Applied Probability Trust (19 May 2014)

ASPECTS OF PREDICTION

N. H. BINGHAM,* *Imperial College London*

BADR MISSAOUI,* *Imperial College London*

Abstract

We survey some aspects of the classical prediction theory for stationary processes, in discrete time (§1). We turn in §2 to continuous time, with particular reference to reproducing-kernel Hilbert spaces and the sampling theorem. We discuss the discrete-continuous theories of ARMA-CARMA, GARCH-COGARCH and OPUC-COPUC in §3. We compare the various models treated in §4 by how well they model volatility, in particular volatility clustering. We discuss the infinite-dimensional case in §5, and turn briefly to applications in §6.

Keywords: Stationary process; Kolmogorov isomorphism theorem; time series; stochastic volatility; volatility clustering; Banach space; covariance operator; locally convex; reproducing-kernel Hilbert space; sampling theorem.

2010 Mathematics Subject Classification: Primary 60:02

Secondary 62:02

1. Prediction theory of stationary sequences

We begin with the most classical aspect: the prediction theory of stationary stochastic processes $X = (X_n)$ in discrete time n and one dimension (we take X complex-valued). There is a good account of this subject up to 1958 in Grenander and Szegő [41], and a great deal of relevant recent work in the books of Simon [91], [92] and [93]. We use as our main reference here the recent surveys by the first author [10] and [11] on the probabilistic side, together with its statistical sequel [12].

We take X zero-mean, with finite variance, (wide-sense) stationary, with autocovariance function $\gamma = (\gamma_n)$, $\gamma_n = E[X_n\overline{X_0}]$ (the variance is constant by stationarity, so we

* Postal address: Department of Mathematics, Imperial College London, South Kensington Campus, LONDON, SW7 2AZ.

may take it as 1 for convenience). Let $\mathcal{H}$ be the Hilbert space spanned by $X = (X_n)$ in the L_2-space of the underlying probability space, with inner product $(X, Y) := E[X\overline{Y}]$ and norm $\|X\| := [E(|X|^2)]^{1/2}$. In the complex plane, we write $\mathbb{T}$ for the unit circle, the boundary of the unit disc $\mathbb{D}$, parametrised by $z = e^{i\theta}$; unspecified integrals are over $\mathbb{T}$.

1.1. The Cramér Representation and Kolmogorov Isomorphism Theorem

There are four key ingredients here.

(i) There is a process Y on $\mathbb{T}$ with orthogonal increments with Fourier coefficients X_n, by the *Cramér representation*

$$X_n = \int e^{in\theta} dY(\theta) \qquad \forall\, n \in \mathbb{Z} \tag{CR}$$

(Cramér [24] in 1942, [25] §§5.6, 5.7, 7.5; cf. [28], X.4); Y is called the *Cramér process*. One can regard Y as a random measure, in which case Y is *orthogonally scattered* (the term is due to Masani in 1968, [72]).

(ii) There is a measure μ on $\mathbb{T}$, the *spectral measure*, with

$$E[Y(A)\overline{Y(B)}] = \mu(A \cap B).$$

In differential notation, this can be written

$$E[dY(\theta)^2] = d\mu(\theta), \tag{SM}$$

which shows the link with Itô calculus. See e.g. Doob [28], IX.1, X.4.

(iii) The autocorrelation function γ then has the *spectral representation*

$$\gamma_n = \int e^{-in\theta} d\mu(\theta) \tag{SR}$$

(Herglotz's theorem of 1911).

(iv) One has the *Kolmogorov isomorphism* between $\mathcal{H}$ (the *time domain*) and $L_2(\mu)$ (the *frequency domain*) given by

$$X_t \leftrightarrow e^{it\cdot} : \theta \in \mathbb{T} \mapsto e^{it\theta} \tag{KIT}$$

(Kolmogorov [61] in 1941) for integer t (as time is discrete). A good historical account of Kolmogorov's work in this area is in Shiryaev [90], 'The forties (1940-49)'. For modern

textbook accounts of the background, see e.g. Janson [51], Ch. 13 (and Notes to Ch. 13), Kallenberg [55], 7.4. Stochastic processes expressible as the Fourier transform of another process are called *harmonizable* (the term is due to Loève [67]; [68], 37.4); see Karhunen [56], Rao [83].

To exclude trivialities, it is necessary to take μ to have infinite support (to be *non-trivial*: Simon [91], p.1); we shall do so in what follows, without further comment.

For reasons of space, we omit state-space models and the Kalman filter; see e.g. [20], Ch. 12.

1.2. Verblunsky's theorem and the partial autocorrelation function (PACF)

By the Kolmogorov Isomorphism Theorem, all the information relevant to the time-evolution of the process is encoded in the spectral measure μ (though other relevant information, such as the distribution of the X_n, is not; see below). Now to model μ one has to proceed non-parametrically, as the situation is infinite-dimensional, and it might be preferable to use a sequence representation rather than a measure (or density-function) representation. This option is available to us, thanks to *Verblunsky's theorem* (S. Verblunsky in 1935 and 1936; the result is named and proved in [91], to which we refer for details), according to which one has a bijection (the *Verblunsky bijection*) between the spectral measures μ (probability measures on $\mathbb{T}$) and sequences $\alpha = (\alpha_n)_{n\in\mathbb{Z}}$, with each $\alpha_n \in \mathbb{D}$ (α is the *Verblunsky sequence*, or *partial autocorrelation function (PACF)*; α_n are the *Verblunsky coefficients*). Verblunsky discovered this result in analysis; it was re-discovered much later in statistics by Barndorff-Nielsen and Schou in 1973 and Ramsey in 1974. It is extremely useful in statistics as it provides an *unrestricted parametrisation* (all sequences with terms in the unit disc are possible; this is not the case for the autocorrelation function (ACF) γ). Then α_n is the correlation between the residuals of X_0 and X_n when regressed on the values $X_1, \ldots, X_{n-1}$ in between. Using the usual notation for Hilbert-space projections,

$$\alpha_n := corr(X_n - P_{[1,n-1]}X_n, X_0 - P_{[1,n-1]}X_0);$$

see e.g. [10] for details and references.

1.3. Orthogonal polynomials on the unit circle (OPUC)

In Szegő's classic book [94] on orthogonal polynomials (on the real line: OPRL), one chapter (Ch. XI) is devoted to orthogonal polynomials on the unit circle (OPUC). Although Verblunsky is not mentioned in [94], his work is the guiding theme in the two-volume book [91], [92]; we summarise here what we need. Recall from OPRL the key role played by the three-term recurrence relation (Favard's theorem); this involves two sequences of coefficients. In OPUC, there is again a three-term recurrence relation, the *Szegő recursion*, but now only one sequence of coefficients (the Verblunsky coefficients above). For each probability measure μ on $\mathbb{T}$ (the spectral measure above), write P_n for the monic orthogonal polynomials they generate (by Gram-Schmidt orthogonalization). For every polynomial Q_n of degree n, write

$$Q_n^*(z) := z^n \overline{Q_n(1/\bar{z})}$$

for the *reversed polynomial*. Then the Szegö recursion is

$$P_{n+1}(z) = zP_n(z) - \bar{\alpha}_{n+1} P_n^*(z).$$

Here the parameters α_n are the Verblunsky coefficients, and lie in $\mathbb{D}$ by Verblunsky's theorem ($|\alpha_n| \leq 1$ as correlations lie in $\mathbb{D}$ automatically; that $|\alpha_n| < 1$ follows from μ being non-trivial).

The Szegő recursion is known in the time-series literature as the *Levinson-Durbin algorithm*. The best linear predictor of the immediate future based on the present and a finite segment of the past is a linear combination of the given values; their coefficients form a triangular matrix $A = (a_{nk})$; the Verblunsky coefficients or PACF are the diagonal elements in A. See e.g. Brockwell and Davis [20] §§3.4, 5.2, [10], §3.

1.4. The Wold decomposition

Write w for the *spectral density* in the Lebesgue decomposition of μ:

$$\mu = \mu_s + \mu_{ac} = \mu_s + w d\theta/2\pi.$$

Write σ^2 for the one-step mean-square prediction error:

$$\sigma^2 := E[(X_0 - P_{(-\infty,-1]} X_0)^2].$$

Call X *non-deterministic* (ND) if $\sigma > 0$, *deterministic* if $\sigma = 0$. The *Wold decomposition* of 1938 ([25] 5.7, [20] §5.7; [10] §3) expresses a process X as the sum of a non-deterministic process U (the 'good' part) and a deterministic process V (the 'bad' part):

$$X_n = U_n + V_n;$$

the process U is a moving average,

$$U_n = \sum_{j=-\infty}^{n} m_{n-j}\xi_j = \sum_{k=0}^{\infty} m_k \xi_{n-k},$$

with the ξ_j – the *innovations* at time n – zero-mean and uncorrelated, with each other and with V; $E[\xi_n] = 0$, $var(\xi_n) = E[\xi_n^2] = \sigma^2$. Thus when $\sigma = 0$ the ξ_n are 0, U is missing and the process is deterministic. When $\sigma > 0$, the spectral measures of U_n, V_n are μ_{ac} and μ_s, the absolutely continuous and singular components of μ (again, the 'good' and 'bad' parts). Then the Wold decomposition agrees with the Lebesgue decomposition of the spectral measure; in view of the Cramér representation, this is often called *Wold-Cramér concordance*. It fails in dimension higher than one; see §5.

1.5. Szegő's theorem

For a process genuinely evolving in time one expects new information as time passes: $\sigma > 0$, non-determinism. The condition for this is *Szegő's condition* (Sz) below. Szegő's theorem gives the equivalence of the following:

(i) $\sigma > 0$;

(ii) $\log w \in L_1(\mathbb{T})$, i.e. (as $\int \log w < \infty$ by Jensen's inequality)

$$\int \log w(\theta) d\theta > -\infty. \tag{Sz}$$

(iii) $\alpha \in \ell_2$.

Then $\sigma^2 = \prod_1^\infty (1 - |\alpha_n|^2)$, and (Kolmogorov's formula [61]) σ^2 is the geometric mean $G(\mu)$ of μ:

$$\sigma^2 = \exp\big(\frac{1}{2\pi}\int \log w(\theta) d\theta\big) =: G(\mu) > 0 \tag{K}$$

(see e.g. [10] §4 for details and references). (The Szegő condition is visibly of entropy type, and is related to the Gibbs variational principle of statistical mechanics.)

1.6. The Szegő function

We now restrict to when (Sz) holds. Then the *Szegő function*

$$h(z) := \exp\Big(\frac{1}{4\pi}\int\Big(\frac{e^{i\theta}+z}{e^{i\theta}-z}\Big)\log w(\theta)d\theta\Big) \qquad (z\in D) \tag{OF}$$

exists, and is an *outer function* (hence (OF)) in the Hardy space $H_2(\mathbb{T})$ on the torus $\mathbb{T}$ (see e.g. [10] §4). It is an 'analytic square root' of w:

$$|h(e^{i\theta})|^2 = w(\theta): \qquad h=[w^{1/2}]$$

(Nikolskii [77], Vol. 1, p.380; [76]). One has (see [49] §2, [48] for background) $h(z) = \sum_0^\infty m_n z^n$: the Maclaurin coefficients $m=(m_n)$ of h are the moving-average $(MA(\infty))$ coefficients in the Wold decomposition. That of $-1/h(z) = \sum_{n=0}^\infty r_n z^n$ $(z\in D)$ gives the $AR(\infty)$ coefficients $r=(r_n)$ in the (infinite-order) *autoregression* $\sum_{j=-\infty}^n r_{n-j}X_j + \xi_n = 0$ $(n\in Z)$.

2. Continuous time

As always, one has to choose from context between modelling our process in discrete or continuous time. Such matters are topical: the ARMA and GARCH models just mentioned have led on to extensive work on continuous-time versions such as CARMA and COGARCH.

Within the context of econometrics: data are discrete, and as a result the discrete-time theory is the more developed. But continuous time offers advantages, such as a setting in which one can use calculus, and model the dynamics directly. The classic text in this area is that of Bergstrom [9]. Continuous-time econometrics, and finance, has flourished as a result of the extensive use of stochastic calculus, in Black-Scholes(-Merton) theory etc. (as the title of [52] bears witness).

First, one has the obvious but strikingly useful fact that the Cramér representation (CR) of §1 allows one to pass directly to continuous time, simply by replacing the discrete time n there by a continuous time t:

$$X_t = \int e^{it\theta}dY(\theta) \qquad (t\in\mathbb{R}). \tag{CR}$$

This emphasises dramatically the key structural role played here by the Cramér process Y. Of course, this procedure commits us to a choice of unit of time; it works best where

the context has a natural one.

Note first the nature of the process $X = (X_t)$ thus defined. It is *harmonizable* (§1); see Lloyd [66], Niemi [75] for early work on such harmonizable processes. In engineering language, the compact support $[-\pi, \pi]$ of Y makes Y *band-limited*, whence X_t is a random analytic function of t ([25] §7.3, Belaev [8]). By the Paley-Wiener theorem ([78]; [58] VI.7), X is a *random entire function*, of *exponential type* π.

2.1. The sampling theorem, Paley-Wiener spaces and RKHS

The Paley-Wiener spaces PW_a (with general $a > 0$ rather than $a = \pi$ as here) are the motivating examples of *de Branges spaces* ([17]; [32] §6.1). They are also classical examples of *reproducing-kernel Hilbert spaces (RKHS)* ([77], Vol. 1, B, §6.5.2). What is needed for a Hilbert space to have a reproducing kernel is that the point-evaluation maps be continuous; thus Hilbert spaces of continuous (and in particular, analytic) functions have reproducing kernels. The reproducing kernel for PW_a is the sinc *function* (sinus cardinalis: this stems from E. T. Whittaker's work of 1915; the name cardinal series is from J. M. Whittaker in the 1920s; the name sinc is from P. M. Woodward in 1953). For $f \in PW_\pi$, one has

$$f(x) = \sum_{n \in \mathbb{Z}} f(n) \operatorname{sinc}(x - n) \quad \forall\, x \in \mathbb{R}, \qquad \operatorname{sinc}(x) := \frac{\sin \pi x}{\pi x}.$$

This recovery of a function everywhere on the line from its values sampled at the integers is called the *sampling theorem* (or Whittaker-Shannon-Kotelnikov sampling theorem: J. M. Whittaker [97] in 1935, C. E. Shannon in 1949, V. I. Kotelnikov in 1933). For background, see e.g. Partington [79] §7.2, Higgins ([44] §2.6.3, [45], [46], [47]).

This works more generally. That any RKHS H with kernel K with a total orthogonal set of point evaluation vectors δ_{x_n} has the sampling property

$$f(x) = \sum_{n \in \mathbb{Z}} f(x_n) K(x, x_n) \qquad (f \in H) \tag{SP}$$

is Kramers' sampling theorem [63]. Martin [71] considers such results from the point of view of operator theory, and shows in particular that many de Branges spaces have this property.

The sampling theorem is clearly closely connected with the Paley-Wiener theorem

from the above, and is in fact an easy corollary of it (Dym and McKean [32], §2.2); similarly for the Poisson summation formula (Boas [14]). Butzer et al. [22] show the equivalence of all three, and of the general Parseval formula, in the band-limited case.

2.2. The Kolmogorov-Wiener filter

The prediction problem given the infinite past was studied in continuous time by Kolmogorov [61] in 1941 (see [90]), M. G. Krein in 1944, and also by Wiener [98]; see Doob [28], XII. Under (Sz), as before the Szegő function h exists, and is in the Hardy space H_2 on the upper half-plane; again, $w = |h|^2$. With $\hat{h}$ the Fourier transform of h, the mean-square error of the best linear predictor of X_T given $\{X_t : t \leq 0\}$ is ([31], §5) $\int_0^T |\hat{h}(t)|^2 dt$.

Wiener's work was done, independently of Kolmogorov, during 1940-41, restricted for security reasons during WWII, circulated subject to this (as the "Yellow Book"), and published in 1949. It contains (p. 55, 59) Wiener's detailed account of the connections between his work and Kolmogorov's, and also two mathematical appendices by Wiener's colleague and former pupil Levinson. Volume III of Wiener's Collected Works [99] contains thirteen of his papers on prediction (published 1949-59; six with Masani, one with Akutowicz), together with commentaries (by Kailath, Akutowicz, Masani, Salehi, Muhly and Kallianpur). These papers contain further discussion by Wiener, of the connections between his work and Kolmogorov's, and between discrete and continuous time.

2.3. Prediction given a finite past

The continuous-time prediction problem given a finite segment of the past is harder, and has been studied by Krein (1944-54) and by Dym and McKean [30], [31] and [32]. Both methods use the Kolmogorov isomorphism $X_t \leftrightarrow e^{it\cdot}$ of §1 to map the given segment $\{X_t : -T \leq t \leq T\}$ to the Paley-Wiener space PW_T. Tools used include Krein's theory of strings and results of Levinson and McKean [64]. Matters are too technical to attempt a summary here; for details, see [30], [31], §§7-23, [32], Ch. 4.

The work of [64] has also been found useful in discrete time ([12], §7.4, [57]).

3. ARMA/CARMA; GARCH/COGARCH; OPUC/COPUC

The general stationary process is an infinite-dimensional object (parametrised several different ways in §1), but in practice one has to truncate and work finite-dimensionally. One familiar way to do this is to use ARMA (autoregressive moving average) models; see e.g. Box et al. [16], [20] Ch. 3, 4, 8. This is flexible and practicable, but the $p+q$ parameters of the $ARMA(p,q)$ model chosen may have little or no meaning in reality.

As in §2, one may prefer to work in continuous time (CARMA). Here the polynomials in the $ARMA(p,q)$ model in the lag operator (equivalently, the (forward or backward) difference operator) are replaced by polynomials in the differentiation operator D. See for example Brockwell [18], [19].

Variance is often not constant as above, but varies, perhaps randomly; ARCH (autoregressive conditionally heteroscedastic) models were introduced to apply time-series methods here, and these were generalised to GARCH models, for which see e.g. [40]. The general case is $GARCH(p,q)$; $GARCH(1,1)$ is fairly widely applicable, and we confine ourselves to it here for simplicity.

As before, one can generalise GARCH to continuous time – COGARCH; see e.g. Klüppelberg et al. [59],[60], Brockwell et al. [21]. Again as before, $COGARCH(1,1)$ suffices for many purposes, so we confine ourselves to it here (Haug et al. [43]).

The theory of OPUC, used in §1 in discrete time, finds its continuous generalisation in the theory of Krein systems [27], [37], and in particular Krein's theory of strings [32].

4. Stochastic volatility and volatility clustering

Not only does volatility vary, so that stochastic volatility models are used (as in the GARCH models of §3), one of the stylised facts of mathematical finance is *volatility clustering*. Economic life typically consists of periods of normality punctuated by crises; the crises give rise to clusters of high volatility. This clustering of extremes happens quite generally [2]; for background on extremes, see e.g. [35].

In the GARCH case, stationarity requires restrictions on the parameters [74]. These involve Liapounov exponents (and Kingman's subadditive ergodic theorem). These can be calculated explicitly in the GARCH(1,1) case, which is usually all that is needed

in practice ([73], §5.3.2). Stationarity also brings in a stochastic recurrence equation (SRE), whose solution again involves Liapounov exponents. One obtains regular variation of the tails of the stationary process ([73] §5.5.2; this uses a result of Kesten in 1973 in the multi-dimensional case and of Goldie in 1991 in one dimension). A stationary GARCH process is strongly mixing and has exponentially decaying correlations, under general conditions (of Markov character: [73], §5.6.1). To recapitulate: for stationary GARCH(1,1), one has heavy (regularly varying) tails and exponentially decreasing correlations.

This connection between the time behaviour (correlation decay) and the distribution at fixed times seems to be a reflection of the Markovian nature of the process (GARCH in this case). In the stationary case of §1, one can have any correlation structure and any distribution; there is no link between them, even when the process is Markovian. For, any distribution can be the stationary distribution of a Markov chain, as in the Metropolis-Hastings algorithm (see e.g. [87]). Although the Cramér process Y encodes the time behaviour of X in (CR), the distribution of $X_0 = \int_{\mathbb{T}} dY = Y(\mathbb{T})$ (the total mass of the complex measure Y) can be anything. One has no way to model volatility clustering, and this restricts the suitability of the prediction theory of §1 for financial time series.

There is, however, no difficulty in capturing volatility clustering in other ways, in one or many dimensions; see e.g. [13]. In finance, the dimensionality d is often high, to reflect the holding of a balanced portfolio of many risky assets, by Markowitzian diversification. The *multivariate elliptic processes* of [13] model the time-evolution of such a d-dimensional time series using the theory of *elliptical distributions* [33]. One works *semi-parametrically*. The *parametric* part of the model is (Σ, μ), where Σ is the $d \times d$ covariance matrix and μ is the d-vector of means, both essential ingredients, from Markowitz' work (think of risk and return together; (co)variances measure risk, means measure return). The *non-parametric* part is a one-dimensional process, the *risk driver* R. Thus the burden of dimensionality is born by (Σ, μ); R models the ambient economic or financial climate. Though this is certainly not one-dimensional in reality, one often works with a one-dimensional proxy for it, such as a stock-exchange index.

5. Higher dimensions

Much of the one- and finite- dimensional theory (for which see e.g. [10], [11]) extends to infinitely many dimensions, but not all of it does. The simplest infinite-dimensional setting is a (separable) Hilbert space; beyond that one has (separable) Banach space; beyond that, locally convex topological vector spaces. Our main references here are the books of Vakhania et al. [95] on Banach spaces, Kakihara [53], Schwartz [89] on cylindrical measures and the Pesi Masani Volume [70], and the papers of (Applebaum and) Riedle [85], [86], [6].

5.1. Hilbert space

The simplest infinite-dimensional setting is Hilbert space, and the theory here was developed by Payen in 1967 [80]. He obtains the Wold decomposition [80, II.3], with two components, as above. He deals with *factorizability* ([80, II.5]; see below), and hence obtains [80, II.6] a three-term decomposition, one component corresponding to the singular component of the spectral measure as before, while the 'absolutely continuous component' splits into two, one part corresponding to the 'largest factorizable part', the other to the residue. He studies the Szegő condition via a series decomposition into rank-one components [80, II.8], and also continuous time [80, III].

The Hilbert-space setting brings closer together the stationary and non-stationary cases (for commentary on the distinction, see [12]). The key concept here is *harmonizability* (due to Loève in 1955, and subsequently studied by Cramér, M.M. Rao and others). A (second-order) process $X = (X_t)$ is *harmonizable* if it is the Fourier-Stieltjes transform of a random measure Y. If Y is orthogonally scattered, X is stationary and so harmonizable. There is no converse, but on a Hilbert-space H, X is harmonizable iff it has a stationary dilation, i.e. it is the projection of a stationary process in some larger Hilbert space K. For details see Richard [84], Kakihara [54].

The Hilbert-space methods used by Kolmogorov in 1941 to prove (KIT) are tantamount to those used by Aronszajn [7] in 1950 in his theory of reproducing-kernel Hilbert spaces (RKHS). Accordingly, the result in the Hilbert-space setting is often called the *Aronszajn-Kolmogorov theorem*. See [95] III.1.3 for positive-definitive functions (covariance kernels), IV.4.3 for orthogonally scattered measures, [69].

The sampling theorem was extended to Hilbert space by Weston in 1949 [96].

5.2. Banach spaces

As mentioned above, our main reference for the Banach case is [95], which while it does not consider prediction theory explicitly is nevertheless well adapted to our purposes here.

In finite-dimensional spaces, linear prediction is always possible. In infinitely many dimensions, this is no longer so. It is possible in the space $L(Y, H)$ of continuous linear operators from a Banach space Y to a Hilbert space H [23, §3].

The results above of Payen [80] on the Wold-Cramér concordance in Hilbert space were extended to the Banach case by Schmidt [88] and Hajduk-Chmielewska [42].

For V, W topological vector spaces, write V^* for the dual of V, $L(V, W)$ for the space of continuous operators (linear maps) from V to W. Then V has the *factorization property* if every non-negative $A \in L(V, V^*)$ can be factorized as

$$A = T^*T$$

with T a continuous operator from V into some Hilbert space (depending on A); cf. [1, §7.3]. The factorization property is equivalent to the continuity of $v \mapsto < Av, v >$ for a non-negative $A \in L(V, V^*)$; this holds for Banach spaces, and for some but not all locally convex topological vector spaces [39]. The Aronszajn-Kolmogorov theorem can be extended to locally convex with the factorization property [39] and so can some aspects of prediction theory (Alpay et al. [3]). For harmonizability, see Richard [84].

5.3. Gaussianity and prediction

Linear prediction is always possible if one restricts attention to the Gaussian case. Even in one dimension, there is a case for doing this: there is a hierarchy of weak-dependence conditions in the Gaussian case, and another in the general case (mixing conditions; there is some economy in working with one rather than two hierarchies; for background on such weak, strong and intermediate conditions, see e.g. [10]). This is why some authors restrict to the Gaussian case throughout, as is done in the classical book by Dym and McKean [32] (Gelfand and Vilenkin [36] discuss this (III.3.4), but defer to probabilistic usage by working generally).

5.4. Gaussianity and covariances

Because of this, the whole question of Gaussian measures in infinite-dimensional settings becomes unavoidable here. This rests on Schwartz's theory of Radon measures and cylindrical measures [89], and in particular on the *covariance* operator (recall that a zero-mean Gaussian process is characterized by its covariance), usually written Q. The theory of covariance operators in infinitely many dimensions is considered in detail in [95, III]. They treat the Aronszajn-Kolmogorov theorem, and they show that in Hilbert space, a measure μ being strong second-order (square-integrable in the strong sense) is equivalent to the covariance operator Q being *nuclear* (*trace class*). As a nuclear operator is a product of two Hilbert-Schmidt operators and covariances, being symmetric and positive, factorize (below), this can be expressed alternatively in the language of Hilbert-Schmidt operators. The covariance Q factorizes as

$$Q = i_Q i_Q^*,$$

with i_Q the inclusion map from the range of Q to the Banach space X. So (Radon) Gaussian measures have nice covariance: Q is nuclear, and the above leads to the reproducing- kernel Hilbert space associated with Q [85, IV].

Returning to the general (non-Gaussian) case, linear prediction fails in general in infinite-dimensional Banach spaces: one can find a (non-Gaussian) random element such that linear prediction of it by Gaussian families fails [23, §7]. One can give spectral representations of stationary processes as before, in the setting of $L(Y, H)$ [23, §8].

5.5. Locally convex topological vector spaces

Spaces of generalized functions are typically not normable, and neither are spaces of holomorphic functions, etc. Instead, one needs to specify the topology in terms of a collection of seminorms; with separability (enough for us here) sequences suffice. Often one can use a sequence of norms and obtains a locally convex topological vector space; the prototype here is Grothendieck's nuclear spaces of 1955, for which see [36, I.3]

Some aspects of prediction theory can be taken over to the locally convex case; see e.g. Alpay et al. [3].

5.6. Random generalized functions

If one passes from functions to generalized functions (or Schwartz distributions, though in probability theory this would overload the word distribution), random variables are replaced by random generalized functions. This was done as early as 1954 by Itô [50]. It is argued by Gelfand and Vilenkin [36, III.1.2] that this setting is actually more natural and realistic ('However, every actual measurement is accomplished by means of an apparatus which has a certain inertia'. So the reading given is not an instantaneous value, but rather an averaged value by a test function ϕ characterizing the apparatus.)

6. Applications

We confine ourselves here to brief comments on the applications of prediction theory in the infinite-dimensional case. This forms part of the extensive area of statistical inference for stochastic processes; see e.g. Bosq [15], Antoniadis and Sapatinas [4] (who consider El Niño), [5]. One principal field of application is *functional data analysis*, for which see e.g. [81], [82], [34]. Statistical analysis of data consisting of curves is well established; one application here is *gait analysis*, used to analyse the motion of walking. Another is *yield curves* in the term structure of interest rates in finance; as with anything in finance, predictive ability is of prime practical importance. The first thing to do is to represent curves economically. Wavelet expansions are often used here; an early and dramatic application was their use to digitize the FBI fingerprint bank . Also used here are Karhunen-Loève expansions – again, in the space variable. Much of functional data analysis is concerned with independent and identically distributed (iid) data of curves. For prediction purposes, one needs to dynamicize these results, and have the expansion coefficients evolving with time. Again, wavelets are well suited to this purpose, as they allow one to handle time and space (or time and frequency) together. As always, theory and applications have much to offer each other here. We close by citing one specific application area – to prediction of electricity consumption [26].

---

without which the US criminal justice system would long since have collapsed

7. Postscript on Whittaker

We have mentioned J. M. Whittaker in §2 in connection with the sampling theorem. John Macnaghten (Jack) Whittaker (1905-1984) was Professor of Pure Mathematics at Liverpool University from 1933-53 (seconded during WWII, when he served on Montgomery's staff in 8th Army). Last semester, the first author lectured in Liverpool, in the Whittaker Room, which has a plaque commemorating him and the sampling theorem over the door. Whittaker went on to be Vice-Chancellor of Sheffield University from 1953-65. There Joe Gani founded the *Journal of Applied Probability* in 1964, whose 50th anniversary this volume celebrates. Joe always speaks well of Whittaker as a Vice-Chancellor whose word could be relied on – not always the case.

Whittaker was the son of E. T. (Sir Edmund) Whittaker (1873-1956), who also worked on cardinal series (in 1915). They were the only father-and-son pair in the Royal Society.

8. Conclusion

The above is a necessarily brief and very partial survey of an enormous field; there is much more to be said, and we will return to these matters elsewhere. We close by pointing out that this illustrates the wonderful and inexhaustably rich interplay between mathematics, probability and statistics, pure and applied. This is at the heart of the raison d'être of the Applied Probability Trust journals, which this volume celebrates.

It is a great pleasure for both authors to contribute to this volume. It is a particular pleasure for the first author, who has known Joe Gani all his career, and who contributed to the 25th anniversary volume in 1988. He was then a new recruit to the Editorial Board, and is now retired from it after many years as a Coordinating Editor. We congratulate the Applied Probability Trust, its journals, its editors and its staff on their first half-century, and look forward to the next one.

References

[1] F. Albiac and N. J. Kalton, *Topics in Banach space theory*. Grad. Texts in Math. **233**, Springer, 2006.

[2] D. Aldous, *Probability approximations via the Poisson clumping heuristic*. Springer, 1989.

[3] D. Alpay, O. Timoshenko and D. Volok, Carathéodory-Fejér interpolation in locally convex topological vector spaces. *Linear Alg. Appl.* **431** (2009), 1257-1266.

[4] A. Antoniadis and T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multiv. Anal.* **87** (2003), 133-158.

[5] A. Antoniadis, E. Paparoditis and T. Sapatinas, A functional wavelet-kernel approach for time-series prediction. *J. Roy. Statist. Soc. B* **68** (2006), 837-857.

[6] D. Applebaum and M. Riedle, Cylindrical Lévy processes in Banahc spaces. *Proc. London Math. Soc.* **101** (2010), 697-726.

[7] N. Aronszajn, Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** (1950), 337-404.

[8] Yu. K. Belyaev, Analytic random processes. *Th. Probab. Appl.* **4** (1959), 437-444.

[9] A. R. Bergstrom, *Continuous-time econometric modelling*. Oxford University Press, 1990.

[10] N. H. Bingham, Szegő's theorem and its probabilistic descendants. *Probability Surveys* **9** (2012), 287-324.

[11] N. H. Bingham, Multivariate prediction and matrix Szegő theory. *Probability Surveys* **9** (2012), 325-339.

[12] N. H. Bingham, Modelling and prediction of financial time series. *Communications in Statistics: Theory and Methods* **43** (2014), 1-10.

[13] N. H. Bingham, J. M. Fry and R. Kiesel, Multivariate elliptic processes.*Statistica Neerlandica* **64** (2010), 352-366.

[14] R. P. Boas, Summation formulas and band-limited signals. *Tôhoku Math. J.* **24** (1972), 121-125.

[15] D. Bosq, *Linear processes in function spaces. Theory and applications*. Lecture Notes in Statistics **149**, Springer, 2000.

[16] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, *Time-series analysis.Forecasting and control* (4th ed.). Wiley, 2008.

[17] L. de Branges, *Hilbert spaces of entire functions: A modern theory of eigenfunction expansions based on the invariant subspace concept*. Prentice-Hall, 1968.

[18] P. J. Brockwell, Continuous-time ARMA models. *Handbook of Statistics* **19** (2001), 249-276 (*Stochastic processes: Theory and Methods*, ed. D. N. Shanbhag and C. R. Rao, North-Holland).

[19] P. J. Brockwell, Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* **53** (2001), 113-124.

[20] P. J. Brockwell and R. A. Davis, *Time series: Theory and methods*(2nd ed.), Springer, New York, 1991 (1st ed. 1987).

[21] P. J. Brockwell, C. Erdenbaatar and A. Lindner, Continuous-time GARCH processes. *Ann. Appl. Probab.* **16** (2006), 790-826.

[22] P. L. Butzer, P. J. S. Ferreira, J. R. Higgins, G. Schmeisser and R. L. Stens, The sampling theorem, Poisson's summation formula, general Parseval formula, reproducing kernel formula and the Paley-Wiener theorem for band-limited signals – their interconnections. *Applicable Analysis* **90** (2011), 431-461.

[23] S. A. Chobanyan and N. N. Vakhania, The linear prediction problem and approximation of weak second-order random elements. Pages 37-60 in [70].

[24] H. Cramér, On harmonic analysis in certain function spaces. *Ark. Mat. Astr. Fys.* **28B** (1942), 1-7, reprinted in *Collected Works of Harald Cramér* Volume II, 941-947, Springer, 1994.

[25] H. Cramér and R. Leadbetter, *Stationary and related stochastic processes*. Wiley, 1967.

[26] J. Cugliari, *Prédiction non-paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité*. PhD Thesis, U. Paris-Sud XI, 2011.

[27] S. A. Denisov, Continuous analogs of polynomials orthogonal on the unit circle: Krein systems. *Int. Math. Res. Surv.* 2006 ID 54517, 148p.

[28] J. L. Doob, *Stochastic processes*. Wiley, 1953.

[29] R. M. Dudley, M. G. Hahn and J. Kuelbs, *Probability in Banach spaces 8*. Birkhäuser, 1992.

[30] H. Dym and H. P. McKean, Application of de Branges spaces of integral functions to prediction of stationary Gaussian processes. *Illinois J. Math.* **14** (1970), 299-343.

[31] H. Dym and H. P. McKean, Extrapolation and interpolation of stationary Gaussian processes. *Ann. Math. Statist.* **41** (1970), 1817-1844.

[32] H. Dym and H. P. McKean, *Gaussian processes, function theory and the inverse spectral problem*. Academic Press, 1976.

[33] K.-T. Fang, S. Kotz and K.-W. Ng, *Symmetric multivariate and relarted distributions*. Chapman and Hall, 1990.

[34] F. Ferraty and Y. Romain (ed.), *The Oxford handbook of functional data analysis*. Oxford University Press, 2011.

[35] B. Finkenstädt & H. Rootzén (ed.), *Extreme values in finance, telecommunications and the environment*. Chapman & Hall, 2004.

[36] I. M. Gelfand and N. Ya. Vilenkin, *Generalized functions. volume 4: Applications of harmonic analysis*. academic Press, 1964.

[37] M. S. Ginovyan and L. V. Mikaelyan, Prediction error for continuous-time stationary processes with singular spectral densities. *Acta Appl. Math.* DOI 10.1007/s10440-008-9414-0.

[38] J. Goldstein, N. Gretssky and J. Uhl, *Stochastic processes and functional analysis: in celebration of M. M. Rao's 65th birthday*. Marcel Dekker, 1997.

[39] J. Gorniak and A. Weron, Aronszajn-Kolmogorov type theorems for positive definite kernels in locally convex spaces. *Studia Math.* **69** (1980), 45-56.

[40] C. Gouriéroux, *ARCH models and financial applications*. Springer, 1997.

[41] U. Grenander and G. Szegő, *Toeplitz forms and their applications*. University of California Press, Berkeley CA, 1958 (2nd ed. Chelsea, New York, 1984).

[42] G. Hajduk-Chmielewska, The Wold-Cramér concordance for Banach-space-valued stationary processes. *Studia Math.* **41** (1988), 31-43.

[43] S. Haug, C. Klüppelberg, A. Lindner and M. Zapp, Method of moment estimation in the COGARCH(1,1) model. *Econometric J.* **10** (2007), 320-341.

[44] J. R. Higgins, *Completeness and basis properties of sets of special functions*. Cambridge Tracts in Math. **72**, Cambridge University Press, 1977.

[45] J. R. Higgins, Five short stories about the cardinal series. *Bull. Amer. Math. Soc.* **12** (1985), 45-89.

[46] J. R. Higgins, *Sampling theory in Fourier analysis and signal analysis: Foundations*. Oxford University Press, 1996.

[47] J. R. Higgins and R. L. Stens (ed.), *Sampling theory in Fourier analysis and signal analysis: Advanced topics*. Oxford University Press, 1996.

[48] A. Inoue, AR and MA representations of partial autocorrelation functions, with applications. *Prob. Th. Rel. Fields* 140 (2008), 523-551.

[49] A. Inoue and Y. Kasahara, Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.* **34** (2006), 973-993.

[50] K. Itô, Stationary random distributions. *Mem. Coll. Sci. Univ. Kyoto Ser. A. Math.* **28** (1954), 209-223 (reprinted in *Kiyosi Itô: Selected Papers* 201-215, Springer, 1987).

[51] S. Janson, *Gaussian Hilbert spaces*. Cambridge Tracts in Math. **129**, Cambridge University Press, 1997.

[52] Y. Kabanov, R. Liptser and J. Stoyanov (ed.), *From stochastic calculus to mathematical finance: The Shiryaev Festschrift*. Springer, 2006.

[53] Y. Kakihara, *Multidimensional second-order stochastic processes*. World Scientific, 1997.

[54] Y. Kakihara, Dilations of Hilbert-Schmidt operator-valued measures and applications. Pages 123-135 in [38].

[55] O. Kallenberg, *Foundations of modern probability*, 2nd ed., Springer, 2002.

[56] K. Karhunen, Über die Struktur stationaren zufälligen Funktionen. *Ark. Mat.* **1** (1950), 141-160.

[57] Y. Kasahara and N. H. Bingham, Verblunsky coefficients and Nehari sequences. *Trans. Amer. Mat. Soc.* **366** (2014), 1363-1378.

[58] Y. Katznelson, *An introduction to harmonic analysis*, 3rd ed. Cambridge University Press, 2004 (1st ed., Wiley, 1968, 2nd ed., Dover, 1976).

[59] C. Klüppelberg, A. Lindner and R. A. Maller, A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. *J. Appl. Probab.* **41** (2004), 601-622.

[60] C. Klüppelberg, A. Lindner and R. A. Maller, Continuous-time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. Pages 393-419 in [52].

[61] A. N. Kolmogorov, Stationary sequences in Hilbert space. *Bull. Moskov. Gos. Univ. Mat.* **2** (1941), 1-40 (in Russian; reprinted, [62], 215-255).

[62] A. N. Kolmogorov, *Selected works of A. N. Kolmogorov, Vol. 2: Theory of probability and mathematical statistics*, Nauka, Moskva, 1986 (in Russian).

[63] H. P. Kramer, A generalised sampling theorem. *J. Math. Phys.* **38** (1959), 68-72.

[64] N. Levinson and H. P. McKean, Weighted trigonometrical approximation on R^1 with application to the germ field of a stationary Gaussian noise. *Acta Math.* **112** (1964), 99-143.

[65] P. Lévy, *Processus stochastiques et mouvement brownien*. Gauthier-Villars, Paris, 1948 (2nd ed. 1965).

[66] S. P. Lloyd, A sampling theorem for stationary (wide sense) stochastic processes. *Trans. Amer. Math. Soc.* **92** (1959), 1-12.

[67] M. Loève, Fonctions aléatoires du second ordre. Note, p.228-352, in [65].

[68] M. Loève, *Probability theory*, 4th ed., Vol I, II. Grad. Texts in Math. 45, 46, Springer, 1977, 1978 (1st ed. 1955, 2nd ed. 1960, 3rd ed. 1963).

[69] V. Mandrekar and H. Salehi, The square-integrability of operator-valued functions with respect to a non-negative operator-valued measure and Kolmogorov's isomorphism theorem. *Indiana Univ. Math. J.* **20** (1972), 545-563.

[70] V. Mandrekar and H. Salehi (ed.), Prediction theory and harmonic analysis. The Pesi Masani Volume. North-Holland, 1983.

[71] R. T. W. Martin, Symmetric operators and reproducing kernel Hilbert spaces. *Complex Analysis and Operator Theory* **4** (2010), 845-880.

[72] P. Masani, Orthogonally scattered measures and their applications. *Adv. Math.* **2** (1968), 61-117.

[73] Th. Mikosch: Modelling dependence and tails of financial time series. Pages 185-286 in [35].

[74] D. B. Nelson, Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6** (1990), 318-334.

[75] H. Niemi, Stochastic processes as Fourier transforms of stochastic measures. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **591** (1975), 1-47.

[76] N. K. Nikolskii, *Treatise on the shift operator: Spectral function theory*. Grundl. math. Wiss. **273**, Springer, 1986.

[77] N. K. Nikolskii, *Operators, functions and systems: an easy reading. Volume 1: Hardy, Hankel and Toeplitz; Volume 2: Model operators and systems*. Math. Surveys and Monographs **92, 93**, Amer. Math. Soc., 2002.

[78] R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex domain*. AMS Colloq. Publ. XIX, AMS, 1934.

[79] J. R. Partington, *Interpolation, identification and sampling*. Oxford University Press, 1997.

[80] R. Payen, Fonctions aléatoires du second ordre à valeurs dans un espace de Hilbert. *Ann. Inst. H. Poincaré Ser. B (NS)* **3** (1967), 323-396.

[81] J. O. Ramsay and B. W. Silverman, *Functional data analysis*. Springer, 1997.

[82] J. O. Ramsay and B. W. Silvberman, Applied functional data analysis, Springer, 2002.

[83] M. M. Rao, *Harmonizable, Cramér and Karhunen classes of processes*. Handbook of Statistics 5 (ed. E. J. Hannan et al.), North-Holland, 1985, Ch. 10, p.279-310.

[84] P. H. Richard, Harmonizability, V-boundedness and stationary dilation for Banach-valued processes. Pages 189-205 in [29].

[85] M. Riedle, Cylindrical Wiener processes. *Sém. Probab.* XLIII, 191-214. Lecture Notes in Math. **2006** (2011), Springer.

[86] M. Riedle, Infinitely divisible cylindrical measures on Banach spaces. *Studia Math.* **207** (2011), 235-256.

[87] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer, 2004.

[88] F. Schmidt, Banach space valued stationary processes with absolutely continuous spectral function. *Probability theory on vector spaces. Lecture Notes in Math.* **656**, Springer, 1978, 237-244.

[89] L. Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*. Oxford Universithy Press, 1973.

[90] A. N. Shiryaev, Kolmogorov – Life and creative activities. *Ann. Prob.* **17** (1989), 866-944.

[91] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory*. AMS Colloquium Publications 54.1, American Math. Soc., Providence RI, 2005.

[92] B. Simon, *Orthogonal polynomials on the unit circle. Part 2: Spectral theory*. AMS Colloquium Publications 54.2, American Math. Soc., Providence RI, 2005.

[93] B. Simon, *Szegő's theorem and its descendants: Spectral theory for L^2 perturbations of orthogonal polynomials*. Princeton University Press, 2011.

[94] G. Szegő, *Orthogonal polynomials*. AMS Colloquium Publications 23, American Math. Soc., Providence RI, 1939.

[95] N. N. Vakhania, V. I. Tarieladze and S. A. Chobanyan, *Probability distributions on Banach spaces*. Reidel, Dordrecht, 1987.

[96] J. D. Weston, The cardinal series in Hilbert space. *Proc. Cambridge Phil. Soc.* **45** (1949), 335-341.

[97] J. M. Whittaker, *Interpolatory function theory*, Cambridge Tracts in Math. **33**, Cambridge University Press, 1935.

[98] N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series. With engineering applications*. MIT Press, 1949.

[99] N. Wiener, *Collected Works, Volume III: The Hopf-Wiener integral equation; prediction and filtering; quantum mechanics and relativity; miscellaneous mathematical papers* (ed. P. Masani), MIT Press, Cambridge MA, 1981.