrm{E}^{\{\alpha\}}\{q(t - nT_0)\} \equiv 0 \tag{63}$$

since $q(t) \in L^1(\mathbb{R})$. Such a result is in accordance with the fact that the infinite-time average $\langle \cdot \rangle_t$ and the almost-periodic component extraction operator $\mathrm{E}^{\{\alpha\}}\{\cdot\}$ are not σ-linear.

In [90, Sec. 2.4.1], it is shown that by properly executing the almost-periodic component extraction operation, the FOT expectation of $x(t)$, which is its almost-periodic component, can be shown to be given by

$$\mathrm{E}^{\{\alpha\}}\{x(t)\} = \sum_{n=-\infty}^{+\infty} \mathrm{E}^{\{\widetilde{\alpha}\}}\{\widetilde{x}_n\} \, q(t - nT_0) \tag{64}$$

where $\mathrm{E}^{\{\widetilde{\alpha}\}}\{\cdot\}$ denotes the *discrete-time* almost-periodic component extraction operator.

Comparison of the procedures used to obtain the expected value (59) of a PAM process and its FOT counterpart (64) shows the required difference in ways of executing the expectation operator in the stochastic and FOT approaches. In fact, even if a summability condition for the discrete-time sequence is satisfied, the FOT expectation operation cannot be freely interchanged with the infinite-summation operation. Thus, the result (59) valid for the stochastic process cannot be immediately extended to its sample paths by replacing the ensemble mean with the infinite time average. In contrast, the result derived in the FOT approach is valid by default for the unique time series at the hand of the experimenter. As an example, the calculation of the infinite time average in the left-hand side of (62) is reported in Appendix I.

Note that every digitally modulated signal, and in particular the PAM process (57), can be modeled as a stochastic process whose sample space (before normalization) has finite measure. Therefore, the existence of a finite-measure sample space is not sufficient to avoid the above-mentioned dichotomy between a stochastic process and its sample paths.

The fact that digitally modulated signals can be modeled as stochastic processes with finite sample space follows from the subsequent considerations. The sequences of symbols belonging to a finite alphabet of size D can easily be modeled as sample paths of a discrete-time stochastic process whose sample space has finite measure and is isomorphic to the interval $[0, 1]$. In fact, every finite or infinite sequence can be seen as the expansion in base D of a number ω taken in the unit interval that plays the role of the sample space Ω. The probability measure is the Lebesgue measure. Sequences corresponding to the rational numbers have zero Lebesgue measure (they occur with probability zero). Typical sequences correspond to Borel's normal numbers, that is, those satisfying the weak law of large numbers: Each letter of the alphabet occurs in the sequence with relative frequency $1/D$ and typical sequences occur with probability one.

Therefore, every digitally modulated signal can be modeled as a stochastic process whose sample space (before normalization) has finite measure since its randomness is due to the randomness of the modulating sequence. This is true even if other parameters are modeled as random variables.

B. EXAMPLE 2 OF DICHOTOMY: LINEAR TRANSFORMATION OF A GAUSSIAN SIGNAL

Let us consider the stochastic process

$$X(t) = G(t) \, \cos(2\pi f_0 t + \phi_0) \tag{65}$$

where $G(t)$ is a zero-mean strict-sense stationary Gaussian stochastic process and f_0, and ϕ_0 are deterministic parameters. The process $X(t)$ is a periodically time-variant linear transformation of a Gaussian process. Thus, it is in turn Gaussian. Due to the periodic modulation, $X(t)$ is strict-sense cyclostationary with period of cyclostationarity $T_0 = 1/(2f_0)$. In particular, its variance is a periodic function of t with period T_0.

Let

$$x(t) = g(t) \cos(2\pi f_0 t + \phi_0) \tag{66}$$

be a single realization of the stochastic process $X(t)$, where $g(t)$ is a single realization of the stochastic process $G(t)$.

Albeit the stochastic process $X(t)$ is Gaussian, its realizations do not have an empirical (stationary) Gaussian distribution. In fact, the time series $g(t)$ and $c(t) \triangleq \cos(2\pi f_0 t + \phi_0)$ have stationary FOT densities

$$f_g(\xi) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\xi^2/2\sigma_g^2} \tag{67}$$

and $f_c(\xi)$ given in (42), respectively. Since the functions $g(t)$ and $x(t)$ are independent in the FOT sense [72, Sec. 5], the FOT density of the product waveform $x(t) = g(t)c(t)$ is given by the classical formula [2, Chap. 2, Problem 14]

$$f_x(\xi) = \int_{-\infty}^{+\infty} f_c(s) f_g(\xi/s) \frac{1}{|s|} \, ds \tag{68}$$

which is a non-Gaussian density. It is the stationary FOT density of $x(t)$ which is coincident with the (stationary) empirical density.

In Fig. 8, results are reported for a band-limited zero-mean Gaussian time series $g(t)$ with bandwidth $1/(64T_s)$, where T_s is the sampling period, which modulates a cosine with $f_0 = 1/(16T_s)$ and $\phi_0 = 0$. The thin line represents theoretical non-Gaussian density obtained by numerically evaluating the integral in (68) as predicted by the FOT model; the dotted line is the density estimated from the data by a kernel-based estimator (with Gaussian kernel); [97, Sec. 2.1.8, pp. 64-65]; the dashed line represents a zero-mean Gaussian density with variance equal to the sample variance estimated from $x(t)$. The data-record length is $T = KT_0$ with $K = 2^{10}$.

In the above result, the stationary FOT model of Section III-C is considered. A more appropriate statistical description of the time series $x(t)$ can be obtained by adopting the almost-periodic FOT model of Section IV (which includes periodic models as a special case).

In Fig. 9 the periodically time-variant distribution estimated by the synchronized average

$$\widehat{F}_{x(t)}^{\{\alpha\}}(\xi) = \frac{1}{K} \sum_{k=0}^{K-1} u(\xi - x(t - kT_0)) \tag{69}$$

is reported for $t = nT_s$, $n = 0, 1, \ldots, 15$. Moreover, asterisks represent the theoretical zero-mean Gaussian distribution

$$\Phi(\xi) \triangleq 1 - \frac{1}{2} \operatorname{erfc}\left(\frac{\xi}{\sqrt{2}\sigma}\right) \tag{70}$$

numerically evaluated in correspondence with the periodically time-variant variance estimated by the synchronized average

$$\widehat{\sigma}_x^2(t) = \frac{1}{K} \sum_{n=0}^{K-1} x^2(t - nT_0). \tag{71}$$
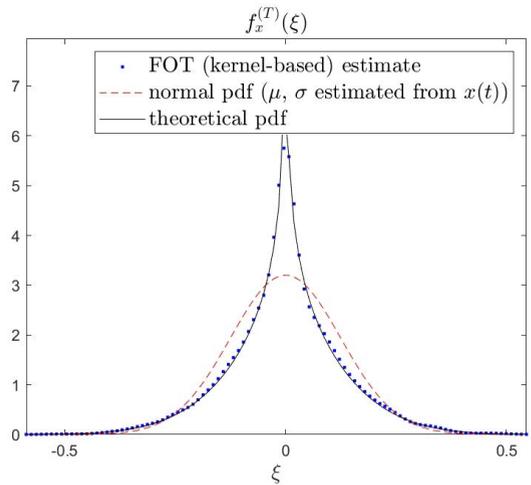
FIGURE 8. Non-Gaussian probability density function of a modulated Gaussian time series as correctly predicted by the stationary FOT model (thin line = theoretical value, dotted line = kernel-based estimate) versus the zero-mean Gaussian density with variance measured from x(t) (dashed line).
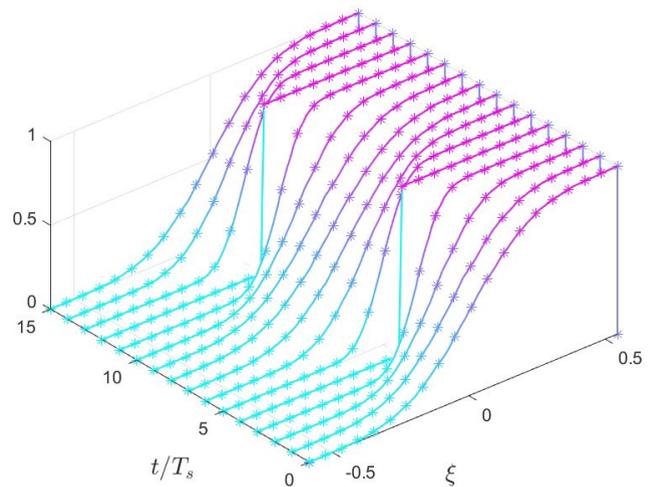
FIGURE 9. Periodically time variant Gaussian distribution of a modulated Gaussian time series as correctly predicted by the periodic FOT model. Smooth lines: Estimated cumulative distribution function for different fixed values of t (see (69)). Asterisks: Gaussian cumulative distribution evaluated with an estimated periodically time-variant variance (see (70), (71)).

The estimated periodically time-variant distribution (69) exactly fits the Gaussian distribution (70) with estimated periodically time-variant variance (71). For $t = 4T_s$ and $t = 12T_s$, $\cos(2\pi f_0 t) = 0$ and one has the degenerate distribution $u(\xi)$ obtained from (70) in the limit as $\sigma \to 0$.

No one would call the stochastic process (65) non-Gaussian. In particular, for every set of fixed values $t_1, \ldots, t_n$, the random variables $X(t_1), \ldots, X(t_n)$ are jointly Gaussian. However, the (stationary) empirical density of $X(t)$ approaches the non-Gaussian density (68) as the data-record length approaches infinity. The reason for this is that the appropriate stochastic stationary distribution

for a cyclostationary process is the time average of the stochastic cyclostationary distribution, which is a mixture of non-identical Gaussians and is therefore non-Gaussian; it is the limit of what the stationary empirical distribution produces. The mismatch between the Gaussianity of the random variables $X(t)$ and the non-Gaussianity of the stationary distribution can be overcome by modeling ϕ_0 as a random variable uniformly distributed in $[0, T_0)$. In such a case, the process $X(t)$ is non-Gaussian, strict-sense stationary, with density given by (68), which is also the limit of the empirical density. The price paid for assuming the stationary model is that the stationary process is not cycloergodic. In fact, every realization is a cyclostationary time series. In particular, the Fourier coefficient (32) with $\gamma = k/T_0$, k integer, and the cyclic autocorrelation functions (46) with $\alpha = k/T_0$, $k \in \{0, \pm 1\}$, are not identically zero for almost all realizations.

In summary, in the stochastic process framework, we have the following. If the stochastic process (65) is modeled as Gaussian cyclostationary (ϕ_0 is modeled as a deterministic parameter), then its stationary empirical distribution approaches a non-Gaussian distribution. If the stochastic process (65) is modeled as a stationary process (ϕ_0 is modeled as a random variable uniformly distributed in $[0, T_0)$), then it is non-Gaussian, its stationary density is given by (68), but it is not cycloergodic.

Such a discrepancy, is not present in the FOT approach. The time series $x(t)$ is cyclostationary. That is, its periodically time variant distribution (30) does not degenerate into the stationary distribution (4a)–(4c). If one decides to exploit the stationary model, then $x(t)$ in (66) is non-Gaussian and the empirical distribution approaches the non-Gaussian distribution whose density is given by (68) with (67) and (42) substituted in. If one decides to exploit the cyclostationary model, then $x(t)$ in (66) is Gaussian and the finite-time FOT distribution (69) approaches the periodically time-variant distribution given by (70) with σ replaced by a periodically time-variant standard deviation which is the square root of the limit in (71) as $K \to \infty$. The non-Gaussian density (68) is the coefficient with zero frequency of the Fourier series expansion of the Gaussian periodically time-variant density function.

A linear periodically time-variant transformation of a Gaussian stochastic process is in turn a Gaussian process, but with realizations whose stationary empirical distribution is non Gaussian. Such a dichotomy is not present in the FOT framework. Both stationary and almost-cyclostationary models give rise to probabilistic functions that are fitted by their estimates obtained from the data.

C. EXAMPLE 3 OF DICHOTOMY: STATISTICAL INDEPENDENCE

Let us consider the two stochastic processes

$$X_1(t) \triangleq B(t) \cos(2\pi t/T_1 + \Theta_1) \qquad (72)$$
$$X_2(t) \triangleq B(t) \cos(2\pi t/T_2 + \Theta_2) \qquad (73)$$

where $B(t)$ is the binary PAM stochastic process

$$B(t) \triangleq \sum_{n=-\infty}^{+\infty} B_n \, p(t - nT_0 + \Theta_0) \qquad (74)$$

for which $\{B_n\}_{n \in \mathbb{Z}}$ is a sequence of independent and identically distributed binary random variables assuming values ± 1, $\{\Theta_i\}$ are random variables uniformly distributed in $[0, T_i)$, $i = 0, 1, 2$, and $p(t) = \text{rect}(t/T_0)$. All the random variables are statistically independent.

It is well known that the random variables Θ_i make the random processes $X_1(t)$ and $X_2(t)$ stationary but not cycloergodic [33]. That is, sample paths of $X_1(t)$ and $X_2(t)$ are cyclostationary time series. In the following, it is shown that the stationary model gives rise to a dichotomy between the statistical dependence property of the stochastic processes $X_1(t)$ and $X_2(t)$ and that of their sample paths.

The stochastic processes $X_1(t)$ and $X_2(t)$ are statistically dependent by virtue of the common factor $B(t)$. Specifically, for every fixed $\bar{t}$, the two random variables $X_1(\bar{t})$ and $X_2(\bar{t})$ are obtained as the product of the same random variable $B(\bar{t})$ with the two random variables $\cos(2\pi \bar{t}/T_1 + \Theta_1)$ and $\cos(2\pi \bar{t}/T_2 + \Theta_2)$, respectively. In general we have

$$
\begin{aligned}
&\mathrm{E}\left\{X_1^{r_1}(\bar{t}) X_2^{r_2}(\bar{t})\right\} \\
&\quad = \mathrm{E}\left\{B^{r_1+r_2}(\bar{t})\right\} \mathrm{E}\left\{\cos^{r_1}(2\pi \bar{t}/T_1 + \Theta_1)\right\} \\
&\qquad \cdot \mathrm{E}\left\{\cos^{r_2}(2\pi \bar{t}/T_2 + \Theta_2)\right\} \qquad (75) \\
&\quad \neq \mathrm{E}\left\{X_1^{r_1}(\bar{t})\right\} \mathrm{E}\left\{X_2^{r_2}(\bar{t})\right\} \quad \forall \text{ integer } r_1, r_2. \quad (76)
\end{aligned}
$$

That is, the necessary and sufficient condition for independence is not satisfied.

Let

$$x_1(t) \triangleq b(t) \cos(2\pi f_1 t + \theta_1) \qquad (77)$$
$$x_2(t) \triangleq b(t) \cos(2\pi f_2 t + \theta_2) \qquad (78)$$

be sample paths of $X_1(t)$ and $X_2(t)$, respectively, where $b(t)$ is a sample path of $B(t)$ and θ_1 and θ_2 are realizations of Θ_1 and Θ_2. Almost every sample path $b(t)$ is a *pseudo-random Bass function* [4], [5], [72, Sec. 6.3]. Therefore, it can be shown that $x_1(t)$ and $x_2(t)$ are FOT independent in the sense that their joint FOT stationary distribution factors into the product of the two marginals (see (18)), provided that the periods T_i, $i = 0, 1, 2$, are incommensurate, that is,

$$T_i/T_j \notin \mathbb{Q} \quad i \neq j \qquad (79)$$

where $\mathbb{Q}$ is the set of rational numbers. In fact, in such a case, it is shown in [24, Sec. V.B] that the necessary and sufficient condition for FOT independence [24, Theorem 2.8], [62, Theorem 2],

$$\langle x_1^{r_1}(t) x_2^{r_2}(t)\rangle_t = \langle x_1^{r_1}(t)\rangle_t \langle x_2^{r_2}(t)\rangle_t \quad \forall \text{ integer } r_1, r_2, \qquad (80)$$

is satisfied.

This is an example of a dichotomy between the statistical dependence property of the stationary stochastic processes

$X_1(t)$ and $X_2(t)$ and the statistical independence of their sample paths $x_1(t)$ and $x_2(t)$ according to their time-invariant FOT distributions.

However, if the AP FOT model is adopted, it can be shown that the condition

$$
\begin{aligned}
& \mathrm{E}^{\{\alpha\}} \left\{ x_1^{r_1}(t) \, x_2^{r_2}(t) \right\} \\
& \quad = \mathrm{E}^{\{\alpha\}} \left\{ x_1^{r_1}(t) \right\} \mathrm{E}^{\{\alpha\}} \left\{ x_2^{r_2}(t) \right\} \quad \forall r_1, r_2 \text{ integer} \quad (81)
\end{aligned}
$$

does not hold; that is, these time series are statistically dependent by virtue of their common factor $b(t)$ according to their AP-FOT distributions.

VII. CONCLUDING REMARKS

The fraction-of-time (FOT) probability method for signal analysis is an alternative to the classical probability method that models signals as sample paths of an ensemble which, together with a probability measure, is called a stochastic process. In the FOT approach, the unique function of time at the hand of the experimenter is not modeled as a representative of an ensemble. All the familiar probabilistic functions and parameters in this approach are constructed from this single function of time by adopting the relative measure of subsets of the real line (representing time) as a probability measure.

The relative measure is not σ-additive (additivity of countably infinite numbers of terms). Moreover, when it is adopted to construct the distribution function of the values assumed over time by a persistent function of time, the corresponding expectation operator, the infinite-time average, is not σ-linear (linearity of an operator applied to a linear combination of a countably infinite number of terms). These facts make the relative measure less amenable to mathematical manipulation involving infinite linear combinations than is the classical probability measure. However, when we define a stochastic process $X(t, \omega)$ as a function of two variables $t \in \mathbb{R}$ and $\omega \in \Omega$, for which ω is the ensemble index, the results of operations (e.g., calculations of values of time-distributions or time-averages) made over the time line indexed by t for some fixed ω (which represent empirical measurements) exhibit, by definition, the same properties as those of the FOT probabilistic functions and, therefore, do not have the σ-additivity and σ-linearity properties of the corresponding results of operations made over ω. Hence, if only one function of time is available at the hands of the experimenter, the abstraction of introducing a hypothetical ensemble indexed by ω creates an unnecessary dichotomy between theory and measurements. Directly considering probabilistic functions constructed over t avoids this dichotomy between the properties of the stochastic process and those of its individual sample paths, as illustrated in the examples presented in Section VI. This is the primary motivation for adopting the FOT probability approach for signal analysis.

The deep difference between the relative measure and the probability measure is a result of dictating that the measure of the hypothetical infinite sample space Ω be finite, which enables the introduction of Axiom VI of probability. These assumed properties are at odds with the empirical quantities to be modeled. They artificially give to the probability measure properties that are not shared with the relative measure, thereby rendering it more amenable to mathematical manipulations involving infinite linear combinations. Furthermore, although signals modulated with digital data have sample spaces with finite measure prior to normalization, as explained in connection with the PAM example in (57), many other communications signals do not have sample space with finite measure before normalization.

The conceptual usefulness of the FOT approach is evident in statistical spectral analysis of time series. This is the point of view of the book [37] by Gardner and the early work on generalized harmonic analysis [106]. In most problems of statistical spectral analysis, in fact, there is no need to model the available time series as a representative of an ensemble and, hence, there is no need to willfully incur a dichotomy between theory and measurements.

In Information Theory, compression of individual sequences has been shown to be a practicable and, in some cases, a convenient approach [118]. The same is true of channel coding. For example, the construction of block codes and convolutional codes and the corresponding decoding procedure can be made without involving any stochastic concept. However, in contrast to the practice of source and channel coding, in order to prove the channel coding theorem, one must adopt the concept of typical sequences whose probabilistic characterization is formulated in terms of the classical stochastic process. This is so, even though all the basic quantities used in Information Theory can be defined in terms of FOT probability.

Even if typical sequences can be characterized without introducing the stochastic process [60, Chap. 2], the proof of the channel-coding theorem is based on the characterization of typical sequences using the classical stochastic approach and the concept of the ensemble of all possible codes for a given channel [23, Chap. 8]. In addition, the channel is modeled as random. No proof of the channel coding theorem based on FOT probability is on the horizon.

A field where the FOT approach appears methodologically more appropriate than the classical stochastic approach is that of Monte Carlo simulations. They are FOT simulations, not stochastic simulations [25, Sec. 4.4]. In fact, a computer program for random number generation produces a unique periodic sequence with a very long period. Calling such a routine several times (with different seeds) is equivalent to picking different time segments of the unique sequence. So, the sample space is time indexed, not ensemble indexed. (If the period is sufficiently long, the sequence can be considered aperiodic for practical purposes.)

The extension of the FOT approach to periodic phenomena, that is, those which produce time series through the interaction of periodic mechanisms and random phenomena, is based on the nonobvious result that the almost-periodic component extraction operator is a valid expectation operator. When it is applied to the indicator of the set of val-

ues of t where a time series is below a threshold ξ, one obtains, for every value of t, a valid distribution function in ξ. The obtained function is almost-periodic in t by construction. Therefore, it is suitable for an FOT probabilistic characterization of cyclostationary, poly-cyclostationary, and almost-cyclostationary time series. In particular, a complete temporal probabilistic theory for these time series can be constructed [35], [37], [43], including a theory of higher-order moments and cumulants [44], [99].

The FOT approach for almost-cyclostationary signals can be extended to the class of the generalized almost-cyclostationary (GACS) signals [88, Chap. 2], [90, Chap. 12]. GACS signals have multivariate statistical functions almost periodic with respect to time with (generalized) Fourier series expansions for which not only the coefficients depend on the lag parameters of the time-shifted versions of the time series, but also the frequencies depend on the lag parameters. For these signals, the almost-periodic component extraction operator can be adopted as the expectation operator and a complete FOT higher-order theory is developed in [57].

Other generalizations of the class of almost-cyclostationary processes are the spectrally correlated (SC) processes [88, Chap. 4], [90, Chap. 13] and the oscillatory almost-cyclostationary (OACS) processes [89], [90, Chap. 14]. For these classes of nonstationary stochastic processes, there apparently do not exist FOT-probability counterparts despite the relationships of these processes to almost cyclostationary processes because the nonstationarity of these generalizations is not almost periodic or otherwise of known functional form [38]. However, time-warped almost-cyclostationary signals have been treated in [46] with an approach that mostly avoids stochastic processes and ergodicity.

APPENDIX I. INFINITE TIME AVERAGE OF A PAM TIME SERIES

In this Appendix, the infinite time average of a PAM time series is calculated without resorting to the σ-linearity of the infinite time average that would lead to the wrong result in the right-hand side of (62).

Let us consider the PAM time series (60) and, for the sake of simplicity, let as assume that the pulse $q(t)$ has finite duration and the sequence $\widetilde{x}_n$ is bounded, that is $|\widetilde{x}_n| \leqslant K$, $\forall n \in \mathbb{Z}$.

The finite time average of the PAM time series is

$$m_x^{(T)} \triangleq \frac{1}{T} \int_{-T/2}^{T/2} \sum_{n=-\infty}^{+\infty} \widetilde{x}_n \, q(t - nT_0) \, \mathrm{d}t \quad \text{(A1a)}$$

$$= m_1^{(T)} + m_2^{(T)} \quad \text{(A1b)}$$

where

$$m_1^{(T)} \triangleq \frac{1}{T} \int_{-T/2}^{T/2} \sum_{n=-N}^{N} \widetilde{x}_n \, q(t - nT_0) \, \mathrm{d}t \quad \text{(A2)}$$

and

$$m_2^{(T)} \triangleq \frac{1}{T} \int_{-T/2}^{T/2} \sum_{|n|>N} \widetilde{x}_n \, q(t - nT_0) \, \mathrm{d}t \quad \text{(A3)}$$

with N denoting the integer such that $T = 2NT_0 + 2\epsilon$ with $0 < \epsilon < T_0$.

By making the variable change $t' = t - nT_0$ and interchanging the order of the finite sum and the integral, we have

$$m_1^{(T)} = \frac{(2N+1)T_0}{2NT_0 + 2\epsilon} \frac{1}{(2N+1)T_0} \sum_{n=-N}^{N} \widetilde{x}_n$$

$$\cdot \left[\int_{-NT_0 - nT_0 - \epsilon}^{NT_0 - nT_0 + \epsilon} q(t') \, \mathrm{d}t' - \int_{-\infty}^{+\infty} q(t') \, \mathrm{d}t' \right.$$

$$\left. + \int_{-\infty}^{+\infty} q(t') \, \mathrm{d}t' \right]$$

$$\triangleq \frac{(2N+1)T_0}{2NT_0 + 2\epsilon} \left[\frac{1}{2N+1} \sum_{n=-N}^{N} \widetilde{x}_n \frac{1}{T_0} \int_{-\infty}^{+\infty} q(t) \, \mathrm{d}t \right.$$

$$\left. + r_1^{(T)} \right] \quad \text{(A4)}$$

with

$$|r_1^{(T)}| = \left| \frac{1}{(2N+1)T_0} \sum_{n=-N}^{N} \widetilde{x}_n \left[\int_{-NT_0 - nT_0 - \epsilon}^{NT_0 - nT_0 + \epsilon} q(t) \, \mathrm{d}t \right. \right.$$

$$\left. \left. - \int_{-\infty}^{+\infty} q(t) \, \mathrm{d}t \right] \right|$$

$$\leqslant \frac{1}{(2N+1)T_0} \sum_{n=-N}^{N} |\widetilde{x}_n| \int_{|t| > NT_0 - nT_0 + \epsilon} |q(t)| \, \mathrm{d}t$$

$$\leqslant \frac{K}{(2N+1)T_0} \sum_{n=-N}^{N} \int_{|t| > NT_0 - nT_0 + \epsilon} |q(t)| \, \mathrm{d}t \quad \text{(A5)}$$

Since $q(t)$ has finite support, only a finite number (not depending on N) of terms of the sum in the right-hand side of (A5) are non zero. Consequently, in the limit as $T \to \infty$ (and, hence $N \to \infty$), we have

$$\lim_{T \to \infty} r_1^{(T)} = 0. \quad \text{(A6)}$$

For the term $m_2^{(T)}$ we have the bound

$$\left| m_2^{(T)} \right| \leqslant \frac{K}{T} \int_{-T/2}^{T/2} \sum_{|n|>N} |q(t - nT_0)| \, \mathrm{d}t. \quad \text{(A7)}$$

All the time shifted pulses in the sum in (A7) are centered at points nT_0 outside the integration interval $[-T/2, T2]$. Thus, since $q(t)$ has finite duration, only a finite number of terms (not depending on N) give nonzero contribution to the integral. Consequently, the integral is bounded and

$$\lim_{T \to \infty} m_2^{(T)} = 0. \quad \text{(A8)}$$

From (A1b), and accounting for (A4), (A6), and (A8), it results that

$$\lim_{T\to\infty} m_x^{(T)} = \langle \widetilde{x}_n \rangle_n \; \frac{1}{T_0} \int_{-\infty}^{+\infty} q(t)\, \mathrm{d}t \qquad \text{(A9)}$$

where

$$\langle \widetilde{x}_n \rangle_n \triangleq \frac{1}{(2N+1)} \sum_{n=-N}^{N} \widetilde{x}_n. \qquad \text{(A10)}$$

Finally note that (A9) can be proved even after relaxing the assumption that $q(t)$ has finite duration provided that it has a rate of decay to zero sufficiently fast as $|t| \to \infty$.

REFERENCES

[1] R. P. Agnew and M. Kac, "Translated functions and statistical independence," *Bull. Amer. Math. Soc.*, vol. 47, no. 2, pp. 148–154, 1941.

[2] R. B. Ash, *Basic Probability Theory*. New York, NY, USA: Wiley, 1970.

[3] S. Ashley, "Core concept: Ergodic theory plays a key role in multiple fields," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 7, p. 1914, 2015.

[4] J. Bass, "Suites uniformément denses, moyennes trigonométriques, fonctions pseudo-aléatoires," *Bull. de la Société Mathématique de France*, vol. 87, pp. 1–64, Jan. 1959.

[5] J. Bass, "Espaces de Besicovitch, fonctions presque-périodiques, fonctions pseudo-aléatoires," *Bull. de la Société Mathématique de France*, vol. 91, pp. 39–61, Jan. 1963.

[6] J. Bass, "Fonctions stationnaires. Fonctions de corrélation. Application à la représentation spatio-temporelle de la turbolence," *Annales de l'Inst. Henri Poincaré, Sect. B*, vol. 5, no. 2, pp. 135–193, 1969.

[7] J. Bass, *Cours de Mathématique*, vol. 3. Paris, France: Masson & Cie, 1971.

[8] J. Bass, "Stationary functions and their applications to the theory of turbolence: I. Stationary functions," *J. Math. Anal. Appl.*, vol. 47, no. 2, pp. 354–399, Aug. 1974.

[9] J. Bass, "Stationary functions and their applications to the theory of turbolence: II. Turbolent solutions of the Navier–Stokes equations," *J. Math. Anal. Appl.*, vol. 47, no. 3, pp. 458–503, Sep. 1974.

[10] J. Bass, "Fonctions de corrélation des fonctions pseudo-aléatoires," *Annales de l'Inst. Henri Poincaré, Section B*, vol. 25, no. 4, pp. 503–515, 1989.

[11] J. J. Benedetto, *Harmonic Analysis and Applications*. New York, NY, USA: CRC Press, 1996.

[12] J.-P. Bertrandias, "Espaces de fonctions bornées et continues en moyenne asymptotique d'orde *p*," *Bull. Soc. Math. France (Mémoires de La S. M. F.)*, vol. 5, pp. 3–106, Jan. 1966.

[13] P. Billingsley, *Convergence in Probability Measures*. New York, NY, USA: Wiley, 1968.

[14] G. D. Birkhoff, "Proof of the ergodic theorem," *Proc. Nat. Acad. Sci. USA*, vol. 17, no. 12, pp. 656–660, Dec. 1931.

[15] S. Bochner and B. Jessen, "Distribution functions and positive-definite functions," *Ann. Math.*, vol. 35, no. 2, pp. 252–257, 1934.

[16] R. Boyles and W. Gardner, "Cycloergodic properties of discrete-parameter nonstationary stochastic processes," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 105–114, Jan. 1983.

[17] D. G. Brennan, "Probability theory in communication system engineering," in *Communication System Theory*, E. J. Baghdady, Ed. New York, NY, USA: McGraw-Hill, 1961, ch. 2.

[18] M. Buchanan, "Capturing chaos," *Nature*, vol. 435, no. 7040, p. 281, May 2005.

[19] M. Buchanan, "Gamble with time," *Nature Phys.*, vol. 9, no. 1, p. 3, Jan. 2013.

[20] G. Casinovi, "L^1-norm convergence properties of correlogram spectral estimates," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4354–4365, Sep. 2007.

[21] A. Church, "On the concept of a random sequence," *Bull. Amer. Math. Soc.*, vol. 46, no. 2, pp. 130–135, 1940.

[22] C. Corduneanu, *Almost Periodic Functions*. New York, NY, USA: Chelsea Publishing Company, 1989.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[24] D. Dehay, J. Leskow, and A. Napolitano, "Central limit theorem in the functional approach," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4025–4037, Aug. 2013.

[25] D. Dehay, J. Leśkow, and A. Napolitano, "Time average estimation in the fraction-of-time probability framework," *Signal Process.*, vol. 153, pp. 275–290, Dec. 2018.

[26] J. N. Doctor, P. P. Wakker, and T. V. Wang, "Economists' views on the ergodicity problem," *Nature Phys.*, vol. 16, no. 12, p. 1168, Dec. 2020.

[27] J. L. Doob, *Stochastic Processes*. New York, NY, USA: Wiley, 1953.

[28] Editorial, "Time to move beyond average thinking," *Nature Phys.*, vol. 15, no. 12, p. 1207, Dec. 2019.

[29] A. Einstein, "Method for the determinination of the statistical values of observations concerning quantities subject to irregular fluctuations," *IEEE ASSP Mag.*, vol. 4, no. 4, p. 6, Oct. 1987.

[30] K. Fukuyama, "Some limit theorems of almost periodic function systems under the relative measure," *Kyoto J. Math.*, vol. 28, no. 3, pp. 557–577, Jan. 1988.

[31] H. Furstenberg, *Stationary Processes and Prediction Theory. (AM-44)*. Princeton, NJ, USA: Princeton Univ. Press, 1960.

[32] R. G. Gallager, *Principles of Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[33] W. Gardner, "Stationarizable random processes," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 8–22, Jan. 1978.

[34] W. A. Gardner, *Introduction to Random Processes With Applications to Signals and Systems*, 2nd ed. New York, NY, USA: Macmillan, 1985.

[35] W. A. Gardner, "The spectral correlation theory of cyclostationary time series," *Signal Process.*, vol. 11, pp. 13–36, Jul. 1986.

[36] W. Gardner, "Introduction to Einstein's contribution to time-series analysis," *IEEE ASSP Mag.*, vol. 4, no. 4, pp. 4–5, Oct. 1987.

[37] W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.

[38] W. A. Gardner, "Correlation estimation and time-series modeling for nonstationary processes," *Signal Process.*, vol. 15, no. 1, pp. 31–41, Jul. 1988.

[39] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Process. Mag.*, vol. 8, no. 2, pp. 14–36, Apr. 1991.

[40] W. A. Gardner, "Two alternative philosophies for estimation of the parameters of time-series," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 216–218, Jan. 1991.

[41] W. A. Gardner, "An introduction to cyclostationary signals," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. IEEE Press, 1994, ch. 1, pp. 1–90.

[42] W. A. Gardner, "History and equivalence of two methods of spectral analysis," *IEEE Signal Process. Mag.*, vol. 13, pp. 20–23, Jul. 1996.

[43] W. A. Gardner and W. A. Brown, "Fraction-of-time probability for time-series that exhibit cyclostationarity," *Signal Process.*, vol. 23, no. 3, pp. 273–292, Jun. 1991.

[44] W. A. Gardner and C. M. Spooner, "The cumulant theory of cyclostationary time-series. I. foundation," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3387–3408, Dec. 1994.

[45] W. A. Gardner. (2018). *Cyclostationarity.Com*. Accessed: Nov. 2021. [Online]. Available: https://cyclostationarity.com

[46] W. A. Gardner, "Statistically inferred time warping: Extending the cyclostationary paradigm from regular to irregular statistical cyclicity in scientific data," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, p. 59, Sep. 2018.

[47] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, 2nd ed. Dordrecht, The Netherlands: Springer, 2009.

[48] P. R. Halmos, "The foundations of probability," *Amer. Math. Monthly*, vol. 51, no. 9, pp. 493–510, Nov. 1944.

[49] P. R. Halmos, *Lectures on Ergodic Theory*. New York, NY, USA: Mathematical Society of Japan, American Mathematical Society (AMS) Chelsea Publishing, 1956.

[50] P. Hartman and A. Wintner, "The (L^2)-space of relative measure," *Proc. Nat. Acad. Sci. USA.*, vol. 33, pp. 128–132, May 1947.

[51] P. Hartman, E. R. van Kampen, and A. Wintner, "Asymptotic distributions and statistical independence," *Amer. J. Math.*, vol. 61, no. 2, pp. 477–486, 1939.

[52] E. K. Haviland, "On statistical methods in the theory of almost-periodic functions," *Proc. Nat. Acad. Sci. USA*, vol. 19, no. 5, pp. 549–555, May 1933.

[53] J. Henniger, "Functions of bounded mean square, and generalized Fourier-stieltjes transforms," *Can. J. Math.*, vol. 22, no. 5, pp. 1016–1034, Oct. 1970.

[54] P. P. Hien, "Mesure asymptotique définie par une fonction à valeurs dans $\mathbb{R}^n$ ou dans un espace vectoriel topologique," *Annales de l'Institut Henri Poincaré–Sect. B*, vol. 11, no. 1, pp. 23–107, 1975.

[55] E. M. Hofstetter, "Random processes," in *The Mathematics of Physics and Chemistry*, vol. 2, H. Margenau and G. M. Murphy, Eds. Princeton, NJ, USA: D. Van Nostrand Co., 1964, ch. 3.

[56] H. Hurd and T. Koski, "The wold isomorphism for cyclostationary sequences," *Signal Process.*, vol. 84, no. 5, pp. 813–824, May 2004.

[57] L. Izzo and A. Napolitano, "The higher order theory of generalized almost-cyclostationary time series," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2975–2989, Nov. 1998.

[58] L. Izzo and A. Napolitano, "Linear time-variant transformations of generalized almost-cyclostationary signals.I. Theory and method," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 2947–2961, Dec. 2002.

[59] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Trans. Amer. Math. Soc.*, vol. 38, no. 1, pp. 48–88, 1935.

[60] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*. Washington, DC, USA: Mathematical Association of America, 1959.

[61] M. Kac and H. Steinhaus, "Sur les fonctions indépendantes (II)," *Studia Mathematica*, vol. 6, pp. 59–66, Jan. 1936.

[62] M. Kac and H. Steinhaus, "Sur les fonctions indépendantes (IV) (Intervalle infini)," *Studia Mathematica*, vol. 7, no. 1, pp. 1–15, 1938.

[63] Y. Katznelson and B. Weiss, "A simple proof of some ergodic theorems," *Isr. J. Math.*, vol. 42, no. 4, pp. 291–296, Dec. 1982.

[64] R. Kershner and A. Wintner, "On the asymptotic distribution of almost periodic functions with linearly independent frequencies," *Amer. J. Math.*, vol. 58, no. 1, pp. 91–94, 1936.

[65] F. Kindermann, A. Dechant, M. Hohmann, T. Lausch, D. Mayer, F. Schmidt, E. Lutz, and A. Widera, "Nonergodic diffusion of single atoms in a periodic potential," *Nature Phys.*, vol. 13, no. 2, pp. 137–141, Feb. 2017.

[66] A. N. Kolmogorov, *Foundations of the Theory of Probability*. New York, NY, USA: Chelsea, 1956.

[67] A. N. Kolmogorov, "Three approaches to the definition of the concept 'quantity of information,'" *Problemy Peredachi Informatsii*, vol. 1, no. 1, pp. 3–11, 1965.

[68] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1970.

[69] U. Krengel, *Ergodic Theorems*. Berlin, Germany: Walter de Gruyter, 1985.

[70] Y. W. Lee, *Statistical Theory of Communication*. New York, NY, USA: Wiley, 1960.

[71] J. Leśkow and A. Napolitano, "Quantile prediction for time series in the fraction-of-time probability framework," *Signal Process.*, vol. 82, no. 11, pp. 1727–1741, Nov. 2002.

[72] J. Leśkow and A. Napolitano, "Foundations of the functional approach for signal analysis," *Signal Process.*, vol. 86, no. 12, pp. 3796–3825, Dec. 2006.

[73] J. Leśkow and A. Napolitano, "Nonrelatively measurable functions for secure communications signal design," *Signal Process.*, vol. 87, no. 11, pp. 2765–2780, Nov. 2007.

[74] A. Lyon, "Kolmogorov's axiomatisation and its discontents," in *The Oxford Handbook of Probability and Philosophy*. Oxford, U.K.: Oxford Univ. Press, 2016, pp. 66–155.

[75] P. M. Mäkilä, "On chaotic and random sequences," *Phys. D*, vol. 198, pp. 300–318, 2004.

[76] P. M. Mäkilä, J. R. Partington, and T. Norlander, "Bounded power signal spaces for robust control and modeling," *SIAM J. Control Optim.*, vol. 37, no. 1, pp. 92–117, Jan. 1998.

[77] J. Marcinkiewicz and A. Zygmund, "Sur les fonctions indépendantes," *Fundamenta Math.*, vol. 29, pp. 60–90, Jan. 1937.

[78] J. Marcinkiewicz and A. Zygmund, "Quelques théorémes sur les fonctions indépendantes," *Fundamenta Math.*, vol. 7, pp. 104–120, 1938.

[79] J. Mari, "A counterexample in power signals space," *IEEE Trans. Autom. Control*, vol. 41, no. 1, pp. 115–116, Jan. 1996.

[80] P. Martin-Löf, "The definition of random sequences," *Inf. Control*, vol. 9, no. 6, pp. 602–619, 1966.

[81] P. Masani, "Review: H. Furstenberg, Stationary processes and prediction theory," *Bull. Amer. Math. Soc.*, vol. 69, no. 2, pp. 195–207, 1963.

[82] R. D. Mauldin, "Probability and nonlinear systems," *Los Alamos Sci.*, vol. 15, pp. 52–90, Jan. 1987.

[83] J. C. Mauro, P. K. Gupta, and R. J. Loucks, "Continuously broken ergodicity," *J. Chem. Phys.*, vol. 126, no. 18, May 2007, Art. no. 184511.

[84] J. L. McCauley, "Time vs. Ensemble averages for nonstationary time series," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 22, pp. 5518–5522, Sep. 2008.

[85] H. Miao, F. Zhang, and R. Tao, "A general fraction-of-time probability framework for chirp cyclostationary signals," *Signal Process.*, vol. 179, Feb. 2021, Art. no. 107820.

[86] T. Michaeli, V. Pohl, and Y. C. Eldar, "U-invariant sampling: Extrapolation and causal interpolation from generalized samples," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2085–2100, May 2011.

[87] C. C. Moore, "Ergodic theorem, ergodic theory, and statistical mechanics," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 7, pp. 1907–1911, Feb. 2015.

[88] A. Napolitano, *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*. Hoboken, NJ, USA: Wiley, 2012.

[89] A. Napolitano, "Time-warped almost-cyclostationary signals: Characterization and statistical function measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5526–5541, Oct. 2017.

[90] A. Napolitano, *Cyclostationary Processes and Time Series: Theory, Applications, and Generalizations*. Amsterdam, The Netherlands: Elsevier, 2019.

[91] A. B. Nobel, "Some stochastic properties of memoryless individual sequences," *IEEE Trans. Inf. Theory*, vol. 50, no. 7, pp. 1497–1505, Jul. 2004.

[92] A. B. Nobel, G. Morvai, and S. R. Kulkarni, "Density estimation from an individual numerical sequence," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 537–541, Mar. 1998.

[93] O. Peters, "The ergodicity problem in economics," *Nature Phys.*, vol. 15, no. 12, pp. 1216–1221, Dec. 2019.

[94] O. Peters, "Reply to: Economists' views on the ergodicity problem," *Nature Phys.*, vol. 16, no. 12, p. 1169, Dec. 2020.

[95] E. Pfaffelhuber, "Generalized harmonic analysis for distributions," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 605–611, Nov. 1975.

[96] J. G. Proakis, *Digital Communications*, 3rd ed. New York, NY, USA: McGraw-Hill, 1995.

[97] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. Hoboken, NJ, USA: Wiley, 1980.

[98] P. Shields, "The ergodic and entropy theorems revisited," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 2, pp. 263–266, Mar. 1987.

[99] C. M. Spooner and W. A. Gardner, "The cumulant theory of cyclostationary time-series. II. Development and applications," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3409–3429, Dec. 1994.

[100] H. Steinhaus, "Sur les foncions indépendantes (VI) (Équipartition)," *Studia Math.*, vol. 9, no. 1, pp. 121–132, 1940.

[101] H. Steinhaus, "Sur les foncions indépendantes (VII)," *Studia Math.*, vol. 10, pp. 1–20, Jan. 1948.

[102] K. Urbanik, "Effective processes in the sense of H. Steinhaus," *Studia Math.*, vol. 17, pp. 335–348, Jan. 1958.

[103] K. Urbanik, "Fourier analysis in Marcinkiewicz spaces," *Studia Math.*, vol. 21, no. 1, pp. 93–102, 1961.

[104] V. A. Uspenskii, A. L. Semenov, and A. K. Shen', "Can an individual sequence of zeros and ones be random?" *Russian Math. Surv.*, vol. 45, no. 1, pp. 121–189, Feb. 1990.

[105] J. V. Neumann, "Proof of the quasi-ergodic hypothesis," *Proc. Nat. Acad. Sci. USA*, vol. 18, no. 1, pp. 70–82, Jan. 1932.

[106] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, no. 1, pp. 117–258, Dec. 1930.

[107] N. Wiener, *The Fourier Integral and Certain of its Applications*. London, U.K.: Cambridge Univ. Press, 1933.

[108] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA, USA: MIT Press, 1949.

[109] A. Wintner, "On the asymptotic repartition of the values of real almost periodic functions," *Amer. J. Math.*, vol. 54, no. 2, pp. 339–345, Apr. 1932.

[110] A. Wintner, "Remarks on the ergodic theorem of Birkhoff," *Proc. Nat. Acad. Sci. USA*, vol. 18, pp. 248–251, Mar. 1932.

[111] A. Wintner, "On the asymptotic differential distribution of almost-periodic and related functions," *Amer. J. Math.*, vol. 56, nos. 1–4, pp. 401–406, 1934.

[112] H. Wold, "On prediction in stationary time series," *Ann. Math. Statist.*, vol. 19, no. 4, pp. 558–567, 1948.

[113] A. Yaglom, "Einstein's 1914 paper on the theory of irregularly fluctuating series of observations," *IEEE ASSP Mag.*, vol. 4, no. 4, pp. 7–11, Oct. 1987.

[114] A. H. Zemanian, *Distribution Theory and Transform Analysis*. New York, NY, USA: Dover, 1987.

[115] C. Zhang, "New limit power function spaces," *IEEE Trans. Autom. Control*, vol. 49, no. 5, pp. 763–766, May 2004.

[116] C. Zhang and W. Liu, "Uniform limit power-type function spaces," *Int. J. Math. Math. Sci.*, vol. 2006, pp. 1–14, Dec. 2006.

[117] C. Zhang and C. Meng, "Two new spaces of vector-valued limit power functions," *Studia Scientiarum Mathematicarum Hungarica*, vol. 44, no. 4, pp. 423–443, Dec. 2007.

[118] J. Ziv, "Coding theorems for individual sequences," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 405–412, Jul. 1978.

ANTONIO NAPOLITANO (Fellow, IEEE) was born in Napoli, Italy, in 1964. He received the Ph.D. degree in electronic engineering and computer science from the University of Napoli Federico II, in 1994. In 1997, he was with the Department of Electrical and Computer Engineering, University of California at Davis, Davis, USA, as a Postdoctorate Research Associate. From 1995 to 2001, he was an Assistant Professor and from 2001 to 2005, an Associate Professor at the University of Napoli Federico II. Since 2005, he has been a Full Professor of telecommunications at the University of Napoli "Parthenope." Since 2020, he has been the Affiliate Status at the James Watt School of Engineering, University of Glasgow, U.K. He was a Visiting Professor at the Institute de Recherche Mathématique de Rennes, Haute Bretagne, France (2005); the Laboratoire d'Analyse des Signaux & des Processus Industriels, Universite Jean-Monnet, Roanne, France (2010 and 2013); the Ecole Nationale Superieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique et Telecommunications, Toulouse, France (2014 and 2015); and Politechnika Krakowska, Krakow, Poland (2021–2022). He held visiting appointments at the Universite de Nice Sophia Antipolis, France (1994); the University of North Carolina, Chapel Hill, NC, USA (1997, 2001, and 2002); the Technical University of Wroclaw, Poland (1999 and 2001); the Centro de Investigacion en Matematicas, Guanajuato, Mexico (2000–2002); the Econometric Department, Wyzsza Szkola Biznesu, Nowy Sacz, Poland (2000–2007); the School of Electrical and Information Engineering, University of South Australia (2010); and the Beijing Institute of Technology (2019). He is the author of the books *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications* (John Wiley & Sons Ltd.,—IEEE Press, 2012) and *Cyclostationary Processes and Time Series: Theory, Applications, and Generalizations* (Elsevier, 2019). In 1996, he received the Best Paper of the Year Award from the European Association for Signal Processing (EURASIP) for a paper on higher-order cyclostationarity. In 2007, he was a recipient of the EURASIP Best Paper Award for a paper on the functional approach in signal analysis. In 2008, he received In 2008, he received from Elsevier the Most Cited Paper Award for a review article on cyclostationarity. From 2006 to 2009 and from 2011 to 2015, he was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has been the Editorial Board of *Signal Processing* (Elsevier), since 2008, and *Digital Signal Processing* (Elsevier), since 2015. From 2008 to 2013 and since 2022, he has been an Elected Member of the Signal Processing Theory and Method Technical Committee (SPTMTC) of the IEEE Signal Processing Society. Since 2017, he has been an Elected Member of the Sensor Array and Multichannel Technical Committee (SAMTC) of the IEEE Signal Processing Society. He is an EURASIP Local Liaison Officer.
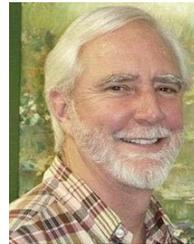
WILLIAM A. GARDNER (Life Fellow, IEEE) received the M.S. degree from Stanford University and the Ph.D. degree from the University of Massachusetts, Amherst, in 1967 and 1972, respectively, both in electrical engineering. He was on the faculty of the Department of Electrical and Computer Engineering, University of California at Davis (UC Davis), from 1972 to 2000. He was the Founder, the President, and the Chief Scientist at Statistical Signal Processing Inc., from 1986 to 2011, and is presently a Professor Emeritus at UC Davis. He is the author or coauthor of chapters in four books. He has authored three books, *Introduction to Random Processes with Applications to Signals and Systems* (Macmillan, 1985) (second edition, McGraw-Hill, 1989), *Statistical Spectral Analysis: A Nonprobabilistic Theory* (Prentice-Hall, 1987), and *The Random Processes Tutor: A Comprehensive Solutions Manual for Independent Study* (McGraw-Hill, 1989). He holds various patents, is the author of over 100 peer-reviewed research papers, and has given about 50 invited lectures at university and industrial research laboratories. He was elected as a Life Fellow of IEEE, in 1991, for his contributions to the development of time-series analysis and stochastic processes with applications to statistical signal processing and communication, and for contributions to engineering education. He received the international Best Paper of the Year Award from the European Association for Signal Processing, in 1986, for the paper "The spectral correlation theory of cyclostationary time-series," the 1987 Distinguished Engineering Alumnus Award from the University of Massachusetts, and the international Stephen O. Rice Prize Paper Award in the Field of Communication Theory from the IEEE Communications Society, in 1988, for the paper titled "Signal interception: A unifying theoretical framework for feature detection." He is the Editor of *Cyclostationarity in Communications and Signal Processing* (IEEE Press, 1994). The present focus of his work is the educational website https://cyclostationarity.com/ and ongoing research on the theory of Fraction-of-Time probability for time-series analysis.

● ● ●