

IX. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee	A. D. Hause	M. Schetzen
Prof. A. G. Bose	I. M. Jacobs	T. G. Stockham, Jr.
D. A. Chesler	K. L. Jordan, Jr.	D. W. Tufts
D. A. George	A. H. Nuttall	C. E. Wernlein
D. L. Haas	T. V. D. Pampoucas	G. D. Zames

Foreword

Dr. Norbert Wiener of the Department of Mathematics has recently analyzed certain interesting and important problems concerning brain waves, frequency modulation, frequency variations in interconnected generators, quantum theory, coding, and non-linear filters and predictors.

Because of the great importance of this analysis in several fields of scientific work, the Statistical Communication Theory group has been helping Dr. Wiener in making it available to research workers in these fields. In order to put the material in print as soon as possible we asked Dr. Wiener to give a series of lectures to a small group – with the understanding that we would undertake to record the lectures by magnetic tape and camera and to transcribe them. We wish to mention, particularly, the cooperation of D. A. Chesler, D. A. George, I. M. Jacobs, A. H. Nuttall, and C. E. Wernlein in this project. These lectures will be published in book form by the Technology Press, M.I.T. Here we report, in preliminary form, the first two of a series of approximately twelve lectures.

We thank Dr. Wiener for the privilege of printing his work in preliminary form.

Y. W. Lee, A. G. Bose

A. NONLINEAR PROBLEMS IN RANDOM THEORY* by Norbert Wiener

Lecture 1.

The Random Function of Time and Phase

I want to discuss methods of handling a random process and, particularly, methods of handling the spectrum of the process, although these methods have a much more general application. In order to do this, I want to bring in a discussion of the Brownian motion. I am considering a Brownian motion in one dimension. Furthermore, for this hour perhaps, I shall discuss the Brownian motion over an interval of time 0 to 1. (Fig. IX-1.)

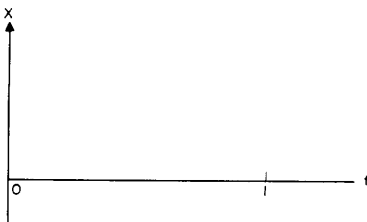


Fig. IX-1.

Let us consider the wandering of a particle. As time goes on, I want this particle to wander in a random way so that the amount that it has departed from a given position at a given time has a Gaussian distribution at

another time and so that in nonoverlapping intervals of time these Gaussian distributions are independent. I shall first discuss these motions at fixed times, and then on the

* Copyright 1958, Massachusetts Institute of Technology.

basis of the discussion of this motion at fixed times, I shall discuss the distribution of curves of this motion – first, over a finite time, and then over an infinite time. Then I am going to consider processes that depend upon this process in linear and in nonlinear ways.

The formula that we want is the formula of composition of Gaussian distributions that are independent over independent intervals of time. Consider a quantity x that has a Gaussian distribution. The probability that this quantity is between x_1 and x_2 is given by

$$P = \int_{x_1}^{x_2} \frac{1}{(2\pi a)^{1/2}} \exp\left(-\frac{x^2}{2a}\right) dx \quad (1.1)$$

I have not yet said how "a" depends upon time, but in order to get a reasonable way of doing this, I shall first discuss the composition of two such motions. That is, I want a particle that at the end of a certain time has a Gaussian distribution. Starting from where it is, the new departure has a Gaussian distribution. What is the distribution at the end of the two times?

So, I start with

$$\left[\frac{1}{(2\pi a)^{1/2}} \exp\left(-\frac{x^2}{2a}\right) \right] \left[\frac{1}{(2\pi b)^{1/2}} \exp\left(-\frac{(y-x)^2}{2b}\right) \right] \quad (1.2)$$

Note that it is $(y-x)$ rather than y which is the parameter of the second probability density in Eq. 1.2. Now, I want the probability distribution for y when x goes to all possible values. What I am interested in is

$$dy \int_{-\infty}^{\infty} \frac{1}{(2\pi a)^{1/2}} \exp\left(-\frac{x^2}{2a}\right) \frac{1}{(2\pi b)^{1/2}} \exp\left(-\frac{(y-x)^2}{2b}\right) dx \quad (1.3)$$

It is easy to compute the integral of Eq. 1.3. It will not be necessary for me to go through it here. It's quite trivial. The answer is given by Eq. 1.4

$$dy \int_{-\infty}^{\infty} \frac{1}{(2\pi a)^{1/2}} \exp\left(-\frac{x^2}{2a}\right) \frac{1}{(2\pi b)^{1/2}} \exp\left(-\frac{(y-x)^2}{2b}\right) dx = dy \frac{1}{[2(a+b)]^{1/2}} \exp\left(-\frac{y^2}{2(a+b)}\right) \quad (1.4)$$

That is the law of composition of Gaussian distributions. Notice that this parameter "a" adds up when we compound two Gaussian distributions. If we then consider that the wandering in nonoverlapping intervals is Gaussian and independent, and that the amount of wandering is dependent only on the time interval and not on the original time, then we see that because a and b add, the distribution starting from a certain time

(IX. STATISTICAL COMMUNICATION THEORY)

and ending at a certain other time is given by

$$\frac{1}{(2\pi k)^{1/2}} \exp\left(-\frac{x^2}{2k}\right) \quad (1.5)$$

where k depends linearly on the time difference, t ; and I shall normalize it so that it is t itself.

$$\frac{1}{(2\pi t)^{1/2}} \exp\left(-\frac{x^2}{2t}\right) \quad (1.6)$$

Notice that if I compound Gaussian distributions then I add the time parameters for the new Gaussian distributions.

For the moment, I am going to consider trying to map all paths of particles on the variable a when a goes from 0 to 1. I want to set up this mapping in detail. The first thing is that all random motions of particles end up somewhere; that is, they will all map on this interval 0 to 1 of a . At the present stage, I have not separated the different ranges of a ; I have taken the whole range from 0 to 1 and assigned it to all of the curves.

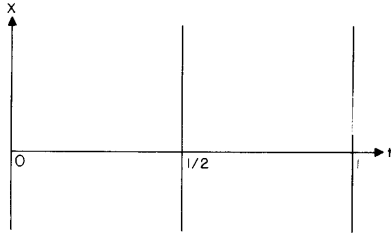


Fig. IX-2.

Let us consider the Brownian motion at two times, 1/2 and 1. (Fig. IX-2.) I shall introduce four classes of Brownian motion. Notice that at time 1/2 we have two possibilities: below the axis and above the axis.

The probability of it being on the axis is 0. It is like tossing a coin, where the probability of standing on end is 0. Given that it is either heads or tails, there is a possibility of heads-heads, heads-tails, tails-heads, and tails-tails. What are these possibilities of heads-heads, tails-tails, and so forth? The two values of a and b here are both 1/2. The distribution is given by

$$\frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{x^2}{2(1/2)}\right) \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{(y-x)^2}{2(1/2)}\right) dx dy \quad (1.7)$$

Expression 1.7 is the probability that the particle at time 1/2 lies between x and $x + dx$; and at time 1 lies between y and $y + dy$. On this basis, the probability that the particle lies, say, below the axis at time 1/2 (that x be negative at time 1/2) and x be negative at time 1 is given by

$$\int_{-\infty}^0 \int_{-\infty}^0 \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{x^2}{2(1/2)}\right) \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{(y-x)^2}{2(1/2)}\right) dx dy \quad (1.8)$$



Fig. IX-3.

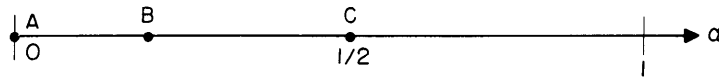


Fig. IX-4.

This expression is a quantity that has a definite value. Well, let us start from 0 and lay out an interval AB of that length on the line of α . (Fig. IX-3.) We shall say that AB corresponds to curves ending below both axes. Then we shall take x between 0 and ∞ , and take y between $-\infty$ and 0; and that probability will be given by

$$\int_{x=0}^{\infty} \int_{y=-\infty}^0 \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{x^2}{2(1/2)}\right) \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{(y-x)^2}{2(1/2)}\right) dx dy \quad (1.9)$$

Let us then lay off another interval, BC, of α , that will correspond to Eq. 1.9. (Fig. IX-4.) [Notice that the first two probabilities will add up to the probabilities of all the curves that end below the axis, and that probability will be 1/2.] Then we integrate Eq. 1.7 as follows:

$$\int_{x=-\infty}^0 \int_{y=0}^{\infty} \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{x^2}{2(1/2)}\right) \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{(y-x)^2}{2(1/2)}\right) dx dy \quad (1.10)$$

That will give us a third segment, CD, of the line. (Fig. IX-5.) Finally, Eq. 1.11 gives the fourth segment, DE. (Fig. IX-6.)

$$\int_{x=0}^{\infty} \int_{y=0}^{\infty} \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{x^2}{2(1/2)}\right) \frac{1}{[2\pi(1/2)]^{1/2}} \exp\left(-\frac{(y-x)^2}{2(1/2)}\right) dx dy_1 \quad (1.11)$$

We now have four segments of the line α ; that is, four probabilities that add up to 1.

Let us take a finer subdivision of the Brownian motions; we shall take the subdivisions where, as you will easily see, every possibility lies in one and only one of the previous possibilities, and where a certain fixed number of these possibilities will add up to give the previous complete probability. We shall get a finer subdivision of the α line; it will correspond to regions of wandering of the Brownian motion. The

(IX. STATISTICAL COMMUNICATION THEORY)

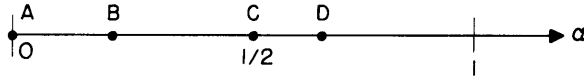


Fig. IX-5.

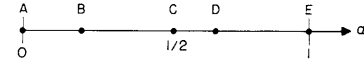


Fig. IX-6.

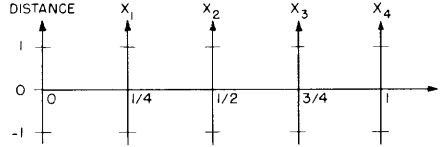


Fig. IX-7.

first thing that I shall do, is to divide the time finer. That is, I shall work by quarter times. (Fig. IX-7.) The probability that a particle lies in a small region specified at each of the four times is given by

$$\left(\frac{1}{[2\pi(1/4)]^{1/2}} \right)^4 \exp\left(-\frac{x^2}{2(1/4)}\right) \exp\left(-\frac{(x_2 - x_1)^2}{2(1/4)}\right) \exp\left(-\frac{(x_3 - x_2)^2}{2(1/4)}\right) \exp\left(-\frac{(x_4 - x_3)^2}{2(1/4)}\right) \times dx_1 dx_2 dx_3 dx_4 \quad (1.12)$$

Then I advance the probability that the particle moves through a hole (region) specified at each of the 4 times. I divide the time into regions 1/4, 1/2, 3/4, and 1, and the regions will now be 4 for each time. At each one of these times, not only do I distinguish positive and negative values but I distinguish values from $-\infty$ to -1 , from -1 to 0 , from 0 to 1 , and from 1 to ∞ . How many types of curves will I have? I will have four possible regions for each time, but there will be four successive times. That is 4^4 regions, which is 256. How do I get the probability for each one of those regions? I simply integrate Eq. 1.12 with respect to the different x 's over that region. I will get 256 quantities that add up to 1; moreover, there will be 64 of these quantities adding up to each of the previous intervals. They will not be the same length, but I will be able to get, with a definite arrangement, 256 subregions, each adding up by 64's to the previous subregions, and I will have a mapping of ranges of my random curve on smaller ranges of a .

Having done this, I take a finer subdivision. (Fig. IX-8.) In the first place, the times will now go by eights, so that for the probability density we shall have

$$\left(\frac{1}{[2\pi(1/8)]^{1/2}} \right)^8 \exp\left(-\frac{x^2}{2(1/8)}\right) \dots \exp\left(-\frac{(x_8 - x_7)^2}{2(1/8)}\right) \quad (1.13)$$

(IX. STATISTICAL COMMUNICATION THEORY)

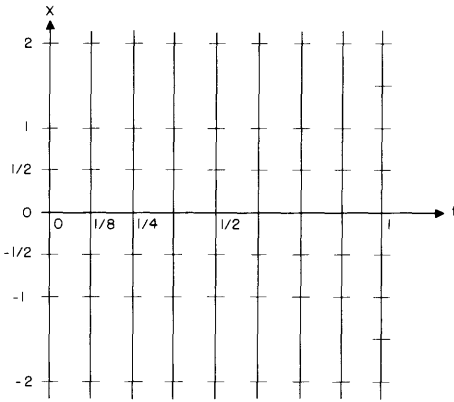


Fig. IX-8.

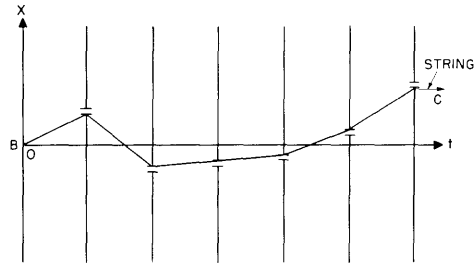


Fig. IX-9.

I shall make the integration over regions in the following way (and this trick I shall continue): I have now eight times. I take all of the previous subdivisions, not only for the times I had before but also for these other intermediate times. Then, for any of these intervals that have limits at both ends, I introduce a new subdivision half-way up. There is no virtue to a half, but using a half is a perfectly good way of doing it. For the regions that run to $+\infty$ the new subdivision will be one unit up; for the regions that run to $-\infty$ the new subdivision will be one unit down.

Now we have not 4^4 regions but 8^8 of these curves. Each one of these will have a length; each one of these will fit into one of the regions of the previous subdivisions, and have finer divisions of the 0-to-1 line of a . So that, as we continue cutting the motion finer and finer, both in time and space, we shall have a larger number of steps of a . With this trick, or with a much more general trick of subdivision, it is easy to show that these intervals of a will all go to 0 in length. If we consider any value of a that is not one of this denumerable set of subdivisions, this value of a will lie in one of the four regions that we got at the first stage, and one of the 256 regions that we got at the second stage, and so on. It will be uniquely determined (except for the boundaries that do not count) by where it is at the different stages of subdivisions. In other words, I am able to box in these wanderings of our point finer and finer.

It will not be true, in general, that as I do this I close down on a particular curve that is continuous; but I want to show you that I can get around this difficulty. I can introduce a certain quantity, which I shall call the straightness of a series of subdivisions. There is no virtue in the 2^n at present, so that I just say that I have a series of holes – one hole at each of an arbitrary number of times. (Fig. IX-9.)

Suppose that we take a string BC fastened at 0 and thread it through the holes. Let me pull that string taut. Then there will be at least one part of the curve with maximum slope. This maximum slope for the taut string will be less than the maximum slope of

(IX. STATISTICAL COMMUNICATION THEORY)

any other curve passing through the same series of holes. We can call this taut string the straightest string that lies in the series of holes. This string will give us the smoothness of the series of holes in the following manner. I call this straightest string through a series of holes, $x = s(t)$, for that series of holes; a particular s will belong to each series of holes. Now consider

$$\max \frac{|s(t+\tau) - s(t)|}{|\tau|^{1/4}} \tag{1.14a}$$

That maximum value I call the smoothness of the series of holes. There is no virtue in one-fourth, except this: it is a definite number less than one-half. Now I have a series of holes at each stage of subdivision, and at each stage of subdivision I throw away all of the series of holes for which the smoothness is greater than A ; that is,

$$\max \frac{|s(t+\tau) - s(t)|}{|\tau|^{1/4}} > A \tag{1.14b}$$

I can now prove the following: The sum of the lengths of the α mappings of the series of holes that I have thrown away at each stage will be finite. At each stage when I throw away these series of holes, I will have an expression $\ell(A)$, the sum of the lengths that I have thrown away. There will be only a denumerable number of series of holes (a denumerable number of intervals); the sum of these lengths will converge and will form what we call a measurable set. A measure of that set will be less than $\ell(A)$.

Although I shall not prove it here, because we are interested in getting on, I can prove that

$$\lim_{A \rightarrow \infty} \ell(A) = 0 \tag{1.14c}$$

That is, the sum of the lengths of all of the series of holes at any stage which are less smooth than a certain amount is finite; and if A increases, this sum goes to 0. Let me discard the series of holes that are less than a certain smoothness. That means that I discard a certain set of values of α of measure less than some small amount. Let us take the remaining series of holes. The remaining series of holes will each contain a continuous curve; moreover, all of these continuous curves will satisfy the same condition of equicontinuity. That is, Δx goes to 0 faster than a certain function of Δt .

There is the following theorem in mathematics. Suppose that I take a set of curves that pass through a series of holes and satisfy a certain condition of equicontinuity. Suppose, also, that I make the holes narrower and narrower, and that I increase the number of time instants at which the holes are defined. This condition of equicontinuity binds my curves tighter and tighter at more and more points. The curves then tend

uniformly to a limit that satisfies the same condition of equicontinuity. That is, if I throw away a certain set of values of a and s such that the measure of s is less than ϵ , then to all of the remaining values of a there will be assigned a limit curve for all the series of holes that correspond successively to this value of a . The limit curve will be continuous; not only that, it will also satisfy the same condition of equicontinuity. Thus, by this process, I have assigned to all values of a (except for a set of zero measure) a curve, which I call $x(t, a)$, that satisfies some condition of equicontinuity; and what is more, to all of these curves except a set of measure zero I have assigned a value of a . These limit curves will prove to be unique. Therefore I will have assigned (except for a set of zero measure) to every value of a , one and only one continuous curve. This I call $x(t, a)$, as I have said. It is a well-defined function of t for almost all values of a . It can easily be shown to be a measurable, bounded function of t and a , and a continuous function of t for almost all a . Furthermore, if the process at each stage of arrangement of the holes is given definitely, which can be done, this is a well-defined function of t and a — as well-defined as any mathematical function. So, I have now introduced what we call the stochastic function $x(t, a)$.

Are there any other things that I can say about $x(t, a)$? I have said that for almost all values of a this is continuous. Is it differentiable? I shall not go into the proof — it is shown by merely taking the different cases, adding them, and counting — but I can say that the following thing can be proved.

Consider the set of curves for which Eq. 1.15a has a limit for at least one value of t :

$$\frac{\Delta_t x(t, a)}{\Delta t} \tag{1.15a}$$

This set of curves has zero measure. That is, almost all of the curves $x(t, a)$ are nowhere differentiable. This is important. We are going to use nondifferentiable continuous curves in the work that we are doing. Not only that; the limit of Eq. 1.15b will exist for no t for almost all a if λ is greater than $1/2$, and will exist for every t for almost all a uniformly if λ is less than $1/2$.

$$\frac{\Delta_t x(t, a)}{(\Delta t)^\lambda} \tag{1.15b}$$

(I am leaving out the $1/2$ case. I am stating the facts here rather than proving.)

The function that I want to use in our further work is $x(t, a)$. I assume that we have established $x(t, a)$. We shall call it the stochastic function. Now to build up some of the integral properties of $x(t, a)$.

(IX. STATISTICAL COMMUNICATION THEORY)

Let us consider the following integral:

$$I = \int_0^1 [x(t_2, a) - x(t_1, a)]^n da \quad [t_1 < t_2] \quad (1.16)$$

Now, $[x(t_2, a) - x(t_1, a)]$ has a Gaussian distribution. That can be proved very easily from our definition. With $[t_2 - t_1]$ as the parameter of the Gaussian distribution, Eq. 1.17 follows.

$$I = \frac{1}{[2\pi(t_2 - t_1)]^{1/2}} \int_{-\infty}^{\infty} u^n \exp\left(-\frac{u^2}{2(t_2 - t_1)}\right) du \quad [t_1 < t_2] \quad (1.17)$$

This equation is certainly true for binary intervals, and by continuity we can extend it easily to nonbinary intervals. That is, these distributions are Gaussian. Equation 1.17 can be computed as follows. Let

$$v = \frac{u}{(t_2 - t_1)^{1/2}} \quad (1.18)$$

Then

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} v^n \exp\left(-\frac{v^2}{2}\right) dv \quad (1.19)$$

Note that $\exp(-v^2/2)$ is an even function. If n is odd, then

$$(t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} v^n \exp\left(-\frac{v^2}{2}\right) dv = 0 \quad (1.20)$$

because the integral from $-\infty$ to $+\infty$ of the product of an odd and an even function is zero.

If n is even, we can compute Eq. 1.19 by an integration by parts. Note that

$$-v \exp\left(-\frac{v^2}{2}\right) dv = d \exp\left(-\frac{v^2}{2}\right) \quad (1.21)$$

Then we have

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} (-1) v^{n-1} d \exp\left(-\frac{v^2}{2}\right) \quad (1.22)$$

Integrating by parts, we get

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} \left\{ \left[-v^{n-1} \exp\left(-\frac{v^2}{2}\right) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (n-1) v^{n-2} \exp\left(-\frac{v^2}{2}\right) dv \right\} \quad (1.23)$$

$$I = (t_2 - t_1)^{n/2} \frac{1}{(2\pi)^{1/2}} (n-1) \int_{-\infty}^{\infty} v^{n-2} \exp\left(-\frac{v^2}{2}\right) dv \quad (1.24)$$

Continuing this method, we get

$$I = (t_2 - t_1)^{n/2} (n-1)(n-3)\dots(1) \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \quad (1.25)$$

The series $(n - 2k)$ goes down to 1 because n is even. At the 1 stage, we can evaluate

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv = 1 \quad (1.26)$$

Hence

$$\int_0^1 [x(t_2, a) - x(t_1, a)]^n da = \begin{cases} 0, & n \text{ odd} \\ (t_2 - t_1)^{n/2} (n-1)(n-3)\dots(1), & n \text{ even} \end{cases} \quad (1.27)$$

The right-hand side of Eq. 1.27 is interesting. Suppose that I have n terms and that I want to consider the number of ways in which I can divide n terms into pairs. If n is odd, there is no way of separating n terms into pairs. If n is even, let us see how many ways there are. We take any term. There are $n - 1$ possible terms we can match it with. When we have done that, there are $n - 2$ terms. Then we take any term, and we can match that with $n - 3$ possibilities. Hence, in all cases, odd or even, the following equation is true.

$$\int_0^1 [x(t_2, a) - x(t_1, a)]^n da = (t_2 - t_1)^{n/2} \times \begin{matrix} \text{(the number of ways of sepa-} \\ \text{rating } n \text{ terms into pairs)} \end{matrix} \quad (1.28)$$

This is the beginning of the calculus of random functions.

Now let us consider the following integral:

$$\int_0^1 f(t) dx(t, a) \quad (1.29)$$

Notice that as an ordinary Stieltjes integral, this integral will not exist, because $x(t, a)$ is almost never differentiable. We can get around this. For the moment, let us suppose

(IX. STATISTICAL COMMUNICATION THEORY)

that $f(t)$ is differentiable and that $f'(t)$ is bounded. (I confine myself at present to that case.) First, we shall define Eq. 1.29 by integration by parts.

$$\int_0^1 f(t) dx(t, a) = f(1) x(1, a) - f(0) x(0, a) - \int_0^1 f'(t) x(t, a) dt \quad (1.30)$$

Now $x(0, a)$ is 0. Remember that the strings that I pick pass through the origin. Therefore

$$\int_0^1 f(t) dx(t, a) = f(1) x(1, a) - \int_0^1 f'(t) x(t, a) dt \quad (1.31)$$

Furthermore, the right-hand side of Eq. 1.31 is a well-defined expression for almost all values of a . $x(t, a)$ is bounded, f' is bounded, and hence the integral of Eq. 1.31 exists. Moreover, if I change $x(t, a)$ into $-x(t, a)$ you will see that the distribution of the Brownian motions is not changed at all. If I just simply turn $x(t, a)$ up and down, not only is Eq. 1.31 an existing expression, but its integral with respect to a will be 0.

The next expression that I want to get is

$$\int_0^1 da \left[\int_0^1 f(t) dx(t, a) \right]^2 \quad (1.32)$$

Integrating by parts and expanding, we have

$$\int_0^1 da \left[\int_0^1 f(t) dx(t, a) \right]^2 = \int_0^1 da \left[f(1) x(1, a) - \int_0^1 f'(t) x(t, a) dt \right]^2 \quad (1.33)$$

$$\begin{aligned} &= \int_0^1 da f^2(1) x^2(1, a) - 2 \int_0^1 da f(1) x(1, a) \int_0^1 f'(t) x(t, a) dt \\ &\quad + \int_0^1 da \int_0^1 f'(t) x(t, a) dt \int_0^1 f'(s) x(s, a) ds \end{aligned} \quad (1.34)$$

Now, by the way, notice the following expression:

$$\int_0^1 x(t_1, a) x(t_2, a) da = \int_0^1 x(t_1, a) [x(t_1, a) + x(t_2, a) - x(t_1, a)] da \quad (1.35)$$

I assume that t_1 is less than t_2 . I am just adding and subtracting the same thing to get the right-hand side of Eq. 1.35. Note that $x(t_1, a)$ and $[x(t_2, a) - x(t_1, a)]$ are independent in distribution. So, when one is plus, the other is equally likely to be plus or minus; and the average of the product will be zero. And the only quantity that survives is

$$\int_0^1 x(t_1, a) x(t_2, a) da = \int_0^1 x^2(t_1, a) da = t_1 \quad (1.36)$$

Now consider the three terms of Eq. 1.34. The first term is given by

$$\int_0^1 da f^2(1) x^2(1, a) = f^2(1) \quad (1.37)$$

Let us suppose that we can interchange the order of integration for the remaining two terms of Eq. 1.34. (It can be fairly easily proved that this interchange is justified.)

Hence, for the second term, we have

$$-2 \int_0^1 da f(1) x(1, a) \int_0^1 f'(t) x(t, a) dt = -2 f(1) \int_0^1 t f'(t) dt \quad (1.38)$$

Equation 1.38 is true, because t is less than 1.

Now consider the remaining term of Eq. 1.34.

$$\begin{aligned} & \int_0^1 da \int_0^1 f'(t) x(t, a) dt \int_0^1 f'(s) x(s, a) ds \\ &= \int_0^1 f'(t) dt \int_0^1 f'(s) ds \int_0^1 x(t, a) x(s, a) da \end{aligned} \quad (1.39)$$

Recall

$$\int_0^1 x(t, a) x(s, a) da = \begin{cases} s, & s \leq t \\ t, & t \leq s \end{cases} \quad (1.40)$$

Equation 1.40 means that we can divide the integral of Eq. 1.39 into two ranges; in one of them s is smaller; in one, t is smaller. But that means that the two integrals are the same integral, because s and t are completely interchangeable. I substitute Eq. 1.40 in Eq. 1.39, keeping s less than t , and I multiply by 2 to account for the two cases. The right-hand side of Eq. 1.39 becomes

$$\int_0^1 f'(t) dt \int_0^1 f'(s) ds \int_0^1 x(t, a) x(s, a) da = 2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds \quad (1.41)$$

Substitution of Eqs. 1.37, 1.38, and 1.41 in Eq. 1.34 gives

$$\int_0^1 da \left[\int_0^1 f(t) dx(t, a) \right]^2 = f^2(1) - 2f(1) \int_0^1 t f'(t) dt + 2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds \quad (1.42)$$

Now

(IX. STATISTICAL COMMUNICATION THEORY)

$$2 \int_0^1 f'(t) dt \int_0^t s f'(s) ds \quad (1.43)$$

$$\begin{aligned} &= 2 \left[\int_0^t s f'(s) ds f(t) \right]_0^1 - 2 \int_0^1 t f(t) f'(t) dt \\ &= 2f(1) \int_0^1 s f'(s) ds - 2 \int_0^1 t f(t) f'(t) dt \end{aligned} \quad (1.44)$$

Therefore

$$\int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 = f^2(1) - 2 \int_0^1 t f(t) f'(t) dt \quad (1.45)$$

Now

$$\frac{d}{dt} [t f^2(t)] = f^2(t) + 2t f(t) f'(t) \quad (1.46)$$

Therefore

$$2 \int_0^1 t f(t) f'(t) dt = [t f^2(t)]_0^1 - \int_0^1 f^2(t) dt \quad (1.47)$$

$$= f^2(1) - \int_0^1 f^2(t) dt \quad (1.48)$$

Thus

$$\int_0^1 d\alpha \left[\int_0^1 f(t) dx(t, \alpha) \right]^2 = \int_0^1 f^2(t) dt \quad (1.49)$$

Notice what we have if we start with the assumption that $f(t)$ is differentiable and belongs to L^2 and we go from that to the function of α given by Eq. 1.49. We have a unitary transformation. This allows us to extend the definition of the integral to any function $F(\alpha)$ which belongs to the Lebesgue class L^2 , and the trick is as follows.

Suppose that we have a sequence $f_n(t)$ of real functions which belong to L^2 such that Eq. 1.50 is satisfied.

$$\int_0^1 [f_n(t) - f(t)]^2 dt \rightarrow 0 \quad (1.50)$$

That is, given any function f in L^2 , I can find such a sequence. Every function of L^2 , Lebesgue-measurable and Lebesgue-integrable-square, can be approximated by functions of bounded derivatives. There is no problem about that.

Then I form $F_n(\alpha)$ defined by Eq. 1.51.

$$F_n(a) = \int_0^1 f_n(t) dx(t, a) \quad (1.51)$$

It follows at once that

$$\int_0^1 [F_n(a) - F_m(a)]^2 da = \int_0^1 [f_n(t) - f_m(t)]^2 dt \quad (1.52)$$

and that

$$\int_0^1 [f_n(t) - f_m(t)]^2 dt \rightarrow 0 \quad (1.53)$$

as m and n tend to infinity independently.

Now we use the Riesz-Fischer theorem. If we have a sequence $\{F_n(a)\}$ belonging to L^2 such that

$$\int_0^1 [F_n(a) - F_m(a)]^2 da \rightarrow 0 \quad (1.54)$$

then there is a function of a to which they converge in the mean.

$$F(a) = \ell. i. m. \lim_{n \rightarrow \infty} F_n(a) \quad (1.55)$$

It can be proved that $F(a)$ does not depend on the sequence $F_n(a)$ that approximates it but that it will be the same for any sequence.

I now define

$$F(a) = \int_0^1 f(t) dx(t, a) \quad (1.56)$$

Equation 1.56 applies to almost all values of a . $F(a)$ is a function of L^2 . There is no problem in verifying that

$$\int_0^1 F^2(a) da = \int_0^1 f^2(t) dt \quad (1.57)$$

So, we have extended our integral to all functions belonging to the Lebesgue class L^2 .

References

1. N. Wiener, Generalized harmonic analysis, Acta Mat. 55, 117-258 (1930).
2. R. E. A. C. Paley and N. Wiener, Fourier transforms in the complex domain (Colloquium Publication No. 19, American Mathematical Society, New York, 1934).

(IX. STATISTICAL COMMUNICATION THEORY)

Lecture 2.

Homogeneous Polynomial Functionals and Their Averages

Let us suppose that

$$\phi \in L^2 \text{ over } [0, 1] \quad (2.1)$$

I have previously defined

$$\int_0^1 \phi(t) dx(t, a) \quad (2.2)$$

As you remember, $x(t, a)$ is the Brownian motion function, which I defined and which is a perfectly explicit function. Now, I proved first that

$$\int_0^1 da \int_0^1 \phi(t) dx(t, a) = 0 \quad (2.3)$$

and that

$$\int_0^1 da \left[\int_0^1 \phi(t) dx(t, a) \right]^2 = \int_0^1 \phi^2(t) dt \quad (2.4)$$

Here, $\phi(t)$ is real.

From Lecture 1, it follows that:

Given $x(t_1, a) x(t_2, a) \dots x(t_n, a)$, where the t 's and a 's all lie between 0 and 1, then for n odd,

$$\int_0^1 x(t_1, a) x(t_2, a) \dots x(t_n, a) da = 0 \quad (2.5)$$

and for n even

$$= \sum \prod \int_0^1 x(t_j, a) x(t_k, a) da \quad (2.6)$$

where the sum is over all ways of dividing n terms t_1, \dots, t_n into pairs, and the product is over all pairs in this way of dividing.

It follows at once that

$$\int_0^1 da \int_0^1 \phi_1(t_1) dx(t_1, a) \int_0^1 \phi_2(t_2) dx(t_2, a) \dots \int_0^1 \phi_n(t_n) dx(t_n, a) \quad (2.7)$$

will be a sum of terms similar to the right-hand side of Eq. 2.6. Evaluating, we get

$$\begin{aligned} & \int_0^1 da \int_0^1 \phi_1(t_1) dx(t_1, a) \int_0^1 \phi_2(t_2) dx(t_2, a) \dots \int_0^1 \phi_n(t_n) dx(t_n, a) \\ &= \sum \prod \int_0^1 \phi_j(t) \phi_k(t) dt \end{aligned} \quad (2.8)$$

Let us assume

$$K_n(\tau_1, \dots, \tau_n) = \phi_1(\tau_1) \dots \phi_n(\tau_n) \quad (2.9)$$

where the ϕ 's all belong to L^2 . More generally, I will use a sum of such products of terms. That is,

$$K_n(\tau_1, \dots, \tau_n) = \sum \phi_1(\tau_1) \dots \phi_n(\tau_n) \quad (2.10)$$

Then, for n odd

$$\int_0^1 da \int_0^1 \dots \int_0^1 dx(\tau_1, a) \dots dx(\tau_n, a) K_n(\tau_1, \dots, \tau_n) = 0 \quad (2.11)$$

and for n even

$$= \sum \int_0^1 d\tau_1 \dots \int_0^1 d\tau_m K_n(\tau_1, \tau_1, \tau_2, \tau_2, \dots, \tau_m, \tau_m) \quad (2.12)$$

where $2m = n$. That is, I simply multiply out the factors of K_n , and the sum is over all ways of dividing in pairs.

I am going to assume that K_n is symmetrical, so that I shall not have to distinguish the order in which I integrate. I think you can see that if K_n is not symmetrical I can make a symmetrical K_n in an expression of this sort by simply taking all the permutations of the τ 's, adding them, and dividing by the number of permutations. The functional that I get by that method will not be changed; I am adding the same functional to itself with a different labeling. Then

$$\begin{aligned} & \int_0^1 da \int_0^1 \dots \int_0^1 dx(\tau_1, a) \dots dx(\tau_n, a) K_n(\tau_1, \dots, \tau_n) \\ &= (2m - 1)(2m - 3) \dots (1) \int_0^1 d\tau_m K_n(\tau_1, \tau_1, \dots, \tau_m, \tau_m) \end{aligned} \quad (2.13)$$

In other words, for a symmetrical K_n , we divide the τ 's into pairs, identify the τ 's in each pair, integrate over them, and add up all the ways of doing that. How many ways are there of doing that? Well, all of these operations will be the same if the K_n are symmetrical. How many ways are there of dividing any $2n$ things into pairs? I showed

(IX. STATISTICAL COMMUNICATION THEORY)

you that at the last lecture: $(2n - 1)(2n - 3) \dots (1)$. So, we have a calculus not merely for linear functions but also for a certain very important class of nonlinear functions.

Let me now take another set of functions $\phi_n(t)$. These are to be normal and orthogonal over $[0, 1]$, and real, for the moment. Let us consider expressions of the form

$$\int_0^1 da \left[\int_0^1 \phi_n(t) dx(t, a) \right]^\nu \quad (2.14)$$

where the ϕ_n belong to L^2 . The ϕ_n must be a normal, orthogonal set, and

$$\int_0^1 da \left[\int_0^1 \phi_n(t) dx(t, a) \right]^\nu = 0, \quad \text{for } \nu \text{ odd} \quad (2.15)$$

$$= (\nu-1)(\nu-3) \dots (1) \int_0^1 \phi_n^2(t) dt, \quad \text{for } \nu \text{ even} \quad (2.16)$$

But

$$\int_0^1 \phi_n^2(t) dt = 1 \quad (2.17)$$

And so, for ν even,

$$\int_0^1 da \left[\int_0^1 \phi_n(t) dx(t, a) \right]^\nu = (\nu-1)(\nu-3) \dots (1) \quad (2.18)$$

Now, the right-hand side of Eq. 2.18 represents again, simply, without any other factor, the number of ways of dividing ν things into pairs; and if I take

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} u^\nu \exp\left(-\frac{u^2}{2}\right) du \quad (2.19)$$

that will be the same thing as Eq. 2.18, as I pointed out last time. In other words, if I consider the moments of the expressions

$$\int_0^1 \phi_n(t) dx(t, a) \quad (2.20)$$

then these moments will be exactly the same as Eq. 2.19. What I have said is equivalent to the following statement: The moments will completely determine the distribution (that, we know), and the distribution of the quantity in Eq. 2.20 is Gaussian. That is, the probability that

$$u \leq \int_0^1 \phi_n(t) dx(t, a) \leq u + du \quad (2.21)$$

is given by

$$\frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{u^2}{2}\right) du \quad (2.22)$$

So, the first thing that I have is that each of these expressions has a Gaussian distribution. Now, I want to consider the simultaneous distribution of two or more expressions of the form of Eq. 2.20. Let us see what we get. Let us take

$$\int_0^1 \left[\int_0^1 \phi_m(t) dx(t, a) \right]^{\nu_1} \left[\int_0^1 \phi_n(t) dx(t, a) \right]^{\nu_2} \left[\int_0^1 \phi_p(t) dx(t, a) \right]^{\nu_3} da \quad (2.23)$$

where ϕ_m , ϕ_n , and ϕ_p are distinct. We take three terms; we are not confining ourselves to two.

Now we want to get the integral of Eq. 2.23. When we do that, remember what we do. We divide the ϕ 's into pairs in all possible ways, integrate each pair, and add. But since the ϕ 's are orthogonal, the integral of the product of two distinct ϕ 's is 0. So, any pair in which the ϕ 's are distinct disappears, and the only ones that remain are those in which the ϕ 's are the same. And so, we find that Eq. 2.23 is equal to the product of the moments of the individual terms. In other words, the moment of the product is the product of the moments. This allows us to prove that the distributions are independent. That is, we can merely multiply the moments; and we can easily show that we then simply multiply the distributions. In other words, the quantities

$$\int_0^1 \phi_n(t) dx(t, a) \quad (2.24)$$

all have the same distribution (which is Gaussian, with the same parameter) and are independent of one another. That is the first thing. To repeat – the quantities

$$\int_0^1 \phi_n(t) dx(t, a)$$

are distributed Gaussianly.

The next thing is this. If the ϕ 's are a closed set it is easy to show that the polynomials in these integrals are a closed set of functions of a and that every function of a can be represented in terms of them. I am not going into the details. The result is that, given any closed normal and orthogonal set of functions of a , we have reduced the problem of the distribution of the functions to the distribution of these integrals, and these integrals all have the same distribution. They are all Gaussian and all independent, as I just explained. Now, this is extremely interesting, and you will see that

(IX. STATISTICAL COMMUNICATION THEORY)

there is something here that is suggestive of Fourier coefficients. Formally,

$$\int_0^1 \phi_n(t) dx(t, a) \tag{2.25}$$

will be the Fourier coefficients of

$$\frac{d}{dt} x(t, a) \tag{2.26}$$

However, this does not exist in the ordinary sense. It is not a differentiable function. If we have a differentiable function here, say $F(t)$, the integral

$$\int_0^1 F(t) \phi_n(t) dt \tag{2.27}$$

is the Fourier coefficient of $F(t)$. Call this a_n , and it follows that

$$\sum_0^\infty |a_n|^2 = \int_0^1 F^2(t) dt \tag{2.27a}$$

Everything here is real, and the space of this set of coefficients is called Hilbert space. This is real Hilbert space. (Later on, I shall go on to complex Hilbert space. But for the moment we are confining ourselves to real Hilbert space.) However, these coefficients

$$a_n(a) = \int_0^1 \phi_n(t) dx(t, a) \tag{2.28}$$

are not a set that is such that the sum of the squares converge. However, they are a set with independent Gaussian distributions. If I take the sum of the squares, the probability that this sum is finite is 0. So here I have something that is quite analogous to Hilbert space but is not Hilbert space. In Hilbert space the sum of the squares of the coefficients is finite. Here the coefficients all have independent Gaussian distributions. This space I call differentiable space.

There are many ways to go ahead. First, suppose that I consider another normal and orthogonal set ψ_n which is also closed, and I consider

$$\int_0^1 \psi_n(t) dx(t, a) \tag{2.29}$$

and compare these expressions with

$$\int_0^1 \phi_n(t) dx(t, a) \tag{2.30}$$

(IX. STATISTICAL COMMUNICATION THEORY)

These two random variables will have the same distribution for the ϕ 's as they had for the ψ 's, because the Gaussian properties are not changed. Now, I make a linear transformation that transforms ϕ_n into ψ_n and any series in one into a series in the other. That is what we call a unitary transformation. It is the general sort of transformation that does not affect the integral of the square of the function. I am talking of real functions now. So, I can write

$$\psi_n(t) = U \phi_n(t) \tag{2.31}$$

which is a unitary transformation; a real unitary transformation. Now what do I do here to the α ? I can work backwards from my distribution of

$$\int_0^1 \phi_n(t) dx(t, \alpha) \tag{2.32}$$

to the values of α . Essentially, the ϕ_n are functions that differ from 0 only over an interval and are 1 over that interval. I could take a closed basis. That's a particular set of ϕ_n . The particular closed basis that I shall take can be obtained as follows. The first of the functions is 1 over the line 0 to 1. The next of the functions is 1 over the first half of the line and -1 over the second half. It is orthogonal to the first. The next function is +1 over the first quarter, -1 over the second quarter, 1 over the third quarter, and -1 over the fourth quarter. That function is orthogonal to both the others. The next function is 1 over the first quarter, -1 over the second quarter, -1 over the third, 1 over the fourth. These functions are known as Walsh functions, and I shall tell you how they are obtained. I take a function which is 1 over the first $\frac{1}{2^n}$ th of the line, -1 over the second $\frac{1}{2^n}$ th, 1 over the third, -1 over the fourth, 1 over the fifth, and so forth. This is the n^{th} Haar function. Now, I take the product of the Haar functions two at a time, three at a time, and so on. The resultant functions, f_n , are the Walsh functions. Any one of these Walsh functions, over a large interval, will always range between 1 and -1 while any other is 1 or -1. The result is that these Walsh functions will be orthogonal to one another, and will be normal.

Then in terms of these Walsh functions of ϕ we take

$$\int_0^1 \omega(t) dx(t, \alpha) \tag{2.33}$$

where $\omega(t)$ is one of the Walsh functions of ϕ . From that I can obtain α , because I have the integrals, and I have the differences between the x 's over every binary interval. From that I can go over the process by which I have obtained the α . Now, if I start with another set of orthogonal functions of the Walsh functions I can say that the integrals of Eq. 2.33 are to be the corresponding integrals of the Walsh functions of another

(IX. STATISTICAL COMMUNICATION THEORY)

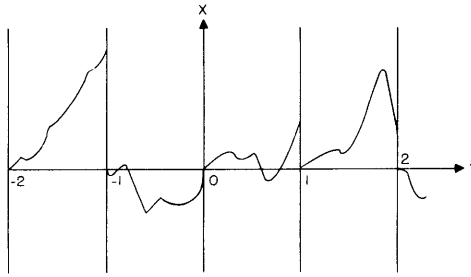


Fig. IX-10.

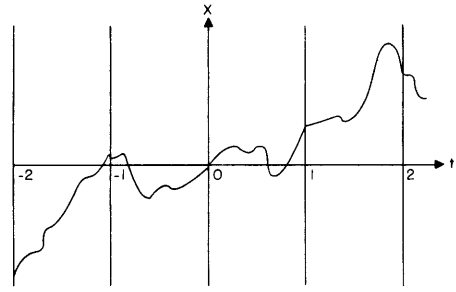


Fig. IX-11.

variable, and I get the a of the other variable. When I have done that, I close down on a , and I get a one-one point-mapping of the new a 's on the old a 's; and this mapping preserves measure because it preserves the integral over any interval. In other words, if I make this transformation I get the following result:

$$\int_0^1 U[\phi_n(t)] dx(t, a) = \int_0^1 \psi_n(t) dx(t, Ta) \quad (2.34)$$

where Ta is a measure-preserving transformation. In other words, I am able to refer unitary transformations on the ϕ space to measure-preserving transformations on the a space. This is a very interesting thing. I am able to go through unitary transformations.

There are several things of a fairly simple character that I can do to generalize this material. There are two generalizations that I will have time to do now: (a) the generalization for a range of t , not 0 to 1, but $-\infty$ to ∞ (which is quite important when we consider spectral problems), and (b) the generalization from a real differentiable space to a complex differentiable space. Now for the first step. Suppose that we take the line of t and divide it from 0 to 1, 1 to 2, 2 to 3, -1 to 0, -2 to -1, and so on. Suppose we have a Brownian motion curve in each region (Fig. IX-10).

It does not matter which way I start the Brownian motion curve in each interval, so I shall start it at the beginning of each interval. Remember, Brownian motion really is symmetrical, because the differences are symmetrical whether we go from the beginning or the end of an interval. The Brownian motion curve here is a set of functions $x(t, a_n)$. Now a_n is a variable that runs between 0 and 1. I want to show you that I can represent this entire sequence of functions in terms of one a that goes between 0 and 1, such that this single a will give us a measure distribution that will generate all of the measure distributions of all the other a 's. (Here, by the way, n goes from $-\infty$ to ∞ .) I shall tell you how I get that. I take a_0 . a_0 can be represented uniquely, except for a set of 0 measure, by a binary fraction. The binary fraction is

$$a_0 = \cdot a_{01} a_{02} a_{03} \dots \tag{2.35}$$

and for a_1 we have

$$a_1 = \cdot a_{11} a_{12} a_{13} \dots \tag{2.36}$$

and so on. In this way, I have a denumerable set of values arranged in a double sequence. I can rearrange this in a single sequence, and there are lots of ways in which I can do this. I can, for example, run along the dotted line

$$\begin{array}{rcccc}
 a_0 = & \rightarrow & \cdot & a_{01} & a_{02} & a_{03} & a_{04} & \dots \\
 & & & \swarrow & \swarrow & \swarrow & \swarrow & \\
 a_1 = & & \cdot & a_{11} & a_{12} & a_{13} & a_{14} & \dots \\
 & & & \swarrow & \swarrow & \swarrow & \swarrow & \\
 a_2 = & & \cdot & a_{21} & a_{22} & a_{23} & a_{24} & \dots
 \end{array}
 \tag{2.37}$$

to give

$$\cdot a_{01} a_{02} a_{11} a_{21} a_{12} a_{03} a_{04} a_{13} a_{22} \dots$$

This sequence will again give me a single binary number between 0 and 1, and this mapping is almost everywhere a one-one mapping. Furthermore, given any set of intervals in the different a 's, the measure for the new a of the set of intervals that I just sketched is simply the same as the measure of the single interval I got before. So, this is a measurement. This is a transformation that preserves measure from an infinite number of single dimensions to a single dimension. I shall call this new variable a , without a label. Then I have a sequence of expressions $x_n(t, a)$, where there is a different x_n for each interval.

Now I do the following thing: From Fig. IX-10 I construct Fig. IX-11. I just slide the curve in a region up or down till the starting point in that region is the same as the end-point of the preceding region for regions with n positive, and vice versa for n negative. The first region is left unchanged. Now, when I have put all these together I call that function $X(t, a)$, and $X(t, a)$ is defined in terms of the old $x_n(t, a)$. The properties of $X(t, a)$ are similar to the properties of $x(t, a)$. It begins at 0 at the origin, and the distribution of the difference in value over any interval is Gaussian. The integral of the square of the difference over an interval is proportional to the length of the interval, and the difference is independent for nonoverlapping intervals.

Suppose that we have a function $\phi(t)$ belonging to L^2 over the infinite interval. Now, $\phi(t)$ can be cut out into a denumerable set of ϕ_n . That is, we can take

$$\phi_1(t) = \left\{ \begin{array}{ll} \phi(t) & t \in [0, 1] \\ 0 & \text{otherwise} \end{array} \right\}
 \tag{2.38}$$

(IX. STATISTICAL COMMUNICATION THEORY)

$$\phi_2(t) = \left\{ \begin{array}{ll} \phi(t) & t \in [1, 2] \\ 0 & \text{otherwise} \end{array} \right\} \quad (2.39)$$

and so on. I can get an infinite set. Then I form

$$\sum \int \phi_n(t) dx_n(t, a) \quad (2.40)$$

and define it as equal to

$$\int_{-\infty}^{\infty} \phi(t) dX(t, a) \quad (2.41)$$

There is no problem about that. It is elementary manipulation. Having done that, I have defined this for every function which is L^2 over the whole infinite interval. Now, I have the following result:

$$\int_0^1 \left[\int_{-\infty}^{\infty} \phi(t) dX(t, a) \right]^2 da = \int_{-\infty}^{\infty} \phi^2(t) dt \quad (2.42)$$

where the a is the new a , and the right-hand side is the sum of the parts. Not only that; we also have that Eq. 2.41 will have a Gaussian distribution. It will be the sum of a large number of terms with Gaussian distributions, so the sum of the squares of these converges to a Gaussian distribution. Also, the terms of this type will be independent for orthogonal ϕ 's over the infinite interval, as they were before; and everything that we have done can be carried over to the infinite interval. In other words, there is no virtue in our working from 0 to 1 with ϕ . We could just as well work from $-\infty$ to ∞ .

Now, having got this behind us, let us, either for the finite or the infinite case – it does not matter which – go over to the complex theory. By the way, in the engineering cases that I am going to deal with, the theory that I want is the real theory. In quantum theory the theory that I want is the complex theory, so this is an appropriate time to introduce the complex theory, and I shall do so now. Let me have

$$\phi_1(t) + i\phi_2(t) \quad (2.43)$$

where ϕ_1, ϕ_2 are real and belong to L^2 . This is the most general ϕ which belongs to L^2 in the complex case. Now let me do the following thing. I am going to build

$$\frac{1}{(2)^{1/2}} \left\{ \left(\int_0^1 \phi_1(t) dx(t, a) - \int_0^1 \phi_2(t) dx(t, \beta) \right) + i \left(\int_0^1 \phi_1(t) dx(t, \beta) + \int_0^1 \phi_2(t) dx(t, a) \right) \right\} \quad (2.44)$$

(IX. STATISTICAL COMMUNICATION THEORY)

When I discuss this complex expression, I am discussing nothing but the real and the imaginary parts separately. In order to get the renormalization, because I am dealing with two functions rather than one, I have put a $1/\sqrt{2}$ outside. The $1/\sqrt{2}$ is there merely because we now have two separate functions at right angles which are normalized. I shall define Eq. 2.44 to be

$$\int_0^1 \phi(t) dy(t, \alpha, \beta) \tag{2.45}$$

That is the first step. With very easy computation, it follows that

$$\int d\alpha \int d\beta \left| \int \phi(t) dy(t, \alpha, \beta) \right|^2 = \int \left| \phi(t) \right|^2 dt \tag{2.46}$$

Notice that there's an absolute value in the square. That is, we add the square of the real part and the square of the imaginary part. These integrals may be over either the finite interval or the infinite interval. That is merely a matter of manipulation. Separate the parts, work with them separately, add them, and everything comes out.

This suggests that we do the following thing. We take $x(t, \alpha) + ix(t, \beta)$ (remember that that is $y(t, \alpha, \beta)$) and we map it. I have given you the infinite mapping, and this is only a "two" mapping. We can write α as a binary fraction

$$\alpha = . a_1 a_2 a_3 \dots \tag{2.47}$$

and also

$$\beta = . \beta_1 \beta_2 \beta_3 \dots \tag{2.48}$$

and form

$$\gamma = . a_1 \beta_1 a_2 \beta_2 a_3 \beta_3 \dots \tag{2.49}$$

This mapping gives us the mapping of an $\alpha\beta$ square onto γ from 0 to 1. I now define $y(t, \alpha, \beta)$ as $y(t, \gamma)$, which gives me a complex Brownian motion function. You will notice that at each stage, in going from the finite to the infinite and going from the real to the complex, I have gone through a specific process, and the functions that I get are perfectly definite in each case. Now, I get the result that

$$\int_0^1 d\gamma \left| \int \phi(t) dy(t, \gamma) \right|^2 = \int \left| \phi(t) \right|^2 dt \tag{2.50}$$

over a finite or infinite range.

Now let us consider

$$\int \phi(t) dx(t, \gamma) \tag{2.51}$$

What is the distribution of this expression? Well, it is easy to show that if we take the real and the imaginary parts of this expression, both of these parts have a Gaussian distribution – the same Gaussian distribution but independent (since if we take two different ones everything is independent). In other words, we have now reduced our complex Brownian motion theory, not to the real independent Gaussian distribution, but to the complex one. The fact that a unitary transformation of functions of t generates the measure-preserving transformation of γ over 0 to 1 is still true. But now we are not confined in our unitary transformations to real unitary transformations. That is the first thing we have. Next time I shall go back to the real theory, now that I have approached the complex theory, which I do want for quantum theory. I shall go back to the theory of orthogonal functions defined in terms of $x(t, a)$.

I am going to give you an analysis of functions of a in terms of functions of varying degrees in $x(t, a)$, and this is the appropriate analysis for nonlinear problems in electrical engineering and similar fields, just as harmonic analysis is the analysis for linear problems. Notice that although I have not given the details of the particular order in which I take the subdivisions of the line 0 to 1, or just which particular mapping I take, I could have made them specific in every case; and $x(t, a)$ or $y(t, \gamma)$ are well-defined functions that can be used quite as readily as sine functions and cosine functions in building up spectra; and that is what I propose to do. Then I shall use this theory in studying the spectra of the response of nonlinear oscillators to random inputs, and I have that in a closed form.

B. THE INVARIANCE PROPERTY OF CORRELATION FUNCTIONS UNDER NONLINEAR TRANSFORMATIONS

Further work on the invariance property (1) has yielded the following generalizations and uses of the separable class of processes. (For convenience, zero means are assumed.)

For a stationary separable Markov process of order one, we can show that the normalized covariance function (or correlation coefficient) must have the form

$$\rho(\tau) = \exp(-c|\tau|)$$

where c is a real constant greater than, or equal to, zero. This is another generalization of a result obtained by Doob (2) and later generalized by Barrett and Lampard (3). Furthermore, for this type of process, the second-order autocorrelation function can be expressed in terms of the first-order autocorrelation function as

$$\overline{x(t) x(t + \tau_1) x(t + \tau_2)} = \overline{x^3(t)} \rho(\tau_2) = \overline{x^3(t)} \exp(-c|\tau_2|)$$

if $0 \leq \tau_1 \leq \tau_2$. Thus the location of the intermediate sample is irrelevant.

For certain classes of nonlinear networks (4, 5), the actual computations are difficult to perform; but an approximation of the output of the nonlinear network can be obtained by substituting an optimum linear network for the nonlinear network so that the mean-square difference of the two outputs is minimized. The over-all system might then be analyzed with the linear network present, instead of the nonlinear network. Consider the class of nonlinear networks whose outputs $r(t)$ are related to their inputs $i(t)$ by

$$r(t) = \int_0^{\infty} F[\sigma, i(t-\sigma)] d\sigma \quad (1)$$

where $F(y, x)$ is a function of the two real variables y and x . This class includes, for instance, all linear networks and all nonlinear no-memory devices. We can then show that if the input $i(t)$ is a separable process, the optimum linear network which should replace the nonlinear network is given by

$$h(t) = \frac{1}{2} \int_0^{\infty} x F(t, x) p(x) dx \quad t \geq 0 \quad (2)$$

where $h(t)$ is the optimum impulse response, $p(x)$ is the input first-order probability density function and σ_i^2 is the input variance. For nonseparable processes, an integral equation for $h(t)$ generally needs to be solved; for separable processes, $h(t)$ is determined directly from Eq. 2.

(IX. STATISTICAL COMMUNICATION THEORY)

As a special case of Eq. 1, consider the particular case

$$F(y, x) = \delta(y) F(x)$$

Then

$$r(t) = f[i(t)]$$

which is a nonlinear no-memory operation. Then from Eq. 2, the optimum linear network for a separable input process is

$$h(t) = \delta(t) \frac{\int_{-\infty}^{\infty} x f(x) p(x) dx}{\sigma_i^2}$$

That is, the optimum linear (memory-capable) network has in fact no memory. It is merely an attenuator. Note that this has been proved true only for the separable class of input processes. Thus, in Booton's equivalent gain networks (6), in which a nonlinear no-memory device is replaced by a linear no-memory device so that the mean-square difference of the two outputs is minimized, there is in fact no linear network which performs any better than that simple attenuator.

A. H. Nuttall

References

1. A. H. Nuttall, The invariance property of correlation functions under nonlinear transformations, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1958, p. 68.
2. J. L. Doob, Brownian motion and stochastic equations, Ann. Math. 43, 351 (1942).
3. J. F. Barrett and D. G. Lampard, An expansion for some second-order probability distributions and its applications to noise problems, Trans. IRE, vol. IT-1, no. 1, p. 12 (March 1955).
4. L. Zadeh, A contribution to the theory of nonlinear systems, J. Franklin Inst. 255, 387-408 (May 1953).
5. L. Zadeh, Optimum nonlinear filters, J. Appl. Phys. 24, 396-404 (April 1953).
6. R. C. Booton, Jr., Nonlinear control systems with statistical inputs, Report No. 61, Dynamic Analysis and Control Laboratory, M.I.T., March 1, 1952.

C. UNIQUENESS OF FIRST-ORDER OUTPUT DISTRIBUTIONS OF LINEAR SYSTEMS BY THE MOMENT METHOD

In the case of a non-Gaussian (ergodic) input process into a (time-invariant) linear system, very little can be said, in general, about the probability density functions of the output process. One way to obtain an estimate of the first-order probability density function of the output of a linear system is to evaluate as many moments of the output

process as is practically possible and substitute in a Gram-Charlier series or Edgeworth expansion (1) or in Pearson's system of approximations (2). However, these series approximations are useful only if the first-order output probability density function is uniquely determined by its moments. There is no sense in attempting to form an approximation of the output first-order probability density function by using the first few moments if the whole set of moments themselves are unable to give the unique output probability density function.

A sufficient condition (3) which tells us that the output moments uniquely determine the output probability density function is that the series

$$\sum_{n=0}^{\infty} \frac{(i\xi)^n}{n!} \mu_n$$

converge absolutely for some $|\xi| > 0$, where $\{\mu_n\}$ are the moments of the output process. That is, if we can show that the radius of convergence, R , of the series

$$\sum_{n=0}^{\infty} \frac{\xi^n}{n!} |\mu_n| \tag{1}$$

is greater than zero, we can then be assured that the output moments uniquely determine the output first-order probability density function. In this case, an expansion in some series form for the output first-order probability density function is on safe ground, in that we know that our approximations will approach a unique function.

Now, from Eq. 1, it might appear that we need the complete set of output moments $\{\mu_n\}$ in order to determine whether or not $R > 0$. We shall show, however, that none of the output moments need be computed, but rather, knowledge of the input moments $\{m_n\}$ is sufficient for this determination.

Let $h(t)$ be the impulse response of the linear network (realizable or unrealizable). We can then show that if

$$\int |h(t)| dt < \infty \tag{2}$$

and if the input moments uniquely determine the input first-order probability density function, then the output moments uniquely determine the output first-order probability function. Thus two easy tests must be performed in order to test uniqueness, one on the network, and one on the input moments. These are sufficient conditions and may not be necessary. The assumption that the input moments determine the input density function is equivalent to requiring that the radius of convergence of the series (4)

(IX. STATISTICAL COMMUNICATION THEORY)

$$\sum_{n=0}^{\infty} \frac{|m_n|}{n!} \xi^n$$

be nonzero, or that

$$\limsup_{n \rightarrow \infty} \frac{|m_n|^{1/n}}{(n!)^{1/n}} < \infty \quad (3)$$

Thus we are justified, under the conditions of Eqs. 2 and 3, in using a series approximation for our unknown (unique) output first-order probability density function. Notice that the conditions in Eqs. 2 and 3 involve only the impulse response of the linear network and the first-order statistics of the input process. Whereas the precise determination of the output first-order probability density function requires knowledge of all the higher-order autocorrelation functions of the input process (5), determination of the uniqueness of the output probability density function by the moment method requires only knowledge of the input first-order probability density function (or alternatively, the moments $\{m_n\}$). Once uniqueness of the moment procedure for the input has been ascertained, approximations are in order for the output probability density function, since all orders of input autocorrelation functions will not, in general, be known.

We now proceed with the derivation to show that Eqs. 2 and 3 are sufficient conditions. For Eq. 1, the radius of convergence R is (4)

$$R = \frac{1}{\limsup_{n \rightarrow \infty} \frac{|\mu_n|^{1/n}}{(n!)^{1/n}}}$$

Therefore if R is to be greater than zero

$$\limsup_{n \rightarrow \infty} \frac{|\mu_n|^{1/n}}{(n!)^{1/n}} \quad (4)$$

must be finite. Now

$$\mu_n = \int \dots \int h(\sigma_1) \dots h(\sigma_n) \phi(\sigma_1 - \sigma_2, \dots, \sigma_1 - \sigma_n) d\sigma_1 \dots d\sigma_n$$

where $\phi(\sigma_1, \dots, \sigma_n)$ is the n^{th} -order autocorrelation function of the input process.

Therefore

$$|\mu_n| \leq \int \dots \int |h(\sigma_1)| \dots |h(\sigma_n)| |\phi(\sigma_1 - \sigma_2, \dots, \sigma_1 - \sigma_n)| d\sigma_1 \dots d\sigma_n \quad (5)$$

Now

$$\phi(\sigma_1 - \sigma_2, \dots, \sigma_1 - \sigma_n) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) x(t + \sigma_1 - \sigma_2) \dots x(t + \sigma_1 - \sigma_n) dt$$

where $x(t)$ is a sample function of the input (ergodic) process. Therefore

$$|\phi(\sigma_1 - \sigma_2, \dots, \sigma_1 - \sigma_n)| \leq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)| |x(t + \sigma_1 - \sigma_2)| \dots |x(t + \sigma_1 - \sigma_n)| dt$$

Then, by employing Holder's inequality (6), we can show that

$$|\phi(\sigma_1 - \sigma_2, \dots, \sigma_1 - \sigma_n)| \leq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)|^n dt = \int |x|^n p(x) dx \equiv a_n$$

for all $\sigma_1, \dots, \sigma_n$. We have interchanged time and ensemble averages on the input process under the ergodic hypothesis. Therefore, from Eq. 5,

$$|\mu_n| \leq \int \dots \int |h(\sigma_1)| \dots |h(\sigma_n)| a_n d\sigma_1 \dots d\sigma_n$$

or

$$|\mu_n| \leq a_n \left[\int |h(t)| dt \right]^n \quad (6)$$

for all n . Thus, if $\int |h(t)| dt < \infty$, and if a_n exists, the n^{th} -order output moment exists. Substituting Eq. 6 in Eq. 4, we have

$$\limsup_{n \rightarrow \infty} \frac{|\mu_n|^{1/n}}{(n!)^{1/n}} \leq \limsup_{n \rightarrow \infty} \frac{a_n^{1/n} \int |h(t)| dt}{(n!)^{1/n}}$$

or

$$\limsup_{n \rightarrow \infty} \frac{|\mu_n|^{1/n}}{(n!)^{1/n}} \leq \int |h(t)| dt \limsup_{n \rightarrow \infty} \frac{a_n^{1/n}}{(n!)^{1/n}}$$

Now, if

$$\int |h(t)| dt < \infty \quad (7)$$

and

(IX. STATISTICAL COMMUNICATION THEORY)

$$\limsup_{n \rightarrow \infty} \frac{a_n^{1/n}}{(n!)^{1/n}} < \infty \quad (8)$$

we have

$$\limsup_{n \rightarrow \infty} \frac{|\mu_n|^{1/n}}{(n!)^{1/n}} < \infty$$

and the output first-order probability density function is uniquely determined by its moments. Thus Eq. 7 is true from our assumptions. But we have yet to show that Eq. 8 follows from Eq. 3. That is done in the following way. Consider the sequence of terms $\left\{ \frac{a_n^{1/n}}{(n!)^{1/n}} \right\}$ in Eq. 8:

$$a_1, \quad \frac{a_2^{1/2}}{(2!)^{1/2}}, \quad \frac{a_3^{1/3}}{(3!)^{1/3}}, \quad \frac{a_4^{1/4}}{(4!)^{1/4}}, \quad \dots$$

which we split into two subsequences:

$$a_1, \quad \frac{a_3^{1/3}}{(3!)^{1/3}}, \quad \frac{a_5^{1/5}}{(5!)^{1/5}}, \quad \dots$$

and

$$\frac{a_2^{1/2}}{(2!)^{1/2}}, \quad \frac{a_4^{1/4}}{(4!)^{1/4}}, \quad \frac{a_6^{1/6}}{(6!)^{1/6}}, \quad \dots$$

For the second subsequence,

$$\limsup_{n \rightarrow \infty} \frac{a_{2n}^{1/2n}}{[(2n)!]^{1/2n}} = \limsup_{n \rightarrow \infty} \frac{m_{2n}^{1/2n}}{[(2n)!]^{1/2n}} < \infty \quad (9)$$

since the terms involved are a subsequence of the terms in Eq. 3. For the first subsequence, (3),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{a_{2n+1}^{1/2n+1}}{[(2n+1)!]^{1/2n+1}} &\leq \limsup_{n \rightarrow \infty} \frac{a_{2n+2}^{1/2n+2}}{[(2n+1)!]^{1/2n+1}} \\ &= \limsup_{n \rightarrow \infty} \frac{m_{2n+2}^{1/2n+2}}{[(2n+1)!]^{1/2n+1}} = \limsup_{n \rightarrow \infty} \frac{m_{2n+2}^{1/2n+2}}{[(2n+2)!]^{1/2n+2}} \cdot \frac{[(2n+2)!]^{1/2n+2}}{[(2n+1)!]^{1/2n+1}} \\ &\leq \limsup_{n \rightarrow \infty} \frac{m_{2n}^{1/2n}}{[(2n)!]^{1/2n}} \cdot \limsup_{n \rightarrow \infty} \frac{[(2n+1)!]^{1/2n+1}}{[(2n)!]^{1/2n}} = a \cdot \beta \end{aligned}$$

if both are finite. Now, $a < \infty$ by Eq. 9, and $\beta = 1$. Therefore

$$\limsup_{n \rightarrow \infty} \frac{a^{1/2n+1}}{[(2n+1)!]^{1/2n+1}} < \infty \quad (10)$$

If we combine Eqs. 9 and 10, we have

$$\limsup_{n \rightarrow \infty} \frac{a^{1/n}}{(n!)^{1/n}} < \infty$$

Thus we have proved that if the linear network weighting function is absolutely integrable, and if the input moments uniquely determine the input first-order probability density function, then the output moments uniquely determine the output first-order probability density function. Of course, this does not give a method for finding an approximation in any given problem but, rather, indicates the cases in which an approximation is worth while. Also, any comment to the effect that computation of the output moments is useless, because of nonuniqueness of the probability density function so determined, can be easily dismissed if the conditions of the present paper are satisfied.

A. H. Nuttall

References

1. H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1951), p. 221.
2. W. P. Elderton, *Frequency Curves and Correlation* (Harren Press, Washington, 4th ed., 1953).
3. H. Cramér, *op. cit.*, p. 176.
4. W. Rudin, *Principles of Mathematical Analysis* (McGraw-Hill Book Company, Inc., New York, 1953), p. 53.
5. J. Y. Hayase, *Quarterly Progress Report*, Research Laboratory of Electronics, M.I.T., Oct. 15, 1956, p. 53.
6. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, London, 1934), p. 139.

D. STATISTICAL ERRORS IN MEASUREMENTS ON AMPLITUDE-DISTORTED GAUSSIAN SIGNALS

Bussgang (1) has suggested that the design of Gaussian correlators might be simplified by deliberately amplitude-distorting one of the input signals, as shown in Figure IX-12. The distortion is performed by the nonlinear no-memory device $g(x)$. The desired input crosscorrelation function is defined as

(IX. STATISTICAL COMMUNICATION THEORY)

$$\phi_{12}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2; \tau) dx_1 dx_2 \quad (1)$$

where $p(x_1, x_2; \tau)$ is the joint Gaussian probability density function of the system inputs. Bussgang (1) also demonstrated that the output crosscorrelation function

$$\psi_{12}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 g(x_2) p(x_1, x_2; \tau) dx_1 dx_2 \quad (2)$$

is identical with the input crosscorrelation function, $\phi_{12}(\tau)$, except for a constant scale factor that depends only upon $g(x)$ and upon the first-order probability density distribution of the inputs, $p(x)$. The objective is to choose $g(x)$ in such a way that a correlator

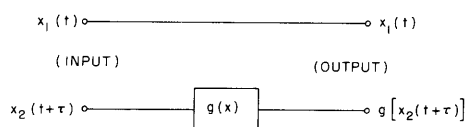


Fig. IX-12. Amplitude distortion of one correlator input signal.

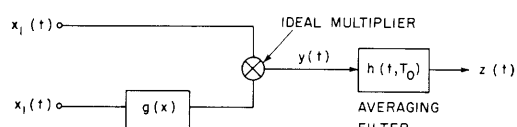


Fig. IX-13. Correlator block diagram.

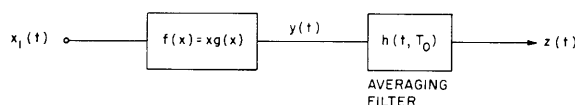


Fig. IX-14. Equivalent block diagram for correlator in Fig. IX-13.

designed to measure $\psi_{12}(\tau)$ would be simpler than that required to measure $\phi_{12}(\tau)$ directly.

The purpose of this analysis is to determine the effect of the distortion $g(x)$ upon the correlator measurement errors. The analysis is restricted to the case of identical Gaussian inputs, $[x_1(t) = x_2(t)]$, with zero means, and to continuous correlators. Thus the desired function is the input autocorrelation function

$$\phi_{11}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2; \tau) dx_1 dx_2 \quad (3)$$

We consider, first, the measurement of $\phi_{11}(0)$, and, then, extend the results to the measurement of $\phi_{11}(\tau)$.

The correlator block diagram is shown in Fig. IX-13. The averaging filter has the impulse response $h(t)$, but the averaging time is restricted to the duration T_0 , so that

the effective filter time response is

$$h(t, T_o) = \begin{cases} h(t) & \text{for } 0 < t < T_o \\ 0 & \text{for } T_o < t \end{cases} \quad (4)$$

The system function of this filter is

$$H(\omega, T_o) = \int_0^{\infty} h(t, T_o) e^{-j\omega t} dt \quad (5)$$

An equivalent representation of the system in Fig. IX-13 is shown in Fig. IX-14, in which $f(x)$ is the equivalent nonlinear function that results from the elimination of the multiplier.

We consider the ideal limiter given by

$$g(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ -1 & \text{for } x < 0 \end{cases} \quad (6)$$

The corresponding equivalent function is

$$f(x) = \begin{cases} x & \text{for } x \geq 0 \\ -x & \text{for } x < 0 \end{cases} \quad (7)$$

which is the full-wave linear detector characteristic. We define the measurement noise-to-signal ratios $\left(\frac{N}{S}\right)_o$ and $\left(\frac{N}{S}\right)_i$ as

$$\left(\frac{N}{S}\right)_o = \frac{\overline{z_o^2(t)} - \overline{z_o(t)}^2}{\overline{z_o(t)}^2} \quad (8)$$

and

$$\left(\frac{N}{S}\right)_i = \frac{\overline{z_i^2(t)} - \overline{z_i(t)}^2}{\overline{z_i(t)}^2} \quad (9)$$

The subscript "o" applies to the case of no distortion, i. e., $g(x) = x$; and the subscript "i" applies if $g(x)$ is the particular distortion function given in Eq. 6.

Davenport (2) has shown that Eq. 8 can be reduced to

$$\left(\frac{N}{S}\right)_o = \frac{4 \int_0^{T_o} \rho^2(\tau) R_h(\tau) d\tau}{\left(\sigma_x^2 H_f\right)^2} \quad (10)$$

(IX. STATISTICAL COMMUNICATION THEORY)

where the following definitions apply:

$$\sigma_x^2 = \overline{x_1^2(t)} - \overline{x_1(t)}^2 \quad (11)$$

$$\rho(\tau) = \frac{\phi_{11}(\tau)}{\sigma_x^2} \quad (12)$$

$$R_h(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega, T_0)|^2 e^{j\omega\tau} d\omega \quad (13)$$

$$H_f = \int_0^{T_0} h(t) dt \quad (14)$$

The determination of $\left(\frac{N}{S}\right)_i$ is simplified by the use of the series expansion of the second-order probability distribution of $y_i(t)$ given in reference 3,

$$p(y_1, y_2; \tau) = \frac{2}{\pi\sigma_x^2} \exp\left[-\frac{y_1^2 + y_2^2}{2\sigma_x^2}\right] \sum_{n=0}^{\infty} [\rho(\tau)]^{2n} \frac{H_{2n}(y_1/\sigma_x) H_{2n}(y_2/\sigma_x)}{(2n)!} \quad (15)$$

for $y_1, y_2 \geq 0$; otherwise, $p(y_1, y_2; \tau) = 0$. Equation 15 can be written as

$$p(y_1, y_2; \tau) = p(y_1) p(y_2) \sum_{n=0}^{\infty} \frac{[\rho(\tau)]^{2n} H_{2n}(y_1/\sigma_x) H_{2n}(y_2/\sigma_x)}{(2n)!} \quad (16)$$

for $y_1, y_2 \geq 0$; otherwise, $p(y_1, y_2; \tau) = 0$. To find the autocorrelation function of $y_i(t)$, which we define as $R_y(\tau)$, we substitute Eq. 16 in

$$R_y(\tau) = \int_0^{\infty} \int_0^{\infty} y_1 y_2 p(y_1, y_2; \tau) dy_1 dy_2 \quad (17)$$

and, after interchanging summation and integration, we have

$$R_y(\tau) = \sum_{n=0}^{\infty} [\rho(\tau)]^{2n} \int_0^{\infty} \int_0^{\infty} y_1 H_{2n}(y_1/\sigma_x) p(y_1) y_2 H_{2n}(y_2/\sigma_x) p(y_2) dy_1 dy_2 \quad (18)$$

or

$$R_y(\tau) = \sum_{n=0}^{\infty} [\rho(\tau)]^{2n} \left[\int_0^{\infty} y H_{2n}(y/\sigma_x) p(y) dy \right]^2 \quad (19)$$

If we define the quantity D_n as

$$D_n = \int_0^{\infty} y H_n(y/\sigma_x) p(y) dy \quad (20)$$

then Eq. 19 may be written as

$$R_y(\tau) = \sum_{n=0}^{\infty} [\rho(\tau)]^{2n} D_{2n}^2 \quad (21)$$

The determination of numerical values for the coefficients D_{2n}^2 becomes increasingly difficult as n increases. The first few values yield

$$R_y(\tau) = \frac{\sigma_x^2}{2\pi} \left[4 + 2\rho^2(\tau) + \frac{1}{6}\rho^4(\tau) + \frac{1}{20}\rho^6(\tau) + \dots \right] \quad (22)$$

To determine the effect of neglecting all terms after the fourth, we find the value for the infinite sum of coefficients. We do this by computing the exact value of the variance, σ_y^2 , of $y_i(t)$, directly from the first-order probability density distribution, $p(y)$, and find that

$$\sigma_y^2 = 0.3634 \sigma_x^2 \quad (23)$$

From the knowledge that

$$\sigma_y^2 = R_y(0) - \overline{y_i(t)^2} \quad (24)$$

and that

$$\overline{y_i(t)^2} = \frac{2}{\pi} \sigma_x^2 \quad (25)$$

we find that we can write

$$\sum_{n=1}^{\infty} D_{2n}^2 = \sigma_y^2 = 0.3634 \sigma_x^2 \quad (26)$$

But

$$D_2^2 + D_4^2 + D_6^2 = 0.3526 \sigma_x^2 \quad (27)$$

so that we can make the approximation

$$R_y(\tau) \approx \frac{\sigma_x^2}{2\pi} \left[4 + 2\rho^2(\tau) + \frac{1}{6}\rho^4(\tau) + \frac{1}{20}\rho^6(\tau) \right] \quad (28)$$

and the error involved will be no more than 3 per cent. By using this result, we find that Eq. 9 yields

(IX. STATISTICAL COMMUNICATION THEORY)

$$\left(\frac{N}{S}\right)_i \approx \frac{\frac{\sigma_x^2}{\pi} \int_0^{T_0} \left[2\rho^2(\tau) + \frac{1}{6}\rho^4(\tau) + \frac{1}{20}\rho^6(\tau) \right] d\tau}{\frac{2}{\pi} \sigma_x^2 H_f^2} \quad (29)$$

If we define

$$b_n = \int_0^{T_0} \rho^n(\tau) R_h(\tau) d\tau \quad (30)$$

then, for $\sigma_x^2 = 1$, Eqs. 10 and 29 become

$$\left(\frac{N}{S}\right)_0 = \frac{4b_2}{H_f^2} \quad (31)$$

and

$$\left(\frac{N}{S}\right)_i \approx \frac{b_2 + \frac{1}{12} b_4 + \frac{1}{40} b_6}{H_f^2} \quad (32)$$

The relative noise-to-signal ratio, R , is now defined as

$$R = \frac{(N/S)_i}{(N/S)_0} \approx \frac{b_2 + \frac{1}{12} b_4 + \frac{1}{40} b_6}{4b_2} \quad (33)$$

or

$$R \approx \frac{1}{4} + \frac{1}{48} \frac{b_4}{b_2} + \frac{1}{160} \frac{b_6}{b_2} \quad (34)$$

For $n > 2$, the quantity b_n/b_2 decreases as n increases. If we assume that the bandwidth of the averaging filter is small compared with the bandwidth of $x_1(t)$, and that

$$\rho(\tau) = e^{-a|\tau|} \quad (35)$$

then we can make the approximation

$$\frac{b_n}{b_2} \approx \frac{2}{n} \quad (36)$$

with little effect on the accuracy of R , and finally obtain

$$R \approx 0.265 \quad (37)$$

Viewed as a measurement technique, the measured quantity $\overline{z_1(t)}$ is a function of the variance, σ_x^2 , which we are seeking. Specifically,

$$\overline{z_i(t)} = \sqrt{\frac{2}{\pi}} \sigma_x \quad (38)$$

so that we are measuring the square root of the variance. An error of ϵ per cent in this measurement is equivalent to an error of 2ϵ per cent in σ_x^2 . This means that the power ratio, Eq. 37, must be corrected by multiplication by a factor of 4 to give an effective ratio, R_{eff} , of

$$R_{\text{eff}} \approx 1.06 \quad (39)$$

We can interpret Eq. 39 as indicating that the measurement signal-to-noise ratio is degraded by approximately 0.24 db by the presence of the ideal limiter, subject to the imposed assumptions.

This same general approach can be applied to the determination of R_{eff} for other distortion functions $f(x)$. A summary of the results obtained for certain power-law devices is given in Table IX-1, in which k is the power law of $f(x)$, as defined in the equation

$$f(x) = |x|^k \quad (40)$$

The results given in Table IX-1 are plotted in Fig. IX-15. This plot shows that all of the nonlinear devices examined produce a degradation of the measurement signal-to-noise ratio, as compared with the linear case represented by $g(x) = x$. It has not been shown analytically that R_{eff} has a minimum value for $k = 2$, yet it seems intuitively plausible that this should be so.

Faran and Hills (4) have analyzed full-wave power-law devices in the detection of Gaussian signals in Gaussian noise. Although their objectives and methods of analysis are different from those presented here, certain of their results are in close agreement

Table IX-1. Relative Noise-To-Signal Ratio R_{eff} versus Power Law k of Distortion Function $f(x)$.

k	R_{eff}	$R_{\text{eff}} - \text{DB}$
1/2	1.13	0.52
1	1.06	0.24
2	1.00	0
3	1.05	0.22
4	1.17	0.65
5	1.37	1.37
6	1.6	2.03

(IX. STATISTICAL COMMUNICATION THEORY)

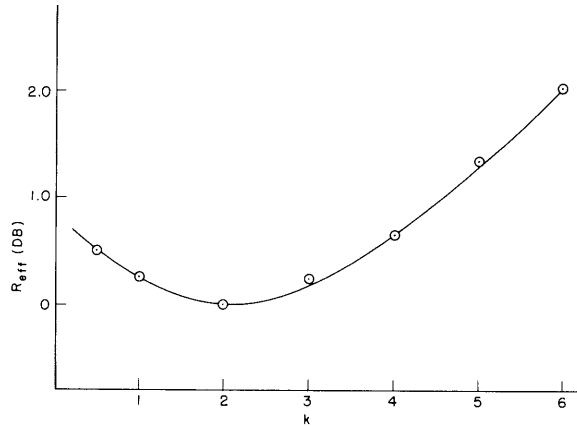


Fig. IX-15. Relative noise-to-signal ratio, R_{eff} versus power-law k , of distortion function $f(x)$.

with those obtained in this analysis. (Compare Fig. IX-15 of this paper with Fig. 4.1 of reference 4.)

The evaluation of the measurement errors in $\phi_{11}(\tau)$ for $\tau \neq 0$ requires the knowledge of fourth- and higher-order probability density distributions of $x_1(t)$. These higher-order statistics are required to compute the autocorrelation function of $y(t)$. Results have been obtained for the two cases,

$$g(x) = x \quad (41)$$

and

$$g(x) = x^2 \quad (42)$$

for an exponential correlation function, as given in Eq. 35. These results indicate that, for $g(x) = x$, the variance of the error fluctuation, σ_z^2 , remains constant as a function of the delay τ , but, for $g(x) = x^2$, σ_z^2 has the form

$$\sigma_z^2 = K(5 + 3 e^{-a|\tau|}) \quad (43)$$

which shows a decrease relative to $\tau = 0$. We cannot say that the same result would hold true for the other distortion functions given in Table IX-1, but at least we would not expect σ_z^2 to change by an order of magnitude as a function of τ .

If we seek the crosscorrelation function of $s_1(t) + x_1(t)$ with $x_1(t)$ (where $s_1(t)$ is a Gaussian signal in the Gaussian noise, $x_1(t)$), it can be shown that, under assumptions similar to those previously stated, if $g(x)$ is an ideal limiter, then the value of R_{eff} is approximately the same as in Eq. 37 for small input signal-to-noise ratios.

Although this analysis was restricted to Gaussian inputs, it should be pointed out

(IX. STATISTICAL COMMUNICATION THEORY)

that the invariance of correlation functions under nonlinear transformation holds for a larger class of inputs (5).

D. L. Haas

References

1. J. J. Busgang, Crosscorrelation functions of amplitude-distorted Gaussian signals, Technical Report 216, Research Laboratory of Electronics, M.I.T., March 26, 1952.
2. W. B. Davenport, Jr., Correlator errors due to finite observation intervals, Technical Report 191, Research Laboratory of Electronics, M.I.T., March 8, 1951.
3. D. L. Haas, Expansions for some second-order probability distributions, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1958.
4. J. J. Faran, Jr. and R. Hills, Jr., Correlators for signal reception, TM-27, Acoustics Research Laboratory, Harvard University, Sept. 15, 1952.
5. A. H. Nuttall, Invariance of correlation functions under nonlinear transformations, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., April 15, 1957, p. 69.