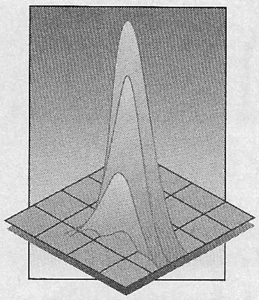

SP Forum

"... most engineers do not understand the need for careful thinking about statistical matters. It is time for the professional statisticians to rise to the challenge of teaching statistics to engineers and scientists."—M.J. Hinich

While *SP Magazine* is not necessarily adverse to the publication of book reviews, for various reasons none has been published for many years. The following articles are included, not so much for their interest as book review material, but for the issues they raise; and also to launch a new department in this Magazine by which readers can express their views on topics of importance in the area of signal processing.

Most of us working in SP would confess to sometimes sloppy or cavalier use of statistical concepts. The reality, is, however, that we rarely have certain justification for any population model assumed, regardless of how mathematically rigorous our analysis based on the model. Professor William Gardner of the University of California at Davis proposes an unorthodox solution to this problem in his recent book in which statistics over ensemble models are replaced by a dual "temporal probability" model over waveforms. In his review of Gardner's book, Professor Melvin Hinich of the University of Texas at Austin takes great exception to this approach in his review written for *SIAM Review*, and reprinted here.

Hinich implies that acceptance of such a model would reveal the engineer's misunderstanding of "the need for careful thinking about statistical matters." Following Hinich's review is Gardner's rejoinder in the form of a humorous piece, "Ensembles in Wonderland," intended to "inform people about both sides of the controversy without trying to discredit either side."
I will let the reader determine whether Professor Gardner has achieved his objective with complete impartiality. Regardless of which side of the debate you find yourself on, the article should be informative.

At a time when several issues centering on statistical signal processing are planned for *SP Magazine*, it is hoped that this set of articles will stimulate debate and interest in the subject. Write to us!
— J. R. Deller

Statistical Spectral Analysis: A Nonprobabilistic Theory. William Gardner. Prentice-Hall, NJ, 1987. xxvi + 566 pages. ISBN 0-13844-572-9.

This book presents a nonstandard treatment of the Fourier analysis approach to the study of deterministic, random continuous, and discrete-time signals. The coverage of the algebra of spectral analysis, and the motivation, examples, and exercises used in his book are basically the same as other modern books in the subfield of electrical engineering, which is usually called "Digital Signal Processing." What is odd about this work is the author's strange "mathematical" approach to stochastic processes.

The subtitle illustrates the bizarre nature of this book, "A Nonprobabilistic Theory" (is nonprobabilistic a word?). In his Preface, Gardner attempts to justify his approach. He writes about the need for an empirical approach to the study of time series (I agree), and then he makes a brief attempt to use past mathematical work on randomness to justify his approach. For example, he uses an out-of-context quote from an obscure book by T. S. Fine (*Theories of Probability: An Examination of Foundations*, Academic Press, New York, 1973).

Gardner does *not* avoid probability and randomness in this exposition. What he does is develop a heuristic treatment of randomness of a realized time series in terms of what he calls the "fraction-of-time probabilistic model." His concept is a confused and confusing attempt to use well-known results from Markov process theory to set up an idiosyncratic theory of probability without the standard discipline from mathematics. Just imagine the confusion that would result in the application of statistical concepts if we had no standard and widely understood theory of probability to rationalize our data analysis and interpretation!

Both the author's Preface and the Forward by Bracewell reveal the confusion that reigns concerning the role of probability models in a statistical analysis of data and experimental design. The science of statistics deals with (1) the design of experiments to control the randomness and uncertainty in any experiment, and (2) the analysis of data as an organized process of scientific inquiry. The basic idea is to take representative samples from some defined population, and then make inferences about the population from the data. In time series analysis, the standard conceptualization is to assume that the time series we observe is a representative sample from an ensemble of series that are mathematically modeled by a joint distribution for the sample times.

One convenient way of thinking about the ensemble (which Gardner abhors) is to think of the observed series as a signal from a machine whose system parameters randomly vary over time in some fashion. The machine that generates the data is one of a set of identical machines that were manufactured at the same plant in the same time. The only difference between the machines is the setting of the varying parameters over time. A probability model is assumed for the random parameters for the set of machines. As long as we believe that the machine that generates the data is a representative randomized selection from the population of identical machines (the ensemble), then it is possible to make inferences about the population from the signal that we observed, provided that the process is stationary and satisfies some probabilistic mixing condition.

The mixing condition is crucial to statistical analysis of time series. It is not enough to assume stationarity and ergodicity. Ergodicity and stationarity are sufficient to develop consistent estimates of system parameters and the spectrum at a set of discrete frequencies. Consistency is convergence in probability. But convergence in probability, or even convergence almost surely to a point, is a very weak condition for an estimation method. In order to make inferences from sample estimates about population parameters, we need to know something about the statistical variability of the sample estimates. This requires the use of probability models and some sort of central limit theorem. Since the interesting cases in time series involve time-dependent random processes, we need to invoke a mixing condition to validate the application of a central limit theorem for asympotic analysis of estimates. If the finite sampling properties of an estimate are known, which may be the case if the process is Gaussian, then we do not need asympotics, but we do need probability models.

The author seems to be aware that he needs to use standard notation in parts of his book, so he writes about expected values using time averages. His approach is very hard to follow because he is "standing on his head" to adhere to this ideology of determinism.

It is too bad that the exposition in his book is harmed by such an odd misdirection of intellectual effort. The book covers a set of topics that are rarely treated in the signal processing literature: cyclostationarity and periodic signals with random components. The treatment of cyclostationarity is very idiosyncratic since Gardner is a pioneer in this field with L. Franks. A cyclostationary process is *not* ergodic, but it may be made so by sampling the process using periodic sampling synchronized with the periodicity of the process.

The simplest such process is a sine function whose amplitude and phase are fixed, but where the phase is drawn at random. Such a process is then stationary but not ergodic. Since it is easy to determine the period of the sinusoid as long as it is not observed with a lot of noise, it is easy to synchronize the sampling. Since the author defines cyclostationarity in a convoluted way using his time averaging formalism, the simplicity of the concept is totally obscured. This is one of many examples in the book where simple concepts are confused by Gardner's compulsion for "home-made" mathematics.

This book should not be used as a single source for a course in statistical spectral analysis because the students will not learn the basic concepts that everyone else learns (or attempts to). There is no reason for students in such a course to learn measure theory, but they should learn the basics of multivariate probability theory. Even social science students are able to learn and properly use these basic mathematical tools. Engineers and scientists have better mathematical background and are constantly drilled in the precise application of mathematical models, so they should be expected to master the simple mathematics required in statistics. The hard part of statistical analysis is the complexities inherent in inference, where there is no *right answer*. It is clear to me from reading this book and its Preface and Forward that most engineers do not understand the need for careful thinking about statistical matters. It is time for the professional statisticians to rise to the challenge of teaching statistics to engineers and scientists.

— Melvin J. Hinich

Author's Comment

Professor Hinich has written a severely critical review [1] of the book [2], and of engineers in scientists in general with regard to their understanding of statistics. As author of this book, I am responding to this criticism, not to try to establish whether or not this is a good book, but rather to counter Professor Hinich's bold claim that the philosophy put forth regarding the utility of a nonstochastic approach to time-series analysis, is misguided and that engineers and scientists who accept this philosophy as a viable conceptual tool do not understand statistics. In Professor Hinich's own words," Both the author's Preface and the Forward by Bracewell reveal the confusion that reigns concerning the role of probability models in a statistical analysis of data...it is clear to me from reading this book that most engineers do not understand the need for careful thinking about statistical matters. It is time for the professional statisticians to rise to the challenge of teaching statistics to engineers and scientists."

Everyone agrees, I think, that stochastic processes have their place—they are indeed useful in some applications. The question before us is, "Must we accept stochastic processes as the only viable approach to dealing with time-series data that is erratic or unpredictable? Or is there a viable alternative that is more useful in some applications?"

The standard approach in statistics is well suited to experimental design, data analysis, and inference for populations: When a population actually exists in the real world, application of the standard conceptual framework of orthodox statistics can be appropriate. However, when no population associated with the available data exists or when no such population can exist (e.g., in astronomy, the concept of a population of universes is not usually considered viable— replicating the "experiment" of creating the universe is rather far-fetched), then the appropriateness of pretending that a population exists should be questioned.

Many statisticians accept and invoke the concept of a population in developing their theories and methods. When one has a single time-series of data to analyze and use as a basis for making inferences about the real-world situation that gave rise to the data, and when one knows that there can be no access to a population of such time-series (because it does not or cannot exist), then one would be remiss in not questioning the appropriateness of the orthodox conceptualization of the time-series as one member of a population, or ensemble, mathematically modeled as a stochastic process.

Many—but by no means all—real-world problems in communications engineering and signal processing involve time-series data for which no population exists; that is, data for which replication of the experiment is impossible or impractical. However, many of these time-series are known to arise from physical phenomena that can be considered to be unchanging in their basic nature of very long periods of time. In such cases, conceptually idealizing this time-invariance by extending the length of the data record without bound enables us to conceive of a model that is derivable from the data in the limit as the amount of data used for measuring the parameters of the model approaches infinity. This leads us to the concept of a fraction-of-time (FOT) probability model that is free from the abstract concept of a population. For example, the FOT probability that a time-series exceeds some specified level is defined to be the fraction of time that this event occurs over the life of the time-series.

Once we have accepted the idea of an infinitely long time-series with an FOT probability model, we can develop a theory of statistical inference and decision that is isomorphic to the theory for stationary stochastic processes.

In summary, I believe Professor Hinich's admitted confusion about the message delivered in [2] results not from any flaw in the philosophy put forth in [2], but rather from his unwillingness to accept this philosophy as a viable alternative to his philosophy—a philosophy to which he clings tightly. Moreover, in defense of engineers and scientists, I mention that the book [3] on stochastic processes, coupled with [2], illustrates that some nonstatisticians are capable of understanding, using, and teaching both the orthodox theory of stochastic processes (for those situations where it is appropriate) and the unorthodox theory of time-series based on FOT probability (for those other situations where it is the more appropriate of the two).

Not all statisticians share Professor Hinich's inflexible position on what he calls "an odd misdirection of intellectual effort." Professor A. M. Yaglom of the Academy of Sciences of Russia, author of well known books on time-series analysis and stochastic processes, states in his review [4] of [2] and [3]:

"It is important, however, that until Gardner's second book was published there was no attempt to present the modern spectral analysis of random processes consistently in language that uses only time-averaging rather than averaging over the statistical ensemble of realizations. Moreover, this book also shows that such a treatment possesses some advantage over the traditional one...Professor Gardner's books are both valuable additions to the available literature on the theory of random processes."

Similarly, Professor Enders A. Robinson of Columbia University, author of thirty books on time-series analysis, states in his review [5]:

"This book can be highly recommended to the engineering profession. Instead of struggling with many unnecessary concepts from abstract probability theory, most engineers would prefer to use methods that are based on the available data. This highly readable book gives a consistent approach for carrying out this task. In this work Professor Gardner has made a significant contribution to statistical spectral analysis, one that would please the early pioneers of spectral theory and especially Norbert Wiener."

Apparently, the distinction to be made is not Professor Hinich's distinction between engineers/scientists and statisticians, but rather it is the distinction between pragmatists (as defined by the American philosophers Charles Sanders Peirce and William James), who can adopt whatever conceptualization best serves the practical purpose at hand, and those others who believe in the sanctity of one particular system of conceptualization, regardless of its practical consequences. The nonpragmatists speak of a controversy over the stochastic and nonstochastic approaches to time-series analysis. But, there is really no basis for controversy. The only real issue is one of judgement—judgement in choosing for each particular time-series analysis problem the most appropriate of two alternative approaches [6].

Making inferences from available data is tricky business for anyone. For example, Professor Hinich makes the following inference about a population numbering in the hundreds of thousands on the basis of data from a single member of this population: "It is clear to me [Hinich] from reading this book that most engineers do not understand the need for careful thinking about statistical matters."

References

1. M. J. Hinich, *Review of Statistical Spectral Analysis: A Nonprobabilistic Theory*, by W. A. Gardner, Prentice-Hall, 1987, Siam Review (1991), pp. 677-678.

2. W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1987.

3. W. A. Gardner, *Introduction to Random Processes with Applications to Signals and Systems*, Macmillian, New York, 1985 (2nd Edition, McGraw-Hill, New York, 1990).

4. A. M. Yaglom, Review of *Introduction to Random Processes with Applications to Signals and Systems*, by W. A. Gardner, Macmillian, 1985, and *Statistical Spectral Analysis: A Nonprobabilistic Theory,* by W. A. Gardner, Prentice-Hall, 1987; *Theory Probab. Appl.* 35, (1990), pp. 405-407.

5. E. A. Robinson, Review of *Statistical Spectral Analysis: A Nonprobabilistic Theory*, by W. A. Gardner, Prentice-Hall, 1987; *Signal Processing* (EURASIP) **20** (1990), p. 99.

6. W. A. Gardner, "Two alternative philosophies for estimation of the parameters of time-series," *IEEE Transactions on Information Theory,* **37** (1991), pp. 216-218.

Ensembles In Wonderland

by William A. Gardner

Alice was an above-average engineering student at the University of the Queen of Hearts, but this semester she was struggling through a course on statistical signal processing, and things were getting curiouser and curiouser. When she was confronted with the topic of spectral analysis, she was perplexed by the mysterious reasoning given for spectrally smoothing the periodogram of a record of data produced by a receiver in a radio astronomy experiment. The argument Alice was presented with went something like this: Nevermind that you have only one record (which could go on forever) and that you want to use this record to learn something about the sources of radio energy impinging on Earth, pretend that your data is one record from an infinite ensemble of possible records (each of which could go on forever) that might have occurred, but didn't. Also, assume that this hypothetical ensemble is governed by a Gaussian probability law. To reduce the variance over the ensemble of these nonexistent records (from nonexistent universes), you should spectrally smooth the periodogram. Why? Because it is mathematically proven in the textbook that the variance over the make-believe ensemble, which defines a Gaussian stochastic process, is reduced by spectral smoothing. Oh, and by the way, this hypothetical variance also can be reduced by time averaging the periodogram instead of spectrally smoothing it. Why? Same reason: It is mathematically proven ...

Alice whispered to her classmate, "It would be so nice if someone would make sense for a change." Professor M. Hatter overheard Alice and turned to her with a stern look on his face and announced "This does make sense." Holding the textbook up for display, he said, "Just start at the beginning and when you reach the end, stop." Troubled by this, Alice decided she would get a second opinion. Later that day, she followed the Queen's Way to the campus repository of knowledge, the Cabbages and Kings Library. Hoping to find a more intuitively satisfying discussion of statistical spectral analysis, she located a score of books on the subject and proudly carried them off to a quiet corner of the reading room and began her quest for enlightenment. But, before long, her enthusiasm began to fade and in its place resignation set in. Alice found that all the books told the same story, whether authored by Tweedle Dee or Tweedle Dum. Stochastic processes were mathematically defined; the property of stationarity was introduced; the autocorrelation function was defined; the power spectral density function was defined to be the Fourier transform of the autocorrelation function; then the periodogram was introduced as a potential estimator of the power spectral density, but it was shown that even though the expected value over the ensemble could be made to approach the power spectral density by using long enough data records, this did not reduce the large variance over the ensemble. Then spectral smoothing was introduced as a means for reducing variance.

At this point, Alice could only conclude that the apparent absurdity of the make-believe ensemble and corresponding stochastic process was a result of her own failure to grasp something very deep and meaningful. After all, engineers do simulate ensembles on computers, and there was a lot of talk about something called virtual reality. So maybe the concept of reality was itself silly she thought, maybe nothing's impossible. Alice put her nose to the grindstone for the rest of the semester and learned the mechanics of stochastic processes, all the while trying to repress the thought "Why is it I give myself very good advice, but very seldom follow it?"

Time passed and Alice went on to earn a Ph.D. degree in, of all things, statistical signal processing. She considered a research position in industry, but was offered a professorship at the prestigious Stanford Institute of Technology in Berkeley, Massachusetts, which she happily accepted. One day after a lecture in which Alice eloquently spun the yarns about ensembles in wonderland that she had learned at Queen of Hearts, one of her students, with the grin of a cheshire cat, pointed out that a book review on this very topic had just appeared in a mathematics journal, and it seemed to suggest that engineers don't always view this topic the same way statisticians do. Alice got a copy of the book review and was amused to find the reviewer claiming that some foolish professor had written an entire book on an apparently ridiculous way of thinking about statistical spectral analysis that avoided stochastic processes and associated ensembles. The review [1] exposed the "bizarre" book [2] as an "odd misdirection of intellectual effort."

Reaching back in her mind, Alice recalled the difficulties she had experienced as a student at Queen of Hearts trying to grasp what then seemed to her to be a mystery, but now she accepted without question. But did she? There still seemed to be some doubt lurking in the recesses of her mind, in spite of the confidence with which she taught the subject of stochastic processes. Alice thought it would be fun to see first hand what this professor had to say. Upon finding the book in the library, she was surprised to see that the binding was broken and the cover was crumpled and soiled. There was a grimy imprint on the title page that read "Nike." It looked as though someone had thrown the book violently to the floor and stomped on it. She skimmed the table of contents and then turned to Chapter 3, where statistical spectral analysis was introduced. This is what she read:

In order to understand why a time--averaged periodogram can be preferable to one that has not been averaged, we must focus our attention not on the data itself but rather on the source of the data—the physical mechanism that generates the data. Generally speaking, data is nothing more than a partial representation of some physical phenomenon—a numerical representation of some aspects of the phenomenon. The fundamental reason for interest in a time-averaged spectrum of some given data is a belief that interesting aspects of the phenomenon being investigated have spectral influences on the data that are masked by uninteresting (for the purpose at hand) random effects and an additional belief (or, at least, hope) that these spectral influences can be revealed by averaging out the random effects. This second belief (or hope) should be based on the knowledge (or, at least, suspicion) that the spectral *influences of the interesting aspects of the phenomenon are time-invariant, so that the corresponding invariant spectral features (such as peaks or valleys) will be revealed rather than destroyed by time-averaging.*

Hmmm, she thought. This doesn't sound all that foolish. Let me go back and look at the introductory section in Chapter I, entitled "Objectives and Motives." She read:

A premise of this book is that the way engineers and scientists are commonly taught to think about empirical statistical spectral analysis of time-series data is fundamentally inappropriate for many applications. The subject is not really as abstruse as it appears to be from the conventional point of view. The problem is that the subject has been imbedded in the abstract probabilistic framework of stochastic processes, and this abstraction impedes conceptualization of the fundamental principles of empirical statistical spectral analysis. Hence, the probabilistic theory of statistical spectral analysis should be taught to engineers and scientists only after they have learned the fundamental deterministic principles—both qualitative and quantitative. For example, one should first learn 1) when and why sine wave analysis of time-series is appropriate, 2) how and why temporal and spectral resolution interact, 3) why statistical (averaged) spectra are of interest, and 4) what the various methods for measuring and computing statistical spectra are and how they are related. One should also learn how simultaneously to control the spectral and temporal resolution and the degree of randomness (reliability) of a statistical spectrum. All this can be accomplished in a nonsuperficial way without reference to the probabilistic theory of stochastic processes.

Alice turned back to Chapter 3 and read on to find that one can obtain approximately the same result by frequency smoothing the periodogram rather than time averaging it. There was a proof of this statement that consisted of manipulating the mathematical expression for the time-averaged periodogram into an expression for the frequency-smoothed periodogram. She jumped ahead to Chapter 5 and found a mathematical proof that the variance over time (not over some make-believe ensemble) was indeed reduced by both time averaging and frequency smoothing the periodogram.

In fact, there were many mathematical results that were essentially the same as those in the books that couch everything in terms of stochastic processes; it was the link between the mathematics and the real world that was different. This book used time averaging everywhere the other books used ensemble averaging. This seemed to make a lot of sense when the real world situation involves a single record of data from a physical phenomenon that is not changing with time.

Alice feverishly read on as her original motive for perusing the book turned into a reborn desire to solve the mystery she first encountered as a student at Queen of Hearts many years ago.

When she finally set the book down hours later with a deep feeling of satisfaction, Alice wondered why all the other books mask such naturally simple ideas behind ramblings about ensembles in wonderland. She suddenly remembered the review which led her to this unique book and she returned to her office to reread the review. She went to the concluding line in the review and read:

It is clear to me from reading this book that most engineers do not understand the need for careful thinking about statistical matters. It is time for the professional statistician to rise to the challenge of teaching statistics to engineers and scientists.

She reflected for a moment and then said to herself "How strange for a professional statistician to draw an inference about a population consisting of hundreds of thousands on the basis of observations on a single member of this population; certainly the concept of ergodicity does not apply here. Then something caught her eye. It was the only technical statement in the review. She read, "the simplest such process is a sine function whose amplitude and phase are fixed [over time], but where the phase is drawn at random [from an ensemble]. Such a process is then stationary but not ergodic. "Gee," Alice thought "maybe engineers have something to offer to statisticians: we can explain to them why this simple process is indeed ergodic, even though it does not pass the standard "mixing" test that they are so fond of.

Alice set the review aside and mused for awhile. It was late and she was tired. Little glimmers of curiousities from the history of engineering came to mind and faded—the heated controversy between physicists and mathematicians over Heaviside's operational calculus, now a standard part of engineering curricula—the mathematician's revulsion over the engineer's best friend, the impulse. She wondered if the current protests from some statisticians over this alternative approach to statistical spectral analysis would subside, and it too would become a standard part of engineering. One thing was clear. It would definitely become a standard part of the signal processing curriculum in her department at Stanford Institute of Technology.

The next day, Alice awoke early and went directly to her desk to prepare a new lecture to replace the one she had given yesterday. She wrote:

"The standard approach in statistics is well suited to experimental design, data analysis, and inference for populations: when a population actually exists in the real world, application of the standard conceptual framework of orthodox statistics can be appropriate. However, when no population associated with the available data exists or when no such population can exist, then one should question the appropriateness of pretending that a population does exist.

"Many statisticians accept and invoke the concept of a population in everything they do in developing theory and method. But unquestioning adherence to orthodoxy is not usually considered good science. It is a comforting approach for many but it is not always defensible on scientific grounds. When one has a single time-series of data to analyze and use as the basis for making inferences about the real-world situation that gave rise to the data, and when one knows that there can be no access to a population of such time-series (because none does or can exist), then one would be remiss in not questioning the appropriateness of the orthodox conceptualization of the time-series as one member of a population, or ensemble, mathematically modeled as a stochastic process.

"Many—but by no means all—real-world problems in communications engineering and signal processing involve time-series data for which no population exists, i.e., for which replication of the "experiment" is impossible or impractical. However, many of these time-series arise from physical phenomena that can be considered to be unchanging in their basic nature for a very long time. In such cases, conceptually idealizing this time-invariance by extending the length of time without bound enables us to conceive of a model that is derivable from the data in the limit as the amount of data used for measuring the parameters of the model "approaches infinity." This leads us to the concept of a fraction-of-time probability model that is free from the abstract concept of a population. (For example, the fraction-of-time probability that a time-series exceeds some specified level is defined to be the fraction of time that this event occurs over the life of the time-series.) ..."

Alice stopped writing for a moment to reflect. It occurred to her that the title of the book [2] *Statistical Spectral Analysis: A Nonprobabilistic Theory* could be misleading since it is shown in Chapter 5 of the book that an empirically motivated inquiry into the problem of quantifying the average behavior of spectral measurements leads naturally to a probabilistic theory. Since this probabilistic theory is nonstochastic (it involves only time averages, not ensemble averages), the title could have been *Statistical Spectral Analysis: A Nonstochastic Theory*. Nevertheless, she thought, the majority of the concepts and methods developed in the book are not only nonstochastic, they are indeed nonprobabilistic and a primary goal of the book is to show that, in an empirically motivated development of the fundamental concepts and methods of statistical spectral analysis, probability does not play a seminal role. It does play an important role in the mechanics of quantifying average behavior, but it plays no role in conceptualizing the objectives and methods of statistical spectral analysis of single time-series.

As Alice continued to write her lecture, she was keenly aware of a new level of enthusiasm that she could bring to her course on statistical signal processing. She would have to make a point of thanking the student with the cheshire cat grin for bringing that book review to her attention.

References

1. M. I. Hinich, Review of Statistical Spectral Analysis: A Nonprobabilistic Theory, by W. A. Gardner, Prentice-Hall, 1987, *SIAM Review*, pp. 677-678, December 1991.

2. W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1987.