A SAMPLING THEOREM FOR STATIONARY (WIDE SENSE) STOCHASTIC PROCESSES

BY

S. P. LLOYD

1. **Introduction.** The interpolation formula[1]

$$f(t) = \sum_n f(nh) \frac{\sin (\pi/h)(t - nh)}{(\pi/h)(t - nh)}, \qquad -\infty < t < \infty,$$

is due to Cauchy [1], and was rediscovered by E. T. Whittaker [2]. The formula is exact if (real or complex) f has a suitably restricted frequency spectrum; namely, if f can be represented as

$$f(t) = \int_{-h^{-1}/2-0}^{h^{-1}/2+0} e^{2\pi i\lambda t} dg(\lambda), \qquad -\infty < t < \infty,$$

where g is of bounded variation on $[-h^{-1}/2, h^{-1}/2]$ and the jumps of g at the endpoints, if any, are equal [3]. This result, or some variant of it, is known in the communications art as the sampling theorem [12]; it is widely used in information theory [4].

In the present paper we seek conditions under which the random variables $x(t)$ of a stationary (wide sense) stochastic process $\{x(t), -\infty < t < \infty\}$ are determined linearly by the "sample" random variables $\{x(nh), -\infty < n < \infty\}$. (The sample spacing $h > 0$ is to be regarded throughout as a fixed preassigned number.) The problem is a simple one, intuitively speaking. Suppose process x contains a component $a_1 e^{2\pi i\lambda_1 t}$ at frequency λ_1; this component contributes to the samples the quantities

(†) $$a_1 e^{2\pi i\lambda_1 nh}, \qquad -\infty < n < \infty.$$

To reproduce x from its samples it must be possible to extract from the samples at least the information that at frequency λ_1 the amplitude is a_1. However, suppose the process contains another component $a_2 e^{2\pi i\lambda_2 t}$, where the second frequency is related to the first by $\lambda_2 = \lambda_1 + rh^{-1}$, with $r \neq 0$ a positive or negative integer. This component contributes to the samples the quantities

$$a_2 e^{2\pi i\lambda_2 nh} = a_2 e^{2\pi i\lambda_1 nh} e^{2\pi i rn} = a_2 e^{2\pi i\lambda_1 nh}, \qquad -\infty < n < \infty,$$

and these are proportional to (†). The best that could be obtained from the samples would be the value of $a_1 + a_2$; there would be no way of disentangling the separate amplitudes a_1 and a_2. Thus if process x is to be determined by

1

its samples it must be the case that if x has nonvanishing power at a frequency λ then x has vanishing power at each of the "alias" frequencies $\lambda + rh^{-1}$, $r = \pm 1$, ± 2, $\cdots$.

Theorem 1, below, is essentially a precise version of this heuristic result. It may be put as follows: the process x is determined linearly by its samples if and only if some set of frequencies Λ containing all of the power of the process is disjoint from each of its translates $\Lambda - rh^{-1}$, $r = \pm 1$, ± 2, $\cdots$ (²). (That is, no two frequencies in Λ differ by a multiple of h^{-1}.)

In the remainder of the paper we are concerned with the more delicate question of when such linear dependence (where it exists) can be given the form $x(t) = \sum_n x(nh) K_n(t)$ for appropriately chosen coefficients $K_n(t)$, and with convergence properties of such series.

2. **Preliminaries.** We follow the notation and terminology of Doob [5, Chapters IV, IX–XII], for the most part(³). We are concerned with a continuous parameter real or complex stochastic process $\{x(t), -\infty < t < \infty\}$, stationary in the wide sense and continuous in mean square. Such a process has a spectral representation

$$(1) \qquad\qquad x(t) = \int_{-\infty}^{\infty} e^{2\pi i \lambda t} dy(\lambda), \qquad\qquad -\infty < t < \infty,$$

where the spectral process y has orthogonal increments. The (continuous) "covariance" function R of process x has the representation(⁴)

$$(2) \qquad R(s - t) = E\{x(s)x(t)^*\} = \int_{-\infty}^{\infty} e^{2\pi i \lambda (s-t)} dF(\lambda), \qquad -\infty < s, t < \infty$$

where F is the spectral distribution function of process x; $dF(\lambda) = E\{|dy(\lambda)|^2\}$, symbolically [5, p. 527]. The (nonnegative, totally finite) Lebesgue-Stieltjes measure induced by F on λ sets we denote by $F\{\ \}$ or simply F(⁵). Any λ set Λ whose complement is of F measure zero, $F\{\Lambda'\} = 0$, is a *support* of F.

Let $\mathfrak{M}$ denote(⁶) the complete normed linear space of random variables spanned by the collection $\{x(t), -\infty < t < \infty\}$, where the norm of $u \in \mathfrak{M}$ is $\|u\| = [E\{|u|^2\}]^{1/2}$. Two random variables are equivalent if their differ-

---

(²) $\Lambda - rh^{-1}$ denotes the set of all λ such that $(\lambda + rh^{-1}) \in \Lambda$.

(³) In the background of the discussion there is a measure space $[\Omega, \mathfrak{F}, P]$, where Ω is a space of points ω, $\mathfrak{F}$ is a Borel field of ω sets, and P is a measure on $\mathfrak{F}$ such that $P\{\Omega\} = 1$. A random variable $u = u(\omega)$ is a measurable ω function; a stochastic process is an indexed family of random variables. We suppress, throughout, explicit mention of $[\Omega, \mathfrak{F}, P]$; $E\{u\}$ denotes the integral $\int_\Omega u(\omega) P\{d\omega\}$, and "with probability 1" (or "w.p.1") replaces "almost everywhere $[P]$."

(⁴) Asterisks denote complex conjugates.

(⁵) The "power" in the Introduction is just F measure.

(⁶) That is, $\mathfrak{M}$ (separable) is the smallest linear manifold containing (for all choices of ν, $\{t_\alpha\}$, complex $\{c_\alpha\}$) all finite linear combinations $\sum_{\alpha=1}^{\nu} c_\alpha x(t_\alpha)$, all limits in norm of sequences of such linear combinations, and all random variables equal to such limits with probability 1.

ence has norm zero; i.e., if they are equal with probability 1. The space $\mathfrak{M}$ is isomorphic (to within equivalences) with the complete normed linear space $\mathfrak{N}$ of (complex) F measurable λ functions v for which the norm $\|v\| = [\int_{-\infty}^{\infty} |v(\lambda)|^2 dF(\lambda)]^{1/2}$ is finite. Two λ functions are equivalent if their difference has norm zero; i.e., if they are equal almost everywhere $[F]$[7]. The (norm preserving) isomorphism under consideration is the usual one, given by

$$(3) \qquad u = \int_{-\infty}^{\infty} v(\lambda) dy(\lambda),$$

where $u \in \mathfrak{M}$ is the image of $v \in \mathfrak{N}$ to within equivalences [5, p. 427] ($\|u\| = \|v\|$ holds for every $u \in \mathfrak{M}$, $v \in \mathfrak{N}$ related by (3)).

The samples $\{x(nh), -\infty < n < \infty\}$ span a certain closed linear subspace of $\mathfrak{M}$; we denote this subspace by $\mathfrak{M}_s$. The random variables in $\mathfrak{M}_s$ are those "determined linearly by the samples," as we shall use this phrase. We denote by $\mathfrak{N}_s$ the image in $\mathfrak{N}$ of $\mathfrak{M}_s$. According to (1) and (3), to $x(nh) \in \mathfrak{M}_s$ there corresponds $e^{2\pi i\lambda nh} \in \mathfrak{N}_s$, $-\infty < n < \infty$. Any linear combination of these latter has period h^{-1} in λ, and standard trigonometric approximation methods show that $\mathfrak{N}_s$ contains all functions in $\mathfrak{N}$ which have period h^{-1} [5, p. 560]. Thus we may characterize $\mathfrak{N}_s$ as those functions in $\mathfrak{N}$ which are periodic with period h^{-1}, together with all functions equivalent to such periodic functions.

Our main result, Theorem 1 of the next section, is based on a certain formula for the projection on $\mathfrak{N}_s$.

LEMMA. *The operator $\mathcal{P}$ given by*

$$(4) \qquad (\mathcal{P}v)(\lambda) = \frac{\sum_n v(\lambda + nh^{-1}) dF(\lambda + nh^{-1})}{\sum_n dF(\lambda + nh^{-1})} \qquad \text{a.e. } [F], v \in \mathfrak{N},$$

is the projection on $\mathfrak{N}_s$.

Proof. We should first define the right side of (4). Given $v \in \mathfrak{N}$, the (σ-additive, totally σ-finite) set functions G_0, G_1, G_2 are determined by their values

$$G_0\{A\} = \sum_n \int_A dF(\lambda + nh^{-1}),$$

$$(5) \qquad G_1\{A\} = \sum_n \int_A v(\lambda + nh^{-1}) dF(\lambda + nh^{-1}),$$

$$G_2\{A\} = \sum_n \int_A |v(\lambda + nh^{-1})|^2 dF(\lambda + nh^{-1})$$

---

[7] We choose to regard $\mathfrak{N}$ as a set of functions, and not as a set of equivalence classes of functions, so we should call $\| \; \|$ a pseudo- (or semi-) norm. (A similar remark applies to $\mathfrak{M}$.) The fact that (depending on F) two different λ functions may be equivalent in $\mathfrak{N}$ lies at the core of our problem, and is not here just an unimportant technical complication.

for bounded Borel sets A. (Let A have diameter less than h^{-1}, so that its translates $\{A_n = A - nh^{-1}, -\infty < n < \infty\}$ are mutually disjoint. Then $G_0\{A\} = F\{\cup_n A_n\}$, and the σ-additivity and σ-finiteness of G_0 follow easily; similarly for G_1 and G_2, using $v \in \mathfrak{R}$ and the totally finite property of F.) The same methods show that G_1 and G_2 are each G_0-continuous[8], and by the right side of (4) we mean the Radon-Nikodym derivative dG_1/dG_0. This derivative is determined almost everywhere $[G_0]$ and is G_0 measurable[9]; a fortiori, it is determined almost everywhere $[F]$ and is F measurable, it being obvious that F is G_0-continuous. Each G has the property $G\{B - nh^{-1}\} = G\{B\}$ for all Borel sets B and integers n, clearly, so that dG_1/dG_0 may be chosen to be periodic with period h^{-1}.

The operator $\mathcal{P}$ is linear; we show that it is bounded. Let Borel set A have diameter less than h^{-1}, so that its translates $A_n = A - nh^{-1}, -\infty < n < \infty$, are mutually disjoint. We have

$$G_1\{A\} = \int_{\cup_n A_n} v(\lambda) dF(\lambda),$$

and Schwarz's inequality gives

$$|G_1\{A\}|^2 \leq \int_{\cup_n A_n} |v(\lambda)|^2 dF(\lambda) \int_{\cup_n A_n} dF(\lambda) = G_2\{A\}G_0\{A\},$$

holding for every such A. It follows that $|dG_1/dG_0|^2 \leq dG_2/dG_0$ a.e. $[G_0]$. Hence

$$\|\mathcal{P}v\|^2 = \int_{-\infty}^{\infty} \left|\frac{dG_1}{dG_0}(\lambda)\right|^2 dF(\lambda) \leq \int_{-\infty}^{\infty} \frac{dG_2}{dG_0}(\lambda) dF(\lambda)$$

$$= \sum_n \int_0^{h^{-1}} \frac{dG_2}{dG_0}(\lambda + nh^{-1}) dF(\lambda + nh^{-1})$$

$$= \sum_n \int_0^{h^{-1}} \frac{dG_2}{dG_0}(\lambda) dF(\lambda + nh^{-1})$$

$$= \int_0^{h^{-1}} \frac{dG_2}{dG_0}(\lambda) G_0\{d\lambda\}$$

$$= \int_0^{h^{-1}} G_2\{d\lambda\} = \|v\|^2,$$

where we have used the fact that dG_2/dG_0 may be chosen to be periodic with period h^{-1}.

---

[8] This term is an abbreviation for "absolutely continuous with respect to G_0."

[9] I.e., measurable with respect to the Borel field generated by the Borel sets and the subsets of Borel sets of G_0 measure zero. A set Λ (function $v(\lambda)$) is G_0 measurable if and only if every translate $\Lambda - nh^{-1}(v(\lambda + nh^{-1}))$ is F measurable.

We show next that $\mathcal{P}$ is (equivalent to) the identity in $\mathfrak{N}_s$. Any $v \in \mathfrak{N}_s$ is equivalent to some $\hat{v} \in \mathfrak{N}_s$ which is periodic with period h^{-1}. From (5) we have

$$G_1\{A\} = \sum_n \int_A \hat{v}(\lambda) dF(\lambda + nh^{-1}) = \int_A \hat{v}(\lambda) G_0\{d\lambda\}$$

for all bounded Borel sets A, whence $dG_1/dG_0 = \hat{v}$ a.e. $[G_0]$[9]. Thus $v \in \mathfrak{N}_s$ implies $\mathcal{P}v = v$ a.e. $[F]$. Since (range $\mathcal{P}) \subset \mathfrak{N}_s$, it follows that $\mathcal{P}$ is the projection on $\mathfrak{N}_s$.

The formula we use in the next section is a modification of (4). For each n, let F_n denote the Lebesgue-Stieltjes measure induced by the distribution function $F_n(\lambda) = F(\lambda + nh^{-1})$, $-\infty < \lambda < \infty$, $-\infty < n < \infty$. Each of these measures may be decomposed (Lebesgue) into an F-continuous part and an F-singular part:

(6) $$F_n\{B\} = \int_B f_n(\lambda) dF(\lambda) + F_n\{B \cap \Gamma_n\}, \text{ Borel } B,$$

where $f_n \geq 0$ is the density relative to F of the F-continuous part of F_n and where the F-singular part of F_n is contained in the set Γ_n, $F\{\Gamma_n\} = 0$. (We choose $f_0 \equiv 1$, $\Gamma_0 = \varnothing$.) If Γ denotes the set $\Gamma = \cup_n \Gamma_n$, then $F\{\Gamma\} = 0$ and

$$F_n\{B\} = \int_B f_n(\lambda) dF(\lambda) + F_n\{B \cap \Gamma\}, \text{ Borel B}, \quad -\infty < n < \infty.$$

The quantities G_0 and G_1 of (5) become

$$G_0\{A\} = \int_A \sum_n f_n(\lambda) dF(\lambda) + G_0\{A \cap \Gamma\},$$

$$G_1\{A\} = \int_A \sum_n f_n(\lambda) v(\lambda + nh^{-1}) dF(\lambda) + G_1\{A \cap \Gamma\},$$

whence the projection $\mathcal{P}v = dG_1/dG_0$ has on Γ' (which is a.e. $[F]$) the value

(7) $$(\mathcal{P}v)(\lambda) = \frac{v(\lambda) + \sum_{n \neq 0} f_n(\lambda) v(\lambda + nh^{-1})}{1 + \sum_{n \neq 0} f_n(\lambda)} \quad \text{a.e. } [F].$$

(The numerator and denominator are each finite a.e. $[F]$; cf. [6, p. 879].) We remark that this version of the projection, in contrast to (4), is no longer formally periodic.

This result, (4) or (7), is (of course) an ergodic theorem. We have presented it in detail because it falls under a case which is usually set aside[10];

[10] However, see [6].

the space at hand (the λ axis) is dissipative with respect to our transformation $\lambda \to \lambda - h^{-1}$ [11].

3. **A sampling theorem.** Now we ask, when is it possible that for some value (or values) of t not of the form nh, the random variable $x(t)$ is determined linearly by the sample random variables $\{x(nh),\ -\infty < n < \infty\}$? A fairly complete answer is provided by

THEOREM 1 (SAMPLING THEOREM). *The following properties of process x, stationary in the wide sense and continuous in mean square, are equivalent:*

(i) *Each random variable $x(t)$ $(-\infty < t < \infty)$ of the process is determined linearly by the samples $\{x(nh),\ -\infty < n < \infty\}$;*

(ii) *For some irrational number ξ, $x(\xi h)$ is determined linearly by the samples;*

(iii) *There exists a support Λ of the spectral distribution of the process whose translates $\{\Lambda - nh^{-1},\ -\infty < n < \infty\}$ are mutually disjoint* [12].

Proof. (i)$\Rightarrow$(ii) is vacuous. We show (ii)$\Rightarrow$(iii) and then (iii)$\Rightarrow$(i). Suppose $x(\xi h)$ is determined linearly by the samples, i.e., $x(\xi h)$ is equal to its projection on $\mathfrak{M}_s$ w.p.1. Then $e^{2\pi i\lambda \xi h}$, which corresponds in $\mathfrak{N}$ to $x(\xi h)$, is equal to its projection on $\mathfrak{N}_s$ a.e. $[F]$. Using (7), we have

$$e^{2\pi i\lambda \xi h} = \frac{e^{2\pi i\lambda \xi h} + \sum\limits_{n\neq 0} f_n(\lambda) e^{2\pi i(\lambda + nh^{-1})\xi h}}{1 + \sum\limits_{n\neq 0} f_n(\lambda)} \qquad \text{a.e. } [F];$$

this simplifies to

$$(8) \qquad \sum_{n\neq 0}(1 - e^{2\pi in\xi}) f_n(\lambda) = 0 \qquad \text{a.e. } [F].$$

Each coefficient $1 - e^{2\pi in\xi}$ here has nonvanishing positive real part, since ξ is assumed to be irrational. It follows that $f_n(\lambda) = 0$ a.e. $[F]$, $n \neq 0$, since the $\{f_n\}$ are nonnegative. But this is to say, F_n is F-singular, $n \neq 0$, from the definition of the $\{f_n\}$.

Thus there exist complementary supports for F and F_n, $n \neq 0$; let $\mathbf{M}_n$ be a support of F such that $F_n\{\mathbf{M}_n\} = 0$, $n \neq 0$. The intersection $\mathbf{N} = \bigcap_{n\neq 0}\mathbf{M}_n$ of

---

(11) Recall that a subset W of some space is wandering with respect to a 1-1 point transformation T if its transforms $\{T^nW,\ -\infty < n < \infty\}$ are mutually disjoint, and a set Y is dissipative with respect to T if it is the sum $Y = \bigcup_n T^n W$ of transforms of some wandering set W. In the present case, the interval $(0, h^{-1}]$, for instance, is wandering with respect to $T\colon T\lambda = \lambda - h^{-1}$, and its translates $\{(-nh^{-1}, -(n-1)h^{-1}],\ -\infty < n < \infty\}$ cover the λ axis. (In the usual ("regular") case of an ergodic theorem an assumption is made to the effect that all (measurable) dissipative sets have measure zero.)

(12) In the language of egodic theory (11), F has a wandering support (with respect to $T\colon T\lambda = \lambda - h^{-1}$). Or, again, there exists a support Λ of F such that no two frequencies in Λ differ by an integral multiple of h^{-1}.

these is a support of F which has the property $F_n\{N\} = 0$, $n \neq 0$. From the nature of the F_n (translates of F) we see that $N_n = N - nh^{-1}$ is a support of F_n which has the property $F_r\{N_n\} = 0$, $r \neq n$, $-\infty < n$, $r < \infty$; in particular, $F\{N_n\} = 0$, $n \neq 0$. Finally, the set $\Lambda = N \cap (\cap_{n \neq 0} N_n')$ is a support of F which is disjoint from each of its translates $\Lambda - nh^{-1}$, $n \neq 0$.

To prove (iii)$\Rightarrow$(i), suppose Λ is a support of F which is disjoint from each of its translates $\Lambda - nh^{-1}$, $n \neq 0$. Clearly, $\Lambda - nh^{-1}$ is a support of F_n, so that F and F_n have disjoint supports, $n \neq 0$; that is, F and F_n are mutually singular, $n \neq 0$. From (6) we have $f_n(\lambda) = 0$ a.e. $[F]$, $n \neq 0$. This implies that each $v \in \mathfrak{N}$ is equivalent to its projection on $\mathfrak{N}_s$, using (7). Hence $\mathfrak{N} = \mathfrak{N}_s$, $\mathfrak{M} = \mathfrak{M}_s$; a fortiori, $\mathfrak{M}_s$ contains for each t ($-\infty < t < \infty$) a random variable equal to $x(t)$ w.p.1.

Before going on to find such random variables explicitly, we derive a corollary and an analog of Theorem 1. The corollary extends a result well known in the case where F has an interval $(-W, W)$ as support. We define the *bandwidth W* of process x as:

$$W = \frac{1}{2} \underset{\text{Borel } B}{\text{g.l.b.}} \{m\{B\} \mid F\{B'\} = 0\}$$

where m denotes Lebesgue measure, and where the g.l.b. is that of the Lebesgue measures of all Borel supports of F, as indicated. Then Theorem 1 has the

COROLLARY. *If process x is determined linearly by its samples then the sample spacing and bandwidth satisfy* $2Wh \leq 1$.

Proof. Any support of F contains a Borel support of F; let Λ be a Borel support of F whose translates $\{\Lambda_n = \Lambda - nh^{-1}, \; -\infty < n < \infty\}$ are mutually disjoint. The (Baire) function k defined by[13]

$$k(\lambda) = \sum_n \chi_\Lambda(\lambda + nh^{-1}), \qquad -\infty < \lambda < \infty,$$

is bounded by unity, $k(\lambda) \leq 1$, $-\infty < \lambda < \infty$, since $\chi_\Lambda(\lambda + nh^{-1}) = \chi_{\Lambda_n}(\lambda)$ and the $\{\Lambda_n\}$ are mutually disjoint. Using this, we have

$$2Wh = h \underset{\text{Borel } B}{\text{g.l.b.}} \{m\{B\} \mid F\{B'\} = 0\}$$

$$\leq hm\{\Lambda\} = h \int_{-\infty}^{\infty} \chi_\Lambda(\lambda) d\lambda$$

$$= h \sum_n \int_0^{h^{-1}} \chi_\Lambda(\lambda + nh^{-1}) d\lambda = h \int_0^{h^{-1}} k(\lambda) d\lambda \leq 1.$$

Theorem 1 tells us nothing if it is known only that $x(t)$ is determined

---

[13] χ_Λ denotes the characteristic function of set Λ.

linearly by the samples for some rational t/h. The next theorem covers this case.

THEOREM 2. *The following properties of process x are equivalent (for each fixed positive integer p):*

(i) *For each integer q $(-\infty < q < \infty)$ the random variable $x(qh/p)$ is determined linearly by the samples;*

(ii) *For some integer q relatively prime to p, $x(qh/p)$ is determined linearly by the samples;*

(iii) *There exists a support Λ of F which is disjoint from each of its translates $\Lambda - nh^{-1}$, where n takes all positive and negative integer values except $0, \pm p, \pm 2p, \cdots$.*

(The proof follows that of Theorem 1; we omit the details. Essentially, both (ii)$\Rightarrow$(iii) and (iii)$\Rightarrow$(i) depend on (8). With $\xi = q/p$ in (8), the coefficient of f_n vanishes when n is a multiple of p, so that the f_n need not vanish for these values of n.)

4. Sampling series. We assume from now on that process x is determined linearly by its samples, and we seek explicit forms for this linear dependence. From Theorem 1, the spectral distribution of the process has some support Λ whose translates $\{\Lambda - nh^{-1}, -\infty < n < \infty\}$ are mutually disjoint. Given such a support, which we may and do assume is a Borel set, consider the λ function k_t defined for each fixed t by

$$(9) \qquad k_t(\lambda) = \sum_n \chi_\Lambda(\lambda + nh^{-1})e^{2\pi i(\lambda + nh^{-1})t}, \qquad -\infty < \lambda < \infty.$$

As in the corollary of Theorem 1, k_t is a bounded Baire function, $|k_t(\lambda)| \leq 1$, $-\infty < \lambda < \infty$, so that $k_t \in \mathfrak{N}$. Moreover, k_t is manifestly periodic in λ with period h^{-1}, so that $k_t \in \mathfrak{N}_s$. Finally, k_t has the property $k_t(\lambda) = e^{2\pi i \lambda t}$ a.e. $[F]$, using $F\{\Lambda'\} = 0$ and (again) the fact that the translates

$$\{\Lambda - nh^{-1}, -\infty < n < \infty\}$$

of Λ are mutually disjoint. It follows that

$$(10) \qquad x(t) = \int_{-\infty}^{\infty} k_t(\lambda)dy(\lambda), \qquad -\infty < t < \infty,$$

is a spectral version of the determination of x by its samples. To complete our efforts we should like to express k_t as a linear combination (or limit in $\mathfrak{N}$ of linear combinations) of the functions $\{e^{2\pi i \lambda nh}\}$, so that the samples $\{x(nh)\}$ themselves would emerge from the right side of (10).

Let us try expanding k_t in its (ordinary) Fourier series. We have formally

$$(11) \qquad k_t(\lambda) \sim \sum_n e^{2\pi i \lambda nh}K(t - nh),$$

where K is given by

(12)
$$K(t) = h \int_\Lambda e^{2\pi i\lambda t}d\lambda, \qquad\qquad -\infty < t < \infty,$$

since the Fourier coefficients of k_t are

$$h \int_0^{h^{-1}} e^{-2\pi i\lambda nh}k_t(\lambda)d\lambda$$

$$= h \int_0^{h^{-1}} e^{-2\pi i\lambda nh} \sum_r \chi_\Lambda(\lambda + rh^{-1})e^{2\pi i(\lambda+rh^{-1})t}d\lambda$$

$$= h \sum_r \int_0^{h^{-1}} \chi_\Lambda(\lambda + rh^{-1})e^{2\pi i(\lambda+rh^{-1})(t-nh)}d\lambda$$

$$= h \int_{-\infty}^{\infty} \chi_\Lambda(\lambda)e^{2\pi i\lambda(t-nh)}d\lambda = K(t - nh), \qquad -\infty < n < \infty.$$

We remark that $|K(t)| \leq 1$, $-\infty < t < \infty$, and that if Λ is the union of mutually disjoint open intervals $\{(\lambda_\alpha', \lambda_\alpha''), \alpha = 1, 2, \cdots\}$ then

(13)
$$K(t) = h \sum_{\alpha=1}^{\infty} \frac{(e^{2\pi i\lambda_\alpha''t} - e^{2\pi i\lambda_\alpha't})}{2\pi it}, \qquad t \neq 0\,(^{14}),$$

$$= hm\{\Lambda\}, \qquad\qquad t = 0.$$

(If Λ is the interval $(-h^{-1}/2, h^{-1}/2)$ then (13) gives the familiar $K(t) = (\pi t/h)^{-1} \sin (\pi t/h)$.)

The Fourier series (11) for k_t may or may not converge in $\mathfrak{N}$ to k_t; if it does we have a *sampling series*(^{15}) for process x:

(14)
$$x(t) = \sum_n x(nh)K(t - nh), \qquad -\infty < t < \infty,$$

substituting (11) in (10) and using (1). However, we cannot expect an orthogonal expansion in one function space to converge in another, in general, and further conditions must be imposed if (14) is to hold. (We return to this question presently.)

We can get around this difficulty by orthogonalizing the

$$\{e^{2\pi i\lambda nh}, -\infty < n < \infty\}$$

in $\mathfrak{N}$. Suppose the Schmidt orthogonalization procedure applied (in $\mathfrak{N}$) to the sequence

(15)
$$1, e^{2\pi i\lambda h}, e^{-2\pi i\lambda h}, e^{2(2\pi i\lambda h)}, \cdots$$

(^{14}) The series is uniformly absolutely convergent on $-\infty < t < \infty$.

(^{15}) We should note at this point that (depending on F) Λ and hence K may be nonunique in a nontrivial way, so that there may be many series of the form (14) for a given process x. (This will certainly be the case if there exists some open wandering support Λ of F such that $m\{\Lambda\} < h^{-1}$; see Theorem 3, below.)

yields the orthonormal sequence θ_0, θ_1, θ_{-1}, θ_2, $\cdots$; the $\{\theta_n\}$ will be of the form

$$\theta_n(\lambda) = \sum_r a_{nr} e^{2\pi i\lambda rh}, \qquad -\infty < \lambda < \infty, \; -\infty < n < \infty,$$

where $a_{nr}=0$ if $|r| > |n|$ ([16]). If k_t (or, what is the same thing, $e^{2\pi i\lambda t}$, since we are now working in $\mathfrak{N}$) is expanded in terms of the $\{\theta_n\}$ there results([17])

$$(16) \qquad x(t) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{n=-N}^{N} \sum_r a_{nr} x(rh) \sum_s \overset{*}{a}_{ns} R(t - sh),$$

where R is the covariance of the process, (2). (In the terminology of Doob [5, Chapter IV], the right side of (16) is the martingale (wide sense) limit

$$\underset{N\to\infty}{\text{l.i.m.}} \, \hat{E}\{x(t) \mid x(0), x(h), x(-h), x(2h), \cdots, x(-Nh)\},$$

and represents the projection of $x(t)$ on $\mathfrak{M}_s$ whether $x(t)$ is determined linearly by the samples or not.)

We are mainly interested in the sampling series (14), however. We have

THEOREM 3. *If the spectral distribution of process x has an open support Λ whose translates $\{\Lambda - nh^{-1}, \; -\infty < n < \infty\}$ are mutually disjoint, then the sampling series is summable $(C, 1)$ in norm to $x(t)$; i.e.,*

$$(17) \qquad x(t) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) x(nh) K(t - nh), \qquad -\infty < t < \infty.$$

Proof. The $(C, 1)$ partial sums of the Fourier series for k_t converge to k_t at every point of Λ, since $k_t(\lambda) = e^{2\pi i\lambda t}$, $\lambda \in \Lambda$, is continuous on open Λ [7, p. 45]. Moreover, these $(C, 1)$ partial sums are bounded, since k_t is bounded [7, p. 46]. Hence, by dominated convergence, the Fourier series for k_t is summable $(C, 1)$ in $\mathfrak{N}$ to k_t, and (17) follows from the isomorphism of $\mathfrak{N}$ and $\mathfrak{M}$.

COROLLARY. *If also Λ is a finite union of intervals, or, more generally, if* l.u.b.$_{-\infty < t < \infty} |tK(t)| < \infty$, *then the sampling series converges in norm to $x(t)$; i.e.,*

$$(18) \qquad x(t) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{n=-N}^{N} x(nh) K(t - nh), \qquad -\infty < t < \infty.$$

Proof. The added hypothesis insures that the ordinary partial sums of the Fourier series for k_t converge to k_t on Λ [7, p. 47], and that the difference

---

([16]) If any member of the sequence (15) is linearly dependent in $\mathfrak{N}$ on preceding members then so will be all subsequent numbers. This occurs when and only when F is increasing at (at most) a finite number of points, under the assumption $\mathfrak{M} = \mathfrak{M}_s$ of the present section. It is not necessary to modify the notation to cover this case.

([17]) We use "l.i.m." to designate a limit in the norm of $\mathfrak{M}$.

between the ordinary and $(C, 1)$ partial sums is bounded [7, p. 43]. By dominated convergence, again, the ordinary partial sums converge in $\mathfrak{N}$ to k_t, and (18) follows.

Our final theorem gives sufficient conditions for convergence of the sampling series with probability 1.

THEOREM 4. *Suppose the spectral distribution of process x has an open support Λ whose translates $\{\Lambda - nh^{-1},\ -\infty < n < \infty\}$ are mutually disjoint. Suppose further that there exist numbers α and β such that*

$$\alpha > 1/2,\ \beta > 0,\ \alpha + \beta/2 > 1,$$

(19)
$$\underset{-\infty < t < \infty}{\text{l.u.b.}}\ \left| t^\alpha K(t) \right| < \infty,$$

$$\underset{-\infty < t < \infty}{\text{l.u.b.}}\ \left| t^\beta R(t) \right| < \infty.$$

Then the sampling series converges to $x(t)$ with probability 1; i.e.,

(20)
$$x(t) = \lim_{N \to \infty} \sum_{n=-N}^{N} x(nh) K(t - nh) \qquad \text{w.p.1}, \qquad -\infty < t < \infty.$$

Proof. Under assumptions (19), convergence of the right side of (20) with probability 1 is a consequence of a generalization of Menchoff's theorem given by Kac, Salem, and Zygmund[18]. Theorem 3, above, shows that the sum is $x(t)$.

An example will indicate the limitations of the sampling series representation (14). We set $h = 1$, to simplify the notation. Let $\xi_0,\ \xi_1,\ \xi_{-1},\ \xi_2,\ \cdots$ be a sequence of numbers which has the property that $\xi_n - \xi_r$ is irrational for all $n \neq r,\ -\infty < n, r < \infty$[19]. Let L_n denote the set of all numbers in the interval $I = (0,\ 1]$ which differ from ξ_n by a rational number, and define $\Lambda_1 = \cup_{n \neq 0} (L_n + n)$, $\Lambda_2 = \cap_{n \neq 0} (I \cap L_n')$, $\Lambda = \Lambda_1 \cup \Lambda_2$. The part Λ_1 of Λ is countable (and dense in I'), while $\Lambda_2 \subset I$ is the complement in I of a countable set. If $\lambda_1,\ \lambda_2,\ \cdots$ is a counting of Λ_1, let F be a Lebesgue-Stieltjes measure which assigns measure $\mu_\alpha > 0$ to λ_α, $\alpha = 1, 2, \cdots$ (with $\sum_1^\infty \mu_\alpha < \infty$), Lebesgue measure to (measurable) subsets of Λ_2, and measure zero to Λ'. Finally, let $\{x(t),\ -\infty < t < \infty\}$ be a stationary (wide sense) stochastic process which has F as its spectral distribution [5, p. 519]. The translates

$$\{\Lambda - n,\ -\infty < n < \infty\}$$

of the support Λ of F are mutually disjoint, clearly, so that process x is determined linearly by its samples. However, no trigonometrical series $\sum_n e^{2\pi i\lambda n} K_n(t)$ which converges in $\mathfrak{N}$ to $e^{2\pi i\lambda t}$ (if such exists) can be a Fourier

---

[18] Lemma I of [8]. We neglect the case $\alpha = 1/2$ which (with an added hypothesis) is covered by Lemma I of [8].

[19] E.g., any sequence of distinct members of a Hamel base.

series. Thus, even if there is a representation $x(t) = \sum_n x(n) K_n(t)$ for this process, it will not be one given by (12) and (14). (Methods for dealing with the atomic part of F are known from smoothing and prediction theory; cf. [5, p. 567].)

5. **Acknowledgments.** Special cases covered by the Corollary of Theorem 3 have appeared previously [9; 10; 11]. The author wishes to thank Professor Kac for bringing reference [8] to his attention.

REFERENCES

1. A.-L. Cauchy, *Mémoire sur diverses formulaes dé analyse*, C.R. Acad. Sci. Paris vol. 12 (1841) pp. 283–298.

2. E. T. Whittaker, *On the functions which are represented by the expansions of the interpolation-theory*, Proc. Roy. Soc. Edinburgh. Section A. vol. 35 (1915) pp. 181–194.

3. H. P. Kramer, *A generalized sampling theorem*, Bull. Amer. Math. Soc. Abstract 63-2-234.

4. C. E. Shannon, *Communication in the presence of noise*, Proc. I.R.E. vol. 37 (1949) pp. 10–21.

5. J. L. Doob, *Stochastic processes*, New York, Wiley, 1953.

6. J. C. Oxtoby, *On the ergodic theorem of Hurewicz*, Ann. of Math. vol. 49 (1948) pp. 872–884.

7. A. Zygmund, *Trigonometrical series*, New York, Dover, 1955 (reprint).

8. M. Kac, R. Salem and A. Zygmund, *A gap theorem*, Trans. Amer. Math. Soc. vol. 63 (1948) p. 235.

9. S. P. Lloyd and B. McMillan, *Linear least squares filtering and prediction of sampled signals*, M.R.I. Symposia Series, vol. 5, Polytechnic Institute of Brooklyn, 1956, pp. 221–247.

10. E. Parzen, Technical Report No. 7, Department of Statistics, Stanford University, December 22, 1956.

11. A. V. Balakrishnan, *A note on the sampling principle for continuous signals*, Institute of Radio Engineers Transactions on Information Theory, vol. IT-3 (1957) pp. 143–146.

12. H. Nyquist, *Certain topics in telegraph transmission theory*, Trans. Amer. Inst. Elec. Engrs. vol. 47 (1928) pp. 617–644.

BELL TELEPHONE LABORATORIES,
    MURRAY HILL, N. J.