INFORMATION AND CONTROL 1, 113–126 (1958)

A Paradox Concerning Rate of Information

I. J. GOOD AND K. CAJ DOOG

Government Communications Headquarters, Cheltenham, England

A natural definition of the rate of transmission of information is given, arising out of the usual theory. We call it the "Riemannian" rate of transmission. It is shown that the definition leads to a paradox if taken in conjunction with the notion of (time-unlimited) band-limited white noise. A mathematical model can hardly contain both these notions at the same time. The Riemannian rate of transmission does however lead to sensible results if used in conjunction with periodic band-limited white noise. In particular it leads to the Hartley-Wiener-Tuller-Sullivan-Shannon formula without the necessity of introducing Shannon's notion of "dimension rate." The discussion refers to matrix signal-to-noise ratios and to the entropy of singular multivariate normal distributions.

I. INTRODUCTION

Let **S** be a population or ensemble of source or sent signals, **N** a statistically independent ensemble of noise, and $\mathbf{R} = \mathbf{S} + \mathbf{N}$ an ensemble of received signals. The addition here may be interpreted as ordinary addition of amplitudes at each instant of time. The expected amount of information concerning **S** provided by **R**, or the entropy concerning **S** provided by **R**, ent $(\mathbf{S}\colon \mathbf{R})$, is formally defined as the expected log-association-factor between **S** and **R**, i.e. the expectation of

$$\log \frac{P(S \,\&\, R)}{P(S)P(R)} \tag{1}$$

where S and R are particular realizations of the source and received signals. (The units are bits, tits, dits, or nits, depending on whether the base of the logarithms is 2, 3, 10, or e.) In practice there can be technical difficulties in giving this formal definition a precise interpretation.

One of the aims of the theory of information is to give finite measures of information. For continuous information it is well known to be necessary to allow for the existence of noise in order to achieve finiteness.

Moreover, if time is allowed to go on forever in the mathematical model, then it is naturally necessary to talk in terms of a rate of transmission,

$$J = \lim_{T\to\infty} \frac{1}{T} \text{ent } (\mathbf{S}: \mathbf{R}), \tag{2}$$

where now $\mathbf{S}$ and $\mathbf{R}$ have to be interpreted as the ensembles restricted to a time-interval of length T.

Nor is this enough to achieve finiteness; the highly abstract notion of dimension rate (Shannon, 1948, Appendix 7) was also introduced in order to cover the most general problems, although this notion does not seem to have been developed in detail. One of our aims is to discuss the problem without explicit mention of dimension rates.

We shall need the familiar formula

$$\text{ent } (\mathbf{S}: \mathbf{R}) = \text{ent } \mathbf{R} - \text{ent } (\mathbf{R} \mid \mathbf{S}), \tag{3}$$

which, when 'addition' is such that N is uniquely determined by R and S and is causally independent of S, gives

$$\text{ent } (\mathbf{S}: \mathbf{R}) = \text{ent } \mathbf{R} - \text{ent } \mathbf{N}. \tag{4}$$

(Cf. Shannon[13], Theorem 16. For a discussion of the notation as used here see Good[8].) The rate of transmission, J, is

$$J = \lim_{T\to\infty} \frac{1}{T} \text{ent } \mathbf{R} - \lim_{T\to\infty} \frac{1}{T} \text{ent } \mathbf{N}, \tag{5}$$

the difference in the entropy rates of the received signal and of the noise. (If these rates are to be calculated separately we may think of the amplitudes as expressed to 10!!!! places of decimals, and use the definition of entropy for discrete sequences; or we may use the definition of entropy adapted to continuous information. In our opinion the second method, although mathematically more convenient, is philosophically less satisfactory.)

Riemannian rate of transmission

Let t be an arbitrary instant of time and let the amplitudes of S, N, R at time t be S_t, N_t, R_t. Suppose that we have made observations of R at times $t_1, t_2, \cdots, t_\nu$, namely $R_{t_1}, R_{t_2}, \cdots, R_{t_\nu}$. The corresponding random variables are $\mathbf{R}_{t_1}, \mathbf{R}_{t_2}, \cdots, \mathbf{R}_{t_\nu}$. If we think of ν as made larger and larger and the instants of time made closer and closer together then we should expect diminishing returns to set in (regarding the amount of information obtained concerning S). In fact if this did not happen we should say that we had run into a paradox. This opinion leads us to make the following definition:

Let

$$\mathfrak{D} = (0 = t_1 < t_2 < \cdots < t_\nu < T)$$

be a dissection of the interval $[0, T)$, closed on the left, open on the right. (We have avoided making $t_\nu = t_1 + T$, in order to prevent misunderstandings later on, when T will be taken as a period of our time series.) Let

$$J(\mathfrak{D}) = \frac{1}{T} \operatorname{ent} (\mathbf{S}: \mathbf{R}_{t_1}, \mathbf{R}_{t_2}, \cdots, \mathbf{R}_{t_\nu}). \tag{6}$$

If $J(\mathfrak{D})$ tends to a limit as $\nu \to \infty$ and the "fineness" of $\mathfrak{D}$ (i.e. its maximum interval) tends to zero, then we call this limit the *Riemannian rate of transmission of information* over the interval T,

$$J_T = \lim J(\mathfrak{D}). \tag{7}$$

(If we allow ν to be enumerably infinite then we should get a definition of the "Lebesguian rate of transmission.")

When we applied this definition to the classical case in which $\mathbf{N}$ is band-limited white noise and $\mathbf{S}$ has the statistical properties of band-limited white noise we found that J_T was infinite. We thought that the logical basis of information theory was collapsing about our sensory organs.

In the next section we shall show how this paradox arises. In Section III we show that the paradox does not arise if we think of our time series as having a (long) period T. The paradox may be regarded as arising through taking too seriously the idea that time goes on forever.

The more use one makes of the mathematically convenient notions of continuity and infinity, the greater the chance of running into a paradox. One of the main implications of modern information theory is to show the truth of this remark. Previously it had been customary to think about noiseless continuous channels and to avoid error by the simple expedient of avoiding rigor. We do not wish to defend extremes of rigor, not even in pure mathematics, but it is not always easy to judge how much rigor is appropriate.

Band-limited noise is certainly more realistic than "purely random" Gaussian noise, which has infinite entropy rate under any reasonable definition. If we try to be still more realistic by assuming time-limited band-limited noise, then we run into trouble. For one thing, anything that is time-limited cannot be a stationary time series, unless it is the constant zero. We can retain stationarity by assuming periodic band-limited noise in our mathematical model, the period, T, being so long that the periodicity has no practical significance.

Periodic random functions have previously been discussed, for example, by Brillouin (1956, pp. 93–97) but we believe that our discussion sheds further light on the subject.

It might be suggested that if we wish to be thoroughly realistic we should use only discrete models. But the theory of numbers is as difficult as any part of mathematics. The real problem in formulating a mathematical model is to find an adequate compromise between realism and mathematical convenience. By using the notion of periodic band-limited noise in the place of time-unlimited band-limited noise we have been neither more nor less realistic, but have traded one metaphysics for another one.

It seems that the definition of the Riemannian rate of transmission must be complemented by a further statement, namely that when we apply the definition we must regard our time series as either being of finite extent (and therefore not stationary) or as having a finite (possibly long) period, T. We can, however, let the period tend to infinity at the end of the calculations, and we define J as the limit of J_T.

II. THE PARADOX

We first recall the Whittaker-Shannon sampling theorem (Whittaker, 1915; Shannon, 1949):

If

$$\int_{-\infty}^{\infty} f(t)e^{2\pi itW'}\,dt = 0$$

whenever $|W'| > W$, then

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2w}\right)\frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)}, \tag{8}$$

so that $f(t)$ is completely determined by its values at a set of points at distances apart of $1/2W$.

White noise limited to the band $(0, W)$ is defined as

$$f(t) = \sum_{n=-\infty}^{\infty} X_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)}, \tag{9}$$

where the random variables X_n have independent normal distributions of mean 0 and variance N, the mean "noise power." (The symbol N previously meant a realisation of $\mathbf{N}$, but no confusion should arise.) The autocovariance function (see, for example, Rice, 1945, §3.2) is

$$\psi_\tau = \frac{N}{W} \cdot \frac{\sin 2\pi W\tau}{2\pi\tau}. \tag{10}$$

Each X_n has entropy $\frac{1}{2} \log (2\pi eN)$ and it is entirely natural to say that the entropy rate is

$$\frac{1}{2} \log (2\pi eN) \div \frac{1}{2W} = W \log (2\pi eN).$$

If the source signal has the statistical properties of white noise with the same band-limitation then the received signal has the entropy rate $W \log (2\pi e(N + S))$, where S is the mean power of the source signal. Then, from (5), we get the Hartley-Wiener-Tuller-Sullivan-Shannon formula

$$W \log (1 + S/N) \tag{11}$$

for the expected rate of transmission of information concerning the source signal.

But what if we use some representation of white noise other than (9)? The set of values of the noise amplitudes at ν instants of time, $t_1, t_2, \cdots, t_\nu$ have a multivariate normal distribution with covariance matrix, or "power matrix" (as we may call it since it is the natural generalization to ν dimensions of the noise power),

$$\mathbf{N} = \{N\psi_{t_i - t_j}\} \tag{12}$$

where ψ_τ is defined by equation (10). (Note that we have now given a new meaning to $\mathbf{N}$, and will later use $\mathbf{S}$ for the covariance or power matrix of $\mathbf{S}$. There seems to be little danger of confusion.) Now the entropy of this multivariate normal distribution is (see[1] Appendix A)

$$\tfrac{1}{2} \log [(2\pi e)^\nu \mid \mathbf{N} \mid], \tag{13}$$

where $\mid \mathbf{N} \mid$ means the determinant of $\mathbf{N}$ (which is non-singular; see Appendix C). If the source signal also has the statistical properties of Gaussian noise, with covariance matrix $\mathbf{S}$, then the covariance matrix

[1] The appendices are an essential part of this paper and are placed at the end only so that the rest can be read without interruption.

of the received signal is $\mathbf{N} + \mathbf{S}$, and the expected amount of information concerning the source signal, provided by the received signal when observed at times $t_1, t_2, \cdots, t_\nu$ is, by equation (4),

$$\frac{1}{2} \log \frac{|\mathbf{N} + \mathbf{S}|}{\mathbf{N}} = \frac{1}{2} \log |\mathbf{I} + \mathbf{S}/\mathbf{N}|. \tag{14}$$

We may call $\mathbf{S}/\mathbf{N}$ the *matrix signal-to-noise ratio.* ($\mathbf{I}$ is the identity matrix.) If, in particular, the source signal has the statistics of white noise, band-limited to the same band as is $\mathbf{N}$, then

$$|\mathbf{N} + \mathbf{S}|/|\mathbf{N}| = (N + S)^\nu/N^\nu \tag{15}$$

and the expected amount of information is $\frac{1}{2} \nu \log [1 + (S/N)]$. This is true however closely packed are the time-instants $t_1, t_2, \cdots, t_\nu$. The Riemannian rate of transmission of information seems to be infinite. The conclusion can be avoided by throwing the argument away and using another one (such as Shannon's argument depending on degrees of freedom) but what is wrong with the argument as it stands? We believe that the answer must be that either the definition of the Riemannian rate of transmission must go, or else the notion of unlimited-time band-limited noise must go. Each of these notions can be used, but they cannot both be used in the same model. In the next section we show that the Riemannian rate can be retained provided that we make use of periodic time series.

III. PERIODIC BAND-LIMITED NOISE

If a function $f(t)$ with period T is such that its Fourier series in $(0, T)$ has no frequency as great as $W = n_0/T$, i.e.

$$\int_0^T f(t) e^{-2\pi i n t/T} \, dt = 0; \tag{16}$$

whenever n_0 is a positive integer, n is an integer, and $|n| \geqslant n_0$, then $f(t)$ is completely determined by its values at *any* set of $(2n_0 - 1)$ points in $(0, T)$. (See Appendix B.) The $(2n_0 - 1 = n_1)$ points do not need to be uniformly spaced, but if they are we have the formula

$$f(t) = \sum_{n=0}^{n_1 - 1} f\left(\frac{nT}{n_1}\right) \frac{\sin \pi \left(n - \dfrac{n_1 t}{T}\right)}{n_1 \sin \pi \left(\dfrac{n}{n_1} - \dfrac{t}{T}\right)}. \tag{17}$$

(See, for example, Goldman (1953, p. 368), Brillouin (1956, pp. 95–96), Good (1955c); and, for applications of the basic idea to pure mathematics, D. G. Kendall (1942–1943), Good (1955b). A short proof is given in Appendix B.)

Periodic band-limited white noise of period T and mean power N, band-limited to $(0, W)$, may be defined as

$$f(t) = \sum_{n=0}^{n_1-1} X_n \frac{\sin \pi \left(n - \frac{n_1 t}{T}\right)}{n_1 \sin \pi \left(\frac{n}{n_1} - \frac{t}{T}\right)}, \tag{18}$$

where the X_n are independently normally distributed with zero means and variances N. Of course $X_n = f(nT/n_1)$. In spite of its periodicity, (18) really does define a stationary time series, that is to say the joint distribution of the values of $f(t)$ at any set of instants is unchanged if a constant is added to all these instants. This follows from the three facts (i) that the joint distribution is multivariate Gaussian, (ii) that the mean of $f(t)$ is always zero, (iii) that an autocovariance function exists. The first two of these facts are obvious, whereas it is proved in Appendix B that the auto-covariance is

$$\psi_\tau = \frac{N \sin (\pi n_1 \tau / T)}{n_1 \sin (\pi \tau / T)}. \tag{19}$$

(Compare Eq. (10).) Naturally $\psi_\tau = N$ if τ is a multiple of T.

Now we can return to our problem in information theory. Any $(2n_0 - 1)$ sampled values of the noise will completely determine it. If we sample ν values, where $\nu > (2n_0 - 1)$, we still get a multivariate normal distribution but it will be a singular (= degenerate) one. (See Appendix C.) Moreover the rank will be $2n_0 - 1 = n_1$, so that the amount of information concerning the source signal is not increased by sampling more than n_1 points. These n_1 points need not be uniformly spaced. In the mathematical model using time-unlimited band-limited noise we do not get a singular covariance matrix no matter how many (distinct) points are sampled, provided that the number of them remains finite. If there is a direct resolution of the paradox it would have to be via the theory of infinite matrices and Hilbert space. The model would then be logically more difficult to handle, though it may sometimes be mathematically easier in a formal sense.

In the periodic model we have just seen that the same amount of information is obtained concerning the source signal whatever the n_1 sampling points may be (in the interval $[0, T)$). In practice it seems likely to be unprofitable to select the sampling points too close together. There is a source of noise not taken into account in the Gaussian model, namely the difficulty of making measurements to a great many places of decimals. This 'rounding off' of the measurements may be regarded as an additional source of noise that applies more or less independently to each measurement.[2] It is reasonable to conjecture that when allowance is made for rounding-off noise then the uniform spacing (or "timing") of the sampling points would be at least approximately optimal, in the sense of maximizing the expected amount of information for a specified number of sampling points. (It is easier to interpolate than to extrapolate.) It may be true rather generally for models in which the source signal and the noise are not both Gaussian (or even if they are but the number of sampling points is less than n_1) that the sampling points should be well spaced or uniformly spaced, in order to maximize the expected amount of information concerning the source signal. It would also be of some interest to know what the effect would be on the variance of the amount of information. (We are here taking seriously the distinction between amounts of information and expected amounts, but we do not know whether it would be better to minimize or to maximize the variance. It may depend on the application: if for some purpose the expected amount of information is less than is required then we should like the variance to be large, whereas if the expected amount is more than is required then we should like the variance to be small. A few further remarks are made in Appendix D concerning the distribution of the amount of information.)

It may be further conjectured that if allowance is made for rounding off then even the ordinary non-periodic model may be used in conjunction with the definition of Riemannian rate of transmission!

We conclude this section with some remarks that are perhaps obvious at this stage. In our periodic model, if we sample at least n_1 points (instants of time), and if the source signal has the same band limitation as the noise, then the entropies of **R** and **N** can easily be seen to differ by

$$\frac{n_1}{2} \log \frac{N + S}{N} = \left(TW - \frac{1}{2}\right) \log \left(1 + \frac{S}{N}\right),$$

[2] The rounding off is not quite independent of the source signal, so that equation (4) is not now accurate.

so that the Riemannian rate of transmission is

$$\left(W - \frac{1}{2T}\right) \log \left(1 + \frac{S}{N}\right). \tag{20}$$

When $T \to \infty$ this expression tends to the Hartley-Wiener-Tuller-Sullivan-Shannon formula (11).

APPENDIX A. THE ENTROPY OF MULTIVARIATE NORMAL DISTRIBUTIONS

It is remarked by Shannon (1948, § 20) that the entropy of the n-dimensional multivariate normal distribution with density

$$p = \frac{|\,\mathbf{a}\,|^{1/2}}{(2\pi)^{(1/2)n}} \exp\left(-\frac{1}{2} \sum_{i,\, j=1}^{n} a_{ij}x_ix_j\right) \tag{21}$$

is

$$\tfrac{1}{2} \log \left((2\pi e)^n \,|\,\mathbf{a}\,|^{-1}\right), \tag{22}$$

where $\mathbf{a} = \{a_{ij}\}$. Perhaps the quickest proof of this result is to quote the more general fact that the distribution of the quadratic form

$$a(x) = \sum_{i,\, j} a_{ij}x_ix_j \tag{23}$$

is that of a gamma-variate (chi-squared) with n degrees of freedom. (See, for example, Wilks, 1946, p. 104.) It follows that the expectation of the quadratic form is n and formula (22) follows at once. We can say more about the distribution of $-\log p$. For

$$-\log p = \tfrac{1}{2} \log \left((2\pi)^n \,|\,\mathbf{A}\,|^{-1}\right) + \tfrac{1}{2}\chi_{[n]}^{\;2} \tag{24}$$

where $\chi_{[n]}^{\;2}$ is a gamma-variate with n degrees of freedom.

The n-dimensional characteristic function (Fourier transform) of (21) is

$$\exp\left(-\frac{1}{2} \sum_{i,\, j=1}^{n} A_{ij}u_iu_j\right) \tag{25}$$

where $u_1, u_2, \cdots, u_n$ are the variables of the characteristic function. (See, for example, Wilks, 1946, p. 70; or Cramér, 1946, p. 311.) Here $\mathbf{A} = \{A_{ij}\} = \mathbf{a}^{-1}$, and is the covariance matrix. The convolution of two statistically independent multivariate normal distributions is obtained by adding their covariance matrices.

A singular multivariate normal distribution is defined as one with a characteristic function of the form (25) but for which the matrix **A** is singular. (See Cramér, 1946, p. 312.) The matrix **A** is then positive semi-definite instead of being positive definite, and the distribution is concentrated in an r-dimensional manifold, where r is the rank of **A**. The entropy of a singular multivariate normal distribution may be defined by restricting our attention to the r-dimensional submanifold. This entropy is

$$\frac{1}{2}\log\left((2\pi e)^r \prod_{i=1}^{r} \lambda_i\right) \tag{26}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_r$ are the positive eigenvalues of the real symmetric positive semi-definite matrix **A**. In the particular case in which $r = n$, i.e. when the distribution is non-singular, formula (26) reduces to

$$\tfrac{1}{2}\log\left((2\pi e)^n \,|\,\mathbf{A}\,|\,\right), \tag{27}$$

which agrees with formula (22).

APPENDIX B. PERIODIC BAND-LIMITED FUNCTIONS

If $f(t)$ has period T and has no frequency as large as $W = n_0/T$ in its Fourier series, then it is equal to $\exp(-2\pi i(n_0 - 1)/T)$ multiplied by a polynomial in $\exp(2\pi i/T)$ of degree $2n_0 - 2$, and therefore it is completely determined by its values at any $n_1 = 2n_0 - 1$ points. If we can conjure up a formula that takes the right values at the right points, and is periodic with period T, and band-limited to frequencies less than W, then it is the right formula. Such a formula is given by the following analogue of Lagrange's interpolation formula. It is given in Whittaker and Robinson (1924, 1940) under the heading of "trigonometrical interpolation."

$$f(t) = \sum_{n=0}^{n_1-1} f(t_n) \times \frac{\sin\frac{\pi}{T}(t-t_1)\sin\frac{\pi}{T}(t-t_2)\cdots\sin\frac{\pi}{T}(t-t_{n-1})\sin\frac{\pi}{T}(t-t_{n+1})\cdots\sin\frac{\pi}{T}(t-t_{n_1})}{\sin\frac{\pi}{T}(t_n-t_1)\sin\frac{\pi}{T}(t_n-t_2)\cdots\sin\frac{\pi}{T}(t_n-t_{n-1})\sin\frac{\pi}{T}(t_n-t_{n+1})\cdots\sin\frac{\pi}{T}(t_n-t_{n_1})} \tag{28}$$

Formula (17), which copes with the case when the points are uniformly spaced, follows by exactly the same argument (once it is written down). The following slight generalization of Formula (17) can also be proved in the same way: For all α,

$$f(t) = \sum_{n=0}^{n_1-1} f(t_n + \alpha) \frac{\sin \pi \left(n - \frac{n_1(t-\alpha)}{T}\right)}{n_1 \sin \pi \left(\frac{n}{n_1} - \frac{t-\alpha}{T}\right)}. \tag{29}$$

We shall now prove Formula (19) for the autocovariance of periodic band-limited white noise. We have, if $f(t)$ is defined by (18),

$$\begin{aligned} &\mathcal{E}(f(t)f(t+\tau)) \\ &= N \sum_{n=0}^{n_1-1} \frac{\sin \pi \left(n - \frac{n_1 t}{T}\right)}{n_1 \sin \pi \left(\frac{n}{n_0} - \frac{t}{T}\right)} \cdot \frac{\sin \pi \left(n - \frac{n_1(t+\tau)}{T}\right)}{n_1 \sin \pi \left(\frac{n}{n_1} - \frac{t+\tau}{T}\right)} \end{aligned} \tag{30}$$

If in (29) we replace t by τ and *then* replace α by $-t$ we obtain an equation which, on putting

$$f(\tau) = \frac{N \sin (\pi n_1 \tau / T)}{n_0 \sin (\pi \tau / T)},$$

is found to sum the right-hand side of Eq. (30). The result is independent of t and may therefore be denoted by ψ_τ, and Formula (19) is established.

APPENDIX C. ON THE RANKS OF COVARIANCE MATRICES OF STATIONARY TIME SERIES

We have just seen that for periodic band-limited white noise we have a covariance of the form of a scalar product:

$$\psi_{t_i - t_j} = \sum_{n=0}^{n_1-1} \varphi_n(t_i)\varphi_n(t_j). \tag{31}$$

The determinant of the covariance matrix is therefore a "Gram determinant" and it vanishes if the number of rows or columns exceeds n_1, the dimensionality of the vectors whose scalar products are the elements of the determinant. (See Hilbert and Courant, 1931, Chapter 1, §5.) This result is a special case of the result proved below.

We need a few preliminary statements.

For stationary time series the covariances are proportional to the correlations; therefore, it makes no difference to the ranks whether we use covariance or correlation matrices. The Wiener-Khintchine theorem [see, for example, Bartlett, 1955, p. 161, equation (4)] states that the correlation function of a stationary time-series is of the form

$$\rho(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega}\, dF(\omega), \tag{32}$$

where $F(\omega)$ has all the properties of a distribution function. It is called the *spectral function* of the time series. We may write

$$F(\omega) = F_1(\omega) + F_2(\omega) \tag{33}$$

where F_1 is absolutely continuous and F_2 is a step function. If F_1 vanishes then we say that the spectrum is discrete, and the values of ω at which F_2 has jumps may be called the *spectral lines.* We can prove that *if a stationary time series has a discrete spectrum with only a finite number,* n_1 *, of lines, then no covariance matrix of the time series can have rank greater than* n_1 .

Proof: The correlation function is of the form

$$\rho(\tau) = \sum_{n=1}^{n_1} a_n e^{i\tau\omega_n}$$

Therefore $\rho(t_1 - t_2)$ is the scalar product of the two vectors

$$[a_1 \exp(it_1\omega_1), \cdots, a_{n_1} \exp(it_1\omega_{n_1})]$$

and

$$[a_1 \exp(it_2\omega_1), \cdots, a_{n_1} \exp(it_2\omega_{n_1})].$$

The result now follows from the same property of Gram determinants used before, generalised in the obvious way to complex numbers. (In a complex scalar product one takes the conjugates of the components of the second vector.) For real time series there is a corresponding proof using cosines and sines instead of complex exponentials.

For periodic band-limited white noise the spectral function can be obtained by expanding Eq. (19) in a (finite) cosine Fourier series.

For time-unlimited band-limited white noise the spectral function increases linearly from 0 to 1 in the interval $(-2\pi W, 2\pi W)$ and is therefore absolutely continuous. It follows that the covariance matrices are never singular (assuming of course that the sampling time instants,

$t_1, t_2, \cdots, t_\nu$ are distinct). We can prove more generally that *if F_1 does not vanish identically then all covariance matrices are non-singular.*

Proof: The determinant of the covariance matrix is non-negative, as in Bartlett, 1955, p. 161. In virtue of the Wiener-Khintchine theorem the determinant can be expressed as a multiple Stieltjes integral of $|\{\cos (t_i - t_j)x_i\}| = C$, say. (We are considering real time series for definiteness.) The required result will be proved if we can show that the integrand, C, cannot vanish identically throughout any ν-dimensional sub-domain. If it did it would vanish identically throughout the entire domain of values of $x_1, x_2, \cdots, x_\nu$ (since it is an integral function). On expanding as a power series in the neighbourhood of the origin we would get $|\{(t_i - t_j)^{m_i}\}| = 0$ for all sets of non-negative integers $m_1, m_2, \cdots, m_\nu$ and hence that $|\{g_i(t_i - t_j)\}|$ for all sets of functions $g_1, g_2, \cdots, g_\nu$ regular at the origin. In particular we would have

$$\left|\left\{\frac{1}{K + t_i - t_j}\right\}\right| = 0$$

for all non-zero values of K; and, if the numbers $t_1, t_2, \cdots, t_\nu$ are distinct this is impossible, as we can see from the evaluation of Cauchy's "alternant." (See, for example, Aitken, 1951, p. 134, example 8.)

APPENDIX D. THE DISTRIBUTION OF AMOUNT OF INFORMATION

Towards the end of Section III we became interested not merely in the expected amount of information concerning the source signal provided by a knowledge of the amplitudes of the received signal, but in the complete distribution of the amount of information. In the notation used by Good[3] (1955a), for ("unexpectated") amounts of information, we have

$$I(S: R) = I(R) - I(R \mid S) = I(R) - I(N). \tag{32}$$

(Compare Eqs. (1) and (2).) For periodic Gaussian processes we see that the question of the distribution of $I(S: R)$ can be expressed in terms of the distribution of a quadratic form in a finite number of variables. (For non-periodic processes the number of variables would be infinite.)

RECEIVED: May 3, 1957.

[3] The paper contained four misprints, one of which may have been misleading. In the discussion of the connection between sufficient statistics and (unexpectated) amount of information the first occurrence of θ should be replaced by $\hat{\theta}$.

REFERENCES

AITKEN, A. C. (1951). "Determinants and Matrices." Oliver and Boyd, New York.

BARTLETT, M. S. (1955). "Stochastic Processes," Cambridge Univ. Press, London and New York.

BRILLOUIN, L. (1956). "Science and Information Theory." Academic Press, New York.

CRAMER, H. (1946). "Mathematical Methods of Statistics." Princeton Univ. Press, Princeton, New Jersey.

GOLDMAN, S. (1953). "Information Theory." Constable, London.

GOOD, I. J. (1955a). Some terminology and notation in information theory. *Proc. Inst. Elec. Engrs.* **103C,** 200–204.

GOOD, I. J. (1955b). A new finite series for Legendre polynomials. *Proc. Cambridge Phil. Soc.* **51,** 385–388.

GOOD, I. J. (1955c). A sampling theorem for television. Unpublished.

HILBERT, D. AND COURANT, R. (1931). "Methoden der mathematischen Physik," Vol. I. Springer, Berlin. English translation, Interscience, New York, 1953.

KENDALL, D. G. (1942–1943). A summation formula for finite trigonometrical integrals. *Quart. J. Math. Oxford* [2] **13,** 172–184.

RICE, S. O. (1944, 1945). Mathematical analysis of random noise. *Bell System Tech. J.* **23,** 282–332; **24,** 46–156. Reprinted in "Noise and Stochastic Processes," Dover, New York, 1954.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27,** 379–423, 623–656. Reprinted by Univ. Illinois Press, Urbana, Illinois, 1949.

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

WHITTAKER, E. T. (1915). On the functions which are represented by the expansions of the interpolatory theory. *Proc. Roy. Soc. Edin.* **35,** 181–194.

WHITTAKER, E. T., AND ROBINSON, G. (1924, 1940). "The Calculus of Observations," 1st and 3rd eds. Blackie, London.

WILKS, S. S. (1946). "Mathematical Statistics." Princeton Univ. Press, London and New York.