BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 5, September 1977

Introduction to statistical time series, by Wayne A. Fuller, Wiley, New York, London, Sydney, Toronto, 1976, ix + 470 pp., $24.95.

Time series analysis is a part of statistics, i.e. the study of the reduction of data. It has individual features that separate it from the central part of statistics, on which it has had relatively little influence. The subject has some attractions for it is mathematically elaborate yet realistic. The complexity of the subject, together with its strong contacts with other parts of science, which make it appear a "foreign land" to a student with a purely mathematical background and too mathematically difficult for a student not mathematically able, also make coherent presentations especially important and there have been many books published on the subject in the last ten years. (I can count ten, leaving aside books emphasising applications in special parts of science or dealing with the underlying probability theory.) In its modern form time series analysis dates from the early fifties, and the advent of high speed computing, and the first reasonably connected account in this sense is probably [**1**], which is still worth studying. One is led to ask what such a treatise, written today, *might* reasonably contain. (What any particular work *will* contain must depend also on the audience to whom it is addressed.) In the first place there will have to be an account of the underlying probability theory which emphasises the theory of stationary processes with finite mean square and thus emphasises the place of Fourier methods in the theory. Of course some account might be given of spatial processes that are homogeneous (i.e. have their probability structure invariant under a group which acts in the space). However it is doubtful how much generality is valuable here for the range of cases of importance is limited and, moreover, from a statistical point of view the process has to be sampled and for most spatial phenomena the sampling is so irregular (e.g. the location of weather stations) that symmetry present in the underlying process is lost. Nevertheless an account sufficient for an understanding of concepts such as wave number spectrum, dispersion, isotropy etc. could be aimed at together with *some* indication of the unifying mathematical ideas (relating to the representation theory of the group). In addition to such a treatment for processes whose state varies continuously, there might also be some account of the theory of point processes. Two other parts of mathematics relate to the statistics. The first of these is ergodic theory. The importance of this can be perceived from the last chapter of [2], for example. The second is (linear) prediction theory. The classical part of this theory is the linear prediction theory for stationary Gaussian processes and though of no real importance in any generality, from the point of view of actually doing prediction, nevertheless it is intimately related to delicate aspects of the structure of stationary Gaussian processes and also to ergodic theory for example (see [**2**]). Apart from this classical part (which emphasises Fourier methods, H_p spaces etc.) there is a somewhat mathematically simpler part that commences from the model of a vector, Gaussian, Markov process (but not necessarily stationary) observed subject to (Gaussian white noise) error. The, so-called, Kalman filter which accomplishes prediction for these models is important not only for prediction itself but also because it relates closely to important statistical questions.

The theory of time series can be presented coherently but such a presentation will require not only the mathematical groundwork already described but also a preliminary treatment of the limit theorems of probability in relation to stationary sequences (at least). The central limit theorem will be essential and in a form that at least allows results to be established for functions of the Fourier coefficients. These are defined, for vectors of data $x(n)$, $n = 1, \ldots, N$, by

$$w(\omega_j) = N^{-1/2} \sum_1^N x(n) e^{in\omega_j}, \qquad \omega_j = 2\pi j/N', \; N' \geqslant N.$$

These quantities, whose cheap computation when N' is highly composite is of great importance, form the basis of a great many statistical calculations. The development of the limit theorems, in some of their most pleasing forms, depends on the theory of square integrable martingales, some part of the theory of which may also, therefore, be needed. This martingale theory will also link with ergodic theory (mixing etc.) and with the theory of the construction of likelihoods, for example for point processes and continuous time linear systems (see below).

The first part of the statistical methods that can be treated in a coherent manner is that using Fourier methods and based, ultimately, on the $w(\omega_j)$. There is a tendency to equate these methods with estimation of spectra and cross spectra (which describe the decomposition of a vector process into orthogonal components via the corresponding decomposition of variances and covariances). This is important and techniques are still developing today but often of greater importance is the use of these quantities in more elaborate statistical calculations. An example is the general field of signal velocity measurement and the measurement of the dispersive properties of the media through which the wave is propagated. Here the basic measurements of spectra (especially coherence and phase) are combined in elaborate ways to make the final measurements. *An underlying idea which gives unity over a much wider range is certainly the following* (which I ascribe to [**3**]). The quantities $w(\omega_j)$ are, for $N' = N$, "nearly" independent Gaussian random vectors (for $x(n)$ a vector measurement) as $N \to \infty$ in the sense that, under suitable conditions, for a fixed ω, and m of the ω_j nearest to ω, this is so. If this independence and normality are fictitiously assumed for all ω_j then $-2N^{-1}$ by the logarithm of the likelihood becomes

$$\log \det\left(\sum\right) + N^{-1} \sum_j \operatorname{tr}\left\{ I(\omega_j) f(\omega_j)^{-1} \right\}.$$

Here $f(\omega)$ is the spectral density matrix and Σ is the covariance matrix of the linear prediction errors. (A further approximation has been introduced that uses the expression of det Σ as the geometric mean of $\det\{2\pi f(\omega)\}$.) Since the physical quantities important in the properties of wave propagation, for example, influence the spectra rather directly, this likelihood, via a good deal of "sleight of hand", leads to many of the most useful final calculations. Of course all of this needs justification, which itself will be dependent on some of the more elaborate developments of the limit theorems sections. (The theory can also be misleading as when

$$x(n) = \rho x(n-1) + \varepsilon(n) + \gamma\varepsilon(n)x(n-1),\ \mathcal{E}\{\varepsilon(n)\} = 0,$$
$$\mathcal{E}\{\varepsilon(m)\varepsilon(n)\} = \delta_{mn},\ \rho^2 + \gamma^2 < 1,$$

and the $\varepsilon(n)$ are Gaussian, then all the above properties and consequent results hold for the $w(\omega_j)$ computed from the $x(n)$ but, of course, the true likelihood is not that being used, even in an approximate sense.)

In all of the statistical sections some details concerning algorithm construction will be needed and in some cases these may be quite extensive for almost always an optimisation of some nonlinear function is called for.

A second large development giving coherence to the subject is that of linear systems. One might reasonably define such a (stationary) system as one generating a random process for which the best linear predictor is the best predictor (in the least squares sense) so that the prediction errors (in the discrete time case) are martingale differences. The linear systems we have in mind here are further restricted to have rational transfer functions (from inputs to outputs). The subject links with the Fourier methods through these transfer functions. The theory of these systems involves an algebraic part (that describes the equivalence classes of systems i.e. classes corresponding to the same transfer function) resting on the properties of matrices of polynomials, a probabilistic part (which is quite elaborate in the continuous time case) and a topological part which arises in connection with the estimation of the equivalence class and in connection with which the manifold structure of the spaces of equivalence classes is important. The theory also links closely with parts of control theory. These systems also lend themselves to generalisations particularly in the direction of nonstationarity. They also lead to special nonlinear models which may prove more important than developments along the lines of the description of the general nonlinear filter with Gaussian input. This is because of the large numbers of parameters inevitably associated with the description of such filters. (Of course the elaborate theory associated with these concepts might nevertheless be included.)

Another range of problems is that associated with the statistical analysis of point processes (e.g. failures in computing apparatus). Here Fourier methods are of minor importance but martingale theory is again central. There is also a large development associated with extreme values of a process (e.g. in connection with the strength and stability of structures) much of which is not very fully developed *from a statistical viewpoint*. Of course the subject also shades off into control theory and the statistical theory of communication but these are rather separated from statistical time series analysis.

It is probable that no one book could adequately cover all of this material, even assuming a sophisticated mathematical audience. The book under review is not aimed at such an audience but rather at a readership possessed only of classical real variable knowledge, elementary linear algebra and some basic statistical technique. (Graduate students in economics were important in a course from which the book developed.) About one-sixth of the 450 pp.

of text is devoted to mathematics and probability theory (mainly Fourier concepts and limit theorems of probability). Some 40% is concerned with the simplest case of linear systems (namely scalar autoregressive-moving average processes), and a quarter with the definition of spectra and their estimation. The remainder is devoted to regression and the estimation of serial covariances. The book is fairly precisely written with results stated as theorems and there is a considerable amount of exemplification. The book is not without interest for a professional researcher in the subject. Sometimes the author shows a lack of depth of knowledge so that an inferior result is established at no saving in simplicity of proof. Thus he proves that, if the autocovariances converge to zero for a second order stationary process, then the sample mean converges to the true mean in mean square. An equally simple proof establishes the more perspicuous result that such convergence takes place if and only if the spectral function has no jump at the origin. This condition is then obviously also necessary and sufficient for almost sure convergence in the strictly stationary (finite mean square) case using the ergodic theorem, which is not mentioned in the book. The choice of material is sometimes, also, strange. For example a fairly lengthy discussion is given of the pointwise convergence of a Fourier series. The author may feel that this gives intuitive feeling but it would seem preferable merely to discuss the topic. One can also wonder whether a student who is interested in this subject at a theoretical level is likely to possess such a meagre mathematical background as that required of him.

Another criticism is that the book has a slightly "old fashioned" air about it, as if it were being published in 1970 not 1976. This relates particularly to the failure to introduce the general linear systems concepts. The Fourier techniques are used only to estimate spectra and cross spectra (and recent developments here are not mentioned) nor are they well linked with other sections.

Nevertheless the book is fairly successful and may obtain a considerable readership basically because it is carefully and correctly written by someone who understands reasonably well what he chooses to write about. The choice of topics is at least wide enough to give a good introduction to the subject and in relation to some, quite elaborate but useful, statistical techniques the book is sufficiently complete (including details of algorithm construction) to provide the information needed to effect the analysis of data.

REFERENCES

1. U. Grenander and M. Rosenblatt, *Statistical analysis of stationary time series*, Wiley, New York, 1957. MR **18** #959.

2. Yu. A. Rozanov, *Stationary random processes*, Holden-Day, San Francisco, 1963. MR **35** #4985.

3. Peter Whittle, *Hypothesis testing in time series analysis*, Thesis, Uppsala Univ., 1951. MR **12** #726.

E. J. HANNAN