BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 21, Number 1, July 1989

0273-0979/89 $1.00 + $.25 per page

Stationary sequences and random fields, by Murray Rosenblatt. Birkhäuser, Boston, Basel, Stuttgart, 1985, 258 pp., $34.95. ISBN 3-7643-3264-6

The statistical theory of stationary sequences (or time series) and random fields has acquired such a vast literature over the last thirty years that finding one's way through its various ramifications is no easy task. This elegantly produced monograph by one of our most distinguished experts on stationary stochastic processes consists of a selection of topics on the frontiers of current research in this field. A great deal of the material is based on the author's own work or a direct development of it.

The point of departure for all statistical work in time series is the Kolmogorov-Wiener theory which provides the mathematical foundations. A brief background of the latter will be helpful to the nonspecialist in understanding and appreciating the more modern developments.

The mathematical background. The setting of the linear prediction problem for stationary sequences in a Hilbert space context is due to Kolmogorov **[6]**. This can be seen for any second order process even without the stationarity assumption. A second order process $x = (x_t)$ is a family of real or complex-valued random variables $x_t(\omega)$ defined on a probability space $(\Omega, \mathscr{F}, \mathscr{P})$ such that $E|x_t|^2 < \infty$ for each t. (The symbol E stands for integration with respect to the measure $P(d\omega)$.) It is convenient to assume $Ex_t \equiv 0$. The index set over which t varies is either $\mathbf{Z}$ or $\mathbf{R}$. In the former, discrete time case, one speaks of a sequence. If $t \in \mathbf{R}$, (x_t) is assumed to be L^2-continuous in t. The terms sequence and process will be used interchangeably in this review.

Each x_t can be regarded as an element of the real (or complex) Hilbert space $L^2 = L^2(\Omega, \mathscr{F}, \mathscr{P})$, with inner product $(u, v) = E(u\overline{v})$. For purposes of linear prediction, the "history" of the process is given by the family of Hilbert spaces $H(x;t) :=$ closed linear subspace of L^2 generated by $\{x_s, s \leq t\}$. $H(x; -\infty) = \bigcap_t H(x;t)$ is called the remote past of x, $H(x;t)$, the past and present up to t and $H(x;\infty) := V_t H(x;t)$, the history of x. The problem is to obtain the optimal (i.e., least squares) predictor $\hat{x}_{t_1|t}$ of a future state x_{t_1} given $\{x_s, s \leq t\}$, $(t_1 > t)$. Since the optimality criterion requires that the prediction error $\|x_{t_1} - \hat{x}_{t_1|t}\| = \inf_{u \in H(x;t)} \|x_{t_1} - u\|$ the solution is given by the orthogonal projection $\hat{x}_{t_1|t} = \mathrm{Proj}_{H(x;t)} x_{t_1}$ which satisfies the equation

$$(\hat{x}_{t_1|t}, x_s) = (x_{t_1}, x_s), \quad \text{for all } s \leq t. \tag{1}$$

Prediction theory, therefore, is concerned with the finding of computable algorithms for the predictor and a computable formula for the error.

(x_t) is called a (weakly) stationary sequence, if $t \in \mathbf{Z}$ (a stationary process if $t \in \mathbf{R}$) and the covariance function $E(x_t\overline{x}_s) = r_{t-s}$. In the discrete time case,

$$r_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dF(\lambda)$$

and in the continuous time case,

$$r_t = \int_{-\infty}^{\infty} e^{it\lambda}\, dF(\lambda)$$

where F is called the spectral distribution function of x. These results, the first due to Herglotz and the second due to Bochner, Khinchin and Wiener, were already in existence before the work on prediction theory.

In 1938, applying the newly emerging theory of stochastic processes to time series analysis, H. Wold obtained a decomposition of a stationary sequence, now bearing his name. It was left to Kolmogorov to exploit the full significance of these developments. In his famous, beautiful 1941 paper, **[6]**, he gave a formulation of the theory in terms of the geometry of Hilbert space and made the Wold decomposition the basis for a deeper analysis of x itself. The process x is called regular or purely nondeterministic (PND) if $H(x; -\infty) = 0$, it is singular or deterministic if $H(x; -\infty) = H(x)$.

Wold decomposition. If $H(x; -\infty) \neq H(x)$, then (x_t) has the unique decomposition

$$x_t = x_t' + x_t'' \tag{2}$$

where (x_t') is PND, (x_t'') is deterministic and the Hilbert subspaces $H(x')$ and $H(x'')$ are mutually orthogonal.

It was shown in **[6]** that the sequence (x_t) is PND iff its spectral distribution is absolutely continuous and the spectral density satisfies the condition

$$\int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda > -\infty$$

(Paley-Wiener criterion for the discrete time case). Kolmogorov also obtained an H^2-factorization of the spectral density $f(\lambda) = |\psi(e^{-i\lambda})|^2$ where ψ is an outer function in Beurling's terminology. The process (x_t) has a one-sided moving average representation

$$x_t = \sum_{j=0}^{\infty} c_j \xi_{t-j} \qquad (E\xi_j \xi_k = \delta_{j-k}) \tag{3}$$

where the coefficients c_j are those of the function

$$\psi(z) = \sum_{j=0}^{\infty} c_j z^j. \tag{4}$$

The expression for the optimal predictor can be immediately written down from (3) and (4). The one-step prediction error is given by the Szegö formula

$$\|x_{t+1} - \hat{x}_{t+1|t}\|^2 = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda \right\}.$$

Wiener, working independently and somewhat later than Kolmogorov, mostly considered continuous parameter stationary processes [**9**]. He reduced the prediction problem to the solution of a "Wiener-Hopf" equation of which equation (1) is a prototype. His treatment even for this case was not as thorough as Kolmogorov's for the discrete time problem. Wiener himself gave only passing attention to stationary sequences or to time domain questions. (See, however, Masani's comments in [**7**].)

Kolmogorov's discussion of the prediction theory for one-dimensional stationary sequences was so complete that attempts to generalize it to the vector valued situation have enriched harmonic analysis, influencing among other areas, the theory of matrix-valued H^p functions and their factorization. A corresponding development of spectral theory for stationary random fields is very much the concern of present-day research.

Time domain analysis. The analysis of the Wold decomposition given in [**6**] can easily be identified as a special case of the more general and abstract form discovered by Halmos twenty years after the appearance of Kolmogorov's paper [**3**]. Let V be an isometry on a separable Hilbert space $\mathscr{H}$. Then $\mathscr{H}$ has the decomposition

$$\mathscr{H} = \mathscr{H}_\infty \oplus \sum_{k=1}^{\infty} \oplus V^k(R^\perp) \tag{5}$$

where $R = V(\mathscr{H})$. The spaces $V^k(R^\perp)$ and $V^j(R^\perp)$ $(k \neq j)$ are mutually orthogonal and the restriction of V to $\mathscr{H}_\infty$ is unitary. The Wold decomposition can be written in the form (5) if we make the following identifications: $\mathscr{H} = H(x, 0)$, $V =$ the restriction of U^{-1} to $H(x, 0)$ where U is the unitary operator associated with (x_t): $Ux_t = x_{t+1}$, $R^\perp = H(x; 0) \ominus H(x; -1)$ (the "innovation" subspace) and $\mathscr{H}_\infty = H(x; -\infty)$.

Extensions of the decomposition (5) to more general processes and its relevance to the invariant subspace problem is yet another link, now well understood, between prediction theory and harmonic analysis.

The development of a nonlinear or nonstationary theory of prediction, interpolation and filtering has, not surprisingly, concentrated on the time domain approach. (See Cramér [**1**] for the beginnings of a spectral theory of nonstationary harmonizable processes and also the recent book of Priestley [**8**].)

The theory of representing a second order PND process (x_t) in terms of innovation processes in the discrete parameter case (or differential innovation processes in the continuous parameter case) is due to Hida and Cramér [**1**, **4**]. The best linear predictor can be directly obtained from such a representation.

The modern nonlinear theory essentially views the processes as stochastic dynamical systems and formulates the problem in terms of stochastic differential equations. Hilbert space techniques are inappropriate and have to be replaced by structural assumptions of a different sort. (See the article by Kallianpur in [**5**].)

Statistical analysis. Some general comments might be in order pertaining to the statistical problems treated in this book. An estimate of an unknown parameter θ appearing in the distribution of the process in question, is any measurable function $T_N(x_1, \ldots, x_N)$ of the observations $x_1, \ldots, x_N$. $\{T_N\}$ is a consistent estimator of θ if $T_N \to \theta$ in probability (or almost surely) as $N \to \infty$. Among consistent estimators, a further desirable property is that $a_N[T_N - \theta]$ have a limiting distribution G (usually Gaussian) as $N \to \infty$, for suitable norming constants $a_N \to \infty$. Both these properties are of an asymptotic character. There are other properties for fixed sample size: The estimate T_N is unbiased if, for each N, $ET_N = \theta$. Among all unbiased estimates, one might look for one whose variance is smallest, i.e., for which $E(T_N - \theta)^2$ is a minimum. In most statistical problems in time series it is the asymptotic properties that are within our reach.

From the point of view of applications, the solution of the prediction problem (the same is true of filtering and interpolation, not discussed here) provided by the Kolmogorov-Wiener theory cannot be used since, in practice, the spectral density (and so also the covariance function) is unknown and has to be estimated from the data. The estimation problems connected with the spectral density may be posed at three different levels:

(a) The functional form of the density $f(\lambda, \theta)$ is known but depends on a finite number of real parameters $\theta = (\theta_1, \ldots, \theta_k)$. An important class of processes that falls under this heading is the ARMA (autoregressive moving average) model.

(b) Estimation of $f(\lambda)$ at a particular frequency λ.

(c) Estimation of the function $f(\cdot)$. This is an example where the parameter is infinite dimensional and is not treated in the book.

The main difference in technique with similar problems in other areas of statistics is the difficulty caused by the fact that the observations are not stochastically independent but are data coming from a stationary sequence. Two assumptions that are most often made to deal with this difficulty (in addition, of course, to other assumptions having to do with smoothness or the existence of moments) are strict stationarity and the strong mixing condition: The process (x_t) is strictly stationary if the joint distribution of every arbitrary vector $(x_{t_1}, \ldots, x_{t_k})$ is invariant under time shifts.

A strictly stationary process (x_t) is strongly mixing if the following condition holds: Let $\mathscr{B}_n$ be the σ-field generated by $\{x_k, k \leq n\}$ and $\mathscr{F}_n$, the σ-field generated by $\{x_k, k \geq n\}$. Then

$$\sup_{\substack{B \in \mathscr{B}_0 \\ F \in \mathscr{F}_n}} |P(B \cap F) - P(B)P(F)| \equiv \alpha(n) \to 0 \quad \text{as } n \to \infty.$$

The strong mixing condition essentially says that the dependence of the sequence is short range, that is, the past and future of (x_t) are asymptotically independent. It was introduced by Rosenblatt in 1956 and has proved indispensable for establishing convergence to a Gaussian distribution. It is used in the book to derive the asymptotic normality of a large class of spectral density estimates in Problem (b). An important example of the latter is the smoothed periodogram

$$f_N(\lambda) = \int_{-\pi}^{\pi} W_N(\mu - \lambda) I_N(\lambda)\, d\lambda$$

where the periodogram

$$I_N(\lambda) = \frac{1}{2\pi N}\left|\sum_{k=0}^{N-1} x_k e^{-ik\lambda}\right|^2 \quad \text{and} \quad W_N$$

is a sequence of suitably chosen weight functions.

Problem (a). Linear processes and ARMA models. A PND stationary sequence has the representation

$$x_t = \sum_{j=-\infty}^{\infty} \alpha_j \xi_{t-j} \quad \text{where} \quad \sum_{j=-\infty}^{\infty} |\alpha_j|^2 < \infty$$

and (ξ_t) is a white noise, i.e. $(\xi_t, \xi_s) = \delta_{t-s}$. (x_t) having the above representation is a linear process if α_j is real and the (ξ_t) are real, independent, identically distributed random variables with $E\xi_t \equiv 0$ and $E\xi_t^2 \equiv 1$. The analysis of such processes throws light on the general theory. A widely studied example of a linear process is the ARMA model. A process (x_t) is ARMA of order (k, l) if it satisfies a relation

$$\sum_{p=0}^{k} \beta_p x_{t-p} = \sum_{q=0}^{l} \alpha_q \xi_{t-q},$$

$\alpha_0, \beta_0 \neq 0$, where (ξ_t) are independent, identically distributed random variables with $E\xi_t \equiv 0$, $E\xi_t^2 = \sigma^2 > 0$. The spectral density of (x_t) has the form

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\alpha(e^{-i\lambda})}{\beta(e^{-i\lambda})}\right|^2$$

where $\alpha(z) = \sum_{q=0}^{l} \alpha_q z^q$ and $\beta(z) = \sum_{p=0}^{k} \beta_p z^p$. The ARMA process is thus a finite parameter model mentioned in (a). The parameters (α_q) and (β_p) are estimated using a modified maximum likelihood method. Consistency and asymptotic normality are also established.

Non-Gaussian linear processes, i.e., where (ξ_t) are non-Gaussian exhibit, in some respects, a strikingly different behavior from the Gaussian linear processes. They are studied in connection with the deconvolution problem, the determination of the sequence (ξ_t) from the (x_t).

Probability density and regression estimates are considered in a separate chapter. The work of Brillinger and Rosenblatt on cumulant spectral estimates is presented in detail in Chapter VI.

The core of the book has to do with the large sample theory of covariance, spectral and cumulant spectral estimates. The limiting distributions are normal when the stationary process in question satisfies a strong mixing condition. This situation covers a wide class of processes met with in practice. However, recent research has uncovered exotic non-Gaussian limiting distributions for a large class of stationary processes exhibiting long range dependence. One example (possibly the first such) due to Professor Rosenblatt himself is described in Chapter II. More recent work in this direction, by him, by Dobrushin and Major and others is briefly alluded to in one of the problems at the end of the chapter. One wishes that this exciting work were given more prominence and space in the book.

Random fields. Let $\mathbf{n} = (n_1, \ldots, n_k) \in \mathbf{Z}^k$, $k \geq 2$. A real or complex-valued stochastic process or sequence $(x_\mathbf{n})$ parametrized by the lattice points $\mathbf{n}$ is called a k-parameter weakly stationary random field if $Ex_\mathbf{n} \equiv a$ and $\operatorname{Cov}(x_\mathbf{n}, x_\mathbf{m}) \equiv E(x_\mathbf{n} - a)\overline{(x_\mathbf{m} - a)} = r_{\mathbf{n}-\mathbf{m}}$ depending only on the difference $\mathbf{n} - \mathbf{m}$.

Many of the above problems—ARMA models, strong mixing, parameter estimation, spectral density estimation, etc., are also considered for random fields. Though random fields form a part of the title of the book, the author has chosen not to treat it as a separate topic. Problems on random fields are scattered in several sections spread over five chapters, and I am not sure that this diffusion does justice to this important area. The spectral theory for Helson and Lowdenslager's prediction for the half-plane is carried out in some detail in Chapter VIII (though without complete proofs).

Since spectral techniques and harmonic analysis are the basic analytical tools of the book, it is understandable that so little space is devoted to time-domain related questions. A discussion of the time-domain approach would have been particularly appropriate for random fields since the relevant spectral theory parallels the definitions one chooses for "past" and "future" (e.g., the half-plane or quarter plane prediction problem).

The section on the Kalman-Bucy filter is too brief and stands in isolation in Chapter II. Here too, an important feature that distinguishes it and nonlinear filtering theory generally from the prediction (or filtering) theory of stationary sequences is the dynamical or time-domain formulation that is essential to the former. Stationarity is neither assumed nor available in general, so spectral techniques cannot be used. Harmonic analysis gives way to (stochastic) differential or difference equations which furnish recursive filters or estimates. Some comments on these points would have made this section serve as a bridge to other vistas of the subject.

As the author indicates in the preface, the book may be used either as a text for a one-semester course for advanced undergraduate students or as a springboard for seminars for research students. In my opinion, it is ideally suited for the latter purpose: the presentation, perfect for use in seminars, focuses directly on the problem under consideration with a minimum of digression. The problems and notes at the end of each chapter provide illuminating information on further developments and related research. Many of the proofs are presented in an informal manner, possibly, in order to avoid cluttering up the book with tedious details. The latter will have to be filled in by the student who might also have to be on his guard against the deceptive simplicity with which some of the deeper ideas are introduced. Though the book is formally self-contained, the introductory material is given in a highly condensed form and students new to the field will derive full benefit from it after they have had a course that takes them in a leisurely fashion through the more traditional parts of the subject. Alternatively, they may supplement Chapter I by studying the basic material given in Chapter 2 of the author's earlier book with U. Grenander [2].

I cannot conclude this review without referring to a novel and pleasing feature of the book—the fairly detailed and interesting discussion of turbulence, a topic in which the author has been interested for many years. As far as I am aware, this is the first book (in English) on stationary processes to include a treatment of this problem.

REFERENCES

1. H. Cramér, *On some classes of nonstationary processes*, Proc. Fourth Berkeley Sympos., vol. II, Univ. of California Press, Berkeley and Los Angeles, 1961, pp. 57–78.

2. U. Grenander and M. Rosenblatt, *Statistical analysis of stationary time series*, John Wiley and Sons, New York, 1957.

3. P. R. Halmos, *Shifts on Hilbert spaces*, J. Reine Angew. Math. **208** (1961), 102–112.

4. T. Hida, *Canonical representations of Gaussian processes and their applications*, Mem. Coll. Sci. Univ. Kyoto Ser. A **33** (1960), 109–155.

5. G. Kallianpur, *Some ramifications of Wiener's ideas on nonlinear prediction* (in Norbert Wiener: Collected Works, Vol. III, MIT Press (P. Masani, ed.), Cambridge, Mass., 1981, pp. 402–425).

6. A. N. Kolmogorov, *Stationary sequences in Hilbert space*, Bull. Math. Univ. Moscow **2** (1941), 40 pp. (Russian)

7. P. Masani, *Commentary on the prediction-theoretic papers*, Norbert Wiener: Collected Works, Vol. III, MIT Press, Cambridge, Mass., 1981, pp. 276–306.

8. M. B. Priestley, *Non-linear and non-stationary time series analysis*, Academic Press, London, New York, 1988.

9. N. Wiener, (A comprehensive survey of Wiener's work on prediction is given in [7].

10. H. Wold, *A study in the analysis of stationary time series*, Almquist and Wiksells, Uppsala, 1938.

G. KALLIANPUR
UNIVERSITY OF NORTH CAROLINA

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 21, Number 1, July 1989

0273-0979/89 $1.00 + $.25 per page

Free rings and their relations, by P. M. Cohn, second edition, Academic Press, London Mathematical Society Monograph No. 19, 1985, xxii + 588 pp., $96.00. ISBN 0-12-179152-1

The fundamental theorem of combinatorial group theory is that a subgroup of a free group is free. In contrast, subalgebras of free (associative) algebras are not well understood, and at the moment defy classification. This is not too surprising: in going from group theory to ring theory the translation of "subgroup" is (one-sided) "ideal." Thus the correct, and fundamental, theorem is that one-sided ideals of free algebras are free submodules. This is a consequence of a "weak algorithm" that holds in the